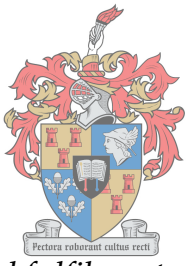
Investigating tumour micro environment dynamics based on cytokine-mediated innate-adaptive immunity

by

Innocenter Moraa Amima



Thesis presented in partial fulfilment of the requirements for the degree of Master of Science in Mathematics in the Faculty of Science at Stellenbosch University

Supervisor: Dr. Gaston Kuzamunu Mazandu

March 2018

Declaration

By submitting this thesis electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

Date: March 2018 .....

Abstract

Investigating tumour micro environment dynamics based on cytokine-mediated innate-adaptive immunity

Innocenter Moraa Amima

Thesis: MSc. (Mathematics)

March 2018

Cancer is a leading cause of death worldwide, yet much is still unknown about its mechanism of establishment, recurrence cycle and destruction. It is known that the successive alterations that occur in a set of specific genes in the cell can trigger carcinogenesis which is the process of transforming a normal cell into a cancer cell. This process is usually done in the following three steps: initiation, proliferation and progression. Cancer stem cells are regulated by complex interactions with the components of the tumour micro-environment (TME) through networks of cytokines and growth factors. Thus, understanding the role of cytokines can be crucial in the fight against cancer in the context of improving diagnostic, prognostic and therapeutic strategies. Several studies have investigated tumour-immune cell dynamics. However, some of these studies are mostly limited to cells that directly kill cancer cells, such as natural killer (NK) and cytotoxic T lymphocytes (CTLs), and they do not explicitly integrate cytokines in the cell dynamics. Furthermore, none of these studies has combined cellular-level mathematical models with molecular-level/signalling pathway analysis, to predict biological processes and enriched pathways associated in cancer disease.

In this study, a new non-linear mathematical model integrating cytokines in the activa-

tion of innate and adaptive immunity was developed to predict the role of cytokines in tumour and immune cell dynamics. Our work complements the role Th2 and Th17 cells play in inhibiting the proliferation of M1 macrophages, CTLs, NK and Th1 cells. Numerical analysis of the model suggested that, lack of TGF-β inhibition effect resulted in tumour clearance, however, the immune cells grew without bound and exceeded the carrying capacity of immune cells. TGF-β is responsible for promoting tumour progression, tumour proliferation and limiting the effectiveness of type 1 immune response. In addition, we established the necessary conditions for tumour clearance by varying parameter values. For an immune (effector) cell to be activated from a resting state, its gene expression must be altered. We used datasets from The Cancer Genome Atlas (TCGA) and an expression level-based model to identify genes specific to cancer patients contributing to the regulation of cytokines in the context of the breast cancer disease. We predicted differentially expressed genes (DEGs) associated with breast cancer disease using a permutation-based significance analysis of micro-array (SAM) approach. Using a selected list of the DEGs, we determined significant pathways and enriched biological processes associated with breast cancer disease. Some of the identified significant biological pathways and processes, happened to be associated with cell differentiation or cell division and the predicted over-expressed genes in tumour samples may contribute to the proliferation of cancer. These genes, pathways and biological processes can be further assessed to check for their suitability as targets for breast cancer disease.

Keywords: Tumour cells, immune cells, cytokines, differentially expressed genes.

Opsomming

Ondersoek van die mikro-dinamika van die tumor mikro-omgewing, gebaseer op sitokien-gemedieerde aangebore-adaptiewe immuniteit

(" Investigating tumour micro environment dynamics based on cytokine-mediated innate-adaptive immunity ")

Innocenter Moraa Amima

Tesis: MSc. (Wiskunde)

Maart 2018

Kanker is wereldwyd een van van die grootste oorsake van dood en steeds is daar nog baie onbekend oor die meganisme van onstaan, herverskynings siklus en vernietiging. Dit is bekend dat die opeenvolgende veranderinge wat in 'n stel spesifieke gene in die sel voorkom, kan karsinogenese veroorsaak, die proses waardeur 'n normale sel in 'n kankersel omskep word en gewoonlik in die volgende drie stappe gedoen word: inisiasie, proliferasie en progressie. Kankerstemselle word gereguleer deur komplekse interaksies met die komponente van die tumor mikro-omgewing deur netwerke van sitokiene en groeifaktore. Die tumor mikro-omgewing (TME) bestaan uit verskillende seltipes, insluitende aangebore en aanpasbare immuun selle, en sitokiene wat die aktivering of inhibisie van die immuun selle en proliferasie van tumorselle reguleer. Om die rol van sitokiene te verstaan, kan 'n belangrike rol speel in die stryd teen kanker. Verskeie navorsingsprojekte het die dinamika van hierdie sitokiene ge-ondersoek in die konteks van kanker om die evolusie van kanker te verstaan vir die verbetering van diagnostiese, prognostiese en terapeutiese strategieë. Hierdie studies is egter meestal beperk tot selle wat kankerselle direk doodmaak, soos natuurlike vernietigers (NV) en sitotoksiese T

limfosiete (CTLs of CD8 + T) selle, en hulle eksplisiet nie sitokien-gemedieerde aangebore aanpassingsimmuniteit in 'n tumor dinamika. Verder het geen van hierdie studies selektiewe wiskundige modelle gekombineer met molekulêre-vlak / seinweganalise, biologiese prosesse en verrykde weë wat verband hou met kankersiekte, voorspel nie.

In hierdie studie is 'n nuwe nie-lineêre wiskundige model wat sitokien-gemedieerde aangebore aanpassingsimmuniteit integreer ontwikkel om die sitokien-gemedieerde tumor en immuun seldinamika te voorspel. Numeriese analise van die model het voorgestel dat 'n gebrek aan TGF-β remmings-effek tot tumorruiming gelei het, maar die immuunselle het sonder gebind gegroei en die dravermoë oorskry. Daarbenewens het ons die nodige toestande vir tumor verwydering deur verskillende parameterwaardes vasgestel. Vir 'n immuun (effektor) sel wat vanuit 'n rustende toestand geaktiveer moet word, moet sy geen uitdrukking verander wor. Ons gebruik datastelle van The Cancer Genome Atlas (TCGA) en 'n uitdrukkingsvlakgebaseerde model om gene wat spesifiek vir kankerpasiënte is, te identifiseer wat bydra tot die regulering van sitokiene in die konteks van die borskanker siekte. Ons het differensieel uitgedrukte gene (DUG) voorspel wat verband hou met borskanker siekte deur gebruik te maak van 'n permutasiegebaseerde betekenisanalise van mikro-skikking benadering. Deur gebruik te maak van 'n geselekteerde lys van die DUG, het ons belangrike paaie en verrykde biologiese prosesse geassosieer met borskanker bepaal. Sommige van die geÃŕdentifiseerde beduidende biologiese weë en prosesse, sowel as oor-uitgedruk gene, wat waargeneem word geassosieer met sel differensiasie of seldeling, kan bydra tot die proliferasie van kankerselle. Hierdie weë en oor-uitgedruk gene kan verder geassesseer word om na te gaan of hulle geskik is as teikens vir die borskanker siekte.

Keywords: Tumor selle, immuun selle, sitokiene, differensieel uitgedruk gene.

Acknowledgements

I would like to express my sincere gratitude and appreciation to my supervisor Dr. Gaston Mazandu whose guidance, continuous encouragements and immense knowledge made my thesis work possible. Secondly, I would like to thank the African Institute for Mathematical Sciences (AIMS) and the International Development Research Centre (IDRC) for the financial support. I would like to thank the AIMS research centre for the opportunity to pursue my studies, AIMS staff for their support and for providing a conducive environment for pursuing research which made my stay memorable. Last but not least, I am greatly indebted to my parents Mr. and Mrs. Amima, siblings and friends for their support and encouragement throughout my studies.

Dedications

To my late baby niece Babra Nyachwaya

Contents

Declaration i

Abstract ii

Opsomming iv

List of Figures x

List of Tables xii

1 Introduction 1
- 1.1 Evolution of cancer disease . . . 2
- 1.2 The immune response to cancer disease . . . 4
- 1.3 Exploring cytokine world during carcinogenesis . . . 5
- 1.4 Overview of RNA-Seq gene expression data analysis . . . 7
- 1.5 Motivation . . . 8
- 1.6 Objectives and significance of this study . . . 8
- 1.7 Outline . . . 9

2 Literature Review 10
- 2.1 Brief review of the biological background . . . 10
- 2.2 Brief review of tumour-immune mathematical models . . . 13

3 Modelling cytokine-mediated tumour-immune cell dynamics 16
- 3.1 Introduction . . . 16
- 3.2 Mathematical model description . . . 16
  - 3.2.1 Model assumptions . . . 18
  - 3.2.2 Description of model equations . . . 19
- 3.3 Model analysis . . . 26

3.3.1 Steady state solutions 29
3.4 Numerical analysis 34
3.4.1 Parameter estimation 34
3.4.2 Numerical solutions 35
3.4.3 Sensitivity analysis 41

4 Expression level based cytokine-regulator gene prediction 44
4.1 Introduction 44
4.2 Data retrieval and pre-processing 44
4.3 Statistical analysis of breast cancer gene expression data 47
4.3.1 The SAM approach 48
4.3.2 Identification of differentially expressed genes 50
4.4 Gene set enrichment and pathway analysis 55
4.4.1 Enriched biological processes associated with breast cancer 57
4.4.2 Significant pathways associated with breast cancer 59

5 Discussion and conclusion 61

List of references 63

List of Figures

1.1 Worldwide estimated breast cancer incidence and mortality rate in 2012. The estimates are age-standardized per 100,000. Adapted from http://globocan.iarc.fr/ 2
1.2 Activated macrophages engulf tumour cells, secrete cytokines to send signals to activate dendritic cells and other immune cells. The outcome can favour either a tumour-promoting response or an anti-tumour immune response. Diagram adapted from De Visser et al. [13] 5

2.1 T cells are activated and can different into various subsets due to the presence of tumour cells, transcription factors and cytokines. 12

3.1 Schematic diagram to represent cytokine-mediated innate-adaptive immunity in the presence of tumour cells. Resting macrophages recognize tumour antigens. Activated macrophages have the ability to engulf tumour cells, secrete cytokines to send signals to activate CTLs, naive T helper (Th0) cells and resting NK (rNK) cells. The outcome of tumour-immune cell dynamics can favour a type 2 immune response at the expense of type 1 immune response or vice versa. 17
3.2 The best fit curve used to estimate unobserved parameter values from data published by Benzekry et al. [76]. 37
3.3 Panel (a-b) shows the positive effect type 2 immune signals (M2 macrophages, Th2 and Th17 cells) has on tumour cells. The depletion of type 1 immune signals (M1 macrophages, CTLS, NK, Th0, Th1 cells) on day 10 resulted in the proliferation and progression of tumour cells. The population of precancerous cells in Panel (e) are proportional to the population of type 1 immune signals. Panel (f) shows the cell dynamics in panel (a-e) on a logarithmic $x-$ and $y-$ axis scale. 38

3.4 Anti-inflammatory cytokines, such as IL-4 (I_4), IL-6 (I_6), IL-10 (I_{10}), IL-23 (I_{23})and TGF-β (I_β) dominate from day 15 after the pro-inflammatory cytokines, such as TNF-α (I_α) and IFN-γ (I_γ) were depleted. 39
3.5 Panel (a) indicates that the removal of IL-23 has no effect on tumour cell population. However, the removal of TGF-β results in tumour clearance but, this caused an over-stimulation of immune cells. 40
3.6 Global sensitivity analysis involved varying some the unobserved parameters π_1, π_2, μ_I, μ_c, θ_I by 40% from the baseline values in Table 3.3. These parameters had a negligible effect on tumour cell population. 41
3.7 Local sensitivity analysis of the intrinsic growth rate α_c by decreasing it by 15% (Panel a) and 25% (Panel b) from its baseline value of 0.502. 42

4.1 RNA-Seq experiments normally result in a large data set containing a long list of genes. The downloaded gene expression data set can be represented as a **G** matrix with p genes and n samples and their respective expression levels X_{ij} and Y_{ij} representing tumour and tumour-free samples, respectively. . . . 46
4.2 Panel (a) and (b) represents standard deviation against expression level mean value plots before and after Miller's test, respectively. The sharp curve (red line) in Panel (a) indicates the presence of genes with low counts. 48
4.3 The SAM approach was used to detect 4,159 differentially expressed genes (DEGs). The plot shows non-DEGs (in black circles) and DEGs (in pink circles). 50
4.4 Slight differences in the 4,159 differentially expressed genes across samples that are with tumour (WT) and tumour-free(TF). 51
4.5 The gene set in Table 4.1 and Table 4.2 had over-expressed genes in tumour samples (WTOE) that were under-expressed in tumour-free samples (TFUE) and under-expressed genes in tumour samples (WTUE) that were over-expressed in tumour-free samples (WTUE). Over-expressed genes in tumour samples have a higher expression level mean compared to the other categories. 55
4.6 The heat map indicates the presence of two distinct clusters across the tumour and tumour-free samples in the selected gene sets. 56
4.7 The biological processes in red represent the over-expressed genes in tumour samples and the ones in blue represent under-expressed genes in tumour samples. 58

List of Tables

3.1 Convention used in naming the parameters in the model 19
3.2 Key cytokines involved in tumour-immune interactions. 20
3.3 Parameter values 36

4.1 Top 20 over-expressed genes in tumour (WT) samples that were under-expressed in tumour-free (TF) samples. 53
4.2 Top 20 under-expressed genes in tumour (WT) samples that were over-expressed in tumour-free (TF) samples. 54
4.3 Enriched biological processes over-represented in tumour samples. 57
4.4 KEGG pathways over-represented in tumour samples. 59
4.5 KEGG pathways over-represented in tumour-free samples. 60

Chapter 1

Introduction

Cancer is a leading cause of morbidity and mortality both in the developed and developing world [1]. Approximately 1.67 million new cases were diagnosed and 0.522 million cancer-related deaths recorded during the year 2012 compared to 2011 [1, 2]. This makes cancer research an interesting public health topic and a lot of effort towards research is being put to understand the mechanisms of cancer evolution and progression. According to the world health organisation (WHO) report, breast cancer is the most (25.2%) common incident site amongst women, making it the leading cause of death among women [1]. Figure 1.1 gives a summary of the estimated incidence and mortality rates of breast cancer in 2012.

The incidence rates in the developed world is greater compared to the developing world and the mortality rates are comparable as shown in Figure 1.1. The increase in breast cancer incidence in the developing world can be associated with the consequences of globalisation, increased urbanization and a transition in behaviour, in terms of adopting western lifestyles which promotes physical inactivity and bad dietary habits [3, 4]. Approximately 21% breast cancer deaths are attributed to risk factors, such as alcohol use (5%), physical inactivity (10%), overweight and obesity (9%) [5]. Although, avoiding exposure to these behavioural risk factors might reduce breast cancer incidence, it is not sufficient to fight breast cancer [3–5]. Therefore, healthcare guidelines for early detection methods, early prognosis and accurate selection of therapy remain the cornerstone of early cancer detection, control and improved overall survival rate [3, 6]. The relative survival rate of breast cancer patients depends on factors, such as the number of axillary lymph node involved, cancer size (diameter), presence of hormone receptors, treatment strategy, stage of the disease and geographical location of the patient [7]. Despite major

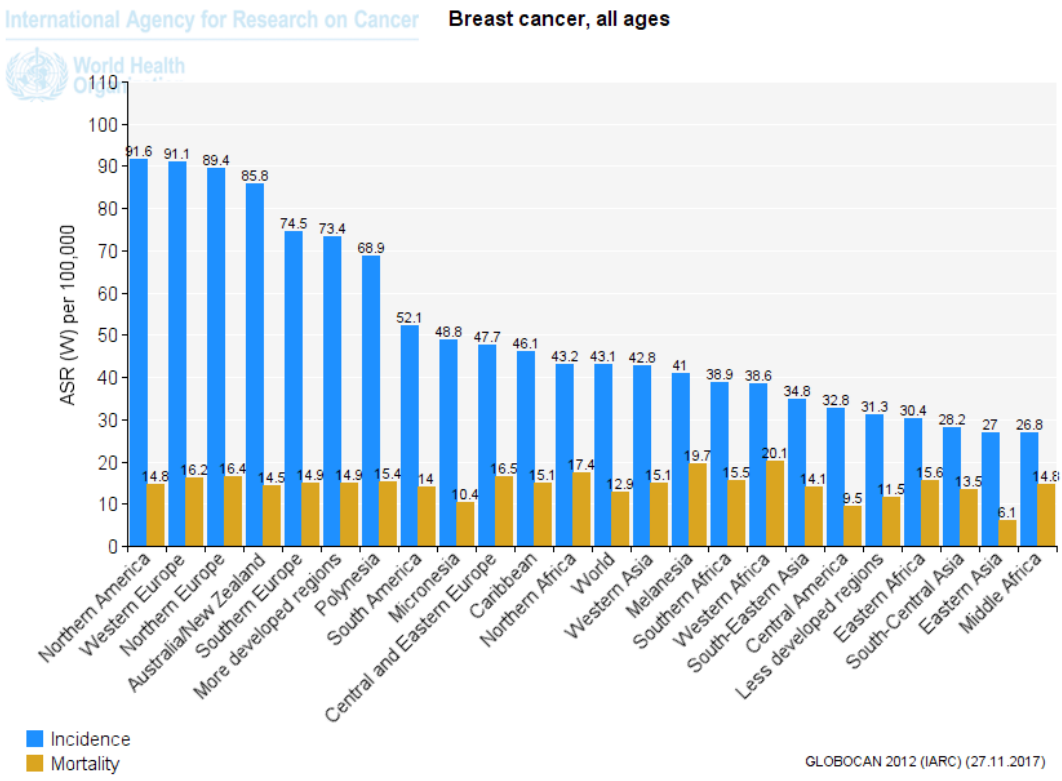


Figure 1.1: Worldwide estimated breast cancer incidence and mortality rate in 2012. The estimates are age-standardized per 100,000. Adapted from http://globocan.iarc.fr/

.

advances in treatment which usually involve both non-surgical and surgical removal of the breast (mastectomy) or the tumour and surrounding tissue (lumpectomy), breast cancer remains a clinical challenge affecting numerous patients both in the developed and developing world [1, 6].

In the next sections, we discuss about carcinogenesis(evolution of cancer), immune response to cancer disease, cytokines promoting carcinogenesis, RNA-Seq expression data analysis, motivation, objectives and significance of this study.

1.1 Evolution of cancer disease

Cancer is a complex tissue made up of different cell types that actively work together. In the 19th Century, Rudolf Virchow hypothesized that the origin of cancer was at sites of chronic inflammation [8]. Tumour micro-environment (TME) contains inflammatory cells which are a useful component of tumour progression. Occurrence of cancer is often

triggered by multiple changes in the genetic make up of normal cells, such as deletion, insertion, substitution, translocation, inversion, or gene amplifications triggered by either internal (genetic) or external (exposure to chemical) carcinogens [9, 10].

Early work on cancer growth focussed on investigating how 'normal' cells mutate into cancer cells [8, 11–13]. The onset of tumour formation happens in the absence of a vascular network. Tumour cells have the ability to secrete tumour angiogenic factors (TAF) that stimulate proliferation of endothelial cells to create their own blood vessels (angiogenesis) and promote its growth [9, 10, 14]. After developing a vascular network, tumour cells will have supply to nutrients and oxygen to support its development and proliferation. TME consist of various cell types such as fibroblasts and epithelial cell types, innate and adaptive immune cells, mesenchymal cell types and cells that form blood and lymphatic vasculature [12, 13]. Cancer cells use the following mechanisms to promote its growth and proliferation as highlighted by Hanahan and Weinberg [9], Bellomo and Preziosi [10]

1. Normal cells in a certain tissue that are programmed to die resist any anti-growth signals by deactivating p53 signals, turning rogue and growing uncontrollably [9].
2. Tumour cells produce angiogenic factors (TAF) that promote angiogenesis and its survival [9, 10, 14].
3. Tumour cells decrease immunogenecity by secreting anti-inflammatory cytokines, loosing major histocompatibility complex (MHC) molecule expressions, and converting anti-tumour immune cell signals to pro-tumour signals [15–18]
4. Tumour cells stimulate immune cells to secrete soluble molecules, such as cytokines, chemokines and other growth signals to promote its development and progression [9].
5. Tumour cells require higher metabolisms compared to normal cells therefore, they reprogram pathways regulating cell metabolism to sustain its rapid proliferation [9].
6. Aggressive tumour cells spread to distant sites by evading immune surveillance, and invading the adjacent local tissues [9].

In summary, cancer cells have acquired the ability to evade immune surveillance [10, 12] and in the next section, we explore how the immune system responds to cancer disease.

1.2 The immune response to cancer disease

The immune system has two main branches, the innate and adaptive immunity [12]. Innate immune cells include macrophages, dendritic cells (DCs), natural killer (NK) cells and adaptive immune cells include B and T cells [12]. B-cells recognize antigens while T cells recognize MHC molecules and APCs [12, 19]. The innate immunity is the first line of defence and the adaptive immunity takes time to respond as it adapts to defend the host's body against specific antigens[12, 20].

The immune system plays three keys roles in destroying tumours; firstly, it suppresses viral infections that may cause viral-induced tumours secondly, it eliminates pathogens quickly or worsens the inflammatory environment not to be conducive for tumour formation and thirdly, it recognizes and eliminates tumour-specific antigens expressed by tumour cells before they cause harm [21]. The presence of a tumour stimulates an immune response and the fate of the tumour is highly dependent on the response of host's immune system, initial number and type of tumour cells, and the surrounding environment in the host [20]. The interaction between cells in an immune response are bi-directional and often controlled by tissues, cytokines, chemokines and other soluble chemical factors [22, 23]. Figure 1.2 illustrates tumour progression in the presence of an immune response and cytokines.

Macrophages and DCs are the most potent antigen presenting cells (APCs) of the immune system [24]. Immature macrophage monocytes are released from the bone marrow into the blood stream, and tissues to undergo maturation into resident macrophages that have proteins on their surface to recognize and directly bind to the surface of tumour cells [12, 25]. Activated macrophages engulf tumour cells, produce battle cytokines that send danger signals to activate DCs and other immune cells by presenting an antigen attached to MHC class II molecules on its surface [12]. When such activated DCs leave the tumour site, they initiate and amplify an immune response [12, 26, 27]. A mature DC (mDC) can express multiple co-stimulatory molecules and cytokines that are important in priming effector T cell responses, activating resting NK (rNK) cells, and CD8 T cells [12, 27] as shown in Figure 1.2. The activity of DCs are regulated by cytokines secreted by macrophages and NK cells [28].

