Hybrid Cellular Automaton for Field Cancerization

by

Karl Deutscher

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science

in

Applied Mathematics

Department of Mathematical and Statistical Sciences

University of Alberta

Abstract

Field Cancerization is a hypothesis for the formation of cancer in certain types of tissues. It proposes the idea that a tumour can form in a "field" of cells that are predetermined for the development of cancer. Further, it is hypothesized that these fields are mainly caused by the onslaught of carcinogens on the tissue. Lastly, field cancerization proposes that tumour recurrence is related to the tumour being excised without fully removing the surrounding field.

The model we propose is a hybrid cellular automaton (CA) used to verify the previously stated propositions and determine how long cancer development will take. The CA is considered to be hybrid due to its' rule depending on the results of partial differential equations (PDEs) and a multi-layer perceptron (MLP). The PDEs represent the spread of carcinogens in the tissue, while the MLP computes the effects of the carcinogens on the gene expression of the genes related to cancer development in the tissue under consideration.

We apply the model to field cancerization of the tongue. Most of the parameters of the model were chosen and are not based upon real data, as the necessary data was not available. This includes the choice of substituting nicotine, which is a mutagen but not currently listed as a carcinogen, to represent the carcinogen impacts of smoking tobacco. According to Health Canada tobacco contains over 4,000 chemicals, of which more than 70 are carcinogens [25]. Many researchers are investigating how nicotine contributes to the development of cancer due to its use in non-tobacco products such as e-cigarettes and nicotine patches. One such study by Sanner & Grimsrud [121]

suggests that nicotine has several cancer-causing effects including speeding up cell growth, it is poisonous to cells, it kick-starts a process that is an important step in the path toward malignant cell growth, and it decreases the tumour suppressor CHK2. Therefore, we considered the readily available data with regards to nicotine as an appropriate choice to substitute for the over 70 carcinogens in tobacco. The other carcinogen considered in this thesis is ethanol to represent alcohol consumption. It was found that nicotine was a more potent carcinogen than ethanol. The combined impact of both ethanol and nicotine resulted in more aggressive cancer growth. It was verified that removing the field results in recurrence taking longer to occur than if the field is not removed. We also tracked cell lineages and found that as the field develops, the number of distinct cell lines decreases. Finally, we found that most tumour masses formed via polyclonal origin instead of monoclonal origin, though both occur within the simulations.

To my wife (Lidna) and daughter (Awrya)
For their loving support during my research and thesis writing.

Acknowledgements

Thanks to my supervisors Thomas Hillen and Jay Newby. I also give thanks to Ion Bica who first led me into research.

Contents

1 Introduction **1**
- 1.1 Biological Introduction . . . 1
  - 1.1.1 Carcinogenesis . . . 1
  - 1.1.2 Cancer Stem Cells . . . 4
  - 1.1.3 Field Cancerization . . . 5
- 1.2 Mathematical Literature Review . . . 11
  - 1.2.1 Review of Foo et al. . . . 11
  - 1.2.2 Review of Ryser et al. . . . 17
  - 1.2.3 Review of Gerlee et al. . . . 19

2 Model **22**
- 2.1 Model Overview . . . 22
- 2.2 Carcinogen Partial Differential Equations . . . 23
- 2.3 Introduction to Neural Networks . . . 35
  - 2.3.1 History . . . 35
  - 2.3.2 Basic ANN Architectures . . . 36
  - 2.3.3 Multi-Layer Perceptron . . . 37
  - 2.3.4 Designing an Artificial Neural Network . . . 39
- 2.4 Gene Expression Neural Network . . . 40
- 2.5 Cellular Automaton . . . 42
  - 2.5.1 Cell Mutation . . . 44
  - 2.5.2 Update Rules for Cell Class . . . 57
  - 2.5.3 Dedifferentiation . . . 58
  - 2.5.4 Cell Fitness . . . 59
  - 2.5.5 CA Rule . . . 61
  - 2.5.6 Tumour Excision . . . 63
  - 2.5.7 Lineage Tracking . . . 63
  - 2.5.8 Field Definition . . . 64
  - 2.5.9 CA Model Timeline . . . 64

3 Application and Implementation **65**
- 3.1 Application . . . 65
  - 3.1.1 Introduction to Head and Neck Squamous Cell Carcinoma 65
- 3.2 Application . . . 67
  - 3.2.1 Carcinogen Parameters . . . 68
  - 3.2.2 Gene Expression Neural Network Parameters . . . 70
  - 3.2.3 CA Parameters . . . 72
- 3.3 Model Implementation . . . 76
  - 3.3.1 Parallel Implementation . . . 76

4 Results **78**
4.1 Equilibrium 78
4.2 General Observations 80
4.2.1 Field Development 82
4.2.2 Tumour Growth Rate 83
4.2.3 Mutational Evolution 84
4.2.4 Phenotypic Evolution 86
4.3 Grid Size Comparisons 87
4.4 Carcinogen Concentration Spatial Distribution 92
4.4.1 Single Carcinogen 99
4.4.2 All Carcinogens 101
4.4.3 Cyclic Carcinogenic Onslaught 102
4.5 Tumour Excision 105
4.6 Cell Lineages 107
4.6.1 Monoclonal versus Polyclonal Origin 110

5 Conclusion **112**
5.1 Discussion of Results 112
5.2 Future Work 113

References **117**

List of Tables

2.1 CA cell classes 43

3.1 Provides the following properties of each gene considered in the model: index for the gene used in the various matrices and vectors required in the model, name, type of the gene, direction the gene must be regulated to become positively mutated, and how phenotypic actions are modified when the gene is positively mutated. 73
3.2 Shows which genes are activated by certain genes. 74

4.1 In this table we compare the time-step the first mutated cell forms and the cell class that cell belongs to between the different grid sizes. 89
4.2 In this table we compare the time-step the first CSC forms and the number of time-steps since the first mutated cell formed between the various grid sizes. 89
4.3 In this table we compare the time-step the first TC forms and the elapsed amount of time it takes a CSC to form the first TC between the various grid sizes. 90
4.4 This table compares the total number of time-steps, number of time-steps since the first TC formed, and the percentage the tumour mass(es) make up of the domain at the end of the simulation between all the grid sizes. 91
4.5 In this table we compare the number of tumour cell lineages at the end of the simulation between each of the grid sizes. . . . 91

List of Figures

2.1 In this figure we show a visual representation of a linear threshold unit (LTU). Firstly, a weighted sum of the inputs is computed, followed by an application of a step function. This figure is adapted from Géron [55]. 36
2.2 In this figure we show a visual representation of a perceptron. Inputs are passed through an input layer (with an added bias neuron), the output of each of the neurons of the input layer are then inputted into each of the LTUs that make up the output layer. This figure is adapted from Géron [55]. 37
2.3 This figure shows a simple Multi-Layer Perceptron with one hidden layer. It is adapted from Géron [55]. 38

4.1 This figure includes three time-steps from a simulation with parameters set at: grid size 128x128 and no active carcinogens. Using the colour map for the cell classes as provided in Table 2.1. These show (a) the initial seed of the simulation, in (b) the domain (tissue) at the halfway point of the simulation, and in (c) the final time-step. 79
4.2 This figure shows the time course of the fraction of cells in the different cell classes NTC, MNTC, NSC, MNSC, CSC, TC, and empty. The parameters of the simulation were as follows: grid size was 128x128 and no carcinogens were active. 79
4.3 This figure shows the time course of the average gene expression for the ten genes we consider. The parameters of the simulation were as follows: grid size was 128x128 and no carcinogens were active. 80
4.4 This figure includes time-steps illustrating the development of a cancer field and cancer cells using the colour map for the cell classes as provided in Table 2.1. These show (a) the initial seed, (b) the cancer field at its early development, (c) the cancer field further developing but prior to cancer, (d) the first stages of cancer development, (e) further cancer growth, and (f) the final time-step. Parameters are as follows: grid size 256x256, carcinogen spatial distribution 2, both carcinogens activated. . 81
4.5 This figure shows the time course of the fraction of mutated cells, thus illustrating the cancer field growth over time. Parameters are as follows: grid size 256x256, carcinogen spatial distribution 2, both carcinogens activated. 82
4.6 This figure shows the time course of the fraction of CSC and TC, thus illustrating the tumour growth over time. Parameters are as follows: grid size of 256x256, carcinogen spatial distribution 2, both carcinogens activated. 83

4.7 In this figure we show the mutational evolution of the genes. The time course of the average gene expression are shown for in (a) the tumour suppressor genes, in (b) the oncogenes. In (c) we show the time course of the fraction of genes that are positively mutated. Parameters were chosen as follows: grid size of 256x256, carcinogen spatial distribution 2 was used, both carcinogens were activated. 85
4.8 This figure illustrates the phenotypic evolution of proliferation, apoptosis, quiescence, and differentiation. In these we show (a) the time course of the fraction of cells that underwent each phenotypic action and (b) the time course of the average probability for each phenotypic action. Parameters are as follows: grid size 256x256, carcinogen spatial distribution 2, both carcinogens activated. 87
4.9 This figure shows a time-step from each of the grid sizes that were considered for comparison. In (a) the grid size is 64x64, in (b) the grid size is 128x128, in (c) the grid size is 256x256, and in (d) the grid size is 512x512. Parameters: both carcinogens were activated and carcinogen spatial distribution 2 was used. 88
4.10 This figure shows a visual representation of each of the carcinogen spatial distributions within a 256x256 domain. In the figures we show (a) carcinogen spatial distribution 1 (CSD1), (b) CSD2, and (c) CSD3. 92
4.11 In figures (a),(b),(c) we show the time course of the fraction of cells in the different cell classes NTC, MNTC, NSC, MNSC, CSC, TC, empty. In (a) we consider carcinogen spatial distribution 1 (CSD1), in (b) CSD2, and in (c) CSD3. Parameters are as follows: grid size 256x256 and both carcinogens activated. 94
4.12 In figures (a),(b),(c) we show the time course of the average gene expression for each of the ten genes. In the plots we consider (a) carcinogen spatial distribution 1 (CSD1), (b) CSD2, and (c) CSD3. Parameters are as follows: grid size 256x256 and both carcinogens activated. 95
4.13 In figures (a),(b),(c) we show the time course of the fraction of cells that underwent each of the phenotypic actions, namely proliferation, apoptosis, quiescence (moved), and differentiation. In the plots we consider (a) carcinogen spatial distribution 1 (CSD1), (b) CSD2, and (c) CSD3. Parameters are as follows: grid size 256x256 and both carcinogens activated. 97
4.14 In this figure we compare various characteristics of the cancer field and cancer development between the carcinogen spatial distributions (CSD). In the plots we show (a) the time course of the fraction of positively mutated genes, (b) the time course of the average cell fitness, (c) the time course of the log of the number of cell lineages, and (d) the time course of the fraction of cells that are part of the tumour mass(es). Parameters are as follows: grid size 256x256 and both carcinogens activated. . 99
4.15 In figures (a),(b) we show the time course of the fraction of cells in the different cell classes NTC, MNTC, NSC, MNSC, CSC, TC, empty. In the plots we consider (a) ethanol and (b) nicotine. Parameters are as follows: grid size 256x256 and carcinogen spatial distribution 2. 100

4.16 In figures (a),(b) we show the time course of the average gene expression for each of the ten genes. In the plots we consider (a) ethanol and (b) nicotine. Parameters are as follows: grid size 256x256 and carcinogen spatial distribution 2. 101
4.17 In figures (a),(b),(c) we show the time course of the fraction of cells in the different cell classes NTC, MNTC, NSC, MNSC, CSC, TC, empty. In the plots we show the case of (a) smoking every day and drinking on weekends, (b) smoking on weekends, and (c) smoking on weekdays. Parameters are as follows: grid size 256x256 and carcinogen spatial distribution 2. 103
4.18 In figures (a),(b) we show the time course of the fraction of cells in the different cell classes NTC, MNTC, NSC, MNSC, CSC, TC, empty. In the plots we consider the case where we (a) remove only TCs (keeping the field) and (b) we remove all mutated cells (removing the field). Parameters are as follows: grid size 256x256, both carcinogens activated, carcinogen spatial distribution 2, and time elapse of excision following first TC appearance was 18 months. 106
4.19 In figures (a),(b) we show how long it takes for cancer to recur after excision has been completed on the tumour cells that were alive for a given number of months. In the plots we consider the case where (a) we remove only TCs (keeping the field) and (b) we remove all mutated cells (removing the field). Parameters are as follows: grid size 256x256, both carcinogens activated, carcinogen spatial distribution 2, and time elapse of excision following first TC appearance was 18 months. 107
4.20 Important time-steps that show the top twenty lineages throughout the development stages of field cancerization. In the figures we show (a) the initial seed, (b) early cancer field formation, (c) later cancer field development, (d)-(f) cancer development. Note that light grey means the cell is not in any of the top twenty lineages and each colour represents a different cell lineage. Parameters are as follows: grid size 256x256, carcinogen spatial distribution 2, both carcinogens activated. 108
4.21 Each colour represents a different tumour cell lineage. Both the light green and light blue lineages have a main mass of tumour cells with other small masses disconnected from it, which implies they both have formed tumour masses by monoclonal origin. However, most of the tumour masses have been formed via polyclonal origin. Parameters are as follows: grid size 256x256, carcinogen spatial distribution 2, and both carcinogens activated. 110

Chapter 1

Introduction

1.1 Biological Introduction

Cancer initialization (carcinogenesis) is an important topic of research due to its applications in early detection of tumours and treatment strategies. There exists many different hypothesis on the matter but we shall discuss just one, namely, field cancerization. However, before delving into field cancerization the concepts of carcinogenesis and cancer stem cells (CSCs) should be introduced.

1.1.1 Carcinogenesis

Most carcinogenesis models consider that cancer is initialized from the result of a multi-step process. A normal cell does not become a cancer cell until multiple genetic alterations accumulate within it. The number of genetic alterations in a cancer cell is an indicator of the level of malignancy of the cell.

Gatenby & Gillies [51] found six micro-environmental barriers for a malignant phenotype: apoptosis with loss of basement membrane contact, inadequate growth promotion, senescence (deterioration of a cells' power of division and growth with age), hypoxia (deficiency in the amount of oxygen reaching the tissues), acidosis (excessively acidic condition of the body fluids or tissues), and ischaemia (restriction of blood supply to tissues, causing hypoxia). The development of cancer occurs when a normal cell overcomes at least one of these barriers. Thus, the micro-environment is an important factor to consider in cancer initialization.

A normal cell lineage can acquire epimutations, termed mutations in Cur-

tius *et al.* [31], that are positively selected in the micro-environment of a healthy organ. Not only do carcinogens cause mutations, the natural aging process of tissue can as well, because mutations accrue in such tissues [15]. Considering that all mutant cells are cancerized, the entire body becomes increasingly cancerized as it grows older [31]. A driver mutation is one that confers growth or survival advantages for tumour cells within the appropriate micro-environment [22, 61, 139]. A passenger (neutral) mutation is one that passively accumulates in cell lineages [22, 61, 139]. It may be that some driver mutations are not currently affecting cancer growth but instead had previously driven the growth of an ancestral lineage [31]. Progression to cancer usually requires the accumulation of multiple driver mutations [157]. A mutant lineage/clone, can grow to produce large patches, or fields, of cells that are predisposed to eventually progress to neoplasm.

The stromal micro-environment is a key regulator of self-renewal in the epithelium [33]. Epithelium is one of the four basic types of animal tissue, along with connective, muscle and nervous tissue [42]. Epithelial tissues line the outer surfaces of organs and blood vessels throughout the body, as well as the inner surfaces of cavities in many internal organs [42]. Epithelial tissue is organized into two different types: glandular and squamous [42]. A gland is one or more cells that produce and secrete a specific product [42]. Glandular tissue has two types: exocrine (secrete their products (enzymes, mucus, milk, *etc.*) into ducts that lead directly to the external environment), endocrine glands (secrete their products (hormones) directly into the bloodstream) [56]. Exocrine glandular epithelium tissues include salivary glands, the esophagus, gastric glands, intestinal glands, pancreatic glands, mammary glands, and sweat glands [56]. Examples of endocrine glandular tissues include pituitary glands, thyroid glands, parathyroid glands, adrenal glands, pancreas, gonads, and pineal glands [56]. Squamous epithelium tissue is a single layer of flat cells in contact with the basal lamina of the epithelium [131]. It is often permeable and occurs where small molecules need to pass quickly through membranes via filtration or diffusion [131]. Some examples of squamous tissue include the skin, walls of capillaries, linings of the pericardial, pleural cavities,

peritoneal cavities, and linings of the alveoli of the lungs [131]. The tissue architecture constrains evolution by limiting the ability of mutant clones to expand [91], so understanding the differences between mutant clone expansions in glandular and squamous tissue is important. In glandular epithelium a mutant clone undergoes niche successions, where a mutant stem cell (SC) in the gland replaces all other SCs; after which the mutant glands produces a field of mutant glands by gland fission [7, 60, 95, 103]. Epithelial cells within a squamous tissue expand by basal replacement of neighbouring SCs [1, 30, 36, 79]. During normal homeostasis basal cells proliferate to produce differentiated progeny that then form the superficial layers of the epithelium [145]. Any new basal cells compete to grow a patch by lateral replacement of one progenitor cell by another [145].

Phenotypic change is caused by either individual high-impact mutations or epistasis among a group of mutations that require each other [31]. As well, large-scale mutational events that simultaneously alter expansive parts of the genome will elicit phenotypic change [84, 135]. Micro-environmental factors provide selective pressures for phenotype adaptation in which cells explore the adaptive landscape (via genetic mutations or phenotypic plasticity), and genotypes of lineages reflect the phenotype that survived [31]. Curtius *et al.* [31] consider only phenotypes that are cancer-related when discussing cancer development and initiation, thus they exclude regions of tissue having DNA damage or passenger mutations and oncogene or tumour suppressor gene mutations that are not currently active. It is important to note that cancerized phenotypes may be subtle and/or lasting only for a short time [31], making them hard to detect. Phenotypic consequences of a driver mutation may be context dependent, so the existence of a driver mutation may not be sufficient to cause a phenotypic change in the current micro-environmental condition [31]. Due to the fact that crypt fission is the main mechanism in glandular tissue development, a mutation that increases upregulation of crypt fission is a cause for carcinoma development or initialization [31]. Mutations that help cells adhere to the basal membrane influence carcinoma development in squamous tissue because they will not allow those cells to migrate and differ-

entiate, thus causing the mutated cells to expand and/or replace neighbouring cells [31].

1.1.2 Cancer Stem Cells

Before discussing cancer stem cells (CSCs) it should be noted that there is no standardized definition of CSCs but instead many slightly different ones, some of which even contradict each other. Typically the definition of CSCs composed by researchers is such that it suits their current work. CSCs are not SCs, they are cells that have some of the characteristics of SCs.

SCs are biological cells that can differentiate into other types of cells and divide to reproduce more of the same type of SCs. There are two types of SCs: embryonic and somatic (adult). In the case of carcinogenesis we only consider somatic SCs, for simplicity we will term a somatic SC as a normal SC (NSC). These are found in all tissues, particularly bone marrow, fat cells, and blood, in which their function is to maintain and repair the tissue. NSCs undergo two types of cell division namely, symmetric (produces two identical SCs) and asymmetric (produces one SC and a progenitor cell) [8]. A cell is a SC if it has the following two properties: self-renewal and potency (potential to differentiate into different cell types).

In the context of SCs, self-renewal is considered the ability to achieve numerous cycles of cell division while maintaining the undifferentiated state. One way that self-renewal can be obtained is through obligatory asymmetric replication. Another mechanism of self-renewal is stochastic differentiation, which occurs when simultaneously one SC develops into two differentiated daughter cells while another SC undergoes symmetric division. Progenitors move through several rounds of cell division before terminally differentiating into a mature cell [8].

Types of potency include totipotent (omnipotent), pluripotent, multipotent, oligopotent, and unipotent. Totipotent SCs can differentiate into embryonic and extra-embryonic cell types, which results in the construction of a complete organism [123]. Pluripotent SCs are descendants of totipotent cells and can differentiate into nearly all cells, *i.e.*, cells derived from any of the

three germ layers (endoderm, mesoderm, ectoderm) [123]. Pluripotent NSCs are rare and small in number but they can be found in tissues. Multipotent SCs can differentiate into a number of cell types, but only those of a closely related family of cells [123]. Unipotent cells can produce only one cell type, their own, but have the property of self-renewal [123]. Most NSCs are multipotent and named based upon their tissue of origin.

Cancer stem cells (CSCs) are multipotent cells in a tumour that like NSCs have self-renewal, but in addition, have the abilities of tumour initiation, migration and metastasis [13, 20]. Another definition, is a small population of cells within the tumour that are tissue specific, slow dividing and with unlimited self-renewal capacity [21]. CSCs are critical in tumour initiation and progression, through their interaction with cancer cells and the extracellular matrix [26]. A CSC differs from an NSC in that it has deregulated proliferative capacity and can have metastatic properties [21, 165].

The origin of CSCs is explained by three possible processes. The first process states that an NSC undergoes several genetic as well as epigenetic alterations to give rise to a CSC [43]. The second process states that CSCs originate from NSCs that acquire a precancerous phenotype during its' development stage [14, 43, 58]. The third process states that the CSC originate from mature tumour cells [35, 69, 81, 98] or epithelial cells [14, 43, 58] that undergo dedifferentiation into a SC through modifications in signaling pathways and regulatory mechanisms. Note that the first and second processes only differ in whether an NSC acquires a genetic alteration when it is fully developed or still in development.

1.1.3 Field Cancerization

The idea of field cancerization was first mentioned by Slaughter *et al.* [133] in 1953 when histologically observing 783 squamous-cell tumours in oral cancers. Within the entire patient population it was found that benign epithelium surrounding the malignant tumour was abnormal. As well some of the patients had multiple separate tumours occur in the same area of the oral cavity. From these observations they proposed a process termed field cancerization, in which

a carcinogenic agent preconditions an area of epithelium towards cancer. If a carcinogenic agent is exposed to an area of epithelium for a sufficient amount of time and with enough intensity then it produces irreversible changes in cells and cell groups, such that the process toward cancer becomes inevitable [133]. Slaughter *et al.* [133] also hypothesized, that a field of preconditioned epithelium may develop cancer at multiple points and possibly lead to multiple tumours. As a result, they do not believe cancer arises from one cell that suddenly becomes malignant but instead from areas of precancerous change. From their hypotheses, Slaughter *et al.* [133] consider that local recurrence after surgery or radiation occurs due to left-over benign epithelium that is preconditioned towards cancer, *i.e.*, from the remaining field. Many papers were written following Slaughter *et al.* [133], that showed field cancerization can be found in colon carcinoma in patients with [49, 80, 83, 150] IBD (irritable bowl syndrome) and without IBD [2, 5, 32, 66, 74, 78, 96, 127], gastric carcinoma [63, 76, 95, 144, 148, 163, 164], aesophogeal squamous carcinoma [27, 75, 82, 92, 106, 118, 162], aesophgeal adenocarcinoma [50, 89, 90, 149], non-small-cell lung squamous carcinoma [28, 47, 72, 93, 110, 134, 137], non-small-cell lung adenocarcinoma [57, 73, 87, 101, 158], small-cell lung carcinoma [57, 73, 87, 101, 161], head and neck squamous cell carcinoma (HNSCC) (oral, oropharanx, hypopharanx, larynx) [4, 16, 18, 23, 102, 108, 126, 133, 151], breast carcinoma [39, 40, 46, 117, 146], cervix [29], prostate carcinoma [64, 104, 147], bladder carcinoma [65, 155, 156], skin carcinoma [71, 77, 138, 140, 154], melanoma [125], and blood cancer [53, 61, 99].

At the time Slaughter *et al.* [133] were conducting their research, the study of genetics was in its infancy, so they could only create hypotheses based on histological observations. The desire to understand cancer from a molecular perspective brought about studies using different molecular analyses on tumour-adjacent tissue to discover biomarkers that would indicate the presence of a field. Biomarkers that were discovered to correlate with the presence of a field are loss of heterozygosity (LOH) [141], micro-satellite alterations [141], chromosomal instability [70], and mutations in the TP53 gene [19, 152]. Braakhuis *et al.* [18] attempted to explain field cancerization from the per-

spective of genetics. They altered Slaughter *et al.* [133] definition of field cancerization to: growth of one or more genetically altered cell(s) that produces a field of cells predisposed to subsequent tumour growth. Braakhuis *et al.* [18] enforce the understanding that a field lesion does not grow invasively nor does it have metastatic properties. They proposed a process for the formation of a field and subsequent tumour within it for head and neck mucosa, esophagus, and bladder carcinomas. First, a SC acquires one or more genetic alterations and forms a group of cells with a mutation in TP53 (clonal unit) that creates genetically altered daughter cells. The SC and its' clonal unit is considered to be a lesion. Following more genetic alterations the SC gains growth advantage and develops into an expanding clone. The lesion, which is gradually becoming a field, displaces the normal epithelium surrounding it due to the enhanced proliferative capacity of a genetically altered clonal unit. As the lesion becomes larger, additional genetic hits create various sub-clones (clonal divergence) within the field. Eventually a sub-clone evolves into invasive cancer due to the presence of a large number of genetically altered SCs, and clonal divergence/selection.

Based on genetic evidence, there currently exists two main hypotheses that explain the underlying cellular basis of field cancerization: polyclonal origin and monoclonal origin. Polyclonal origin proposes that mutations occur in multiple sites of the epithelium due to continuous carcinogen exposure which leads to multi-focal carcinomas or lesions of independent origin [153]. Monoclonal origin proposes that the mutant cells from the initial lesion migrate and develop multiple lesions that share a common clonal origin. Three theories have been proposed to explain the mechanisms involved in monoclonal origin. The first two theories suggest that some tumour or tumour progenitor cells from the primary site either migrate through the submucosa or shed in the lumen of an organ (*e.g.*, the oral cavity or the bladder) which in both cases leads to the formation of a tumour at an adjacent secondary site [9, 24]. The third theory suggests that the continuous genetically altered lesions in the epithelium lead to the development of clonally related neoplastic lesions that develop via lateral spreading in the same or adjacent anatomical areas

[4, 112, 129, 143]. Some biological researchers believe both hypotheses hold, while others that just one holds. Braakhuis *et al.* [18] are in the group that believes only monoclonal origin is the correct hypothesis. Histologically it has been suggested that monoclonal origin holds true due to multiple biopsies sharing "early markers of carcinogenesis" [24, 142, 152]. Also the "late" markers being heterogeneously mixed within the tissue in and surrounding the tumour implies that clonal divergence, *i.e.*, development of multiple sub-clones occurred [105], which means that monoclonal origin likely occurred. Though monoclonal and polyclonal origin are the standard hypotheses to explain field cancerization origin based upon genetic evidence, these do not fully address all mechanisms for field formation. In fact, further exploration of biological mechanisms would likely elicit expanded theories.

Another breakthrough in biology since Slaughter *et al.* [133] was the discovery of CSCs and their importance in cancer initiation, progression, and treatment. Simple *et al.* [130] came up with a model to explain field cancerization using Braakhuis *et al.* [18] model plus the addition of CSCs. They define field cancerization as the occurrence of molecular abnormalities in the tumour adjacent mucosal field. They consider both monoclonal and polyclonal origin within their model. Simple *et al.* [130] model for oral cancer includes the following steps. First, continuous exposure of the oral mucosa to carcinogens results in molecular alterations that lead to the induction of CSC-like behaviour in a step-wise manner. Second, CSCs originate either by transformation of the NSCs or by dedifferentiation of the tumour cells and migrate through normal mucosa to develop the field. Third, initial hits at 17p (TP53) and 3p/9p (p16/FHIT) lead to transformation of the NSCs into transient amplifying cells (TACs). Fourth, these transformed cells divide and expand to create a field of neoplastic cells. Fifth, a genetic hit in the cells within the field at 13q, location of the Rb gene, allows a carcinoma to develop. Note that alteration to the Rb gene is known to release CSCs from their quiescent stage such that proliferation, self-renewal and formation of tumours can occur.

The development of the field mentioned at the second and fourth steps of the process either occurs polyclonally or monoclonally. In the case of poly-

clonal origin the following process occurs. First, NSCs at different sites in the mucosa undergo step-wise transformation into CSCs through independent carcinogen-mediated molecular alterations. Second, the CSCs proliferate leading to the development of clones at different sites. Third, additional genetic hits give rise to further divergence in the sub-clones within the field. Fourth, one of the sub-clones obtains the final genetic hit at 13q to develop into carcinoma. Considering the monoclonal process of field cancerization the following occurs. An initial lesion originates from the NSCs and gradually expands to become a field. Next, either the resident CSCs of the lesion or those that originate by dedifferentiation of the tumour cell migrate from the parental lesion. The dedifferentiation process can be driven by mutations in TP53 and over expression of OCT4, SOX2. The CSCs that migrate have gained growth advantage and thus can displace the normal epithelium by either lateral intra-epithelial migration or submucosal spread. It is known that the migratory CSCs can then switch to the non-migratory form and generate secondary tumours in the adjacent mucosal field.

Recently Curtius *et al.* [31] decided to study field cancerization from an evolutionary perspective. They define a cancerized field to be a single cell or group of cells that are further along an evolutionary path towards cancer. Since a cancerized field has mutational diversity it is a great candidate for natural selection, meaning that over time the fittest mutant clone will dominate the field. A cancerized field can be described by the following phenotypic properties: growth and death rate, and immune evasion capacity. Curtius *et al.* [31] define field cancerization drivers as mutations that drive phenotypic changes that cause a cancerized field. Driver mutations have been found in both the carcinoma and the cancerized field thus indicating that a driver mutation may also be a field cancerization driver.

Curtius *et al.* [31] define field cancerization as a somatic evolutionary process that produces cells that are close to cancer. Braakhuis *et al.* [18] definition of field cancerization implies that the mutant clone that grows has an altered phenotype that drives its expansion. From an evolutionary perspective this is achieved by a mutant clone being fitter than the resident cell population.

During field cancerization multiple phenotype states are achieved. Per Curtius *et al.* [31] a cancerized fields' formation is driven by exposure of a prevalent carcinogen and promoter of clonal expansion and/or subsequent convergent evolution of the epigenome. As a result field cancerization can occur because of multiple independent clonal expansions, *i.e.*, polyclonal origin. Thus, both Simple *et al.* [130] and Curtius *et al.* [31] consider that a cancerized field can be formed via monoclonal or polyclonal origin.

The general framework that Curtius *et al.* [31] propose for the initiation of a cancerized field is as follows. A group of cells, gradually becoming a cancerized field, undergo mutations that occur due to DNA replication errors during aging and/or carcinogens, resulting in many genetically distinct clones within. As more time passes, daughter clones with phenotypes that increase their fitness will dominate the group. Finally, if the mutagenic insult is ongoing then new clones will be continually generated and the cancerized field will appear genetically diverse. At this point carcinoma will first occur as described by Simple *et al.* [130], *i.e.*, when one of the clones acquires a genetic hit at 13q.

Considering all the literature, a field will be considered as a region of tissue that has genetic and phenotypic change that preconditions it towards the possible formation of one or multiple tumours within it. The genetic and phenotypic change can be caused by carcinogenic onslaught, genetic defects at birth, chance mutations, or a combination thereof. Field cancerization is then described as the process that results in the formation of a field that successfully yields one or more tumours within it. Note that the causes of genetic change will be provided solely by chance mutations and carcinogens. The steps of the process of field cancerization that will be considered here are as follows:

1. a region of tissue is affected by one or more carcinogen over time;
2. the carcinogen(s) cause genetic mutations in the cells of the tissue which in turn influence the phenotype of the cell;
3. as the cells start to proliferate and differentiate, the field expands;
4. eventually a CSC will be created, which will finally create the first TC.

It is important to note that not all the cells in the field have to be clonally related they just have to be considered preconditioned towards tumour formation, so both monoclonal and polyclonal origins are possible.

1.2 Mathematical Literature Review

There exists an extensive amount of literature that study cancer initiation [38, 107, 122], progression [10, 41], metastasis [48], treatment (chemotherapy, immunotherapy, radiation) [34, 41, 114], and effects of various microenvironmental and external factors on cancer development [54] from the perspective of mathematical analysis. However, the only mathematical analysis on field cancerization that could be found at the time of writing, were spatial stochastic models in Foo *et al.* [45] and Ryser *et al.* [120].

1.2.1 Review of Foo et al.

Foo *et al.* [45] describe field cancerization as the process of primary tumours forming from genetically altered fields of premalignant cells that have high chances of progression to malignancy. Also they state that the premalignant fields can cause recurrent tumours if not excised with the primary tumour during surgery. The main objectives of their study were to:

1. develop a spatial evolutionary framework for field cancerization;
2. describe the size and geometry of the premalignant fields at the moment of tumour initiation;
3. determine the risk of multi-focal lesions, recurrence timing, and clonal origin of recurrent tumours;
4. discover the effects of different characteristics of tissue and cancer type on 2 and 3.

The domain of their model is a lattice of dimension (d) 1, 2, or 3 wherein each lattice point is occupied by one cell. The dimension is generally either 1 or 2 because although epithelial tissues are 3D it is sufficient to consider just

1 or 2D approximations of the areas of interest. For example due to the ratio of tube radius to the length of mammary ducts of the breast, renal tubules of the kidney and bronchi tubes of the lung can be approximated with d=1. Also cancer initiation in the squamous epithelium of the cervix, bladder, and oral cavity can be considered a d=2 process as it occurs in the basal layer which is only 1-2 cells thick. In the lattice model, each cell reproduces by placing its offspring randomly in one of the $2d$ neighbours at the rate equal to their fitness, $s \geq 0$, at exponential waiting time. A cell having a fitness advantage has increased reproductive rate or avoidance of apoptotic signals. The cell type is determined by the fitness, s, which they also describe as the number of genetic hits a cell has accumulated. A cell of type-0 have fitness normalized to 1, s=1, and are considered to be wild-type or normal. The type-0 cell can become a type-1 cell by acquiring the first type of mutation at the rate u_1. In general a type-i cell has a fitness advantage of $1 + s_i$ relative to the type-(i-1) and is assumed to have acquired all the mutations up to mutation i. If it acquires the $(i+1) - th$ mutation at rate u_{i+1} then it will become a type-(i+1) cell. The mutation rates are quite small, $u_i << 1$, and so the domain has to be large to increase the chance a cell will acquire a mutation. The previously stated definitions and properties make up what is called the spatial Moran model.

The lattice is initialized with cells having type-0. The starting time is at the end of tissue development and start of the tissue renewal phase. Since this time is difficult to acquire it is challenging to find the time of cancer initiation, σ_k. It is assumed that when a cell has developed k mutations and so is of type-k that cancer has initialized and the simulation is stopped. The critical number of mutations k is set based upon the type of cancer because different cancers have different mutation thresholds. It is important to note that the model assumes all k-mutations positively influence cancer initialization and have the effect of increased cell growth and/or reduction in apoptotic signalling. Another assumption they used was, that all their mutations come from random mutations, which occur very rarely and are generally fixed by the bodies DNA repair process, before they have an effect. Foo *et al.* [45] ignore the genetic hits that cause selective disadvantage because the result is that the

cell and its' progeny die too quickly. Cell death and reproduction dynamics are accomplished similarly to the biased voter model [86].

To determine the probability of a mutant clone population of type-1 surviving they considered a one mutation model with cells of type-0 having fitness 1 and type-1 having fitness 1+s. The lattice is initialized with all type-0 cells except the cell at x=0, this allows the lattice to be either finite or infinite. In the finite case a Williams-Bjerknes model [160] is used, otherwise a biased voter model is used [86]. The number of type-1 cells is a jump process in which a discrete time random walk moves one up with probability

$$\frac{s}{1+s}$$

or one down with probability

$$\frac{1}{1+s},$$

since the only possible events are type-0 which is replaced by type-1 (jump up) and type-1 which is replaced by a type-0 (jump down). Thus the probability that a mutant clone population of type-1 survives is

$$\frac{s}{1+s} \approx s$$

(when $s << 1$). A mutant clone population with fitness s is successful if it reaches size $>> \frac{1}{s}$ as this results in a negligible chance of its extinction. Unsuccessful type-1 mutant clone populations typically have a space-time volume of order:

$$l(s)=\begin{cases} s^{-2}, & d=1 \\ s^{-1}\ln(s^{-1}), & d=2 \\ s^{-1}, & d=3 \end{cases},$$

which represents the magnitude of the domain a mutant clone population will occupy. The probability of a mutant clone population of type-1 surviving still holds for spatial Moran models as long as

$$\frac{1}{u_1} >> l(s)^{\frac{d+2}{2}},$$

where u_1 is the mutation rate for type-0 cells to become type-1 cells.As a result of the previous condition if the number of type-1 cells is significantly less than

the total number of cells, N, for all time, then successful type-1 mutations arrive as a Poisson process at rate

$$\frac{Nu_1 s}{1+s}.$$

Upon considering a mesoscopic model (a hybrid model that includes elements from both the macroscopic and microscopic scales) they allow type-1 mutations arrive as Poisson with rate Nu_1, distributed uniformly at random in the domain. Each mutation event has two potential outcomes:

(a) With probability $\frac{s}{1+s}$ the mutation is successful and the clonal expansion is approximated with a ball whose radius grows deterministically with macroscopic growth rate $c_d(s)$.

(b) With probability $\frac{1}{1+s}$ the mutation is unsuccessful and the clone is conditioned to go extinct.

Having acquired the basic probabilistic and size aspects of the model they moved to a two mutation model so that it could be fit in the context of field cancerization whereby type-0 cells have fitness 1, type-1 cells are premalignant with fitness $1+s_1$ relative to type-0, and type-2 are cancer cells with fitness $1+s_2$ relative to type-1. If $s_1=s_2=s > 0$ the timing of cancer initiation is controlled by the limiting value of the meta-parameter:

$$\Gamma=(Nu_1 s)^{d+1}(c_d^d(s)u_2 s)^{-1},$$

which is the number of type-1 clones needed to generate a successful type-2 mutation. If $\Gamma \to 0$ then the first successful type 2 mutation will occur in the first successful type 1 clone. Varying the meta-parameter Γ results in the following three regimes:

(a) Regime 1 (R1): When $\Gamma<1$ the first successful type-2 mutation occurs within the expanding clone of the first successful type-1 mutation. The time at cancer initiation, σ_2, is exponential and does not depend on the spatial dimension.

(b) Regime 2 (R2): For $\Gamma \in (10, 100)$ the first successful type-2 mutation occurs within one of the several successful type-1 clones. The time at cancer initiation, σ_2, is not exponential and depends on the spatial dimension.

(c) Regime 3 (R3): When $\Gamma > 1000$ the first successful type-2 mutation occurs after many successful type-1 mutations. It can arise from a successful or unsuccessful type-1 mutation. The time at cancer initiation, σ_2, is a mixture distribution of these events.

The borderline regimes $R1/R2$ and $R2/R3$ occur for $\Gamma \in [1, 10]$ and $\Gamma \in [100, 1000]$, respectively. For all further analysis Foo *et al.* [45] assumed that in $R3$ successful type-2 mutation arise only from successful type-1. As the number of cells increases, the time that the first type-2 mutation arises, σ_2, decreases, and thus cancer initiation occurs earlier. When σ_2 is small, there is a diffuse premalignant field and a large number of independent lesions. However, when σ_2 is large there is a single premalignant field that contains the initial tumour cell. Assuming that the time of diagnosis, T_D, is independent of σ_2 then the premalignant field at time of diagnosis, $\sigma_2 + T_D$, can be characterized by the field at σ_2 together with the distribution of T_D.

