Modelling Chromosome Missegregation in Tumour Evolution

Arturo Araujo

A dissertation submitted in partial fulfilment
of the requirements for the degree of
Doctor of Philosophy
of
University College London

CoMPLEX

UCL Computer Science

January 2013

I, *Arturo Araujo,* confirm that the work presented in this dissertation is my own. Where information has been derived from other sources, I confirm that this has been indicated in the dissertation.

Acknowledgments

Finally, this work is dedicated to my mother Elia Gutierrez, my brother Ricardo Araujo, my sister Marlene Araujo and my late father Rodulfo Araujo. Although my father did not live to see this dream of mine come true, a part of him lives through this new piece of original science. I thank them for making me feel loved and for their unconditional support. They have done more for me that can ever be said, or that can ever be fully repaid.

All of these people are seared into my heart. I like to think that they have taken a piece of me and that I in turn carry a piece of them wherever I go and whatever I do. And to all the people that said that this research couldn't be done, this work is proof that not only it could be done, but that it was the necessary first step into a promising new path in the fight against cancer.

-Arturo Araujo

October 2013

Abstract

Cancer is a disease in which the controls that usually ensure the coordinated behaviour of individual cells break down. This rarely happens all at once. Instead, the clone of cells that grows into a developing tumour is under high selection pressure, leading to the evolution of a complex and diverse population of related cells that have accumulated a wide range of genetic defects. One of the most evident but poorly characterized of these genetic abnormalities is a disorder in the number of chromosomes, or aneuploidy. Aneuploidy can arise though several different mechanisms. The project explores one such mechanism - chromosome missegregation during cell division- and its role in oncogenesis.

To address the role that chromosome missegregation may have in the development of cancer a computational model was devised. We then defined the behaviour of individual cells, their genomes and a tissue niche, which could be used in simulations to explore the different types of cell behaviour likely to arise as the result of chromosome missegregation. This model was then used to better understand how defects in chromosome segregation affect cancer development and tumour evolution during cancer therapy. In stochastic simulations, chromosome missegregation events at cell division lead to the generation of a diverse population of aneuploid clones that over time exhibit hyperplastic growth. Significantly, the course of cancer evolution depends on genetic linkage, as the structure of chromosomes lost or gained through missegregation events and the level of genetic instability function in tandem to determine whether tumour growth is driven primarily by the loss of tumour suppressors or by the overexpression of oncogenes. As a result, simulated cancers differ in their level of genetic stability and in their growth rates. We then used this system to investigate the consequences of these differences in tumour heterogeneity for anti-cancer therapies based on surgery and anti-mitotic drugs that selectively target proliferating cells.

Results show that simulated treatments induce a transient delay in tumour growth, and reveal a significant difference in the efficacy of different therapy regimes in treating genetically stable and unstable tumours. These data support clinical observations in which a poor prognosis is correlated with a high level of chromosome missegregation. However, simulations run in parallel also exhibit a wide range of behaviours, and the response of individual simulations (equivalent to single tumours) to anti-cancer therapy prove extremely variable. The model therefore highlights the difficulties of predicting the outcome of a given anti-cancer treatment, even in cases in which it is possible to determine the genotype of the entire set of cells within the developing tumour.

Contents

List of Figures

15

List of Tables

1. Introduction

Cancer is not a single disease. It is a group of diseases, all involving unregulated cell growth (R. A. Weinberg, 2007). Typically in cancer, a single clone of cells divides and grows uncontrollably before spreading to colonize distant parts of the body (Hanahan & Weinberg, 2011). Although some cancers are caused by specific external agents (e.g. mesothelioma – asbestos (D. W. Berman, 2011)/ skin cancer - UV light (Wang et al., 2005)/ lung cancer- tobacco smoke (Ha et al., 2004) / Burkitts lymphoma – EBV (S. Lee et al., 2007)) the precise cause of most cancers remains unknown. Instead, a large number of factors increase the risk of cancer (Anand et al., 2008; Marcucci, Haferlach, & Döhner, 2011; R. A. Weinberg, 2007); the greatest of these being age (Meza, Jeon, Moolgavkar, & Luebeck, 2008). As a result, the probability of surviving this disease varies greatly depending on the type of cancer and the genetics of each patient. Moreover, each individual case of cancer is unique and changes during its life history. Once cancer starts, several factors act in tandem to shape the subsequent evolution of cancer. Amongst them, mutation and selection constantly change key properties of cancer, such as proliferation rates and resistance to death (R. A. Weinberg, 2007). Cancer can affect people of all ages, and the risk of developing cancer increases with age. This means that everyday each one of us is at higher risk of acquiring this disease.

Much effort has been dedicated to understand tumours through studying their individual components: altered genes that break down the cellular cooperation to maintain an organism. Although some important cancer hallmarks have been proposed (Hanahan & Weinberg, 2011), fundamental questions about the development of the disease have not yet been answered. Are there rules that dictate the pathways that led to different kinds of cancer? How do cell behaviours, usually regulated by a robust network of interconnected genes, change in such a way that the

31

organismic integrity becomes compromised? What are the interactions and dynamics that take place as cancer evolves? As more research is carried out, it is clear that evolution within the tumour is a key aspect of the disease (Merlo, Pepper, Reid, & Maley, 2006). If we could study the development of cancer cells in a living organism, could we recognize a key evolutionary transition while it is happening?

Cancer research seeks to understand this complex disease at a fundamental level and to develop better methods for the effective prevention, diagnosis and treatment of cancer. While much of our understanding of cancer has come from studying the individual constituents involved, the genes, gene products and other factors, like carcinogens, phenomena such as the emergent behaviours that appear when a number of simple agents, like cancer cells, interact challenge the limits of this reductionist approach. Fortunately, recent advances in mathematics and computer science have provided methods for exploring these phenomena. The aim of this work is to use these methods to explore an important, but poorly understood biological phenomenon in cancer: Aneuploidy.

This original research takes place at the interphase between computer science and biology. Cutting edge materials and methods from both fields, from *in silico* agent-based modelling to wet lab immunohistochemistry techniques, will be combined to create a useful framework to study the role of aneuploidy in cancer. Although these techniques are well established on their own, this is to our knowledge the first time they have been applied in such an integrated way. We prepose that this kind of integration will offer a unique opportunity to better understand the phenomenon of aneuploidy and thus result in the formulation of first principles; a first at the time of writing for this complex phenomenon. Furthermore, new wet-lab protocols will be designed to form a bridge between the *in silico* models and the biological models.

1.1. Research Problem: Aneuploidy in Cancer

For many types of cancer, a complex and diverse array of different kinds of genetic abnormalities accumulates within individual tumours as they evolve (Jefford & Irminger-Finger, 2006). One of the most evident but poorly characterized of these chromosomal abnormalities is the presence of an abnormal number of chromosomes, also known as aneuploidy, in the cancer genomes (Jacobs, Luce, & Cailleau, 2006). Such alterations in the number of chromosomes have been shown to correlate with the dysregulation of gene expression (Hertzberg et al., 2007) (Pavelka, Rancati, & Li, 2010a). This can have similar effects to the loss of tumour suppressors or the amplification of oncogenes that encourage proliferation. Recent research has started to throw light on the possible mechanisms that generate missing or additional chromosomes in cells (Ganem, Godinho, & Pellman, 2009). Experimental evidence shows how defects arising during cell division can give rise to the missegregation of entire chromosomes between the two daughter cells (Crasta et al., 2012), (Thompson, Bakhoum, & Compton, 2010), (Cimini, 2008), (Feldser, Hackett, & Greider, 2003). Whether an abnormal number of chromosomes is a consequence or a cause of malignant transformation is a hotly debated topic in cancer research (B. Weaver & Cleveland, 2007), (Nigg, 2002), (Torres et al., 2007).

An argument for chromosome missegregation having a causal role in tumourigenesis is that it creates novel genotypes, some of which may promote unregulated cell proliferation (B. A. A. Weaver, Silk, Montagna, Verdier-Pinard, & Cleveland, 2007). For evolution to occur, there must be heredity, variation and selection. Evolution involves tinkering at each generation; an opportunity provided by chromosome missegregation. Conversely, excessive chromosome missegregation can work against the process of tumourigenesis by making cell division so unstable that genotypes are not faithfully replicated at each round of cell division, reducing cell fitness (Holland & Cleveland, 2009). Excessive chromosome missegregation can also prevent

cells from carrying out basic essential functions, resulting in cell death (Thompson & Compton, 2010). Currently, there is no clear consensus on the role of chromosome missegregation and cancer (Gibbs, 2003). Because of this, a holistic computational approach that integrates the accumulated knowledge of this phenomenon may help in shedding light on the role of chromosome missegregation in tumourigenesis.

The objective of this work is to assess whether a computational model of chromosome missegregation is an effective approach to study the role of aneuploidy in cancer evolution. The first step towards such a model is the abstraction of key biological objects such as genes and chromosomes; and processes such as cellular division and chromosome segregation. Within the framework of an abstract homeostatic tissue where cellular processes are regulated by genes, several *in silico* experiments can be carried out. The model presented in this work will be used to explore whether differences in the path by which cancers evolve can account for the different, seemingly conflicting, observations regarding **chromosome missegregation** in cancer (B. Weaver & Cleveland, 2007).

1.2. Hypothesis

Whether chromosome missegregation is sufficient to initiate carcinogenesis, or if it has a key role in shaping the evolutionary dynamics of cancer progression is still unknown (B. Weaver & Cleveland, 2007). By integrating current biological data and theories that describe the evolution and behaviour of chromosome missegregation in cancer, a computational model could capture the dynamics of such a complex system. The hypothesis presented in this work is that:

An integrated computational model of chromosome missegregation during cell division provides an effective approach to assess the role of aneuploidy in the evolution of cancer.

where we will use the following definitions:

- **Aneuploidy** is the cellular state of having an abnormal number of chromosomes.
- **Chromosome missegregation** is the phenomenon in which cells make errors at the time of the distribution of their duplicated genetic content during cell division. Problems during chromosome separation may result in daughter cells with different numbers of chromosomes. The resulting cells, known as aneuploid cells, are a condition associated with cancer.
- **An integrated computational model** is the set of software that constitutes a virtual lab. Such a set would comprise not only of the model of chromosome missegregation, but also the tools needed to extract useful information from it.