In the presence of tumour antigens, $CD8^+$ T cells differentiate into cytotoxic T lymphocytes (CTLs), $CD8^+$ T cells differentiate into type 1 (Th1), type 2 (Th2), type 9 (Th9), type 17 (Th17), type 22 (Th22), and $FoxP3^+$ cells which have very distinct biological

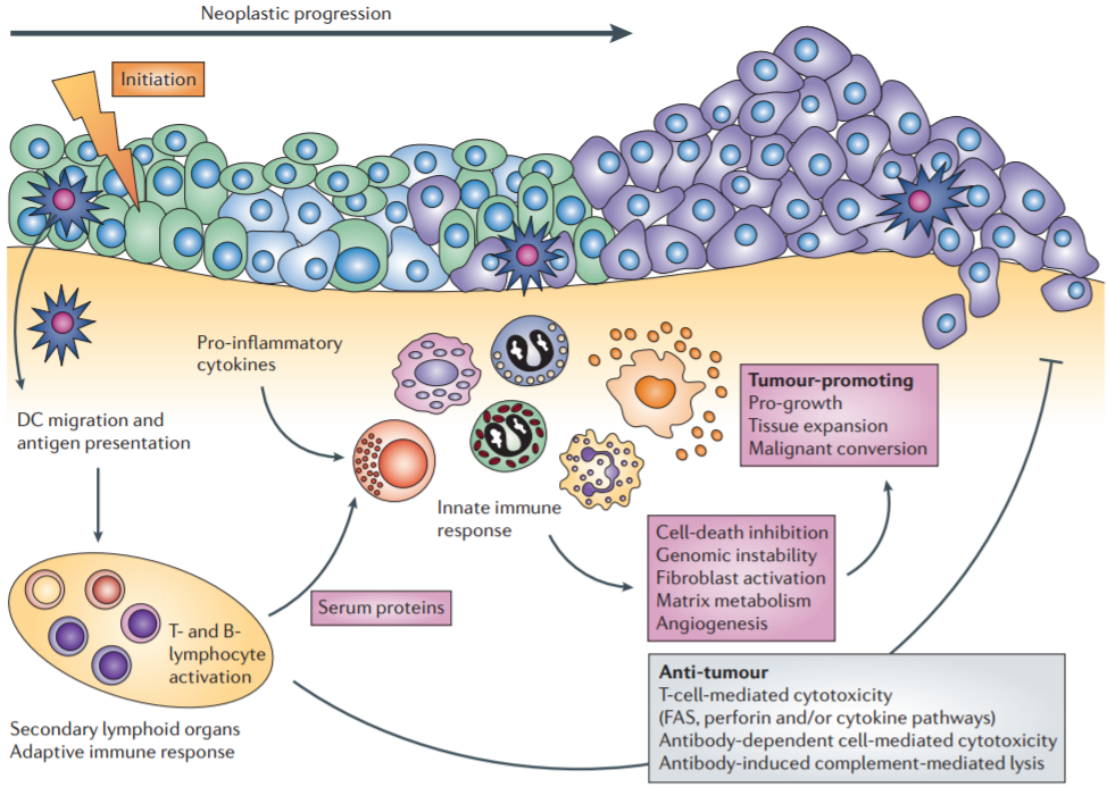


Figure 1.2: Activated macrophages engulf tumour cells, secrete cytokines to send signals to activate dendritic cells and other immune cells. The outcome can favour either a tumour-promoting response or an anti-tumour immune response. Diagram adapted from De Visser et al. [13]

roles [29, 30]. The crucial difference between $CD4^+$ T cells and $CD8^+$ T cells is that they recognize antigens in different pathways. $CD8^+$ T cells recognise endogenous antigens presented by MHC class I molecules while $CD4^+$ T cells recognise antigens presented by MHC class II molecules [17, 31]. Cytokines are considered the main determining factor in the initial differentiation of T cell subsets and they are used to activate an immune response. In the next section, we look into the role of cytokines in promoting tumour development.

1.3 Exploring cytokine world during carcinogenesis

Cytokine signals regulate the activation or inhibition of immune cells and proliferation of tumour cells [10]. Cytokines have three functional classes; pro-inflammatory, anti-inflammatory, and lymphocytes growth factors or cytokines that can polarize an

immune response due to the presence of an antigen [32, 33]. Type 1 immune signals including type 1 macrophages, CTLs, NK and Th1 cells highly produce pro-inflammatory cytokines whereas anti-inflammatory cytokines are secreted by tumour cells and type 2 immune signals, such as type 2 macrophages, Th2, and Th17 cells [12, 16, 29, 33]. The balance of cytokines in a TME is critical for the human body's internal stability. Understanding the role of cytokines during tumour development, proliferation, and progression is not easy because of its pleiotropic, antagonistic, redundant and multifunctional nature [19, 32]. For example, IFN-γ and IL-2 stimulate the proliferation and differentiation of effector cells, such as Th1 cells but, at a later stage it can promote apoptosis of these effector cells [34].

Tumour cells provoke changes in the local cytokine environment to promote its development and progression [19]. A majority of solid (aggressive) tumours secrete anti-inflammatory cytokines to promote apoptosis of effector cells, inhibit proliferation and activation of immune cells [12, 16, 29, 33, 35]. For example, (i) transforming growth factor (TGF)-β inhibits the activation and expansion of CTLs and B cells and this reduces tumour antigen expression [36], (ii) IL-10 inhibits anti-tumour cytotoxic T-cells, antigen presentation, and MHC class II expression on tumour cells [37], (iii) IL-6 and TGF-β inhibits IL-2 production, favours angiogenesis, proliferation, and invasion of breast cancer [33, 36], (iv) IL-23 reduces tumour-suppressing effects of macrophages and cytotoxic T lymphocytes [14, 37]. To counter the effect of tumour cells, type 1 immune signals secrete TNF-α to activate macrophages, NK cells and enhance the cytotoxicity of NK cells [12], IL-12 and TNF-α to stimulate the differentiation of $CD8^+$ T cells, IL-15 to promote NK cell differentiation *in vivo*, and IL-1 and IL-18 to increase the potential effect of IL-12 receptor on NK cells [18, 38]. The review in Dranoff [19] provides an extensive overview of cytokines during the development of cancer.

Cytokines have been approved as diagnostic, prognostic and therapeutic agents in various diseases [32]. An example is IL-2 which is the only pro-inflammatory cytokine that has been approved for cancer treatment. On the downside, its effects are harsh and not easily tolerated by most patients and it has a low efficacy in treating melanoma and renal cell carcinoma [32]. Cytokine activate immune cells and for an immune cell to be activated from a resting state, its gene expression must be altered [12]. In the next section, we provide an overview of RNA-Seq experiments and analysis of gene expression data.

1.4 Overview of RNA-Seq gene expression data analysis

The evolution of a cell is regulated by the genes contained in its nucleus [10]. According to UniProt Knowledgebase database (http://www.uniprot.org), the human genome contains 20,230 manually reviewed protein-coded genes. A correct gene expression is essential for normal cellular function, gene alteration can lead to over-expression or under-expression of corresponding gene, suppression of proteins and translation of new protein-causing disease. When the right combination of genes are mutated then one may develop cancer disease.

Due to advancements in technology, the next generation sequencing (NGS) application to transcriptomics (RNA-Seq) is increasingly being used to simultaneously monitor the behaviour patterns of thousands of nucleic acid sequences or proteins [39]. The sequencing framework of RNA-Seq enables investigation of all RNAs in a sample, characterizing their sequences and quantifying their abundances (count) at the same time [40]. Statistical approaches have been proposed by many studies to analyse RNA-Seq gene expression data in pursuit of finding differentially expressed genes (DEGs), that can be used to explain phenotypic differences between groups(conditions), cell types and tissue samples [41].

Statistical methods have been proposed to predict bio-marker genes that are expressed differently between samples or conditions. A gene is said to be differentially expressed if there is a difference in RNA-Seq per cell produced under different conditions. The goal of gene expression analysis is to identify biological pathways and processes associated with a selected gene set. Gene-set based methods are preferred to single-gene based methods when investigating the phenotypic differences at pathway and functional level [42]. Genes have multiple annotations in various databases, implying that a gene may have multiple notations, this often leads to gene annotation ambiguity. Resources, such as Gene Ontology (GO) [43], Gene Ontology Annotations (GOA) [44] or Kyoto Encyclopedia of genes and genomes (KEGG) [45], provide information about biological processes, functions and pathways of DEGs. Understanding biological processes and pathways of DEGs associated with breast cancer might lead to a more targeted and optimal approach to breast cancer treatment [2].

1.5 Motivation

The immune system represents a complex interacting network with most of the interactions between innate and adaptive immunity relying on the presentation of antigens, cytokines and chemokines [12, 23]. Cytokines have been used in medicine, as diagnostic, prognostic and therapeutic agents in various diseases [32]. However, cytokines have multifunctional characteristics, for example TGF-β promotes healthy (tumour-free) cell growth and function, but also enhances tumour growth and metastasis by inhibiting immune response [36]. Thus, understanding the role of cytokines in tumour-immune cell dynamics can be crucial in the fight against cancer. To our best knowledge, we did not find (i) a model that implicitly investigates cytokine-mediated innate-adaptive immunity in the presence of a tumour and (ii) research studies that combine cellular-level mathematical models with molecular-level analysis to predict DEGs that regulate cytokine production.

1.6 Objectives and significance of this study

In this study, we will

1. Predict known cytokines that play a crucial role in the activation or inhibition of an immune response and development of tumour.
2. Develop and analyse a mathematical model to investigate the role of cytokines in tumour-immune cell dynamics.
3. Establish conditions necessary for tumour clearance.
4. Predict DEGs that can be used for breast cancer bio-marker discovery and regulation of cytokine production.
5. Determine enriched pathways and biological processes associated with breast cancer.

Cancer cells are highly heterogeneous, they have different genetic make-up and have developed resistance to drugs during and after treatment. The aim of this study is to provide insights into cytokine-mediated tumour-immune cell dynamics and DEGs, enriched pathways, and biological processes across breast cancer samples that promote

carcinogenesis. These results can be useful for cancer drug discovery and will contribute to the current progress in cancer treatments, such as chemotherapy and immunotherapy.

1.7 Outline

The thesis is structured as follows

1. Chapter 1 is this introduction that briefly reviewed cancer evolution, immune cell response, cytokines during carcinogenesis, gene expression data set analysis, and gave the motivation and objectives of this study.
2. Chapter 2 gives a brief review of the mathematical models that have been proposed to explore tumour-immune cell dynamics with and without cytokine-mediated effects.
3. Chapter 3 provides a review of a new cellular-level mathematical model proposed in this study. The model is used to predict cytokine-mediated tumour-immune cell dynamics. In this Chapter, we investigate steady states, and their stability and numerical simulations of the model.
4. Chapter 4 analyses gene expression data set for breast cancer patients and healthy (tumour-free samples) to predict DEGs that are either over-expressed or under-expressed across the two samples. Here, we use the DEGs to predict enriched pathways and biological processes associated with breast cancer disease.
5. Chapter 5 discusses the findings of this study and future directions.

Chapter 2

Literature Review

2.1 Brief review of the biological background

The main function of the immune system is to monitor tissue homoeostasis, to protect against invading or infectious pathogens and to eliminate damaged cells [12]. Over the past decade, studies have indicated that the immune system can recognize and partially suppress nascent and immunogenic tumours. The promptitude of an immune response is a consequence of the availability of innate immune cells that express their responding receptors before exposure to the stimulants [12].

Innate immune cells are a population of lymphocytes that are always present and control cancer cells. Human NK cells are made up of approximately 10% of the peripheral blood lymphocytes, these cells are phenotypically characterised by the presence of CD56 and absence of CD3 [46]. NK cells can be activated due to low levels of MHC I, over expression of CD27, gp96 or NKG2D ligands [28] or up regulating ligands for stimulatory and co-stimulatory molecules expressed on NK cells, such as NKG2D, CD244, CD28 and CD137 [18]. An infiltration of NK cells in a TME has been linked to favourable prognosis in cancer patients [47]. Because cancers often express stress-related genes MICA and MICB, which function as ligands for NKG2D receptors expressed by NK cells, and cytotoxic lymphocytes [19]. The maturity and apoptosis of DCs is dependent on the NK cell activating receptor NKp30 although, this process can be counter-regulated by killer-cell immunoglobulin-like (KIR) and NKG2A inhibitory receptors [28, 48]. NK are known to be the main supply of IFN-γ at the early stage of tumour growth. Production of IFN-γ secreted during NK cell-mediated tumour rejection is critical for priming of Th1 cells, activation or induction of MHC I molecules in DCs and tumour cells, and proliferation

of CTLs particularly when tumours express CD70 or CD80 and CD8649 [28]. A mature DC (mDC) can express multiple co-stimulatory molecules, growth factors and cytokines that are important in priming of effector T cell response, activating of resting NK (rNK) cells, enhancing the cytotoxicity of NK cells, promoting the survival of NK cells and $CD8^+$ T cells and that affect endothelial, epithelial and mesenchymal cells in the local TME [12, 18, 27, 38].

The adaptive immune cells are the mediators of immunity [12]. An adult human being has about 300 billion of T-cells [12]. During the early stage of tumour growth, $CD4^+$ T cells secrete cytokines, such as IFN-γ to help expand and promote adequate functioning of $CD8^+$ T cells [17, 30]. However, $CD8^+$ T cells lack the ability to orchestrate a broad anti-tumour response which some of the $CD4^+$ T cells subsets in Figure 2.1 have. Sometimes $CD8^+$ T cells fail to function properly in the absence of $CD4^+$ T cells [17] because tumour cells have the ability to frequently change by processing and secreting endogenous antigens or immune-suppressive soluble factors, activating co-stimulatory molecules or losing of key molecules required for antigen recognition and presentation [31].

At a later tumour stage naive $CD4^+$ T cells can become polarized into various subsets that are directly involved in mediating *in vivo* tumour regression or evasion [17, 29]. There exists many subsets of T cells but Th1, Th2 and Th17 cells are commonly studied as they have an immune response towards/during an invasion. The first subsets to be identified were Th1 and Th2 cells [49]. A new $CD4^+$ T cell subset, Th17 was identified in 2015, it plays a central role in inflammation and auto-immune diseases but it is also known to support tumour growth [30]. The presence of Th2 and Treg promote tumour growth by suppressing the commitment of both the $CD4^+$ and $CD8^+$ T cells. Th1 results in the regression of tumour and supports the differentiation of $CD8^+$ T cells whereas Th2 suppresses it. Studies indicate an increase in tumour incidence with inflammation and several cytokines such as TNF-α, TGF-β, IL-6 and IL-23 have been linked to tumour promoting inflammation and Th17 lineage [14]. T cell subsets are regulated by different signal transducer and activator of transcription (STAT) and transcription factor [27]. Figure 2.1 illustrates the activation and differentiation of these T cell subsets.

$CD4^+$ T cells secrete cytokines like IL-2 when induced by APCs to stimulate and promote the proliferation and differentiation of T cells [36]. Using Figure 2.1, at the transcriptional level, Th1 development is induced by pathogens that stimulate production of IFN-γ and IL-2 through signals from STAT4, STAT1 which then activates T-box tran-

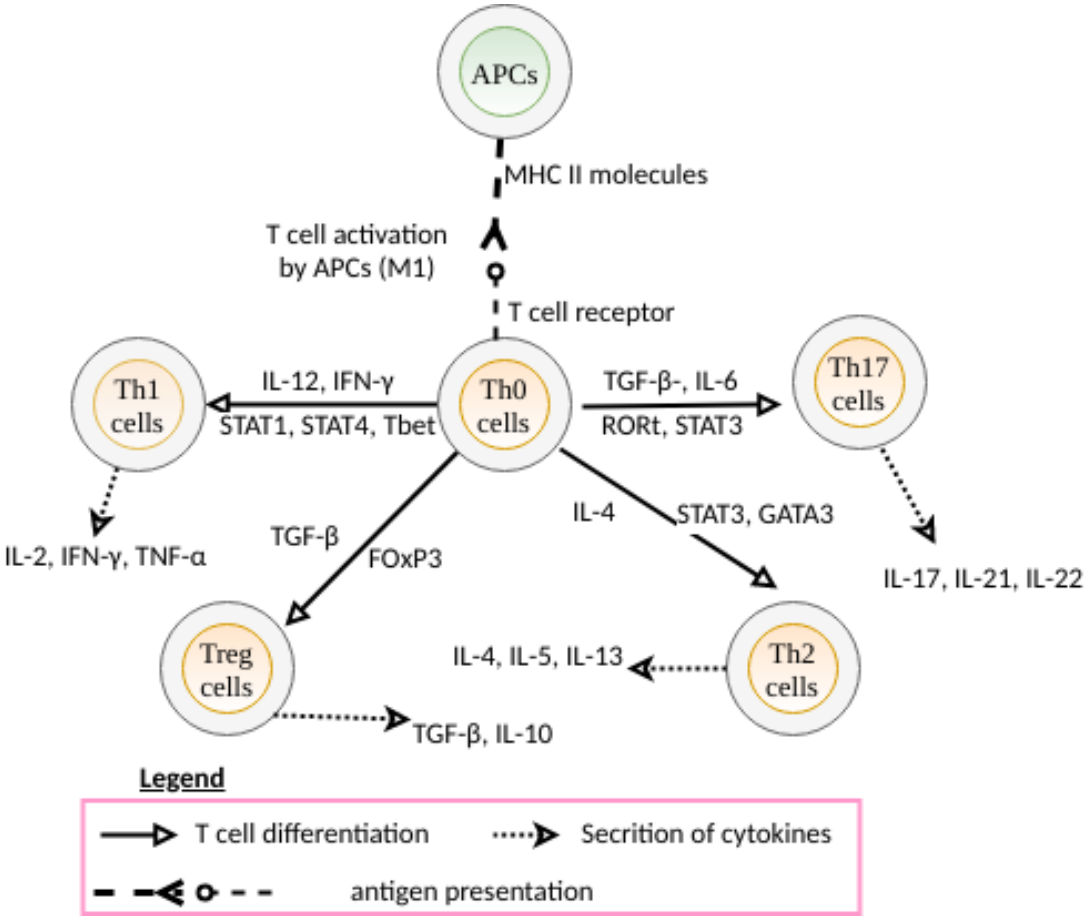


Figure 2.1: T cells are activated and can different into various subsets due to the presence of tumour cells, transcription factors and cytokines.

scription factor T-bet/Tbx21 [30, 50]. Th2 is produced when IL-4 stimulates the expression of GATA3 transcription factor through signals from STAT6 and Th17 require STA3 and the retinoic-acid-related orphan receptor ($ROR_{\gamma}t$) for its commitment and differentiation [30, 50]. $CD4^{+}$ T helper cells can be differentiated into Treg that are defined by expression of forkhead box P3 (FoxP3) and they play an anti-inflammatory role [50]. Treg and Th17 have shown to be dependent on TGF-β for their differentiation [51]. The addition of IL-6 or IL-21 to TGF-β results in the differentiation of Th17 [52].

Although, it is still unclear whether both IL-6 and TGF-β influence the commitment and development of Th17 cell lineage. For instance, experiments on mice and even Mangan et al. [51] demonstrated that TGF-β is required for the differentiation of Th17 commitment independently of IL-23 but Wilson et al. [53] demonstrated that TGF-β and IL-6 do

not play any key role in the commitment and differentiation of Th17 T cells and demonstrated that IL-23 and IL-1β are responsible for the commitment of Th17 T cells *in vitro*. Wilson et al. [53] further observed that presence of IL-6 and TGF-β even blocked IL-23 induced development of Th17 in humans. Although, similar results were observed in the inhibiting effect of IL-4 and IL-12 on the IL-23 Th17 development in humans and mice. Contrarily to observations made in mice experiment, Wilson et al. [53] showed that TGF-β may actually inhibit the differentiation of Th17.

Majority of solid cancer cells do not have MHC class II molecules on their surface, but tumour cells up regulate MHC class II molecules upon exposure to IFN-γ produced by $CD4^+$ T in the presence of interleukin (IL)-12 produced by mature DCs and macrophages [35, 54]. MHC expression on tumour cells increases their immunogenecity and cell-surface NK receptors recognizes MHC I molecule signal which activates NK cell functions to selectively lyse tumour or infected pre-cancerous cells [18].

Tumour-immune cell interactions can be complex and mathematical models are developed to (i) understand the dynamics of tumour cells and the components of its local TME, (ii) uncover basic mechanisms of these dynamics and (iii) identify new hypotheses or ideas that can be tested experimentally both *in vivo* and *in vitro* without incurring a lot of cost [55]. In the next section, we briefly discuss mathematical models that have been developed to explore tumour-immune cell dynamics.

2.2 Brief review of tumour-immune mathematical models

Tumour-immune models have been around for almost 3 decades [55]. Significant efforts have been made towards understanding cancer evolution using mathematical models. These models can be categorised as (i) molecular-level models developed to understand signalling pathway involved in the activation of an immune response, (ii) cellular-level models developed to understand the role of innate and adaptive cells, and (iii) tissue-scale models developed to understand the distribution of immune cells inside solid tumours [55]. We refer the reader to Eftimie et al. [55] for an extensive review describing these categories of mathematical models.

Much of the original work on investing tumour-immune cell dynamics using mathematical models was done by Kuznetsov et al. [56] and colleagues. Over the past 20 years, intensive research has focused on the interactions between tumour cells and immune effector cells, such as (i) innate immune cells [56], (ii) adaptive immune cells [56], (iii)

combination of innate and adaptive immune cells [57–59] (iv) cytokines and the adaptive immune cells [34, 36, 60, 61]. Reviews in Eftimie et al. [62], Adam and Bellomo [63] provide a comprehensive summary of mathematical models proposed to investigate and understand tumour-immune dynamics.

Cytokines play a crucial role in activating an immune response, priming of T helper cells, and promoting tumour development. Tumour-immune models have been proposed to investigate cytokines implicitly [57], as state variables [36, 60, 61], or as quasi-steady states because cytokines evolve (are produced and decay) at a much faster time scale than the immune cells [34]. The first model to explore the role of cytokines, specifically IL-2 was proposed by Kirschner and Panetta [61], who developed a three-equation model referred to as Kirschner-Panetta (KP) system as shown in Equation 2.2.1

$$\frac{dE}{dt} = cT - \mu_2 E + \frac{p_1 E I_L}{g_1 + I_L} + s_1, \tag{2.2.1a}$$

$$\frac{dT}{dt} = r_2(1 - bT) - \frac{aET}{g_2 + T}, \tag{2.2.1b}$$

$$\frac{dI_L}{dt} = \frac{p_2 ET}{g_2 + T} - \mu_3 I_L + s_2, \tag{2.2.1c}$$

with initial conditions, $E(0) = E_0$, $T(0) = T_0$, $I_L(0) = I_{L_0}$, and where E is the activated immune effector cells that destroy tumour cells T, stimulated by IL-2 I_L. The first Equation 2.2.1a represents the rate of change of immune (effector) cells population. Effector cells are recruited at a rate c by the antigenicity of tumour cells and s_1 by external sources (treatment effect). The proliferation of effector cells is simulated at a rate p_1 by IL-2. The saturated effects of an immune response g_1 is modelled as the third term using Michaelis-Menten form. Effector cells have a natural death rate μ_2. The second Equation 2.2.1b represents the rate of change of tumour cell population. Tumour cells have an intrinsic growth rate r_2 and a carrying capacity b. The strength of the immune response a is dependent on the saturation effect of tumour cells g_2. The third Equation 2.2.1c represents the concentration level of IL-2. IL-2 is recruited at a rate p_2 by its interaction with tumour cells with a half saturation constant g_2 and by external supply source s_2 (treatment term). IL-2 degrades at a rate μ_3.