The size of an initiating clone population is picked from a distribution generated around the sizes of the clones at the time the initiated mutation arose. If the premalignant field is left during surgery, they found that tumours appearing later have a higher recurrence probability. As u_1 increases, the premalignant field is made up of an increasing number of independent type-1 patches and the model moves towards R2 and R3. As the number of type-1 patches increases, so does the chance of a type-2 cell being formed and hence σ_2 (time of first type-2 cell forming) decreases. As u_2 increases, the model moves towards R1 in which fewer type-1 clones are required to produce the first successful type-2 cell, and the size of the type-1 field decreases. In R1 and R2 the total distant field size is of the same order of magnitude as the local field size, whereas in R3 it is significantly larger. Thus, secondary tumour recurrences for cancer types in R3 are more likely to come from the distant field and as a result are not directly related to the primary tumour because the

distant fields will be genetically distinct. In R1 the expected number of small cell groups peaks and then declines as larger cell groups begin to dominate, whereas in R2 and R3 small and large cell groups coexist for a longer period of time. A larger premalignant field increases the chance of fast recurrence. In R1 local recurrence is more likely, however when σ_2 is large, then it is slightly more likely to come back in the distant fields. In R2 and R3 the overall probability of local and distant recurrences is comparable, however when σ_2 is small, recurrence is more likely to occur in distant fields and when σ_2 is larger local recurrence is more likely.

Foo *et al.* [45] also found the distribution of the local field radius, distribution of the area of the local field at σ_2, size-distribution of the distant field clones at σ_2, distribution of the number of field patches, probability of a second field tumour having formed before a given time, and probability that the distant field at the time of initiation gives rise to a second primary tumour at some point in time.

A limitation of Foo *et al.* [45] is that they use a specific sequence of mutations, although it is well known that mutations are not sequential and they occur in random order. They assume the micro-environment is static and uniform, which brings about differences in the timing and intensity of malignancy. They use circular growth for the cell groups which does not allow them to grow in every direction or shapes other than circular. They assume a two-hit mutation model, when in reality it is dependent on the cancer as to how many mutations are required. There is no genetic heterogeneity within the clone groups. They don't consider carcinogens as mutagens but instead consider only random mutations, which would greatly dwindle the process of field cancerization since carcinogenic onslaught is considered a significant mutagen. The phenotypic actions they consider are proliferation and apoptosis, but they do not include differentiation and quiescence. They ignore tissue structures, thus boundaries and growth limitations will not occur. Finally, they include only normal tissue cells, where as, most tissue also includes blood cells, fat cells, and stem cells.

1.2.2 Review of Ryser et al.

Ryser *et al.* [120] applied the model described in Foo *et al.* [45] to head and neck squamous cell carcinoma (HNSCC) which is commonly a result of field cancerization. They specifically investigated HPV-negative cases, patients that had a history of smoking, and age of diagnosis. Ryser *et al.* [120] attribute the difficulty of detecting the premalignant fields surrounding the primary tumour in clinical practise to a lack of understanding of the dynamics and geometry of the fields. Since HNSCC are formed in stratified epithelia, the growth and renewal of the tissue is accomplished by stem cells (SC), which they call progenitor cells (PC). The PC are located in the basal layer of the tissue, and they renew the tissue by producing transit amplifying cells (TAC), which have limited proliferative potential. These TAC move upwards through the tissue generating new cells and within a few weeks or less are sloughed off. As a result of the short lifespan of the TAC, they are not good candidates for becoming mutated and thus do not create neoplastic lesions. Therefore Ryser *et al.* [120] only consider the PC in their microscopic model. Cells mutate from normal to cancerous through accumulation of genetic aberrations. When a normal PC acquires growth advantage its' progeny start spreading through the epithelium. Since the number and timing of genetic alterations changes from patient to patient Ryser *et al.* [120] decided to look at phenotypic progression instead of genotypic progression.

The three histopathological stages of epithelial dysplasia (precancerous stages) are mild, moderate and severe (carcinoma in situ [CIS]). Based on these three stages Ryser *et al.* [120] consider the following four type of cells: normal cells (type 0), mildly dysplastic cells (type 0*), moderately dysplastic cells (type 1), and severely dysplastic cells (type 2). They use the stochastic Moran model on a regular two-dimensional lattice as described in Foo *et al.* [45]. Ryser *et al.* [120] initialize the model with normal PCs (type 0) with proliferative rate f_0. A type 0 cell becomes a type 0* cell at the rate $u_{1,a}$, type 0* becomes type 1 at the rate $u_{1,b}$, and type 1 becomes type 2 at the rate u_2. Biologically the proliferative rate of type 0 and type 0* is the same so type

0* has proliferative rate f_0. Based on the previous result they computed the probability of a type 0* cell becoming a type 1 cell, thus resulting in a rate v_{01} that a type 0 cell becomes a type 1 cell. Type 1 cells have a proliferative advantage over type 0 and type 0* cells due to having a fitness advantage s_1, and so its' proliferative rate is given by $f_1 = f_0(1 + s_1)$. Similarly the type 2 cells have a proliferative rate of $f_2 = f_1(1 + s_2)$.

Ryser *et al.* [120] model the time between onset of carcinoma in situ (CIS) and diagnosis using an exponentially distributed random variable with rate Ψ. To analytically compute the waiting times and field geometries they use the mesoscopic approximation to the spatial model from Foo *et al.* [45]. In this model the arrival of expanding type 1 clones is a stochastic Poisson process with rate $Nu_1\frac{s_1}{1+s_1}$, where N is the total number of cells. The factor $\overline{s_1} = \frac{s_1}{1+s_1}$ represents the idea that progeny of a new type 1 cell either go to extinction with probability $1 - \overline{s_1}$ or expand indefinitely with probability $\overline{s_1}$.

According to a theorem, expanding type 1 clones asymptotically grow as a convex symmetric shape with constant radial growth rate c_2. In Foo *et al.* [45] it was found in particular that the convex symmetric shape can be approximated as a disk or circle. The rate c_2 depends on the selective advantage s_1. For small s_1 it scales as $c_s(s_1) \sim \sqrt{4\pi s_1/\log(1/s_1)}$, where $f(s) \sim g(s)$ means that $f(s)/g(s) \to 1$ as $s \to 0$.

For larger values of s_1 the relationship had to be numerically computed, in particular for $s_1 > 0.5$, Ryser *et al.* [120] found an approximately linear dependence given by $c_2(s_1) \approx 0.6s_1+0.22$. Ryser *et al.* [120] allow the existence of multiple precancer fields of type 1 cells within the model.

To estimate and compute the parameters for their model they use age-specific incidence rates from the Surveillance, Epidemiology, and End Results (SEER) program of the National Cancer Institute (18 registries, 2000-2012) in a Bayesian framework. Since Ryser *et al.* [120] only considered HPV-negative cancers they restricted to only HNSCC within the lip, tongue, floor of the mouth, gum and other mouth, hypopharynx, and larynx. They computed the number of susceptible individuals and the number of cancer cases diagnosed for the following age groups: 15-19, 20-24, ..., 80-84, and 85+. The final reduction

to the data used was to consider only smokers within each considered age groups, as tobacco consumption is a major cause of HPV-negative HNSCC.

Ryser *et al.* [120] computed the survival function, the probability density function of the local field radius, and the probability of harboring at least two clonally unrelated fields in the head and neck region with respect to the mean age at smoking initiation to diagnosis with invasive cancer. They found that there is a strong dependence of the local field size on age at diagnosis, with a doubling of the expected field diameter between ages at diagnosis of 50 and 90 years. Further the probability of harboring multiple clonally unrelated fields at the time of diagnosis were found to increase substantially with patient age. As a result of these discoveries Ryser *et al.* [120] suggest that patient age at diagnosis is a critical predictor of the size and multiplicity of precancerous lesions.

1.2.3 Review of Gerlee et al.

The next set of literature discusses mathematical techniques that were used for the model that will be proposed in this thesis. The first paper by Gerlee & Anderson [54] inspired the general framework of the mathematical model. Gerlee & Anderson [54] created a hybrid cellular automaton to model the effect of various micro-environmental factors on solid tumour growth. Their model is a hybrid cellular automaton because the rule of the automaton depends upon the output of a neural network and partial differential equations. The cellular automaton comprises of two cell types; an empty cell (normal cell) and a tumour cell. It is initialized by setting all the automaton elements to empty except the middle four cells which are occupied by tumour cells. The neural network is used to approximate the relationship between the micro-environmental variables and the phenotype of a cell. The partial differential equations are used to model the spread of the various micro-environmental variables in the domain of consideration.

For the neural network they use a multi-layer perceptron (MLP), with input being the output of the partial differential equation for the cell at a location (x, y) and output being a vector of likelihoods of a phenotype and movement

occurring at a time-step. The hidden layer of the MLP represents the genes and hence the neural network attempts to replicate the genotype-phenotype relationship. They consider the phenotypes proliferation (P), quiescence (Q), and apoptosis (A). Each time-step represents a cell cycle so that a single phenotypic action will occur for each cell. The maximum of the likelihoods between P, Q, and A determines which phenotypic action occurs. If the likelihood of movement is sufficiently large then the cell is allowed to move.

Each of the actions also has some restrictions based upon the cells metabolism and adhesion but these will be ignored due to both of these aspects of the cell not being modelled in this thesis. All of the partial differential equations are chemical field equations of the form:

$$\frac{\partial c(x,y,t)}{\partial t} = D\Delta c(x,y,t) \pm f(x,y,t).$$

Where $c(x,y,t)$ is the concentration of the micro-environmental variable, D is the diffusion coefficient, Δ is the Laplacian operator, and

$$f(x,y,t) = \begin{cases} 0, & \text{If the automaton element at } (x,y) \text{ is empty} \\ rF(x,y), & \text{If the automaton element is occupied} \end{cases};$$

where r is the base consumption/production rates and $F(x,y)$ is the modulated energy consumption of the individual cell occupying the automaton element at (x,y). The hybrid cellular automaton goes through the following process for each cell and time-step:

1. The input to the neural network is sampled from the local environment.
2. The MLP is computed to obtain the phenotype and likelihood of movement.
3. The cell consumes nutrients according to the chosen phenotype and the metabolic pathway is chosen.
4. The chosen phenotype is carried out.
5. If movement is activated and the cell has not divided it tries to move to a neighbouring cell location.

This thesis aims to answer the following questions around field cancerization:

1. What degree of carcinogenic onslaught is necessary for field cancerization to occur? Which carcinogens are the most aggressive?

2. How long before a cancer field is formed? How long before the first tumour cell within the field is formed?

3. Is the field formed via monoclonal origin, polyclonal origin, or a mixture of both? Which type of origin is the most common?

4. How long does it take for a tumour to be large enough such that it is detectable by physicians? Once a tumour is detected, what size is the surrounding field?

5. How long does it take a recurrence to occur after removal of the field versus the field remaining intact?

Chapter 2

Model

2.1 Model Overview

Here we develop a hybrid cellular automaton (CA) model for the cancer field effect. The CA is hybrid since its rule depends on the output of other mathematical objects. In this case the mathematical objects are partial differential equations (PDE) and neural networks (NN). The PDE model spreads one or more carcinogen(s) within the domain of the CA. The NN is used to compute the change in gene expression of the genes under consideration for each cell with respect to the amount of carcinogen at the cell's location and age of the cell. The CA includes states that are used to represent the following biological cell types: normal tissue cells (NTC), mutated normal tissue cells (MNTC), normal stem cells (NSC), mutated normal stem cells (MNSC), cancer stem cells (CSC), and tumour cells (TC). Evolution of the model occurs in the following basic steps:

1. Carcinogens spread via a reaction diffusion PDE or a given function.

2. Changes in gene expressions resulting from carcinogenic exposure and age of the cell are computed by the NN causing gene mutations to occur.

3. The state of each cell is updated using the CA rule, which includes spatial translocation, genetic mutations, phenotypic drift, mitosis, cell death through the process of apoptosis, and dedifferentiation.

2.2 Carcinogen Partial Differential Equations

Consider a carcinogen that is in the spatial domain
$\Omega=\{\boldsymbol{x}=(x_1,x_2)|0<x_1<L, 0<x_2<M\}$ and evolves in the domain
$\Omega_T=\Omega\times(0,T], T>0$. The concentration for the carcinogen is computed by the function $c(\boldsymbol{x},t)$. In which $c(\boldsymbol{x},t)$ is a solution to the following initial boundary value problem (IBVP)

$$\text{PDE} \quad c_t(\boldsymbol{x},t)=D\Delta c(\boldsymbol{x},t)+F(\boldsymbol{x},t), (\boldsymbol{x},t)\in\Omega; \tag{2.1}$$

$$\text{BCs} \quad c(0,x_2,t)=g_1(x_2,t), c(L,x_2,t)=g_2(x_2,t), (x_2,t)\in\partial\Omega\times[0,T]; \tag{2.2}$$

$$\text{BCs} \quad c(x_1,0,t)=g_3(x_1,t), c(x_1,M,t)=g_4(x_1,t), (x_1,t)\in\partial\Omega\times[0,T]; \tag{2.3}$$

$$\text{IC} \quad c(\boldsymbol{x},0)=f(\boldsymbol{x}), \boldsymbol{x}\in\Omega; \tag{2.4}$$

$$\text{Source} \quad F(\boldsymbol{x},t)=I(\boldsymbol{x},t)-O(\boldsymbol{x},t), \tag{2.5}$$

where $\Delta=\frac{\partial^2}{\partial x_1^2}+\frac{\partial^2}{\partial x_2^2}$; $F(\boldsymbol{x},t)\in\mathbb{R}$ is the source term with $I(\boldsymbol{x},t)\in\mathbb{R}_+$ being the input and $O(\boldsymbol{x},t)\in\mathbb{R}_+$ being the loss of the carcinogen;
$g_1(x_2,t), g_2(x_2,t), g_3(x_1,t), g_4(x_1,t), f(\boldsymbol{x})\in\mathbb{R}_+$. The IBVP (2.1)-(2.4) is non-homogeneous both in the boundary conditions (BC) and PDE.

Let us non-dimensionalize the equation so that the solution of the PDE is unit-less and can be later used as input into a neural network. To non-dimensionalize a PDE all the dependent and independent variables need to be made dimensionless. This is achieved by dividing all the independent and dependent variables by some characteristic value that is denoted with a subscript c. Let us define the dimensionless variables

$$\hat{x_1}=\frac{x_1}{x_c^{(1)}}, \hat{x_2}=\frac{x_2}{x_c^{(2)}}, \hat{t}=\frac{t}{t_c} \tag{2.6}$$

$$\hat{c}=\frac{c}{c_c}, \hat{F}=\frac{F}{F_c} \tag{2.7}$$

$$\hat{g_1}=\frac{g_1}{g_c^{(1)}}, \hat{g_2}=\frac{g_2}{g_c^{(2)}}, \hat{g_3}=\frac{g_3}{g_c^{(3)}}, \hat{g_4}=\frac{g_4}{g_c^{(4)}}, \tag{2.8}$$

$$\hat{f}=\frac{f}{f_c}. \tag{2.9}$$

We choose the characteristic values of the source term, boundary conditions and initial condition to be the absolute value of the maximum value of the

functions so that

$$F_c = \max_{\boldsymbol{x},t} |F(\boldsymbol{x},t)| \tag{2.10}$$

$$g_c^{(1)} = \max_{x_2,t} |g_1(x_2,t)|, g_c^{(2)} = \max_{x_2,t} |g_2(x_2,t)|, \tag{2.11}$$

$$g_c^{(3)} = \max_{x_1,t} |g_3(x_1,t)|, g_c^{(4)} = \max_{x_1,t} |g_1(x_1,t)| \tag{2.12}$$

$$f_c = \max_{\boldsymbol{x}} |f(\boldsymbol{x})|. \tag{2.13}$$

To discover the characteristic space, time, and concentration values we need to plug the dimensionless variables into the PDE, which gives us

$$\begin{aligned}\frac{c_c}{t_c}\hat{c}_{\hat{t}}(\hat{x_1},\hat{x_2},\hat{t}) = Dc_c \left(\frac{\hat{c}_{\hat{x_1}\hat{x_1}}(\hat{x_1},\hat{x_2},\hat{t})}{{x_c^{(1)}}^2} + \frac{\hat{c}_{\hat{x_2}\hat{x_2}}(\hat{x_1},\hat{x_2},\hat{t})}{{x_c^{(2)}}^2} \right) \\ + \hat{F}(\hat{x_1},\hat{x_2},\hat{t}) F_c.\end{aligned}$$

Multiplying the above by $\frac{t_c}{c_c}$ and letting $x_c^{(1)} = x_c^{(2)} := x_c$ gives us

$$\begin{aligned}\hat{c}_{\hat{t}}(\hat{x_1},\hat{x_2},\hat{t}) = \frac{Dt_c}{x_c^2} \left(\hat{c}_{\hat{x_1}\hat{x_1}}(\hat{x_1},\hat{x_2},\hat{t}) + \hat{c}_{\hat{x_2}\hat{x_2}}(\hat{x_1},\hat{x_2},\hat{t}) \right) \\ + \hat{F}(\hat{x_1},\hat{x_2},\hat{t}) \frac{F_c t_c}{c_c}.\end{aligned} \tag{2.14}$$

By convention of non-dimensionalization we make the coefficients equal to one so to simplify the equation, this gives us

$$\frac{Dt_c}{x_c^2} = 1 \implies t_c = \frac{x_c^2}{D}, \tag{2.15}$$

$$\frac{F_c t_c}{c_c} = 1 \implies c_c = F_c t_c = \frac{x_c^2 F_c}{D}. \tag{2.16}$$

Thus we have the non-dimensional PDE given by

$$\hat{c}_{\hat{t}}(\hat{x_1},\hat{x_2},\hat{t}) = \Delta\hat{c}(\hat{x_1},\hat{x_2},\hat{t}) + \hat{F}(\hat{x_1},\hat{x_2},\hat{t}) \tag{2.17}$$

For convenience we choose

$$x_c = \max(L, M). \tag{2.18}$$

We can now write out the boundary conditions in the new dimensionless vari-

ables

$$\hat{c}(0,\hat{x_2},\hat{t})=\frac{g_c^{(1)}}{c_c}\hat{g_1}(\hat{x_2},\hat{t})=\frac{g_c^{(1)}D}{x_c^2F_c}\hat{g_1}(\hat{x_2},\hat{t}), \tag{2.19}$$

$$\hat{c}\left(\frac{L}{x_c},\hat{x_2},\hat{t}\right)=\frac{g_c^{(2)}}{c_c}\hat{g_2}(\hat{x_2},\hat{t})=\frac{g_c^{(2)}D}{x_c^2F_c}\hat{g_2}(\hat{x_2},\hat{t}), \tag{2.20}$$

$$\hat{c}(\hat{x_1},0,\hat{t})=\frac{g_c^{(3)}}{c_c}\hat{g_3}(\hat{x_1},\hat{t})=\frac{g_c^{(3)}D}{x_c^2F_c}\hat{g_3}(\hat{x_1},\hat{t}), \tag{2.21}$$

$$\hat{c}\left(\hat{x_1},\frac{M}{x_c},\hat{t}\right)=\frac{g_c^{(4)}}{c_c}\hat{g_4}(\hat{x_1},\hat{t})=\frac{g_c^{(4)}D}{x_c^2F_c}\hat{g_4}(\hat{x_1},\hat{t}). \tag{2.22}$$

Finally we write out the initial condition in the new dimensionless variables

$$\hat{c}(\hat{x_1},\hat{x_2},0)=\frac{f_c}{c_c}\hat{f}(\hat{x_1},\hat{x_2})=\frac{f_cD}{x_c^2F_c}\hat{f}(\hat{x_1},\hat{x_2}). \tag{2.23}$$

Thus we have the non-dimensional PDE problem given by

$$\hat{c_{\hat{t}}}(\hat{x_1},\hat{x_2},\hat{t})=\Delta\hat{c}(\hat{x_1},\hat{x_2},\hat{t}),(\hat{\boldsymbol{x}},\hat{t})\in\left(0,\frac{L}{x_c}\right)\times\left(0,\frac{M}{x_c}\right)\times\left(0,\frac{T}{t_c}\right] \tag{2.17}$$

$$+\hat{F}(\hat{x_1},\hat{x_2},\hat{t}),$$

$$\hat{c}(0,\hat{x_2},\hat{t})=\frac{g_c^{(1)}D}{x_c^2F_c}\hat{g_1}(\hat{x_2},\hat{t}),(\hat{x_2},\hat{t})\in\left(0,\frac{M}{x_c}\right)\times\left(0,\frac{T}{t_c}\right], \tag{2.19}$$

$$\hat{c}\left(\frac{L}{x_c},\hat{x_2},\hat{t}\right)=\frac{g_c^{(2)}D}{x_c^2F_c}\hat{g_2}(\hat{x_2},\hat{t}),(\hat{x_2},\hat{t})\in\left(0,\frac{M}{x_c}\right)\times\left(0,\frac{T}{t_c}\right], \tag{2.20}$$

$$\hat{c}(\hat{x_1},0,\hat{t})=\frac{g_c^{(3)}D}{x_c^2F_c}\hat{g_3}(\hat{x_1},\hat{t}),(\hat{x_1},\hat{t})\in\left(0,\frac{L}{x_c}\right)\times\left(0,\frac{T}{t_c}\right], \tag{2.21}$$

$$\hat{c}\left(\hat{x_1},\frac{M}{x_c},\hat{t}\right)=\frac{g_c^{(4)}D}{x_c^2F_c}\hat{g_4}(\hat{x_1},\hat{t}),(\hat{x_1},\hat{t})\in\left(0,\frac{L}{x_c}\right)\times\left(0,\frac{T}{t_c}\right], \tag{2.22}$$

$$\hat{c}(\hat{x_1},\hat{x_2},0)=\frac{f_cD}{x_c^2F_c}\hat{f}(\hat{x_1},\hat{x_2}),(\hat{x_1},\hat{x_2})\in\left(0,\frac{L}{x_c}\right)\times\left(0,\frac{M}{x_c}\right), \tag{2.23}$$

$$x_c=\max(L,M), \tag{2.18}$$

$$t_c=\frac{x_c^2}{D}, \tag{2.15}$$

$$F_c=\max_{\boldsymbol{x},t}|F(\boldsymbol{x},t)|, \tag{2.10}$$

$$g_c^{(1)}=\max_{x_2,t}|g_1(x_2,t)|,g_c^{(2)}=\max_{x_2,t}|g_2(x_2,t)|, \tag{2.11}$$

$$g_c^{(3)}=\max_{x_1,t}|g_3(x_1,t)|,g_c^{(4)}=\max_{x_1,t}|g_1(x_1,t)| \tag{2.12}$$

$$f_c=\max_{\boldsymbol{x}}|f(\boldsymbol{x})|. \tag{2.13}$$

Since the diffusion equation is a linear differential equation of the parabolic type then we can solve the IBVP via the method of Green's Function as described by Polyanin & Nazaikinskii [111].

First let us go through the theory of solving a non-homogeneous linear differential equation of the parabolic type in n space variables of the form

$$u_t - L_{\boldsymbol{x}}[u]=\overline{\Phi}(\boldsymbol{x},t), \tag{2.24}$$

where $u=u(\boldsymbol{x},t), \boldsymbol{x}=V\subset\mathbb{R}^n$, with V being a simply connected region with boundary $S=\partial V$. The symbol $L_{\boldsymbol{x}}$ is a second-order partial differential operator having non-divergence form so

$$L_{\boldsymbol{x}}[u]=\sum_{i,j=1}^{n}\overline{a}_{ij}(\boldsymbol{x},t)u_{x_ix_j}+\sum_{i=1}^{n}\overline{b}_i(\boldsymbol{x},t)u_{x_i}+\overline{c}(\boldsymbol{x},t)u. \tag{2.25}$$

Also the partial differential operator $\frac{\partial}{\partial t}-L_{\boldsymbol{x}}$ is uniformly parabolic, that is there exists a constant $\theta>0$ such that

$$\sum_{i,j=1}^{n}\overline{a}_{ij}(\boldsymbol{x},t)\xi_i\xi_j\geq\theta\sum_{i=1}^{n}\xi_i^2$$

for all $(\boldsymbol{x},t)\in V\times(0,T]$. The initial condition for the PDE is given by

$$u=\overline{f}(\boldsymbol{x}) \text{ at } t=0. \tag{2.26}$$

Let $S_k, k=1,...,p$ be distinct portions of the surface S such that $S=\sum_{k=1}^{p}S_k$ then the non-homogeneous linear boundary condition is given by

$$\Gamma_{\boldsymbol{x}}^{(k)}[u]=\overline{g_k}(\boldsymbol{x},t), (\boldsymbol{x},t)\in S_k\times[0,T], k=1,...,p. \tag{2.27}$$

In the general case, $\Gamma_{\boldsymbol{x}}$ is a first-order linear differential operator in the space coordinates with coefficients depending on $\boldsymbol{x}$ and t. The three main important forms of $\Gamma_{\boldsymbol{x}}$ include

1. Dirichlet: $\Gamma_{\boldsymbol{x}}^{(k)}[u]=u$,
2. Neumann: $\Gamma_{\boldsymbol{x}}^{(k)}[u]=u_{M_x}$,
3. Robin: $\Gamma_{\boldsymbol{x}}^{(k)}[u]=u_{M_x}+\overline{v}(\boldsymbol{x},t)u$,

where $u_{M_x}=\sum_{i,j=1}^{n}\overline{a}_{ij}(\boldsymbol{x},t)N_jG_{x_i}$ with $\boldsymbol{N}=\{N_1,...,N_n\}$ being the unit outward normal to the surface $S_k, k=1,...,p$. The boundary condition allows the carcinogen concentration to be influenced by some source at the boundary. A Dirichlet boundary condition is used in cases where it is known how the carcinogen concentration is distributed at the boundary. For example, it might be known that the surrounding tissue releases carcinogen at some fixed rate. Another possibility is that surrounding blood vessels or saliva release the carcinogen into the domain. A Neumann boundary condition is utilized when only the normal derivative of the carcinogen concentration at the boundary is known or in other words the flux of the carcinogen. It can be interpreted as the carcinogen concentration spatial distribution at the boundary changing at some known rate at each time-step. Biologically, this might be the carcinogen entering the domain via some porous medium or biological gate that regulates and controls what passes through, such as a protein gate. The Robin boundary condition is a linear combination of Dirichlet and Neumann conditions. Biologically, multiple sources would influence the carcinogen concentration at the boundary some of which would be constant and others that change the concentration spatial distribution at some rate. By the theory of Green's function the solution of the non-homogeneous linear boundary value problem defined by (2.24)-(2.27) is given by

$$u(\boldsymbol{x},t)=\int_0^t\int_V\overline{\Phi}(\boldsymbol{y},\tau)G(\boldsymbol{x},\boldsymbol{y},t,\tau)dV_yd\tau+\int_V\overline{f}(\boldsymbol{y})G(\boldsymbol{x},\boldsymbol{y},t,0)dV_y \qquad (2.28)$$
$$+\sum_{k=1}^{p}\int_0^t\int_{S_k}\overline{g_k}(\boldsymbol{y},\tau)H_k(\boldsymbol{x},\boldsymbol{y},t,\tau)dS_yd\tau.$$

$G(\boldsymbol{x},\boldsymbol{y},t,\tau)$ is the Green's function; for $t>\tau\geq 0$, it satisfies the homogeneous equation

$$G_t-L_{\boldsymbol{x}}[G]=0 \qquad (2.29)$$

with the non-homogeneous initial condition of the special form

$$G=\delta(\boldsymbol{x}-\boldsymbol{y}) \text{ at } t=\tau \qquad (2.30)$$

and the homogeneous boundary condition

$$\Gamma_{\boldsymbol{x}}^{(k)}[G]=0, \boldsymbol{x}\in S_k, k=1,..,p. \qquad (2.31)$$

The vector $\boldsymbol{y}=\{y_1,...,y_n\}\in V$ from problem (2.29)-(2.31) is an n-dimensional free parameter, and $\delta(\boldsymbol{x}-\boldsymbol{y})=\delta(x_1-y_1)...\delta(x_n-y_n)$ is the n-dimensional Dirac delta function. The Green's function G is independent of the in-homogeneity's in the IBVP (2.24)-(2.27), so it is independent to functions $\overline{\Phi}$, $\overline{f}$, and $\overline{g}$. In the solution (2.28) the integrations are performed everywhere with respect to $\boldsymbol{y}$ with $dV_y=dy_1...dy_n$. The functions $H_k, k=1,...,p$ involved in the integrand of the last term in solution (2.28) can be expressed in terms of Green's function G and its choice is based upon the three main types of boundary conditions as follows

1. Dirichlet: $H_k=-G_{M_y}:=-\sum_{i,j=1}^{n}\overline{a}_{ij}(\boldsymbol{y},\tau)N_jG_{y_i}$, N_j is the j-th component of the unit outward normal to S_k;

2. Neumann: $H_k=G$;

3. Robin: $H_k=G$.

Note if the coefficients of (2.29) and the boundary condition (2.31) are independent of t, then the Green's function reduces to only three arguments and $G(\boldsymbol{x},\boldsymbol{y},t,\tau)=G(\boldsymbol{x},\boldsymbol{y},t-\tau)$.

For the problem (2.17)-(2.23) we have that $u=\hat{c}$, $n=2$, $V=\Omega$, $p=4$, $S_1=\{\hat{\boldsymbol{x}}|\hat{x_1}=0, 0<\hat{x_2}<\frac{M}{x_c}\}$, $S_2=\{\hat{\boldsymbol{x}}|\hat{x_1}=\frac{L}{x_c}, 0<\hat{x_2}<\frac{M}{x_c}\}$, $S_3=\{\hat{\boldsymbol{x}}|\hat{x_2}=0, 0<\hat{x_1}<\frac{L}{x_c}\}$, $S_4=\{\hat{\boldsymbol{x}}|\hat{x_2}=\frac{M}{x_c}, 0<\hat{x_1}<\frac{L}{x_c}\}$, $\overline{a}_{11}=\overline{a}_{22}=1$, $\overline{a}_{12}=\overline{a}_{21}=\overline{b}_1=\overline{b}_2=\overline{c}=0$, $L_{\hat{\boldsymbol{x}}}=\Delta\hat{c}$, $\overline{\Phi}=\hat{F}$, $\Gamma_{\hat{\boldsymbol{x}}}=\hat{c}$, $\overline{f}=\hat{f}$, $\overline{g_1}=\hat{g_1}$, $\overline{g_2}=\hat{g_2}$, $\overline{g_3}=\hat{g_3}$, $\overline{g_4}=\hat{g_4}$. Since the coefficients in $L_{\hat{\boldsymbol{x}}}$ and $\Gamma_{\hat{\boldsymbol{x}}}$ are independent of $\hat{t}$ then $G=G(\hat{\boldsymbol{x}},\boldsymbol{y},\hat{t}-\tau)$. Also since the boundary conditions are Dirichlet then

$$H_1=-\sum_{i,j=1}^{2}\overline{a}_{ij}(\boldsymbol{y},\tau)N_jG_{y_i}=-(N_1G_{y_1}+N_2G_{y_2})=G_{y_2}|_{y_1=0}, \tag{2.32}$$

$$H_2=-\sum_{i,j=1}^{2}\overline{a}_{ij}(\boldsymbol{y},\tau)N_jG_{y_i}=-(N_1G_{y_1}+N_2G_{y_2})=-G_{y_2}|_{y_1=\frac{L}{x_c}}, \tag{2.33}$$

$$H_3=-\sum_{i,j=1}^{2}\overline{a}_{ij}(\boldsymbol{y},\tau)N_jG_{y_i}=-(N_1G_{y_1}+N_2G_{y_2})=G_{y_1}|_{y_2=0}, \tag{2.34}$$

$$H_4=-\sum_{i,j=1}^{2}\overline{a}_{ij}(\boldsymbol{y},\tau)N_jG_{y_i}=-(N_1G_{y_1}+N_2G_{y_2})=-G_{y_1}|_{y_2=\frac{M}{x_c}}. \tag{2.35}$$

Next we solve for Green's function via the IBVP (2.29)-(2.31) which can be written as

$$G_t=\Delta_{\hat{\boldsymbol{x}}}G, \tag{2.29}$$

$$G=\delta(\hat{x_1}-y_1)\delta(\hat{x_2}-y_2) \text{ at } t=\tau, \tag{2.30}$$

$$G=0, \hat{\boldsymbol{x}}\in S_k, k=1, ..., 4. \tag{2.31}$$

Assume by separation of variables that

$$G:=\phi(\hat{\boldsymbol{x}}, \boldsymbol{y})\eta(\hat{t}-\tau)$$

and insert into (2.29) to acquire

$$\phi(\hat{\boldsymbol{x}}, \boldsymbol{y})\eta'(\hat{t}-\tau)=\Delta_{\hat{\boldsymbol{x}}}\phi(\hat{\boldsymbol{x}}, \boldsymbol{y})\eta(\hat{t}-\tau).$$

Divide the above by $\phi(\hat{\boldsymbol{x}}, \boldsymbol{y})\eta(\hat{t}-\tau)$ to obtain

$$\frac{\eta'(\hat{t}-\tau)}{\eta(\hat{t}-\tau)}=\frac{\Delta_{\hat{\boldsymbol{x}}}\phi(\hat{\boldsymbol{x}}, \boldsymbol{y})}{\phi(\hat{\boldsymbol{x}}, \boldsymbol{y})}.$$

Since the left hand side (LHS) of the above only depends on $\hat{t}-\tau$ and the right hand side (RHS) only on ($\hat{\boldsymbol{x}}$, $\boldsymbol{y}$) then set each side equal to some separation constant, -λ. This results in the differential equations

$$\eta'(\hat{t}-\tau)+\lambda\eta(\hat{t}-\tau)=0, \tag{2.36}$$

$$\Delta_{\hat{\boldsymbol{x}}}\phi(\hat{\boldsymbol{x}}, \boldsymbol{y})+\lambda\phi(\hat{\boldsymbol{x}}, \boldsymbol{y})=0. \tag{2.37}$$

Taking under consideration the assumed form of G and $\eta(\hat{t}-\tau)\neq 0$, the BC (2.31) becomes

$$\phi(\hat{\boldsymbol{x}}, \boldsymbol{y})=0, \hat{\boldsymbol{x}}\in S_k, k=1, ..., 4 \tag{2.38}$$

The equations (2.37)-(2.38) form a BVP which can be solved using separation of variables. Assume

$$\phi(\hat{\boldsymbol{x}}, \boldsymbol{y}):=\chi(\hat{x_1}, y_1)\psi(\hat{x_2}, y_2)$$

and insert into (2.37) to acquire

$$\chi_{\hat{x_1}\hat{x_1}}(\hat{x_1}, y_1)\psi(\hat{x_2}, y_2)+\chi(\hat{x_1}, y_1)\psi_{\hat{x_2}\hat{x_2}}(\hat{x_2}, y_2)+\lambda\chi(\hat{x_1}, y_1)\psi(\hat{x_2}, y_2)=0.$$

Subtracting the above by the second and third terms of the LHS and then dividing the result by $\chi(\hat{x_1}, y_1)\psi(\hat{x_2}, y_2)$ yields

$$\frac{\chi_{\hat{x_1}\hat{x_1}}(\hat{x_1}, y_1)}{\chi(\hat{x_1}, y_1)} = -\left(\frac{\psi_{\hat{x_2}\hat{x_2}}(\hat{x_2}, y_2)}{\psi(\hat{x_2}, y_2)} + \lambda\right).$$

The LHS of the above equation depends only on $(\hat{x_1}, y_1)$ while the RHS only on $(\hat{x_2}, y_2)$. Thus setting both sides equal to a separation constant, -μ, acquires the following ordinary differential equations (ODEs)

$$\chi_{\hat{x_1}\hat{x_1}}(\hat{x_1}, y_1) + \mu\chi(\hat{x_1}, y_2) = 0, \tag{2.39}$$

$$\psi_{\hat{x_2}\hat{x_2}}(\hat{x_2}, y_2) + (\lambda - \mu)\psi(\hat{x_2}, y_2) = 0. \tag{2.40}$$

Insert $\phi(\hat{\boldsymbol{x}}, \boldsymbol{y})$ into the BC (2.38), with the assumption that the trivial solutions of the ODEs are undesirable because it would lead to the spatial part of the PDE problem being zero and hence we would be left with the trivial solution of the PDE problem, to acquire

$$\chi(\hat{x_1}, y_1) = 0, \hat{x_1} \in S_k, k = 1, 2 \tag{2.41}$$

$$\psi(\hat{x_2}, y_2) = 0, \hat{x_2} \in S_k, k = 3, 4. \tag{2.42}$$

Assume by separation of variables that $\chi(\hat{x_1}, y_1) = \hat{X_1}(\hat{x_1})Y_1(y_1)$, $\psi(\hat{x_2}, y_2) = \hat{X_2}(\hat{x_2})Y_2(y_2)$ then (2.43) and (2.44) become

$$Y_1(y_1)(\hat{X_1}''(\hat{x_1}) + \mu\hat{X_1}(\hat{x_1})) = 0$$

$$Y_2(y_2)(\hat{X_2}''(\hat{x_2}) + (\lambda - \mu)\hat{X_2}(\hat{x_2})) = 0.$$

Since $\chi(\hat{x_1}, y_1) = 0, \psi(\hat{x_2}, y_2) = 0$ is undesirable, as it would lead to a trivial solution for the PDE problem, then it must be that $Y_1(y_1), Y_2(y_2) \in \mathbb{R}$ and

$$\hat{X_1}''(\hat{x_1}) + \mu\hat{X_1}(\hat{x_1}) = 0 \tag{2.43}$$

$$\hat{X_2}''(\hat{x_2}) + (\lambda - \mu)\hat{X_2}(\hat{x_2}) = 0 \tag{2.44}$$

Insert $\chi(\hat{x_1}, y_1), \psi(\hat{x_2}, y_2)$ into the BC (2.41) and (2.42), respectively, with the assumption that the trivial solutions of the ODE's are undesirable since it would lead to a trivial solution for the PDE problem, to acquire

$$\hat{X_1}(\hat{x_1}) = 0, \hat{x_1} \in S_k, k = 1, 2 \tag{2.45}$$

$$\hat{X_2}(\hat{x_2}) = 0, \hat{x_2} \in S_k, k = 3, 4 \tag{2.46}$$

Two Sturm-Liouville problems (SLP), namely (2.43, 2.45) and (2.44, 2.46) have resulted from the separation of variables. Since (2.44, 2.46) depends on two separation constants, (λ, μ), and (2.43, 2.45) only on the constant μ then (2.43, 2.45) must be solved first. The equation (2.43) is a homogeneous second order linear ODE with constant coefficients in the canonical form: $\tilde{a}v''(\xi)+\tilde{b}v'(\xi)+\tilde{c}v(\xi)=0$. Therefore the solution can be classified using the discriminant of the characteristic equation $\tilde{a}\nu^2+\tilde{b}\nu+\tilde{c}=0$, whereby, if it is:

$$\begin{aligned}
&\text{strictly positive then } && v(\xi)=K_1\cosh(\nu_1\xi)+K_2\sinh(\nu_2\xi),\\
&\text{strictly negative then } && v(\xi)=e^{\mathrm{Re}(\nu)\xi}\left(K_1\cos(\mathrm{Im}(\nu)\xi)+K_2\sin(\mathrm{Im}(\nu)\xi)\right),\\
&\text{zero then} && v(\xi)=K_1e^{\nu\xi}+K_2\xi e^{\nu\xi}.
\end{aligned}$$

Applying the above solution method to equation (2.43) we have that $\tilde{a}=1$, $\tilde{b}=0$, $\tilde{c}=\mu$, therefore the discriminant is strictly negative and $\nu=\pm\sqrt{\mu}i, \mu>0$. Thus the general solution is given by

$$\hat{X_1}(\hat{x_1})=K_1\cos(\sqrt{\mu}\hat{x_1})+K_2\sin(\sqrt{\mu}\hat{x_1}),\ K_1, K_2\in\mathbb{R}.$$