To test the hypothesis the following methodology will be used:

1. Carry out a literature review of cancer biology and cancer modelling.
2. From the literature review of cancer biology, extract the first principles of chromosome missegregation.
3. Informed by the cancer modelling literature review, build an integrated model of chromosome missegregation.
4. Assess the usefulness of that model.

The model will be used to investigate whether chromosome missegregation has a significant role in the evolution of cancer. By making use of biological abstractions and behaviour, we seek to shed light on the kind of evolutionary pathways that cancers may take. This is a key step in elucidating some of the governing principles that may in turn help guide biological experiments

and theory. Such a model would be useful if it is able to provide us with insights regarding the evolution of cancer, and how that evolution may be affected by different forms of intervention (Endy & Brent, 2001).

1.3. Scope of the Project

The scope of the project will not include the modelling of other kinds of genetic change (such as mutations in key nucleotides or the breakage of chromosomes, which will be discussed in the next chapter) or an explicit physical model of a tissue context (such as hypoxia due to the lack of oxygen at the centre of a solid tumour or the different densities that act as barriers between distinct kinds of tissues). We will not address other phenomena such as the phenomenon of metastasis (the colonization of other organs by cancer cells).

1.4. Objectives

The main objectives in this work are:

1. Investigate in the literature possible pathways for cancer to originate through, or for its subsequent evolution to be affected by the phenomenon of chromosome missegregation.
2. Research which modelling paradigms would be better suited to model the biological phenomenon of aneuploidy.
3. Study how computational models could be validated or corroborated.
4. Summarise and abstract relevant biological concepts for computational implementation and incorporate current theories and known facts into a working computational model.
5. Identify key emergent states and behaviours in both the model and those reported in the literature for the evolution of cancer.
6. Compare with results from the model two key properties of chromosome missegregation reported in the literature, and demonstrate that the model can be used to support or refute the following theories of chromosome missegregation:

- Chromosome missegregation can lead to further genetic instability, which may be advantageous to cancer progression (B. A. A. Weaver et al., 2007).
- The levels of aneuploidy may change at different stages of cancer progression, shaping the subsequent tumour evolution (Jefford & Irminger-Finger, 2006).

7. Investigate through the implementation of *in silico* experiments the role of chromosome missegregation in cancer treatments.

8. Design experimental tools and techniques that can enable biologists to design experiments informed by the model.

1.5. Thesis Contributions

This work presents a number of original contributions to the fields of computer science, molecular biology and cancer research:

1. Results from *in silico* experiments highlight the existence of two distinct types of aneuploidy. In the first type, dominant proliferating clones within the tumour exhibit a relatively stable state of aneuploidy. In the second, selection for the loss of the aneuploidy gene results in tumours that continually generate increasing levels of heterogeneity and ever-more malignant subclones. Both have been found in real tumours.

2. Using the tool developed, organizing principles of chromosome missegregation were elucidated. It was found that cancer evolution depends on genetic linkage, as the structure of chromosomes lost or gained through missegregation events and the level of genetic instability function in tandem to determine whether tumour growth is driven primarily by the loss of tumour suppressors or by the overexpression of oncogenes. As a result, simulated cancers differ in their level of genetic stability and in their growth rates.

3. The role of chromosome missegregation in cancer therapies was investigated, and results highlight that treatments induce a transient delay in tumour growth. Results also reveal a significant difference in the efficacy of different therapy regimes in treating genetically stable and unstable tumours. These data support clinical observations in which a poor prognosis is correlated with a high level of chromosome missegregation.

4. The model places a key role for tumour heterogeneity in the response to treatments. Simulations run in parallel exhibit a wide range of behaviours, and the response of individual simulations, equivalent to single tumours, to anti-cancer therapy prove extremely variable. The model therefore highlights the difficulties of predicting the outcome of a given anti-cancer treatment, even in cases in which it is possible to determine the genotype of the entire set of cells within the developing tumour.

5. An experimental protocol that provides the proof of principle for an immunostaining technique as a measurement of Ploidy in real cancer samples was developed and tested successfully in the wet lab.

6. Aa algorithm for the implementation of an immunostaining-based ploidy measuring technique , which could be useful in designing *in vitro* experiments that make use of the model presented in this work, was developed and tested successfully.

7. A protocol for inducing aneuploidy in cell lines through Aurora Inhibition was developed and tested successfully *in vitro*.

1.6. Publications

The research described in this thesis led to a number of publications:

Araujo, A., Bentley, P. J. and Baum, B. (2010) Modelling the Role of Aneuploidy in Tumour Evolution. 12th International Conference on the Synthesis and Simulation of Living Systems. 2010. MIT Press.

Araujo, A., Bentley, P. J. and Baum, B. (2010) Modelling the Role of Chromosome Missegregation in Cancer Therapies. 3rd Complex Systems Modelling and Simulation Workshop. 2010. Luniver Press.

Araujo, A., Bentley, P. J. and Baum, B. (2013) The Role of Chromosome Missegregation in Cancer Development: A Theoretical Approach Using Agent-Based Modelling. PLoS ONE 8(8): e72206

1.7. Thesis Overview

The next chapter provides a Background section in which the relevant biological concepts are explained. The chapter starts with a concise review of the current biological ideas, data and concepts in cancer research. This is then followed by a review on the most representative methodologies that have been used to model this disease. This section provides information on how experiments have informed the models and how in turn these models have helped drive knowledge and experiments forward.

Chapter 3 focuses on the methodology employed for the development of the model. The decisions behind the selection of the modelling paradigm are explained and justified. The methodology that will be used for the assessment of the model is also presented.

Chapter 4 describes the modelling process. In this chapter, the conceptualizations of key biological aspects modelled are carried out. The different iterations of the model are presented and assessed. The final version of model is proposed and justified.

Chapter 5 describes the experiments carried out with the model. The goal of the chapter is to test whether the model can provide insights regarding the role of chromosome missegregation in the evolution of cancer. Analysis tools created for the analysis of the model are presented and used.

Results will show that the model suggests that chromosome missegregation may give rise to two clinically observed kinds of aneuploidy: a stable aneuploidy and an unstable one.

Chapter 6 takes the results obtained in the previous chapter to investigate the role of chromosome missegregation on cancer therapies. Scenarios that model two types of cancer treatments: Surgery and Chemotherapy are presented. For the analysis of the results of these experiments, new analysis tools are introduced. Several key insights generated by the model are highlighted, including the suggestion of some predictive markers for a successful therapy.

Chapter 7 describes the experimental work carried out during the development of the computational model in the wet lab. This section complements the body of work by offering a solid grounding in reality for the model through immunostaining experiments. A tool for the measurement of aneuploidy by means of image cytometry is proposed, as well as contributing with a novel protocol for the generation of aneuploid cells in the wet lab.

Chapter 8 concludes this work by discussing the body of work, and putting the results in context. The approach taken and the inaccuracies and insufficiencies in the model are evaluated and discussed. The chapter concludes with future work, which contains plans to address many of these deficiencies.

2. Background

In this chapter, the current biological and computational literature, key biological discoveries and existing methods of biological modelling will be reviewed and discussed. The chapter begins with an introduction to the relevant biology of cancer, focusing on biological definitions and current theories of key biological processes such as cancer genes and genetic instability. This section is then followed by a literature review on computational cancer modelling, with focus on the most relevant models. In this chapter, evidence will be provided for the existence of a gap in the current knowledge that the work presented in this thesis seeks to fill.

2.1. The Biology of Cancer

Cancer is a disease that will affect one out of every three people on the planet (Jemal et al., 2008). Although the likelihood of cancer striking correlates with age, it is a disease that may affect young people, and even children *in utero* (Kontopoulos, Gualtieri, & Quintero, 2012). The earliest recorded case of human cancer dates back to Ancient Egypt in 1500 B.C., where 8 cases of breast tumours were documented on papyrus (interestingly, even back then they were treated by cauterization) (Sakorafas & Safioleas, 2009), (Hajdu, 2010). Even Greek philosophers like Hippocrates, who believed that black bile was responsible for the condition, were concerned about this disease (Hajdu, 2004). It was not until autopsies began to be carried out more regularly in the 1700's that it became clear that lymph nodes (replacing the black bile theory) played a key part in the spreading of the disease (Sakorafas & Safioleas, 2009). It was then theorized that abnormalities in lymph nodes were the cause of cancer. It was in the late 1800's that a German doctor, Rudolph Carl Virchow (also known as the father of modern pathology), proposed the cellular origins of cancer (Sakorafas & Safioleas, 2010). After the discovery of the cell, Virchow reasoned that all cells come from other cells; even cancer cells. This means that a single cell with

oncogenic properties has the potential to form a tumour. It was then, and still is today a very frightening concept. Cancer comes from within. It is our own body turned against us.

The 20th century saw the greatest progress in cancer research. The discovery of DNA, the development of the computer, the faster dissemination of knowledge throughout the globe and, key amongst those, the birth of molecular biology, contributed greatly to our current understanding of cancer as a genetic disease. Today, some cancers have become curable, and rational treatments and preventions for many others have been developed (Sakorafas, Safioleas, & Safioleas, 2010). Better treatments are currently being designed, taking into account everything that it is currently known about this disease. In reality, however, few rationally designed drugs have proved useful such as Imatinib and Herceptin (S. E. Bates, Amiri-Kordestani, & Giaccone, 2012). Throughout this chapter, a general framework in which to view the disease will be described. This work will explore the biology of cancer, review the main theories regarding the origin of cancer, and discuss the biological issues that surround the hypothesis of this work.

2.1.1. Introduction to Cancer Biology

Although there are hundreds of different kinds of cancers, each with its own individual properties, there is a consensus amongst cancer researchers regarding six hallmarks of cancer (Hanahan & Weinberg, 2000), (Hanahan & Weinberg, 2011). Unlike normal cells, cancer cells share the following properties:

1. Have a limitless potential to replicate
2. Avoid suicide by apoptosis
3. Are insensitive to anti-growth signals
4. Become self-sufficient with respect to growth signals

Once cancer cells have acquired these key properties that enable tumour growth, the cancer cells within the crowded tumour can acquire additional properties:

5. Recruitment of blood vessels to meet the oxygen demands of the growing tumour
6. Acquire the ability to break away from original tissue, survive and proliferate in foreign tissue (invasion and metastasis).