The study in Kirschner and Panetta [61] identified that high amounts of IL-2 together with immunotherapy can lead to tumour clearance, but at the expense of uncontrolled immune over-stimulation, a condition where the immune system grows without bound [61]. The study in Arciero et al. [36] included one more equation for TGF-β to the Kirschner-Panetta (KP) system in Equation 2.2.1. The study in Arciero et al. [36] found that aggres-

sive tumours secreted TGF-β which inhibited IL-2 production, reduced antigen expression, activation of immune effector cells and tumour detection by effector cells. These mechanisms promoted tumour development and progression. Arciero et al. [36] included a fifth equation for small interfering RNA (siRNA) therapy to the KP system of equation. This therapy inhibited the production of TGF-β by blocking TGF-β synthesis. Unfortunately, the therapy did not yield a persistent tumour dormancy, but it offered mechanisms to counter the effects of TGF-β such as IL-2 inhibition, tumour escape and growth [36]. Aggressive tumours were more destructive compared to their passive counterparts that were not secreting TGF-β [36].

In the next Chapter, we propose and analyse a new ten-equation non-spatial mathematical model. The model builds on the structure of models proposed by Yates et al. [34], Kuznetsov et al. [56], Eftimie and Hamam [57] and it adds tumour-promoting effect of a third subset of T helper cells known as Th17 cells, and tumour-suppressing effect of NK cells together with the role cytokines play in activating an immune response.

Chapter 3

Modelling cytokine-mediated tumour-immune cell dynamics

3.1 Introduction

Mathematical models can be used to improve our understanding of dynamics that underlie an immune response. The idea here, is to introduce anti-inflammatory cytokines (IL-6, IL-10, IL-23, and TGF-β) and pro-inflammatory cytokines (IL-12, IFN-γ, TNF-α) as quasi-steady states, model there influence in activating innate (type 1 and 2 macrophages, NK cells) and adaptive immune cells (CTLs, Th0, Th1, Th2, Th17 cells). In the next sections, we develop a new mathematical model, analyse the model and draw conclusions from the analysis. Due to lack of spatial data from literature and clinical experiments, the model developed will not capture spatial dynamics.

3.2 Mathematical model description

Figure 3.1 represents a schematic diagram describing cytokine-mediated tumour-immune cell dynamics considered in this study.

Cancer "danger signals" and antigens activate the innate and adaptive immunity mediated by cytokines as shown in Figure 3.1. Resting macrophages are activated by the presence of tumour antigens and they can differentiate into two subsets type 1 (M1) and type 2 (M2) macrophages [12]. According to Figure 3.1, the activated M1 macrophages can activate naive $CD4^+$ and $CD8^+$ T cells stimulated by cytokines, such as IL-12 and tumour necrosis factor (TNF)-α [12, 64]. In the presence of tumour antigens, $CD8^+$ T cells

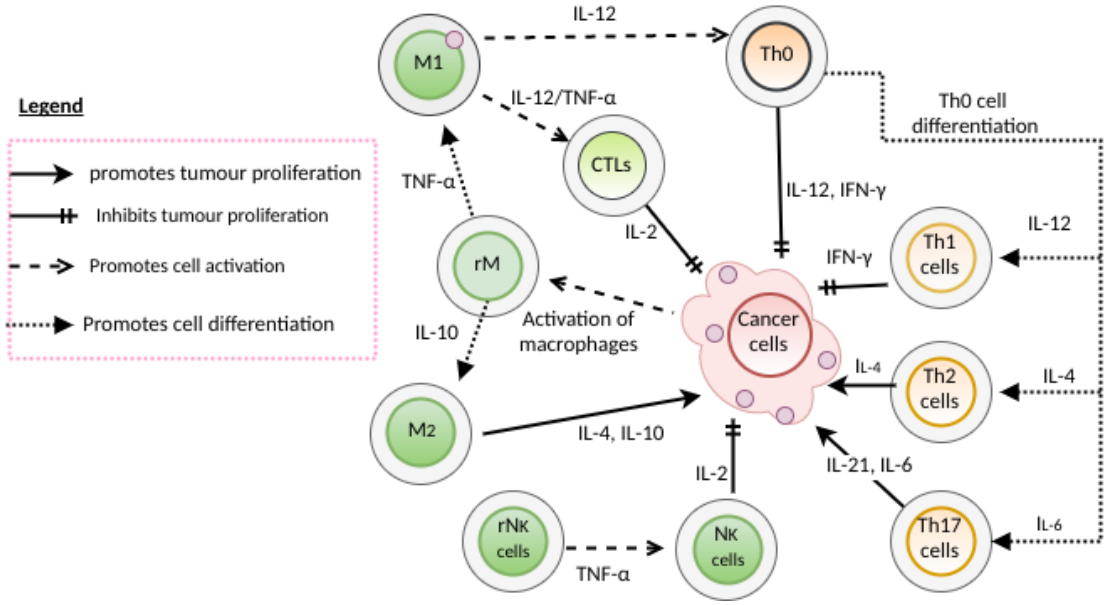


Figure 3.1: Schematic diagram to represent cytokine-mediated innate-adaptive immunity in the presence of tumour cells. Resting macrophages recognize tumour antigens. Activated macrophages have the ability to engulf tumour cells, secrete cytokines to send signals to activate CTLs, naive T helper (Th0) cells and resting NK (rNK) cells. The outcome of tumour-immune cell dynamics can favour a type 2 immune response at the expense of type 1 immune response or vice versa.

differentiate into CTLs and T helper cells can differentiate into various subsets including type 1 (Th1), type 2 (Th2), type 17 (Th17) cells which have very distinct biological roles [29, 30]. Th1 cells can activate M1 macrophages, CTLs and B cells [12]. Type 2 immune signals (M2 macrophages, Th2 and Th17 cells) promote tumour development by producing tumour-promoting cytokines, such as IL-6, IL-10, IL-23 and TGF-β that inhibit the activation of type 1 immune signals (M1 macrophages, CTLs, NK and Th1 cells) and proliferation of tumour-suppressing cytokines, such as IL-12, interferon (IFN)-γ, IL-2, TNF-α. NK cells can kill/destroy without being activated but, they require TNF-α and IFN-γ to increase there cytotoxicity and cell population [12].

The ODE model developed for this study describes dynamics of the following variables as they vary through time t

1. Cytokine; tumour-promoting cytokines including IL-6, IL-10, IL-23, TGF-β and tumour-suppressing cytokines including IL-12, TNF-α and IFN-γ secreted by immune and tumour cells.

2. Innate cell; activated type 1 macrophages (M1) $M_1(t)$, type 2 macrophages (M2) $M_2(t)$ and natural killer (NK) cells $N_k(t)$.

3. Adaptive cell; cytotoxic T lymphocytes (CTLs) cells $T_c(t)$, activated naive T helper (Th0) cells $T_0(t)$, type 1 T helper (Th1) cells $T_1(t)$, type 2 T helper (Th2) cells $T_2(t)$ and type 17 T helper (Th17) cells $T_{17}(t)$.

4. Infective or infected cell; cancer $C(t)$ and infected pre-cancerous cells $P_c(t)$.

To simplify the model, we made the following assumptions

3.2.1 Model assumptions

1. Tumour cells and immune cells grow logistically up to a carrying capacity β_i in the absence of an immune response or tumour cells, respectively [56, 65, 66]. This concept was first explained by Gompertz [66], the study observed a sigmoidal population growth curve when modelling cell replication and death.

2. The adaptive immunity is dependent on signals from the innate immunity [12]. Innate immune cells (NK cells, M1 and M2 macrophages) are always present and can kill tumour cells but, adaptive immune (CTLs, Th0, Th1, Th2, Th17) cells have to be activated to kill.

3. Cytokines can be produced by innate immune cells, T cells and other cells within the TME, such as neutrophils, eosinophils, and basophils [57]. We made an assumption that these other cells within the TME have a negligible influence to the production of cytokines.

4. Tumour cells, Tregs, M2 macrophages, Th17 cells produce large amounts of TGF-β, IL-6, IL-10 and IL-23 to inhibit any type 1 immune signals [37]. Suppressive effects of Tregs and Th17 are modelled implicitly [37]. We assumed that Tregs increase the population of IL-10 and TGF-β by a factor S_2 whereas Th17 increase the population of IL-6, IL-10 and TGF-β by a factor S_1 [14, 29, 67].

5. Cytokines act as growth factors to stimulate proliferation and activation of immune cells [12]. We considered the stimulating effects of cytokines in recruiting/activating immune cells but, ignored/implicitly modelled its effect in promoting the proliferation of immune cells.

6. IFN-γ inhibits the proliferation of Th17 cells but, not Th2 cells [51]. IL-4, IL-23 and TGF-β can inhibit the proliferation of type 1 immune signals [27, 34, 37, 57].

3.2.2 Description of model equations

Parameters are used to describe the rate of change of cell population. Table 3.1 summarizes the notations used in describing the model.

Table 3.1: Convention used in naming the parameters in the model

Parameter	Description of terms
ξ_i	Production rate of type i signalling molecule (cytokine)
ρ_i	Activation rate of effector cell i by the presence of tumour cells which is determined by cytokine signals
α_i	Intrinsic growth rate of cell i
β_i	Carrying capacity of cell i
δ_i	Inactivation rate of effector cell i due to its interactions with tumour cells
Λ_c	Inactivation rate of tumour cells due to its interaction with effector cells
η_i	Inhibition rate of cell i by cytokine i
μ_i	Natural death/degradation rate of cell i
θ_i	Rate of producing new tumour cells by cell i
S_i	Production rate of cytokines by Th17 cells S_1 and Tregs S_1

The rate of change of immune and tumour cell population can be described as

$$Type\ 1\ immune\ signal = activation + proliferation - necrosis(inactivation) - apoptosis$$

$$Type\ 2\ immune\ signal = activation + proliferation - apoptosis(natural\ death)$$

$$Cancer\ cell = proliferation + recruitment - necrosis - apoptosis$$

$$Infected(pre - cancerous)\ cell = production - apoptosis$$

The equations of the model follow the above description and can be described as shown below

- Equation for cytokines: Cytokines evolve at a much faster time scale than the immune cells thus, we made an assumption that cytokine dynamics can be described by a quasi-steady state [34]. Pro-inflammatory (tumour-suppressing) cytokines are produced at a rate ξ_1 by type 1 immune signals. Anti-inflammatory (tumour-promoting) cytokines are produced at a rate ξ_2 by type 2 immune signals and

tumour cells. We made an assumption that, Th17 cells boost the production of IL-6, IL-10 and TGF-β by a factor S_1 whereas T regulatory cells boost the production of IL-10 and TGF-β by a factor S_2. Table 3.2 lists quasi-steady states of these cytokines, where superscript i, a and c denote cytokines secreted by the innate, adaptive and tumour cells, respectively. For simplicity, we considered all cytokines at a population-level, and made an assumption that cells, such as neutrophils and eosinophils have negligible effect to the concentration of cytokines.

Table 3.2: Key cytokines involved in tumour-immune interactions.

Tumour-promoting cytokines	Tumour-suppressing cytokines
$I_4^i = \xi_2 M_2$	$I_\gamma^i = \xi_1 N_k$
$I_\beta^c = \xi_2 S_1 S_2 C$	$I_{12}^i = \xi_1 M_1$
$I_{23}^c = \xi_2 C$	$I_\alpha^i = \xi_1 M_1$
$I_6^i = \xi_2 S_2 M_2$	$I_\gamma^a = \xi_1 (T_1 + T_c)$
$I_{10}^c = \xi_2 S_1 S_2 C$	$I_\alpha^a = \xi_1 T_1$

- Equation for Type 1 macrophages:

  The first term of the equation describes the recruitment rate. Type 1 macrophages are activated at a rate ρ_m due to the presence of cancer cells stimulated by TNF-α [12, 64]. Where $C_i = C/(\kappa + C)$ represents the saturation effect of cancer cells and κ is the half saturation level of cancer cells that stimulates an immune response [60]. The second term describes the proliferation rate. M1 and M2 grow logistically at a rate α_m up to a carrying capacity β_m. The proliferation of effector cells, such as M1 macrophages is inhibited at a rate η_3 by IL-23 and TGF-β [27, 34, 37, 57]. The interaction between tumour cells and M1 macrophages might lead to its deactivation at a rate δ_m. Macrophages (M_1 and M_2) have a half-life of $1/\mu_m$. The dynamics of M1 macrophages can be represented by the following equation

$$\frac{dM_1}{dt} = \rho_m I_\alpha C_i + \frac{\alpha_m M_1 \left(1 - \frac{M_1}{\beta_m}\right)}{1 + \eta_3 (I_\beta + I_{23})} - \delta_m M_1 C - \mu_m M_1.$$

- Equation for Type 2 macrophages:

  Type 2 macrophages proliferate at a rate ρ_m in the presence of cancer cells stimulated by IL-4 and IL-10 which can be secreted by Th2 cells [57, 58]. IFN-γ antagonizes the secretion of TGF-β, IL-6, and IL-10 [27]. Therefore, the proliferation of

M2 macrophages is inhibited by IFN-γ at a rate η_1 [57]. The following equation describes the rate of change of the population of M2 macrophages

$$\frac{dM_2}{dt} = \rho_m I_{10} C_i + \frac{\alpha_m M_2 \left(1 - \frac{M_2}{\beta_m}\right)}{1 + \eta_1 I_\gamma} - \mu_m M_2.$$

- Equation for Natural Killers:

  NK cells' killing ability is stimulated by IL-2 and IFN-γ secreted by Th1 and itself [12]. Natural killer cells are activated at a rate ρ_K due to the presence of cancer cells and TNF-α produced by Th1 cells and macrophages [12, 37]. NK cells grow at a rate α_k up to a carrying capacity β_k stimulated by IL-2 and IFN-γ secreted by Th1 and NK cells [12]. TGF-β and IL-23 inhibit the proliferation of NK cells at a rate η_3 [27, 37]. NK cells' interaction with cancer cells results in its inactivation at a rate δ_k and they have a half-life of $1/\mu_k$. The following equation describes the rate of change of the population of NK cells

$$\frac{dN_K}{dt} = \rho_k I_\alpha C_i + \frac{\alpha_k N_k \left(1 - \frac{N_k}{\beta_k}\right)}{1 + \eta_3 (I_\beta + I_{23})} - \delta_k N_k C - \mu_k N_k.$$

- Equation for cytotoxic T cells (CTLs):

  IL-12 plays a major role in the activation and differentiation of T cells [68]. $CD8^+$ are activated at a rate ρ_8 due to the presence of tumour cells stimulated by IL-12 which can be secreted by M1 macrophages [12, 68]. CTLs proliferate at a rate α_t up to a carrying capacity β_t promoted by IL-2 and IFN-γ secreted by NK and Th1 cells [17, 69]. TGF-β inhibits the proliferation of CTLs at a rate η_3 [14, 34, 53]. Tumour cells kill CTLs at a rate δ_t and they have a half-life of $1/\mu_8$. The following equation describes the rate of change of the population of CTLs

$$\frac{dT_c}{dt} = \rho_8 I_{12} C_i + \frac{\alpha_t T_c \left(1 - \frac{T_c}{\beta_t}\right)}{(1 + \eta_3 I_\beta)} - \delta_t T_c C - \mu_8 T_c.$$

- Equation for activated helper T (Th0) cells:

  Naive $CD4^+$ T cells are activated at a rate ρ_t in the presence of tumour cells stimulated by IL-12 [35, 68]. Th0 cells proliferate at a rate α_t up to a carrying capacity β_t in the presence of IFN-γ produced by NK cells [30, 67]. The proliferation of Th0 cells is inhibited at a rate η_3 by TGF-β [14, 34, 53]. Helper T cells can differentiate into three major subsets including Th1, Th2 and Th17 cells [28, 30] depending

on the cytokines in their environment. We implicitly modelled the differentiation of T cell subtypes. Tumour cells destroy/kill helper T cells at a rate δ_t and they decay at a rate μ_t. Activated Th0 cells differentiate into various subsets including Th1, Th2 and Th17 cells [28, 30] that is highly dependent on cytokines in the local environment. The differentiation of T cell subtypes was modelled implicitly. The following equation describes the rate of change of the population of Th0 cells

$$\frac{dT_0}{dt} = \rho_t I_{12} C_i + \frac{\alpha_t T_0 \left(1 - \frac{T_0}{\beta_t}\right)}{(1 + \eta_3 I_\beta)} - \delta_t T_0 C - \mu_t T_0.$$

- Equation for type 1 T helper (Th1) cells:

  Helper T cells differentiate into Th1 cells at a rate ρ_t if the tumour site is rich in IL-12 [58, 60, 67]. IL-2 stimulates the proliferation of Th1 effector cells [12]. TGFβ inhibits the proliferation of Th1 cells at a rate η_3 and by IL-4 at a rate η_2 [14, 16, 37, 53]. The following equation describes the rate of change of the population of Th1 cells

$$\frac{dT_1}{dt} = \rho_t I_{12} T_0 + \frac{\alpha_t T_1 \left(1 - \frac{T_1}{\beta_t}\right)}{1 + \eta_3 I_\beta + \eta_2 I_4} - \delta_t T_1 C - \mu_t T_1.$$

- Equation for type 2 T helper (Th2) cells:

  Th0 effector cells commit to Th2 cells at a rate ρ_t if the tumour site is rich in IL-4 [60, 67]. Proliferation of Th2 effector cells is inhibited at a rate η_3 by TGF-β [14, 16].The following equation describes the rate of change of the population of Th2 cells

$$\frac{dT_2}{dt} = \rho_t I_4 T_0 + \frac{\alpha_t T_2 \left(1 - \frac{T_2}{\beta_t}\right)}{1 + \eta_3 I_\beta} - \mu_t T_2.$$

- Equation for type 17 T helper (Th17) cells:

  IL-6 is necessary for the commitment of Th17 cell lineage in breast cancer tissue in humans [29]. The proliferation of Th17 cells is inhibited at a rate η_1 by IFN-γ cytokine [51]. The following equation describes the rate of change of the population of Th17 cells

$$\frac{dT_{17}}{dt} = \rho_t I_6 T_0 + \frac{\alpha_t T_{17} \left(1 - \frac{T_{17}}{\beta_t}\right)}{1 + \eta_1 I_\gamma} - \mu_t T_{17}.$$

- Equation for infected pre-cancerous cells:

  The first term in the equation describes the rate of appearance of new infections in the system due to the interaction of cancer cells with type 1 immune signals like M1 macrophages, CTLs, NK and Th1 cells. Infected pre-cancerous cells can become cancerous at a rate θ_I. This compartment is not usually monitored because of the huge costs involved in measuring the infected cells. Activated CTLs and NK cells can recognize and kill infected cells at a rate Λ_c [34]. Infected cells have a half-life of $1/\mu_I$. The following equation describes the rate of change of the population of infected pre-cancerous cells

$$\frac{dP_c}{dt} = (\delta_m M_1 + \delta_k N_k + \delta_t(T_0 + T_c + T_1))C - (\Lambda_c\,(T_c + N_k))P_c - (\theta_I + \mu_I)P_c.$$

- Equation for cancer cells:

  Cancer cells grow at a rate α_c up to a carrying capacity β_c [56, 65, 70]. The presence of M2 macrophage boosts the population of cancer cells at a rate θ_c [58]. Cancer cells can be engulfed by M1 macrophages and destroyed by effector cells like CTLs, Th1, and NK cells at a rate Λ_c [15, 34, 58]. Cancer cells can inhibit its inactivation at a rate η_3 by producing IL-10 and TGF-β [37]. Apoptosis of cancer cells occur at a rate μ_c. The following equation describes the rate of change of the population of cancer cells

$$\frac{dC}{dt} = \alpha_c C\left(1 - \frac{C}{\beta_c}\right) + \theta_c M_2 C + \theta_I P_c - \left(\frac{\Lambda_c\,(M_1 + N_k + T_c + T_1)}{1 + \eta_3 I_\beta} + \mu_c\right) C.$$

Putting all these equations together, the initial value problem is defined as

$$\left.\begin{aligned}
\frac{dM_1}{dt} &= \rho_m I_\alpha C_i + \frac{\alpha_m M_1\left(1-\frac{M_1}{\beta_m}\right)}{1+\eta_3(I_\beta+I_{23})} - \delta_m M_1 C - \mu_m M_1,\\
\frac{dM_2}{dt} &= \rho_m I_{10} C_i + \frac{\alpha_m M_2\left(1-\frac{M_2}{\beta_m}\right)}{1+\eta_1 I_\gamma} - \mu_m M_2,\\
\frac{dN_K}{dt} &= \rho_k I_\alpha C_i + \frac{\alpha_k N_k\left(1-\frac{N_k}{\beta_k}\right)}{1+\eta_3(I_\beta+I_{23})} - \delta_k N_k C - \mu_k N_k,\\
\frac{dT_c}{dt} &= \rho_8 I_{12} C_i + \frac{\alpha_t T_c\left(1-\frac{T_c}{\beta_t}\right)}{(1+\eta_3 I_\beta)} - \delta_t T_c C - \mu_8 T_c,\\
\frac{dT_0}{dt} &= \rho_t I_{12} C_i + \frac{\alpha_t T_0\left(1-\frac{T_0}{\beta_t}\right)}{(1+\eta_3 I_\beta)} - \delta_t T_0 C - \mu_t T_0,\\
\frac{dT_1}{dt} &= \rho_t I_{12} T_0 + \frac{\alpha_t T_1\left(1-\frac{T_1}{\beta_t}\right)}{1+\eta_3 I_\beta+\eta_2 I_4} - \delta_t T_1 C - \mu_t T_1,\\
\frac{dT_2}{dt} &= \rho_t I_4 T_0 + \frac{\alpha_t T_2\left(1-\frac{T_2}{\beta_t}\right)}{1+\eta_3 I_\beta} - \mu_t T_2,\\
\frac{dT_{17}}{dt} &= \rho_t I_6 T_0 + \frac{\alpha_t T_{17}\left(1-\frac{T_{17}}{\beta_t}\right)}{1+\eta_1 I_\gamma} - \mu_t T_{17},\\
\frac{dP_c}{dt} &= (\delta_m M_1 + \delta_k N_k + \delta_t(T_0+T_c+T_1))C - \Lambda_c\left(T_c+N_k\right)P_c - (\theta_I+\mu_I)P_c,\\
\frac{dC}{dt} &= \alpha_c C\left(1-\frac{C}{\beta_c}\right) - \left(\frac{\Lambda_c\left(M_1+N_k+T_c+T_1\right)}{1+\eta_3 I_\beta}+\mu_c\right)C + \theta_c M_2 C + \theta_I P_c,
\end{aligned}\right\} \tag{3.2.1}$$

with initial conditions

$$M_1(0) = M_{10},\ M_2(0) = M_{20},\ N_k(0) = N_{k0},\ C(0) = C_0 > 0 \quad \text{and}$$
$$T_c(0) = T_{c0},\ T_0(0) = T_{00},\ T_1(0) = T_{10},\ T_2(0) = T_{20},\ T_{17}(0) = T_{170},\ P_c(0) = P_{c0} = 0.$$

The initial population of the innate and cancer cells is not zero and its zero for the adaptive immune cells. On substitution of the quasi-steady state cytokines in Table 3.2 into

model (3.2.1) we have

$$\left.\begin{aligned}
\frac{dM_1}{dt} &= \varphi_1 T_1 C_i + \frac{\alpha_m M_1 \left(1 - \frac{M_1}{\beta_m}\right)}{1 + (\pi_1 + \pi_2)C} - \delta_m M_1 C - \mu_m M_1,\\
\frac{dM_2}{dt} &= \varphi_1 T_2 C_i + \frac{\alpha_m M_2 \left(1 - \frac{M_2}{\beta_m}\right)}{1 + \pi_0 T_1 + \pi_1 C} - \mu_m M_2,\\
\frac{dN_K}{dt} &= \varphi_1 (T_1 + M_1) C_i + \frac{\alpha_k N_k \left(1 - \frac{N_k}{\beta_k}\right)}{1 + (\pi_1 + \pi_2)C} - \delta_k N_k C - \mu_k N_k,\\
\frac{dT_c}{dt} &= \varphi_2 M_1 C_i + \frac{\alpha_t T_c \left(1 - \frac{T_c}{\beta_t}\right)}{1 + \pi_2 C} - \delta_t T_c C - \mu_8 T_c,\\
\frac{dT_0}{dt} &= \varphi_3 M_1 C_i + \frac{\alpha_t T_0 \left(1 - \frac{T_0}{\beta_t}\right)}{1 + \pi_2 C} - \delta_t T_0 C - \mu_t T_0,\\
\frac{dT_1}{dt} &= \varphi_3 M_1 T_0 + \frac{\alpha_t T_1 \left(1 - \frac{T_1}{\beta_t}\right)}{1 + \pi_1 T_2 + \pi_2 C} - \delta_t T_1 C - \mu_t T_1,\\
\frac{dT_2}{dt} &= \varphi_3 M_2 T_0 + \frac{\alpha_t T_2 \left(1 - \frac{T_2}{\beta_t}\right)}{1 + \pi_2 C} - \mu_t T_2,\\
\frac{dT_{17}}{dt} &= \varphi_3 M_2 T_0 + \frac{\alpha_t T_{17} \left(1 - \frac{T_{17}}{\beta_t}\right)}{1 + \pi_0 T_1} - \mu_t T_{17},\\
\frac{dP_c}{dt} &= (\delta_m M_1 + \delta_k N_k + \delta_t (T_0 + T_c + T_1))C - \Lambda_c \left(T_c + N_k\right) P_c - (\theta_I + \mu_I) P_c,\\
\frac{dC}{dt} &= \alpha_c C \left(1 - \frac{C}{\beta_c}\right) - \left(\frac{\Lambda_c \left(M_1 + N_k + T_c + T_1\right)}{1 + \pi_2 C} + \mu_c\right) C + \theta_c M_2 C + \theta_I P_c,
\end{aligned}\right\} \tag{3.2.2}$$

with initial conditions

$$M_{10},\ M_{20},\ N_{k0},\ C_0 > 0 \quad \text{and} \quad T_{c0},\ T_{00},\ T_{10},\ T_{20},\ T_{170},\ P_{c0} = 0,$$

where $\varphi_1 = \rho_m \xi_1 = \rho_m \xi_2 S_1 S_2 = \rho_k \xi_1$, $\varphi_2 = \rho_8 \xi_1$, $\varphi_3 = \rho_t \xi_1 = \rho_t \xi_2$, $\pi_0 = \eta_1 \xi_1$, $\pi_1 = \eta_2 \xi_2 = \eta_3 \xi_2$ and $\pi_2 = \eta_3 \xi_2 S_1 S_2$.