The first BC in (2.45) implies that $K_1=0$, so that

$$\hat{X_1}(\hat{x_1})=K_2\sin(\sqrt{\mu}\hat{x_1}).$$

The second BC in (2.45) results in $K_2\sin\left(\sqrt{\mu}\frac{L}{x_c}\right)=0$ which implies that either $K_2=0$ or $\sin\left(\sqrt{\mu}\frac{L}{x_c}\right)=0$. Choose $K_2=1$ and set $\sin\left(\sqrt{\mu}\frac{L}{x_c}\right)=0$, so that $\hat{X_1}(\hat{x_1})\neq 0$, to arrive at the solution

$$\begin{aligned}
&\hat{X_1}^{(n)}(\hat{x_1})=\sin(\sqrt{\mu_n}\hat{x_1}),\\
&\mu_n=\frac{n^2\pi^2x_c^2}{L^2}, n=1,2,3,....
\end{aligned}$$

Note that the SLP (2.44, 2.46) differs from the SLP (2.43, 2.45) only in that the characteristic equation has $\tilde{c}=\lambda-\mu_n$ and thus the solution is

$$\begin{aligned}
&\hat{X_2}^{(m)}(\hat{x_2})=\sin(\sqrt{\lambda_{nm}-\mu_m}\hat{x_2}),\\
&\lambda_{nm}=\pi^2x_c^2\left(\frac{n^2}{L^2}+\frac{m^2}{M^2}\right), n,m=1,2,3,....
\end{aligned}$$

Therefore the solution of the BVP (2.37)-(2.38) is given by

$$\phi_{nm}(\hat{\boldsymbol{x}},\boldsymbol{y})=\sin\left(\frac{n\pi\hat{x_1}x_c}{L}\right)Y_1^{(n)}(y_1)\sin\left(\frac{m\pi\hat{x_2}x_c}{M}\right)Y_2^{(m)}(y_2), \tag{2.47}$$

$$\lambda_{nm}=\pi^2x_c^2\left(\frac{n^2}{L^2}+\frac{m^2}{M^2}\right), n,m=1,2,3,... \tag{2.48}$$

Thus the solution of the BVP (2.29, 2.31) is

$$G_{nm}(\hat{\boldsymbol{x}},\boldsymbol{y},\hat{t}-\tau)=\phi_{nm}(\hat{\boldsymbol{x}},\boldsymbol{y})\eta_{nm}(\hat{t}-\tau).$$

Since $G_{nm}(\hat{\boldsymbol{x}},\boldsymbol{y},\hat{t}-\tau)$ is a set of infinite solutions and a superposition of solutions is also a solution then $G_{nm}(\hat{\boldsymbol{x}},\boldsymbol{y},\hat{t}-\tau)$ can be written as an infinite sum, *i.e.*,

$$G(\hat{\boldsymbol{x}},\boldsymbol{y},\hat{t}-\tau)=\sum_{n,m=1}^{\infty}\phi_{nm}(\hat{\boldsymbol{x}},\boldsymbol{y})\eta_{nm}(\hat{t}-\tau). \tag{2.49}$$

Now $\eta_{nm}(\hat{t}-\tau)$ must be solved via the ODE

$$\eta'_{nm}(\hat{t}-\tau)+\lambda_{nm}\eta_{nm}(\hat{t}-\tau)=0. \tag{2.36}$$

This equation is a first order linear ODE which can be solved by multiplying it by an integrating factor, so that the LHS is the result of an application of the product rule and thus can be written as a derivative of the product between the dependent term, $\eta_{nm}(\hat{t}-\tau)$, and the integrating factor, after which it can be integrated w.r.t. the independent variable, $\hat{t}$. Using this technique with the integrating factor $e^{\lambda_{nm}(\hat{t}-\tau)}$ obtains

$$e^{\lambda_{nm}(\hat{t}-\tau)}\eta_{nm}(\hat{t}-\tau)=K_{nm}, K_{nm}\in\mathbb{R}.$$

Multiplying the above by $e^{-\lambda_{nm}(\hat{t}-\tau)}$ results with the general solution

$$\eta_{nm}(\hat{t}-\tau)=K_{nm}e^{-\lambda_{nm}(\hat{t}-\tau)}, K_{nm}\in\mathbb{R} \tag{2.50}$$

Applying the initial condition (IC) (2.30) to acquire a formulation of K_{nm} which in turn obtains the particular solution, results in

$$G(\hat{\boldsymbol{x}},\boldsymbol{y},0)=\sum_{n,m=1}^{\infty}\eta_{nm}(0)\phi_{nm}(\hat{\boldsymbol{x}},\boldsymbol{y})=\delta(\hat{\boldsymbol{x}}-\boldsymbol{y}).$$

By the rules of exponents e^0=1 it is easily shown that $\eta_{nm}(0)$=K_{nm} and therefore

$$\sum_{n,m=1}^{\infty} K_{nm}\phi_{nm}(\hat{\boldsymbol{x}},\boldsymbol{y})=\delta(\hat{\boldsymbol{x}}-\boldsymbol{y}).$$

If we let $K_{nm}=\frac{4x_c^2}{LM}$, $Y_1(y_1)=\sin\left(\frac{n\pi x_c y_1}{L}\right)$, $Y_2(y_2)=\sin\left(\frac{m\pi x_c y_2}{M}\right)$ then

$$\frac{4x_c^2}{LM}\sum_{n,m=1}^{\infty}\sin\left(\frac{n\pi x_c\hat{x_1}}{L}\right)\sin\left(\frac{m\pi x_c\hat{x_2}}{M}\right)\sin\left(\frac{n\pi x_c y_1}{L}\right)\sin\left(\frac{m\pi x_c y_2}{M}\right)$$
$$=\delta(\hat{\boldsymbol{x}}-\boldsymbol{y})$$

is the Fourier sine series of $\delta(\hat{\boldsymbol{x}}-\boldsymbol{y})$. Thus Green's function is given by

$$G(\hat{\boldsymbol{x}},\boldsymbol{y},\hat{t}-\tau)=\frac{4x_c^2}{LM}\sum_{n,m=1}^{\infty}\phi_{nm}(\hat{\boldsymbol{x}},\boldsymbol{y})\eta_{nm}(\hat{t}-\tau), \tag{2.51}$$

$$\phi_{nm}(\hat{\boldsymbol{x}},\boldsymbol{y})=\sin\left(\frac{n\pi x_c\hat{x_1}}{L}\right)\sin\left(\frac{m\pi x_c\hat{x_2}}{M}\right)\sin\left(\frac{n\pi x_c y_1}{L}\right)\sin\left(\frac{m\pi x_c y_2}{M}\right).$$

Therefore the solution of the IBVP (2.17)-(2.23) is

$$\begin{aligned}\hat{c}(\hat{\boldsymbol{x}},\hat{t})=&\int_0^{\hat{t}}\int_0^{\frac{L}{x_c}}\int_0^{\frac{M}{x_c}}\hat{F}(\boldsymbol{y},\tau)G(\hat{\boldsymbol{x}},\boldsymbol{y},\hat{t}-\tau)dy_1dy_2d\tau\\&+\frac{f_cD}{F_cx_c^2}\int_0^{\frac{L}{x_c}}\int_0^{\frac{M}{x_c}}f(\boldsymbol{y})G(\hat{\boldsymbol{x}},\boldsymbol{y},\hat{t})dy_1dy_2\\&+\frac{g_c^{(1)}D}{F_cx_c^2}\int_0^{\hat{t}}\int_0^{\frac{M}{x_c}}\hat{g}_1(y_2,\tau)H_1(\hat{\boldsymbol{x}},\boldsymbol{y},\hat{t}-\tau)dy_2d\tau\\&+\frac{g_c^{(2)}D}{F_cx_c^2}\int_0^{\hat{t}}\int_0^{\frac{M}{x_c}}\hat{g}_2(y_2,\tau)H_2(\hat{\boldsymbol{x}},\boldsymbol{y},\hat{t}-\tau)dy_2d\tau\\&+\frac{g_c^{(3)}D}{F_cx_c^2}\int_0^{\hat{t}}\int_0^{\frac{L}{x_c}}\hat{g}_3(y_1,\tau)H_3(\hat{\boldsymbol{x}},\boldsymbol{y},\hat{t}-\tau)dy_1d\tau\\&+\frac{g_c^{(4)}D}{F_cx_c^2}\int_0^{\hat{t}}\int_0^{\frac{L}{x_c}}\hat{g}_4(y_1,\tau)H_4(\hat{\boldsymbol{x}},\boldsymbol{y},\hat{t}-\tau)dy_1d\tau\end{aligned}\tag{2.52}$$

$$H_1(\hat{\boldsymbol{x}},\boldsymbol{y},\hat{t}-\tau)=G_{y_2}(\hat{\boldsymbol{x}},\boldsymbol{y},\hat{t}-\tau)|_{y_1=0},\tag{2.32}$$

$$H_2(\hat{\boldsymbol{x}},\boldsymbol{y},\hat{t}-\tau)=-G_{y_2}(\hat{\boldsymbol{x}},\boldsymbol{y},\hat{t}-\tau)|_{y_1=\frac{L}{x_c}},\tag{2.33}$$

$$H_3(\hat{\boldsymbol{x}},\boldsymbol{y},\hat{t}-\tau)=G_{y_1}(\hat{\boldsymbol{x}},\boldsymbol{y},\hat{t}-\tau)|_{y_2=0},\tag{2.34}$$

$$H_4(\hat{\boldsymbol{x}},\boldsymbol{y},\hat{t}-\tau)=-G_{y_1}(\hat{\boldsymbol{x}},\boldsymbol{y},\hat{t}-\tau)|_{y_2=\frac{M}{x_c}},\tag{2.35}$$

$$G(\hat{\boldsymbol{x}},\boldsymbol{y},\hat{t}-\tau)=\frac{4x_c^2}{LM}\sum_{n,m=1}^{\infty}\phi_{nm}(\hat{\boldsymbol{x}},\boldsymbol{y})\eta_{nm}(\hat{t}-\tau),\tag{2.51}$$

$$\phi_{nm}(\hat{\boldsymbol{x}},\boldsymbol{y})=\sin\left(\frac{n\pi x_c\hat{x_1}}{L}\right)\sin\left(\frac{m\pi x_c\hat{x_2}}{M}\right)\sin\left(\frac{n\pi x_cy_1}{L}\right)\sin\left(\frac{m\pi x_cy_2}{M}\right),\tag{2.47}$$

$$\eta_{nm}(\hat{t})=e^{-\lambda_{nm}\hat{t}},\tag{2.50}$$

$$\lambda_{nm}=\pi^2x_c^2\left(\frac{n^2}{L^2}+\frac{m^2}{M^2}\right),\tag{2.48}$$

$$F_c=\max_{\boldsymbol{x},t}|F(\boldsymbol{x},t)|,\tag{2.10}$$

$$g_c^{(1)}=\max_{x_2,t}|g_1(x_2,t)|,\ g_c^{(2)}=\max_{x_2,t}|g_2(x_2,t)|,\tag{2.11}$$

$$g_c^{(3)}=\max_{x_1,t}|g_3(x_1,t)|,\ g_c^{(4)}=\max_{x_1,t}|g_1(x_1,t)|\tag{2.12}$$

$$f_c=\max_{\boldsymbol{x}}|f(\boldsymbol{x})|,\tag{2.13}$$

$$x_c=\max(L,M),\tag{2.18}$$

$$t_c=\frac{x_c^2}{D}.\tag{2.15}$$

When considering $C>1$ carcinogens each carcinogen evolves using the same PDE model described above with each solution and the parameters of the

model being distinguished by an index i=1, ..., C.

2.3 Introduction to Neural Networks

This section is adapted from Géron [55].

2.3.1 History

Artificial neural networks (ANNs) were created to build an intelligent machine, more commonly known as artificial intelligence (AI), based upon the brain's architecture. In 1943 the neurophysiologist Warren McCulloch and mathematician Walter Pitts invented the first ANN architecture which was a simplified computational model for how biological neurons might work in animal brains to perform complex computations using propositional logic [94]. The initial success of these ANNs up until the 1960s led to the hope that there would soon be truly intelligent machines that could be conversed with. However when this hope was not met the development of ANNs stopped due to loss of funding and interest. In the early 1980s development of new network architectures and better computational techniques led to a renewed interest in ANNs. By the 1990s more powerful alternative machine learning techniques such as support vector machines were favoured by researchers as they had better results and stronger theoretical foundations to understand and further develop. Until more recently, the last decade or so, ANNs were not very prevalent but there is now a prominent wave of development and research into them that has led to AI being used almost everywhere. This new wave likely will not stop due to the following reasons:

- a significant quantity of data available to train neural networks,
- ANNs frequently outperform other machine learning (ML) techniques,
- computational power improvements since the 1990s which reduces both training and execution time,
- new training algorithms are perpetually being developed and existing algorithms have been perfected,

- When ANNs are applied to real problems some of the theoretical limitations are not present,
- ANNs have acquired significant funding and progression due to prominent ongoing development of applications for ANNs.

2.3.2 Basic ANN Architectures

McCulloch & Pitts [94] proposed a very simple model of the biological neuron, which later became known as an artificial neuron. An artificial neuron is made up of one or more binary (on/off) inputs and one binary output. The output is activated when a certain number of its inputs is active. They showed that even with this seemingly simple model it was possible to build a network of artificial neurons that could compute any possible logical proposition.

Another simple ANN architecture was invented by Frank Rosenblatt in 1957 that was called the Perceptron. It was based upon an artificial neuron called a linear threshold unit (LTU) in which the inputs and output are now real numbers instead of binary on/off values and each input connection is associated with a weight. The LTU computes a weighted sum of its inputs $z=\boldsymbol{w}^T\cdot\boldsymbol{x}$, then applies a step function to that sum and outputs the result: $h_{\boldsymbol{w}}(\boldsymbol{x})=\text{step}(z)=\text{step}(\boldsymbol{w}^T\cdot\boldsymbol{x})$, reference Figure 2.1 for a visual representation.

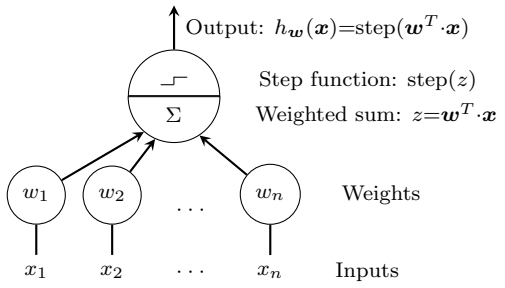


Figure 2.1: In this figure we show a visual representation of a linear threshold unit (LTU). Firstly, a weighted sum of the inputs is computed, followed by an application of a step function. This figure is adapted from Géron [55].

The most common step function used in Perceptrons is the Heaviside step function although sometimes the sign function is used in its place. A single LTU is typically used for linear binary classification as it computes a linear

combination of the inputs and outputs the positive or negative class dependent on whether the inputs exceed some threshold.

A Perceptron is composed of a single layer of LTUs, with each LTU connected to all the inputs, see Figure 2.2 for a visual representation.

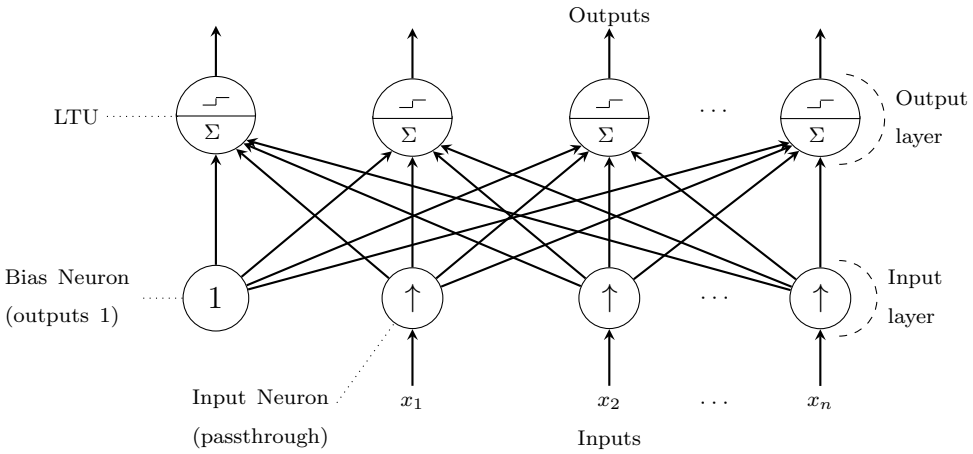


Figure 2.2: In this figure we show a visual representation of a perceptron. Inputs are passed through an input layer (with an added bias neuron), the output of each of the neurons of the input layer are then inputted into each of the LTUs that make up the output layer. This figure is adapted from Géron [55].

Perceptrons make predictions based off some threshold, thus they do not output a class probability. In Minsky & Papert [97] they discuss a number of serious weaknesses of Perceptrons, in particular Perceptrons are incapable of solving some trivial problems. Some of the limitations of Perceptrons can be eliminated by stacking multiple Perceptrons to create what is called a Multi-Layer Perceptron (MLP).

2.3.3 Multi-Layer Perceptron

A Multi-Layer Perceptron (MLP) is composed of one input layer, one or more layers of LTUs called hidden layers, and one final layer of LTUs called the output layer, an example MLP can be seen in Figure 2.3. The input and hidden layers have a bias neuron and are fully connected to the next layer, meaning each output node of one layer is connected to every node of the input of the next layer. Note that when an ANN has two or more hidden layers, it is called a deep neural network (DNN).

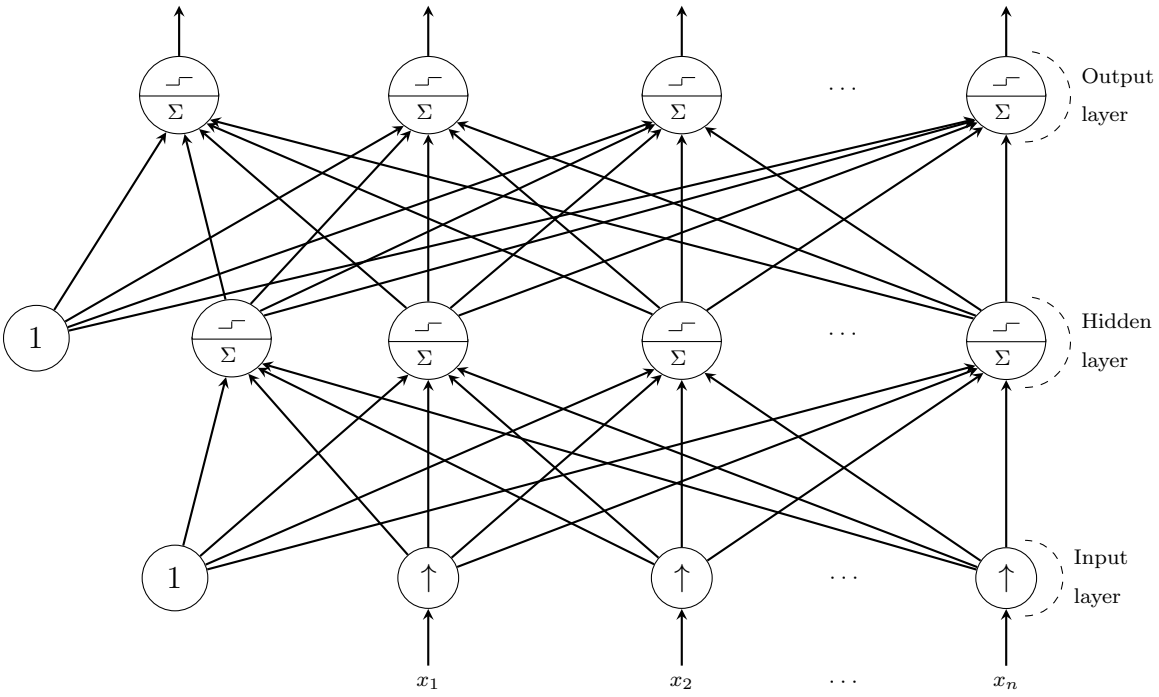


Figure 2.3: This figure shows a simple Multi-Layer Perceptron with one hidden layer. It is adapted from Géron [55].

To train, the weights for each LTU using training data, MLPs use an algorithm called back-propagation, which was created by Rumelhart *et al.* [119]. Back-propagation attempts to minimize the weights in a network so that the measure of the difference between the actual output and the networks output is minimized. The measure for the difference, typically called the loss function, is minimized by computing the gradient of the loss function with respect to each weight by chain rule, computing the gradient one layer at a time, starting at the last layer to avoid redundant calculations of intermediate terms in the chain rule. Gradients are used as it shows how much the input of a function needs to change to minimize the function. Since LTUs use step functions which are not differentiable everywhere Rumelhart *et al.* [119] replaced the step function with the logistic function, $\sigma(z)=\frac{1}{1+\exp(-z)}$. Many other functions can be used and are known as activation functions. Some examples of activation functions include the hyperbolic tangent function $\tanh(z)=2\sigma(2z)-1$ and the ReLU function $\text{ReLU}(z)=\max(0,z)$.

An MLP is typically used for classification, where each output corresponds

to the log likelihood of a different class. The probability that the input belongs to class k can be estimated by computing the exponential of every log likelihood, and normalizing them. The probabilities are thus given by:

$$\hat{p}_k = \frac{\exp(s_k(\boldsymbol{x}))}{\sum_{j=1}^{K} \exp(s_j(\boldsymbol{x}))}.$$

If only the class with the highest probability is desired then compute $\hat{y} = \underset{k}{\operatorname{argmax}}\ \hat{\boldsymbol{p}} = \underset{k}{\operatorname{argmax}}\ s_k(\boldsymbol{x})$.

2.3.4 Designing an Artificial Neural Network

One of the important decisions when designing an ANN architecture is how many hidden layers to include. If the problem at hand does not have inherent structures, such as hats in photos of people, then one hidden layer is sufficient for complex functions, provided it has enough neurons. A deep network can model complex functions using fewer neurons than shallow nets, making them much faster to train so they are more parameter efficient. Sometimes the complex patterns in data can be broken down into a combination of simpler patterns, whereby lower layers model low-level patterns, intermediate layers combine the low-level patterns to model intermediate-level patterns, and the highest layers along with the output layer combine the intermediate patterns to model the high-level patterns. The hierarchical architecture helps DNNs generalize to new data-sets. A trained DNN can be used for another similar task by reusing the parameters of the lower hidden layers and retraining the intermediate and higher-level layers. Thus for many problems when designing an ANN we start with just one or two hidden layers on a simple problem and then for more complex problems we gradually increase the number of hidden layers until over-fitting of the training set starts to occur. Another decision to make when creating an ANN architecture is how many neurons to include in each hidden layer. A prior common technique was to set the number of neurons for each layer such that each successive layer had fewer neurons than the previous layer, which represented the fact that higher-level structures required fewer individual classes to distinguish. However, nowadays,

one simply sets all the layers to have the same number of neurons so that there is just one parameter instead of a parameter for each layer. Another technique is to gradually increase the number of neurons until the network starts over-fitting the training data-set. If the number of layers over the number of neurons per layer is increased the model will become more accurate. Another simpler technique is to pick a model with more layers and neurons than needed, then use early stopping when training to prevent it from over-fitting.

The final important question to consider is what activation function to use for each layer. For the hidden layers, it is common practice to use the ReLU activation function as it is faster to compute than most other functions and gradient descent does not become stuck as much on plateaus. For the output layer, the softmax activation function (which normalizes a vector into some probability distribution so that the sum of the output vector sums to one and each component of the vector can be interpreted as a probability), is generally a good choice for classification tasks that have mutually exclusive classes. When the classes are not mutually exclusive the choice is typically to use the logistic function. When dealing with some tasks, like a regression, that do not involve classification, it is sometimes useful to have no activation function at all in the output layer.

2.4 Gene Expression Neural Network

In this section we describe the neural network we use to mutate genes through cell age and carcinogenic onslaught. We consider $G \in \mathbb{N}$ genes that are biomarkers to the considered cancer type. The gene expression of each gene is represented by the function

$$e_j(\boldsymbol{x}, t) \in \mathbb{R}, j = 1, 2, ..., G. \tag{2.53}$$

The gene expression is a non-dimensional value that is zero when the expression is normal, negative when it is under-expressed, and positive when it is over-expressed. The gene expression of each gene changes over time based upon a

simple multi-layer perceptron (MLP). The input of the MLP is the vector

$$\boldsymbol{X}(\boldsymbol{x},t):=[\{c_i(\boldsymbol{x},t-1)\}_{i=1,\ldots,C},\alpha(\boldsymbol{x},t-1)]^T\in\mathbb{R}_+^{C+1}, \tag{2.54}$$

where $c_i(\boldsymbol{x},t)$ are the carcinogen concentrations and $\alpha(\boldsymbol{x},t)$ is the age of the cell. This choice is such that changes in gene expression is based upon the carcinogens in the environment of the cell and the age, which essentially means we are looking at the effects of the carcinogens and replication errors as a cell ages. The output of the MLP is given by

$$\boldsymbol{Y}(\boldsymbol{x},t):=[\{\overline{\delta}_j(\boldsymbol{x},t)\}_{j=1,2,\ldots,G}]^T\in\mathbb{R}^G, \tag{2.55}$$

where $\overline{\delta}_j(\boldsymbol{x},t)$ is the computed maximum possible change in gene expression for gene j. The amount the gene j will be mutated in a time-step is a random sample from the uniform distribution multiplied by $\overline{\delta}_j(\boldsymbol{x},t)$.

$\boldsymbol{Y}(\boldsymbol{x},t)$ is computed using matrix multiplication, addition and application of a non-linear transform. The hidden layer is computed by

$$\boldsymbol{H}(\boldsymbol{x},t):=\gamma(W_X\boldsymbol{X}(\boldsymbol{x},t))\in\mathbb{R}^G, \tag{2.56}$$

where

$$\gamma(\xi):=\frac{\xi}{\sqrt{1+\nu\xi^2}},\in\left(\frac{\text{-1}}{\sqrt{\nu}},\frac{1}{\sqrt{\nu}}\right) \tag{2.57}$$

is the non-linear transform (also known as an activation function) that is applied element wise to a vector and $W_X\in\mathbb{R}^{G\times C+1}$ is a weight matrix. Note that the activation function is chosen to ensure $|\overline{\delta}_j(\boldsymbol{x},t)|<\frac{1}{\sqrt{\nu}}$, hence allowing us to control the maximum amount the expression of gene j can change in a time-step via ν. After the hidden layer is computed the output is computed by

$$\boldsymbol{Y}(\boldsymbol{x},t)=\gamma(W_Y\boldsymbol{H}(\boldsymbol{x},t)+\boldsymbol{b}_Y(\boldsymbol{x},t)), \tag{2.58}$$

where $W_Y\in\mathbb{R}^{G\times G}$ is a weight matrix and $\boldsymbol{b}_Y(\boldsymbol{x},t)\in\mathbb{R}^G$ is a bias vector.

Biologically speaking $W_X^{(i,j)}, i\in[1,G], j\in[1,C]$ represents how carcinogen i influences gene j, $W_X^{(i,C+1)}, i\in[1,G]$ represents whether cell age influences gene j, $W_Y^{(i,j)}$ represents whether gene i influences gene j, and $b_Y^{(i)}(\boldsymbol{x},t)$ is whether gene i has a higher chance of gene expression changes relative to other genes.

Note that if a value in the weight matrices is negative it means there is a negative relationship, if it is positive it means there is a positive relationship, and finally if it is zero it means there is no relationship. In the case of determining how age affects each gene, the values of $W_X^{(i,C+1)}, i{\in}[1, G]$ are randomly made positive or negative at every time-step based upon sampling from the uniform distribution and setting it positive if the sample is less than 0.5 and negative otherwise. We randomly choose the direction cell age regulates gene expression because as a cell ages there are higher changes of gene replication errors, thus the direction will depend on the type of error.

Let U(0, 1) be the uniform distribution. The gene expression, $e_j(\boldsymbol{x}, t)$, of a gene is updated by

$$e_j(\boldsymbol{x}, t){=}e_j(\boldsymbol{x}, t{-}1){+}z\bar{\delta}_j(\boldsymbol{x}, t), z{\sim}U, \tag{2.59}$$

where we use $z{\sim}U$ to indicate that z is sampled from U. A gene j is considered to be mutated if its' gene expression is above the threshold value $\overline{M}{\in}\mathbb{R}_+$, *i.e.*, $|e_j(\boldsymbol{x}, t)|{\geq}\overline{M}$. The bias for a gene j, $b_Y^{(j)}(\boldsymbol{x}, t)$, is updated through the relation

$$b_Y^{(j)}(\boldsymbol{x}, t){=}\begin{cases}\beta & , e_j(\boldsymbol{x}, t{-}1){\geq}\overline{M} \\ -\beta & , e_j(\boldsymbol{x}, t{-}1){\leq}{-}\overline{M} \\ 0 & , \text{otherwise}\end{cases}, \beta{\in}\mathbb{R}_+. \tag{2.60}$$

2.5 Cellular Automaton

As mentioned in the model overview, the six cell classes that we consider are normal tissue cells (NTC), mutated normal tissue cells (MNTC), normal stem cells (NSC), mutated normal stem cells (MNSC), cancer stem cells (CSC), and tumour cells (TC). Each cell class is specified in the cellular automata (CA) using a numerical value between 0 and 5. More precisely, we have that 0=NTC, 1=MNTC, 2=NSC, 3=MNSC, 4=CSC, 5=TC. Since biological cells can move, proliferate, differentiate, and go through apoptosis then we must also introduce an empty cell class which is represented by the value 6. The cell class in the CA is represented by $s(\boldsymbol{x}, t){\in}\{0, 1, ..., 6\}$. Note that when visualizing the CA each value of $s(\boldsymbol{x}, t)$ also has a colour associated to it.

Though we can choose any arbitrary colour for each cell class the colours chosen for the results of the model can be seen in table 2.1.

Cell class	$s(\boldsymbol{x},t)$	Colour
normal tissue cell (NTC)	0	brown
mutated normal tissue cell (MNTC)	1	green
normal stem cell (NSC)	2	blue
mutated normal stem cell (MNSC)	3	yellow
cancer stem cell (CSC)	4	purple
tumour cell (TC)	5	red
empty cell	6	white

Table 2.1: CA cell classes

Each cell in the CA tracks the gene expression of the G genes in a vector defined by

$$\boldsymbol{E}(\boldsymbol{x},t)=[\{e_j(\boldsymbol{x},t)\}_{j=1,...,G}]. \tag{2.61}$$

The phenotype of a cell is tracked by a vector that contains probabilities for each type of phenotypic action occurring in a given time-step and is defined by

$$\boldsymbol{P}(\boldsymbol{x},t)=[p(\boldsymbol{x},t),q(\boldsymbol{x},t),a(\boldsymbol{x},t),d(\boldsymbol{x},t)], \tag{2.62}$$

where $p(\boldsymbol{x},t)$ represents proliferation, $q(\boldsymbol{x},t)$ represents quiescence, $a(\boldsymbol{x},t)$ represents apoptosis, and $d(\boldsymbol{x},t)$ represents differentiation. The probabilities are set such that $\boldsymbol{P}(\boldsymbol{x},t)$ generates a probability distribution, so that

$$\sum_{i=1}^{4} P_i(\boldsymbol{x},t)\equiv p(\boldsymbol{x},t)+q(\boldsymbol{x},t)+a(\boldsymbol{x},t)+d(\boldsymbol{x},t)=1 \text{ and } P_i(\boldsymbol{x},t)\geq 0, \forall t. \tag{2.63}$$

At a time-step in the CA a phenotypic action is chosen to occur by sampling from the probability distribution generated from $\boldsymbol{P}(\boldsymbol{x},t)$. Hence, since we do not want a cell to reproduce more than once in a time-step, each time-step represents the length of a typical cell cycle for the type of tissue under consideration.

When a NSC, MNSC, or CSC differentiate the resultant cell initially is a transient amplifying cell (TAC) for a set number of generations, Θ, after which it turns respectively into a NTC, MNTC, or TC. As a result of this each cell has two parameters $\overline{\tau}(\boldsymbol{x},t)\in\{0,1\}$ and $\overline{n}(\boldsymbol{x},t)\in\{0,...,\Theta\}$, where $\overline{\tau}(\boldsymbol{x},t)$ is a

binary parameter used to determine if a cell is currently a TAC or not and $\overline{n}(\boldsymbol{x},t)$ is the number of generations a TAC cell lineage has produced. The parameters $\overline{\tau}(\boldsymbol{x},t)$ and $\overline{n}(\boldsymbol{x},t)$ are copied from parent to child cell and once $\overline{n}(\boldsymbol{x},t)$=$\Theta$ then $\overline{\tau}(\boldsymbol{x},t+1)$=0, $\overline{n}(\boldsymbol{x},t+1)$=0.

The final aspect of the cell that is tracked and represented in the overall cell state is the age of the cell, $\alpha(\boldsymbol{x},t)\in\mathbb{N}$. The state of a cell in the CA is given by the vector

$$\boldsymbol{S}(\boldsymbol{x},t)=[s(\boldsymbol{x},t),\alpha(\boldsymbol{x},t),\boldsymbol{E}(\boldsymbol{x},t),\boldsymbol{P}(\boldsymbol{x},t),\overline{\tau}(\boldsymbol{x},t),\overline{n}(\boldsymbol{x},t)] \tag{2.64}$$

The domain could theoretically be either two-dimensional or three-dimensional, however for simplicity and computational purposes, the domain considered will be two-dimensional, so $\boldsymbol{x}\in\mathbb{R}^2$. Each cell has a neighbourhood that contains itself, the cardinal directions around it, and the cells directly NE, SE, SW, and NW of the cell. In CA theory this is called the Moore neighbourhood and is mathematically defined for two-dimensional grids as

$$N^M_{(x_0,y_0)}=\{(x,y)||x-x_0|\leq r,|y-y_0|\leq r\}, \tag{2.65}$$

where r is the range of the Moore neighbourhood and (x_0,y_0) is the cell that the neighbourhood surrounds [59]. Most commonly the Moore neighbourhood used is with r=1 and it will be the one used in this thesis.

The boundary conditions of the grid are standard periodic boundary conditions. Periodic boundaries were chosen because they are easier to work with and the size of the domain physically speaking is minuscule so they will not effect the results. We could easily also use non-periodic boundaries by dealing with edge cases.

2.5.1 Cell Mutation

A cell can become mutated through changes in both the likelihood of a phenotype occurring and the gene expression. The process of these changes is described in this section.

2.5.1.1 How Mutations Effect Gene Expression

The chosen G genes are known genes related to the type of cancer being studied. Thus, each of the G genes is either a tumour suppressor gene or an oncogene. Define the vector $\boldsymbol{T}\in\{0,1\}^G$, where each element, T_j, represents gene j's type, whereby T_j=0 represents a tumour suppressor gene and T_j=1 represents an oncogene.

A gene j is positively mutated towards cancer (positively mutated) if it is mutated and either it is a tumour suppressor gene and its gene expression is downregulated, T_j=0 and $e_j(\boldsymbol{x},t)\leq -\overline{M}$, or it is an oncogene and its gene expression is upregulated, T_j=1 and $e_j(\boldsymbol{x},t)\geq\overline{M}$, where $\overline{M}$ is the given threshold.

At each time-step the gene expression of each gene is updated from the results of the gene expression neural network. The changes in the gene expression allow the gene to become mutated or even go from mutated to non mutated (normally expressed). The following function is used as an indicator to determine if a gene j is mutated, positively mutated or normally expressed:

$$\Psi(\boldsymbol{x},t,j)=\begin{cases}0, & |e_j(\boldsymbol{x},t)|<\overline{M}(\text{normally expressed})\\ 1-2T_j, & e_j(\boldsymbol{x},t)\leq-\overline{M}(\text{underexpressed gene})\\ -2+4T_j, & e_j(\boldsymbol{x},t)\geq\overline{M}(\text{overexpressed gene})\end{cases} \tag{2.66}$$

Notice that when $\Psi(\boldsymbol{x},t,j)>0$ the gene j is positively mutated and when $\Psi(\boldsymbol{x},t,j)<0$ it is mutated away from cancer. Further take note that if

$$|\Psi(\boldsymbol{x},t,j)|=1$$

the gene j is downregulated and when

$$|\Psi(\boldsymbol{x},t,j)|=2$$

it is upregulated.

2.5.1.2 Update Rules for Gene Expression

A given gene can influence the expression of another gene as follows. A positively mutated gene will cause a positive mutation of a related gene. A non-positively mutated gene will cause a negative mutation (mutation that regulates a gene towards normal expression) of a related gene. Define the matrix

$R \in \{0,1\}^{G \times G}$, where each entry, R_{ij}, represents whether gene i is related to gene j with 0=unrelated and 1=related. Note that the matrix R is not necessarily symmetric as a gene i might regulate gene j but not vice versa. To update gene j according to formula (2.68) below, we chose a random number $z \sim U(0,1)$ and an update occurs only if $z \leq \overline{\gamma}$ and $R_{ij}=1$. The process of a gene i changing the gene expression of another gene j is represented by the formula:

$$\varsigma(\boldsymbol{x},t,i,j)=\begin{cases} 1, & (\Psi(\boldsymbol{x},t,i)>0 \text{ and } T_j=1 \text{ and } e_j(\boldsymbol{x},t)<M) \text{ or} \\ & (\Psi(\boldsymbol{x},t,i)\leq 0 \text{ and } e_j(\boldsymbol{x},t)<0) \\ -1, & (\Psi(\boldsymbol{x},t,i)>0 \text{ and } T_j=0 \text{ and } e_j(\boldsymbol{x},t)>-M) \text{ or} \\ & (\Psi(\boldsymbol{x},t,i)\leq 0 \text{ and } e_j(\boldsymbol{x},t)>0) \\ 0, & \text{otherwise} \end{cases}, \tag{2.67}$$

$$e_j(\boldsymbol{x},t)=e_j(\boldsymbol{x},t-1)+\varsigma(\boldsymbol{x},t-1,i,j)\varepsilon, \tag{2.68}$$

where $\varepsilon \sim U(\varepsilon_1,\varepsilon_2)$ is a randomly chosen increment between ε_1 and ε_2. The previous update is used by fixing a gene i and applying the function to all the other genes, then repeating the process on the next gene and so forth, until all the genes have been processed. In each of these steps new random numbers z, ε are generated.

In addition if a gene j is mutated then there is a chance that the gene expression is negatively mutated, so to replicate the bodies attempting to revert mutated genes. The following function is applied only when some random variable $z \sim U(0,1)$ is less than or equal to a threshold $\overline{\phi}$. This is represented by the gene repair function:

$$e_j(\boldsymbol{x},t)=e_j(\boldsymbol{x},t-1)+\begin{cases} \varepsilon, & |\Psi(\boldsymbol{x},t-1,j)|=1 \\ -\varepsilon, & |\Psi(\boldsymbol{x},t-1,j)|=2 \end{cases}, \tag{2.69}$$

where $\varepsilon \sim U(\varepsilon_1,\varepsilon_2)$ is a random increment. In a given time-step the previous function (2.69) is applied once to all G genes.

2.5.1.3 Update Rules for Phenotypic Action

The probability of a phenotypic action occurring can change at each time-step based upon the gene expression of a gene. When a gene is mutated it can

modify the probability of a phenotypic action occurring, $P_i(\boldsymbol{x},t)$. We define the matrix $\overline{U}\in\mathbb{R}^{4\times G}$, where each entry, $\overline{U}_{ij}$, is an increment to the probability of phenotypic action i, $P_i(\boldsymbol{x},t)$, under the circumstance that gene j is mutated and its' expression is upregulated. Similarly, we define the matrix $\overline{D}\in\mathbb{R}^{4\times G}$, where each entry, $\overline{D}_{ij}$, is an increment to the probability of phenotype action i, $P_i(\boldsymbol{x},t)$, under the circumstance that gene j is mutated and its' expression is downregulated. We define updates for the phenotypic actions in (2.72) and (2.73) in a way such that the probability of a phenotypic action occurring is kept bounded between 0 and 1 and $\sum_{i=1}^{4} P_i(\boldsymbol{x},t)=1$.