Although relatively simple, this conceptual framework has provided cancer researchers a cell biological foundation for the understanding of the disease. These properties are acquired via different mechanisms and at different times. However, these cellular phenotypes result from changes at the epigenetic and gene level, from point mutations to deletions, large-scale rearrangements of the genome and changes in chromosome copy number. The mechanisms of cancer initiation (oncogenesis) and tumour evolution, however, are a matter of heated debate (Shackleton, Quintana, & Fearon, 2009). To unravel this mystery, it is necessary to understand the first principles of cancer biology.

Research carried out in the 1970´s revealed that functional changes in the activity of genes that control proliferation play a key role in cancer. Typically these cancer-enabling genes, also known as proto-oncogenes, become true oncogenes as the result of mutation or rare gene-fusion events, and can contribute to oncogenesis when overexpressed as the result of changes in gene copy number and epigenetic changes. Oncogene activation, if left unchecked, can lead to cellular over proliferation (Weinstein & Joe, 2008). Cells however appear to have safety mechanisms in place that prevent against excessive proliferation. These genes are known as tumour-suppressor genes. Some function to dampen or inhibit cell proliferation (Rb), while others help detect and respond to damage and/or promote cell death (p53) (Dean, Mcclendon, Stengel, & Knudsen, 2010). When these tumour-suppressor genes are lost or mutated, cells become more vulnerable to oncogenic reprograming. A cell also has failsafe mechanisms, such as the Hayflick limit, which

restrict the number of times a cell can divide. It is thought that telomeres associated with each cell's DNA will get slightly shorter with each new cell division until they shorten to a critical length. Empirical evidence shows that typically cancer cells have managed to circumvent this by re-expressing telomerase and this makes cells practically immortal (R. A. Weinberg, 2007).

With the advent of whole genome sequencing it is now possible to study the entire set of evolving changes in genome sequence and architecture that accompany cancer development. This type of analysis reveals an enormous number of genetic changes (Meyerson, Gabriel, & Getz, 2010). These likely include a small number of driver mutations that directly contribute to cancer development, together with a very large number of mutations that are considered passenger mutations that may perform no function or that may functionally compromise tumour cell function. In addition there is enormous level of genetic heterogeneity in DNA sequence and genome architecture (Gerlinger et al., 2012). This genetic heterogeneity makes it difficult to identify changes in oncogenes and tumour suppressor genes that drive cancer development (recent work uses a quantitative analyses of genetic data from tumours to generate tumour cell lineage trees, identifying early and late mutational events (Frumkin et al., 2008)). Despite the genetic heterogeneity, one key feature that many cancer cells share is an aneuploid genotype (Swanton & Caldas, 2009)–genomes composed of an abnormal number of chromosomes.

2.1.2. On the Origins of a Tumour

Although many carcinogens have been identified, no one really knows the initial trigger that leads to most common cancers. There are, however, several theories regarding the likely origins of most cancers (Gibbs, 2003). Although cancer origins are likely to begin with the birth of multicellularity (Saul & Schwartz, 2007), the dominant theory suggests that DNA damage over decades leads to many thousands of random mutations in the genome of each cell which first

trigger the onset of overproliferation (Muzny et al., 2012) (Gavrilov & Gavrilova, 2002). These mutations are likely to arise from carcinogens like ionizing radiation (X-rays, etc.) as well as chemicals in foodstuffs and the highly reactive by-products of our own metabolism (Ding et al., 2008). This theory of stochastic damage however raises the question: How can such a complicated genome reprogramming be carried out so consistently for the development of a cancer phenotype by means of stochastic mutations and in such a limited time span? (Huang & Ingber, 2007). The answer is that this theory lacks the lens of evolution, which, by generating variation, selection and heredity, can create consistent pathways (Merlo et al., 2006).

A second theory suggests that specific mutations in a few key genes are sufficient to reprogram cell behaviour leading to tumourigenesis (Hanahan & Weinberg, 2011). This is based upon experimental evidence that structural changes in a small number of genes that have a direct impact on the regulation of growth, death and division can disrupt internal and external cell regulation and reprogram cell fate (Gatenby et al., 2007). Such evidence first emerged from the isolation of specific genetic sequences from viruses that induce cancer-like behaviour (Rous sarcoma virus in chickens (Laan et al., 2004) and oncoviruses such as the human papilloma virus in humans (Tzenov et al., 2013)), and from research on familial cancers (inherited genetic susceptibility to diseases like retinoblastoma). It has since been shown that specific mutations in a small number of genes when combined are sufficient to cause cancer in model various systems from fruit flies to mice (R. A. Weinberg, 2007). However, these include mutations, such as the RasV12 mutation, that are likely to be exceedingly rare. Moreover, the tumours that arise in these models typically contain a heterogeneous mix of cells, with large numbers of genetic abnormalities and high levels of genetic instability (Wagner, 2004). Thus, this simplistic view does not capture the underlying evolutionary and selective forces acting within developing

tumours, the impact that a mutated phenotype has on the cell population, nor the interaction with a particular microenvironment in which phenotype selection takes place (Gatenby, 2006).

A third theory, the aneuploid theory of cancer, suggests that errors in cell division cycle (Figure 2.1) can give rise to changes in chromosome numbers and in turn may result in tumourigenesis (Gibbs, 2003). An aberrant chromosome number that deviates from a multiple of the haploid, also known as **aneuploidy**, is a remarkably common feature of human cancers. The theory proposes that defects in cell division may give rise to the missegregation of entire chromosomes; resulting in the creation of two daughter cells with different sets of chromosomes (Figure 2.2). The loss or gain of entire chromosomes could generate new viable genomes, with a probability exceeding that of conventional mutations by several orders of magnitude (Rajagopalan & Lengauer, 2004). Such events may generate variation in an initially homogeneous population upon which selection can act. Through chromosome missegregation, hundreds of genes are affected simultaneously. The aggregation effect of many genes being mutated, e.g. the loss of function of genes that suppress growth and the over expression of genes that promote growth, may act in tandem to provide cancer cells with powerful oncogenic traits. Genetic variability combined with heredity and natural selection may enable cells to rapidly evolve into progressively more aggressive genotypes and could explain the diversity of tumours observed in the clinic.
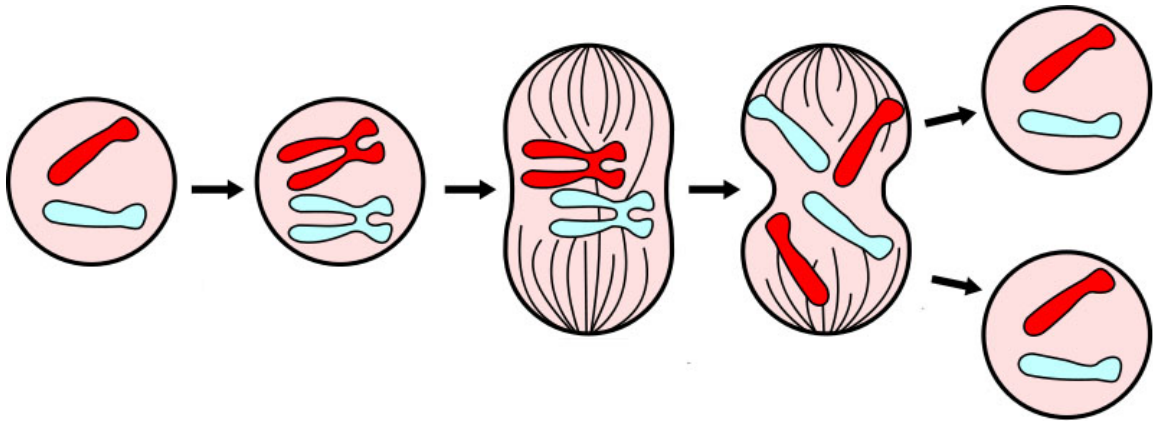
Figure 2.1- During the process of cellular mitosis a eukaryotic cell segregates the chromosomes in its cell nucleus into two identical sets. This process normally results in two daughter cells with the same DNA content
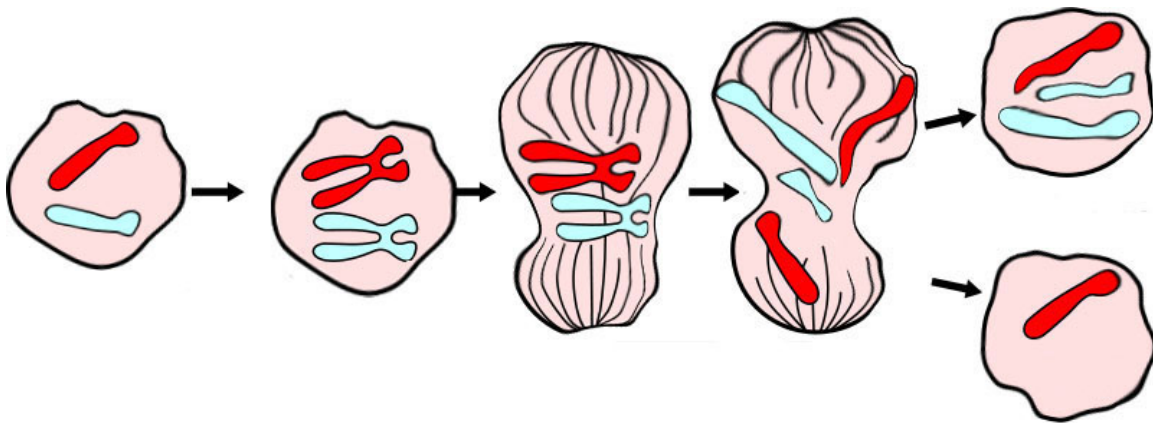
Figure 2.2- The missegregation of chromosomes during mitosis may lead to the creation of novel genotypes. The two daughter cells may not have the same DNA content, thus generating a state of aneuploidy. The genomic instability arising from this process may be crucial for the origin of cancerous cells.