We prefer to work with the non-dimensionalized form of the model for all the analyses done on Equation (3.2.2). The model will make sense biologically if parameter values and solutions are non-negative and unique.

3.3 Model analysis

For model(3.2.2) to be biologically meaningful, we require the that solutions of the model exist, and are unique and non-negative.

3.3.0.1 Existence of Unique, Positive Solutions

The non-linear IVP in Equation (3.2.2) can be represented in vector form as

$$\begin{aligned}\frac{d\mathbf{X}}{dt} &= \mathbf{F}(t,\mathbf{X}) \\ &= (f_1(t,\mathbf{X}), f_2(t,\mathbf{X}), f_3(t,\mathbf{X}), f_4(t,\mathbf{X}), f_5(t,\mathbf{X}), f_6(t,\mathbf{X}) f_7(t,\mathbf{X}), f_8(t,\mathbf{X}), f_9(t,\mathbf{X}), f_{10}(t,\mathbf{X}))^T\end{aligned}$$

with initial conditions

$$M_1(0),\ M_2(0),\ N_k(0),\ C(0) > 0 \text{ and } T_c(0),\ T_0(0),\ T_1(0),\ T_2(0),\ T_{17}(0),\ P_c(0) = 0,$$

and where

$$\begin{aligned}&f_1(t,\mathbf{X}) = dM_1/dt,\ f_2(t,\mathbf{X}) = dM_2/dt,\ f_3(t,\mathbf{X}) = dN_k/dt,\ f_4(t,\mathbf{X}) = dT_0/dt,\\ &f_5(t,\mathbf{X}) = dT_c/dt,\ f_6(t,\mathbf{X}) = dT_1/dt,\ f_7(t,\mathbf{X}) = dT_2/dt,\ f_8(t,\mathbf{X}) = dT_{17}/dt,\\ &f_9(t,\mathbf{X}) = dP_c/dt,\ f_{10}(t,\mathbf{X}) = dC/dt.\end{aligned}$$

Theorem 3.3.1 (Existence and uniqueness)**.** *Unique solutions exist for model* (3.2.2) *for all* $t \in \mathbb{R}_+$ *with initial conditions* $\mathbf{X}(0)$.

Proof. Consider an initial value problem (IVP)

$$\frac{d\mathbf{X}}{dt} = \mathbf{F}(t,\mathbf{X}), \quad \text{with initial conditions} \quad \mathbf{X}(0), \tag{3.3.1}$$

Equation (3.2.2) can be re-written as $\mathbf{F} : \mathbb{R}_+ \times \mathbb{R}_+^{10} \mapsto \mathbb{R}_+^{10}$.

The Picard Lindelof theorem discussed by Teschl [64] was used to show existence of unique solutions. Suppose $\mathbf{F} \in C(U, \mathbb{R}^n)$ where U is an open subset of $\mathbb{R}^{n+1}$ and initial conditions $(t_0, \mathbf{X}_0) \in U$. Since the IVP in Equation (3.2.2) and (3.3.1) is differentiable, this means that $\mathbf{F}$ is Lipschitz continuous in $\mathbf{X}$ and uniformly continuous with respect to t. Hence there exists a unique local solution $\mathbf{X}(t) \in C^1(t_0 - \epsilon, t_0 + \epsilon)$ to the IVP in Equation (3.2.2) where $\epsilon > 0$. □

The solutions for model Equation (3.2.2) are unique. Since we are working with cells, the unique solutions are expected to be non-negative.

Lemma 3.3.2 (Positivity of solutions)**.** *If the initial tumour and immune cell densities at time* $t = 0$ *are non-negative* $M_1(0), M_2(0), N_k(0), T_c(0), T_1(0), T_2(0), T_{17}(0), C(0) \geq 0$ *then the solutions of the model* (3.2.2) *are non-negative for all time* $t \geq 0$.

Proof. Suppose the model has negative solutions, this implies that a time $t_0 < \infty$ exists which can be defined as

$$t_0 = inf\{t > 0 \mid M_1(t),\ M_2(t),\ N_k(t),\ T_c(t),\ T_0(t),\ T_1(t),\ T_2(t),\ T_{17}(t),\ P_c(t),\ C(t) < 0$$

Assuming this relation, model Equation (3.2.2) can be integrated from time interval $[t_0, t]$ to get the following solutions for

- Type 1 Macrophages

$$\begin{aligned}\frac{dM_1}{dt} &= \varphi_1 T_1 C_i + \frac{\alpha_m M_1\left(1 - \frac{M_1}{\beta_m}\right)}{1 + \pi_1 C} - (\delta_m C + \mu_m) M_1 \\ &\geq -(\delta_m C + \mu_m) M_1, \quad \text{for } t \leq t_0.\end{aligned}$$

  Since $M_1(t_0) \geq 0$ then the following unique positive solution exists,

$$M_1(t) \geq M_1(t_0) \exp\Big(-\int_{t_0}^{t} (\delta_m C + \mu_m) ds\Big) \geq 0 \ \text{ for } t \in [t_0 - \epsilon_1, t_0 + \epsilon_1].$$

- Type 2 Macrophages

$$\frac{dM_2}{dt} \geq -\mu_m M_2, \ \text{ for } t \leq t_0.$$

  Since $M_2(t_0) \geq 0$ then a unique positive solution exists as shown below,

$$M_2(t) \geq M_2(t_0) \exp\Big(\int_{t_0}^{t} -\mu_m ds\Big) \geq 0 \ \text{ for } \ t \in [t_0 - \epsilon_2, t_0 + \epsilon_2].$$

- Natural Killer cells

$$\frac{dN_k}{dt} \geq -N_k(\delta_k C + \mu_k), \ \text{ for } t \leq t_0.$$

  Since $N_k(t_0) \geq 0$ then a unique positive solution exists as shown below,

$$N_k(t) \geq N_k(t_0) \exp\Big(-\int_{t_0}^{t} (\delta_k C + \mu_k) ds\Big) \geq 0 \ \text{ for } \ t \in [t_0 - \epsilon_3, t_0 + \epsilon_3].$$

- CTLs

$$\frac{dT_c}{dt} \geq -T_c(\delta_t C + \mu_8), \text{ for } t \leq t_0.$$

Since $T_c(t_0) \geq 0$ then a unique positive solution exists as shown below,

$$T_c(t) \geq T_c(t_0) \exp\Big(-\int_{t_0}^{t} (\delta_t C + \mu_8) ds \Big) \geq 0 \text{ for } t \in [t_0 - \epsilon_4, t_0 + \epsilon_4].$$

- Naive T helper (Th0) cells

$$\frac{dT_0}{dt} \geq -T_0(\delta_t C + \mu_t), \text{ for } t \leq t_0.$$

Since $T_0(t_0) \geq 0$ then a unique positive solution exists as shown below,

$$T_0(t) \geq T_0(t_0) \exp\Big(-\int_{t_0}^{t} (\delta_t C + \mu_t) ds \Big) \geq 0 \text{ for } t \in [t_0 - \epsilon_5, t_0 + \epsilon_5].$$

- Type 1 T helper (Th1) cells

$$\frac{dT_1}{dt} \geq -T_1(\delta_t C + \mu_t), \text{ for } t \leq t_0.$$

Since $T_1(t_0) \geq 0$ then a unique positive solution exists as shown below,

$$T_1(t) \geq T_1(t_0) \exp\Big(-\int_{t_0}^{t} (\delta_t C + \mu_t) ds \Big) \geq 0 \text{ for } t \in [t_0 - \epsilon_6, t_0 + \epsilon_6].$$

- Type 2 T helper (Th2) cells

$$\frac{dT_2}{dt} \geq -\mu_t T_2, \text{ for } t \leq t_0.$$

Since $T_2(t_0) \geq 0$ then a unique positive solution exists as shown below,

$$T_2(t) \geq T_2(t_0) \exp\Big(-\int_{t_0}^{t} (\mu_t) ds \Big) \geq 0 \text{ for } t \in [t_0 - \epsilon_7, t_0 + \epsilon_7].$$

- Type 17 T helper (Th17) cells

$$\frac{dT_{17}}{dt} \geq -\mu_t T_{17}, \text{ for } t \leq t_0.$$

Since $T_{17}(t_0) \geq 0$ then a unique positive solution exists as shown below,

$$T_{17}(t) \geq T_{17}(t_0) \exp\Big(-\int_{t_0}^{t} (\mu_t) ds \Big) \geq 0 \text{ for } t \in [t_0 - \epsilon_8, t_0 + \epsilon_8].$$

- Infected cells

$$\frac{dP_c}{dt} \geq -P_c(\theta_I + \mu_I + \Lambda_c\,(T_c + N_k)\,,\ \text{for } t \leq t_0.$$

Since $P_c(t_0) \geq 0$ then a unique positive solution exists as shown below,

$$P_c(t) \geq P_c(t_0)\exp\Big(-\int_{t_0}^{t}(\theta_I + \mu_I + \Lambda_c\,(T_c + N_k)\,ds\Big) \geq 0 \ \text{ for } \ t \in [t_0 - \epsilon_9, t_0 + \epsilon_9].$$

- Tumour cells

$$\frac{dC}{dt} \geq -\frac{C\,(\mu_C + \Lambda_c(M_1 + N_k + T_c + T_1)}{1 + \pi_2 C},\ \text{for } t \leq t_0.$$

Since $C(t_0) \geq 0$ then a unique positive solution exists as shown below,

$$C(t) \geq C(t_0)\exp\left(-\int_{t_0}^{t}\frac{\mu_C + \Lambda_c(M_1 + N_k + T_c + T_1)}{1 + \pi_2 C}ds\right) \geq 0$$

for $t \in [t_0 - \epsilon_{10}, t_0 + \epsilon_{10}]$.

The solutions of Equation (3.2.2) are non-negative for the time interval $t \in [t_0 - \epsilon, t_0 + \epsilon]$ with $\epsilon = min\{\epsilon_1, \epsilon_2, \epsilon_3, \epsilon_4, \epsilon_5, \epsilon_6, \epsilon_7, \epsilon_8, \epsilon_9, \epsilon_{10}\}$. Which contradicts with the initial assumption on t_0. Since the initial conditions and solutions to the model are always positive, this implies that the model is biologically well posed and we can draw meaningful findings from the analysis.

□

3.3.1 Steady state solutions

Steady states are solutions that are time invariant and they are often used to investigate long-term behaviour of models. Because of the highly non-linear terms in the model, we will only consider the tumour-free steady states (TFSS).

3.3.1.1 Tumour-free steady state

The TFSS can be categorized as either type 1 or 2 immune signal as shown below

1. Type-1 tumour-free (T1TF) steady state can be characterized by

$$(M_1^*, M_2^*, N_k^*, T_c^*, T_0^*, T_1^*, T_2^*, T_{17}^*, C^*, P_c^*) = (M_1^*, 0, N_k^*, T_c^*, T_0^*, T_1^*, 0, 0, 0, 0)$$

where

$$M_1^* = \frac{\beta_m}{\alpha_m}(\alpha_m - \mu_m), \quad N_k^* = \frac{\beta_k}{\alpha_k}(\alpha_k - \mu_k), \quad T_c^* = \frac{\beta_t}{\alpha_t}(\alpha_t - \mu_8), \quad T_0^* = \frac{\beta_t}{\alpha_t}(\alpha_t - \mu_t)$$
$$T_1^* = \frac{1}{2\alpha_t}\left(\beta_t(\alpha_t - \mu_t) \pm \beta_t\sqrt{(\alpha_t - \mu_t)(\alpha_m\alpha_t + 4\alpha_m\beta_m\varphi_3 - \alpha_m\mu_t - 4\beta_m\mu_m\varphi_3)}\right)$$

This steady state can exist if the proliferation rate of cell i is greater than its decay rate, that is, $\alpha_i > \mu_i$. We consider the positive root for Th1 steady state T_1^* because a negative root will not be biologically meaningful.

2. Type 2 tumour-free (T2TF) steady state can be characterized by

$$(M_1^*, M_2^*, N_k^*, T_c^*, T_0, T_1^*, T_2^*, T_{17}^*, C^*, P_c^*) = (0, M_2^*, 0, 0, 0, 0, T_2^*, T_{17}^*, 0, 0)$$

where

$$M_2^* = \frac{\beta_m}{\alpha_m}(\alpha_m - \mu_m), \quad T_2^* = \frac{\beta_t}{\alpha_t}(\alpha_t - \mu_t), \quad T_{17}^* = \frac{\beta_t}{\alpha_t}(\alpha_t - \mu_t)$$

The T2TF steady state exists only if the proliferation rate of cell i is greater than its decay rate, $\alpha_i > \mu_i$.

3.3.1.2 Local stability analysis of the tumour-free steady state

The next generation matrix (NGM) approach was used to compute the reproduction number R_0 and investigate stability of the TFSS [71] as shown in Equation (3.3.2). R_0 can be defined as the average number of infected immune cells produced by a single tumour cell.

$$R_0 = \rho(FV^{-1}) \quad \text{with} \quad F = \frac{\partial \mathcal{F}_i}{\partial X_j}(X^*), \quad V = \frac{\partial \mathcal{V}_i}{\partial X_j}(X^*), \quad 1 \leq i,\ j \leq m, \tag{3.3.2}$$

where ρ is the spectral radius, X^* is the tumour-free steady state of model (3.2.2), matrix $\mathcal{F}_i$ represents the rate of appearance of new infections and $\mathcal{V}_i$ describes the difference between transfer rates of cells into and out of any compartment i. The rate of change of immune and tumour cells can be represented as $dX/dt = (F - V)X$. Instead of using full vectors $\mathcal{F}$ and $\mathcal{V}$, we only considered the infected/infective classes, cancer C and infected pre-cancerous P_c cells since $m = 2$.

Theorem 3.3.3 (Local stability analysis)**.** *The tumour-free steady state is asymptotically stable if $R_0 < 1$ and unstable otherwise.*

Proof. Given the two infected classes, the rate of new infections is defined as

$$\mathcal{F} = \begin{pmatrix} (\delta_m M_1 + \delta_k N_k + \delta_t (T_c + N_k))C \\ \theta_I P_c \end{pmatrix},$$

and difference between the transfer rates of the two classes as

$$\mathcal{V} = \begin{pmatrix} P_c \left(\theta_I + \mu_I + \Lambda_c (T_c + N_k)\right) \\ -\alpha_C C \left(1 - \frac{C}{\beta_C}\right) - \theta_c M_2 C + \frac{\Lambda_c (N_k + M_1 + T_c + T_1)C}{1 + \pi_2 C} + \mu_c C \end{pmatrix}.$$

Differentiating $\mathcal{F}$ and $\mathcal{V}$ with respect to the infected cells P_c and tumour C at the tumour-free steady state X^* gives

$$F = \begin{pmatrix} 0 & M_1^* \delta_m + N_k^* \delta_k + \delta_t \left(N_k^* + T_c^*\right) \\ \theta_I & 0 \end{pmatrix},$$

which is a non-negative transmission matrix and,

$$V = \begin{pmatrix} \Lambda_c \left(N_k^* + T_c^*\right) + \mu_I + \theta_I & 0 \\ 0 & \Lambda_c \left(M_1^* + N_k^* + T_1^* + T_c^*\right) - M_2^* \theta_c - \alpha_c + \mu_c \end{pmatrix},$$

the non-singular transition matrix. The inverse matrix of V can be given as

$$V^{-1} = \begin{pmatrix} \frac{1}{\Lambda_c \left(N_k^* + T_c^*\right) + \mu_I + \theta_I} & 0 \\ 0 & \frac{1}{\Lambda_c \left(M_1^* + N_k^* + T_1^* + T_c^*\right) - M_2^* \theta_c - \alpha_c + \mu_c} \end{pmatrix}.$$

The next generation matrix (NGM) is given by

$$FV^{-1} = \begin{pmatrix} 0 & \frac{M_1^* \delta_m + N_k^* \delta_k + \delta_t \left(N_k^* + T_c^*\right)}{\Lambda_c \left(M_1^* + N_k^* + T_1^* + T_c^*\right) - M_2^* \theta_c - \alpha_c + \mu_c} \\ \frac{\theta_I}{\Lambda_c \left(N_k^* + T_c^*\right) + \mu_I + \theta_I} & 0 \end{pmatrix}. \tag{3.3.3}$$

The reproduction number is the off-diagonals in Equation (3.3.3) which can be split into two partial reproduction numbers, R_{01} for the infected and R_{02} for the tumour cells

$$R_{01} = \frac{M_1^* \delta_m + N_k^* \delta_k + \delta_t \left(N_k^* + T_c^*\right)}{\Lambda_c \left(M_1^* + N_k^* + T_1^* + T_c^*\right) - M_2^* \theta_c - \alpha_c + \mu_c} \quad \text{and}$$
$$R_{02} = \frac{\theta_I}{\Lambda_c \left(N_k^* + T_c^*\right) + \mu_I + \theta_I}.$$

The spectral radius is given as

$$\rho(FV^{-1}) = \sqrt{R_0} = \sqrt{R_{01} R_{02}}. \tag{3.3.4}$$

Theorem 2 proposed by Van den Driessche and Watmough [71], states that the TFSS is locally asymptotically stable whenever $R_0 < 1$ and unstable otherwise. When $R_0 < 1$ it means that on average, less than 1 normal (tumou-free) cell become cancerous for every interaction it has with tumour cells. A combination of higher values in parameters, such as Λ_c, μ_c, and μ_I, decreases the value of R_0, whereas higher values in parameters, such as δ_m, δ_k, δ_t, θ_c, and α_c increases the value of R_0. This makes sense because, higher values in the inactivation rate Λ_c, and natural death rate μ_c of tumour cells and infected cell μ_I increases the rate of necrosis and apoptosis of tumour and infected cells, this might lead to a tumour-free steady state. It was inconclusive to say if the TFSS is stable $R_0 < 1$ or unstable $R_0 > 1$.

□

3.3.1.3 Global stability of tumour-free steady state

Theorem discussed by Castillo-Chavez et al. [72] was used to investigate the global stability of the tumour-free steady state. Considering the components of vector X as the uninfected class (immune cells) and vector Y as the infected class (infected (pre-cancerous) and tumour cells). Then the system can be defined as

$$\begin{aligned} \frac{dX}{dt} &= F(X,Y), \quad \text{for all} \quad X \in \mathbb{R}^n, \quad \text{and} \quad Y \in \mathbb{R}^m, \\ \frac{dY}{dt} &= G(X,Y), \quad \text{with} \quad G(X,0) = 0 \end{aligned}$$

Suppose $V_0 = (X^*, 0)$ is a tumour-free steady state of the model, then V_0 is globally asymptotically stable if

1. $R_0 < 1$,
2. For $dX/dt = F(X,0)$, the TFSS X^* is globally asymptotically stable and,
3. $G(X,Y) = AY - \hat{G}(X,Y)$, with $\hat{G}(X,Y) \geq 0$ for X, $Y \in \Omega$. Where $A = D'_Y G(X^*, 0)$ is an n-matrix who's off-diagonals are non-negative and Ω is a region where the system is biologically meaningful.

It is worth noting that, in our case, the number of uninfected classes, $n = 8$ and the number of infected classes, $m = 2$.