The sum of the phenotype vector equaling one is maintained by balancing the probability of each phenotype action against the probability of quiescence and quiescence equally against all the other phenotypic actions. Note that usually $\overline{U}=-\overline{D}$ since the effects of the up-regulation of a gene j has the opposite effect on the phenotypic action i relative to the down-regulation of the gene. For each update we choose a random sample $\varpi\sim U(0,1)$ which denotes the magnitude of the change to P_i based on a change in e_j. To ensure that updates remain between 0 and 1 we use the following modified expression for the increment:

$$\Xi(\epsilon,\xi)=\begin{cases}-\xi\times 0.99, & \xi+\epsilon\leq 0\\ (1-\xi)\times 0.99, & \xi+\epsilon\geq 1\\ \epsilon, & \text{otherwise}\end{cases}, \tag{2.70}$$

Given the mutated gene j we change the phenotypic probability P_i with the following formulas, where Φ defines the maximum increment or detriment amount:

$$\Phi(\boldsymbol{x},t,i)=\begin{cases}\overline{D}_{ij}, & |\Psi(\boldsymbol{x},t,j)|=1\\ \overline{U}_{ij}, & |\Psi(\boldsymbol{x},t,j)|=2\end{cases} \tag{2.71}$$

Then the update for P_i ensures that the new probabilities remain between 0 and 1.

$$\begin{aligned}P_i(\boldsymbol{x},t)=&P_i(\boldsymbol{x},t-1)\\ &+\begin{cases}\operatorname{sgn}(\Phi(\boldsymbol{x},t,i))\min(|\Xi(\varpi\Phi(\boldsymbol{x},t,i),P_i(\boldsymbol{x},t-1))|, & ,i\neq 2\\ \qquad\qquad |\Xi(-\varpi\Phi(\boldsymbol{x},t,i),P_2(\boldsymbol{x},t-1))|) & \\ \Xi(\varpi\Phi(\boldsymbol{x},t,i),P_2(\boldsymbol{x},t-1)) & ,i=2\end{cases},\end{aligned} \tag{2.72}$$

These changes are balanced with changes to the other phenotype actions $k{\neq}i$ to ensure full probability.

$$P_k(\boldsymbol{x},t){=}P_k(\boldsymbol{x},t{-}1) - \begin{cases} \operatorname{sgn}(\Phi(\boldsymbol{x},t,i))\min(|\Xi(\varpi\Phi(\boldsymbol{x},t,i),P_i(\boldsymbol{x},t{-}1))|, & ,k{=}2 \\ \qquad\qquad |\Xi(-\varpi\Phi(\boldsymbol{x},t,i),P_k(\boldsymbol{x},t{-}1))|) & \\ \dfrac{\Xi(\varpi\Phi(\boldsymbol{x},t,i),P_i(\boldsymbol{x},t{-}1))P_k(\boldsymbol{x},t{-}1)}{\sum_{l\neq 2}P_l(\boldsymbol{x},t{-}1)} & ,k{\neq}2 \end{cases}, \quad (2.73)$$

The equation (2.72) modifies a phenotypic action with respect to a gene j, with equation (2.73) balancing those changes so that the sum of the probability distribution generated by $\boldsymbol{P}(\boldsymbol{x},t)$ remains one. Thus, to mutate a phenotypic action i with respect to gene j, we use equation (2.72) then if $i{\neq}2$, we are not modifying quiescence, and we use equation (2.73) with $k{=}2$, otherwise we use equation (2.73) with $k{=}1,3,4$. Since $|\Psi(\boldsymbol{x},t,j)|{=}1$ means that gene j is down-regulated, then we know that in that case the increment for the phenotypic action i with respect to gene j will be D_{ij}. Similarly, whenever $|\Psi(\boldsymbol{x},t,j)|{=}2$ then the increment for the phenotypic action i will be U_{ij}. Mathematically speaking this update process is accomplished through a composition of applications of (2.72) and (2.73). We define the functional $\overline{P}^j(\boldsymbol{x},t,i)$ where i is the phenotypic action being modified using (2.72) and (2.73) with respect to gene j. Thus the update can be written as $P_i(\boldsymbol{x},t) = \overline{P}^1(\overline{P}^2(...\overline{P}^G(\boldsymbol{x},t,i)))$, where in each step we re-sample random variables z,ϖ as described above.

Lemma 2.5.1. *Let* $P_l(\boldsymbol{x},t{-}1){\in}(0,1), l{=}1,2,3,4$ *and* $\sum_{l=1}^4 P_l(\boldsymbol{x},t{-}1){=}1$. *If the probability* P_i *of phenotypic action* $i \in \{1,2,3,4\}$ *is modified using the rules* (2.72) *and* (2.73) *with respect to gene* j *then we will still have* $\sum_{l=1}^4 P_l(\boldsymbol{x},t){=}1$, $\forall t{>}0$.

Proof. Assume $\Psi(\boldsymbol{x},t,j){\neq}0$ and the threshold for gene j to modify the probability of phenotypic action i is met, *i.e.*, $z{\sim}Z(0,1){\leq}\overline{\rho}$. Also since the increment being applied to the probability of phenotypic action i has only two possible values related to whether $|\Psi(\boldsymbol{x},t,j)|{=}1$ or $|\Psi(\boldsymbol{x},t,j)|{=}2$, then we can assume without loss of generality that the increment is some value ϵ. If $\epsilon{=}0$ then trivially we have that $\sum_{l=1}^4 P_l(\boldsymbol{x},t){=}1$ since no changes will occur to $P_i(\boldsymbol{x},t{-}1)$.

Case 1: First we assume that $i \neq 2$ and $k=2$ as we are balancing the change in the probability of phenotypic action i with the probability of quiescence. Further assume that $P_l(\boldsymbol{x},t)=P_l(\boldsymbol{x},t-1)$, $l \neq i,2$.

Case A: Assume $\epsilon<0$.

Case a: Assume $0<P_i(\boldsymbol{x},t-1)+\epsilon<1$ and $0<P_2(\boldsymbol{x},t-1)-\epsilon<1$. We have that

$$\begin{aligned}&\operatorname{sgn}(\epsilon)\min(|\Xi(\epsilon,P_i(\boldsymbol{x},t-1))|,|\Xi(-\epsilon,P_2(\boldsymbol{x},t-1))|)\\&=\operatorname{sgn}(\epsilon)\min(|\epsilon|,|-\epsilon|)=\operatorname{sgn}(\epsilon)|\epsilon|=\epsilon. \qquad (1)\end{aligned}$$

Using equation (2.72) and considering (1) we have that

$$P_i(\boldsymbol{x},t)=P_i(\boldsymbol{x},t-1)+\epsilon. \qquad (2)$$

Also using equation (2.73) and considering again (1) we have that

$$P_2(\boldsymbol{x},t)=P_2(\boldsymbol{x},t-1)-\epsilon. \qquad (3)$$

Therefore, using equations (2) and (3) we have

$$\begin{aligned}\sum_{l=1}^{4}P_l(\boldsymbol{x},t)&=\sum_{\substack{l=1\\l\neq i,2}}^{4}P_l(\boldsymbol{x},t)+(P_i(\boldsymbol{x},t-1)+\epsilon)+(P_2(\boldsymbol{x},t-1)-\epsilon)\\&=\sum_{\substack{l=1\\l\neq i,2}}^{4}P_l(\boldsymbol{x},t)+P_i(\boldsymbol{x},t-1)+P_2(\boldsymbol{x},t-1)\\&=\sum_{\substack{l=1\\l\neq i,2}}^{4}P_l(\boldsymbol{x},t-1)+P_i(\boldsymbol{x},t-1)+P_2(\boldsymbol{x},t-1)\\&=\sum_{l=1}^{4}P_l(\boldsymbol{x},t-1)=1.\end{aligned}$$

Case b: Assume $P_i(\boldsymbol{x},t-1)+\epsilon \leq 0$ and $0<P_2(\boldsymbol{x},t-1)-\epsilon<1$. Since $P_i(\boldsymbol{x},t-1) \leq -\epsilon$ we have

$$\begin{aligned}&\operatorname{sgn}(\epsilon)\min(|\Xi(\epsilon,P_i(\boldsymbol{x},t-1))|,|\Xi(-\epsilon,P_2(\boldsymbol{x},t-1))|)\\&=-\min(|-0.99P_i(\boldsymbol{x},t-1)|,|-\epsilon|)\\&=-\min(0.99P_i(\boldsymbol{x},t-1),-\epsilon)=-0.99P_i(\boldsymbol{x},t-1). \qquad (4)\end{aligned}$$

Using equation (2.72) and considering (4) we have that

$$P_i(\boldsymbol{x},t)=P_i(\boldsymbol{x},t-1)-0.99P_i(\boldsymbol{x},t-1)=0.01P_i(\boldsymbol{x},t-1). \quad (5)$$

Also using equation (2.73) and considering again (4) we have that

$$P_2(\boldsymbol{x},t)=P_2(\boldsymbol{x},t-1)+0.99P_i(\boldsymbol{x},t-1). \quad (6)$$

Therefore, using equations (5) and (6) we have

$$\begin{aligned}\sum_{l=1}^{4} P_l(\boldsymbol{x},t)&=\sum_{\substack{l=1\\ l\neq i,2}}^{4} P_l(\boldsymbol{x},t)+0.01P_i(\boldsymbol{x},t-1)\\ &\quad+(P_2(\boldsymbol{x},t-1)+0.99P_i(\boldsymbol{x},t-1))\\ &=\sum_{\substack{l=1\\ l\neq i,2}}^{4} P_l(\boldsymbol{x},t)+P_i(\boldsymbol{x},t-1)+P_2(\boldsymbol{x},t-1)\\ &=\sum_{\substack{l=1\\ l\neq i,2}}^{4} P_l(\boldsymbol{x},t-1)+P_i(\boldsymbol{x},t-1)+P_2(\boldsymbol{x},t-1)\\ &=\sum_{l=1}^{4} P_l(\boldsymbol{x},t-1)=1.\end{aligned}$$

Case c: Assume $0<P_i(\boldsymbol{x},t-1)+\epsilon<1$ and $P_2(\boldsymbol{x},t-1)-\epsilon\geq 1$. Since $1-P_2(\boldsymbol{x},t-1)\leq-\epsilon$ we have

$$\begin{aligned}&\operatorname{sgn}(\epsilon)\min(|\Xi(\epsilon,P_i(\boldsymbol{x},t-1))|,|\Xi(-\epsilon,P_2(\boldsymbol{x},t-1))|)\\ &=-\min(|\epsilon|,|0.99(1-P_2(\boldsymbol{x},t-1))|)\\ &=-\min(-\epsilon,0.99(1-P_2(\boldsymbol{x},t-1)))=0.99(P_2(\boldsymbol{x},t-1)-1). \quad (7)\end{aligned}$$

Using equation (2.72) and considering (7) we have that

$$\begin{aligned}P_i(\boldsymbol{x},t)&=P_i(\boldsymbol{x},t-1)+0.99(P_2(\boldsymbol{x},t-1)-1)\\ &=P_i(\boldsymbol{x},t-1)+0.99P_2(\boldsymbol{x},t-1)-0.99. \quad (8)\end{aligned}$$

Also using equation (2.73) and considering again (7) we have that

$$\begin{aligned}P_2(\boldsymbol{x},t)&=P_2(\boldsymbol{x},t-1)-0.99(P_2(\boldsymbol{x},t-1)-1)\\ &=0.01P_2(\boldsymbol{x},t-1)+0.99. \quad (9)\end{aligned}$$

Therefore, using equations (8) and (9) we have

$$\begin{aligned}\sum_{l=1}^{4} P_l(\boldsymbol{x},t) &= \sum_{\substack{l=1\\ l\neq i,2}}^{4} P_l(\boldsymbol{x},t)+(P_i(\boldsymbol{x},t-1)+0.99P_2(\boldsymbol{x},t-1)-0.99)\\ &\quad +(0.01P_2(\boldsymbol{x},t-1)+0.99)\\ &= \sum_{\substack{l=1\\ l\neq i,2}}^{4} P_l(\boldsymbol{x},t)+P_i(\boldsymbol{x},t-1)+P_2(\boldsymbol{x},t-1)\\ &= \sum_{\substack{l=1\\ l\neq i,2}}^{4} P_l(\boldsymbol{x},t-1)+P_i(\boldsymbol{x},t-1)+P_2(\boldsymbol{x},t-1)\\ &= \sum_{l=1}^{4} P_l(\boldsymbol{x},t-1)=1.\end{aligned}$$

Case d: Assume $P_i(\boldsymbol{x},t-1)+\epsilon\leq 0$ and $P_2(\boldsymbol{x},t-1)-\epsilon\geq 1$. We compute

$$\begin{aligned}&\operatorname{sgn}(\epsilon)\min(|\Xi(\epsilon,P_i(\boldsymbol{x},t-1))|,|\Xi(-\epsilon,P_2(\boldsymbol{x},t-1))|)\\ &=-\min(|\text{-}0.99P_i(\boldsymbol{x},t-1)|,|0.99(1-P_2(\boldsymbol{x},t-1))|)\\ &=-\min(0.99P_i(\boldsymbol{x},t-1),0.99(1-P_2(\boldsymbol{x},t-1))).\end{aligned}\tag{10}$$

There are two possible situations for the minimum in equation (10) but both are similar respectively to Case b and c and so in either situation $\sum_{l=0}^{4} P_l(\boldsymbol{x},t)=1$.

Case B: Assume $\epsilon>0$.

Case a: Assume $0<P_i(\boldsymbol{x},t-1)+\epsilon<1$ and $0<P_2(\boldsymbol{x},t-1)-\epsilon<1$. Same as Case (1, A, a) so $\sum_{l=1}^{4} P_l(\boldsymbol{x},t)=1$.

Case b: Assume $P_i(\boldsymbol{x},t-1)+\epsilon\geq 1$ and $0<P_2(\boldsymbol{x},t-1)-\epsilon<1$. Since $1-P_i(\boldsymbol{x},t-1)\leq\epsilon$ we have

$$\begin{aligned}&\operatorname{sgn}(\epsilon)\min(|\Xi(\epsilon,P_i(\boldsymbol{x},t-1))|,|\Xi(-\epsilon,P_2(\boldsymbol{x},t-1))|)\\ &=\min(|0.99(1-P_i(\boldsymbol{x},t-1))|,|\epsilon|)\\ &=\min(0.99(1-P_i(\boldsymbol{x},t-1)),\epsilon)=0.99(1-P_i(\boldsymbol{x},t-1)).\end{aligned}\tag{11}$$

Using equation (2.72) and considering (11) we have that

$$\begin{aligned}P_i(\boldsymbol{x},t)&=P_i(\boldsymbol{x},t-1)+0.99(1-P_i(\boldsymbol{x},t-1))\\ &=0.01P_i(\boldsymbol{x},t-1)+0.99.\end{aligned}\tag{12}$$

Also using equation (2.73) and considering again (11) we have that

$$\begin{aligned} P_2(\boldsymbol{x},t) &= P_2(\boldsymbol{x},t-1) - 0.99(1-P_i(\boldsymbol{x},t-1)) \\ &= P_2(\boldsymbol{x},t-1) + 0.99P_i(\boldsymbol{x}) - 0.99. \end{aligned} \tag{13}$$

Therefore, using equations (12) and (13) we have

$$\begin{aligned} \sum_{l=1}^{4} P_l(\boldsymbol{x},t) &= \sum_{\substack{l=1 \\ l\neq i,2}}^{4} P_l(\boldsymbol{x},t) + (0.01P_i(\boldsymbol{x},t-1)+0.99) \\ &\quad + (P_2(\boldsymbol{x},t-1)+0.99P_i(\boldsymbol{x},t-1)-0.99) \\ &= \sum_{\substack{l=1 \\ l\neq i,2}}^{4} P_l(\boldsymbol{x},t) + P_i(\boldsymbol{x},t-1) + P_2(\boldsymbol{x},t-1) \\ &= \sum_{\substack{l=1 \\ l\neq i,2}}^{4} P_l(\boldsymbol{x},t-1) + P_i(\boldsymbol{x},t-1) + P_2(\boldsymbol{x},t-1) \\ &= \sum_{l=1}^{4} P_l(\boldsymbol{x},t-1) = 1. \end{aligned}$$

Case c: Assume $0<P_i(\boldsymbol{x},t-1)+\epsilon<1$ and $P_2(\boldsymbol{x},t-1)-\epsilon\leq 0$. Since $P_2(\boldsymbol{x},t-1)\leq\epsilon$ we have

$$\begin{aligned} &\operatorname{sgn}(\epsilon)\min(|\Xi(\epsilon,P_i(\boldsymbol{x},t-1))|,|\Xi(-\epsilon,P_2(\boldsymbol{x},t-1))|) \\ &= \min(|\epsilon|,|-0.99P_2(\boldsymbol{x},t-1)|) \\ &= \min(\epsilon,0.99(1-P_2(\boldsymbol{x},t-1))) = 0.99P_2(\boldsymbol{x},t-1). \end{aligned} \tag{14}$$

Using equation (2.72) and considering (14) we have that

$$P_i(\boldsymbol{x},t) = P_i(\boldsymbol{x},t-1) + 0.99P_2(\boldsymbol{x},t-1) \tag{15}$$

Also using equation (2.73) and considering again (14) we have that

$$\begin{aligned} P_2(\boldsymbol{x},t) &= P_2(\boldsymbol{x},t-1) - 0.99P_2(\boldsymbol{x},t-1) \\ &= 0.01P_2(\boldsymbol{x},t-1). \end{aligned} \tag{16}$$

Therefore, using equations (15) and (16) we have

$$\begin{aligned}\sum_{l=1}^{4} P_l(\boldsymbol{x},t) &= \sum_{\substack{l=1\\ l\neq i,2}}^{4} P_l(\boldsymbol{x},t) + (P_i(\boldsymbol{x},t-1)+0.99P_2(\boldsymbol{x},t-1)) \\ &\quad +0.01P_2(\boldsymbol{x},t-1) \\ &= \sum_{\substack{l=1\\ l\neq i,2}}^{4} P_l(\boldsymbol{x},t) + P_i(\boldsymbol{x},t-1)+P_2(\boldsymbol{x},t-1) \\ &= \sum_{\substack{l=1\\ l\neq i,2}}^{4} P_l(\boldsymbol{x},t-1) + P_i(\boldsymbol{x},t-1)+P_2(\boldsymbol{x},t-1) \\ &= \sum_{l=1}^{4} P_l(\boldsymbol{x},t-1) = 1.\end{aligned}$$

Case d: Assume $P_i(\boldsymbol{x},t-1)+\epsilon \geq 1$ and $P_2(\boldsymbol{x},t-1)-\epsilon \leq 0$. We compute

$$\begin{aligned}&\operatorname{sgn}(\epsilon)\min(|\Xi(\epsilon,P_i(\boldsymbol{x},t-1))|,|\Xi(-\epsilon,P_2(\boldsymbol{x},t-1))|) \\ &= \min(|0.99(1-P_i(\boldsymbol{x},t-1))|,|-0.99P_2(\boldsymbol{x},t-1)|) \\ &= \min(0.99(1-P_i(\boldsymbol{x},t-1)),0.99P_2(\boldsymbol{x},t-1)). \qquad (17)\end{aligned}$$

There are two possible situations for the minimum in equation (17) but both are similar respectively to Case b and c and so in either situation $\sum_{l=0}^{4} P_l(\boldsymbol{x},t)=1$.

Case 2: Now assume that $i=2$ and $k=1,3,4$ as we are balancing modifications in the probability of quiescence equally against the probabilities of all the other phenotypic actions.

Case A: Assume $\epsilon<0$.

Case a: Assume $0<P_2(\boldsymbol{x},t-1)+\epsilon<1$. We have that

$$\Xi(\epsilon,P_2(\boldsymbol{x},t-1))=\epsilon \qquad (18)$$

Using equation (2.72) and considering (18) we have that

$$P_2(\boldsymbol{x},t)=P_2(\boldsymbol{x},t-1)+\epsilon. \qquad (19)$$

Also using equation (2.73) and considering again (18) we have that

$$P_k(\boldsymbol{x},t)=P_k(\boldsymbol{x},t-1)-\frac{\epsilon P_k(\boldsymbol{x},t-1)}{\sum_{\substack{m=1\\m\neq 2}}^{4}P_m(\boldsymbol{x},t-1)}, \qquad (20)$$

$$k=1,3,4.$$

Therefore, using equations (19) and (20) we have

$$\begin{aligned}\sum_{l=1}^{4}P_l(\boldsymbol{x},t)&=\sum_{l=1}^{4}P_l(\boldsymbol{x},t-1)-\epsilon\sum_{\substack{l=1\\l\neq 2}}^{4}\frac{P_l(\boldsymbol{x},t-1)}{\sum_{\substack{m=1\\m\neq 2}}^{4}P_m(\boldsymbol{x},t-1)}+\epsilon\\&=\sum_{l=1}^{4}P_l(\boldsymbol{x},t-1)+\epsilon\left(1-\frac{\sum_{\substack{l=1\\l\neq 2}}^{4}P_l(\boldsymbol{x},t-1)}{1-P_2(\boldsymbol{x},t-1)}\right)\\&=\sum_{l=1}^{4}P_l(\boldsymbol{x},t-1)+\epsilon\left(1-\frac{1-P_2(\boldsymbol{x},t-1)}{1-P_2(\boldsymbol{x},t-1)}\right)\\&=\sum_{l=1}^{4}P_l(\boldsymbol{x},t-1)+\epsilon(1-1)\\&=\sum_{l=1}^{4}P_l(\boldsymbol{x},t-1)=1.\end{aligned}$$

Case b: Assume $P_2(\boldsymbol{x},t-1)+\epsilon\leq 0$. We have that

$$\Xi(\epsilon,P_2(\boldsymbol{x},t-1))=\text{-}0.99P_2(\boldsymbol{x},t-1) \qquad (21)$$

Using equation (2.72) and considering (21) we have that

$$\begin{aligned}P_2(\boldsymbol{x},t)&=P_2(\boldsymbol{x},t-1)-0.99P_2(\boldsymbol{x},t-1)\\&=0.01P_2(\boldsymbol{x},t-1). \qquad (22)\end{aligned}$$

Also using equation (2.73) and considering again (21) we have that

$$P_k(\boldsymbol{x},t)=P_k(\boldsymbol{x},t-1)+\frac{0.99P_2(\boldsymbol{x},t-1)P_k(\boldsymbol{x},t-1)}{\sum_{\substack{m=1\\m\neq 2}}^{4}P_m(\boldsymbol{x},t-1)}, \qquad (23)$$

$$k=1,3,4.$$

Therefore, using equations (22) and (23) we have

$$\begin{aligned}
\sum_{l=1}^{4} P_l(\boldsymbol{x},t) &= \sum_{\substack{l=1\\l\neq 2}}^{4} P_l(\boldsymbol{x},t-1) \\
&\quad +0.99P_2(\boldsymbol{x},t-1)\sum_{\substack{l=1\\l\neq 2}}^{4}\frac{P_l(\boldsymbol{x},t-1)}{\sum_{\substack{m=1\\m\neq 2}}^{4} P_m(\boldsymbol{x},t-1)} \\
&\quad +0.01P_2(\boldsymbol{x},t-1) \\
&= \sum_{\substack{l=1\\l\neq 2}}^{4} P_l(\boldsymbol{x},t-1) \\
&\quad +0.99P_2(\boldsymbol{x},t-1)\frac{\sum_{\substack{l=1\\l\neq 2}}^{4} P_l(\boldsymbol{x},t-1)}{1-P_2(\boldsymbol{x},t-1)}+0.01P_2(\boldsymbol{x},t-1) \\
&= \sum_{\substack{l=1\\l\neq 2}}^{4} P_l(\boldsymbol{x},t-1) \\
&\quad +0.99P_2(\boldsymbol{x},t-1)\frac{1-P_2(\boldsymbol{x},t-1)}{1-P_2(\boldsymbol{x},t-1)}+0.01P_2(\boldsymbol{x},t-1) \\
&= \sum_{\substack{l=1\\l\neq 2}}^{4} P_l(\boldsymbol{x},t-1)+0.99P_2(\boldsymbol{x},t-1)+0.01P_2(\boldsymbol{x},t-1) \\
&= \sum_{\substack{l=1\\l\neq 2}}^{4} P_l(\boldsymbol{x},t-1)+P_2(\boldsymbol{x},t-1)=\sum_{l=1}^{4} P_l(\boldsymbol{x},t-1)=1.
\end{aligned}$$

Case B: Assume $\epsilon>0$.

Case a: Assume $0<P_2(\boldsymbol{x},t-1)+\epsilon<1$. Same as Case (2, A, a) so

$$\sum_{l=1}^{4} P_l(\boldsymbol{x},t)=1.$$

Case b: Assume $P_2(\boldsymbol{x},t-1)+\epsilon\geq 1$. We have that

$$\Xi(\epsilon, P_2(\boldsymbol{x},t-1))=0.99(1-P_2(\boldsymbol{x},t-1)) \tag{24}$$

Using equation (2.72) and considering (24) we have that

$$\begin{aligned}
P_2(\boldsymbol{x},t) &= P_2(\boldsymbol{x},t-1)+0.99(1-P_2(\boldsymbol{x},t-1) \\
&= 0.01P_2(\boldsymbol{x},t-1)+0.99.
\end{aligned} \tag{25}$$

Also using equation (2.73) and considering again (24) we have that

$$\begin{aligned} P_k(\boldsymbol{x},t) &= P_k(\boldsymbol{x},t-1) - \frac{0.99(1-P_2(\boldsymbol{x},t-1))P_k(\boldsymbol{x},t-1)}{\sum_{\substack{m=1\\m\neq 2}}^{4} P_m(\boldsymbol{x},t-1)} \\ &= P_k(\boldsymbol{x},t-1) \\ &\quad + \frac{(0.99P_2(\boldsymbol{x},t-1)-0.99)P_k(\boldsymbol{x},t-1)}{\sum_{\substack{m=1\\m\neq 2}}^{4} P_m(\boldsymbol{x},t-1)}, \end{aligned} \tag{26}$$

$k=1,3,4$.

Therefore, using equations (25) and (26) we have

$$\begin{aligned} \sum_{l=1}^{4} P_l(\boldsymbol{x},t) &= \sum_{\substack{l=1\\l\neq 2}}^{4} P_l(\boldsymbol{x},t-1) \\ &\quad + (0.99P_2(\boldsymbol{x},t-1)-0.99)\sum_{\substack{l=1\\l\neq 2}}^{4} \frac{P_l(\boldsymbol{x},t-1)}{\sum_{\substack{m=1\\m\neq 2}}^{4} P_m(\boldsymbol{x},t-1)} \\ &\quad + 0.01P_2(\boldsymbol{x},t-1)+0.99 \\ &= \sum_{\substack{l=1\\l\neq 2}}^{4} P_l(\boldsymbol{x},t-1) \\ &\quad + (0.99P_2(\boldsymbol{x},t-1)-0.99)\frac{\sum_{\substack{l=1\\l\neq 2}}^{4} P_l(\boldsymbol{x},t-1)}{1-P_2(\boldsymbol{x},t-1)} \\ &\quad + 0.01P_2(\boldsymbol{x},t-1)+0.99 \\ &= \sum_{\substack{l=1\\l\neq 2}}^{4} P_l(\boldsymbol{x},t-1) \\ &\quad + (0.99P_2(\boldsymbol{x},t-1)-0.99)\frac{1-P_2(\boldsymbol{x},t-1)}{1-P_2(\boldsymbol{x},t-1)} \\ &\quad + 0.01P_2(\boldsymbol{x},t-1)+0.99 \\ &= \sum_{\substack{l=1\\l\neq 2}}^{4} P_l(\boldsymbol{x},t-1) + 0.99P_2(\boldsymbol{x},t-1) - 0.99 \\ &\quad + 0.01P_2(\boldsymbol{x},t-1)+0.99 \\ &= \sum_{\substack{l=1\\l\neq 2}}^{4} P_l(\boldsymbol{x},t-1) + P_2(\boldsymbol{x},t-1) = \sum_{l=1}^{4} P_l(\boldsymbol{x},t-1) = 1. \end{aligned}$$

Since the two cases above are exhaustive and all cases resulted in $\sum_{l=1}^{4} P_l(\boldsymbol{x},t)=1$ then it must be that modifying phenotypic action i with respect to some gene j will maintain the summation given by $\sum_{l=1}^{4} P_l(\boldsymbol{x},t)=1$. □

2.5.2 Update Rules for Cell Class

The difference between the mutated class of a cell and the non-mutated class is that the mutated class has $\Upsilon \in \mathbb{N}$ positively mutated genes. Note that to delay the arrival of the first CSC not only does there have to be Υ positively mutated genes but for a SC or MNSC to transition into a CSC a random variable has to be less than the threshold $\iota \in \mathbb{R}_+(0,1)$. This can be described by the following

$$\overline{\Psi}(\boldsymbol{x},t,j) = \begin{cases} 1, & \Psi(\boldsymbol{x},t,j)=1 \text{ or } \Psi(\boldsymbol{x},t,j)=2 \\ 0, & \text{otherwise} \end{cases}, \tag{2.74}$$

$$s(\boldsymbol{x},t)= \begin{cases} 0(\text{NTC}), & s(\boldsymbol{x},t-1)=1(\text{MNTC}), \\ & \sum_{j=1}^{G} \overline{\Psi}(\boldsymbol{x},t,j)<\Upsilon \\ 1(\text{MNTC}), & s(\boldsymbol{x},t-1)=0(\text{NTC}), \\ & \sum_{j=1}^{G} \overline{\Psi}(\boldsymbol{x},t,j)\geq\Upsilon \\ 2(\text{NSC}), & s(\boldsymbol{x},t-1)=3(\text{MNSC}), \\ & \sum_{j=1}^{G} \overline{\Psi}(\boldsymbol{x},t,j)<\Upsilon \\ 3(\text{MNSC}), & s(\boldsymbol{x},t-1)=2(\text{NSC}), \\ & \sum_{j=1}^{G} \overline{\Psi}(\boldsymbol{x},t,j)\geq\Upsilon \\ 4(\text{CSC}), & s(\boldsymbol{x},t-1)=2(\text{NSC}), \\ & \sum_{j=1}^{G} \overline{\Psi}(\boldsymbol{x},t,j)\geq\Upsilon, \\ & z\leq\frac{\iota}{2} \\ 4(\text{CSC}), & s(\boldsymbol{x},t-1)=3(\text{MNSC}), \\ & \sum_{j=1}^{G} \overline{\Psi}(\boldsymbol{x},t,j)\geq\Upsilon, \\ & z\leq\iota \end{cases}, \tag{2.75}$$

where $z \sim U(0,1)$, $\Psi(\boldsymbol{x}, t, j)$ is the mutation indicator function given by (2.66), and G is the number of genes.

Since each cell class has different probabilities for each phenotypic action, then there exists an initial phenotype matrix, $\tilde{P} \in \mathbb{R}^{6\times 4}(0,1)$ with $\sum_{i=1}^{4} \tilde{P}_{ki}=1$, $k=0,1,2,3,4,5$. Therefore, whenever a non-empty cell changes class, the phenotype vector must be set to its' new cell class's initial values, while also keeping all the changes that have happened to the phenotype vector. This is achieved through the following process

$$P_i(\boldsymbol{x}, t)=\tilde{P}_{ki}+\Xi(P_i(\boldsymbol{x}, t-1)-\tilde{P}_{s(\boldsymbol{x},t-1)i}, \tilde{P}_{ki}), \quad \text{where } k \text{ is the new cell state}, i=1,2,3,4. \tag{2.76}$$

If the probabilities of the phenotypic actions no longer sum to one after the changes, then standard normalization is used to fix the sum.

2.5.3 Dedifferentiation

Dedifferentiation is the process of a specialized cell reverting back to a non-specialized cell. In our model this is accomplished by a non stem cell becoming a stem cell. Dedifferentiation is used to help maintain the proper ratio of stem cells to non stem cells in the grid by dedifferentiating whenever the number of stem cells in the neighbourhood of a non stem cell is less than or equal to some chosen value, $\hat{S}$, or if the number of empty cells in the neighbourhood of a non stem cell is less than or equal to some chosen value, $\hat{E}$, or by random chance based on a sample from the uniform distribution being less than a threshold. To help reduce the number of cells dedifferentiating, the process is completed only when a random sample from the uniform distribution is less than or equal to some threshold, $\hat{D} \in \mathbb{R}_{+}(0,1)$. If the non stem cell is not a TAC then the threshold is set at $\frac{\hat{D}}{2}$. In the case of random chance the threshold is set at $\frac{\hat{D}}{4}$. This process is represented by the function:

$$s(\boldsymbol{x}, t)=\begin{cases} 2(\text{NSC}), & s(\boldsymbol{x}, t-1)=0(\text{NTC}) \\ 3(\text{MNSC}), & s(\boldsymbol{x}, t-1)=1(\text{MNTC}) \\ 4(\text{CSC}), & s(\boldsymbol{x}, t-1)=5(\text{TC}) \end{cases}. \tag{2.77}$$

2.5.4 Cell Fitness

Each cell has a fitness value associated to it, so that the cells can compete and the population contains only the healthiest, or in the case of mutated cells, the most positively mutated cells. The characteristics that affect the fitness are based upon work by Bowling *et al.* [17], in which they point to the following important characteristics:

1. if a cell has a high apoptotic rate, it is less fit;
2. if a cell has a high proliferation rate, it is more fit;
3. if a cell is older, then it is less fit;
4. if a gene in a cell is mutated towards cancer, than it is less fit unless the cell is cancerous, in which case it is more fit;
5. if a cell is what they call super-competitive, then it is more fit than any type of cell.

Based upon this the fitness of a cell in the CA is computed using the functions $\hat{M}$ which is the gene expression ratio and the fitness function $\hat{F}$:

$$\hat{M}(\boldsymbol{x},t,j)=\begin{cases} -\dfrac{E_j(\boldsymbol{x},t)}{\overline{M}} & ,(T_j=0 \text{ and } E_j(\boldsymbol{x},t)<0) \\ & \text{ or } (T_j=1 \text{ and } E_j(\boldsymbol{x},t)>0) \\ \dfrac{E_j(\boldsymbol{x},t)}{\overline{M}} & ,(T_j=0 \text{ and } E_j(\boldsymbol{x},t)>0) \\ & \text{ or } (T_j=1 \text{ and } E_j(\boldsymbol{x},t)<0) \\ 0 & ,\text{otherwise} \end{cases}, \tag{2.78}$$

$$\begin{aligned} \hat{F}(\boldsymbol{x},t)=&\begin{cases} \dfrac{P_1(\boldsymbol{x},t)}{\tilde{P}_{s(\boldsymbol{x},t)1}} & ,\overline{\tau}(\boldsymbol{x},t)=0 \\ \dfrac{P_1(\boldsymbol{x},t)}{\tilde{P}_{s(\boldsymbol{x},t)1}+\omega} & ,\overline{\tau}(\boldsymbol{x},t)=1 \end{cases} \\ &-\frac{P_3(\boldsymbol{x},t)}{\tilde{P}_{s(\boldsymbol{x},t)3}}-\alpha(\boldsymbol{x},t)P_3(\boldsymbol{x},t) \\ &+\begin{cases} \sum_{j=1}^{G}\hat{M}(\boldsymbol{x},t,j) & ,s(\boldsymbol{x},t)=0 \text{ (NTC) or } s(\boldsymbol{x},t)=2\text{(NSC)} \\ -\sum_{j=1}^{G}\hat{M}(\boldsymbol{x},t,j) & ,\text{otherwise} \end{cases}, \end{aligned} \tag{2.79}$$

where $E_j(\boldsymbol{x},t)$ is the gene-expression of gene j, $\overline{M}$ is the threshold that determines where gene j is mutated, T_j indicates if the gene j is either a tumour suppressor gene or oncogene, $P_1(\boldsymbol{x},t)$ is the probability that proliferation will occur in a time-step, $P_3(\boldsymbol{x},t)$ is the probability that apoptosis will occur in a time-step, $\alpha(\boldsymbol{x},t)$ is the age of the cell, $\overline{\tau}(\boldsymbol{x},t)$ indicates if the cell is a TAC or not, ω is the amount the probability of proliferation is increased when a cell is a TAC, and $\tilde{P}$ is the initial phenotype matrix that provides the initial values for each of the phenotypic actions for each cell class. If the cell is from a SC class (SC, MNSC, CSC) then its' fitness is multiplied by a factor $\Pi\in\mathbb{R}_+$ if $\hat{F}(\boldsymbol{x},t)>0$ or Π^{-1} if $\hat{F}(\boldsymbol{x},t)<0$, so that a SC type cell has a higher fitness than a non-stem cell.

2.5.5 CA Rule

Each non-empty cell in the CA grid chooses a phenotypic action to execute for that time-step and attempts to complete such action. The phenotypic action is chosen by taking a random sample from the probability distribution that is generated by the phenotype vector. Consider that the cell that is performing the phenotypic action is located at $\boldsymbol{x}^{(p)} \in \Omega$. The cell the action is being applied to will be located at $\boldsymbol{x}^{(c)} \in \Omega$, which is a randomly chosen location in the cell's neighbourhood. The randomly chosen cell is an empty cell or a cell with a lower fitness in the case of proliferation, and differentiation. It is important to note that only NSC, MNSC, and CSC can differentiate.

CSCs and TCs are the only class of cells that can kill other cells when moving during quiescence. If the parent cell is a CSC or TC and the chosen cell has a higher fitness then the phenotypic action is accomplished only if a sample from some random variable is less than a threshold to kill, $\kappa \in \mathbb{R}_+(0,1)$. A CSC can kill a TC and TC a CSC only if the fitness is lower, as otherwise the tumour cell population and/or CSC population would die out before they have had the chance to thrive.

2.5.5.1 Proliferation

Assuming all the conditions for the phenotypic action to occur are met then the following changes occur to the states of the parent and child cell upon proliferation:

$$\boldsymbol{S}(\boldsymbol{x}^{(k)},t) = \begin{cases} [s(\boldsymbol{x}^{(p)},t-1), 0, \boldsymbol{E}(\boldsymbol{x}^{(p)},t-1), \boldsymbol{P}(\boldsymbol{x}^{(p)},t-1), & ,\overline{\tau}(\boldsymbol{x}^{(p)},t-1)=0 \\ \quad \overline{\tau}(\boldsymbol{x}^{(p)},t-1), \overline{n}(\boldsymbol{x}^{(p)},t-1)] & \\ & \\ [s(\boldsymbol{x}^{(p)},t-1), 0, \boldsymbol{E}(\boldsymbol{x}^{(p)},t-1), \boldsymbol{P}(\boldsymbol{x}^{(p)},t-1), & ,\overline{\tau}(\boldsymbol{x}^{(p)},t-1)=1, \\ \quad \overline{\tau}(\boldsymbol{x}^{(p)},t-1), \overline{n}(\boldsymbol{x}^{(p)},t-1)+1] & \overline{n}(\boldsymbol{x}^{(p)},t-1) \neq \Theta \\ & \\ [s(\boldsymbol{x}^{(p)},t-1), 0, \boldsymbol{E}(\boldsymbol{x}^{(p)},t-1), & ,\overline{\tau}(\boldsymbol{x}^{(p)},t-1)=1, \\ \quad [P_1(\boldsymbol{x}^{(p)},t-1) - \Xi(\omega, P_1(\boldsymbol{x}^{(p)},t-1)), & \overline{n}(\boldsymbol{x}^{(p)},t-1)=\Theta \\ \quad P_2(\boldsymbol{x}^{(p)},t-1) + \Xi(\omega, P_2(\boldsymbol{x}^{(p)},t-1)), & \\ \quad P_3(\boldsymbol{x}^{(p)},t-1), P_4(\boldsymbol{x}^{(p)},t-1)], 0, 0] & \end{cases}, \tag{2.80}$$

where $k \in \{p, c\}$; $\omega \in \mathbb{R}_+(0,1)$; $\boldsymbol{S}(\boldsymbol{x}, t)$ is defined by equation (2.64); $\Xi(\epsilon, \xi)$, defined by equation (2.70), ensures that the increment ω does not cause the phenotypic action being modified to leave the interval $[0, 1]$.