Aneuploidy has long been known to be a hallmark of cancer; but because it arises as the result of chromosomal instability it is difficult to study and therefore it has been largely ignored (Holland & Cleveland, 2009). Aneuploidy can present itself in many forms and may exhibit different kinds of emergent behaviours, making its evolution at the cellular level very difficult to trace (Jefford & Irminger-Finger, 2006).

47

Although it is clear that aneuploidy is a hallmark of most cancers (Chin et al., 2004), (Austin et al., 2008), (Elzagheid et al., 2008), and that whole chromosome rearrangements (each chromosome containing thousands of genes) could generate differences in phenotype that could then be the subject of selection for cancer cell proliferation, the ideas and concepts behind the aneuploidy theory of cancer have caused much controversy amongst cancer researchers. There are several reasons for this. I. Few tumours have been identified that have a stable and characteristic cancer-specific aneuploid chromosomal organisation (due largely to the heterogeneity observed in many cancers) (Duesberg & Rasnick, 2000). II. With the identification of oncogenes in the 1970's (Sakorafas et al., 2010), specific genetic lesions became the predominant focus of most cancer research rather than the variation in the genetic architecture of cells between and within cancers. III. Because it is difficult to study the consequences of aneuploidy *in vivo* or its development during cancer progression (Holland & Cleveland, 2009; Pavelka et al., 2010b). However, progress has recently been made that makes it worth reconsidering this theory. It has been recently show that aneuploidy causes quantitative changes in the entire set of proteins expressed by the genome of budding yeast; accompanied also by phenotypic variation (Pavelka et al., 2010b)).

2.1.3. Oncogenes and Tumour Suppressors

From a recent census of the genes that are causally implicated in cancer development (Stratton, Campbell, & Futreal, 2009), more than 2000 have been reported, making up more than 1% of all the genes in humans (Futreal et al., 2004); more being discovered each year. Many initiatives exist to further identify new cancer genes by means of massive collaborations like the United Kingdom's "Cancer Genome Project" (Stephens et al., 2004) and the American "Cancer Genome Atlas" (Hudson et al., 2010) amongst others (Strachan, Abitbol, Davidson, & Beckmann, 1997).

Of those cancer genes that have been identified, some key genes have been classified as either proto-oncogenes or tumour suppressor genes, according to the types of mutation generated and their predicted role in cancer progression.

From all the known cancer genes, those classified as *proto-oncogenes* are those which, once mutated to their oncogenic form, play an important role in the regulation of cell growth and proliferation (R. A. Weinberg, 2007). Oncogenes typically code for important molecules used in signalling such as growth factors, increase transcription of other growth promoters or deregulate the cell cycle in such a way to make division more likely. *Tumour suppressor genes,* on the other hand, are those genes that suppress increases in cell number, either by inhibiting proliferation (Collins, Napoli, Ribeiro, Roberts, & Lloyd, 2012) or by causing cell death (R. A. Weinberg, 2007). Their inhibitory function typically becomes important once cells begin to overproliferate (Evan, 2006). They then restrain tumour growth, until two copies of each tumour suppressor gene within the diploid genome of a human cell are lost through epigenetic silencing, e.g. through methylation of the gene promoter region, and/or mutation, e.g. disabling the function of p53 in inducing cellular suicide, also known as apoptosis (Matlashewski et al., 1984), (Kimura et al., 2008) in response to the stressful conditions associated with overproliferation.

Although alteration in some of these cancer genes may account for specific cellular misbehaviours, the evolutionary pathway through which cells become cancerous (i.e. acquire the hallmarks of cancer discussed in Chapter 2.1.1) remains unknown, although it is starting to be revealed by the sequencing of populations of cells within tumours (Navin et al., 2011). Different types of cancer may have different evolutionary trajectories. In the absence of sexual recombination, the possible paths to cellular evolution are through genetic abnormalities including epigenetic changes, mutations and events with higher probability such as chromosome

rearrangements, and whole chromosome aneuploidy (Cimini & Degrassi, 2005), (Stratton et al., 2009). Moreover, because there is no recombination genomes are inherited as indivisible blocks. In this case, once the cells with the least mutated genomes begin to carry at least one mutation, no genomes with fewer such mutations can be inherited to future generations. This results in an eventual accumulation of passenger mutations, known as genetic load, which cannot be cleared of the population..

2.1.4. Aneuploidy in Cancer

There have been recent attempts to use clinical data and novel experimental models to determine whether chromosome missegregation is a cause or a consequence of malignant transformation, and the role it plays in tumour evolution (Duesberg & Rasnick, 2000). One such attempt dealt with the study of families with known history of *mosaic variegated aneuploidy*, an autosomal recessive condition where cells missegregate chromosomes with high risk of malignancy due to genetic alterations (Hanks & Rahman, 2005). The study focused on the gene BUB1B, whose function in animal studies is to delay chromosome segregation until all the chromosomes have been attached to the mitotic spindle, which then pulls and separates the paired chromosomes in opposite directions (Hede, 2005). Although this study directly links chromosome missegregation with cancer development, the size of the study was small and the effects of this particular kind of germ line predisposition for aneuploidy include many other diseases besides cancer, such as growth retardation, microcephaly and eye anomalies amongst others (Hanks et al., 2004).

Some scientists have argued that chromosome missegregation is unlikely to function in tumour initiation (Hahn et al., 1999), (X. Li, Schimenti, & Tye, 2009). Instead it may be an indirect effect of the altered activity of proto-oncogenes and tumour suppressor genes (Cahill, Kinzler, Vogelstein, & Lengauer, 1999), (Nigg, 2002),, including genes like p53 which has been

50

considered a "guardian of the genome". These arguments gained support from studies carried out in the mid 1990's which showed that there was a link between cellular transformation and genomic instability (Denko, Giaccia, Stringer, & Stambrook, 1994), (Felsher & Bishop, 1999). This however may not be true of all tumours. Genomic instability has been observed to be prevalent in epithelial tumours but not in blood cell cancers (Johansson, Mertens, & Mitelman, 1996)..

Mutations in special oncogenes and tumour suppressor genes as well as microenvironment selection could also cooperate with aneuploidy to promote tumour progression (A. R. A. Anderson, Weaver, Cummings, & Quaranta, 2006). Conversely other experiments with similar mouse models of chromosome missegregation have suggested that it may limit tumour growth (B. A. A. Weaver, Silk, Montagna, Verdier-Pinard, & Cleveland, 2007). Such findings help fuel the controversy but are hardly surprising in a field in which DNA damaging agents like those that are thought to cause cancer are routinely used in the clinic as cancer therapies.

Recent research suggests that mutations in checkpoint genes (genes responsible for the surveillance mechanisms that ensure faithful DNA replication and segregation during cell division) can contribute to oncogenesis. Experiments using the functional disruption of checkpoint genes such as those from the Mad (mitotic arrest deficient) or Bub (budding uninhibited by benzimadazoles) families in mouse models suggest a strong relationship between chromosome missegregation and the increased rate of carcinogenesis (Baker, Chen, & van Deursen, 2005). Mutations in these checkpoint genes appear to continually generate moderate levels of aneuploidy that are compatible with cell viability. If this is the case, subsequent clonal division could lead to the effective propagation of increasingly aneuploid genotypes through the population that are more susceptible to malignant transformation.

51

Furthermore, if genes that regulate chromosome segregation became deregulated, aneuploidy could become a main drive for cancer evolution (Cimini & Degrassi, 2005). Many recent experiments have shown that aneuploidy can occur through many different mechanisms (B. Weaver & Cleveland, 2006). Currently, RNAi screens are being carried out experimentally in well-established cell lines such as HeLa and RPE1 cells to discover the mechanisms through which aneuploidy may arise (Kittler et al., 2007), (Kwon et al., 2008)..

Aneuploidy has also been recently suggested to confer selective advantages for cells under high-stress conditions, such as the drugs used in cancer treatments (Selmecki, Dulmage, Cowen, Anderson, & Berman, 2009). Cancer therapies may act as a form of selection on cells and exert even more selective pressure on the cancer cells. Because drugs would destroy most cancer cells, those cancer cells that remain may be of a more aggressive and resistant kind and thus make the administered drugs ineffective (Rajagopalan & Lengauer, 2004). A deeper understanding of the effects of aneuploidy on cancer initiation and progression may also help us devise new cancer treatments that account for drug resistance due to aneuploidy.

2.1.5. Summary of Biological Background

There are several theories regarding the origin of cancer. A compelling theory suggests that the loss or gain of entire chromosomes (gaining or losing hundreds of genes through whole chromosome missegregation) might give rise to different kinds of viable aneuploid cells (Manchester, 1995). There are currently many different views on the role of aneuploidy, some of which give aneuploidy a central role in cancer development and others that suggest that aneuploidy is actually the outcome of oncogenesis (Hede, 2005). The change in number in many of these genes could deregulate the cell functions by changing the dosage of the gene products

and de-regulating pathways that may confer selective advantages to individual cells. However, the key properties that chromosome missegregation events have on cells, the viable evolutionary pathways for oncogenesis through aneuploidy and the properties of the resulting aneuploidy genotypes are currently unknown (Dobles, Liberal, Scott, Benezra, & Sorger, 2000).

2.2. The Modelling of Cancer

It is the hypothesis of this work that a computational model could aid in the understanding of this phenomenon. In this section it will be discussed how other researchers have modelled key aspects of cancer. Some of the most relevant mathematical and computational models will be reviewed in order to understand how it is decided which kind of approach is best for modelling a particular biological process.

2.2.1. Introduction to Cancer Modelling

Due to a recent explosion in molecular biology techniques, the exponential increase in computational power availability and the development of vast biological databases (Furney, Higgins, Ouzounis, & López-Bigas, 2006), cancer modelling has become a practical tool to understand fundamental principles and specific features in cancer (Macklin & Lowengrub, 2006). These kinds of models have been proven useful in the development of new theories, the tackling of complex interactions, as well as testing many new therapies. Many of the current techniques and concerns faced by cancer modelling have been addressed in insightful reviews such as the ones written by Nagl (Nagl, 2006), Wodarz (Wodarz & Komarova, 2005), and Deisboeck (Deisboeck, 2011) highlighting the need for new kinds of models that are able to cope with the overwhelming biological complexity. Models have emerged from different areas of science, such

as physics, mathematics, computer science, biology and medicine; beginning to converge in the more recent fields of complex systems, systems biology, systems medicine and artificial life. In this section, some of these kinds of models will be described. This section will also explore some of the most important paradigms of the field and offer an insight into why it was decided to use an agent based model to tackle the phenomenon of chromosome missegregation.