Theorem 3.3.4. *The tumour-free steady state of model (3.2.2) exists and is globally asymptotically stable if it satisfies the above conditions.*

Proof. Assuming that the first assumption is met $R_0 < 1$. The system

$$F(X,Y) = \left(\frac{dM_1}{dt}, \frac{dM_2}{dt}, \frac{dN_k}{dt}, \frac{dT_c}{dt}, \frac{dT_0}{dt}, \frac{dT_1}{dt}, \frac{dT_2}{dt}, \frac{dT_{17}}{dt}\right)^T,$$

$$G(X,Y) = \left(\frac{dP_c}{dt}, \frac{dC}{dt}\right)^T,$$

$$= \begin{pmatrix} (\delta_m M_1 + \delta_k N_k + \delta_t(T_0 + T_c + T_1))C - P_c(\theta_I + \mu_I + \Lambda_c(T_c + N_k)) \\ \alpha_C C\left(1 - \frac{C}{\beta_C}\right) + \theta_I P_c + \theta_c M_2 C - \frac{\Lambda_c(N_k + M_1 + T_c + T_1)C}{1+\pi_2 C} - \mu_c C \end{pmatrix},$$

with

$$A = \begin{pmatrix} -(\theta_I + \mu_I + \Lambda_c(T_c + N_k) & M_1\delta_m + N_k\delta_k + \delta_t\left(T_0 + T_1 + T_c\right) \\ \theta_I & -\Lambda_c\left(M_1 + N_k + T_1 + T_c\right) + M_2\theta_c + \alpha_c - \mu_c \end{pmatrix}.$$

Given that all the parameter values are always positive, it is easy to show that the off-diagonals of matrix $A = G(X^*, 0)$ are non-negative. Since the infected class $Y = (P_c, C)^T$ then,

$$AY = \begin{pmatrix} \left(\delta_m M_1 + \delta_k N_k + \delta_t(T_0 + T_1 + T_c)\right)C - P_c(\theta_I + \mu_I + \Lambda_c(T_c + N_k)) \\ \theta_I P_c - \Lambda_c C\left(N_k + M_1 + T_c + T_1\right) + \theta_c M_2 C + \alpha_c C - \mu_c C \end{pmatrix}.$$

$\hat{G}(X,Y)$ is given as

$$\hat{G}(X,Y) = AY - G(X,Y),$$

$$= \begin{pmatrix} 0 \\ C\left(\frac{\alpha_c C}{\beta_c} + \Lambda_c\left(N_k + M_1 + T_c + T_1\right)\left(\frac{1}{1+\pi_2 C} - 1\right)\right) \end{pmatrix}$$

The term $\frac{1}{1+\pi_2 C} - 1$ is negative. Therefore, the TFSS $V_0(X^*, 0)$ will be globally asymptotically stable when,

$$\frac{\alpha_c C}{\beta_c} + \frac{\Lambda_c\left(N_k + M_1 + T_c + T_1\right)C}{1 + \pi_2 C} \geq \Lambda_c(N_k + M_1 + T_c + T_1)$$

because this will imply that $\hat{G}(X,Y) \geq 0$, and unstable otherwise.

□

The steady state of the TFSS cannot be obtained analytically due to the non linearity of the model and numerical analysis is preferred. In the next section, we perform numerical simulations for model (3.2.2) using *R* version 3.4.3 (www.r-project.org).

3.4 Numerical analysis

The non-linear IVP equations in Equation (3.2.2) cannot be integrated to find exact solutions, we solved the system of equations numerically using *deSolve* package developed by Soetaert et al. [73].

3.4.1 Parameter estimation

The innate immune cells are always present in the lymph nodes, but it takes a couple of hours to days for the adaptive immune cells to become activate [12]. We assumed that the infected and adaptive immune cells (CTLs, Th0, Th1, Th2 cells) were inactive and absent at the initial time. Whereas cancer and innate immune cells were present, that is approximately 0.1×10^6 cancer cells, 2×10^4 M1 macrophages, 10^4 M2 macrophages and 10^4 NK cells were present at the initial time.

We estimated parameter values from experimental data published by Kuznetsov et al. [56], Uhr et al. [65], Diefenbach et al. [70], Wilkie and Hahnfeldt [74], Ribeiro et al. [75], Benzekry et al. [76]. In a case where a parameter had a wide range of values, we chose the most appropriate values for the model or an average. For some of the unobserved parameters, we estimated them by fitting the model in Equation (3.2.2) to the average breast cancer volume data published by Benzekry et al. [76]. Where a total of 5 experiments were conducted on 6- to 8-week-old 34 female mice to investigate tumour growth [76]. Benzekry et al. [76] implanted approximately 1×10^6 of LM2-4$^{\text{LUC}}+$ cells originally derived from MDA-MD-231 cells (triple negative breast cancer cell line), into the right inguinal mammary fat pads and measured the tumour volume (size) regularly from day 18 to 38. We used the least square method to estimate parameters by minimizing the sum of squared differences/errors (SSE) between the data and model output.

The parameter values shown in Table 3.3 were estimated as follows

- Death rate (day^{-1}): Precursor macrophages in the lymph nodes have a half-life of approximately 17.4 hours to 5 days [58], the NK cells and CD8$^+$ T cells can be depleted within 14-27 days [70], and the half-life of T lymphocytes from the spleen has been approximated to be 30 days [56]. The formula, $t_{1/2} = \ln(2)/\mu_i$ is used to calculate the death rate where $t_{1/2}$ describes the half-life of cells. Therefore, using this formulae, macrophages, CTLs, and CD4$^+$ T cells have a death

rate of $\mu_m = (\ln(2)/0.725, \ln(2)/5) \approx (0.725, 0.13)$, $\mu_8 = (\ln(2)/14, \ln(2)/27) \approx (0.0495, 0.0256)$ and $\mu_t = \ln(2)/30 \approx 0.023$, respectively. We choose $\mu_m = 0.131$, $\mu_8 = 0.034$, $\mu_t = 0.023$ and $\mu_k = 0.0412$.

- Activation rate (day^{-1}): According to de Pillis et al. [59] the maximum recruitment rate of CD8$^+$ T cells is 0.0375 and 0.025 day^{-1} for NK cells. We assumed that the activation rate $\varphi_1 = 0.02$ for the innate cells, $\varphi_2 = 0.0375$ for CTLs and activated naive T (Th0) cells, and $\varphi_3 = 0.024$ for CD4$^+$ T (Th1, Th2, Th17) cells.

- Intrinsic growth rate (day^{-1}): The median breast cancer growth rate was $\alpha_c = 0.502$ for the 34 animals [76]. The intrinsic growth rate for immune cells was estimated from Kuznetsov et al. [56], Uhr et al. [65]. We choose $\alpha_t = 0.22$, $\alpha_m = 0.2$, $\alpha_k = 0.22$.

- Carrying capacity (cell^{-1}): The average tumour volume for the 34 animals was between a range of 202-1902 mm^3 (1 mm^3 = 1000 cells) [76]. The carrying capacity for the immune cells was estimated from Kuznetsov et al. [56], Uhr et al. [65]. We assumed $\beta_k = 2.02 \times 10^9$, $\beta_t = 10^9$, $\beta_c = 10^9$ and $\beta_m = 10^9$.

- Inactivation rate (cell^{-1} day^{-1}): Tumour cells can be killed by M1 macrophages at a rate 10^{-6} and 5.3×10^{-8} by CD4 T cells [58]. NK cells are killed by tumour cells at a rate $\delta_k = 3.5 \times 10^{-6}$ [70]. We assumed $\delta_m = 10^{-6}$, $\delta_t = 10^{-7}$ and $\Lambda_c = 10^{-5}$.

- Unobserved parameter values, such as $\theta_I, \theta_c, \mu_c, \mu_I, \pi_0, \pi_1, \pi_2$ were estimated by fitting the model in Equation (3.2.2) to average data provided by [76], as shown in Figure 3.2.

Table 3.3 provides a summary of the parameter values used for numerical analysis.

3.4.2 Numerical solutions

The system in Equation (3.2.2) was solved numerically using the baseline parameter values shown in Table 3.3 and initial conditions stated above. The results are shown in Figure 3.3.

According to Figure 3.3 type 1 immune signals (M1 macrophages, NK cells, CTLs, Th0 and Th1 cells) managed to control the proliferation of type 2 immune signals (Th2, Th17 cells, M2 macrophages) and tumour cells for the first 10 days, after which type 2 immune signals over-powered type 1 immune signals and continued to proliferate for some time.

Table 3.3: Parameter values

Parameter	Baseline value	Units	Source
φ_1	0.02	day^{-1}	[58]
φ_2	0.0375	day^{-1}	[59]
φ_3	0.024	day^{-1}	[56]
π_0	0.021	day^{-1}	[76]**
π_1	0.021	day^{-1}	[76]**
π_2	0.0231	day^{-1}	[76]**
α_m	0.2	day^{-1}	[74]*
α_k	0.22	day^{-1}	[74]
α_t	0.22	day^{-1}	[74]
α_c	0.502	day^{-1}	[76]
β_m	1×10^9	cell^{-1}	[65]*
β_k	2.02×10^9	cell^{-1}	[56][65]
β_t	1×10^9	cell^{-1}	[65]*
β_c	1×10^9	cell^{-1}	[56, 65, 76]
δ_m	1×10^{-6}	$\text{cell}^{-1}\ \text{day}^{-1}$	[70]*
δ_k	3.5×10^{-6}	$\text{cell}^{-1}\ \text{day}^{-1}$	[70]
δ_t	1×10^{-7}	$\text{cell}^{-1}\ \text{day}^{-1}$	[58, 65, 77]
μ_k	0.0412	day^{-1}	[59]
μ_m	0.131	day^{-1}	[58]
μ_8	0.034	day^{-1}	[70, 75]
μ_t	0.023	day^{-1}	[56]
μ_I	0.0272	day^{-1}	[76]**
μ_c	0.0115	day^{-1}	[76]**
κ	10^3	cell	[60]
Λ_c	1×10^{-5}	$\text{day}^{-1}\ \text{cell}^{-1}$	[58]
θ_c	4.5117×10^{-12}	$\text{day}^{-1}\ \text{cell}^{-1}$	[76]**
θ_I	1×10^{-4}	$\text{day}^{-1}\ \text{cell}^{-1}$	[76]**
ξ_1	3.72×10^{-5}	pg/cell/day	[78]
ξ_2	1.5×10^{-5}	pg/cell/day	[78]
** estimated	values from	experimental	data [76]

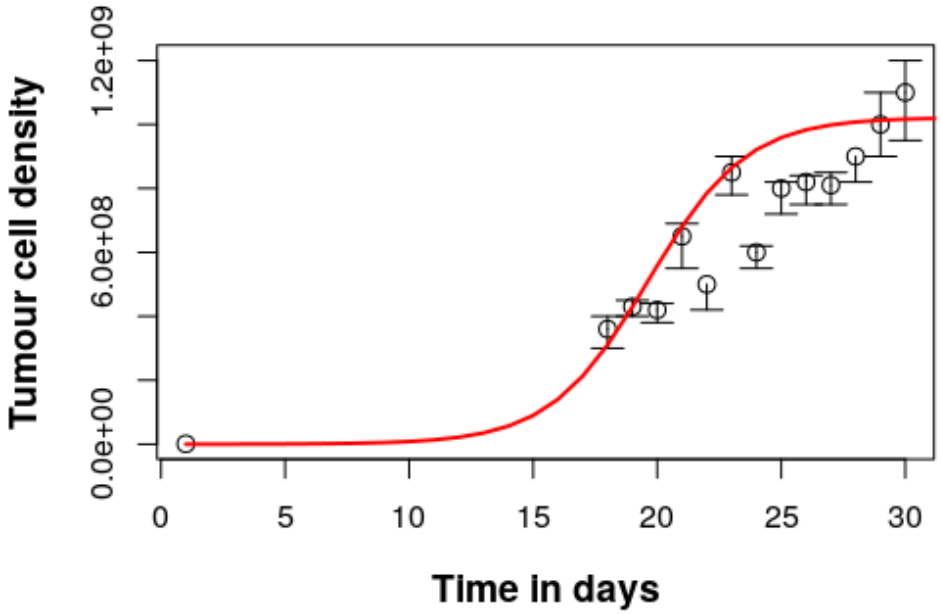


Figure 3.2: The best fit curve used to estimate unobserved parameter values from data published by Benzekry et al. [76].

Tumour cells reached a dormant steady state during this phase. These findings are similar with the findings of den Breems and Eftimie [58], Diefenbach et al. [70], but they observed that CTLs were depleted after 20 days. Panel (b-c) of Figure 3.3 indicate that M1 macrophages, NK and Th1 cells stimulate each other. The decrease in M2 macrophage cell density might have been as a result of the decrease in Th2 cell population as shown in panel (a) of Figure 3.3. This observation shows that the continuous proliferation of tumour cells is highly dependent on tumour-promoting cytokines, such as IL-4, IL-6, IL-10, IL-13, and IL-23 secreted predominantly by type 2 immune signals. Figure 3.4 visualizes the rate of change of these cytokines on day 5, 10, 15 and 40.

Numerical results in Figure 3.4 emphasized the role of cytokines in activating an immune response and promoting tumour progression [12]. The depletion of type 1 cytokines 3.4 lead to a decrease in type 1 immune cells efficiency, cytotoxicity and population, this fallout promoted tumour development as shown in panel (a) of Figure 3.3. IL-10, IL-23 and TGF-β dominated from day 25, a time when type 1 signals had been depleted. Furthermore, we analysed the effect of IL-23 and TGF-β on the population of

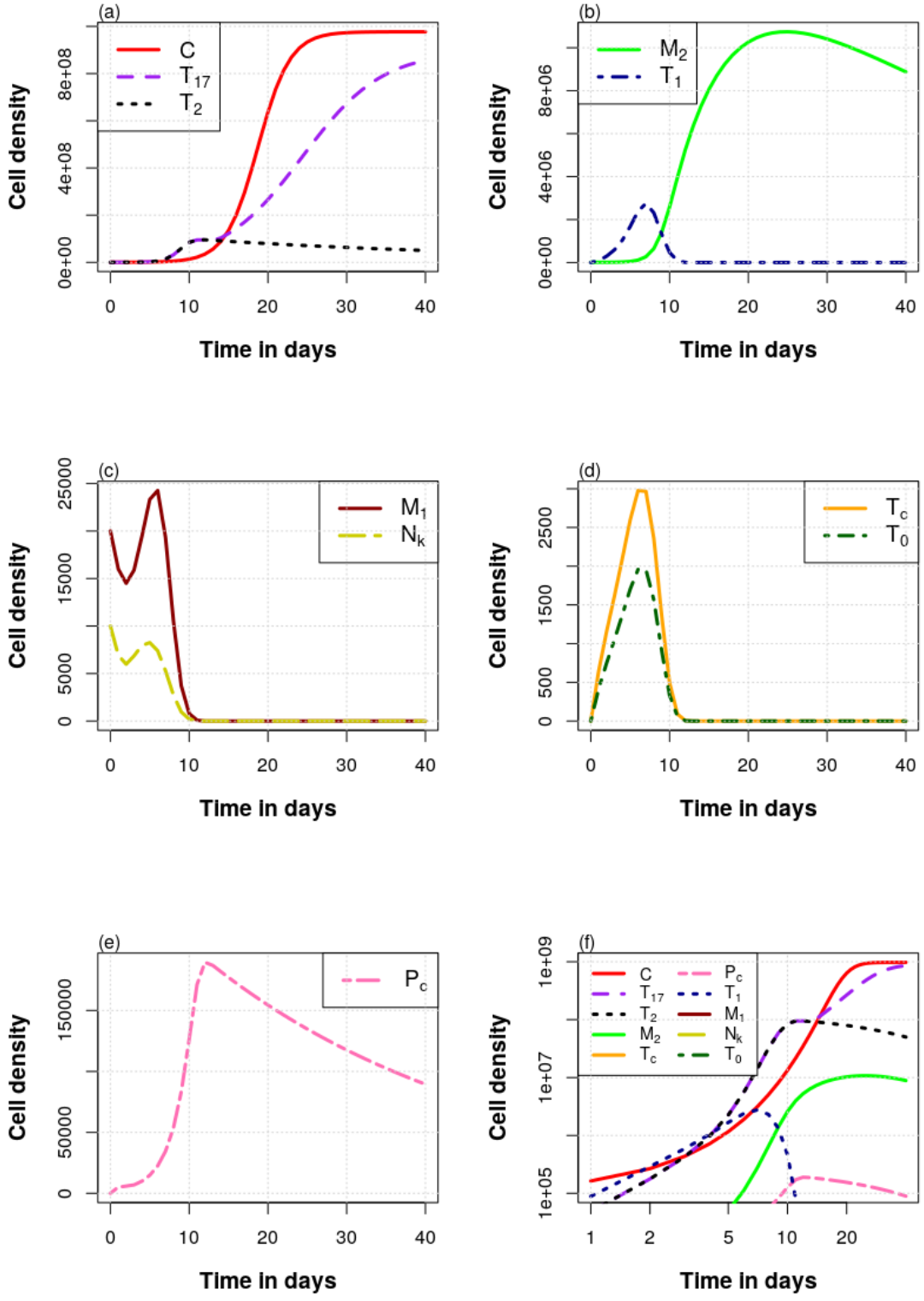


Figure 3.3: Panel (a-b) shows the positive effect type 2 immune signals (M2 macrophages, Th2 and Th17 cells) has on tumour cells. The depletion of type 1 immune signals (M1 macrophages, CTLS, NK, Th0, Th1 cells) on day 10 resulted in the proliferation and progression of tumour cells. The population of precancerous cells in Panel (e) are proportional to the population of type 1 immune signals. Panel (f) shows the cell dynamics in panel (a-e) on a logarithmic $x-$ and $y-$ axis scale.

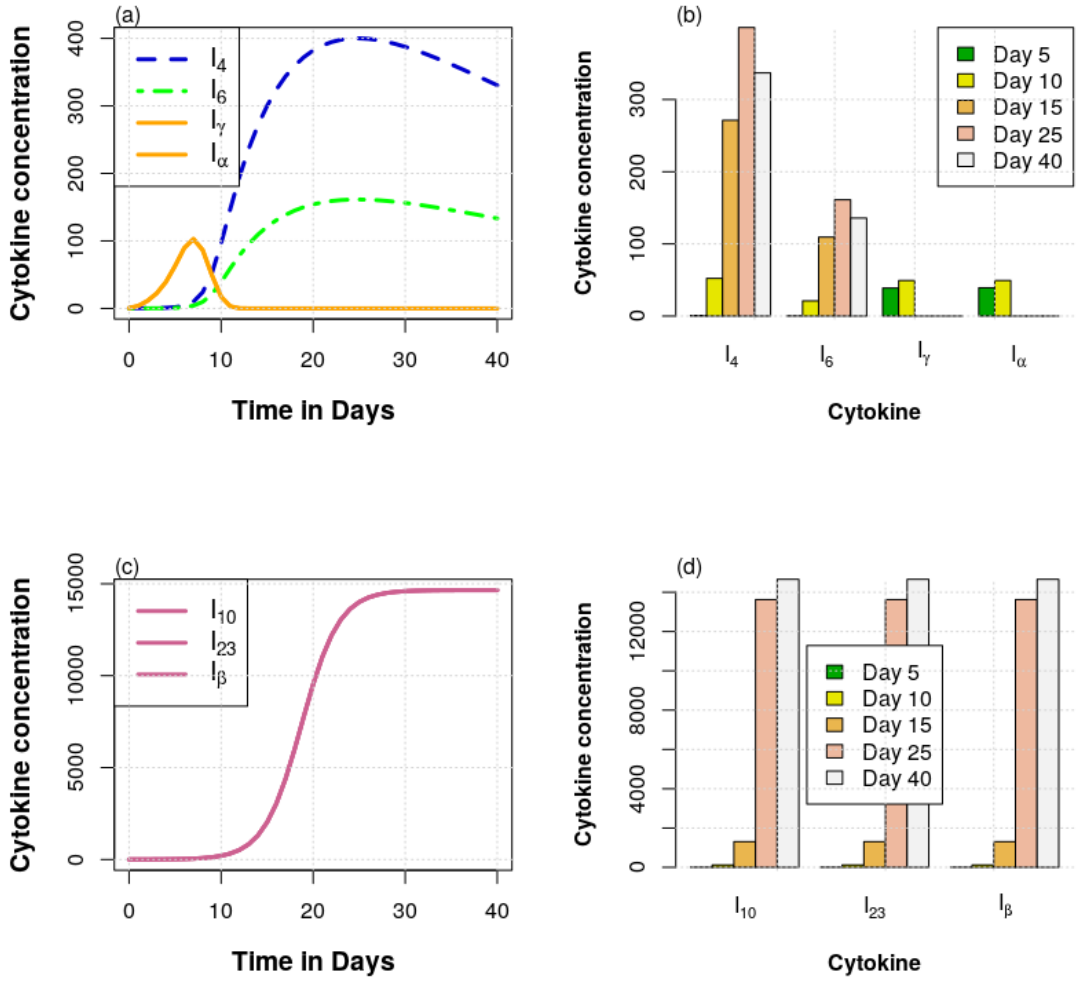


Figure 3.4: Anti-inflammatory cytokines, such as IL-4 (I_4), IL-6 (I_6), IL-10 (I_{10}), IL-23 (I_{23})and TGF-β (I_β) dominate from day 15 after the pro-inflammatory cytokines, such as TNF-α (I_α) and IFN-γ (I_γ) were depleted.

tumour and immune cells as shown in Figure 3.5.

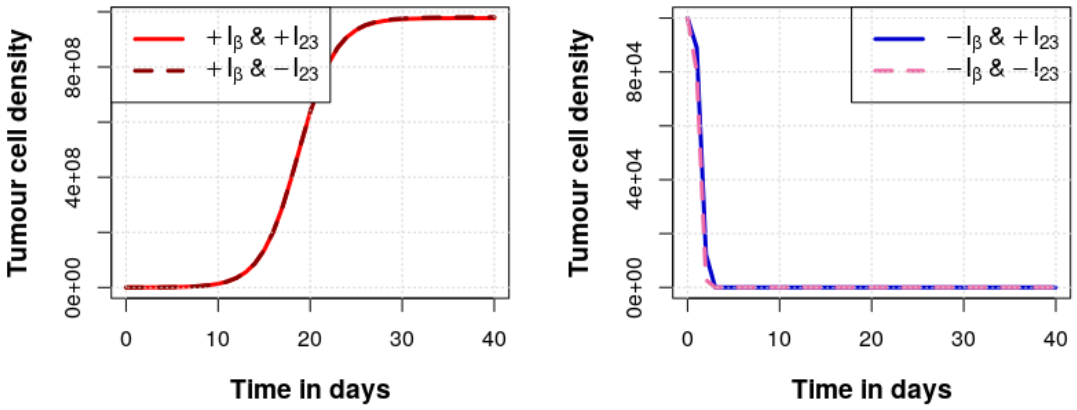


Figure 3.5: Panel (a) indicates that the removal of IL-23 has no effect on tumour cell population. However, the removal of TGF-β results in tumour clearance but, this caused an over-stimulation of immune cells.

The absence of TGF-β resulted in the elimination of tumour cells by day 4. We noticed that by day 7, all the immune cells had exceeded the carrying capacity of 10^{10} cells, this can cause a series of serious dysfunctional organs [14]. Similar results were also obtained by Kirschner and Panetta [61]. This finding illustrate the crucial role of TGF-β in controlling and regulating explosive proliferation of immune cells [14]. Arciero et al. [36] observed that TGF-β inhibited IL-2 production, reduced antigen expression, activation of immune effector cells and tumour detection by effector cells. Limited anti-inflammatory signals can lead to increased cytotoxicity and tumour regression. This might explain why the tumour cells in this study was not cleared.