2.5.5.2 Quiescence

Assuming all the conditions for the phenotypic action to occur are met then the following changes occur to the states of the parent and child cell upon cell movement:

$$\boldsymbol{S}(\boldsymbol{x}^{(p)}, t) = \begin{cases} [6, 0, \boldsymbol{0}, \boldsymbol{0}, 0, 0] & , z < \zeta \\ [s(\boldsymbol{x}^{(p)}, t-1), \alpha(\boldsymbol{x}^{(p)}, t-1), \boldsymbol{E}(\boldsymbol{x}^{(p)}, t-1), & \\ \quad \boldsymbol{P}(\boldsymbol{x}^{(p)}, t-1), \overline{\tau}(\boldsymbol{x}^{(p)}, t-1), \overline{n}(\boldsymbol{x}^{(p)}, t-1)] & , \text{otherwise} \end{cases}, \tag{2.81}$$

$$\boldsymbol{S}(\boldsymbol{x}^{(c)}, t) = \begin{cases} [s(\boldsymbol{x}^{(p)}, t-1), \alpha(\boldsymbol{x}^{(p)}, t-1), \boldsymbol{E}(\boldsymbol{x}^{(p)}, t-1), & \\ \quad \boldsymbol{P}(\boldsymbol{x}^{(p)}, t-1), \overline{\tau}(x_1, y_1, t-1, \overline{n}(\boldsymbol{x}^{(p)}, t-1)] & , z < \zeta \\ [s(\boldsymbol{x}^{(c)}, t-1), \alpha(\boldsymbol{x}^{(c)}, t-1), \boldsymbol{E}(\boldsymbol{x}^{(c)}, t-1), & \\ \quad \boldsymbol{P}(\boldsymbol{x}^{(c)}, t-1), \overline{\tau}(\boldsymbol{x}^{(c)}, t-1), \overline{n}(\boldsymbol{x}^{(c)}, t-1)] & , \text{otherwise} \end{cases}, \tag{2.82}$$

where $z \sim U(0,1)$ and $\zeta \in \mathbb{R}_+(0,1)$ is the threshold that must be met for the cell to move.

2.5.5.3 Apoptosis

The following changes occur to the state of the cell at $\boldsymbol{x}^{(p)}$ upon apoptosis:

$$\boldsymbol{S}(\boldsymbol{x}^{(p)}, t) = [6, 0, \boldsymbol{0}, \boldsymbol{0}, 0, 0]. \tag{2.83}$$

2.5.5.4 Differentiation

Assuming all the conditions for the phenotypic action to occur are met then the following changes occur to the states of the parent and child cell upon

differentiation:

$$\boldsymbol{S}(\boldsymbol{x}^{(p)},t)=[s(\boldsymbol{x}^{(p)},t-1),0,\boldsymbol{E}(\boldsymbol{x}^{(p)},t-1), \tag{2.84}$$
$$\boldsymbol{P}(\boldsymbol{x}^{(p)},t-1),\overline{\tau}(\boldsymbol{x}^{(p)},t-1),\overline{n}(\boldsymbol{x}^{(p)},t-1)]$$

$$\overline{\Sigma}(\boldsymbol{x}^{(p)},t)=\begin{cases}0(\text{NTC}) & ,s(\boldsymbol{x}^{(p)},t-1)=2(\text{NSC})\\ 1(\text{MNTC}) & ,s(\boldsymbol{x}^{(p)},t-1)=3(\text{MNSC})\\ 5(\text{TC}) & ,s(\boldsymbol{x}^{(p)},t-1)=4(\text{CSC})\end{cases} \tag{2.85}$$

$$\boldsymbol{S}(\boldsymbol{x}^{(c)},t)=[\overline{\Sigma}(\boldsymbol{x}^{p},t),0,\boldsymbol{E}(\boldsymbol{x}^{(p)},t-1), \tag{2.86}$$
$$[P_1(\boldsymbol{x}^{(p)},t-1)+\Xi(\omega,P_1(\boldsymbol{x}^{(p)},t-1)),$$
$$P_2(\boldsymbol{x}^{(p)},t-1)-\Xi(\omega,P_2(\boldsymbol{x}^{(p)},t-1)),$$
$$P_3(\boldsymbol{x}^{(p)},t-1),P_4(\boldsymbol{x}^{(p)},t-1)],1,0],$$

where $\omega\in\mathbb{R}_+(0,1)$ is the amount the probability of proliferation changes when the cell is a TAC.

2.5.6 Tumour Excision

For the purpose of exploring the length of time it takes for a tumour to recur, we include three simple types of excisions in the model. One whereby, all the tumour cells in the grid and their neighbours up-to a certain depth are removed. Another, where all the mutated cell class cells and their neighbours up-to a certain depth are removed. Lastly, where a circular sub-region of the grid is removed.

2.5.7 Lineage Tracking

The lineage of each cell is tracked for the purpose of following tumour cell lineages from their origin, checking how many independent tumour masses form throughout the simulation, and whether the origin is monoclonal or polyclonal. The lineage is tracked using the following methods. Each cell has a parameter $\Lambda(\boldsymbol{x},t)\in\{-1,0,1,...,L\times M-1\}$, where L is the length in the x-direction and M is the length in the y-direction, that when set to -1 implies no information is known or the cell is empty, and if set to any other value represents the spatial index of the parent at the time it created the first cell in the lineage.

During cell movement the parameter is passed down from parent ($\boldsymbol{x}^{(p)}$) to child cell ($\boldsymbol{x}^{(c)}$). In the case of apoptosis the parameter is set to -1. Finally, during proliferation and differentiation the parameter is set as follows for the cells $\boldsymbol{x}^{(p)}$ and $\boldsymbol{x}^{(c)}$:

$$\Lambda(\boldsymbol{x},t)=\begin{cases} x_1^{(p)}M+x_2^{(p)}, & \Lambda(\boldsymbol{x}^{(p)},t-1)=-1 \\ \Lambda(\boldsymbol{x}^{(p)},t-1), & \text{otherwise} \end{cases}. \tag{2.87}$$

$$\tag{2.88}$$

2.5.8 Field Definition

A field in this model is defined to be any groupings of mutated cells, thus a field is a section of the grid that contains few or no NTC or NSC. We can also further define a field based around the lineages, so that a grouping of mutated cells that share the same lineage would be considered a field.

2.5.9 CA Model Timeline

Each time-step has the following order of actions:

1. update the carcinogen sensitivity function,
2. run the gene expression neural network,
3. update the gene expressions based upon the output of the neural network in step 2,
4. update the gene expressions via the gene instability process,
5. update the phenotype vector based upon the gene expressions of each gene,
6. update the states of each cell using the state transition process,
7. apply the dedifferentiation process,
8. apply the phenotypic action chosen by the cell for that time-step,
9. possibly perform tumour excision.

Chapter 3

Application and Implementation

Though field cancerization is found in many types of tissue throughout the body, the most commonly studied case is head and neck squamous cell carcinoma (HNSCC). HNSCC is the case study we will use to test the model proposed in the previous chapter. Ethanol and nicotine are the most commonly associated carcinogens to HNSCC, thus we consider them in this chapter. These two carcinogens typically enter the body through smoking and/or chewing tobacco and drinking alcohol, respectively.

In this chapter we will discuss how the parameters for the model were chosen for this case study. As well as how the model was implemented in parallel and the difficulties that arose in that process.

3.1 Application

3.1.1 Introduction to Head and Neck Squamous Cell Carcinoma

We review head and neck squamous cell carcinoma (HNSCC) as presented in [68]. Squamous cell carcinoma (SCC) is cancer that originates from squamous cells. These cells are found in the outer layer of skin and in the mucous membranes, which are the moist tissues that line body cavities such as the airways and intestines. There is a particular type of SCC that develops in the mucous membranes of the mouth, nose, and throat, it is called head and neck SCC. HNSCC is classified relative to its location, namely, there are the following main types: oral cavity (occurs in the mouth), oropharynx (middle part of the

throat near the mouth), nasal cavity and para-nasal sinuses (space behind the nose), nasopharynx (upper part of the throat near the nasal cavity), larynx (voice-box), and the hypopharynx (lower part of the throat near the larynx). Symptoms of HNSCC include abnormal patches or open sores (ulcers) in the mouth and throat, unusual bleeding or pain in the mouth, sinus congestion that does not clear, sore throat, earache, pain when swallowing or difficulty swallowing, a hoarse voice, difficulty breathing, or enlarged lymph nodes. It can metastasize to other parts of the body such as the lymph nodes or lungs. There is about a 50% chance of surviving another 5 years after initial diagnosis. HNSCC is the seventh most common type of cancer worldwide with approximately 600,000 new diagnoses each year, including about 50,000 in the US alone. It most often occurs in men in their 50s or 60s.

Tobacco use, including cigarettes, cigars, pipes, chewing tobacco, and snuff, is the largest risk factor for HNSCC, since it is linked to 85% of cases [67]. In the US smoking more than 2 packs of cigarettes per day is the main tobacco-related risk factor for mouth and throat cancer [116]. Pipe smoking in particular has been linked to cancer in the part of the lips that touch the pipe stem [116]. Chewing tobacco or snuff is associated with a 50% increase in the risk of developing cancer in areas of the mouth that comes most in-contact with the tobacco. This includes the cheeks, gums, and inner surface of the lips [116]. Secondhand smoke may also increase a person's risk of head and neck cancer [116]. Frequent and heavy alcohol consumption increases the risk of head and neck cancer with the risk increasing proportional to the amount of alcohol a person consumes [116]. Using alcohol and tobacco together is known to increase the risk of developing head and neck cancer by two to three times more than just one of them alone [116]. Human papillomavirus (HPV) infection, which induces cancer to develop in the tonsils and base of the tongue, increases the risk of developing throat cancer 16-fold and causes 60% of throat cancers [116]. HPV is causing an increase in the number of incidences of HNSCC among younger individuals [116]. Prolonged sun exposure can increase the risk of cancer in the lip area due to UV radiation [116]. Gender seems to be a factor in the risk as well since men are more likely to develop HNSCC

then women [116]. People with fairer skin also seem to have an increased risk to develop HNSCC [67]. Generally people older than 45 have an increased risk for oral cancer, though it can develop in people of any age [67]. Poor dental care and not following regular oral hygiene practises may cause an increased risk of oral cavity cancer, this risk is even further increased for people that use alcohol and tobacco products [116]. A diet low in fruits and vegetables and a vitamin A deficiency may increase the risk of oral and oropharyngeal cancer [67]. Chewing betel nuts, a nut containing a mild stimulant that is popular in Asia, also raises a person's risk of developing oral and oropharyngeal cancer [67]. It has also been seen that people that have a weakened immune system may have a higher risk of developing HNSCC [67]. Finally, it has been shown that people that use marijuana may be at a higher-than-average risk for HNSCC [67]. HNSCC is generally not inherited so as result it arises from mutations in the body's cells that occur during an individuals lifetime [68]. The best ways to prevent the risk of developing HNSCC is to not use any tobacco products and to try to prevent acquiring HPV infection.

There are many genes related to HNSCC but the top nine genes are: ING1, PTEN, TNFRSF10B, TP53, MIR21, MIR210, MIR205, MIR98, and ING3 [52]. Looking specifically at HNSCC in the tongue the top twenty genes that are related are: TP53, FAT1, CDKN2A, NOTCH1, PIK3CA, KMT2D, FAT4, CASP8, MYH9, EP300, NSD1, HRAS, NOTCH2, MLLT4, FBXW7, NFE2L2, AKAP9, GRIN2A, RB1, and CDH11 [52].

The reason we use HNSCC for a case study of the model proposed in this thesis is twofold, first, it is the first type of cancer that field cancerization was discovered, and secondly, it is a widely studied case. This thesis will consider the case of tobacco and alcohol as carcinogens as related to HNSCC of the tongue.

3.2 Application

As mentioned we will be using HNSCC of the tongue for our particular case study. Ethanol (indexed as 1) and tobacco are the most commonly associated

carcinogens to HNSCC, thus we consider these in this chapter. Recall that we are substituting nicotine (indexed as 2) to represent the carcinogens of tobacco. We will be considering ten genes in our model that are correlated with HNSCC of the tongue.

3.2.1 Carcinogen Parameters

For each carcinogen there are two required parameters, the diffusion coefficient and the influx of the carcinogen. Since the length of each time-step in the CA is based upon the length of the cell cycle, which is typically in the magnitude of hours, then our time unit will be in hours.

First let us look at the diffusion coefficients which are computed using the Reddy-Doraiswamy equation [115] which is given by

$$D=\begin{cases} 10^{-7}\frac{M^{\frac{1}{2}}T}{\mu(V_1V_2)^{\frac{1}{3}}}, & \frac{V_2}{V_1}\leq 1.5 \\ \\ 8.5\times 10^{-8}\frac{M^{\frac{1}{2}}T}{\mu(V_1V_2)^{\frac{1}{3}}}, & \frac{V_2}{V_1}>1.5 \end{cases}, \tag{3.1}$$

where M is the molar mass of the solvent, T is the absolute temperature, μ is the solvent viscosity, V_1 is the molar volume of the solute, and V_2 is the molar volume of the solvent. We assume the solvent is water for which the molecular weight is M=18.01528g/mol. We will use normal body temperature of 37^oC which is equivalent to $310.15K$, so that T=$310.15K$. We will compute the viscosity of the solvent using the equation

$$\mu=2.4152\times 10^{-5}\text{Pa}s\exp\left(\frac{4742.8\text{J/mol}}{R(T-139.86\text{K})}\right), \tag{3.2}$$

where R=8.31441J/(mol K) is the gas constant [44]. Using (3.2) we determine that the viscosity of water at body temperature is

$$\mu=6.882\times 10^{-4}\text{Pa}s=6.882\times 10^{-3}\frac{g}{cms}.$$

The molar volumes, V_1 and V_2, will be computed using

$$V_m=\frac{M}{\rho}, \tag{3.3}$$

where ρ is the mass density of the substance. Using (3.3) we acquire the molar volume of ethanol as

$$V_1^e = \frac{46.07 g/\text{mol}}{0.789 g/cm^3} = 58.39 cm^3/\text{mol}.$$

From (3.3) we acquire the molar volume of nicotine as

$$V_1^n = \frac{162.23 g/\text{mol}}{1.01 g/cm^3} = 160.624 cm^3/\text{mol}.$$

Finally, from (3.3) we acquire the molar volume of water as

$$V_2 = \frac{18.01528}{0.997 g/cm^3} = 18.07 cm^3/\text{mol}.$$

Upon inserting these values into (3.1) we acquire the diffusion coefficient as $D_e = 2.18 \times 10^{-2} \frac{cm^2}{h}$ for ethanol. Similarly, we acquire the diffusion coefficient of nicotine as $D_n = 1.56 \times 10^{-2} \frac{cm^2}{h}$.

The influx of the carcinogens for each time-step may now be computed. If males consume 5 or more drinks a day and women 4 or more drinks a day then they are considered heavy drinkers [37]. If males consume 2 drinks or less in a day and women 1 drink or less in a day, they are considered moderate drinkers [37]. Both moderate and heavy drinkers have a higher risk of developing particular head and neck cancers [6, 88]. For instance it has been found that moderate drinkers have a 1.8-fold and heavy drinkers a 5-fold higher risk of oral cavity cancer and pharynx cancer, as compared to non-drinkers [6, 88]. It has also been found that moderate drinkers have a 1.4-fold and heavy drinkers a 2.6-fold higher risk of developing larynx cancers, as compared to non-drinkers [6, 88]. A standard alcoholic beverage contains 14g of pure alcohol [136], so a moderate drinking male would be consuming 28g or less a day and a female 14g or less a day. Similarly, a heavy drinking male would be consuming 70g or more a day and a female 56g or more per day. Let us consider a heavy drinker and take the average amount per day between males and females to obtain 63g per day. If we consider that a person is typically awake 15.65 hours a day, then that means they would be consuming $4.026 \frac{g}{h}$ of alcohol. Only about 5% of the alcohol is absorbed in the mouth [113], therefore we can assume that the tongue absorbs about $0.201 \frac{g}{h}$. The average volume of the tongue is

$79.5cm^3$ [85] so the concentration per hour of alcohol absorbed by the tongue is $2.532\times10^{-3}\frac{g}{cm^3h}$. Since most of the ethanol is metabolized by the liver and none of it is metabolized by the oral cavity, then we set the outflux to be $0\frac{g}{cm^3h}$.

An individual that smokes two packs a day is more likely to develop cancer [116], we assume a typical pack contains 20 cigarettes so that they smoke 40 cigarettes a day. Most cigarettes contain $1.45\times10^{-3}g$ of nicotine, so again assuming a person is awake 15.65 hours a day, we determine that the person consumes $3.71\times10^{-3}\frac{g}{h}$. Now we use the decay formula of nicotine given by:

$$\overline{g}(x,t)=\overline{g}_0+x\left(\frac{1}{2}\right)^{\frac{2t}{3}}, \tag{3.4}$$

where t is time, $\overline{g}_0$ is the initial amount of nicotine, and x is the accumulative amount of nicotine. Using equation (3.4) as an iterator until equilibrium is reached, we calculate that the amount of nicotine that is left in the body after decay has occurred is $1.543\times10^{-4}g$. Therefore, considering the initial consumption and remaining amount the body metabolizes $3.70\times10^{-3}\frac{g}{h}$ of nicotine. Since only 15% of nicotine is metabolized by saliva then $5.55\times10^{-4}\frac{g}{h}$ is metabolized in the oral cavity. Thus, again using the volume of the tongue of $79.5cm^3$, we obtain the influx of nicotine as $7.01\times10^{-6}\frac{g}{cm^3h}$, assuming 15% is absorbed in the oral cavity, and the outflux of nicotine is $6.98\times10^{-6}\frac{g}{cm^3h}$. Thus F_c^e=$2.009\times10^{-3}\frac{g}{cm^3h}$ and F_c^n=$3.00\times10^{-8}\frac{g}{cm^3h}$.

We set the characteristic length as x_c=$N1.45\times10^{-3}cm$, where $1.45\times10^{-3}cm$ is the size of an epithelial cell [132]. We let the boundary and initial conditions be zero. The boundary condition is set at zero as we don't know how the carcinogen enters the domain at the boundary and thus assume all the influx comes from the source term. The initial condition is zero as the body doesn't naturally produce the carcinogens considered so there would be no initial base concentration within the body.

3.2.2 Gene Expression Neural Network Parameters

The four main parameters for the gene expression neural network are the two weight matrices, activation function parameter, and the mutation bias. The

weight matrix associated with the input of the neural network (2.56) is given by:

$$\tilde{\alpha}(\boldsymbol{x},t,z)=\begin{cases}1, & z\leq 0.5\\ \text{-}1, & z>0.5\end{cases}, \tag{3.5}$$

$$W_X(\boldsymbol{x},t)=\begin{bmatrix}1 & \text{-}1 & \tilde{\alpha}(\boldsymbol{x},t,z_1)10^{\text{-}7}\\ 0 & 0 & \tilde{\alpha}(\boldsymbol{x},t,z_2)10^{\text{-}7}\\ 0 & \text{-}1 & \tilde{\alpha}(\boldsymbol{x},t,z_3)10^{\text{-}7}\\ 1 & \text{-}1 & \tilde{\alpha}(\boldsymbol{x},t,z_4)10^{\text{-}7}\\ 1 & 0 & \tilde{\alpha}(\boldsymbol{x},t,z_5)10^{\text{-}7}\\ 1 & 1 & \tilde{\alpha}(\boldsymbol{x},t,z_6)10^{\text{-}7}\\ 1 & 1 & \tilde{\alpha}(\boldsymbol{x},t,z_7)10^{\text{-}7}\\ 0 & 1 & \tilde{\alpha}(\boldsymbol{x},t,z_8)10^{\text{-}7}\\ 0 & 1 & \tilde{\alpha}(\boldsymbol{x},t,z_9)10^{\text{-}7}\\ 1 & 1 & \tilde{\alpha}(\boldsymbol{x},t,z_{10})10^{\text{-}7}\end{bmatrix}, \tag{3.6}$$

where $z_i \sim U(0,1), i=1,...,10$. As insufficient data was unavailable we assumed that each carcinogen has a weight of 1, -1, or 0 for each gene depending on how the carcinogen effects that gene. For example since ethanol tends to upregulate TP53 then $W_X^{11}=1$. We assume that each gene has the same mutation rate which causes the last column in W_X, that is associated with mutations caused by replication errors due to cell age, to have one value. The mutation rate was chosen based upon the human genomic mutation rate being approximately $2.5\times10^{\text{-}8}$ per base per generation [100]. The weight matrix associated with the output of the neural network (2.58) is given by:

$$W_Y=\begin{bmatrix}1.00 & 0 & 0.01 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\ 0.01 & 0.1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\ 0.01 & 0 & 0.3 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\ 0.01 & 0 & 0 & 0.1 & 0 & 0 & 0.01 & \text{-}0.01 & 0 & 0\\ 0.01 & 0 & 0 & 0 & 0.1 & 0 & 0 & 0 & 0 & 0\\ 0.01 & 0 & 0 & 0 & 0 & 0.1 & 0 & 0 & 0 & 0\\ 0.01 & 0 & 0.01 & 0 & 0 & 0 & 0.2 & 0 & 0 & 0.01\\ 0.01 & 0 & 0 & 0 & 0 & 0 & 0 & 0.3 & 0 & 0.01\\ 0.01 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.1 & 0\\ 0.01 & 0 & 0 & 0 & 0 & 0 & 0 & 0.01 & 0 & 0.3\end{bmatrix}. \tag{3.7}$$

The main diagonal of the above matrix gives the main weights for each gene with W_Y^{11} being the highest as it is TP53. Each diagonal value was given a default of 0.1 and it is increased by 0.1 for each gene it calls or is related to, so TP53 gets a value of 1 because it is assumed all the genes relate to

TP53. Each column describes the relations between the other genes and the gene associated with the main diagonal value of that column, where if the gene is upregulated by the diagonal gene it gets a value of 0.01 and when it downregulates the gene it gets a value of -0.01. The magnitude of the values in the matrix were chosen by trial and error since there is not sufficient data to complete the matrix with accurate values. The activation function (2.57) parameter is given by:

$$\nu=10^6. \tag{3.8}$$

The value of ν results in the neural network outputting values in the range $(\frac{-1}{\sqrt{\nu}}, \frac{1}{\sqrt{\nu}}) = (-1\times10^{-3}, 1\times10^{-3})$ and was chosen so to keep the maximum amount each gene can change to a reasonable figure. Finally the mutation bias vector update function (2.60) parameter is given by:

$$\phi = 10^{-3}. \tag{3.9}$$

The value of ϕ was chosen to correspond with the maximum output value of the neural network, so that when a gene is mutated, the neural network will always output the maximum value.

3.2.3 CA Parameters

The initial seed is set such that the domain has the following breakdown of each cell type: 64.5% normal tissue cells (NTC; green), 6.5% normal stem cells (NSC; yellow), and 29% empty cells (white). The maximum number of TAC generations is given by Θ=2. The chance a cell moves when it is quiescent is 0.25. The chance a tumour cell (TC; red) or cancer stem cell (CSC; purple) randomly kills another cell during movement, proliferation, or differentiation is 0.2. The chance that an SC or MSC becomes a CSC is 2.5×10^{-6}. The chance a non stem cell becomes a stem cell through dedifferentiation is 10^{-4}. If either there are no stem cells or there are at least six empty cells in the neighbourhood of a non stem cell, then the process of dedifferentiation will be attempted. The threshold that has to be met for dedifferentiation to randomly occur is 10^{-4}. When an excision is performed the number of neighbourhoods around a TC removed is two.

Index	Gene	Gene-type	Regulation	Phenotypes
1	TP53	tumour-suppressor	down	$\uparrow$: p $\downarrow$: a, q
2	TP73	tumour-suppressor	down	$\downarrow$: a
3	RB	tumour-suppressor	down	$\uparrow$: p, d $\downarrow$: q
4	TP21	tumour-suppressor	down	$\uparrow$: p
5	TP16	tumour-suppressor	down	$\uparrow$: p
6	EGFR	oncogene	up	$\uparrow$: p
7	CCDN1	oncogene	up	$\downarrow$: a
8	MYC	oncogene	up	$\uparrow$: p, d $\downarrow$: a
9	PIK3CA	oncogene	up	$\downarrow$: a
10	RAS	oncogene	up	$\uparrow$: p, d $\downarrow$: a

Table 3.1: Provides the following properties of each gene considered in the model: index for the gene used in the various matrices and vectors required in the model, name, type of the gene, direction the gene must be regulated to become positively mutated, and how phenotypic actions are modified when the gene is positively mutated.

We consider ten genes which are given in Table 3.1. We set the mutation threshold to $\overline{M}$=0.1 and the minimum number of positively mutated genes for a cell to be considered mutated to be four [3]. Using the last two columns of Table 3.1 and assuming each phenotypic action is modified at the same magnitude we obtain the phenotypic action increment matrices (2.71) given by:

$$\overline{D}=\begin{bmatrix} 10^{-6} & -10^{-6} & -10^{-6} & 0 \\ 0 & 0 & -10^{-6} & 0 \\ 10^{-6} & -10^{-6} & 0 & 10^{-6} \\ 10^{-6} & 0 & 0 & 0 \\ 10^{-6} & 0 & 0 & 0 \\ -10^{-6} & 0 & 0 & 0 \\ 0 & 0 & 10^{-6} & 0 \\ -10^{-6} & 0 & 10^{-6} & -10^{-6} \\ 0 & 0 & 10^{-6} & 0 \\ -10^{-6} & 0 & 10^{-6} & -10^{-6} \end{bmatrix}, \tag{3.10}$$

$$\overline{U}=\begin{bmatrix} -10^{-6} & 10^{-6} & 10^{-6} & 0 \\ 0 & 0 & 10^{-6} & 0 \\ -10^{-6} & 10^{-6} & 0 & -10^{-6} \\ -10^{-6} & 0 & 0 & 0 \\ -10^{-6} & 0 & 0 & 0 \\ 10^{-6} & 0 & 0 & 0 \\ 0 & 0 & -10^{-6} & 0 \\ 10^{-6} & 0 & -10^{-6} & 10^{-6} \\ 0 & 0 & -10^{-6} & 0 \\ 10^{-6} & 0 & -10^{-6} & 10^{-6} \end{bmatrix}. \tag{3.11}$$

Using Table 3.1 we can create the gene type vector, $\boldsymbol{T}$, that is used in (2.66), (2.67), and (2.78) which is given by:

$$\boldsymbol{T}=\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}^T, \tag{3.12}$$

where 0 means the gene is a tumour-suppressor and 1 means it is an oncogene.

Gene	Gene Activation's
TP53	TP21,TP16, RB
RB	TP53, CCDN1
CCDN1	TP21
MYC	TP21 (de-activates), Ras
RAS	CCDN1, MYC

Table 3.2: Shows which genes are activated by certain genes.

Using Table 3.2 we can create the gene relationship matrix, R, that is used in (2.67) which is given by:

$$R=\begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}, \tag{3.13}$$

where 0 means the genes are not related and 1 means the genes are related. Note that in the above matrix we assumed that TP53 is related to all the genes.

The main diagonal is zero so that genes cannot modify themselves during the genetic instability phase of the model. The chance that a gene modifies the gene expression of another or that the body tries to fix the gene expression is 0.45. The maximum amount a gene expression can be changed during the gene instability stage is $\frac{1}{\sqrt{\nu}}$.

We let

$$\overline{a}_1=\frac{\tilde{c}}{\overline{c}_1}, \tag{3.14}$$

$$\overline{a}_2=\frac{\tilde{c}}{\overline{c}_2} \tag{3.15}$$

be the initial probabilities of apoptosis for a normal tissue cell and normal stem cell. Where $\tilde{c}$ is the length of the cell cycle in hours, $\overline{c}_1$ is the life span of a cell, and $\overline{c}_2$ is the life span of a stem cell. The initial phenotype matrix that is used in equation (2.76) is given by:

$$\tilde{P}=\begin{bmatrix} \overline{p}_1\overline{a}_1 & \overline{a}_1 & 1-\overline{a}_1(\overline{p}_1+1) & 0 \\ \overline{p}_1\overline{a}_1 & \frac{\overline{a}_1}{\overline{\alpha}} & 1-\overline{a}_1(\overline{\alpha}^{-1}+\overline{p}_1) & 0 \\ \overline{p}_2\overline{a}_2 & \overline{a}_2 & 1-\overline{a}_2(\overline{p}_2+1)-\overline{d}\tilde{d} & \overline{d}\tilde{d} \\ \overline{p}_2\overline{a}_2 & \frac{\overline{a}_2}{\overline{\alpha}} & 1-\overline{a}_2(\overline{\alpha}^{-1}+\overline{p}_2)-\overline{d}\tilde{d} & \overline{d}\tilde{d} \\ \overline{p}_2\overline{a}_2 & \frac{\overline{a}_2}{5\overline{\alpha}^2} & 1-\overline{a}_2((5\overline{\alpha})^{-2}+\overline{p}_2)-\overline{d}\tilde{d} & \overline{d}\tilde{d} \\ \overline{p}_1\overline{a}_1 & \frac{\overline{a}_1}{5\overline{\alpha}^2} & 1-\overline{a}_1((5\overline{\alpha})^{-2}+\overline{p}_1) & 0 \end{bmatrix}, \tag{3.16}$$

where $\overline{p}_1$ is the proliferation factor for normal tissue cell types, $\overline{p}_2$ is the proliferation factor for normal stem cell types, $\overline{\alpha}$ is the apoptotic factor, $\overline{d}$ is the differentiation factor, and $\tilde{d}$ is the probability of differentiation occurring neglecting competition between cells. The cell cycle length can range anywhere between 8 and 24 hours for the various cells in the body, since we are analyzing the tongue we will use $\tilde{c}$=10h [11]. The lifespan of a taste bud is 250±50 hours [11], so $\overline{c}_1$=250h. The lifespan of a typical stem cell is around 25550

hours [128], so $\overline{c}_2$=25550h. We set $\overline{\alpha}$=1.625, $\overline{p}_1$=0.65, $\overline{p}_2$=14.75, and $\overline{d}$=1.485 so that equilibrium in the tissue is maintained when there are no carcinogens in the domain. Note that $\overline{p}_1$ is less than 1, since we want most of the new cells to come from TACs created by SCs, because, biologically speaking, normal tissue cells rarely proliferate. Since each TAC produces a certain number of generations, given by Θ, then it will produce $2^{\Theta+1}-2$ new cells so we set

$$\tilde{d}=\frac{1}{2^{\Theta+1}-2}. \tag{3.17}$$

When a cell is a TAC the probability of proliferation increases by $\frac{1}{3}$, so that it will create its' Θ generations in as few time-steps as possible, assuming there is enough available space. The chance that a gene modifies the probability of a phenotypic action is given by 0.35. The maximum value a gene can modify the phenotypic action by is 10^{-6}.

3.3 Model Implementation

3.3.1 Parallel Implementation

Due to the vast number of cells that would need to be updated each time-step we utilized GPU parallelization to implement the model resulting in the ability to run the model in real time and vastly speeding up the generation of the results. It has been shown that utilizing asynchronous updates doesn't impact the results of a CA as compared to synchronous updates [124]. One problem with using parallelization is that if a cell has been acted upon then it should not itself be able to complete an action due to the fact that its state has been changed. Although, as long as the new state of the cell is not empty, the results are not impacted.

Issues can occur when a cell is searching for another to perform an action upon, due to the mechanism used in our CA allowing the cell to randomly choose a neighbor on which to perform an action. As the cells are updating concurrently, methods had to be developed to prevent multiple neighbours attempting to perform an action upon the same cell. We also had to ensure that while a cell is attempting to perform an action, another cell cannot be

attempting to perform an action on that cell. As well we had to ensure that as soon as an action succeeds or fails the search for a new neighbour is ceased. This issue was resolved using a lock mechanism. When a cell is attempting an action, it locks itself to any other cell attempting an action upon it in the same time-step. Similarly, if a cell currently has an action being attempted upon it by another cell, that cell is also locked. When a cell succeeds in an action, that cell is labelled as having completed an action and the cell being acted upon is updated. A cell will search its neighbours until it locates one that is unlocked upon which it attempts an action on that cell. To prevent a cell indefinitely searching it completes the task in a loop until either it succeeds or fails or it reaches a maximum number of searches. The simulation used a maximum value of 100 searches through the neighbours.

Chapter 4

Results

In this chapter we will be discussing the results of simulations of the CA model as applied to the case study discussed in the previous chapter. We will explore what impact the following have on the results: grid size, number of carcinogens, ethanol and nicotine concentration, excising the tumour versus excising the entire field, and earlier excision versus later excision.

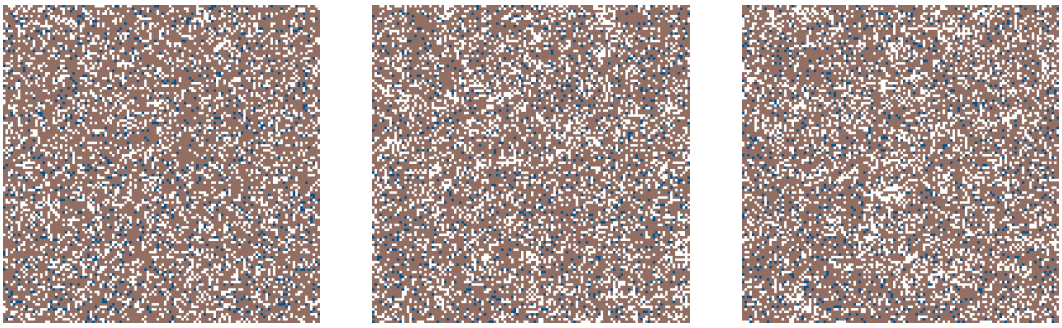
We will discuss field growth, changes in probabilities of phenotypic actions over time, mutation spread rates, and the number of lineages. By tracking lineages we will also check monoclonal versus polyclonal origins.

4.1 Equilibrium

To test that the model can maintain tissue such that the cell numbers stay at equilibrium and tumours do not sporadically appear without carcinogens, we run a simulation that does not include any carcinogens. This will show that in our base model, random mutations alone cannot cause cancer to form due to the low mutational rate of genes in the body, and the fact that the body is well adept at fixing mutations as they occur. We run the simulation on a 128×128 grid for 8766 time-steps, and as stated above, with no carcinogens.

In Figure 4.1 we present three time-steps from a simulation where no carcinogens were included, with NTC as the brown cells, NSC as the blue cells, and empty cells are white. The Figures 4.1 show (a) the initial seed, (b) the domain (tissue) at the halfway point of the simulation, and (c) the final time-step. We observe that the tissue stayed in equilibrium. The changes through

time are due to cell movement and the natural birth and death processes. The figures show that, as desired, no mutated cells (green, yellow) arise and thus no cancer stem cells (purple) or tumour cells (red) are formed.



(a) Initial Seed (b) Time-step 4383 (5 years) (c) Time-step 8766 (10 years)

Figure 4.1: This figure includes three time-steps from a simulation with parameters set at: grid size 128x128 and no active carcinogens. Using the colour map for the cell classes as provided in Table 2.1. These show (a) the initial seed of the simulation, in (b) the domain (tissue) at the halfway point of the simulation, and in (c) the final time-step.

In Figure 4.2 we present the time evolution of the fraction of cells in the different cell classes. We see that the fraction of normal tissue stays constant (with small fluctuations) and mutated cell classes never form. Figure 4.2 shows us that, as desired, the number of NSC and NTC stay approximately constant over time.

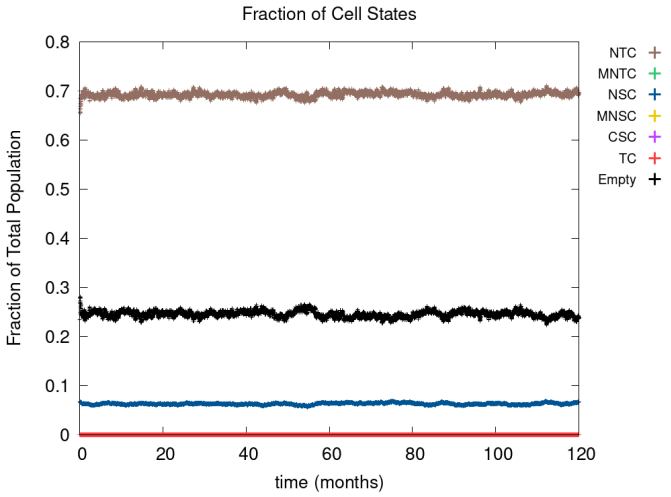


Figure 4.2: This figure shows the time course of the fraction of cells in the different cell classes NTC, MNTC, NSC, MNSC, CSC, TC, and empty. The parameters of the simulation were as follows: grid size was 128x128 and no carcinogens were active.

In Figure 4.3 we present the time evolution of the average gene expression

for each of the ten genes. We see that all of the genes maintain a normal gene expression of zero. This does not necessarily mean that the cells had a zero gene expression, but those that did were negligible due to the averaging process. We can see that none of the genes are mutated. Our model is able to maintain regular healthy tissue.