2.2.2. Spectrum of Cancer Models

It is not practical (and currently not feasible) to simulate every single aspect of cancer biology. Cancer modelling has therefore addressed particular features of cancer, employing mathematical and computational concepts and language. There is a whole spectrum of cancer models that range from computationally low level, specified models to computationally high-level, abstracted models (Ideker & Lauffenburger, 2003). At the lowest level (coincidentally, the first cancer models historically), highly specified models–models where the system is broken down in order to gain insight into its compositional sub-systems–are found. These models usually deal with the structure and the influences from one component to another. Through many of these low-level cancer models, important relationships between the different elements of tumours were elucidated. The most widely used techniques in low-level cancer models have represented the system's relationships and dynamics as a mathematical system in terms of differential equations, statistical models and Markov chains. A difficulty that most of these traditional mathematical models have to deal with is the fact that incorporating features in order to exhibit more realistic behaviour increases the difficulty of solving mathematical equations (Busenberg & Mahaffy, 1985). Because of this, it is difficult to address the biological mechanisms that underlie the causes and the key factors that determine the time course of tumour development with low-level models. Because of these important limitations, computational models, that would in some cases be complementary to the mathematical models, were also developed. From the fields of complex

54

systems, dynamical systems, and systems biology, models that address the flow of information and the connections between the different components were developed. Amongst these mid-level models, Boolean Networks and Bayesian Networks are amongst the most successful (Shmulevich, Dougherty, Kim, & Zhang, 2002), (Kampis, Karsai, & Szathmáry, 2011). Many of these models have been proven useful for describing, and sometimes even predicting the behaviour of many of the elements of cancer (Nagl, Williams, & Williamson, 2007).

Although it is possible to use the computer to model a system with mathematical equations, another interesting paradigm became possible with recent advances in computer science: piecing together elemental units to describe an emergent system. Computers have become fast enough that it is now feasible and practical to directly address the interactions between the individual components (Macal & North, 2005). This high-Level paradigm allows for expression of a rich variety of behaviour (C. Chen et al., 2007). In these kinds of models, the logical and physical structure of the components of a system can be abstractly represented. Some of the most successful models coming from the recently formed fields of Artificial Life, Systems Biology and Systems Medicine have used Cellular Automata and Agent-based models to model key aspects of cancer dynamics.

To decide between the different kinds of models that are found in this spectrum, it is important to determine first the kind of question that will be the focus of study. Depending on our knowledge of the system, one modelling paradigm will have an advantage over another. Some of the key models that have helped define this field will be reviewed.

2.2.3. Differential Equations Models

A traditional approach to mathematical modelling of biological systems by means of ordinary differential equations has been widely applied in the past decades for the analysis of the dynamics of these complex systems. Many of these offer a description of the complex system as a whole. This global system level behaviour can be directly analysed, but a downside is that these phenomenological models rarely address the particular role that individual components play in generating the global behaviour (Heidtke & Schulze-Kremer, 1998).

One of the first and most successful models of cancer was created by Burton in 1966 to investigate the rate of growth of solid tumours (Burton, 1966). It was experimentally observed that tumours showed an exponential retardation in their growth rate. At the time, there was much debate regarding the cause. One of the most interesting theories was that the formation of necrosis in the centre of the tumour reduced the active growing part of the tumour to the exterior shell (Mayneord 1932). Burton, based on this explanation, decided to model tumour growth in terms of its diffusion, setting a trend that would prevail until the mid 1980s. In his model, he examined the distribution of oxygen in an idealized spherical tumour. The blood supply (source of nutrients and growth factors) would be confined physically only to the surface of the tumour. Burton's model exhibited exponential retardation following the Gompertz growth equation, hinting at the formation of a necrotic region in the centre of a tumour. His model offered a mathematical explanation, of the experimentally observed deceleration of tumour growth. The equation as first developed by Benjamin Gompertz in 1825 (Araujo & McElwain, 2004), describes tumour size Z as a function of time t through the relationship:

$$Z(t) = ae^{be^{ct}},$$

where a is the upper asymptote or limit of the curve, b and c are negative constants for adjustment of the displacement and the scaling on the x axis respectively, and e is Euler's number. Burton's

model was successful as being one of the first models that offered both a mechanistic model of biological behaviour, as well as generating predictions that could be tested experimentally. Burton's work was extended by Greenspan in 1972 (Greenspan, 1972) by introducing differential equations that model surface tension and the disintegration of necrotic cellular debris along with the original reaction-diffusion equations, and then by Deakin in 1975 (Deakin, 1975), by incorporating a differential equation that expressed a non-uniform oxygen consumption throughout the tumour. This model enabled cancer researchers to contemplate different possibilities of tumour growth and showed a glimpse of the complexity of tumours.

Models of this kind that describe tumour-signal regulation have described the system in terms of coupled equations that express the rate of production of one element of the system as a function of the concentrations of other elements such as tumour growth with respect to nutrient availability (Macklin & Lowengrub, 2007) or the level of tissue invasion with respect to cell adhesion (Gerisch & Chaplain, 2008). While useful, this approach requires a high understanding of the relationship between the different variables of the system (Gierer, 1981). The incomplete understanding of many genetic regulatory mechanisms and the general absence of quantitative knowledge usually means that the best that can be achieved are highly idealized, qualitative results.

For the many models that use Partial Differential equations to describe tumour growth, like those of Burton, Greenspan and Deakin, a direct analytical solution is not possible in general. Numerical approximations have to be used to estimate the parameters of the system. One of the major drawbacks of modelling with differential equations is also the need to drastically change the system of equations when an extension of the model is required. Also, deterministic descriptions of cancer may fail to accurately reflect the non-linear, complex behaviour typically

57

found in cancer. Because of this, in parallel with the development of Differential Equations models, stochastic cancer models were also created.

2.2.4. Stochastic Models

During the 1970s some models of tumour growth also employed a stochastic approach. Scientists like Wette in some of the first stochastic models for solid tumour kinetics, argued that the random fluctuations were fundamental to all biological phenomena (Wette, Katz, & Rodin, 1974a). In his model, Wette focused on the probabilistic aspect of the growth kinetics of a population of cancer, highlighting the importance of understanding the behaviour of the cell population at a given time over simply measuring the average growth of the tumour (Wette, Katz, & Rodin, 1974b). Wette estimated that the probability of birth of cells in a solid tumour would be best estimated as being be proportional to the 2/3 power of the tumour cell population size and the death rate proportional to the population size; both rates being age-independent. Using the relatively simple assumption that tumours survive on nutrients that have to enter through the tumour's surface, he was able to calculate the probability of extinction (P_0) and the expected survival time of a solid tumour given different scenarios.

Liotta followed this work with the first stochastic model of metastasis, in which he proposed a mechanism for the dynamics of the process (Liotta, 1976). Based on quantitative experimental evidence from tumour transplantation experiments, he focused on estimating the entry rate of tumour cells into the blood circulation. For this, Liotta used a non-homogeneous, two-dimensional Markov process to describe experimental evidence in which that less than 0.1% of the cells released into the circulation from the tumour survived to be able to metastasize at a distant site (Liotta, 1976). The model focused on calculating the relationships between tumour clump sizes and the random variation of the populations of clumps, yielding the probability of the

generation of metastatic foci. This model was the first to provide a framework describing the metastatic process and allowed the individual assessment of key steps in metastasis and their interdependences. Liotta's model highlighted the importance of the size on the clumps by correlating directly them through a Markov process to the probability of generating metastatic foci, being one of the first models to reflect a good match with clinical observations.

As more biological data became available, more realistic stochastic models addressing complex phenomenon such as that of vessel growth and its role in angiogenesis began to emerge. One of the most successful probabilistic models of vessel growth was developed by Stokes and Lauffenburger to address the role of endothelial cells in angiogenesis (Stokes & Lauffenburger, 1991). Stokes and Lauffenburger's model focused on the random motility of endothelial cells and chemotaxis, to test the hypothesis that the behaviour of endothelial cells determined the rate of capillary growth and its structure. Their model focuses on the roles of endothelial cells' random motility and chemotaxis, to test the hypothesis that these cell behaviours are critical in determining the rate of capillary growth and the fractal structure. The model starts with vessels branching off from pre-existing "parent" vessels. The authors assume that a source of angiogenic stimulus is located a distance away from the existing vasculature to which the buds are attracted. Cell proliferation, migration (both random and due to chemotaxis) and budding of new sprouts off newly formed vessels are all behaviours governed by a series of stochastic ordinary differential equations that cannot be solved analytically, but can be simulated in a computer. The key insights gained from this kind of stochastic model are those gained by the adjustment of the different probabilities, simulating key scenarios. A major prediction of this model was that a moderate chemotactic response would be necessary to provide directed growth of the vasculature, but that excessive levels of chemotaxis would not be useful or would be even detrimental to this process because a vascular network would not form adequately. Many of these insightful

predictions were later corroborated to be experimentally valid under certain conditions (Stokes & Lauffenburger, 1991).