Mathematical models solemnly depend on parameter values. A change in parameter values can make a lot of difference in the model output. Sensitivity analysis measures the change in a mathematical model output associated with small perturbations in the model input parameters. We performed sensitivity analysis to determine parameters that influence the output of model (3.2.2).

3.4.3 Sensitivity analysis

Sensitivity analysis measures the change in a mathematical model output associated with small perturbations in the model input parameters. We used the Flexible Modelling Environment (FME) package developed by Soetaert and Petzoldt [79] to determine the important parameters in model (3.2.2). The method used generated sensitivity functions for all variables in model (3.2.2). A summary of these sensitivity functions revealed that the (i) immune cell activation rate φ_1, φ_2, φ_3, (ii) inhibition rate π_1, π_2, (iii) half saturation constant κ, (iv) CTLs and T helper cell intrinsic growth rate α_t, (v) CTLs death rate μ_8, and T helper cells μ_t, and (vi) the carrying capacity of CTLs and T helper cells β_t were more sensitive on the output variables in model (3.2.2).

Some of the unobserved parameters including π_1, π_2 were identified as the most sensitive parameters and others including μ_I, μ_c, and θ_I were involved in the reproduction number R_0 in Equation (3.3.4). We performed global sensitivity analysis to explore the effect of varying these unobserved parameters on tumour cell population as shown in Figure 3.6.

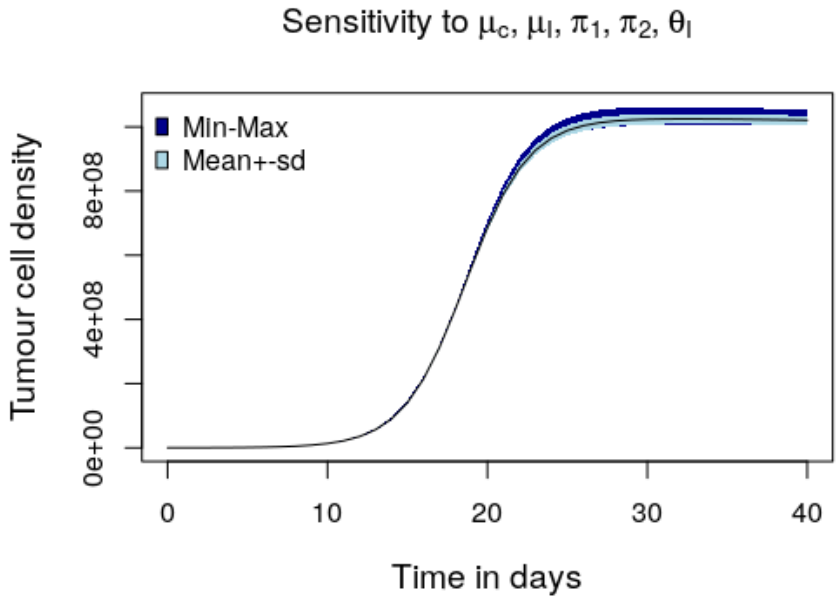


Figure 3.6: Global sensitivity analysis involved varying some the unobserved parameters π_1, π_2, μ_I, μ_c, θ_I by 40% from the baseline values in Table 3.3. These parameters had a negligible effect on tumour cell population.

Figure 3.6 indicates that the perturbations of the unobserved parameters made a negligible difference in the output of tumour cells. Parameter μ_c representing the natural death rate for tumour cells had more effect on the global sensitivity outcome compared to the other unobserved parameters.

To determine conditions necessary for tumour clearance, we varied the intrinsic growth rate α_c. This parameter can be varied experimentally by including treatment or converting the local tumour environment to be unfavourable to tumour progression, for example, by increasing tumour-suppressing cytokines (IL-2, IFN-γ) in the local tumour site. Figure 3.7 shows the local sensitivity analysis of α_c.

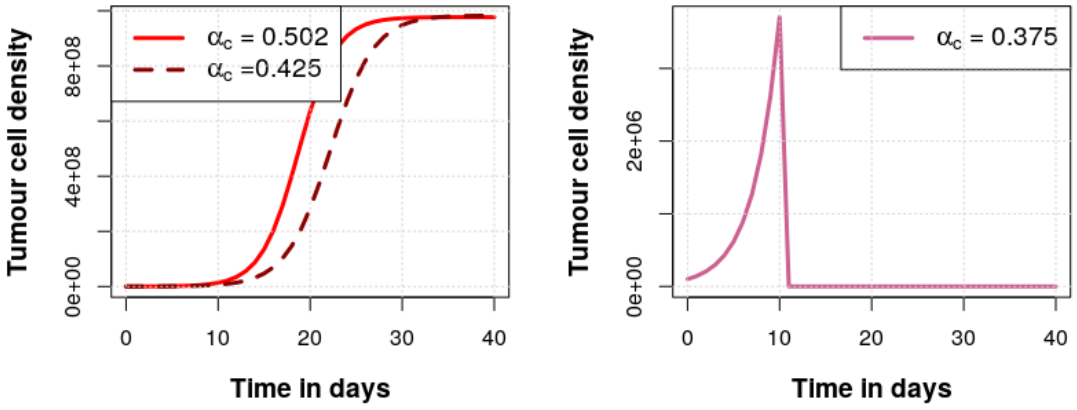


Figure 3.7: Local sensitivity analysis of the intrinsic growth rate α_c by decreasing it by 15% (Panel a) and 25% (Panel b) from its baseline value of 0.502.

The tumour intrinsic growth rate α_c was varied over a range of $15\% - 25\%$ from its baseline value of 0.502. Decreasing the intrinsic growth rate by 15% (0.425) resulted in a 60% decrease of the population of tumour cells by day 20 as shown in Panel (a) of Figure 3.7. Decreasing intrinsic growth rate by 25% (0.375) resulted in tumour clearance by day 11 as shown in Panel (b) of Figure 3.7.

From the numerical analysis, we observed that the quality and effectiveness of an immune response relies on the interactions between cytokines produced and the host immune cells [12]. The activation of an immune cell alters its gene expression. Since, we are entering an era for cell-based and gene-based therapies, it is important to study the

genes that regulate cytokine production. The next chapter covers analyses RNA-Seq data to identify DEGs, pathways and biological processes that can be used to explain breast cancer disease.

Chapter 4

Expression level based cytokine-regulator gene prediction

4.1 Introduction

Gene expression data sets can be analysed in two ways: a gene-based approach where genes play the role of features or sample-based approach where samples play the role of features. We used a gene-based approach to analyse RNA-Seq breast cancer data set. Since high RNA-Seq data can be very noisy, we applied the Miller's test to filter out genes with high variation. The significance analysis of micro-arrays (SAM) approach was applied to the filtered data set to predict differentially expressed genes (DEGs) across breast cancer (tumour) and healthy (tumour-free) samples. A selected gene set containing DEGs was used to predict enriched pathways and biological processes associated with breast cancer. In the next sections, we perform pre-processing and statistical analysis of the gene expression data.

4.2 Data retrieval and pre-processing

RNA-Seq techniques based on next generation sequencing (NGS) technologies are used for gene expression profiling. A sequencer uses image analysis on micro-array chips to measure individual gene expression at high resolution [40]. The sequencing machine generates a list of reads (millions of short strings) from random positions of each sample (input RNAs). Each read in the list can be computationally mapped on a reference genome, to a particular gene, to reveal a transcriptional map. A count is made for the

number of reads mapped to each particular gene, this represents its level of expression [39, 40].

Gene expression data sets can be obtained from The Cancer Genome Atlas (TCGA) (https://cancergenome.nih.gov/) project which contains harmonized genomic data sets for over 20 types of cancer. For the purpose of this study, analysis-ready breast cancer gene expression data set was downloaded from GDAC Firehose and FireBrowse (http://gdac.broadinstitute.org/). Gene expression data sets are represented by a large matrix with gene expression levels (count data) over a number of RNA-Seq experiments (samples) and usually, the genes measured are more compared to the number of samples. Suppose n experiments (samples) were performed on p regions of interest (genes). Let X_{ij} be the expression level of gene i in sample j under condition 1, Y_{ik} be the expression level of gene i in sample k under condition 2 and the response variable $z = \{1, 2\}$ represents the two conditions, that is the tumour and tumour-free samples, respectively. We preferred to work with the transposed matrix and denote the $n \times p$ matrix as $\mathbf{G}$, which can be formalized as a numerical vector

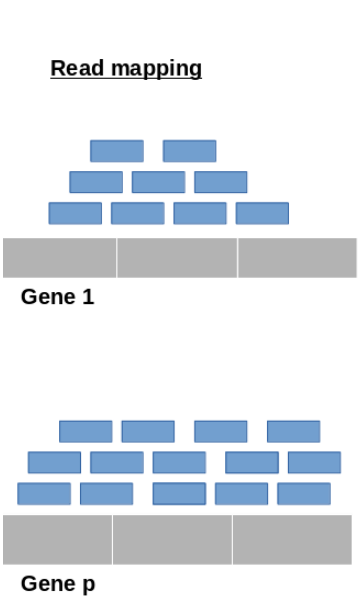
$$\mathbf{G} = \begin{bmatrix} X_{ij} \\ Y_{ij} \end{bmatrix}_{1 \leqslant i \leqslant p,\ 1 \leqslant j \leqslant n_1,\ 1 \leqslant k \leqslant n_2}$$

where n_1 is the tumour sample size, n_2 is the tumour-free sample. Figure 4.1 illustrates the untransposed $p \times n$ matrix.

RNA-Seq experiments result in large data sets containing a long list of genes and their respective expression levels. The downloaded breast cancer data set contained 20,531 genes from 1,100 tumour samples and 112 tumour-free samples. The data had been normalized using the upper-quartile (UQ) normalization approach given as

$$\text{Normalized count} = \frac{\text{raw count values} \times 100}{75^{th}\text{percentile of the column (gene) i}}.$$

The UQ approach removes the bias highly expressed genes have over lowly expressed genes without introducing additional noise. One could use the median for normalization but due to preponderance of low-count genes, that statistic will be biased and uninformative for DEGs thus, the UQ approach was preferred. Normalized RNA-Seq data are highly skewed and noisy, assuming that the gene expression data follows a log-normal distribution, it is therefore reasonable to transform the data to reduce the



	Tumour samples: Sample 1		Sample n_1	Tumour-free samples: Sample $n_1+1=m_1$	...	Sample $n_1+n_2=n$
Gene 1	X_{11}	...	X_{1n1}	Y_{1m1}	...	Y_{1n}
Gene 2	X_{21}	...	X_{2n1}	Y_{2m1}	...	Y_{2n}
...	...	...	...	...	...	...
Gene p	X_{p1}		X_{pn1}	Y_{pm1}		Y_{pn}

Figure 4.1: RNA-Seq experiments normally result in a large data set containing a long list of genes. The downloaded gene expression data set can be represented as a **G** matrix with p genes and n samples and their respective expression levels X_{ij} and Y_{ij} representing tumour and tumour-free samples, respectively.

variability across samples

$$\text{transformed data} = \log_2(\text{normalized count} + 1),$$

a value of 1 is added to the normalized counts to offset it from zero and to avoid missing values or large variances (Infs/-Infs). The transformed data for tumour samples X_{ij} and the tumour-free samples Y_{ik} have a geometric mean and standard deviation defined as

$$\bar{x}_i = \frac{\sum_j^{n_1} x_{ij}}{n_1} = \frac{\sum_j^{n_1} \log_2(X_{ij}+1)}{n_1} \quad \text{and} \quad s_1 = \sqrt{\frac{\sum_j^{n_1}(x_{ij}-\bar{x}_i)^2}{n_1-1}} \quad \text{for tumour samples and}$$

$$\bar{y}_i = \frac{\sum_k^{n_2} y_{ik}}{n_2} = \frac{\sum_k^{n_2} \log_2(Y_{ik}+1)}{n_2} \quad \text{and} \quad s_2 = \sqrt{\frac{\sum_k^{n_2}(y_{ik}-\bar{y}_i)^2}{n_2-1}} \quad \text{for tumour-free samples.}$$

Geometric mean is approximately equal to the median and as an estimate of the population mean, it is less likely to be affected by extremely large values compared to the arithmetic mean. P-values obtained in logarithmic scale remain valid in original scale [80]. Since RNA-Seq experiments involve a multi-step process, there is a potential of biological, experimental/technical and residual variation [40]. It is therefore necessary to filter

genes with high variations, that is genes with low counts before performing the actual statistical analysis. Filtering genes with high variation will reduce the dimensionality of the data set given that the data set has more genes/features compared to sample size.

Normally distributed data sets have a mean μ and variance σ^2. The standard deviation is not an appropriate measure of dispersion in this study, because the gene expression mean values across the samples are not identical [81]. The coefficient of variation (CV), a unit-free measure of dispersion [82], was used for quantifying heterogeneity of gene expression levels across the two samples. The population CV defined as $\gamma = \sigma/\mu$ measures the variability of a population relative to its mean and standard deviation. When the distribution is unknown, the CV can be estimated from observed gene expression data as $\hat{\gamma} = s_i/|\bar{x}_i|$ and $\hat{\gamma} = s_i/|\bar{y}_i|$ for tumour and tumour-free samples, respectively. To investigate whether a gene can be considered for further analyses (gene with $\hat{\gamma} < 1$), we set the following hypothesis

$$H_0 : \hat{\gamma} < 1 \quad \text{versus} \quad H_A : \hat{\gamma} > 1. \tag{4.2.1}$$

Statistical tests such as Wald, Miller, or likelihood ratio are used to test the significance of the differences and verify that the differences are not by random chance. We used Miller's test which has a Gaussian distribution $Z \sim N(0,1)$ with a mean equal to 0 and variance equal to 1 as described in Equation (4.2.2).

$$MiL = \frac{\hat{\gamma} - \gamma_0}{s_m} \sim Z(0,1) \quad \text{where} \quad s_m = \sqrt{\frac{\hat{\gamma}^4 - 0.5\hat{\gamma}^2}{n}} \quad \text{and} \quad \gamma_0 = 1. \tag{4.2.2}$$

The CV of each gene in the data set was computed, and genes across the samples with high variation that were less likely to be interesting were removed at a level of significance $\alpha < 0.05$. Removing genes with high variation reduced the dimension of the data set by approximately 19%. Figure 4.2 illustrates the influence of lowly expressed genes.

The sharp curve in panel (a) of Figure 4.2 indicates the presence of genes with low counts that should be filtered out before performing the actual statistical analysis. In the next section we predict DEGs by applying the significance analysis of micro-arrays (SAM) approach.

4.3 Statistical analysis of breast cancer gene expression data

It is expected that most genes across the tumour and tumour-free samples are normally expressed and only a few are differentially expressed. A gene is identified as differentially expressed if its expression level of transcripts (counts) between two or more

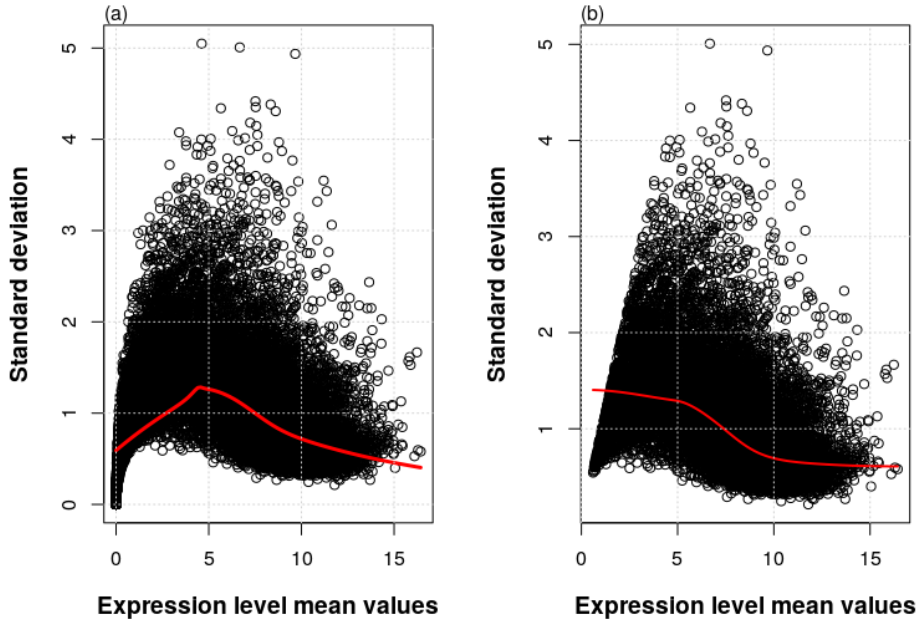


Figure 4.2: Panel (a) and (b) represents standard deviation against expression level mean value plots before and after Miller's test, respectively. The sharp curve (red line) in Panel (a) indicates the presence of genes with low counts.

conditions is different. Statistical methods used to detect DEGs can be grouped into two (i) parametric methods used when the distribution is known, such as t-test and its variations, empirical Bayes methods, and (ii) non-parametric methods used when the distribution is unknown, such as Mann-Whitney test, Wilcoxon signed-rank test, Kruskal-Wallis test, and non-empirical Bayes method. Useful approaches have been proposed for individual-based gene analyses, both Seyednasrollah et al. [83] and Rapaport et al. [41] have compared various statistical algorithms and they noted that array-based methods adapted to RNA-Seq data performed well compared to the methods developed for analysing RNA-Seq data. A non-parametric SAM-permutation based approach was found to detect a relatively large number of DEGs Chu et al. [84]

4.3.1 The SAM approach

For each gene, we conducted a SAM permutation-based approach to identify DEGs. Each gene is considered independently when performing the statistical analysis, the test is used to detect the significance of any observed differences between the two samples.

We used the following steps to compute the SAM d_i statistic as proposed by Tibshirani et al. [85].

1. For each gene $i = 1, 2, \ldots, p$ the following statistic was computed

$$d_i = \frac{r(i)}{s(i) + s_0}, \quad \text{with} \quad r(i) = \bar{x}_i - \bar{y}_i, \;\; s(i) = s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \quad \text{and}$$

$$s_p = \sqrt{\frac{\sum_{k}^{n_2}(x_{ij} - \bar{x}_i)^2 + \sum_{j}^{n_1}(y_{ik} - \bar{y}_i)^2}{n_1 + n_2 - 2}}, \tag{4.3.1}$$

   where x_{ij} is the expression level for gene i in tumour sample j, y_{ik} is the expression level for gene i in tumour-free sample k, $\bar{x}_i$ is the expression level mean for gene i for tumour samples and $\bar{y}_i$ is the expression level mean for gene i for tumour-free samples, $s(i)$ is the gene-specifier scatter, s_p is a pooled standard deviation, and the parameter s_0 is a non-negative exchangeability factor.

2. Order the d_i statistic based on the statistical significance of the observed differences between the two groups such that $d_{(1)} \leqslant d_{(2)} \leqslant \ldots \leqslant d_{(p)}$.

3. Perform 10,000 sets of permutations on the sample labels (tumour and tumour-free). For each permutation, re-calculate steps (i)-(ii). The estimated expected null score for each permutation b was given as

$$\text{Expected } d_{(i)} = \bar{d}_{(i)} = \frac{\sum_{b=1}^{B} d^b_{(i)}}{B}, \quad \text{where} \quad B = 10,000.$$

4. Plot the ordered observed d_i scores against the expected d_i null scores.

5. Pick a threshold/cut-off Δ to identify genes expressed differently across the two conditions. A gene is significant if

$$\left| d_i - \bar{d}_i \right| > \Delta,$$

   where d_i is the observed null score and $\bar{d}_i$ is the expected null score calculated from step (iii).

4.3.2 Identification of differentially expressed genes

The exchangeability factor s_0 was used to minimize the coefficient of variation of d_i. Parameter s_0 was estimated from pooling expression data of all the genes. Using an s_0 equal to 0.0208, the observed d_i and expected d_i null score for each gene i was computed. Our interest is to find significant genes that are differently expressed between the two samples. We selected DEGs genes with a p-value < 0.05 using both the SAM T-statistic and the SAM permutation-based statistic. These genes were above a fixed threshold $\Delta = 8.1538$, the upper quartile of $|d_i - \bar{d}_i|$. The p-values obtained gives confidence that the same association will be found for a new sample with that particular phenotype or expression profile. The DEGs (in pink) and non-DEGs (in black) can be visualized as shown in Figure 4.3.

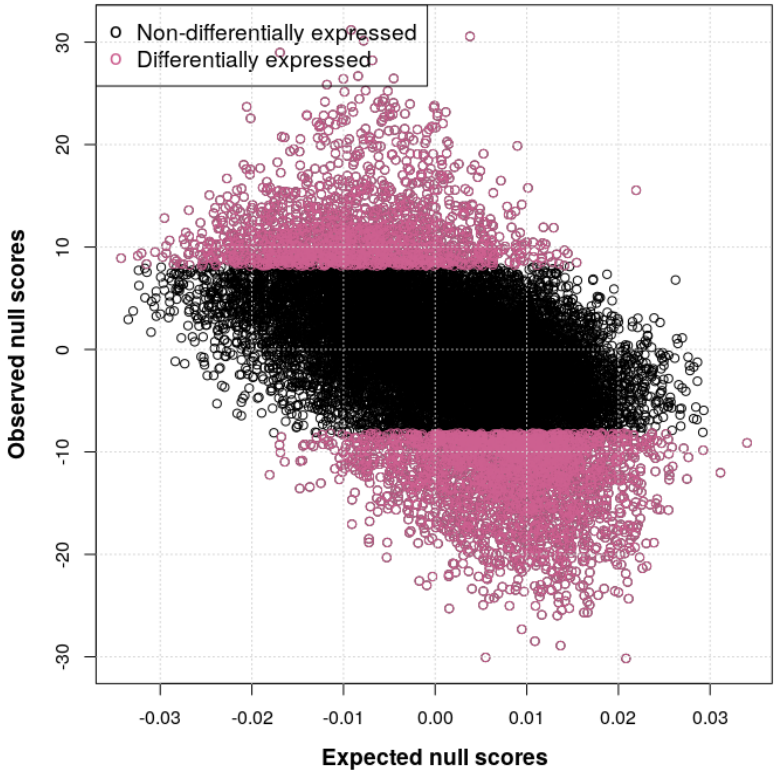


Figure 4.3: The SAM approach was used to detect 4,159 differentially expressed genes (DEGs). The plot shows non-DEGs (in black circles) and DEGs (in pink circles).