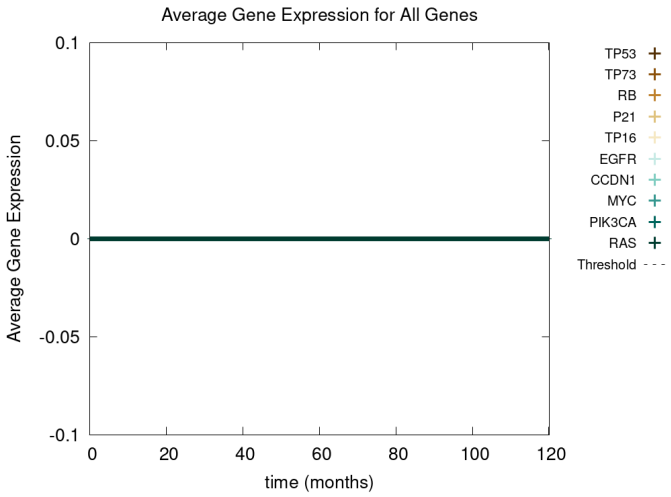
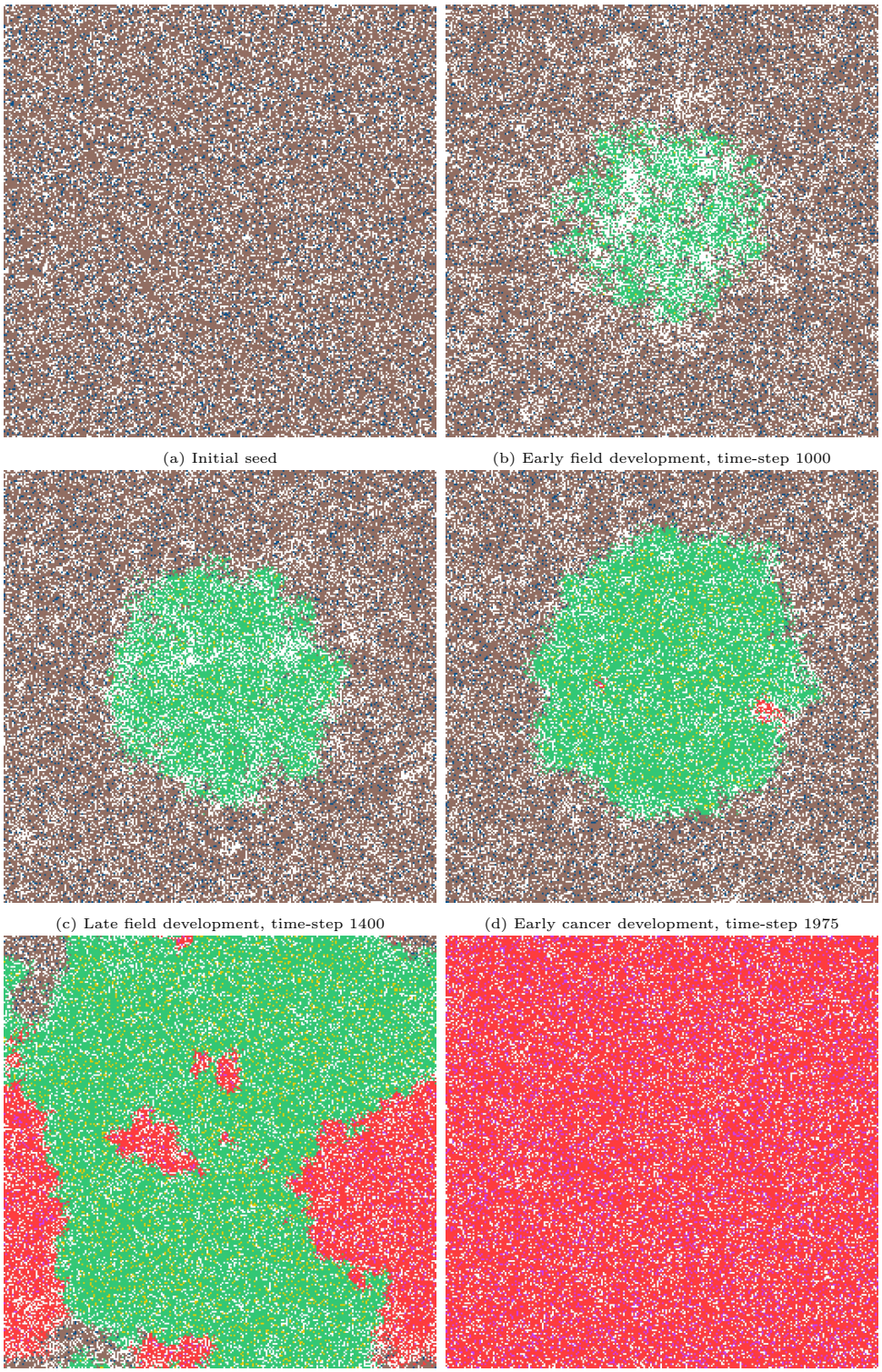


Figure 4.3: This figure shows the time course of the average gene expression for the ten genes we consider. The parameters of the simulation were as follows: grid size was 128x128 and no carcinogens were active.

4.2 General Observations

Now we study simulations where carcinogens are present and cause mutations and, ultimately, cancer. Figure 4.4 illustrates the development of a cancer field and tumours within it, where nicotine and ethanol are simulated using carcinogen spatial distribution 2 equation (4.2). The various time-steps show (a) the initial seed, (b) the cancer field at its early development, (c) the cancer field further developing prior to cancer, in (d)-(f) the multiple stages of cancer development. The colour map for the cell classes is as provided in Table 2.1. The cancer field is initially minimal and undeveloped, but over time it evolves and matures, eventually forming tumours. These tumours grow and outpace the growth rate of the cancer field, as observed from the time-steps in Figures 4.4e and 4.4f. Note that in the time-step within Figure 4.4e the tumour masses near the edge of the field begin to explore beyond the field.



(a) Initial seed

(b) Early field development, time-step 1000

(c) Late field development, time-step 1400

(d) Early cancer development, time-step 1975

(e) Late cancer development, time-step 3947

(f) Final time-step, time-step 8766

Figure 4.4: This figure includes time-steps illustrating the development of a cancer field and cancer cells using the colour map for the cell classes as provided in Table 2.1. These show (a) the initial seed, (b) the cancer field at its early development, (c) the cancer field further developing but prior to cancer, (d) the first stages of cancer development, (e) further cancer growth, and (f) the final time-step. Parameters are as follows: grid size 256x256, carcinogen spatial distribution 2, both carcinogens activated.

4.2.1 Field Development

Regardless of changes to parameters, other than carcinogens activated, we observe that the field begins to form where the carcinogen is most concentrated; this can be verified with Figure 4.4b (early field development) and 4.10b (visual representation of the carcinogen spatial distributions used in 4.4b). Initially, the field is made up of only mutated normal tissue cells (green cells) and mutated normal stem cells (yellow cells), as shown in Figure 4.4b. Typically, the first mutated cell is a MNTC due to there being a higher number of NTC compared to NSC. The field grows outwards as it takes over normal tissue.

In Figure 4.5 we show the time evolution of the fraction of mutated cells. We defined the cancer field as the areas of the domain that contain cells from the mutated cell classes thus, the figure shows the cancer field growth over time.

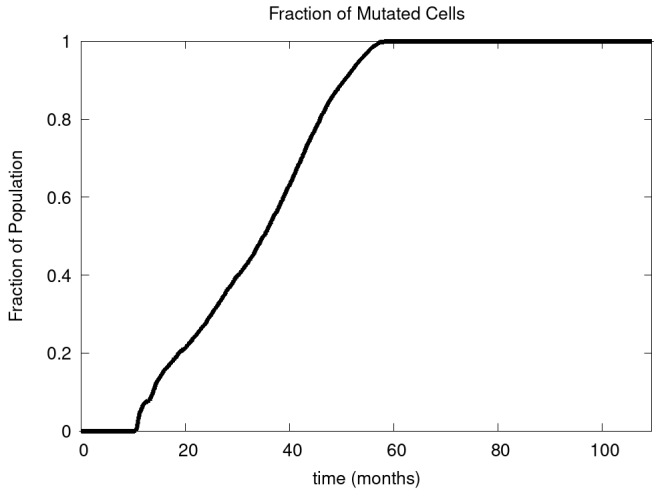


Figure 4.5: This figure shows the time course of the fraction of mutated cells, thus illustrating the cancer field growth over time. Parameters are as follows: grid size 256x256, carcinogen spatial distribution 2, both carcinogens activated.

The point at which the curve begins its rapid linear growth is that time for which the first mutated cell appears, this line flattens at the time in which the carrying capacity of the domain is achieved. Field growth rate initially starts off moderately, briefly slows, finally becoming aggressive. The period of slow growth is related to empty cells being unusually abundant because the mutated cells are not yet aggressive enough to overtake the empty cells, nor

is the stem cell to normal tissue cell ratio stabilized at this point. Further evidenced by Figure 4.4b which shows patches of empty cells (white) within the field and very few stem cells (yellow) at this point. Figure 4.5 illustrates that once the field has developed and grown large enough, the odds of a NSC or MNSC becoming a CSC increases. Soon after the emergence of the first CSC, TCs begin to form. Note that the TCs will die off if the first TC's fitness is too low compared to neighbouring cells, or if the CSC that created the TCs dies off. Once the tumour mass within the field itself starts to form, TCs and CSCs eventually take over the entire field, as shown in Figures 4.4d-4.4f. This phenomenon is a result of having limited space, and other mutated cells not being as aggressive and fit as the tumour cells.

4.2.2 Tumour Growth Rate

In Figure 4.6 we show the time evolution of the fraction of CSC and TCs which together form the tumour mass. We observe that the tumour follows logistic growth which is due to domain spatial limitations.

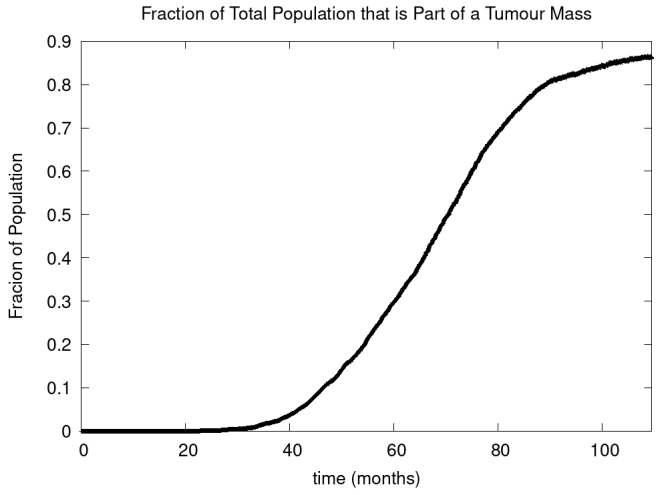


Figure 4.6: This figure shows the time course of the fraction of CSC and TC, thus illustrating the tumour growth over time. Parameters are as follows: grid size of 256x256, carcinogen spatial distribution 2, both carcinogens activated.

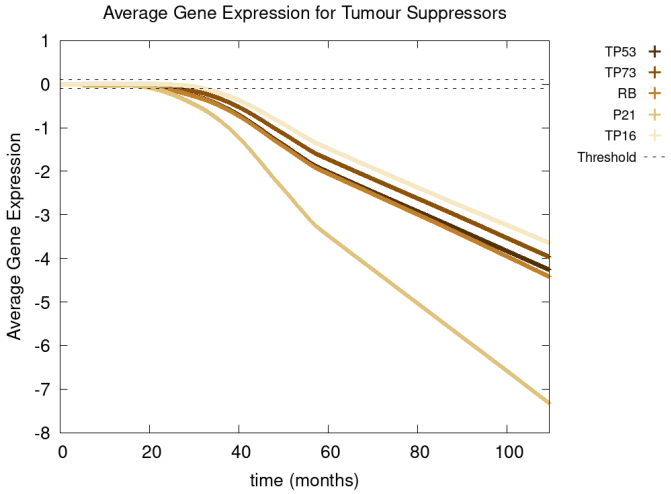
Growth is initially quite slow but rapidly increases as the cells become more aggressive from the accumulated mutations. The growth flattens, never quite achieving 100% due to the aforementioned spatial limitations of the domain

and the domain still also containing some CSCs and empty cells. Theoretically, with an infinite or growing domain the growth curves would be exponential. However, a fixed domain is more realistic biologically speaking due to the fact that the tissue surrounding the tumour has limited space. Therefore, the growth of a tumour within a given tissue should follow logistic growth, which our model confirms.

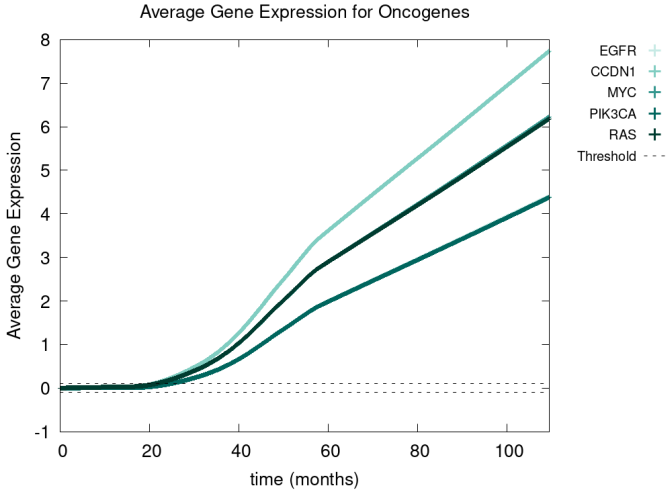
4.2.3 Mutational Evolution

Recall we use the term "positive mutation" to portend that a mutation promotes cancer, (i.e., upregulation of an oncogene or downregulation of a tumour suppressor gene). In Figure 4.7 we show various graphs that represent the mutational evolution of the genes over time. In (a) and (b) the average gene expression is illustrated for first the tumour suppressors and secondly the oncogenes. In (c) we show the time evolution of the fraction of genes that are positively mutated. In Figure 4.7a all the tumour suppressor genes are downregulated, hence positively mutated. Other than RB, which decreases at a faster rate, the gene expressions of all the other tumour suppressor genes decrease at a similar rate. Figure 4.7b displays that the oncogenes are upregulated, therefore positively mutated. The gene expressions of the oncogenes increase at a similar rate except CCDN1 and RAS, which increase at a faster rate. In Figures 4.7a and 4.7b we see that the gene expressions between the genes can vary significantly, principally with P21 and CCDN1. These two genes mutate because they are related to the most genes, and therefore have a higher weight in the MLP output weight matrix, W_y, in equation (3.7).

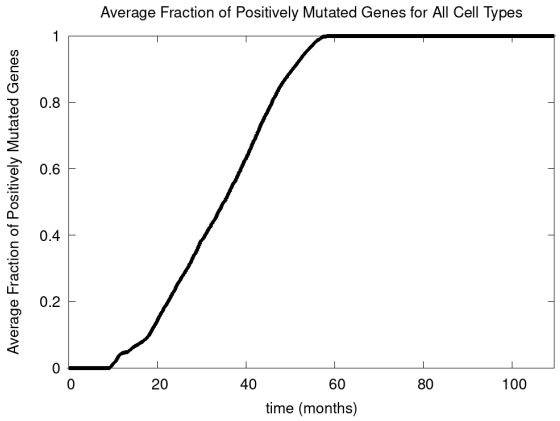
Figure 4.7c shows a lag of time before the first positively mutated genes occur, this is due to the low mutational rate, and competition against the body trying to revert mutations. The initial spike in mutational rate at the onset of the first mutated cell is due to the relative size of the domain versus the mutated cells. Once multiple genes become positively mutated the progression accelerates, due to changes in the expression of other genes caused by genetic instability, as displayed in the period starting at about 20 to 25 months. Referring to Figures 4.7a and 4.7b we can observe that all the genes



(a) Tumour suppressor gene expression over time



(b) Oncogene gene expression over time



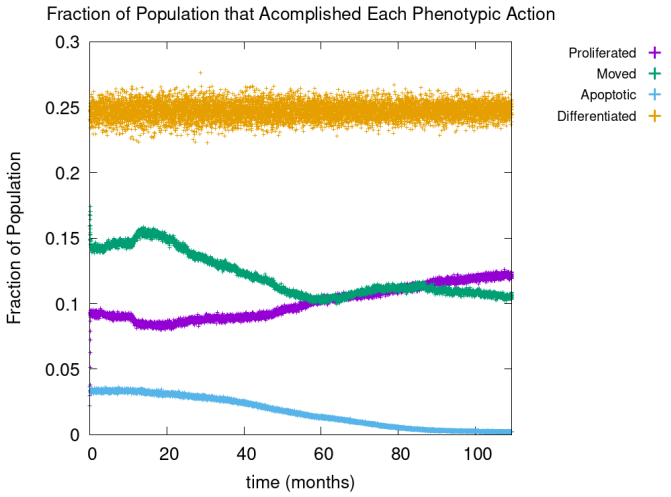
(c) Fraction of positively mutated genes over time

Figure 4.7: In this figure we show the mutational evolution of the genes. The time course of the average gene expression are shown for in (a) the tumour suppressor genes, in (b) the oncogenes. In (c) we show the time course of the fraction of genes that are positively mutated. Parameters were chosen as follows: grid size of 256x256, carcinogen spatial distribution 2 was used, both carcinogens were activated.

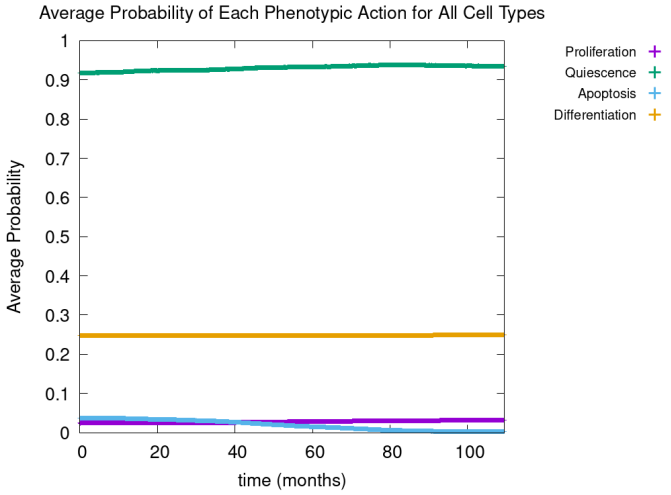
are positively mutated by about 25 months. Once TP53 is positively mutated, genes will become cancerous, as it correlates to cell cycle arrest and can make a cell apoptotic, due to genetic mutations.

4.2.4 Phenotypic Evolution

Next we consider the evolution of a phenotypic action, which includes proliferation, quiescence, apoptosis, and differentiation as functions of time. Figure 4.8 illustrates the phenotypic evolution of these actions as (a) the time evolution of the fraction of cells that underwent each phenotypic action and in (b) the time evolution of the average probability for each of these. Figure 4.8b demonstrates the probability of apoptosis occurring decreases as the cell population moves towards being cancerous. This occurs mathematically with time as the majority of the genes become positively mutated causing apoptosis to decrease as expected due to the genes modifying the probability of apoptosis. While the probability of apoptosis decreases, the chance of proliferation and differentiation increases, this again is caused by how the positively mutated genes influence the probability of proliferation and differentiation. Probability of differentiation increases at a slower rate than proliferation because fewer of the genes we consider influence differentiation. Finally, for the most part, the probability of quiescence remains stable - it goes slightly up and down, due to being balanced against the other phenotypic actions, and not many genes are influencing it, but otherwise it is at equilibrium. Figure 4.8a shows us that apoptosis and proliferation change the most over time, in particular, as apoptosis decreases, we see that proliferation increases, due to less cells dying before they can become more cancerous. This figure reveals that quiescence (movement) initially increases when the field begins forming as the field contains an insufficient number of mutated stem cells to thrive, consequently there will be more empty cells.



(a) Fraction of cells undergoing each phenotypic action

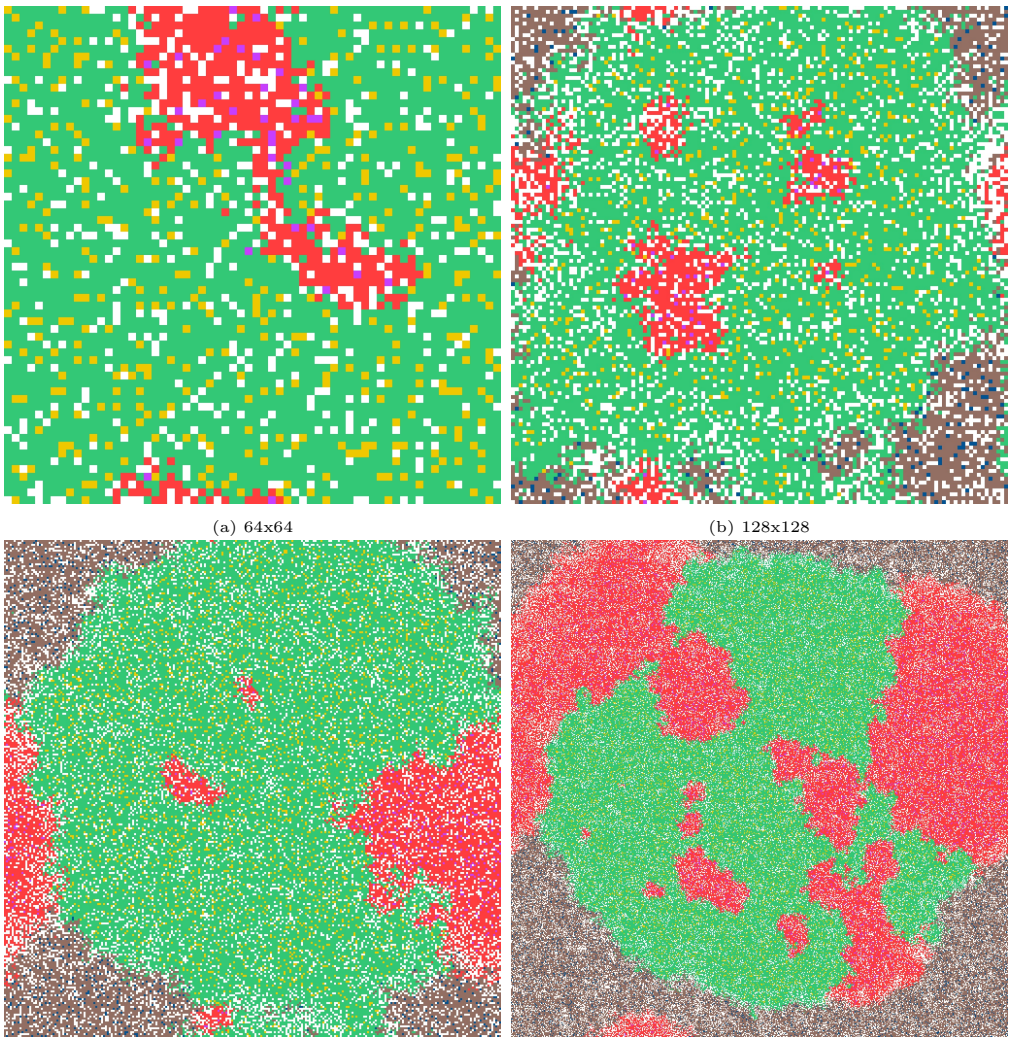


(b) Average probability for each phenotypic action over time

Figure 4.8: This figure illustrates the phenotypic evolution of proliferation, apoptosis, quiescence, and differentiation. In these we show (a) the time course of the fraction of cells that underwent each phenotypic action and (b) the time course of the average probability for each phenotypic action. Parameters are as follows: grid size 256x256, carcinogen spatial distribution 2, both carcinogens activated.

4.3 Grid Size Comparisons

In Figure 4.9 we show a sample time-step for each grid size we compare. In (a) the grid size is 64x64, in (b) the grid size is 128x128, in (c) the grid size is 256x256, and in (d) the grid size is 512x512. When comparing the grid sizes all the other parameters were the same, both carcinogens were activated and carcinogen spatial distribution 2 was used. We see that as the domain size increases, the tumour masses within it also increase.



(a) 64x64 (b) 128x128

(c) 256x256 (d) 512x512

Figure 4.9: This figure shows a time-step from each of the grid sizes that were considered for comparison. In (a) the grid size is 64x64, in (b) the grid size is 128x128, in (c) the grid size is 256x256, and in (d) the grid size is 512x512. Parameters: both carcinogens were activated and carcinogen spatial distribution 2 was used.

The various grid sizes show slight differences in four ways, all predominantly due to the increase in the number of cells. Most of the events in the CA are probabilistic, as a result, almost automatically, as we increase the size of the grid, the chance of a probabilistic event increases as well. Therefore, the overall dynamics are the same for each grid size, but the timing of various main events differ slightly as will be illustrated with the following tables.

In Table 4.1 we show the time-step at which the first mutated cell forms and the cell class of that cell. The time-step indicates no pattern to explain the differences in the values as a consequence of the random effects. We note

that the first mutated cell is always an MNTC, due to the fact that a higher ratio of NTC than NSC exists in the domain.

Grid size	Time-step first mutated cell forms	Cell class of first mutated cell
64x64	817	MNTC
128x128	748	MNTC
256x256	744	MNTC
512x512	735	MNTC

Table 4.1: In this table we compare the time-step the first mutated cell forms and the cell class that cell belongs to between the different grid sizes.

In Table 4.2, we compare the time-step the first CSC forms and the amount of time since the first mutated cell formed for each grid size. The table shows that the variation between the smallest grid size as compared to the remaining three is substantially larger. Due to the fact that the probability of a CSC forming is minuscule, attempting such an action within such a small grid size reduces the chances of these formations drastically in comparison to the larger grid sizes. Similarly, it would be expected that the opposite would be true for the largest grid size in which we observe the shortest time-step to the formation of the first CSC. With regards to the other two grid sizes we observe these to be insignificantly different demonstrating the probabilistic effects influencing the differences. We note that the time-steps since the first mutated cell was formed to the first CSC forms follow the same pattern in that the smallest grid has the largest value and the largest grid has the smallest value. This is expected due to the fact that all four grids formed the first mutated cells at very close time-steps. Overall, grid size in the two outer cases, the smallest and largest, does impact the time it takes for the first CSC to form.

Grid size	Time-step first CSC forms	# of time-steps since first mutated cell
64x64	2872	2055
128x128	1402	654
256x256	1650	906
512x512	1082	347

Table 4.2: In this table we compare the time-step the first CSC forms and the number of time-steps since the first mutated cell formed between the various grid sizes.

Table 4.3 compares the time-step the first TC forms and the amount of time since the first CSC formed for each grid size. Note that we observe basically no

time lapse between the formation of the first CSC and the first TC, therefore we observe the same differences with the formation of the first TC as we did in the prior observations of the formation of the first CSC. The 128x128 grid formed the first CSC and the first TC simultaneously, this is possible due to a cell's ability to change cell class and perform a phenotypic action in the same time-step.

Grid size	Time-step first TC forms	# of time-steps since first CSC
64x64	2875	3
128x128	1402	0
256x256	1653	3
512x512	1084	2

Table 4.3: In this table we compare the time-step the first TC forms and the elapsed amount of time it takes a CSC to form the first TC between the various grid sizes.

In Table 4.4 we compare the total number of time-steps to the final endpoint, number of time-steps since the first TC formed, and the percentage of the domain occupied by the tumour mass(es) at the endpoint for each of the grid sizes. Recall that the model is set with a loop in which the cells are acting and being acted upon either until a period of 10 years (8776 time-steps) is reached or the grid is composed of only CSC, TC, and empty cells. Note that the two grid sizes that were stopped due to the 10 years rule were the 64x64 and the 512x512. The smaller of these only reached a very small percentage of occupied tumour mass whereas the larger grid size reached similar percentages as the other two. Since the smallest grid size has insufficient cell numbers, the law of large numbers does not come into play and thus the chance of various events is too low, as a result it does not form a tumour aggressive enough to fill the majority of the domain within 10 years. With regards to the 512x512 grid size the 10 year rule stopped the tumour from further developing simply due to the vast number of cells within the domain increasing the number of time-steps. When comparing the 128x128 to the 256x256 in which the simulation ended before the 10 years, then the smaller grid size took less time to reach a similar percentage occupied of tumour mass due to a combination of less space to overcome and less competition between cell lineages. A smaller total

number of time-steps is correlated with a cancer field and tumour mass(es) being more aggressive as observed with the 128x128 grid showing the lowest number of time-steps from the first TC being formed to the endpoint.

Grid size	Total # time-steps	# time-steps since first TC	% occupied by tumour mass(es)
64x64	8766	5891	28.44%
128x128	5398	3996	84.45%
256x256	7995	6342	86.05%
512x512	8766	7682	86.51%

Table 4.4: This table compares the total number of time-steps, number of time-steps since the first TC formed, and the percentage the tumour mass(es) make up of the domain at the end of the simulation between all the grid sizes.

In Table 4.5 we show the number of tumour cell lineages at the end of the simulation for each grid size. Comparing the number of distinct tumour cell lineages between the grid sizes in Table 4.5, we see an increase relative to grid size. Since all the other lineages either died out or were overthrown via competition, the first two grid sizes end with a domain that has only one tumour cell lineage. Overall, for all the grid sizes, the number of tumour cell lineages decreases over time as the system reaches an equilibrium, where the most fit lineages survive.

Grid size	# of tumour cell lineages
64x64	1
128x128	1
256x256	6
512x512	24

Table 4.5: In this table we compare the number of tumour cell lineages at the end of the simulation between each of the grid sizes.

Due to how the parameters for the gene expression neural network were set, such as TP53 influencing all the other genes, comparing all the grid sizes result in all the genes becoming positively mutated.

4.4 Carcinogen Concentration Spatial Distribution

We consider three different carcinogen spatial distributions (CSD) given by the equations below and shown in Figure 4.10:

$$\text{CSD1: } \frac{1}{2}\left(\sin\left(\frac{2\pi x}{N}\right)\cos\left(\frac{\pi y}{N}\right)+1\right); \tag{4.1}$$

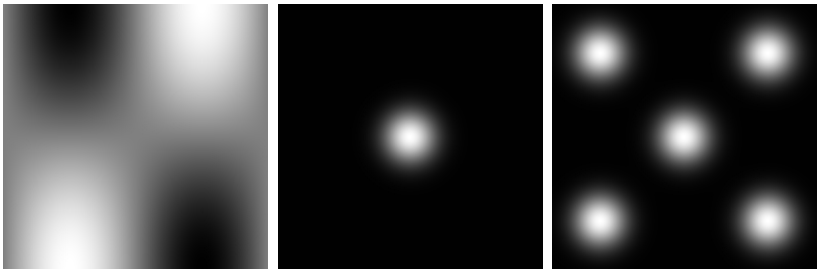
$$\text{CSD2: } \exp\left(-\frac{1}{2}\frac{(x-\mu)^2+(y-\mu)^2}{\sigma}\right), \mu=\frac{N}{2}-1, \sigma=\frac{N}{15}; \tag{4.2}$$

$$\text{CSD3: } \sum_{i=1}^{5}\exp\left(-\frac{1}{2}\frac{(x-\mu_x^i)^2+(y-\mu_y^i)^2}{\sigma}\right), \sigma=\frac{N}{15}, \tag{4.3}$$

$$\mu_x=\{\frac{N}{2}-1, \frac{N}{4}-\sigma-1, N-\frac{N}{4}+\sigma-1, \frac{N}{4}-\sigma-1, N-\frac{N}{4}+\sigma-1\},$$

$$\mu_y=\{\frac{N}{2}-1, \frac{N}{4}-\sigma-1, N-\frac{N}{4}+\sigma-1, N-\frac{N}{4}+\sigma-1, \frac{N}{4}-\sigma-1\},$$

where N is the domain size, assuming a square domain is used. The carcinogen is present within the domain starting at the initial seed (time-step zero). Since the solution of a diffusion equation and CSD2 are both Gaussian distributions the dynamics and results are comparable. Therefore, we will not discuss the case of using the diffusion equation since it only changes the timing of events relative to CSD2 because it covers a larger region of the domain. Throughout this section we consider CSD2 to be the default spatial distribution, thus other than when comparing between the spatial distributions all the comparisons are against CSD2.

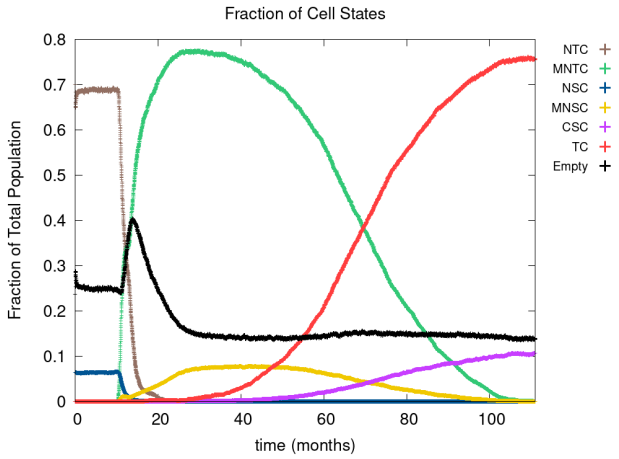


(a) CSD1 (b) CSD2 (c) CSD3

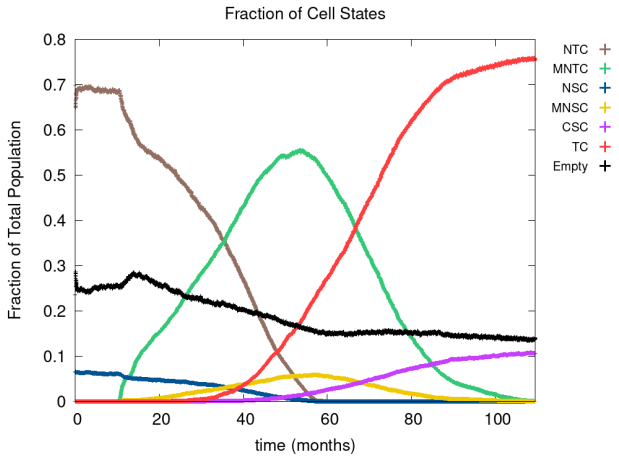
Figure 4.10: This figure shows a visual representation of each of the carcinogen spatial distributions within a 256x256 domain. In the figures we show (a) carcinogen spatial distribution 1 (CSD1), (b) CSD2, and (c) CSD3.

In Figure 4.11 we present the time evolution of the fraction of cells in the

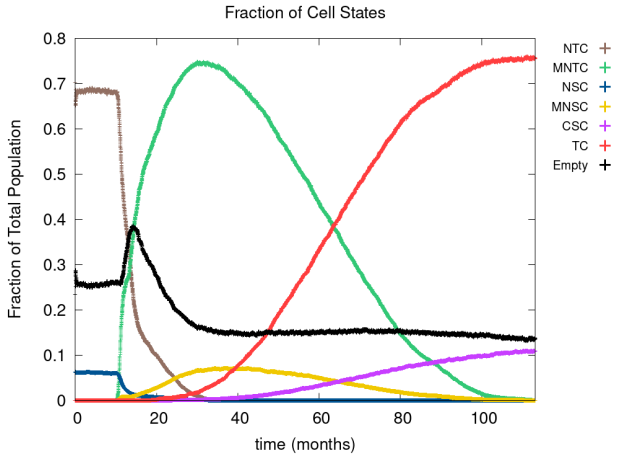
different cell classes. In (a) we use CSD1, in (b) CSD2, and in (c) CSD3. Videos of the scenarios CSD2 and CSD3 are provided at Hybrid Cellular Automata of Field Cancerization Example 1 (https://youtu.be/eKxsrSoDiKs) and Hybrid Cellular Automata of Field Cancerization Example 2 (https://youtu.be/Gtf6MoxXCkM), which each contain three simultaneous videos including from left to right the carcinogen spatial distribution, the CA grid, and a visualization of the top twenty cell lineages. We observe in all graphs that the fraction of normal tissue (brown, blue) decreases to be replaced by the cancer field (green, yellow, and purple). The cancer field is later replaced by the tumour cells (red). The fraction of MNSC decreases along with the mutated cells, while the fraction of CSCs increases as the cancer increases. The field starts to form at around 10 months as it correlates to four genes being positively mutated in at least one cell. Note that a spike in empty cells occurs soon after the beginning of field formation due to insufficient MNSC to create TAC which rejuvenate the MNTC. CSD1 results in the most expeditious development of a field and tumours, because the function covers most of the domain and has the highest average intensity within the domain. CSD3 is the next to develop a field and tumours, it is ahead of CSD2 because CSD3 has five Gaussian distributions to CSD2's one, thus events have a higher probability of occurring due to the carcinogen(s) covering more of the domain. Figure 4.11 shows us that the field develops at approximately the same rate for CSD1 and CSD3. The field does not reach as high of a fraction of the domain for CSD2, compared to CSD1 and CSD3, before the tumour cells take over.



(a) CSD1



(b) CSD2

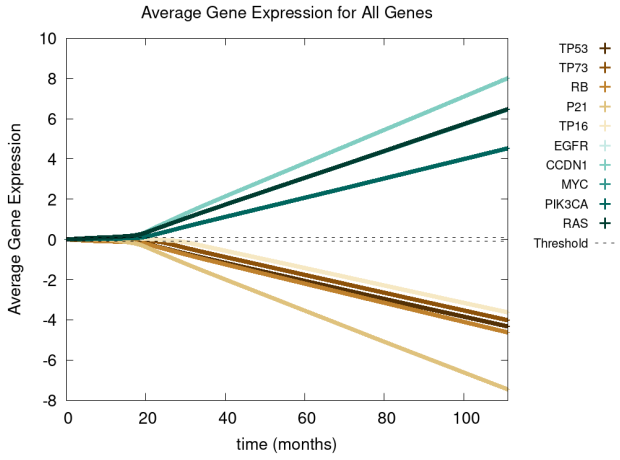


(c) CSD3

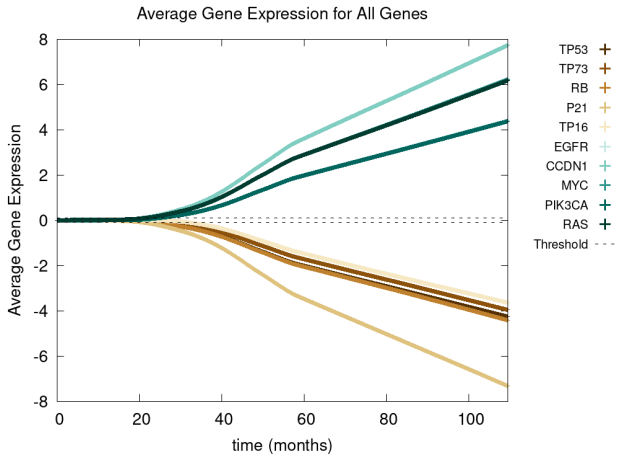
Figure 4.11: In figures (a),(b),(c) we show the time course of the fraction of cells in the different cell classes NTC, MNTC, NSC, MNSC, CSC, TC, empty. In (a) we consider carcinogen spatial distribution 1 (CSD1), in (b) CSD2, and in (c) CSD3. Parameters are as follows: grid size 256x256 and both carcinogens activated.

In Figure 4.12 we show the time evolution of the average gene expression

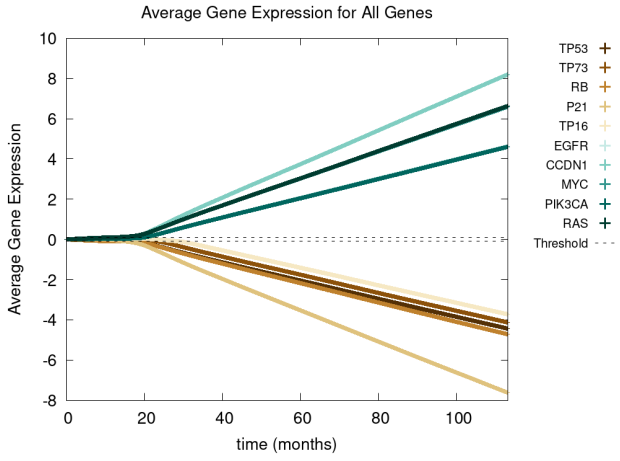
for each of the ten genes.



(a) CSD1



(b) CSD2

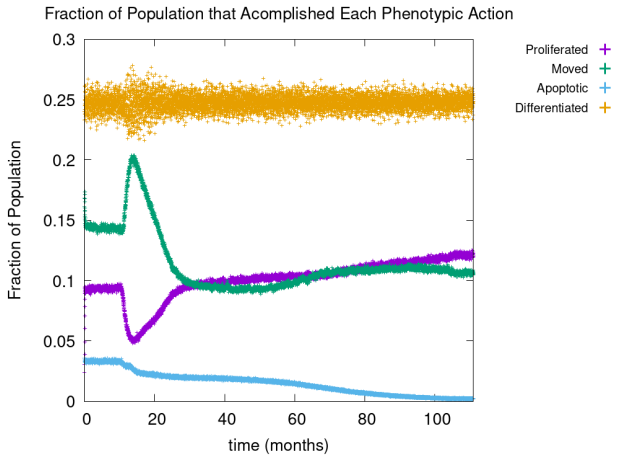


(c) CSD3

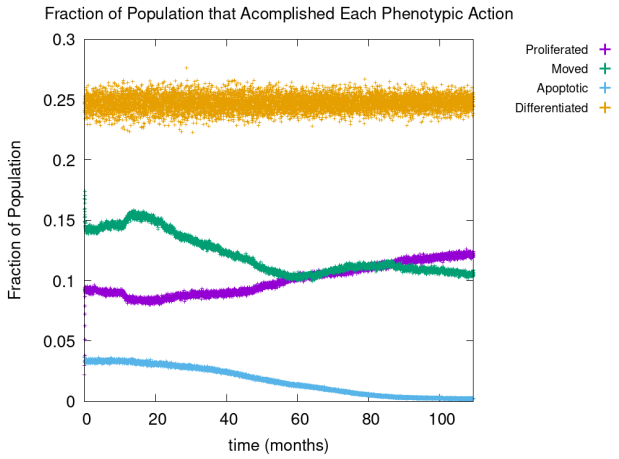
Figure 4.12: In figures (a),(b),(c) we show the time course of the average gene expression for each of the ten genes. In the plots we consider (a) carcinogen spatial distribution 1 (CSD1), (b) CSD2, and (c) CSD3. Parameters are as follows: grid size 256x256 and both carcinogens activated.