2.2.5. Markov Chain Models

As the models of cancer started describing the cellular and the molecular behaviours of the system, the randomness within the probability of biological processes was recognized to play an important role in the behavioural dynamics. Because of this, the use of Markov chains for the description of cancer dynamics has helped significantly to bring models closer to reality, enabling the discovery of key theoretical relationships. A Markov chain is a system that undergoes transitions from one state to another (Rabiner, 1999). A system modelled through Markov chain has a finite or countable number of possible states where the next state depends only on the current state and not on previous states. Markov chains allow for the probabilistic description of a process of production, loss and interchange among the elements of the systems. One such process in cancer was recently described by Alarcón and Page. The authors developed a stochastic model of VEGF (Vascular Endothelial Growth Factor) signalling in endothelial cells (Alarcón & Page, 2006) to investigate the role that the overexpression of VEGFRs (Vascular Endothelial Growth Factor Receptors) in endothelial cells of tumour vessels on angiogenesis. With this model the authors were able to show that, counter intuitively, the overexpression of VEGFRs actually increases their sensitivity to low levels of VEGF. Their model was formulated in terms of a Markov process given by a master equation:

$$\frac{\partial \Psi(\boldsymbol{X}, t)}{\partial t} = \sum_{\boldsymbol{r}} (W(\boldsymbol{X} - \boldsymbol{r}, \boldsymbol{r}, t) \Psi(\boldsymbol{X} - \boldsymbol{r}, t))$$
$$- W(\boldsymbol{X}, \boldsymbol{r}, t) \Psi(\boldsymbol{X}, t).$$

This stochastic model was specified in terms of three quantities: *X*, *W* and *r*. The state vector *X* represents the number of molecules of each of the species involved. *W* is the probability per unit time corresponding to each of the reactions involved in the process being modelled, and *r* represents increments in the number of molecules (of a given kind, depending on the scenario) when each reaction occurs. This stochastic model describes the probability density of the system being in state *X* at time *t*.

With their stochastic model, the authors have explored the angiogenic roles of a ligand/receptor binding, ligand- induced receptor activation, and the binding of enzymes members of the Src tyrosine kinase family to activated receptors (all early stage processes of angiogenesis) (Alarcón & Page, 2006). The key discovery of this model, however, was made through the modelling of the effect of overexpression of surface VEGFR on anti-angiogenic therapy. Paige discovered that the effect of receptor overexpression by inhibition of endocytosis or by up-regulation of receptor synthesis led to a substantially increased resistance to treatment with an anti-VEGF drug. They suggested that even the transient activation of VEGF receptor can eventually be replaced by a slower and more sustained response, wich leads to an increased sensitivity to low values of the concentration of VEGF (Alarcón & Page, 2006).

A key contribution of the model is that, through it, the authors provided plausible mechanisms for resistance to anti-angiogenic therapy. However, there is an important concern: their equations quickly became too difficult to solve for this complex phenomenon. Because of this, an approximate solution had to be found. They employed the Wentzel–Kramers–Brillouin approximation (March & Plaskett, 1956) (traditionally used for quantum mechanic probabilities), as a means to analyse their scenarios. This forced the authors to introduce additional requirements

and oversimplifications (such as the rate of receptor degradation being precisely balanced by receptor synthesis) that could have an impact on the final results.

As has been shown, when the capture of more detail of the behaviour of complex biological systems with mathematical equations is attempted, these equations become exponentially complicated and ultimately unsolvable to the point where extreme idealizations and limits have to be imposed. A different, but useful approach to these kinds of limitations has been to model the complex system, not as a series of mathematical relations, but as a network.

2.2.6. Boolean Network Models

Network models started with Euler in 1736; however, it wasn't until stochastic methods were introduced by Erdős and Rényi that the real versatility of networks became apparent (Barabási, 2003). By representing a complex system as a network, it is possible to examine the interconnections among diverse elements of the system, at different levels of abstraction. Although networks had been more extensively used in an economic and social context, as well as means to model the Internet, it has been through the framework established by Stewart Kauffman (Kauffman, 1969) that Network models have been applied in a biological context.

Boolean networks consist of a set of Boolean variables, whose state is determined by other variables in the network. Kauffman used this to model genetic regulatory networks. In his model, genes are represented as binary-state nodes with K inputs to each node (representing regulatory mechanisms). Genes are able to carry out one of the possible Boolean functions of its K inputs. The two states *1,0* (on/off) represent respectively, the status of a gene being expressed or inactive. Time t is viewed as discrete steps. At each step, the new state of a gene is a Boolean function of the prior states of the K input nodes. If the activity of a gene at any time is 1, then the value of all

its output lines at time *t+1* is simultaneously *1*. Thus, the state of the outputs of a gene at *t+1* depends on its activity at time *t* alone.

With this simple model, Kaufmann constructed large networks of randomly interconnected binary on-off genes. He studied the behaviour that these networks displayed, as well as key properties. Kaufmann arrived at the surprising conclusion that randomly constructed networks (studying the case in which each element is directly affected by two others) undergo short stable behaviour cycles in the states of their constituents. He then proposed that the time course of these behaviour cycles parallels (and in some cases, predicts) the time required for cell replication in many phyla (depending on the size of their DNA). Kauffman argued that this kind of stability is essential for life. Because these kinds of models had initially to be worked out by hand, as more and more modellers began to understand the different programing languages and the computer as a useful tool to work out all the numerical computations, more models were developed. One such model comes from the field of system biology (Sahin et al., 2009).

With this Boolean network framework, Sahin el al. developed a Boolean model that linked protein ERBB2 to a key phosphorylated retinoblastoma protein (a surrogate for cell cycle progression) in breast cancers that are resistant to ErbB2-targeted inhibitors (like cancer drug trastuzumab). They modelled regulatory interactions as Boolean nodes and simulated single and multiple protein loss-of-functions as Boolean operators. This model was able to model protein interactions, which were later validated experimentally, and predict the results from knocking down proteins in the network (Sahin et al., 2009). The authors focused on two major cell signalling pathways and two key transcription factors. They were able to model a sufficiently abstract interaction network that allowed for the identification of novel targets in the treatment of trastuzumab-resistant breast cancer. A key result from this model was the elucidation that ERBB

63

receptors alone would not be able to halt cell cycle progression in cells that had acquired resistance. The authors, based on their simulations, instead suggested C-MYC as a potential alternative target (Sahin et al., 2009).

Although this methodology is very promising, there are few cancer models that take advantage of this framework. A common criticism of these models is that they can only be used at the genetic level (unless coupled with another kind of model), because they usually do not take into account spatial differences and thus may not provide us with a key feature of biological systems like the differences in local and global characteristics crucial to evolution and selection (M. Bissell, Rizki, & Mian, 2003). The effects that a microenvironment has with respect to cells and their gene networks may generate unexpected behaviour that is not addressed by these models. They may be useful however to understand how some of the key dynamics of the system work. Nevertheless, it has recently been shown that Boolean logic can be used to model cell types in protein signalling networks (Saez-Rodriguez et al., 2009). From a combination of interactome databases and with empirical data, a transduction of signalling networks into logical models can be derived. If combined with cancer disruption databases, Boolean Networks could be a promising cancer-modelling paradigm for future research.

2.2.7. Bayesian Network Models

In cancer modelling, several aspects of network theory have been used to tackle important questions by modelling the physical space (such as tumours and arteries) and the flow of cancer cells as a physical network with different degrees of connectivity (Barabasi, 2005). From these different aspects of network science, Bayesian networks are beginning to play a more prominent role in cancer modelling. Bayesian networks describe pathway interactions (direct or indirect) in terms of the probability that a certain component will affect other components. The Bayesian

network approach towards biological modelling can be especially compelling because of its solid basis in statistics. Because of this, Bayesian networks can be very useful when only incomplete information about the system is available (Endy & Brent, 2001). This is particularly useful when modelling seemingly stochastic aspects like gene expression and experimental noise. Another very interesting feature of these models is the way in which genes and their interdependencies can be represented in intuitive graphs, although again, this approach does not deal explicitly with the dynamical aspects of gene regulation and evolution.

Bayesian network modellers have focused on genes and their interactions. Informed by the recent advances in bioinformatics, these kinds of models have been used to represent the interplay between key genes, yielding interesting results, like the discovery of causal relationships, signalling cascades and feedback loops in those networks. One of the most effective models was developed by Nagl et al. (Nagl et al., 2007). The authors were able to reverse engineered a Bayesian network from a clinical dataset, calculating the probability of the linking of the different variables. The authors, focusing on breast cancer, used three sources of data: one clinical and two genomic. The clinical data came from a subset of the American Surveillance, Epidemiology and End Results (SEER) study (over 3 million patients). The large study offers data on both cancer diagnosis and survival in the USA between 1975 and 2003. The authors also used two karyotype datasets from the progenetix database (www.progenetix.de), consisting of a compilation of published data from comparative genome hybridisation (CGH), array CGH, and matrix CGH experiments. The authors determined the conditional independencies through an algorithm for maximising the entropy of the probability distributions. As a result, the authors were able to examine the influence of karyotype pattern on clinical parameters such as tumour size, grade, receptor status and the likelihood of lymph node involvement (Nagl, Williams, El-Mehidi, Patkar,

& Williamson, 2006). Interestingly, their model allowed examining the influence of those parameters on karyotype pattern in an iterative manner.

2.2.8. Cellular Automata Models

From the early days of computing, another kind of model has been slowly but steadily becoming a part of the arsenal of modellers: cellular automata. Cellular automata are a class of spatially and temporally discrete dynamical systems. (Wolfram, 2002). They can incorporate many features of self-organizing, complex systems, and are able to exhibit emergent behaviour (as can be appreciated from Conway's Game of Life). Cancer models that rely on cellular automata consider the simulation of a biological process that abstracts known biological facts as some conditions set down in rules of interaction (Ribba et al., 2004). Actions in the model are carried out by these predefined rules whose conditions may or may not be satisfied by a given situation. The future state of a cell could, for instance, be defined by the state of its neighbours, depending on a certain configuration or rule; an abstraction mimicking the way cells in real organisms can regulate each other. Depending on the kind of rules, this kind of modelling can yield a deterministic outcome, as demonstrated *in silico* experiments where the development of new vasculature by means of angiogenesis is successfully addressed (Merks & Glazier, 2006), or a stochastic one (McDougall, Anderson, Chaplain, & Sherratt, 2002).

One of the earliest Cellular Automata models of cancer was developed by Düchting et al. to address various aspects of pre-angiogenic tumour development (Düchting & Dehl, 1980). The goal of the model was to study the regulation of disturbed cell renewal, through the analysis of two competing populations of cells (truly one of the first proper *in silico* experiments). The cellular automata used a two-dimensional regular 10×10 square lattice and considered a von Neumann neighbourhood. In the CA model, each lattice site corresponds to a biological cell in a

discrete state (if a cell dies the lattice site becomes empty). The transition rules are simple, deterministic and local: On each time step, a cell can survive, divide or die with an asynchronous update.