A total of 4,159 genes were identified as differentially expressed. These genes constitute about 20% of the original data set. Out of the 4,159 genes, 2,816 genes were over-expressed and 1,343 genes were under-expressed in tumour samples whereas 3,022

genes were over-expressed and 1,137 genes were under-expressed among tumour-free samples. Figure 4.4 shows a density distribution and box plot to examine any differences in the 4,159 DEGs across tumour and tumour-free samples

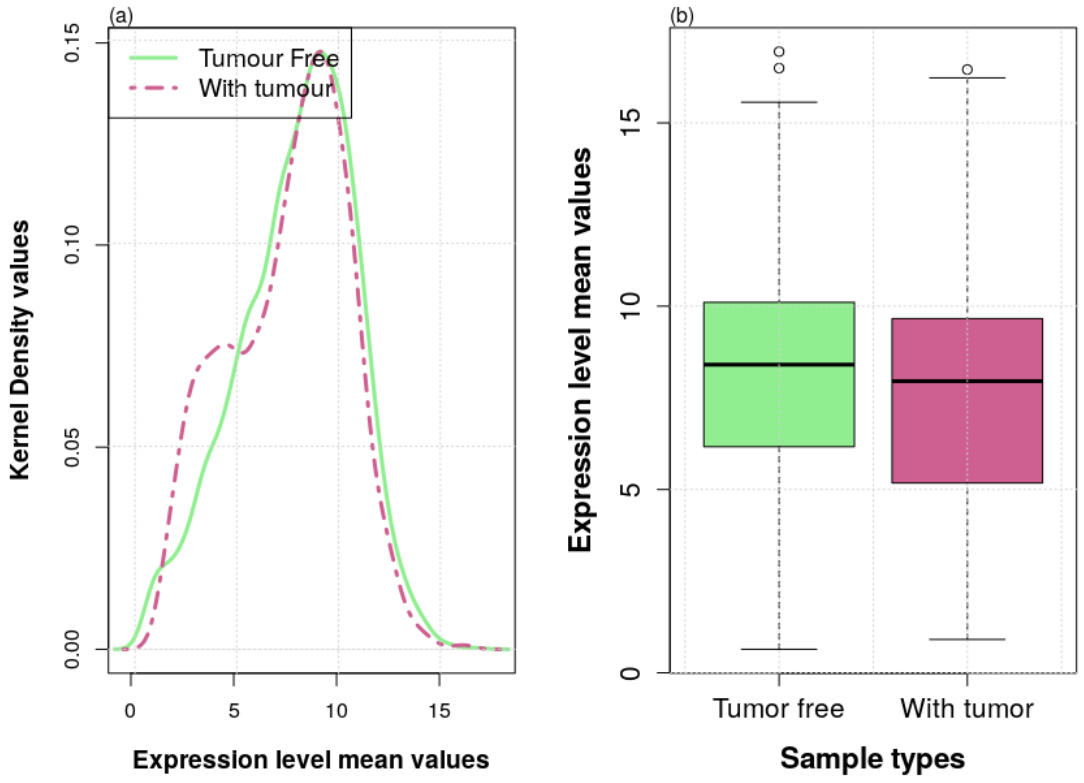


Figure 4.4: Slight differences in the 4,159 differentially expressed genes across samples that are with tumour (WT) and tumour-free(TF).

There is a very slight difference in the expression level means of the 4,159 DEGs across the two samples. Some of the DEGs in Figure 4.4 can be used to identify bio-marker genes that explain phenotypic differences between two conditions. Our interest is to find genes that are common and expressed differently between the two samples that are either over-expressed or under-expressed. Therefore, it will be reasonable to reduce the gene set of interest to a small number of genes that can be used as marker genes for breast cancer detection.

4.3.2.1 Identification of breast cancer marker genes

The ability to successfully distinguish between samples using gene expression data is an important aspect of bio-marker discovery. Genes that are either active or inactive and expressed differently across breast cancer and tumour-free samples can be used to characterise the disease. A re-sampling method was used to identify odd-shared common DEGs across tumour and tumour-free samples. A total of 298 over-expressed genes in tumour samples were under-expressed in tumour-free samples and 504 under-expressed genes in tumour samples were over-expressed in tumour-free samples. Since we are interested in keeping genes that might be of interest as small as possible, we selected 40 genes which might explain the phenotypic differences between tumour and tumour-free samples. Table 4.1 and 4.2 lists the gene set of interest with description terms obtained from the UniProt knowledgebase database (http://www.uniprot.org).

There is a clear difference in expression level means across tumour and tumour-free samples as shown in Table 4.1 and Table 4.2. The expression level mean values for the over-expressed genes in tumour samples were higher compared to the under-expressed genes in tumour-free samples. The expression level mean values for the under-expressed genes in tumour samples were lower compared to the over-expressed genes in tumour-free samples. We used a density distribution and box plot to further examine the differences in the selected gene set as shown in Figure 4.5.

There is a difference in the expression level mean values across tumour and tumour-free samples for the selected gene set in Table 4.1 and 4.2. WTOE genes have a higher expression level mean compared to TFUE, WTUE and TFOE. Although, random genes can be identified as differentially expressed, we can expect that some of the selected genes in Table 4.1 and 4.2 are associated with the phenotypic differences between the two samples. Furthermore, the hierarchical clustering plot (heatmap) in Figure 4.6 gives a visualization of the patterns in the selected list of DEGs across the two samples.

The heatmap in Figure 4.6 supports the suspicion that two clusters exist. The over-expressed genes (in pink row names) clustered into two groups, with the tumour samples having higher expression levels compared to tumour-free samples. The under-expressed genes (in black row names) clustered into two groups, with tumour-free samples having higher expression levels compared to tumour samples. It is possible that some of the genes in Table 4.1 and 4.2 can be used as marker genes for breast cancer detection, classification and targets for treatment. One of the main goals of analysing

Table 4.1: Top 20 over-expressed genes in tumour (WT) samples that were under-expressed in tumour-free (TF) samples.

	Protein	Mean		95% Confidence Interval	
ID	name	WT	TF	WT	TF
ANLN	Anillin	9.32	5.99	9.2447–9.4058	5.75901–6.216
ASF1B	Histone chaperone ASF1B	8.87	5.84	8.80477–8.92742	5.64039–6.03701
ASPM	Abnormal spindle-like microcephaly-associated protein	8.86	5.16	8.77988–8.94681	4.88994–5.43269
C19orf21	Mitotic interactor and substrate of PLK1	8.84	5.62	8.71199–8.97698	5.26192–6.0021
CDC20	Cell division cycle protein 20 homolog	9.04	5.58	8.9544–9.12285	5.33165–5.83411
COL10A1	Collagen alpha-1(X) chain	10.56	3.78	10.43949–10.68534	3.53817–4.01912
COL11A1	Collagen alpha-1(XI) chain	10.62	4.77	10.46703–10.76848	4.4818–5.05414
COMP	Cartilage oligomeric matrix protein	10.35	5.57	10.22631–10.48075	5.18387–5.97755
DTL	Denticleless protein homolog	8.93	5.85	8.86027–8.9939	5.624–6.06788
EEF1A2	Elongation factor 1-alpha 2	8.94	6.30	8.75738–9.13237	5.86191–6.74585
FOXM1	Forkhead box protein M1	9.60	6.15	9.52051–9.68584	5.92631–6.37061
GJB2	Gap junction beta-2 protein	8.90	5.15	8.7825–9.01485	4.92367–5.36686
IQGAP3	Ras GTPase-activating-like protein IQGAP3	9.12	5.10	9.04288–9.19718	4.84876–5.35248
KIF23	Kinesin-like protein KIF23	8.82	6.18	8.75628–8.88539	5.98043–6.37748
KIFC1	Kinesin-like protein KIFC1	9.05	5.93	8.98354–9.12346	5.71193–6.15298
MYBL2	Myb-related protein B	9.36	5.62	9.25767–9.46516	5.34726–5.88712
NUSAP1	Nucleolar and spindle-associated protein 1	9.53	6.40	9.46109–9.59094	6.18072–6.61277
PLK1	Serine/threonine-protein kinase PLK1	8.82	5.32	8.73789–8.89456	5.11399–5.51372
RRM2	Ribonucleoside-diphosphate reductase subunit M2	9.68	6.14	9.60533–9.76053	5.86555–6.4105
UBE2C	Ubiquitin-conjugating enzyme E2 C	8.99	4.77	8.89906–9.07478	4.49894–5.04416

Table 4.2: Top 20 under-expressed genes in tumour (WT) samples that were over-expressed in tumour-free (TF) samples.

	Protein	Mean		95% Confidence Interval	
ID	name	WT	TF	WT	TF
ACADL	Long-chain specific acyl-CoA dehydrogenase, mitochondrial	2.77	6.96	2.64042–2.89347	6.63677–7.27304
AQPEP	Aminopeptidase Q	2.65	7.12	2.5608–2.73286	6.68181–7.55625
CLDN19	Claudin-19	2.95	6.79	2.82596–3.07339	6.42963–7.1861
DEFB132	Beta-defensin 132	2.93	6.56	2.81696–3.03227	6.08507–7.0255
FAM196B	Protein FAM196B	2.92	7.02	2.80996–3.01877	6.64757–7.40501
FGFBP2	Fibroblast growth factor-binding protein 2	2.94	6.55	2.82017–3.04667	6.17551-6.91063
FIGF	Vascular endothelial growth factor D	3.15	9.19	3.03902–3.25991	9.00798–9.37411
GRIA4	Glutamate receptor 4	3.01	7.21	2.8798–3.13604	6.80405–7.63982
HIF3A	Hypoxia-inducible factor 3-alpha	2.99	7.88	2.88013–3.10529	7.67081–8.07789
HSD17B13	17-beta-hydroxysteroid dehydrogenase 13	1.89	6.54	1.81435–1.96978	6.1628–6.91454
KCNB1	Potassium voltage-gated channel subfamily B member1	2.95	6.70	2.84455–3.05093	6.46077–6.94607
KY	Kyphoscoliosis peptidase	2.81	6.71	2.70017–2.90674	6.37727–7.06032
LRRC2	Leucine-rich repeat-containing protein 2	3.17	6.50	3.09608–3.25162	6.30511–6.68414
MASP1	Mannan-binding lectin serine protease 1	2.64	6.71	2.53572–2.7319	6.45511–6.97115
NRG2	Pro-neuregulin-2, membrane-bound isoform	2.96	6.52	2.84606–3.0735	6.34775–6.68827
PFKFB1	6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase 1	3.21	6.71	3.11716–3.30299	6.32161–7.10695
SCN3A	Sodium channel protein type 3 subunit alpha	3.21	6.96	3.11047–3.30684	6.77919–7.15352
SLC7A10	Asc-type amino acid transporter 1	2.80	7.25	2.6886–2.92001	6.76915–7.73683
TNMD	Tenomodulin	2.55	7.15	2.42869–2.66261	6.83969–7.46191
TSLP	Thymic stromal lymphopoietin	2.72	6.52	2.63166–2.80915	6.41222–6.62226

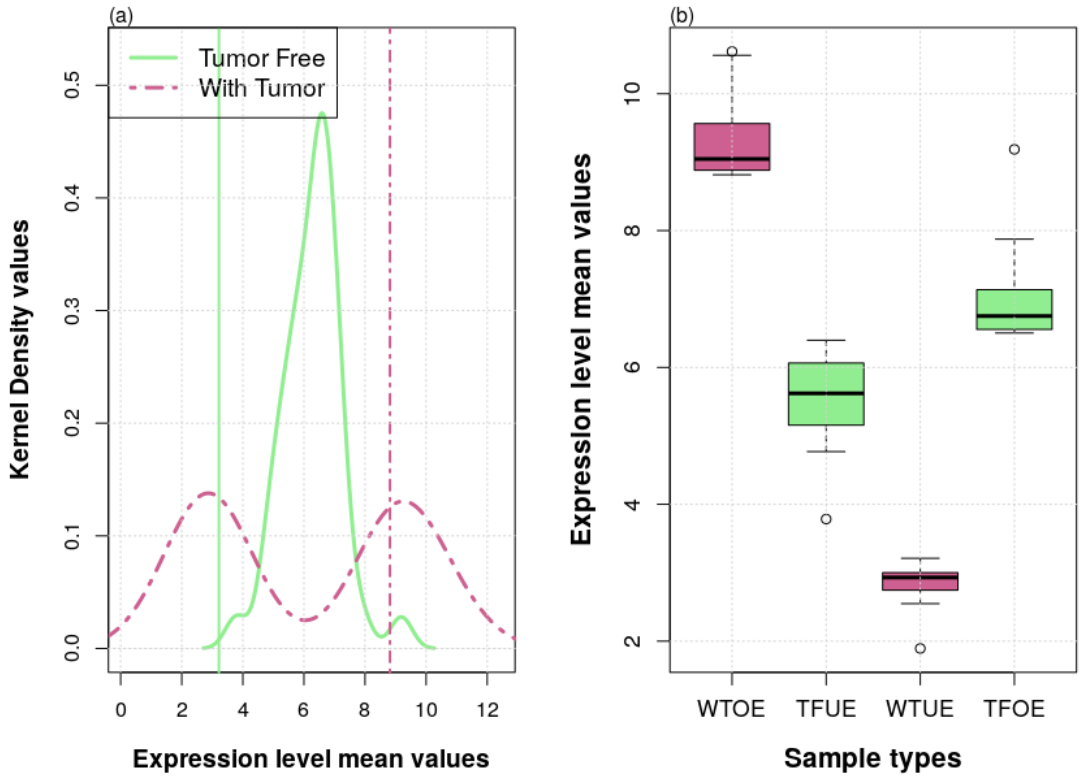


Figure 4.5: The gene set in Table 4.1 and Table 4.2 had over-expressed genes in tumour samples (WTOE) that were under-expressed in tumour-free samples (TFUE) and under-expressed genes in tumour samples (WTUE) that were over-expressed in tumour-free samples (WTUE). Over-expressed genes in tumour samples have a higher expression level mean compared to the other categories.

gene expression data is to determine pathways and biological processes of interesting gene sets associated with a disease outcome. Instead of focussing on individual genes, we performed enrichment and pathway analysis on the selected gene sets in Table 4.1 and Table 4.2 in the next sections.

4.4 Gene set enrichment and pathway analysis

Annotation databases like Gene Ontology (GO) [43], Gene Ontology Annotation [44] (GOA) and Kyoto Encyclopedia of Genes and Genomes (KEGG) [45] are used to interpret interesting DEGs. GO contains evidence-based annotations and biological de-

Figure 4.6: The heat map indicates the presence of two distinct clusters across the tumour and tumour-free samples in the selected gene sets.

scriptions of individual gene products (proteins) [43]. The domains of GO include biological processes, molecular functions or cellular components [43]. In this study, we focus on predicting enriched biological processes of the selected gene set. KEGG is a knowledge-base that maps a set of DEGs with biological functional information or interacting molecules in a cells [45].

Enrichment and pathway analysis uses statistical methods to determine enriched biological processes and significant pathways associated with the outcome of a disease [42]. Hyper-geometric, Fisher's exact, χ^2 test and Jaccard index are used to perform enrichment and pathway analysis. In this study, a hyper-geometric test was used because it appropriately models the probability that a certain annotation term occurs i times in the list of DEGs. The p-values obtained were corrected using the Bonferroni method. The adjusted p-values was used to determine the probability of observing at least i number

of genes out of the total gene list annotated to a particular term, given the proportion of genes in the whole genome that are annotated to that term. A pathway or biological process was said to be enriched at a significance level 0.05.

4.4.1 Enriched biological processes associated with breast cancer

Several GO similarity measures have been proposed. We applied a topology-based similarity measure proposed by Mazandu and Mulder [86] to predict biological processes over-presented in the selected gene set mapped from 17,292 GO human biological processes. Level is a term in the GO directed acyclic graph. Unfortunately, the under-expressed genes in tumour samples in Table 4.2 that were over-expressed in tumour-free samples had no enriched biological processes. The over-expressed genes in tumour samples that were under-expressed in tumour-free samples in Table Table 4.1 were associated with the enriched biological processes in Table 4.3.

Table 4.3: Enriched biological processes over-represented in tumour samples.

GO ID	GO names	Level	p-value	Adjusted p-value
GO:0000281	mitotic cytokinesis	7	0.0	0.0
GO:0007076	mitotic chromosome condensation	8	0.0	0.0013
GO:0051437	positive regulation of ubiquitin-protein ligase activity involved in regulation of mitotic cell cycle transition	13	0.0002	0.0209
GO:0042787	protein ubiquitination involved in ubiquitin-dependent protein catabolic process	10	0.0003	0.0423

The enriched BPs are associated with cell division, differentiation and cell cycle progression. GO:0000281 (mitotic cytokinesis), is the physical separation of two daughter cells during cell division, and CDC20 which is over-expressed in tumour samples plays a central role in this biological process [87]. Studies have indicated that the failure to complete cytokinesis can promote tumorigenesis and chromosomal/genome instability [87]. GO:0007076 (mitotic chromosome condensation), contributes to the pathogenesis of cancer and is a necessary process that prevents chromosomes from being entangled during chromosome segregation [88]. Conjugation of ubiquitin to a substrate is performed by a three-step process; ubiquitin activating enzyme (E_1), ubiquitin conjugating enzyme (E_2),

and ubiquiting ligase (E_3) [45, 89]. Ubiquitin system is responsible for processing MHC I antigens and altered function/mutation of ubiquitin-proteosome especially E_3 process (ubiquitin ligase) can directly or indirectly lead to the onset of cancer [89, 90]. Known genes involved in ubiquitin pathway include UBE2C for E2 process, BRCA1 and p53 for E3 process [90]. p53 is a known tumour suppressor protein often found to be mutated or depleted in breast cancer and is responsible for chromosomal instability [87, 91].

Hierarchical clustering was used to visualize the patterns of the biological processes associated with breast cancer genes in Table 4.1 and 4.2. The hierarchical structure of biological processes of the selected gene list is shown in Figure 4.7.

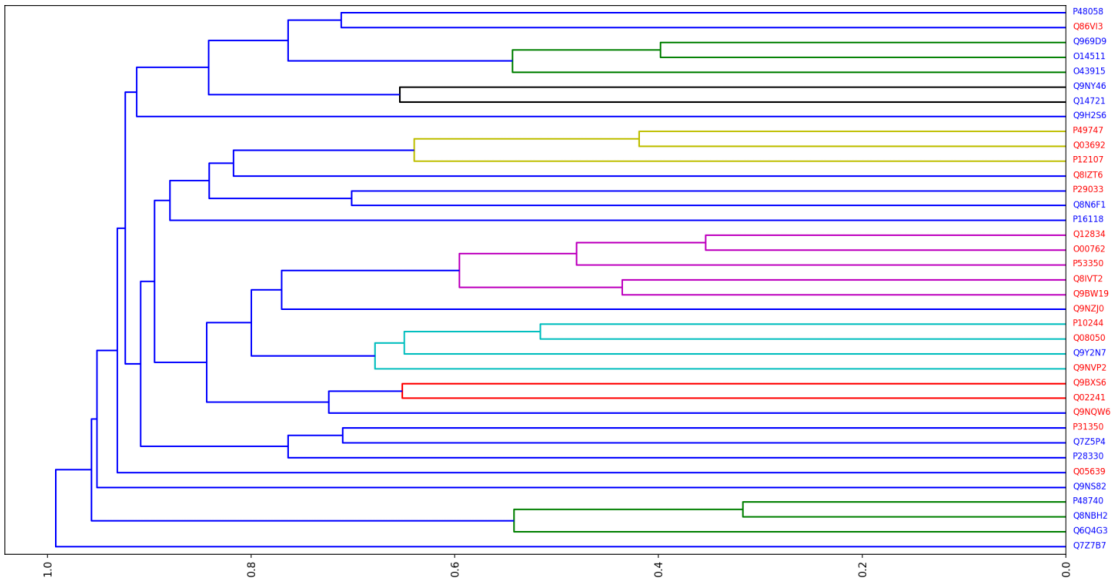


Figure 4.7: The biological processes in red represent the over-expressed genes in tumour samples and the ones in blue represent under-expressed genes in tumour samples.

Most biological processes of the over-expressed genes in tumour samples that were under-expressed in tumour-free samples in Table 4.1 clustered together (in red) and biological processes of the under-expressed genes in tumour samples that were over-expressed in tumour-free samples in Table 4.2 clustered together (in blue), except for a few. The diagram shows groups of biological processes in breast cancer that might be working together to promote its development and progression. In the next section, we predicted pathways associated with the DEG set.

4.4.2 Significant pathways associated with breast cancer

A pathway is a series of biological events occurring sequentially to maintain or change cell environment for its stability. Pathways with a Bonferroni-adjusted p-value < 0.05 were selected as significant. The closer the p-value is to zero, the less likely it is for the observed annotation to a group of genes to occur by random chance. Table 4.4 lists significant KEGG pathways associated with the genes in Table 4.1 which were over-expressed genes in tumour samples and under-expressed in tumour-free samples.

Table 4.4: KEGG pathways over-represented in tumour samples.

Pathway ID	Pathway name	Protein name	P-value	Adjusted p-value
hsa04218	Cellular senescence	Forkhead box protein M1 (FOXM1), Myb-related protein B (MYBL2)	0.0	0.0001
hsa04974	Protein digestion and absorption	Collagen alpha-1 (X) and (XI) chain (COL10A1 and COL11A1)	0.0007	0.0163

Cellular senescence and protein digestion and absorption are the most significant KEGG pathways associated with the over-expressed genes in breast cancer patients. Normally, cell replication occurs till death however, some cells may stop multiplying and reach a terminal cell proliferation arrest which can be triggered by an uncontrolled multiple cell division, un-repaired DNA damage or cellular stresses, such as oncogenes [92]. These naturally occurring senescent cells that accumulate with age can create an environment that encourages tumour growth. Studies have shown that senescence cells in human epidermal growth factor receptor 2 (HER2)-positive breast cancer cell lines, secrete IL-6, IL-8 and other cytokines that promote tumour development and proliferation [92]. FOXM1 and MYBL2 involved in the cellular senescence pathway are proto-oncogenes that code for proteins which regulate cell growth, differentiation and cell cycle progression. MYBL2 and FOXM1 can be used as a marker for cellular senescence and COL10A1, and COL11A1 for protein digestion and absorption.

Table 4.5 lists significant KEGG pathways associated with genes in Table 4.2 which were under-expressed genes in tumour samples and over-expressed in tumour-free samples.

Table 4.5: KEGG pathways over-represented in tumour-free samples.

Pathway ID	Pathway name	Protein name	P-value	Adjusted p-value
hsa05150	Staphylococcus aureus infection	Mannan-binding lectin serine protease (MASP1)	0.0	0.0002
hsa04610	Complement and coagulation cascades	Mannan-binding lectin serine protease (MASP1)	0.0	0.0006
hsa05033	Nicotine addiction	Glutamate receptor (GRIA4)	0.0016	0.0476

Staphylococcus aureus infection, nicotine addiction, and complement and coagulation cascades were the most significant pathways associated with under-expressed genes in tumour samples which were over-expressed in tumour-free samples. The compliment system is an immune surveillance system and a mediator of the innate and adaptive immunity, and the coagulation cascade is responsible for ensuring homoeostatic maintenance [45, 68, 93]. The experiment by Bianco et al. [93] found that the rapid proliferation of macrophages can be induced by the activation of the complement system. Rapid proliferation of macrophages might lead to adaptive immunity (T cell responses) activation and tumour clearance, this explains why the protein MASP1 that regulates the complement system [68] is under-expressed in breast cancer patients and over-expressed in tumour-free samples. Bacteria products, such as staphylococcus aureus infection activate the complement and coagulation cascades [94]. Staphylococcus bacteria is a common cause for breast abscesses (an accumulation of pus and inflammation in breast tissue) and is associated with development of rare kinds of breast cancer [45]. The bacteria can produce proteins that inhibit complement system activation, and it can express super-antigens that prevent a normal immune response [45].