In (a) we show CSD1, in (b) we show CSD2, and in (c) we show CSD3. We see that no matter what carcinogen spatial distribution is used, all the genes become positively mutated. Figure 4.12 shows that the mutation rate for CSD1 and CSD3 is linear, whereas for CSD2 it is initially exponential before becoming linear; less cells would be mutated and as a result averaging would affect the curve.

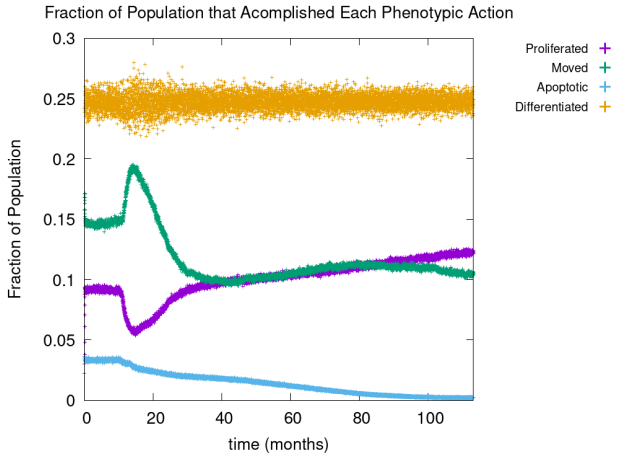
In Figure 4.13 we show the time evolution of the fraction of cells that underwent each of the phenotypic actions, namely proliferation (purple), apoptosis (blue), quiescence (moved, green), and differentiation (orange). In (a) we consider CSD1, in (b) CSD2, and in (c) CSD3. For all the carcinogen spatial distributions we see that as the cancer field develops, proliferation increases, apoptosis decreases, and cell movement and differentiation remain approximately constant. When the field begins forming there is a spike in cell movement and differentiation caused by more empty cells being available. Figure 4.13 shows that all the spatial distributions have similar phenotypic evolution. Cell movement spikes less for CSD2, as compared to CSD1 and CSD3, due to the smaller coverage of the domain. Proliferation declines more for CSD1 and CSD3, as compared to CSD2, during the early stages of the field, due to competition for empty cells from cell movement caused by higher probability of quiescence.



(a) CSD1



(b) CSD2

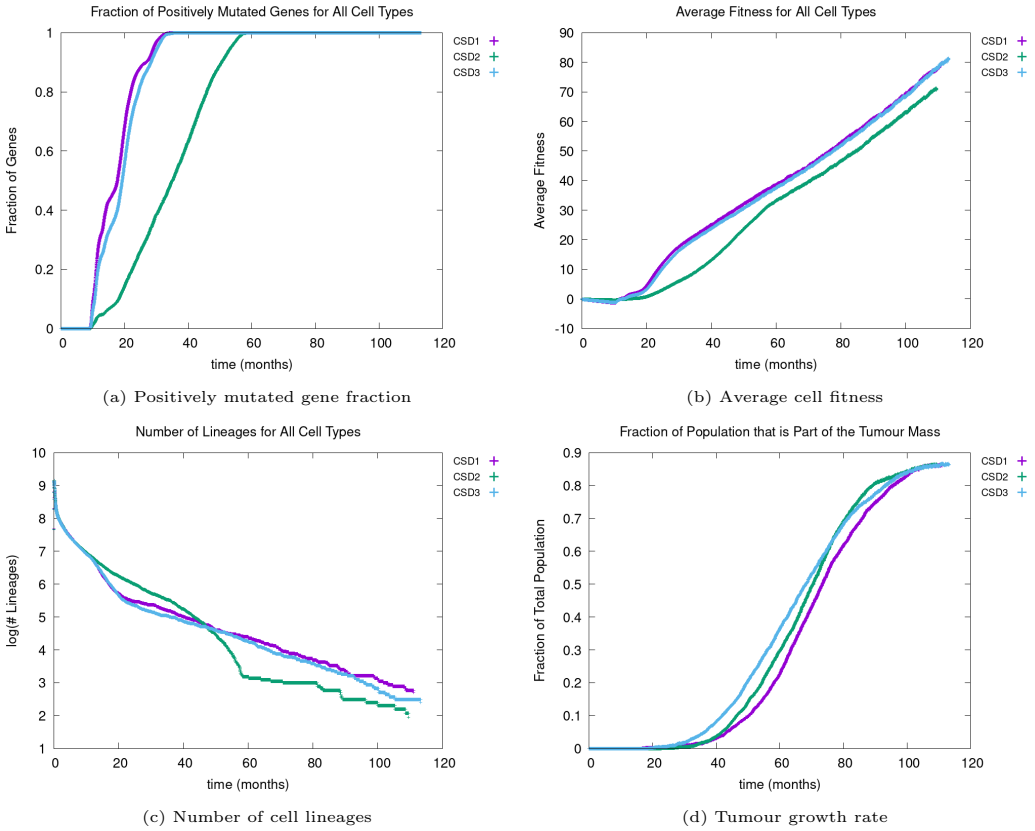


(c) CSD3

Figure 4.13: In figures (a),(b),(c) we show the time course of the fraction of cells that underwent each of the phenotypic actions, namely proliferation, apoptosis, quiescence (moved), and differentiation. In the plots we consider (a) carcinogen spatial distribution 1 (CSD1), (b) CSD2, and (c) CSD3. Parameters are as follows: grid size 256x256 and both carcinogens activated.

In Figure 4.14 we compare various characteristics of the cancer field and

cancer development between the carcinogen spatial distributions (CSDs). In the plots, we show (a) the time evolution of the fraction of positively mutated genes, (b) the time evolution of the average cell fitness, (c) the time evolution of the log of the number of cell lineages, and (d) the time evolution of the fraction of cells that are part of the tumour mass(es). Figure 4.14a shows us that all the genes become positively mutated for the three distributions, however, it takes CSD2 longer than the rest. Figure 4.14b reveals to us that the fitness increases at a similar rate for the three distributions, with CSD2 being more delayed in the increase and having a lower fitness overall. Figure 4.14c indicates that the number of cell lineages decreases, however, at the period 40-60 months CSD2 decreases rapidly until it stabilizes and continues decreasing at the same rate as the other two. This occurs because the point at which the first TC is formed the field in CSD2 is significantly smaller, thus the TCs overtake the field more quickly than the other two cases, as demonstrated in Figure 4.11. Finally, Figure 4.14d shows us that the tumour growth rate curve has a similar shape for all the distributions, with CSD1 requiring the most time to fill the domain, followed by CSD3, and CSD2 requiring the least amount of time. This phenomenon is explained by less competition occurring between lineages for CSD2. More carcinogen in the domain leads to faster development of the field.



(a) Positively mutated gene fraction

(b) Average cell fitness

(c) Number of cell lineages

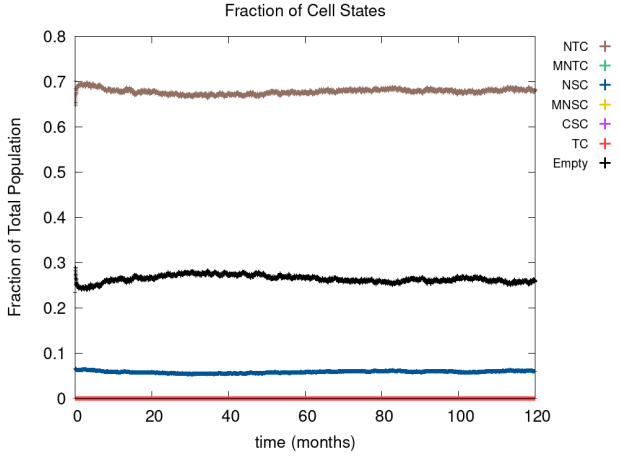
(d) Tumour growth rate

Figure 4.14: In this figure we compare various characteristics of the cancer field and cancer development between the carcinogen spatial distributions (CSD). In the plots we show (a) the time course of the fraction of positively mutated genes, (b) the time course of the average cell fitness, (c) the time course of the log of the number of cell lineages, and (d) the time course of the fraction of cells that are part of the tumour mass(es). Parameters are as follows: grid size 256x256 and both carcinogens activated.

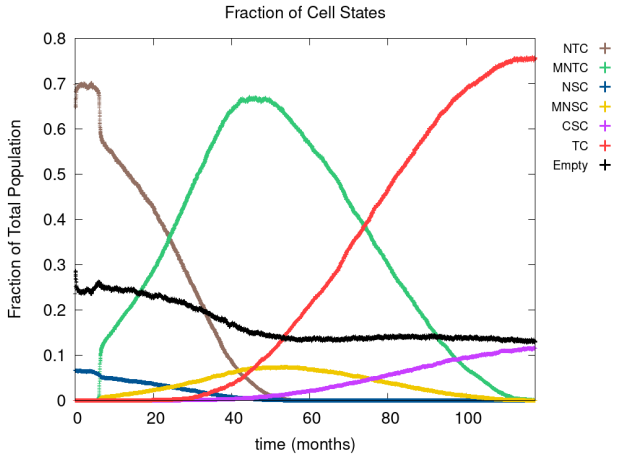
4.4.1 Single Carcinogen

In Figure 4.15 we present the time evolution of the fraction of cells in the different cell classes. In Figure 4.16 we present the time evolution of the average gene expression for each gene. For both figures in (a) we use ethanol and in (b) nicotine. We see that for ethanol (alcohol) the dynamic is very similar to the equilibrium case, except the genes are mutated slightly. Alcohol alone does not cause a field to develop due to the carcinogen causing the majority of the genes to mutate away from cancer, as shown in Figures 4.15a and 4.16a. This is in accordance with the biology that states alcohol alone rarely causes oral cancer. Nicotine alone does cause a field to develop because, unlike alcohol, the genes are positively influenced towards cancer, as is evidenced by Figures

4.15b and 4.16b. This again agrees with the actual effects of nicotine on the body, as smoking is a major cause of oral and lung cancers.

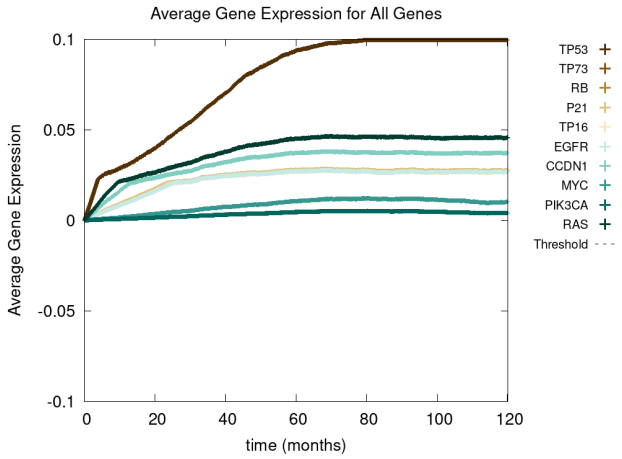


(a) Ethanol

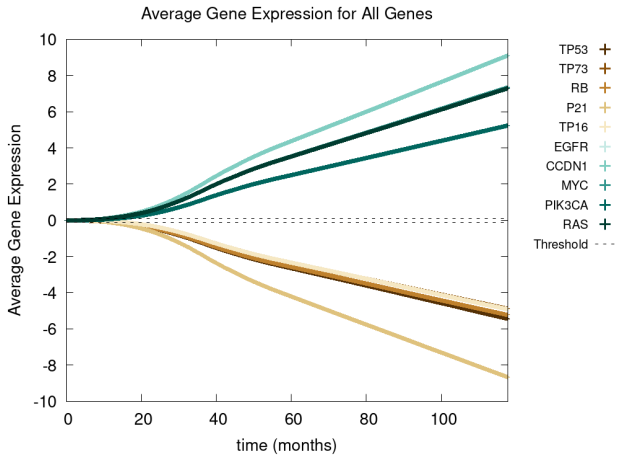


(b) Nicotine

Figure 4.15: In figures (a),(b) we show the time course of the fraction of cells in the different cell classes NTC, MNTC, NSC, MNSC, CSC, TC, empty. In the plots we consider (a) ethanol and (b) nicotine. Parameters are as follows: grid size 256x256 and carcinogen spatial distribution 2.



(a) Ethanol



(b) Nicotine

Figure 4.16: In figures (a),(b) we show the time course of the average gene expression for each of the ten genes. In the plots we consider (a) ethanol and (b) nicotine. Parameters are as follows: grid size 256x256 and carcinogen spatial distribution 2.

If we had considered a different tissue such as the liver and/or included more genes that are positively mutated by alcohol, then it is probable that alcohol would cause a field and, eventually, cancer. In particular, alcohol would likely develop cancer in the liver as it has a high mutagenic effect on the liver [6, 62, 109].

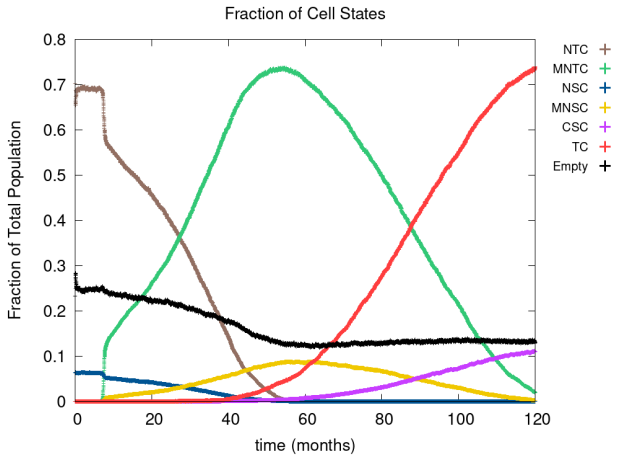
4.4.2 All Carcinogens

When both carcinogens are included, a cancer field and tumour always develop. Tumour development is faster when both carcinogens are active compared to nicotine alone, as is demonstrated by comparing Figures 4.11b and 4.15b.

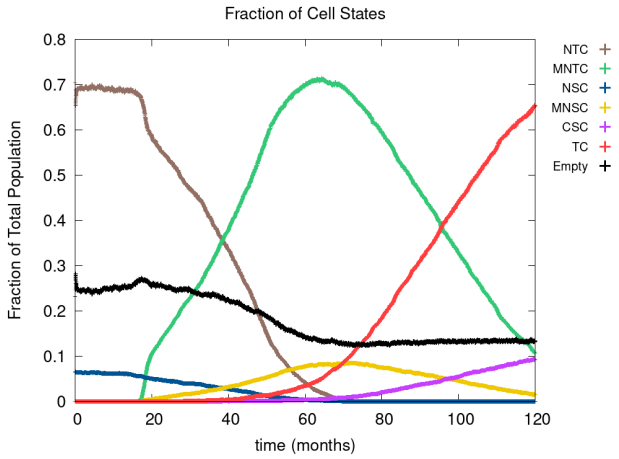
Analyzing the figures, we can conclude that although the field initiates first in the nicotine case and becomes larger, the tumour grows at a faster rate in the case that both carcinogens are active. This agrees with the clinical data, which alludes to drinking and smoking in combination increasing the chances of cancer forming relative to one of them alone [67]. Further it is known that alcohol in combination with smoking increases the chances of developing cancer because the alcohol weakens the cells, which allows the tobacco to be more mutagenic [159]. This combined impact of the carcinogens is what causes the field to be smaller than the nicotine alone. In the case of nicotine being the sole carcinogen, the field initiates faster since it does not have to fight against the mutagenic effects of alcohol. The tumour growth rate is quicker in the two-carcinogen case because there is more carcinogen within the domain, thus more gene mutations will accrue.

4.4.3 Cyclic Carcinogenic Onslaught

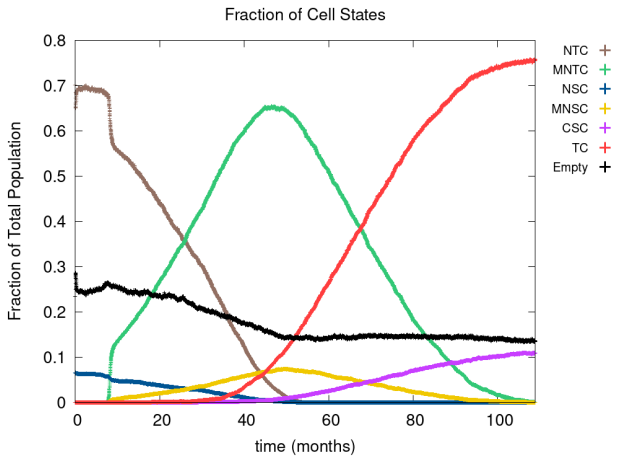
We analyze various realistic scenarios where the carcinogen(s) are removed from the domain in a cyclic fashion. For all the simulations we use a grid size of 256x256 and CSD2 (equation 4.2). In Figure 4.17 we show the time evolution of the fraction of cells in the different cell classes. In the plots, we have the case of (a) smoking every day and drinking on weekends, (b) smoking on weekends, and (c) smoking on weekdays. A video of the scenario of smoking on weekends is provided at Hybrid Cellular Automata of Field Cancerization Example 3 (https://youtu.be/4CTrhoddFOw), which contains three simultaneous videos including from left to right the carcinogen spatial distribution, the CA grid, and a visualization that shows the top twenty cell lineages. We can see the field starts to develop earlier in cases (a) and (c) relative to case (b), as evidenced by the MNTC (green) and MNSC (yellow) curves. Additionally, in case (a) the field reaches a larger size, relative to cases (b) and (c), as is established by the MNTC and MNSC curves. The first TC forms earlier in cases (a) and (c) compared to case (b). The rate of tumour growth is slightly faster in (a) and (c) as compared to (b), although (c) is slightly faster than (a).



(a) Smoke every day and drink on weekends



(b) Smoke on weekends



(c) Smoke on weekdays

Figure 4.17: In figures (a),(b),(c) we show the time course of the fraction of cells in the different cell classes NTC, MNTC, NSC, MNSC, CSC, TC, empty. In the plots we show the case of (a) smoking every day and drinking on weekends, (b) smoking on weekends, and (c) smoking on weekdays. Parameters are as follows: grid size 256x256 and carcinogen spatial distribution 2.

As noted above, the development of the field in Figure 4.17b in which the

only carcinogen is nicotine and smoking occurs on weekends only, is vastly delayed and develops much less rapidly as compared to the other two figures. This is to be expected since in this case there is a large reduction in the carcinogenic onslaught due in part to the lack of alcohol consumption and to the large reduction of nicotine. Further, we note that it takes the body up-to four days after smoking stops for it to be cleared of nicotine [12], thus nicotine will still continue to mutate the genes during the days of non-smoking.

When comparing these three scenarios it was noted that onset of cancer is slower and grows at a slightly lower rate in the scenario of smoking every day and drinking on weekends versus the scenario in which smoking occurs on weekdays only. This contradicts what was observed in the comparison of Figures 4.11b (smoking and drinking every day) and 4.15b (smoking every day) in which the combined impact of the two carcinogens caused more rapid cancer growth. Thus, it is shown that the negative impacts of the alcohol in combination with nicotine are negated somewhat when alcohol is only consumed on the weekend. In fact it would appear in this case that this small amount of alcohol fights against the mutations being caused by the nicotine.

Comparing the difference of this impact of the amount of alcohol consumption is also illustrated by comparing Figure 4.17a with 4.11b in which the alcohol consumption occurred every day along with smoking. The field develops faster when less alcohol is consumed due to the lack of alcohol during the week which would otherwise be fighting the positive mutations caused by the nicotine. The onset of cancer and the rate of cancer growth when drinking every day with smoking is significantly greater than when only drinking on the weekends. Further, confirming the impact of cancer growth when two carcinogens are introduced and the impact of the onslaught of carcinogens.

The final case is the impact of the quantity of nicotine alone, thus we compare Figure 4.17c (smoking on weekdays) versus Figure 4.15b (smoking every day). The dynamics between the two are near identical illustrating that if a heavy smoker only ceases smoking for two days per week, it is not a sufficient reduction in onslaught to impact the results. This is due to the fact that in the model once carcinogen onslaught has caused some positively

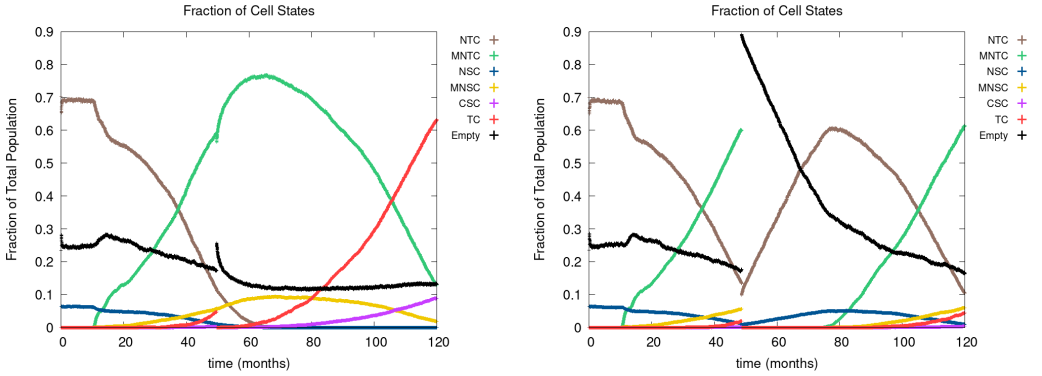
mutated genes to occur, then those genes will continue to mutate throughout the two day period.

4.5 Tumour Excision

We will be looking at how long recurrence takes after a tumour excision is performed and the dynamics of the field development after excision. We consider two types of excision, one where we only remove the TCs - which we will call keeping the field, and the other where we kill all mutated cell classes - which we will call removing the field.

In Figure 4.18 we show the time evolution of the fraction of cells in the different cell classes. In (a) we consider the case of keeping the field and in (b) the case of removing the field. Videos of the scenarios of keeping the field and removing the field are provided at Hybrid Cellular Automata of Field Cancerization Example 4 (https://youtu.be/zngGzjSlPwU) and Hybrid Cellular Automata of Field Cancerization Example 5 (https://youtu.be/EOFI4Ai1A9U), which each contain three simultaneous videos including from left to right the carcinogen spatial distribution, the CA grid, and a visualization that shows the top twenty cell lineages. The excision occurs in the period of 40-60 months, prior to this period we observe normal cancer field and tumour development. As the field develops, the number of normal tissue cells decreases as the number of mutated cells increases, with TC just starting to form and accelerate its growth and a very small uptake in CSCs beginning. At the point of excision there is a spike in the number of empty cells, which is more prominent in the removing the field case, and the number of tumour cells goes to zero. In the case for which the field is removed, Figure 4.18b, the number of mutated cells is reduced to zero and after an extended lag the field restarts its growth at about the same rate as prior. This lag in field growth is due to the extensive tissue damage that occurs. Tumour growth rate after excision including the field, is reset back to the initial time-step of zero with an increased lag before the first TC appears and a slower initial growth rate once it restarts. Recall that the TC growth lags the field growth, thus as shown in Figure 4.18b we

observe the relationship in growth between these cells. Whereas when the field is kept intact it continues to grow as shown in Figure 4.18a. Thus, when the cancer returns in this case, we observe that the tumour growth is at the same rate as before the excision, or even slightly more aggressive. Referring to Figure 4.18 we can see that the longer the delay in excision in either case, the larger the field will have grown and in the case of a) the tumour growth will be in a more accelerated state and thus when it returns will be in this more accelerated growth rate.



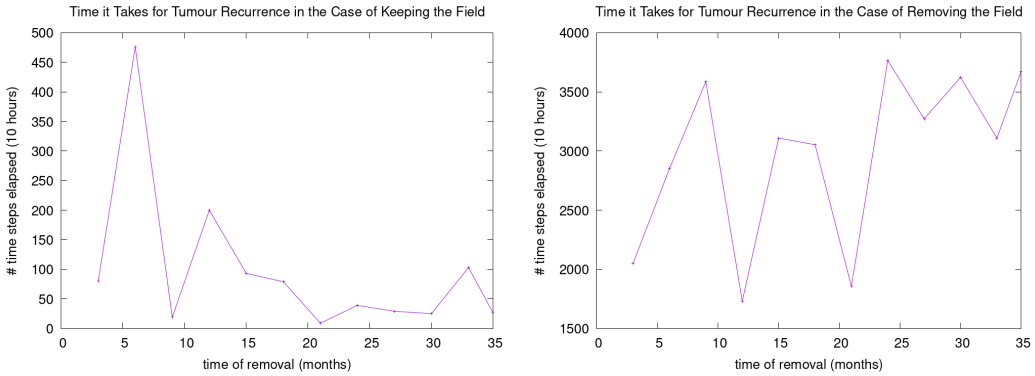
(a) Remove only tumour cells, i.e., keeping the field

(b) Remove all mutated cells, i.e., removing the field

Figure 4.18: In figures (a),(b) we show the time course of the fraction of cells in the different cell classes NTC, MNTC, NSC, MNSC, CSC, TC, empty. In the plots we consider the case where we (a) remove only TCs (keeping the field) and (b) we remove all mutated cells (removing the field). Parameters are as follows: grid size 256x256, both carcinogens activated, carcinogen spatial distribution 2, and time elapse of excision following first TC appearance was 18 months.

In Figure 4.19 we show the elapsed time it takes for cancer to recur after an excision has been performed some set amount of time after the first successful tumour cell(s) formed. In (a) we consider the case of keeping the field and in (b) we consider removing the field. Figure 4.19 in either case does not result with a linear graph due to the probabilistic nature of the CA model. In the case of keeping the field a linear downward trend is observed. This illustrates that the longer the elapsed time after the first TC forms and before excision, the less time it takes for the cancer to recur. This is to be expected because the remaining field will have developed more mutations over the course of time, is larger, and is more aggressive, which allows a tumour cell to recur faster. In the removing the field case there is no correlation between length of time before performing an excision and elapsed time before recurrence, there does,

however, seem to be an upward trend, the values are in a similar range and can be attributed to random effects. Comparing keeping the field versus removing the field it is evident that it takes substantially longer for the cancer to recur in the case of removing the field.



(a) Remove only tumour cells, i.e., keeping the field

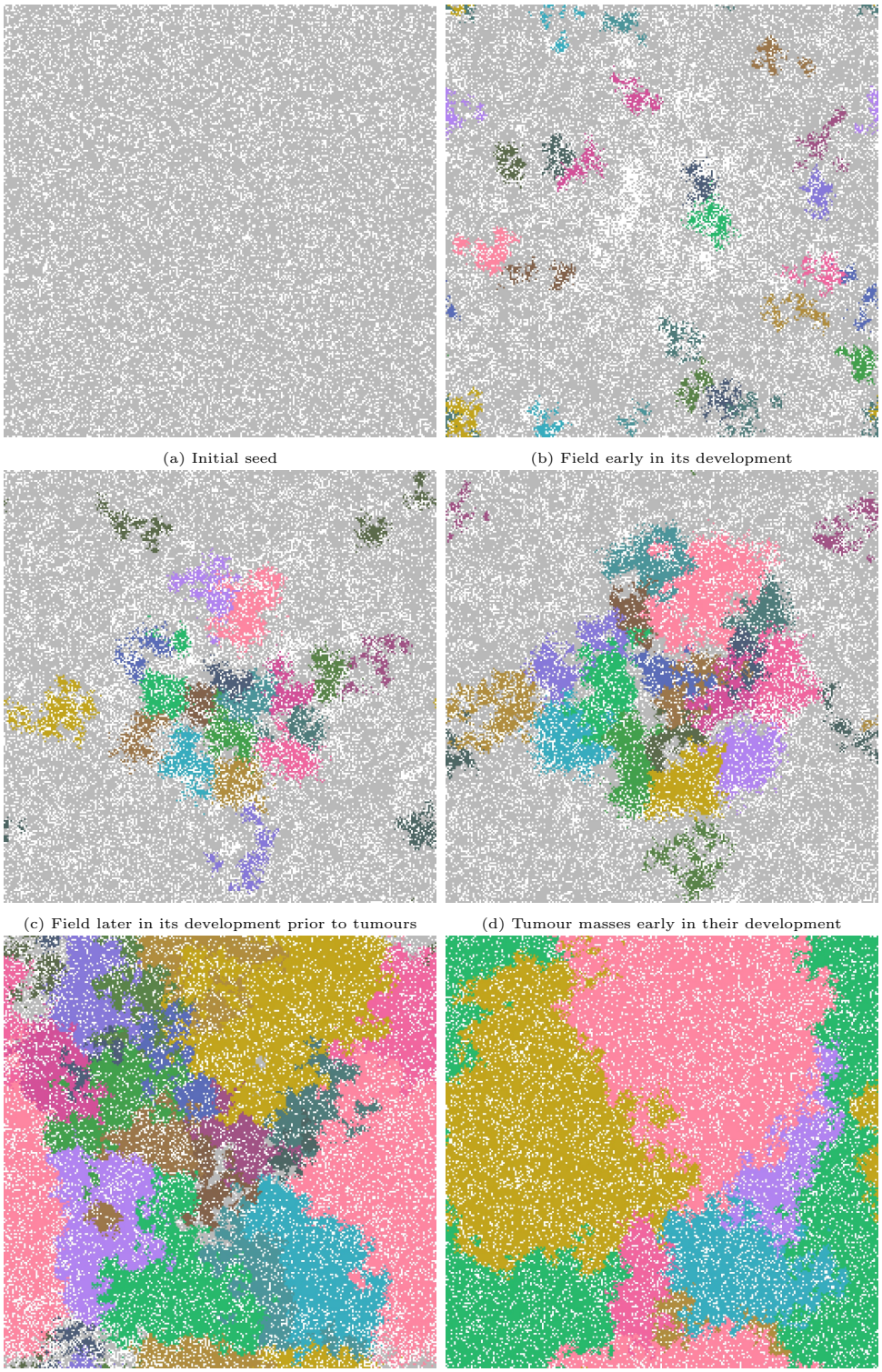
(b) Remove all mutated cells, i.e., removing the field

Figure 4.19: In figures (a),(b) we show how long it takes for cancer to recur after excision has been completed on the tumour cells that were alive for a given number of months. In the plots we consider the case where (a) we remove only TCs (keeping the field) and (b) we remove all mutated cells (removing the field). Parameters are as follows: grid size 256x256, both carcinogens activated, carcinogen spatial distribution 2, and time elapse of excision following first TC appearance was 18 months.

If we were to include a wound healing mechanism in the model, the tissue would heal more rapidly and allow the first TC to recur faster, whether keeping the field or removing it. It is also important to note that in the case of keeping the field, if the domain at the time of excision is only made up of a large tumour mass, then all the cells will be killed off. Similarly, in the case of removing the field, if the field makes up the whole domain, then all the cells in the domain will be killed. These cases would be biologically realistic only in situations in which the excision removed the whole tumour mass and/or its surrounding field within a tissue or if the whole organ was removed. Recurrence would not happen at all if after the excision the individual was no longer exposed to the carcinogenic onslaught that originated the cancer field.

4.6 Cell Lineages

In Figure 4.20 we show the top twenty cell lineages at time-steps that are relevant to the major developmental stages of field cancerization.



(a) Initial seed

(b) Field early in its development

(c) Field later in its development prior to tumours

(d) Tumour masses early in their development

(e) Tumour masses later in their development

(f) Final time-step

Figure 4.20: Important time-steps that show the top twenty lineages throughout the development stages of field cancerization. In the figures we show (a) the initial seed, (b) early cancer field formation, (c) later cancer field development, (d)-(f) cancer development. Note that light grey means the cell is not in any of the top twenty lineages and each colour represents a different cell lineage. Parameters are as follows: grid size 256x256, carcinogen spatial distribution 2, both carcinogens activated.

In the figures we show (a) the initial seed, (b) early cancer field formation, (c) later cancer field development, in (d)-(f) cancer development. The initial seed does not have any lineages because we can not track lineages until after the first phenotypic actions occur. At the beginning of the field formation the largest cell lineages are scattered everywhere, but as the field develops and grows, the largest cell lineages are concentrated at the location of the field, as shown in Figure 4.20d. Note that the largest cell lineages increase in size as the field and tumour cells become more aggressive. Figure 4.20 shows that the cell lineages that are part of the field are the largest cell lineages in the domain, which goes to show that the field contains the most fit cells in the domain. By the final time-step the remaining lineages, less than ten, are all TC lineages. Initially, there are several cell lineages, but over time the number of cell lineages decreases as the higher fitness cell lineages overcome less fit lineages.

In Figure 4.21 we illustrate a time-step from a simulation that shows the largest tumour cell lineages, each represented by a different colour. We observe that most of the lineages are concentrated near the location of the carcinogen, with smaller lineages dispersed throughout the domain. Cell lineages are competitive as is demonstrated by tumour mass cell lines enveloping or infringing on other tumour mass cell lines, displayed in Figure 4.21. Theoretically, given enough time the number of cell lineages would be reduced to one or very near to one, because certain cell lines become more and more competitive. Since we use periodic boundary conditions, we should be able to observe convergence to one lineage as time goes to infinity, this means our model will maintain multiple lineages only on sufficiently small space and time scales.

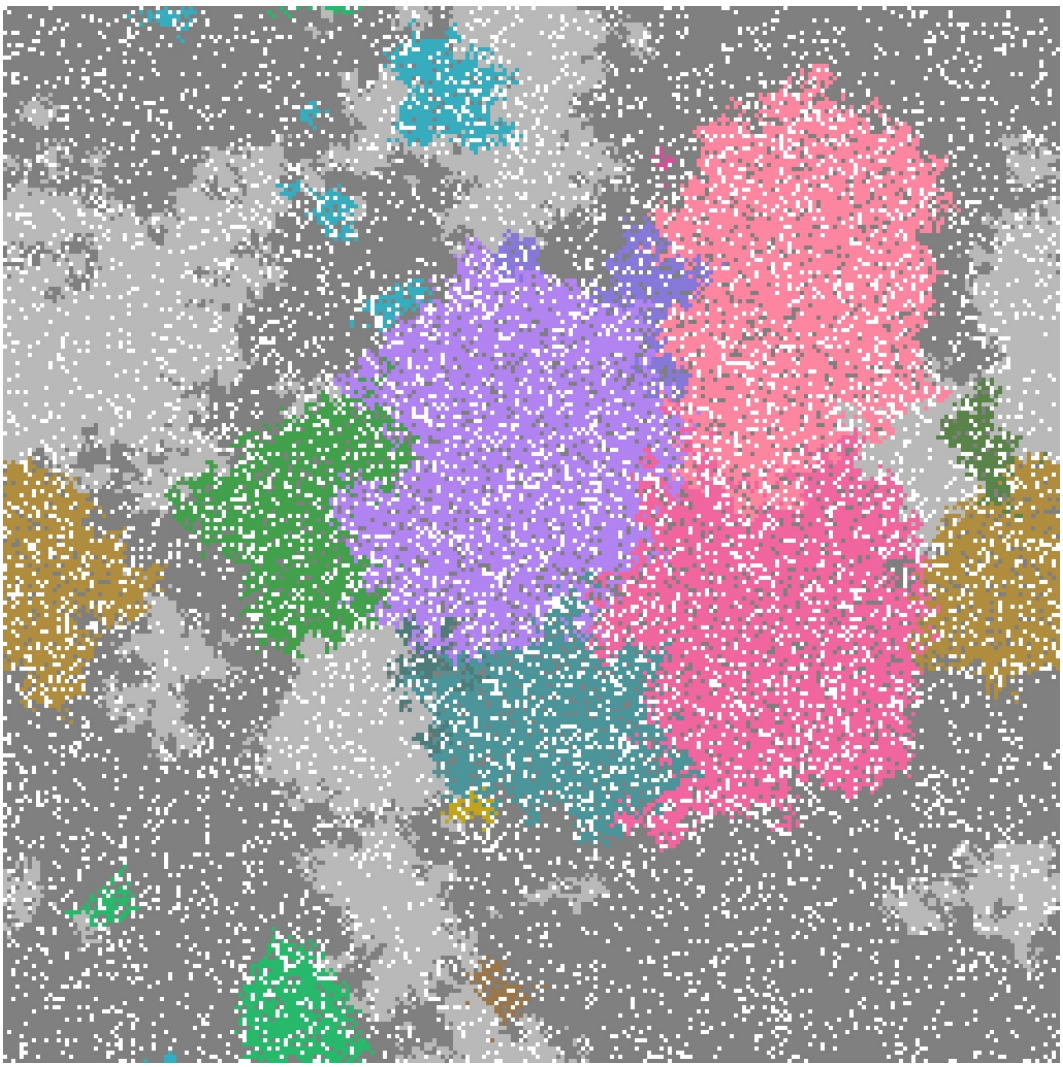


Figure 4.21: Each colour represents a different tumour cell lineage. Both the light green and light blue lineages have a main mass of tumour cells with other small masses disconnected from it, which implies they both have formed tumour masses by monoclonal origin. However, most of the tumour masses have been formed via polyclonal origin. Parameters are as follows: grid size 256x256, carcinogen spatial distribution 2, and both carcinogens activated.

4.6.1 Monoclonal versus Polyclonal Origin

We investigate whether a typical cancer field is formed via monoclonal or polyclonal origin. Our model demonstrated both, as evidenced by examining the top twenty cell lineages over time in Figure 4.20, where different cancer lineages are indicated in different colours. However, generally it seems that polyclonal origin is the more common origin, as can be verified by Figure 4.21.

The chance of monoclonal origin is associated with the amount of cell movement that occurs, since a cell has to move far enough away from its origin

to form a new tumour mass. As a result, it is possible to increase the amount of monoclonal origin by increasing the chance of movement. Polyclonal origin has no relation to movement, thus it is much more likely to occur due to the vast amount of cells and competition between cell lineages.

Chapter 5

Conclusion

5.1 Discussion of Results

In this thesis we developed a sophisticated cellular automata model for the cancer field effect. The model is an extension to existing cellular automata models [54] since I include the effect of carcinogens (ethanol and tobacco) on the gene expression of oncogenes and tumour suppressor genes. The gene expression was modelled by a multi-layer neural network, which can be trained once more data is available.

Due to the fact that the chosen genes are not positively mutated towards cancer by ethanol, our model shows that ethanol alone will not cause a field to develop. Nicotine is the most effective carcinogen in our simulations, since it causes the most positive mutations. A field does not develop without carcinogens in the domain, thus verifying the original hypothesis that field cancerization is caused by carcinogenic onslaught, and in particular that it needs consistent carcinogenic onslaught. As well, when investigating cyclic carcinogenic onslaught, we verify that the frequency of carcinogenic onslaught is important, because the less time the carcinogen is in the body, the more time the field takes to develop.

We determined that grid size does not significantly have an impact on the dynamics of the model, other than that it seems the domain size needs to be a certain size for cancer to thrive. We observed, by tracking the cell lineages, that the model shows that the tumour masses form both from monoclonal and polyclonal origin, with more polyclonal origin than monoclonal.

We demonstrated that when an excision is performed that removes only the tumour cells but leaves the remaining surrounding tissue intact, the cancer recurs faster than when removing the entire field of mutated tissue. When the field is not removed during excision the cancer that recurs is more aggressive than when the field was removed. We found that the more time the field has to develop before excision, the faster and more aggressive the recurrence.

Our research shows that for heavy smokers, the development of a cancer field is expected, which may or may not lead to cancer.