Düchting's model simulates an unrealistically low number of cells within the lattice. Two clones of cells (with normal and fast growth) can coexist in competition with different scenarios included the simulation of normal growth as well as malignant growth. The results were not validated at the time because of general lack of experimental data, but have since proven to be simple to describe avascular growth (Roose et al., 2007). The model suffered primarily from the limited computational power existing at that time. The key ideas proposed in this model, however, have had a strong impact on all the Cellular Automata models that followed.

Depending on the kind of behaviour one wishes to study, it is possible to abstract enough information in terms of cellular automata to carry out *in silico* experiments. An important example of this can be appreciated in the more contemporary model proposed by Anderson for the investigation of the geometry of tumour growths (A. R. A. Anderson, 2005).

Cellular Automaton have also been successfully used to model complex cancer phenomena such as angiogenesis. An important CA model that explored the mechanism of cell migration was proposed by Chaplain and Anderson (A. R. A. Anderson & Chaplain, 2012). In their model, the authors explored the hypothesis that cell migration was influenced by gradients of tumour angiogenic factors (such as chemoattractants secreted by tumour cells) and fibronectin (a component of the Extra Cellular Matrix that enhances cell adhesion). Their model consisted of a two-dimensional regular lattice with 200 × 200 sites, modelling a tissue. Only two discrete states were considered: whether cells were present or absent. For this model, the authors considered a

von Neumann neighbourhood with a synchronous update. Experiments were carried out with a continuous field of cell density, tumour angiogenic factors and fibronectin; following probabilistic migration rules and deterministic branching rules. In their model, branching depends on the state of the neighbours (whether they are responding to angiogenic factors or fibronectin), and whether there is space available. The continuous fields of cells, angiogenic factors and fibronectin are described by coupled partial differential equations. Once these equations are estimated for a given configuration of the lattice, five coefficients (proportional to the cell density at each site and its neighbours) are calculated. The probability of migration is calculated proportionally to these coefficients. After some parameter recalibration, the authors were able to successfully reproduce results from reported experiments of solid tumours in animals.

A delicate issue with these kinds of model systems however is that of a devising an appropriate control strategy, which should be established to determine the order and hierarchy in which the rules are to be evaluated. Although this is mainly implemented to resolve the conflicts that arise when several rules match the given conditions at the same time, as can be seen in some of the *in silico* experiments carried out to investigate the evolution of homeostasis in artificial organisms (Basanta, Miodownik, & Baum, 2008), there is a potential risk of implicitly defining the system's behaviour. Many difficulties may also arise when trying to maintain consistency in the overall model when adding new biological knowledge and quantitative information (Gerlee & Anderson, 2007).

Some cellular automata models have successfully addressed situations in which the environments of interest are not spatially homogeneous. This kind of simulation can yield interesting, feature-rich models appropriate for the study of the tumour growth in physically restricted environments (Gevertz, Gillies, & Torquato, 2008), or even to address important cell-cell interactions (Ghaemi

& Shahrokhi, 2006). The main advantage of these kinds of models is that the spatial distribution and other important physical features can be explicitly represented in simulations of the interaction of individuals that change in time and space (Hogeweg, 2002).

Other cellular automata approaches consist of tissue modelling, as proposed by Merks and Glazier through their cell modelling software CompuCell3D (Merks & Glazier, 2005). By making a generalization of the cellular automata, Merks and Glazier have defined interesting active parameters that simulate the behaviour of individual cells in time. One difficulty with this approach however is the process of distilling useful information from the considerable amount of data offered by a simulation of this magnitude (Mahoney et al., 2008).

A criticism of Cellular Automata models such as those of Anderson et al. is the assumption of diffusive cell motion on all timescales, and that other cellular phenomena such as cell-cell adhesion and cell-ECM interactions are described stochastically (Jeon 2010). This may impact the inclusion of any abstraction of mechanistic cell behaviour that the researcher may want to include. To tackle this, Jeon el al. propose an off-lattice approach to model tumour growth and invasion (Jeon 2010). In their work, they combine a continuum part that describes micro-environmental components (nutrients, oxygen, enzymes and other molecular components), and a discrete part, which represents individual cell behaviour (cell cycle, cellular interactions, and the interactions of cells with the Extracellular matrix). Their model is able to incorporate an experimentally measured supper diffusive behaviour of mammalian cells during motion/migration on short timescales. Their model shows yet another facet of tumour development; namely that cell-cell adhesion and cell motility strongly affect both tumour growth and its morphology, generating finger-like shapes characteristic of invasive tumour.

As many of the off-lattice strengths are disadvantages in square lattice Cellular Automata models and vice versa, Gallaher et al. have recently tried to bridge some of these modelling paradigms to bring balance and robustness to their research (Gallaher 2013). In their study of the evolution of intra-tumoral phenotypic heterogeneity, the researchers focused on investigating the role of trait inheritance in cells. Taking advantage of the less restrictive motion that off-lattice models provide, she considers individual cells with assigned trait values for proliferation and migration rates. Heterogeneity in these cells is modelled with frequency distributions and combinations of traits with density maps. With this off lattice model, the researchers have developed a platform in which heterogeneity changes over time, depending on how traits are passed on to progeny cells. The researchers discuss how the spatial and temporal evolution of phenotypic heterogeneity may have significant implications for the treatment of tumours and highlight how samples taken in the clinic may not account for the entire heterogeneity of the tumour.

2.2.9. Agent-Based Models

Agent-based models have been proven efficient tools to simulate evolution in complex systems. An agent in a computer simulation is considered to be an identifiable, self-contained individual with a set of characteristics and rules governing its behaviours and decision-making capability (Macal & North, 2005). An agent has the ability to learn and adapt its behaviours based on experience, which can be stored in an agent's internal memory. An agent may contain both base-level rules for overall initial behaviour as well as a higher-level set of rules that allow the agent to modify the base rules and thus generate novel behaviour (C. Chen et al., 2008).

Making use of this paradigm, many agent-based models have considered agents to represent biological cells. Some of these models may also consider cells to be in discrete cellular states (as in Cellular Automata), for example, normal, necrotic, quiescent, or proliferating state, depending

on the questions the model is trying to answer (Roose et al., 2007). An interesting result of these kinds of models is the successful integration of the interaction between cells with the effect that nutrient concentrations and physical constraints have on a tumour (C. Chen et al., 2007). Agent-based models have been quite versatile in describing different kinds of very specific biological features (Ribba et al., 2004).

Spencer and Forrest used an agent-based model to address key aspects of somatic evolution in tumourigenesis (Spencer et al., 2006). The authors focused on the evolutionary dynamics that govern the "Hallmarks of Cancer", discussed in Chapter 2.2. For this purpose, they developed a three-dimensional 100x100x100 grid, containing a maximum of 10^6 cells initialized with a single normal cell and a single blood source. A small set of underlying rules governs the transformation of normal cells to tumour cells, implemented in a stochastic multistep model. In their model, normal cells that proliferate far from the blood supply signal for neovasculatization (a key step in angiogenesis) to acquire the resources they need to divide. The blood supply then creates a branching pattern of neovascularization, extending in the direction of the requesting cells. Normal cells are able to signal for neovascularization until they reach a limit (an angiogenic boundary) beyond which they are not able to signal, and are able to divide only within the region containing supplied growth factors. The modellers focused on the biological phenomenon of telomere shortening, a cellular mechanism that limits the number of times that a cell can be copied (eliminating cells once their telomere length reaches zero), thus preventing the accumulation of mutations.

The simulation of the model is as follows: a normal cell will initiate cell division if there is unoccupied adjacent space. If dividing, the "genome" of the parent cell is copied with a small probability of mutations that enable the hallmarks of cancer, the telomere length of both cells is

decreased by a single unit, and the new daughter cell fills the empty neighbouring space. There is a small probability of random death and a probability of death associated with acquired mutations. This process continues until either 50,000 time steps have passed or the tissue has developed cancer-like behaviour (by filling more than 90% of the grid), as defined by the Hallmarks of Cancer.

The researchers focused on how the sequence of acquired mutations affects the timing and cellular makeup of the resulting tumour and how the cellular-level population dynamics drive neoplastic evolution. Their model makes interesting predictions. Key amongst those are that early-onset cancers may go through a different sequence in the acquisition of mutations than late-onset cancers; that tumour heterogeneity varies with acquisition of genetic instability and the mutation acquisition pathway; and that there may exist an optimal initial telomere length which lowers cancer incidence and raises time of cancer onset.

It is interesting to see the dynamical trade-offs that may occur between early incidence and late incidence of cancer such as the interplay between clonal expansions (in the presence of blood supply) and competition between cells with selective advantages. The model raises key questions still unanswered by cell biology, such as, what is the precise role of telomere length in cancer? Genetic instability was modelled as an enabling characteristic of cancer and thus, often one of the first mutations acquired on the early onset tumours. However, the precise role that this genetic instability may have is not completely explored. Because this is a controversial issue in cancer biology, the authors propose that this question be answered empirically (Abbott et al., 2006).

Others have successfully addressed other key dynamics and properties of cancer systems with agents with the right kind of abstractions. Bentley demonstrated the usefulness and versatility of

agents for the purpose of modelling of angiogenesis, providing a new mechanistic view of vascular patterning as well (K. Bentley et al., 2008). Agent-based models for the evolution of entire cancer ecosystems have also been proposed (Gerlee & Anderson, 2007). Although highly ambitious, an agent-based model that can address both the global evolution and the individual dynamics of tumours could be used to develop a more general model of cancer.