The odd-shared DEGs across tumour and tumour-free samples share pathways including metabolic (hsa01100), cytokine-cytokine receptor interaction (hsa04060), focal adhesion (hsa04510) and phosphatidylinositol 3' -kinase(PI3K)-Akt signalling (hsa04151). Cytokine-cytokine receptor interaction pathway stimulate responses through binding to specific receptors. Proteins, such as Vascular endothelial growth factor D (FIGF) and Thymic stromal lymphopoietin (TSLP) were found to be involved in the cytokine-cytokine receptor pathway. Cancer cells require high metabolism, and pathways regulating cell metabolism are often reprogrammed to sustain the rapid cell proliferation of cancer cells [9].

Chapter 5

Discussion and conclusion

In our study, we developed a new non-linear ODE model to predict the role of cytokines in tumour-immune cell dynamics. The numerical analysis in Chapter 3 emphasized the role of cytokines in activating an immune response, and promoting tumour progression [12]. From the numerical analysis, TGF-β predominantly secreted by tumour cells and Tregs was found to be crucial for tumour proliferation and progression. The inhibiting effect of TGF-β, limited the effectiveness of type 1 immune response. The absence of TGF-β resulted in the immune cells growing without bound and this can result in a series of serious dysfunctional organs [14]. We observed that tumour regression occurred when (i) type 1 immune response was not inhibited by TGF-β and (ii) the intrinsic growth rate was decreased by 25% from its baseline value 0.502. The intrinsic growth rate of tumour cells can be decreased by use of cancer therapies or changing the local TME to favour pro-inflammatory cytokines and type 1 immune signals. The persistence of type 2 immune response after type 1 signals were depleted on day 10 promoted tumour progression, because of the rich anti-inflammatory cytokines available at the tumour site. Our work complements the role Th2 and Th17 cells play in inhibiting the proliferation of M1 macrophages, CTLs, NK and Th1 cells.

For an immune cell to be activated from a resting state, its gene expression has to be altered. Thus we downloaded and analysed breast cancer RNA-Seq data from The cancer Genome Atlas project. The goal for analysing gene expression data is to predict DEGs, pathways and biological processes that can explain phenotypic differences across samples. Significance analysis of micro-arrays (SAM) approach was used to predict DEGs across breast cancer and tumour-free samples. We predicted and selected 40 genes that were differentially expressed and common between the two samples that might explain

the phenotypic differences. These 40 genes were either over-expressed in tumour samples, but under-expressed in tumour-free samples or under-expressed in tumour samples, but over-expressed in tumour-free samples.

A hyper-geometric test was used to predict enriched biological processes and significant pathways of the selected gene set. Enriched biological processes, such as ubiquitin system, mitotic cytokinesis and chromosome condensation were over-represented in the genes that were over-expressed in tumour samples. These enriched processes can be used for future research of breast cancer progression and regulation of APCs. The over-expressed genes in tumour samples were associated with pathways including cellular senescence and protein digestion and absorption. These pathways can be used for future research of breast cancer growth and proliferation. DEGs including MYBL2 and FOXM1 can be used as markers for cellular senescence pathway while COL10A1 and COL11A1 can be used as markers for protein digestion and absorption pathway. The over-expressed genes in tumour-free samples were associated with pathways, such as staphylococcus aureus infection and complement and coagulation cascades. These pathways are involved in activating a normal immune response.

In future we will, (i) study the effect of TGF-β without the logistic proliferation terms and with treatment, (ii) investigate the influence of tumour cells on dendritic cells that favours the differentiation of M2 macrophages at the expense of M1 macrophages, (iii) predict the overall survival rates of breast cancer patients with the identified enriched pathways and biological processes and (iv) predict the suitability of the identified over-expressed DEGs in tumour samples, enriched biological processes and significant pathways as target for breast cancer diagnosis, prognosis and treatment.

List of references

[1] C.P. Wild (Eds.) B.W. Stewart. *World Cancer Report 2014*. IARC Nonserial Publication. World Health Organization, 1 edition, 2014.

[2] S. McGuire. World cancer report 2014. Geneva, switzerland: World health organization, international agency for research on cancer, who press, 2015. *Advances in Nutrition: An International Review Journal*, 7(2):418–419, 2016.

[3] B. O. Anderson, C.-H. Yip, R. A Smith, R. Shyyan, S. F. Sener, A. Eniu, R. W. Carlson, E. Azavedo, and J. Harford. Guideline implementation for breast healthcare in low-income and middle-income countries. *Cancer*, 113(S8):2221–2243, 2008.

[4] P. Vineis and C. P. Wild. Global cancer patterns: causes and prevention. *The Lancet*, 383(9916):549–557, 2014.

[5] G. Danaei, S. V. H., A. D. Lopez, C. J. Murray, M. Ezzati, Comparative Risk Assessment collaborating group (Cancers, et al. Causes of cancer in the world: comparative risk assessment of nine behavioural and environmental risk factors. *The Lancet*, 366(9499):1784–1793, 2005.

[6] S. Koscielny, M. Tubiana, M. Le, A.J. Valleron, H. Mouriesse, G. Contesso, and D. Sarrazin. Breast cancer: relationship between the size of the primary tumour and the probability of metastatic dissemination. *British journal of cancer*, 49(6):709, 1984.

[7] C. L. Carter, C. Allen, and D. E. Henson. Relation of tumor size, lymph node status, and survival in 24,740 breast cancer cases. *Cancer*, 63(1):181–187, 1989.

[8] L. M. Coussens and Z. Werb. Inflammation and cancer. *Nature*, 420(6917):860–867, 2002.

[9] D. Hanahan and R. A. Weinberg. The hallmarks of cancer. *cell*, 100(1):57–70, 2000.

[10] N. Bellomo and L. Preziosi. Modelling and mathematical problems related to tumor evolution and its interaction with the immune system. *Mathematical and Computer Modelling*, 32(3-4):413–452, 2000.

[11] A. G. Knudson. Mutation and cancer: statistical study of retinoblastoma. *Proceedings of the National Academy of Sciences*, 68(4):820–823, 1971.

[12] L. M. Sompayrac. *How the immune system works*. John Wiley & Sons, 2015.

[13] K. E. De Visser, A. Eichten, and L. M. Coussens. Paradoxical roles of the immune system during cancer development. *Nature reviews cancer*, 6(1):24–37, 2006.

[14] J. Mumm and M. Oft. Cytokine-based transformation of immune surveillance into tumor-promoting inflammation. *Oncogene*, 27(45):5913–5919, 2008.

[15] K. N. Heller, C. Gurer, and C. Münz. Virus-specific CD4+ T cells: ready for direct attack. *Journal of Experimental Medicine*, 203(4):805–808, 2006.

[16] A. Yates and R. Callard. Cell death and the maintenance of immunological memory. *Discrete and continuous dynamical systems series series b*, 1(1):43–60, 2001.

[17] P. Muranski and N. P. Restifo. Adoptive immunotherapy of cancer using CD4+ T cells. *Current opinion in immunology*, 21(2):200–208, 2009.

[18] M. A. Cooper, T. A. Fehniger, A. Fuchs, M. Colonna, and M. A. Caligiuri. Nk cell and DC interactions. *Trends in immunology*, 25(1):47–52, 2004.

[19] G. Dranoff. Cytokines in cancer pathogenesis and cancer therapy. *Nature Reviews Cancer*, 4(1):11–22, 2004.

[20] M. Kolev, E. Kozłowska, and M. Lachowicz. A mathematical model for single cell cancer-immune system dynamics. *Mathematical and computer modelling*, 41(10): 1083–1095, 2005.

[21] J. B. Swann and M. J. Smyth. Immune surveillance of tumors. *The Journal of clinical investigation*, 117(5):1137–1146, 2007.

[22] J.L. Banyer, N.H. Hamilton, I.A. Ramshaw, and A.J. Ramsay. Cytokines in innate and adaptive immunity. *Reviews in immunogenetics*, 2(3):359–373, 1999.

[23] G. S. Getz. Bridging the innate and adaptive immune systems, 2005.

[24] B. Pulendran, K. Palucka, and J. Banchereau. Sensing pathogens and tuning immune responses. *Science*, 293(5528):253–256, 2001.

[25] K.S. Siveen and G. Kuttan. Role of macrophages in tumour progression. *Immunology letters*, 123(2):97–102, 2009.

[26] F. Belardelli and M. Ferrantini. Cytokines as a link between innate and adaptive antitumor immunity. *Trends in immunology*, 23(4):201–208, 2002.

[27] H. A. Alshaker and K. Z. Matalka. IFN-γ, IL-17 and TGF-β involvement in shaping the tumor microenvironment: The significance of modulating such cytokines in treating malignant solid tumors. *Cancer cell international*, 11(1):33, 2011.

[28] T. Walzer, M. Dalod, S. H. Robbins, L. Zitvogel, and E. Vivier. Natural killer cells and dendritic cells: "l'union fait la force".

[29] L. Yang, Y. Qi, J. Hu, L. Tang, S. Zhao, and B. Shan. Expression of Th17 cells in breast cancer tissue and its association with clinical parameters. *Cell biochemistry and biophysics*, 62(1):153–159, 2012.

[30] C. T. Weaver and K. M. Murphy. T-cell subsets: the more the merrier. *Current Biology*, 17(2):R61–R63, 2007.

[31] H. T. Khong and N. P. Restifo. Natural selection of tumor variants in the generation of "tumor escape" phenotypes. *Nature immunology*, 3(11):999–1005, 2002.

[32] C. A. Dinarello. Historical Review of Cytokines. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3140102/, 2011. [available in PMC 2011 Jul 20].

[33] A. Nicolini, A. Carpi, and G. Rossi. Cytokines in breast cancer. *Cytokine & growth factor reviews*, 17(5):325–337, 2006.

[34] A. Yates, C. Bergmann, J. L. Van Hemmen, J. Stark, and R. Callard. Cytokine-modulated regulation of helper T cell populations. *Journal of theoretical biology*, 206 (4):539–560, 2000.

[35] n. F. Koch, U. Stanzl, P. Jennewein, K. Janke, C. Heufler, E. Kämpgen, N. Romani, and G. Schuler. High level IL-12 production by murine dendritic cells: upregulation via MHC class II and CD40 molecules and downregulation by IL-4 and IL-10. *Journal of Experimental Medicine*, 184(2):741–746, 1996.

[36] J.C. Arciero, T.L. Jackson, and D.E. Kirschner. A mathematical model of tumor-immune evasion and siRNA treatment. *Discrete and Continuous Dynamical Systems Series B*, 4(1):39–58, 2004.

[37] B. E. Lippitz. Cytokine patterns in patients with cancer: a systematic review. *The lancet oncology*, 14(6):e218–e228, 2013.

[38] F. Belardelli. Role of interferons and other cytokines in the regulation of the immune response. *Apmis*, 103(1-6):161–179, 1995.

[39] S. Tarazona, F. García-Alcalde, J. Dopazo, A. Ferrer, and A. Conesa. Differential expression in RNA-Seq: a matter of depth. *Genome research*, 21(12):2213–2223, 2011.

[40] F. Finotello and B. D. Camillo. Measuring differential gene expression with RNA-Seq: challenges and strategies for data analysis. *Briefings in functional genomics*, 14 (2):130–142, 2015.

[41] F. Rapaport, R. Khanin, Y. Liang, M. Pirun, A. Krek, P. Zumbo, C. E. Mason, N. D. Socci, and D. Betel. Comprehensive evaluation of differential gene expression analysis methods for RNA-Seq data. *Genome biology*, 14(9):3158, 2013.

[42] E. J. Edelman, J. Guinney, J.-T. Chi, P. G. Febbo, and S. Mukherjee. Modeling cancer progression via pathway dependencies. *PLoS computational biology*, 4(2):e28, 2008.

[43] G. O. Consortium et al. Jingqi chengene ontology consortium: going forward. *Nucleic acids research*, 43(D1):D1049–D1056, 2015.

[44] E. Camon, M. Magrane, D. Barrell, V. Lee, E. Dimmer, J. Maslen, D. Binns, N. Harte, R. Lopez, and R. Apweiler. The gene ontology annotation (GOA) database: sharing knowledge in uniprot with gene ontology. *Nucleic acids research*, 32(suppl_1):D262–D266, 2004.

[45] M. Kanehisa and S. Goto. Kegg: kyoto encyclopedia of genes and genomes. *Nucleic acids research*, 28(1):27–30, 2000.

[46] T. A. Fehniger, M. A. Cooper, G. J. Nuovo, M. Cella, F. Facchetti, M. Colonna, and M. A. Caligiuri. Cd56bright natural killer cells are present in human lymph nodes and are activated by T cell-derived IL-2: a potential new link between adaptive and innate immunity. *Blood*, 101(8):3052–3057, 2003.

[47] F. R. Villegas, S. Coca, V. G. Villarrubia, R. Jiménez, M. J. Chillón, J. Jareño, M. Zuil, and L. Callol. Prognostic significance of tumor infiltrating natural killer cells subset CD57 in patients with squamous cell lung cancer. *Lung cancer*, 35(1):23–28, 2002.

[48] M. Vitale, M. Della Chiesa, S. Carlomagno, D. Pende, M. Aricò, L. Moretta, and A. Moretta. Nk-dependent DC maturation is mediated by TNFα and IFNγ released upon engagement of the NKP30 triggering receptor. *Blood*, 106(2):566–571, 2005.

[49] T. R. Mosmann, H. Cherwinski, M. W. Bond, M. A. Giedlin, and R. L. Coffman. Two types of murine helper T cell clone. I. definition according to profiles of lymphokine activities and secreted proteins. *The Journal of immunology*, 136(7):2348–2357, 1986.

[50] E. Bettelli and V. K. Kuchroo. IL-12-and IL-23-induced T helper cell subsets. *Journal of Experimental Medicine*, 201(2):169–171, 2005.

[51] P. R. Mangan, L. E. Harrington, D. B. O'Quinn, W. S. Helms, D. C. Bullard, C. O. Elson, R. D. Hatton, S. M. Wahl, T. R. Schoeb, and C. T. Weaver. Transforming growth factor-b induces development of the th17 lineage. 2006.

[52] V. Bronte. Th17 and cancer: friends or foes? *Blood*, 112(2):214–214, 2008.

[53] N. J. Wilson, K. Boniface, J. R. Chan, B. S. McKenzie, W. M. Blumenschein, J. D. Mattson, B. Basham, K. Smith, T. Chen, F. Morel, et al. Development, cytokine profile and function of human interleukin 17–producing helper T cells. *Nature immunology*, 8(9):950–957, 2007.

[54] M. Cella, D. Scheidegger, K. Palmer-Lehmann, P. Lane, A. Lanzavecchia, and G. Alber. Ligation of CD40 on dendritic cells triggers production of high levels of interleukin-12 and enhances T cell stimulatory capacity: TT help via APC activation. *Journal of Experimental Medicine*, 184(2):747–752, 1996.

[55] R. Eftimie, J. J. Gillard, and D. A. Cantrell. Mathematical models for immunology: Current state of the art and future research directions. *Bulletin of mathematical biology*, 78(10):2091–2134, 2016.

[56] V. A. Kuznetsov, I. A. Makalkin, M. A. Taylor, and A. S. Perelson. Nonlinear dynamics of immunogenic tumors: parameter estimation and global bifurcation analysis. *Bulletin of mathematical biology*, 56(2):295–321, 1994.

[57] R. Eftimie and H. Hamam. Modelling and investigation of the CD4+ T cells–macrophages paradox in melanoma immunotherapies. *Journal of Theoretical Biology*, 420:82–104, 2017.

[58] N. Y. den Breems and R. Eftimie. The re-polarisation of M2 and M1 macrophages and its role on cancer outcomes. *Journal of theoretical biology*, 390:23–39, 2016.

[59] L. G. de Pillis, A. E. Radunskaya, and C. L. Wiseman. A validated mathematical model of cell-mediated immune response to tumor growth. *Cancer research*, 65(17): 7950–7958, 2005.

[60] R. Eftimie, J. L. Bramson, and D. J Earn. Modeling anti-tumor Th1 and Th2 immunity in the rejection of melanoma. *Journal of theoretical biology*, 265(3):467–480, 2010.

[61] D. Kirschner and J. C. Panetta. Modeling immunotherapy of the tumor–immune interaction. *Journal of mathematical biology*, 37(3):235–252, 1998.

[62] R. Eftimie, J. L. Bramson, and D. J. Earn. Interactions between the immune system and cancer: a brief review of non-spatial mathematical models. *Bulletin of mathematical biology*, 73(1):2–32, 2011.

[63] J. Adam and N. Bellomo. A survey of models for tumor-immune system dynamics. Springer Science & Business Media, 2012.

[64] G. Teschl. Ordinary differential equations and dynamical systems, volume 140. American Mathematical Society Providence, 2012.

[65] J. W. Uhr, T. Tucker, R. D. May, H. Siu, and E. S. Vietta. Cancer dormancy: studies of the murine BCL1 lymphoma. *Cancer research*, 51(18 Supplement):5045s–5053s, 1991.

[66] B. Gompertz. On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies. *Philosophical transactions of the Royal Society of London*, 115:513–583, 1825.

[67] B. Burkholder, R.-Y. Huang, R. Burgess, S. Luo, V. S. Jones, W. Zhang, Z.-Q. Lv, C.-Y. Gao, B.-L. Wang, Y.-M. Zhang, et al. Tumor-induced perturbations of cytokines and immune cell networks. *Biochimica et Biophysica Acta (BBA)-Reviews on Cancer*, 1845 (2):182–201, 2014.

[68] D. Ricklin, G. Hajishengallis, K. Yang, and J. D. Lambris. Complement: a key system for immune surveillance and homeostasis. *Nature immunology*, 11(9):785–797, 2010.

[69] S. M. Santini, C. Lapenta, M. Logozzi, S. Parlato, M. Spada, T. Di Pucchio, and F. Belardelli. Type I interferon as a powerful adjuvant for monocyte-derived dendritic cell development and activity in vitro and in Hu-PBL-SCID mice. *Journal of Experimental Medicine*, 191(10):1777–1788, 2000.

[70] A. Diefenbach, E. R. Jensen, A. M. Jamieson, and D. H. Raulet. Rae1 and H60 ligands of the NKG2d receptor stimulate tumour immunity. *Nature*, 413(6852):165–171, 2001.

[71] P. Van den Driessche and J. Watmough. Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission. *Mathematical biosciences*, 180(1):29–48, 2002.

[72] C. Castillo-Chavez, Z. Feng, and W. Huang. On the computation of R0 and its role on. *Mathematical approaches for emerging and reemerging infectious diseases: an introduction*, 1:229, 2002.

[73] K. Soetaert, T. Petzoldt, and R. W. Setzer. Solving differential equations in R: Package desolve. *Journal of Statistical Software*, 33(9):1–25, 2010. ISSN 1548-7660. doi: 10.18637/jss.v033.i09. URL http://www.jstatsoft.org/v33/i09.

[74] K. P. Wilkie and P. Hahnfeldt. Tumor-immune dynamics regulated in the microenvironment inform the transient nature of immune-induced tumor dormancy. *Cancer research*, 73(12):3534, 2013.

[75] R. M. Ribeiro, H. Mohri, D. D. Ho, and A. S. Perelson. In vivo dynamics of T cell activation, proliferation, and death in HIV-1 infection: why are CD4+ but not CD8+ T cells depleted? *Proceedings of the National Academy of Sciences*, 99(24):15572–15577, 2002.

[76] S. Benzekry, C. Lamont, A. Beheshti, A. Tracz, J. M. Ebos, L. Hlatky, and P. Hahnfeldt. Classical mathematical models for description and prediction of experimental tumor growth. *PLoS computational biology*, 10(8):e1003800, 2014.

[77] H. Siu, E.S. Vitetta, R.D. May, and J.W. Uhr. Tumor dormancy. I. regression of BCL1 tumor and induction of a dormant tumor state in mice chimeric at the major histocompatibility complex. *The Journal of Immunology*, 137(4):1376–1382, 1986.

[78] K.-L. Liao, X.-F. Bai, and A. Friedman. Mathematical modeling of interleukin-27 induction of anti-tumor t cells response. *PloS one*, 9(3):e91844, 2014.

[79] K. Soetaert and T. Petzoldt. Inverse modelling, sensitivity and monte carlo analysis in R using package FME. *Journal of Statistical Software*, 33(3):1–28, 2010. doi: 10.18637/jss.v033.i03. URL http://www.jstatsoft.org/v33/i03/.

[80] J. Olivier, W. D. Johnson, and G. D. Marshall. The logarithmic transformation and the geometric mean in reporting experimental IgE results: what are they and when and why to use them? *Annals of Allergy, Asthma & Immunology*, 100(4):333–337, 2008.

[81] G. Edward Miller. Asymptotic test statistics for coefficients of variation. *Communications in Statistics-Theory and Methods*, 20(10):3351–3363, 1991.

[82] S. Banik, B.M. Kibria, and D. Sharma. Testing the population coefficient of variation. *Journal of Modern Applied Statistical Methods*, 11(2):5, 2012.

[83] F. Seyednasrollah, A. Laiho, and L. L. Elo. Comparison of software packages for detecting differential expression in RNA-Seq studies. *Briefings in bioinformatics*, 16 (1):59–70, 2013.

[84] G. Chu, J. Li, B. Narasimhan, R. Tibshirani, and V. Tusher. Significance analysis of microarrays users guide and technical document, 2001.

[85] R. Tibshirani, G. Chu, B. Narasimhan, and J. Li. *samr: SAM: Significance Analysis of Microarrays*, 2011. URL https://CRAN.R-project.org/package=samr. R package version 2.0.

[86] G. K. Mazandu and N. J. Mulder. A topology-based metric for measuring term similarity in the gene ontology. *Advances in bioinformatics*, 2012, 2012.

[87] A. P. Sagona and H. Stenmark. Cytokinesis and cancer. *FEBS letters*, 584(12):2652–2661, 2010.

[88] W. Antonin and H. Neumann. Chromosome condensation and decondensation during mitosis. *Current opinion in cell biology*, 40:15–22, 2016.

[89] A. Ciechanover. The ubiquitin–proteasome pathway: on protein death and cell life. *The EMBO journal*, 17(24):7151–7160, 1998.

[90] R. Layfield, A. Alban, R. J. Mayer, and J. Lowe. The ubiquitin protein catabolic disorders. *Neuropathology and applied neurobiology*, 27(3):171–179, 2001.

[91] C. Lengauer, K. W. Kinzler, and B. Vogelstein. Genetic instabilities in human cancers. *Nature*, 396(6712):643–649, 1998.

[92] M. F. Zacarias-Fluck, B. Morancho, R. Vicario, A. L. Garcia, M. Escorihuela, J. Villanueva, I. T. Rubio, and J. Arribas. Effect of cellular senescence on the growth of HER2-positive breast cancers. *JNCI: Journal of the National Cancer Institute*, 107(5), 2015.

[93] C. Bianco, A. Eden, and Z. A. Cohn. The induction of macrophage spreading: role of coagulation factors and the complement system. *Journal of Experimental Medicine*, 144(6):1531–1544, 1976.

[94] F. Lupu, R. S. Keshari, J. D. Lambris, and K. M. Coggeshall. Crosstalk between the coagulation and complement systems in sepsis. *Thrombosis research*, 133:S28–S31, 2014.