5.2 Future Work

With regards to the genes, several possible enhancements could be explored. It is possible to increase accuracy of gene expression by accounting for the fact that there are two sets of each gene, positive and negative. A dynamic mutation threshold that would alter on predefined parameters, such as the number of genes that are mutated, or on cell age, could be added. The mutation threshold could also be specific for each gene or each type of gene. In our model we assumed TP53 is related to all other genes and as a result once it is mutated all other genes will become mutated as well. However, it might be better to have an order the gene mutations must occur in such as for RAS to become mutated TP53 must first become mutated. Another possibility would be to only allow a gene to mutate when a related gene is mutated. Currently we consider ten genes in our model, in the future we want to examine using genes specifically associated with cancer formation caused by the carcinogens we consider, especially for ethanol. We could use data from lab experiments to train the gene expression neural network by looking at how each carcinogen effects the expression of each gene. It would be of benefit to include viral infections to the model such as human papillomavirus (HPV) as input to the gene expression neural network. We would want to include genetic precursors towards cancer as input to the gene expression neural network as well. Non-carcinogen mutagen factors to consider adding to the gene expression neural network or a new neural network or through a correlation matrix, include;

gender, age and lifestyle. It would be beneficial to be able to use a patient specific data set, such as genetic sequencing to improve accuracy of the values in the gene expression neural network and carcinogen exposure information such as how often the patient smokes.

Future biological mechanisms to add to our model would incorporate wound healing, cell metabolism, micro-environment variables, and cell adhesion. Telomeres are at the end of the DNA strands and with each cell division they get cut shorter, eventually becoming so short that the cell can no longer proliferate and so will enter senescence. Senescent cells are similar to quiescent cells except they can't perform any actions and at some point undergo apoptosis. Therefore, the model could be enhanced by introducing telomeres. How stem cells are distributed in the domain could be more accurately represented by using a stem cell niche instead of allowing stem cells to distribute randomly in the grid. Currently the model only forms tumour cells from cancer stem cells, whereas we could allow tumour cells to randomly appear, or transition other cell classes into tumour cells. Currently a non-mutated cell class can transition to its associated mutated class as long as it has a certain number of positively mutated genes, it would be more accurate to specify certain genes that must be positively mutated before a transition to a mutated class is allowed. A Transient amplifying cell (TAC) currently doesn't fully differentiate until it succeeds in producing its allotted number of cell generations, it would be more realistic to have the cell fully differentiate after some rule has been met, such as there being a sufficient number of cells surrounding it or if it has been a certain number of time-steps since first being formed. The SC classes could be given an extended neighbourhood to search for dead tissue and make a TAC to regenerate it. Other cells types could be included such as muscle, fat, and white blood cells. Carcinogens could be permitted to hinder phenotypic actions from occurring, or allow the direct killing of cells. We could include immune-response mechanisms that would randomly kill mutated cells to replicate the body eliminating rouge cells. Carcinogens could also be allowed to weaken cells so that more genetic and phenotypic mutations can be accomplished by other carcinogens. The carcinogens currently spread

via carcinogen spatial distributions that don't depend on time, it would be interesting to consider spatial distributions that can vary over time, such as allowing it to increase in size or move location. We currently consider only one type of carcinogen spatial distribution for each carcinogen in a simulation, instead we could allow each carcinogen to have different carcinogen spatial distributions depending on the type of carcinogen. Cyclic carcinogen onslaught could be improved by including a rule for a carcinogen where it is permanently deactivated, for example due to a person quitting smoking. Instead of substituting nicotine as the carcinogen for tobacco we could consider the key carcinogens of tobacco. The chance of occurrence of each phenotypic action could be estimated from lab experiments, by recording how often cells differentiate, proliferate, and move. A case analysis between different types of cancers could be achieved by applying our model to other cancer types. The CA grid could be initialized with mutations and mutated cells already within the domain. Grid initialization could utilize patient-specific parameters such as age or genetic tests.

Running the model through more simulations and parameters would help to determine a rough estimate of the field size and the parameter values. The ability to tune the grid size and type of lattice to match the tissue and cancer type would be beneficial. More simulations for a broader base of cases could be run such that we are able to determine averages and trends. Some CA parameter analysis could be accomplished by modifying for example the chance a cell moves. Eventually we would want to run the model in a three-dimensional domain. One of the questions we originally wanted to answer was how long it would take for a tumour to become large enough to be detected by physicians, however, we were not able to answer this question due to the size of the cells requiring at least a domain size 1024x1024 to represent the required $1cm$ detection size. A few simulations at 1024x1024 were run and we found it would take more than 10 years to fill in the space, thus it would take at least 10 years for the tumour to be detectable. Another possibility would be to use a three-dimensional domain, which would require a grid size of only 128x128x64. With regards to the CA we could consider various types of neighbourhoods

such as Neumann or extended Moore. All the random samples in the CA are currently taken from a uniform distribution, thus investigation could be conducted on other distributions and potentially a mixture of distributions based upon the use case of the sample point.

With regards to efficiency of running the model, as the complexity increases, the speed of the calculations involved in the gene expression neural network could be improved with linear algebra libraries in CUDA. Using texture memory in the GPU to store cell neighbourhoods would make calculations both faster and easier, as it has faster bandwidth and built in boundary conditions. The code could be made more cross compatible by allowing parallel computation on the CPU and switching from CUDA to OpenCL. Running the model with cloud computing would be a better option for a program built for hospital use.

References

1. Alcolea, M. P. *et al.* Differentiation imbalance in single oesophageal progenitor cells cause clonal immortalization and field change. *Nature Cell Biology* **16,** 612–619 (2014) (cit. on p. 3).

2. Alonso, S. *et al.* Methylation of MGMT and ADAMTS14 in normal colon MUCOSA: biomarkers of a field defect for cancerization preferentially targeting elder African-Americans. *Oncotarget* **6,** 3420–3431 (2015) (cit. on p. 6).

3. Anandakrishnan, R. *et al.* Estimating the number of genetic mutations (hits) required for carcinogenesis based on the distribution of somatic mutations. *PLoS computational biology* **15,** e1006881 (2019) (cit. on p. 73).

4. Angadi, P. V. *et al.* Oral field cancerization: current evidence and future perspectives. *Oral and Maxillofacial Surgery* **16,** 171–180 (2012) (cit. on pp. 6, 8).

5. Asada, K. *et al.* FHL1 on chromosome X is a single-hit gastrointestinal tumor-suppressor gene and contributes to the formation of an epigenetic field defect. *Oncogene* **32,** 2140–2149 (2013) (cit. on p. 6).

6. Bagnardi, V. *et al.* Alcohol consumption and site-specific cancer risk: a comprehensive dose-response meta-analysis. *Br J Cancer* **112,** 580–593 (2015) (cit. on pp. 69, 101).

7. Baker, A.-M. *et al.* Quantification of Crypt and Stem Cell Evolution in the Normal and Neoplastic Human Colon. *Cell Reports* **8,** 940–947 (2014) (cit. on p. 3).

8. Beckmann, J. *et al.* Asymmetric cell division within the human hematopoietic stem and progenitor cell compartment: identification of asymmetrically segregating proteins. *Blood* **109,** 5494–5501 (2007) (cit. on p. 4).

9. Bedi, G. C. *et al.* Multiple head and neck tumors: evidence for a common clonal origin. *Cancer Research* **56,** 2484–2487 (1996) (cit. on p. 7).

10. Beerenwinkel, N. *et al.* Cancer Evolution: Mathematical Models and Computational Inference. *Systematic Biology* **64,** e1–e25 (2014) (cit. on p. 11).

11. Beidler, L. M. & Smallman, R. L. Renewal of cells within taste buds. *The Journal of cell biology* **27,** 263–272 (1965) (cit. on p. 75).

12. Benowitz, N. L., Hukkanen, J. & Jacob, P. Nicotine chemistry, metabolism, kinetics and biomarkers. *Nicotine psychopharmacology,* 29–60 (2009) (cit. on p. 104).

13. Biddle, A. *et al.* Cancer Stem Cells in Squamous Cell Carcinoma Switch between Two Distinct Phenotypes That Are Preferentially Migratory or Proliferative. *Cancer Research* **71,** 5317–5326 (2011) (cit. on p. 5).

14. Bjerkvig, R. *et al.* The origin of the cancer stem cell: current controversies and new insights. *Nature Reviews Cancer* **5,** 899–904 (2005) (cit. on p. 5).

15. Blokzijl, F. *et al.* Tissue-specific mutation accumulation in human adult stem cells during life. *Nature* **538,** 260–264 (2016) (cit. on p. 2).

16. Boscolo-Rizzo, P. *et al.* Telomere shortening in mucosa surrounding the tumor: Biosensor of field cancerization and prognostic marker of mucosal failure in head and neck squamous cell carcinoma. *Oral Oncology* **51,** 500–507 (2015) (cit. on p. 6).

17. Bowling, S., Lawlor, K. & Rodríguez, T. A. Cell competition: the winners and losers of fitness selection. *Development* **146,** 1–12 (2019) (cit. on p. 59).

18. Braakhuis, B. J. M. *et al.* A Genetic Explanation of Slaughter's Concept of Field Cancerization. *Cancer Research* **63,** 1727–1730 (2003) (cit. on pp. 6–9).

19. Brennan, J. A. *et al.* Molecular Assessment of Histopathological Staging in Squamous-Cell Carcinoma of the Head and Neck. *New England Journal of Medicine* **332,** 429–435 (1995) (cit. on p. 6).

20. Bu, Y. & Cao, D. The origin of cancer stem cells. *Frontiers in Bioscience* **Scholar,** 819–830 (2012) (cit. on p. 5).

21. Cabanillas, R. & Llorente, J. L. The Stem Cell Network model: clinical implications in cancer. *European Archives of Oto-Rhino-Laryngology* **266,** 161–170 (2009) (cit. on p. 5).

22. Calabrese, P., Tavaré, S. & Shibata, D. Pretumor progression: clonal evolution of human stem cell populations. *The American Journal of Pathology* **164,** 1337–1346 (2004) (cit. on p. 2).

23. Califano, J. *et al.* Genetic Progression Model for Head and Neck Cancer: Implications for Field Cancerization. *Cancer Research* **56,** 2488–2492 (1996) (cit. on p. 6).

24. Califano, J. *et al.* Second Esophageal Tumors in Patients with Head and Neck Squamous Cell Carcinoma: An Assessment of Clonal Relationships. *Clinical Cancer Research* **5,** 1862–1867 (1999) (cit. on pp. 7, 8).

25. *Carcinogens in Tobacco Smoke* `https://www.canada.ca/en/health-canada/services/publications/healthy-living/carcinogens-tobacco-smoke.html` (2022) (cit. on p. ii).

26. Catalano, V. *et al.* Tumor and its microenvironment: A synergistic interplay. *Seminars in Cancer Biology* **23,** 522–532 (2013) (cit. on p. 5).

27. Cense, H. A. *et al.* A patient with seven primary tumors of the upper aerodigestive tract: the process of field cancerization versus distant monoclonal expansion. *Diseases of the Esophagus* **10,** 139–142 (1997) (cit. on p. 6).

28. Chang, Y.-L. *et al.* Clonality and prognostic implications of p53 and epidermal growth factor receptor somatic aberrations in multiple primary lung cancers. *Clinical Cancer Research* **13,** 52–58 (2007) (cit. on p. 6).

29. Chu, T.-Y. *et al.* Monoclonality and surface lesion-specific microsatellite alterations in premalignant and malignant neoplasia of uterine cervix: a local field effect of genomic instability and clonal evolution. *Genes, Chromosomes and Cancer* **24,** 127–134 (1999) (cit. on p. 6).

30. Clayton, E. *et al.* A single type of progenitor cell maintains normal epidermis. *Nature* **446,** 185–189 (2007) (cit. on p. 3).

31. Curtius, K., Wright, N. A. & Graham, T. A. An evolutionary perspective on field cancerization. *Nature Reviews Cancer* **18,** 19–32 (2017) (cit. on pp. 1–4, 9, 10).

32. Damania, D. *et al.* Nanocytology of rectal colonocytes to assess risk of colon cancer based on field cancerization. *Cancer Research* **72,** 2720–2727 (2012) (cit. on p. 6).

33. Davis, H. *et al.* Aberrant epithelial GREM1 expression initiates colonic tumorigenesis from cells outside the stem cell niche. *Nature Medicine* **21,** 62–70 (2015) (cit. on p. 2).

34. de Pillis, L. *et al.* Mathematical model creation for cancer chemo-immunotherapy. *Computational and Mathematical Methods in Medicine* **10,** 165–184 (2009) (cit. on p. 11).

35. Di Fiore, R. *et al.* Mutant p53 gain of function can be at the root of dedifferentiation of human osteosarcoma MG63 cells into 3AB-OS cancer stem cells. *Bone* **60,** 198–212 (2014) (cit. on p. 5).

36. Doupé, D. P. *et al.* A Single Progenitor Population Switches Behavior to Maintain and Repair Esophageal Epithelium. *Science* **337,** 1091–1093 (2012) (cit. on p. 3).

37. *Drinking Levels Defined* `https://www.niaaa.nih.gov/alcohol-health/overview-alcohol-consumption/moderate-binge-drinking` (2021) (cit. on p. 69).

38. Durrett, R., Foo, J. & Leder, K. Spatial Moran models, II: cancer initiation in spatially structured tissue. *Journal of mathematical biology* **72,** 1369–1400 (2016) (cit. on p. 11).

39. Dworkin, A. M., Huang, T. H.-M. & Toland, A. E. Epigenetic alterations in the breast: Implications for breast cancer detection, prognosis and treatment. *Seminars in Cancer Biology* **19,** 165–171 (2009) (cit. on p. 6).

40. Ellsworth, D. L. *et al.* Outer Breast Quadrants Demonstrate Increased Levels of Genomic Instability. *Annals of Surgical Oncology* **11,** 861–868 (2004) (cit. on p. 6).

41. Enderling, H. & Chaplain, M. Mathematical modeling of tumor growth and treatment. *Current pharmaceutical design* **20,** 4934–4940 (2014) (cit. on p. 11).

42. *Epithelium* `https://my.clevelandclinic.org/health/articles/22062-epithelium` (2022) (cit. on p. 2).

43. Feller, L. L. *et al.* Oral squamous cell carcinoma in relation to field precancerisation: pathobiology. *Cancer Cell International* **13,** 1–8 (2013) (cit. on p. 5).

44. FogelŚon, R. L. & Likhachev, E. R. Temperature dependence of viscosity. *Technical Physics* **46,** 1056–1059 (2001) (cit. on p. 68).

45. Foo, J., Leder, K. Z. & Ryser, M. D. Multifocality and recurrence risk: a quantitative model of field cancerization. *Journal of Theoretical Biology* **355,** 170–184 (2014) (cit. on pp. 11, 12, 15–18).

46. Foschini, M. P. *et al.* Genetic clonal mapping of in situ and invasive ductal carcinoma indicates the field cancerization phenomenon in the breast. *Human Pathology* **44,** 1310–1319 (2013) (cit. on p. 6).

47. Franklin, W. A. *et al.* Widely dispersed p53 mutation in respiratory epithelium. A novel mechanism for field carcinogenesis. *The Journal of Clinical Investigation* **100,** 2133–2137 (1997) (cit. on p. 6).

48. Franssen, L. C. *et al.* A mathematical framework for modelling the metastatic spread of cancer. *Bulletin of mathematical biology* **81,** 1965–2010 (2019) (cit. on p. 11).

49. Galandiuk, S. *et al.* Field Cancerization in the Intestinal Epithelium of Patients With Crohn's Ileocolitis. *Gastroenterology* **142,** 855–864 (2012) (cit. on p. 6).

50. Galipeau, P. C. *et al.* Clonal expansion and loss of heterozygosity at chromosomes 9p and 17p in premalignant esophageal (Barrett's) tissue. *Journal of the National Cancer Institute* **91,** 2087–2095 (1999) (cit. on p. 6).

51. Gatenby, R. A. & Gillies, R. J. A microenvironmental model of carcinogenesis. *Nature Reviews Cancer* **8,** 56–61 (2008) (cit. on p. 1).

52. *Genes related to Squamous Cell Carcinoma, Head and Neck* `https://www.malacards.org/card/squamous_cell_carcinoma_head_and_neck?limit[RelatedGenes]=286%5C#RelatedGenes-table` (2021) (cit. on p. 67).

53. Genovese, G. *et al.* Clonal Hematopoiesis and Blood-Cancer Risk Inferred from Blood DNA Sequence. *New England Journal of Medicine* **371,** 2477–2487 (2014) (cit. on p. 6).

54. Gerlee, P. & Anderson, A. R. An evolutionary hybrid cellular automaton model of solid tumour growth. *Journal of theoretical biology* **246,** 583–603 (2007) (cit. on pp. 11, 19, 112).

55. Géron, A. in. 1st ed. Chap. Chapter 10. Introduction to Artificial Neural Networks (O'Reilly Media, Inc., 2017) (cit. on pp. 35–38).

56. *Glandular tissue* https://www.histology.leeds.ac.uk/glandular/ (2022) (cit. on p. 2).

57. Gomperts, B. N. *et al.* Enriching the Molecular Definition of the Airway "Field of Cancerization": Establishing New Paradigms for the Patient at Risk for Lung Cancer. *Cancer Prevention Research* **6,** 4–7 (2013) (cit. on p. 6).

58. González-Moles, M. A. *et al.* The cancer stem cell hypothesis applied to oral carcinoma. *Oral Oncology* **49,** 738–746 (2013) (cit. on p. 5).

59. Gray, L. A mathematician looks at Wolfram's new kind of science. *Notices of the American Mathematical Society* **50,** 200–211 (2003) (cit. on p. 44).

60. Greaves, L. C. *et al.* Mitochondrial DNA mutations are established in human colonic stem cells, and mutated clones expand by crypt fission. *Proceedings of the National Academy of Sciences* **103,** 714–719 (2006) (cit. on p. 3).

61. Greaves, M. F. *et al.* Leukemia in twins: lessons in natural history. *Blood* **102,** 2321–2333 (2003) (cit. on pp. 2, 6).

62. Grewal, P. & Viswanathen, V. A. Liver cancer and alcohol. *Clin Liver Dis* **16,** 839–850 (2012) (cit. on p. 101).

63. Gutierrez–Gonzalez, L. *et al.* The Clonal Origins of Dysplasia From Intestinal Metaplasia in the Human Stomach. *Gastroenterology* **140,** 1251–1260 (2011) (cit. on p. 6).

64. Haaland, C. M. *et al.* Differential gene expression in tumor adjacent histologically normal prostatic tissue indicates field cancerization. *International Journal of Oncology* **35,** 537–546 (2009) (cit. on p. 6).

65. Hafner, C. *et al.* Clonality of multifocal urothelial carcinomas: 10 years of molecular genetic studies. *International Journal of Cancer* **101,** 1–6 (2002) (cit. on p. 6).

66. Hawthorn, L., Lan, L. & Mojica, W. Evidence for field effect cancerization in colorectal cancer. *Genomics* **103,** 211–221 (2014) (cit. on p. 6).

67. *Head and neck cancer - risk factors and prevention* https://www.cancer.net/cancer-types/head-and-neck-cancer/risk-factors-and-prevention (2021) (cit. on pp. 66, 67, 102).

68. *Head and neck squamous cell carcinoma* https://medlineplus.gov/genetics/condition/head-and-neck-squamous-cell-carcinoma/ (2021) (cit. on pp. 65, 67).

69. Herreros-Villanueva, M. *et al.* SOX2 promotes dedifferentiation and imparts stem cell-like features to pancreatic cancer cells. *Oncogenesis* **2,** 1–12 (2013) (cit. on p. 5).

70. Hittelman, W. N. Genetic instability in epithelial tissues at risk for cancer. *Annals of the New York Academy of Sciences* **952,** 1–12 (2001) (cit. on p. 6).

71. Hu, B. *et al.* Multifocal epithelial tumors and field cancerization from loss of mesenchymal CSL signaling. *Cell* **149,** 1207–1220 (2012) (cit. on p. 6).

72. Kadara, H. & Wistuba, I. I. Field Cancerization in Non–Small Cell Lung Cancer: Implications in Disease Pathogenesis. *Proceedings of the American Thoracic Society* **9,** 38–42 (2012) (cit. on p. 6).

73. Kadara, H. *et al.* Transcriptomic architecture of the adjacent airway field cancerization in non–small cell lung cancer. *Journal of the National Cancer Institute* **106,** 1–9 (2014) (cit. on p. 6).

74. Kamiyama, H. *et al.* DNA demethylation in normal colon tissue predicts predisposition to multiple cancers. *Oncogene* **31,** 5029–5037 (2012) (cit. on p. 6).

75. Kammori, M. *et al.* Squamous cell carcinomas of the esophagus arise from a telomere-shortened epithelial field. *International Journal of Molecular Medicine* **20,** 793–799 (2007) (cit. on p. 6).

76. Kang, G. H. *et al.* Genetic evidence for the multicentric origin of synchronous multiple gastric carcinoma. *Laboratory Investigation; a journal of technical methods and pathology* **76,** 407–417 (1997) (cit. on p. 6).

77. Kanjilal, S. *et al.* p53 mutations in nonmelanoma skin cancer of the head and neck: molecular evidence for field cancerization. *Cancer Research* **55,** 3604–3609 (1995) (cit. on p. 6).

78. Kaz, A. M. *et al.* Patterns of DNA methylation in the normal colon vary by anatomical location, gender, and age. *Epigenetics* **9,** 492–502 (2014) (cit. on p. 6).

79. Klein, A. M. *et al.* Stochastic fate of p53-mutant epidermal progenitor cells is tilted toward proliferation by UV B during preneoplasia. *Proceedings of the National Academy of Sciences* **107,** 270–275 (2010) (cit. on p. 3).

80. Koizumi, K. *et al.* Array-based identification of common DNA methylation alterations in ulcerative colitis. *International Journal of Oncology* **40,** 983–994 (2012) (cit. on p. 6).

81. Kumar, S. M. *et al.* Acquired cancer stem cell phenotypes through Oct4-mediated dedifferentiation. *Oncogene* **31,** 4898–4911 (2012) (cit. on p. 5).

82. Lee, Y.-C. *et al.* Revisit of field cancerization in squamous cell carcinoma of upper aerodigestive tract: better risk assessment with epigenetic markers. *Cancer Prevention Research* **4,** 1982–1992 (2011) (cit. on p. 6).

83. Leedham, S. J. *et al.* Clonality, founder mutations, and field cancerization in human ulcerative colitis–associated neoplasia. *Gastroenterology* **136,** 542–550 (2009) (cit. on p. 6).

84. Li, X. *et al.* Temporal and spatial evolution of somatic chromosomal alterations: a case-cohort study of Barrett's esophagus. *Cancer Prevention Research* **7,** 114–127 (2014) (cit. on p. 3).

85. Liégeois, F., Albert, A. & Limme, M. Comparison between tongue volume from magnetic resonance images and tongue area from profile cephalograms. *European Journal of Orthodontics* **32,** 381–386 (2009) (cit. on p. 70).

86. Liggett, T. M. *Stochastic interacting systems: contact, voter and exclusion processes* (Springer Science & Business Media, 1999) (cit. on p. 13).

87. Lin, M.-W. *et al.* Clinicopathology and genetic profile of synchronous multiple small adenocarcinomas: implication for surgical treatment of an uncommon lung malignancy. *Annals of Surgical Oncology* **21,** 2555–2562 (2014) (cit. on p. 6).

88. LoConte, N. K. *et al.* Alcohol and cancer: a statement of the American Society of Clinical Oncology. *Journal of Clinical Oncology* **36,** 83–93 (2018) (cit. on p. 69).

89. Maley, C. C. *et al.* Selectively advantageous mutations and hitchhikers in neoplasms: p16 lesions are selected in Barrett's esophagus. *Cancer Research* **64,** 3414–3427 (2004) (cit. on p. 6).

90. Maley, C. C. *et al.* The combination of genetic instability and clonal expansion predicts progression to esophageal adenocarcinoma. *Cancer Research* **64,** 7629–7633 (2004) (cit. on p. 6).

91. Martens, E. A. *et al.* Spatial structure increases the waiting time for cancer. *New Journal of Physics* **13,** 1–22 (2011) (cit. on p. 3).

92. Matsuda, Y. *et al.* Hypomethylation of Alu repetitive elements in esophageal mucosa, and its potential contribution to the epigenetic field for cancerization. *Cancer Causes & Control* **23,** 865–873 (2012) (cit. on p. 6).

93. McCaughan, F. *et al.* Genomic evidence of pre-invasive clonal expansion, dispersal and progression in bronchial dysplasia. *The Journal of Pathology* **224,** 153–159 (2011) (cit. on p. 6).

94. McCulloch, W. S. & Pitts, W. A logical calculus of the ideas immanent in nervous activity. *The Bulletin of mathematical biophysics* **5,** 115–133 (1943) (cit. on pp. 35, 36).

95. McDonald, S. A. C. *et al.* Mechanisms of field cancerization in the human stomach: the expansion and spread of mutated gastric stem cells. *Gastroenterology* **134,** 500–510 (2008) (cit. on pp. 3, 6).

96. Milicic, A. *et al.* Ectopic expression of P-cadherin correlates with promoter hypomethylation early in colorectal carcinogenesis and enhanced intestinal crypt fission in vivo. *Cancer Research* **68,** 7760–7768 (2008) (cit. on p. 6).

97. Minsky, M. L. & Papert, S. A. *Perceptrons: expanded edition* 1988 (cit. on p. 37).

98. Moon, J.-H. *et al.* Nanog-induced dedifferentiation of p53-deficient mouse astrocytes into brain cancer stem-like cells. *Biochemical and Biophysical Research Communications* **412,** 175–181 (2011) (cit. on p. 5).

99. Mori, H. *et al.* Chromosome translocations and covert leukemic clones are generated during normal fetal development. *Proceedings of the National Academy of Sciences* **99,** 8242–8247 (2002) (cit. on p. 6).

100. Nachman, M. W. & Crowell, S. L. Estimate of the Mutation Rate per Nucleotide in Humans. *Genetics* **156,** 297–304 (2000) (cit. on p. 71).

101. Nakachi, I. *et al.* Application of SNP microarrays to the genome-wide analysis of chromosomal instability in premalignant airway lesions. *Cancer Prevention Research* **7,** 255–265 (2014) (cit. on p. 6).

102. Narayana, A. *et al.* Second primary tumors in laryngeal cancer: results of long-term follow-up. *International Journal of Radiation Oncology •Biology •Physics* **42,** 557–562 (1998) (cit. on p. 6).

103. Nicholson, A. M. *et al.* Barrett's metaplasia glands are clonal, contain multiple stem cells and share a common squamous progenitor. *Gut* **61,** 1380–1389 (2012) (cit. on p. 3).

104. Nonn, L., Ananthanarayanan, V. & Gann, P. H. Evidence for field cancerization of the prostate. *The Prostate* **69,** 1470–1479 (2009) (cit. on p. 6).

105. Nowell, P. C. The clonal evolution of tumor cell populations. *Science* **194,** 23–28 (1976) (cit. on p. 8).

106. Oka, D. *et al.* The presence of aberrant DNA methylation in noncancerous esophageal mucosae in association with smoking history: a target for risk diagnosis and prevention of esophageal cancers. *Cancer: Interdisciplinary International Journal of the American Cancer Society* **115,** 3412–3426 (2009) (cit. on p. 6).

107. Paterson, C., Clevers, H. & Bozic, I. Mathematical model of colorectal cancer initiation. *Proceedings of the National Academy of Sciences* **117,** 20681–20688 (2020) (cit. on p. 11).

108. Pentenero, M. *et al.* Field effect in oral precancer as assessed by DNA flow cytometry and array-CGH. *Journal of Oral Pathology & Medicine* **41,** 119–123 (2012) (cit. on p. 6).

109. Petrick, J. L. *et al.* Tobacco, alcohol use and risk of hepatocellular carcinoma and intrahepatic cholangiocarcinoma: The Liver Cancer Pooling Project. *Br J Cancer* **118,** 1005–1012 (2018) (cit. on p. 101).

110. Pipinikas, C. P. *et al.* Cell migration leads to spatially distinct but clonally related airway cancer precursors. *Thorax* **69,** 548–557 (2014) (cit. on p. 6).

111. Polyanin, A. D. & Nazaikinskii, V. E. in *Handbook of Linear Partial Differential Equations for Engineers and Scientists* 2nd ed., 1199–1231 (Chapman and Hall/CRC, 2016) (cit. on p. 26).

112. Prevo, L. J. *et al.* p53-mutant clones and field effects in Barrett's esophagus. *Cancer Research* **59,** 4784–4787 (1999) (cit. on p. 8).

113. *ProServe Liquor Training* `https://proserve.aglc.ca` (2022) (cit. on p. 69).

114. Radunskaya, A., Kim, R. & Woods II, T. Mathematical modeling of tumor immune interactions: a closer look at the role of a PD-L1 inhibitor in cancer immunotherapy. *Spora: A Journal of Biomathematics* **4,** 25–41 (2018) (cit. on p. 11).

115. Reddy, K. A. & Doraiswamy, L. K. Estimating liquid diffusivity. *Industrial & Engineering Chemistry Fundamentals* **6,** 77–79 (1967) (cit. on p. 68).

116. *Risk Factors of Mouth and Throat Cancer* `https://www.merckmanuals.com/home/ear,-nose,-and-throat-disorders/mouth,-nose,-and-throat-cancers/mouth-and-throat-cancer` (2021) (cit. on pp. 66, 67, 70).

117. Rivenbark, A. G. & Coleman, W. B. Field cancerization in mammary carcinogenesis—Implications for prevention and treatment of breast cancer. *Experimental and Molecular Pathology* **93,** 391–398 (2012) (cit. on p. 6).

118. Roesch-Ely, M. *et al.* Proteomic analysis of field cancerization in pharynx and oesophagus: a prospective pilot study. *The Journal of Pathology* **221,** 462–470 (2010) (cit. on p. 6).

119. Rumelhart, D. E., Hinton, G. E. & Williams, R. J. in *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Vol. 1: Foundations* 318–362 (MIT Press, Cambridge, MA, USA, 1986) (cit. on p. 38).

120. Ryser, M. D. *et al.* Quantifying the dynamics of field cancerization in tobacco-related head and neck cancer: a multiscale modeling approach. *Cancer Research* **76,** 7078–7088 (2016) (cit. on pp. 11, 17–19).

121. Sanner, T. & Grimsrud, T. K. Nicotine: Carcinogenicity and Effects on Response to Cancer Treatment – A Review. *Frontiers in Oncology* **5,** 1–10 (2015) (cit. on p. ii).

122. Sara N. Gentry, S. N. & Jackson, T. L. A Mathematical Model of Cancer Stem Cell Driven Tumor Initiation: Implications of Niche Size and Loss of Homeostatic Regulatory Mechanisms. *PLOS ONE* **8,** 1–15 (2013) (cit. on p. 11).

123. Schöler, H. R. in *Humanbiotechnology as Social Challenge: An Interdisciplinary Introduction to Bioethics* (eds Schipanski, D., Knoepffler, N. & Sorgner, S. L.) 1st ed., 27–55 (Ashgate Publishing, 2007) (cit. on pp. 4, 5).

124. Schönfisch, B. & de Roos, A. Synchronous and asynchronous updating in cellular automata. *BioSystems* **51,** 123–143 (1999) (cit. on p. 76).

125. Shain, A. H. *et al.* The genetic evolution of melanoma from precursor lesions. *New England Journal of Medicine* **373,** 1926–1936 (2015) (cit. on p. 6).

126. Shaw, R. J. *et al.* Molecular staging of surgical margins in oral squamous cell carcinoma using promoter methylation of $p16^{INK4A}$, cytoglobin, E-cadherin, and TMEFF2. *Annals of Surgical Oncology* **20,** 2796–2802 (2013) (cit. on p. 6).

127. Shen, L. *et al.* MGMT promoter methylation and field defect in sporadic colorectal cancer. *Journal of the National Cancer Institute* **97,** 1330–1338 (2005) (cit. on p. 6).

128. Sieburg, H. B., Rezner, B. D. & Muller-Sieburg, C. E. Predicting clonal self-renewal and extinction of hematopoietic stem cells. *Proceedings of the National Academy of Sciences* **108,** 4370–4375 (2011) (cit. on p. 76).

129. Simon, R. *et al.* Cytogenetic analysis of multifocal bladder cancer supports a monoclonal origin and intraepithelial spread of tumor cells. *Cancer Research* **61,** 355–362 (2001) (cit. on p. 8).

130. Simple, M. *et al.* Cancer stem cells and field cancerization of oral squamous cell carcinoma. *Oral Oncology* **51,** 643–651 (2015) (cit. on pp. 8, 10).

131. *Simple Squamous Epithelium* `https://www.biologyonline.com/dictionary/simple-squamous-epithelium` (2022) (cit. on pp. 2, 3).

132. *Size of epithelia cell* https://bionumbers.hms.harvard.edu/bionumber.aspx?id=108917&ver=0&trm=cell+volume&org= (2022) (cit. on p. 70).

133. Slaughter, D. P., Southwick, H. W. & Smejkal, W. "Field cancerization" in oral stratified squamous epithelium. Clinical implications of multicentric origin. *Cancer* **6,** 963–968 (1953) (cit. on pp. 5–8).

134. Sozzi, G. *et al.* Genetic evidence for an independent origin of multiple preneoplastic and neoplastic lung lesions. *Cancer Research* **55,** 135–140 (1995) (cit. on p. 6).

135. Stachler, M. D. *et al.* Paired exome analysis of Barrett's esophagus and adenocarcinoma. *Nature Genetics* **47,** 1047–1055 (2015) (cit. on p. 3).

136. *Standard Alcoholic Drink* https://pubs.niaaa.nih.gov/publications/Practitioner/pocketguide/pocket_guide2.htm (cit. on p. 69).

137. Steiling, K. *et al.* The field of tissue injury in the lung and airway. *Cancer Prevention Research* **1,** 396–403 (2008) (cit. on p. 6).

138. Stern, R. S. *et al.* p53 mutation in nonmelanoma skin cancers occurring in psoralen ultraviolet a-treated patients: evidence for heterogeneity and field cancerization. *Journal of Investigative Dermatology* **119,** 522–526 (2002) (cit. on p. 6).

139. Stratton, M. R., Campbell, P. J. & Futreal, P. A. The cancer genome. *Nature* **458,** 719–724 (2009) (cit. on p. 2).

140. Szeimies, R. M. *et al.* Clinical, histopathological and immunohistochemical assessment of human skin field cancerization before and after photodynamic therapy. *British Journal of Dermatology* **167,** 150–159 (2012) (cit. on p. 6).

141. Tabor, M. P. *et al.* Persistence of genetically altered fields in head and neck cancer patients: biological and clinical implications. *Clinical Cancer Research* **7,** 1523–1532 (2001) (cit. on p. 6).

142. Tabor, M. P. *et al.* Discordance of genetic alterations between primary head and neck tumors and corresponding metastases associated with mutational status of the TP53 gene. *Genes, Chromosomes and Cancer* **33,** 168–177 (2002) (cit. on p. 8).

143. Tabor, M. P. *et al.* Multiple head and neck tumors frequently originate from a single preneoplastic lesion. *The American Journal of Pathology* **161,** 1051–1060 (2002) (cit. on p. 8).

144. Takeshima, H. *et al.* Frequent involvement of chromatin remodeler alterations in gastric field cancerization. *Cancer Letters* **357,** 328–338 (2015) (cit. on p. 6).

145. Teixeira, V. H. *et al.* Stochastic homeostasis in human airway epithelium is achieved by neutral competition of basal cell progenitors. *Elife* **2,** 1–18 (2013) (cit. on p. 3).

146. Trujillo, K. A. *et al.* Breast field cancerization: isolation and comparison of telomerase-expressing cells in tumor and tumor adjacent, histologically normal breast tissue. *Molecular Cancer Research* **9,** 1209–1221 (2011) (cit. on p. 6).

147. Trujillo, K. A. *et al.* Markers of field cancerization: proposed clinical applications in prostate biopsies. *Prostate Cancer* **2012,** 1–12 (2012) (cit. on p. 6).

148. Ushijima, T. & Hattori, N. Molecular pathways: Involvement of Helicobacter pylori–triggered inflammation in the formation of an epigenetic field defect, and its usefulness as cancer risk and exposure markers. *Clinical Cancer Research* **18,** 923–929 (2012) (cit. on p. 6).

149. van Dekken, H. *et al.* Clonal analysis of a case of multifocal oesophageal (Barrett's) adenocarcinoma by comparative genomic hybridization. *The Journal of Pathology* **188,** 263–266 (1999) (cit. on p. 6).

150. van Dekken, H. *et al.* Genomic analysis of a case of multifocal adenocarcinoma in ulcerative colitis. *Virchows Archiv* **449,** 716–721 (2006) (cit. on p. 6).

151. van der Vorst, S. *et al.* Assessment of p53 functional activity in tumor cells and histologically normal mucosa from patients with head and neck squamous cell carcinoma. *Head & Neck* **34,** 1542–1550 (2012) (cit. on p. 6).

152. van Houten, V. M. M. *et al.* Mutated p53 as a molecular marker for the diagnosis of head and neck cancer. *The Journal of Pathology: A Journal of the Pathological Society of Great Britain and Ireland* **198,** 476–486 (2002) (cit. on pp. 6, 8).

153. van Oijen, M. G. C. T. & Slootweg, P. J. Oral field cancerization: carcinogen-induced independent events or micrometastatic deposits? *Cancer Epidemiology and Prevention Biomarkers* **9,** 249–256 (2000) (cit. on p. 7).

154. Vatve, M. *et al.* Management of field change in actinic keratosis. *British Journal of Dermatology* **157,** 21–24 (2007) (cit. on p. 6).

155. Vriesema, J. L. J. *et al.* Superficial and metachronous invasive bladder carcinomas are clonally related. *International Journal of Cancer* **93,** 699–702 (2001) (cit. on p. 6).

156. Wang, Y. *et al.* Comparison of the clonality of urothelial carcinoma developing in the upper urinary tract and those developing in the bladder. *SpringerPlus* **2,** 1–6 (2013) (cit. on p. 6).

157. Weaver, J. M. J. *et al.* Ordering of mutations in preinvasive disease stages of esophageal carcinogenesis. *Nature Genetics* **46,** 837–843 (2014) (cit. on p. 2).

158. Weichert, W. & Warth, A. Early lung cancer with lepidic pattern: adenocarcinoma in situ, minimally invasive adenocarcinoma, and lepidic predominant adenocarcinoma. *Current Opinion in Pulmonary Medicine* **20,** 309–316 (2014) (cit. on p. 6).

159. Weinberg, R. A. in. Chap. 11: Multi-step Tumorigenesis (Garland science, 2013) (cit. on p. 102).

160. Williams, T. & Bjerknes, R. Stochastic model for abnormal clone spread through epithelial basal layer. *Nature* **236,** 19–21 (1972) (cit. on p. 13).

161. Wirtschafter, E. *et al.* Diffuse idiopathic pulmonary neuroendocrine cell hyperplasia of the lung (DIPNECH): current best evidence. *Lung* **193,** 659–667 (2015) (cit. on p. 6).

162. Yakoub, D. *et al.* Metabolic profiling detects field effects in nondysplastic tissue from esophageal cancer patients. *Cancer Research* **70,** 9129–9136 (2010) (cit. on p. 6).

163. Yamanoi, K. *et al.* Epigenetic clustering of gastric carcinomas based on DNA methylation profiles at the precancerous stage: its correlation with tumor aggressiveness and patient outcome. *Carcinogenesis* **36,** 509–520 (2015) (cit. on p. 6).

164. Zaky, A. H. *et al.* Clinicopathologic implications of genetic instability in intestinal-type gastric cancer and intestinal metaplasia as a precancerous lesion: proof of field cancerization in the stomach. *American Journal of Clinical Pathology* **129,** 613–621 (2008) (cit. on p. 6).

165. Zhou, Z.-T. & Jiang, W.-W. Cancer stem cell model in oral squamous cell carcinoma. *Current Stem Cell Research & Therapy* **3,** 17–20 (2008) (cit. on p. 5).