2.2.10. Modelling Chromosome Instability

In the past decade, chromosome instability and the role that it plays in cancer has become a subject of interest for cancer modellers. Each of the models has explored an interesting feature of aneuploidy. Amongst the most prominent of the models that addresses this phenomenon, the one developed by Komarova et al., which focused on colon cancer, explored the role that chromosome instability could have if it were an early event in cancer progression (Komarova et al., 2003). The authors use an evolutionary lens to tackle important questions regarding the conditions for chromosomal instability to precede the loss or inactivation of the first tumour suppressor gene. The researchers modelled the evolution of a homogeneous population of cells with respect of the acquisition of mutations in time as a stochastic process that considers a linear Kolmogorov forward equation. By solving this equation analytically, the researchers were able to elucidate the way in which homogeneous states are connected by 'stochastic tunnels'. These tunnels occurred for certain parameter values that allowed for a jump from one cell type to a highly mutated cell type by skipping the normal intermediate mutation. The researchers were able to quantified the rate of tunnelling and their results suggest that the likelihood of chromosome instability being an early mutation depends on the kind of cell (mutations in stem cells having more impact as heir progeny inherits the mutation), the rate of point mutations that would lead to loss of tumour suppression in cells (assuming tumour suppression after both copies of the gene

73

are lost) and the number of genes that regulate chromosome instability.

Solé et al explore in their model the possibility that there may be a threshold in the levels of genetic instability, assumed to be costly to cancer cells, in order to overcome selection barriers (Solé et al., 2004). In their work they present a quasi-species model of competition between cancer cell populations, and how these dynamics affect tumour growth. Their model suggests that once this threshold is reached, the tumour cell populations become too unstable to maintain their genetic information, and thus becomes detrimental to its growth. They considered a tumour cell population that exhibits three degrees of an increasing genetic instability phenotype. They also considered that transitions from one stage to the other are given by different mutation rates for each degree of chromosomal instability. Furthermore, they assume that high genetic instability has antagonistic effects on the fitness of the genetically unstable cell populations. In their model, when genetic instability exceeds a certain threshold, the replication rate of the more malignant subpopulation offers no competitive advantage anymore and in some cases, become a detriment to the cell population.

Michor et al. propose in their model that chromosomal instability is a major drive for the loss of heterozygosity; the initial loss of a copy of a tumour suppressor gene (Michor et al., 2005). Following the two hit hypothesis, the loss of the second copy would inactivate tumour suppression for that gene. They consider a mutational network of inactivating two tumour suppressor genes, where one node is separated from other nodes through paths describing mutations in tumour suppressor genes as a stochastic process. The authors investigate a path to tumourigenesis via chromosomal instability before the loss of heterozygosity. They explored a

number of plausible parameters and propose that in any pathway of cancer progression where at least two tumour suppressor genes need to be eliminated in rate limiting steps, chromosomal instability is likely to be an initial event. As is the case in Komarova's model, Michor's model suggests that the existence of early chromosomal instability depends on the number of genes that regulate this behaviour. This however would depend on the cost associated with the activation of chromosomal instability.

These models of chromosomal instability discuss situations where inactivation of one or two tumour suppressor genes is required for tumourigenesis. However, due to necessity, they do not incorporate the effects that oncogene activation may have on the evolution and development of the tumour. The inclusion of genes that regulate proliferation, in tandem with chromosome instability, may playa key role in the evolution of cancer and give rise to unintuitive behaviours. These models consider an abstraction of chromosome instability that does not include a mechanistic process through which chromosomal instability occurs. The genes that regulate chromosomal instability are abstracted in rates that change by means of a governing equation. However, it is important to acknowledge that there are many kinds of viable cells that may result from chromosomal instability. Some cells may have more chromosomes, affecting the number of genes and in some cases even countering a loss of heterozygosity. Also, because some of these models describe the process mathematically through compartments, information about the precise genetic structure, the different sized of the population and the evolution of novel genotypes that are not accounted for in the initial mathematical description is lacking. A mechanistic process that underlies the simulations may shed some light on elusive, but important events that happen at the cellular level that may have an impact on the development of new kinds of phenotypes and the emergence of different behaviours.

75

2.2.11. Summary on Cancer Modelling

In the past few decades there has been an explosion in cancer modelling techniques due to the advances in biological techniques, the development of complex mathematical and computational frameworks and the increasing availability of computational power. There is a whole spectrum of cancer models that goes from highly specified, low-level models such as differential equations and stochastic models; to the abstract, high-level models like the cellular automata and agent-based models. In this section, the spectrum of cancer models–from low to high level–has been explored and some of the key models that have shaped the field have been reviewed.

It has been shown how low-level models have dealt with the structure and the influences of variables over the components of the system. In the middle of the modelling spectrum, Networks and how they address the flow of information and the way the elements of the system are connected. Finally, the high-level end of the spectrum, where the biological mechanisms and the components themselves were the subject of interest, was addressed. This chapter highlights the need of models that are able to adequately put into context what it is known, and what we wish to know about cancer. The decision of each group of researchers to use a given kind of model, or a hybrid of two or more of the reviewed categories, to describe a particular feature of cancer was determined by the kind of question they were addressing. The key result of this review is the absence of a suitable model for the phenomenon of aneuploidy and its effects on cancer. This gap in our knowledge is result of a combination of technological limitations and lack of data until recently.

2.3. Conclusions

In this chapter, it was explained why the role of aneuploidy in cancer is a very controversial topic and that models are needed to shed some much needed light on the subject. To model the most recent biological knowledge of the mechanics of cell division, our review suggests that an abstract model of behaviour such as an agent-based model would be appropriate.

In this chapter, the complex relationship between the theory, experiment and clinical reality in cancer research was explored. Key research on cancer biology and summarized some of the most important frameworks on cancer modelling was highlighted. By taking both into account, a solid background for the understanding of the research problem addressed in this thesis was provided, and evidence in support for the hypothesis proposed was offered.

Throughout the years, cancer modelling has played a key role in the elucidation of a general framework and terminology for cancer researchers to use. Cancer modelling has become one of our primary sources of understanding regarding the origins, the evolution and the fundamental principles of life. As important pieces of knowledge are being discovered in the laboratory and in the clinic, modelling has helped to place them in context. This in turn has enabled other cancer researchers to develop new theories, gain insights and devise new therapies. Cancer research, both biological and theoretical has helped us to gain a deeper understanding, not only about this terrible disease, but also about life itself.

3. Research Methodology

This chapter describes the research methodology used in this work, and the tools used to address and evaluate the research hypothesis. In order to provide evidence to support the hypothesis that an integrated computational model of chromosome missegregation during cell division provides an effective approach to assess the role of aneuploidy in the initiation and evolution of cancer, a computational model needs to be developed. The process of cancer modelling incorporates concepts from mathematics, computer science and cancer biology to simulate and analyse biological behaviour *in silico*. This chapter begins with an exploration of the computational modelling methodology that will be used through this process. The next section examines the process that was used in the assessment of the model. This is followed by a description of the biological modelling methodology that will be followed. In the final section, a discussion on the selection of modelling paradigm will be presented, ending with an evaluation and a summary of this chapter.

3.1. Introduction

The previous chapter explored the controversy surrounding the role of aneuploidy in cancer, and highlighted a knowledge gap in the field. This work seeks to fill that gap with the creation of a computational tool to assess the role of aneuploidy in cancer with a rigorous methodology. To create this tool, an investigation of the most recent cancer theories and modelling paradigms was carried out. This chapter describes four aspects of methodology for this work: how models should be constructed in a rigorous way, how the models can be assessed, how the appropriate biological components and mechanisms can be determined, and how to determine which modelling paradigm is most appropriate.

3.2. Computational Modelling Methodology

Computational cancer models represent information of biological areas of interest. The right assumptions and simplifications have to be made in order to make a useful model of the known and theorized biological properties, components and relationships. The main problems that this kind of biological modelling has faced are computational constraints and the lack of data (P. J. Bentley, 2009). In tandem, these limitations sometimes make it infeasible to accurately model key aspects of biology. Fortunately, the best models are not always the most accurate ones. Simplifications and abstractions can help cope with both computational constraints and lack of data.

Although many new fields in the interface between computer science and the biological sciences are dedicated to the simulation and modelling of biological processes, there is no universally agreed methodology for abstracting properties and building a computational model. Although there are many ways to design and implement models, two modelling methodologies stand out: top-down and bottom-up modelling (Silverston et al., 1997).

A top-down approach attempts to break down of a system to gain insight into its compositional elements (Minsky, 1974), (Silverston et al., 1997). In this paradigm, which dates from the births of computers and AI, there is a substantial analysis and decomposition of the major level into sub levels (Minsky, 1961). Each sublevel can be decomposed further until the base elements are reached. A bottom-up approach is the piecing together of constitutional elements to give rise to the system (Brooks, 1990). In this approach the base elements of the system are first specified and then linked together to synthesise a larger system, where the resultant behaviour of that system is

emergent and unpredictable, and thus is not predefined by the developer of the model. This can also be done on several levels, making the original set of elements a subsystem of a greater emergent system.

Both paradigms require the use of abstractions and simplifications and they have the risk of simplifying elements and interaction in an inappropriate way. For the model to work as a virtual lab in which to test hypotheses, the abstractions have to capture the essence of the biological process while ensuring that what is modelled is supported by good data and what is not modelled will not impact the results in a significant way. Considerable analysis, scientific debate and creativity are usually present in the early stages of the model design and implementation. It is in this critical step that poor abstractions can be made, which will impact the ability of the model to provide meaningful or significant results at later stages. Key questions need answering (Le Martelot & Bentley, 2009):

- What are the key biological elements that will be modelled?
- How do the elements interact?
- What is the context in which this interaction takes place, and how does this affect the interaction?
- How should the abstractions of these elements and interactions be designed and implemented?

Like any software engineering task, the creation of models may follow several methods, such as:

- **Waterfall**: a linear design process, in which progress is seen as flowing steadily downwards from conception to construction and then implementation (Royce, 1970). This highly structured kind of modelling, however, leaves little room for changes after a process has been completed.

- **Prototyping**: an iterative framework where only a few aspects (sometimes as an incomplete version) of the system are simulated (Grimm, 1998). Although this kind of modelling needs more time to become fully functional than the waterfall, one of the main benefits is that it is possible to incorporate feedback early on.

- **Incremental**: a combined linear-iterative framework where the model is designed, implemented, and then, based on feedback, it is augmented to provide more functionality (Pressman, 2009). This is done iteratively until a final goal is reached. One of the main difficulties in this approach is that bringing new functionality to the model may conflict with the original model architecture. Because of this, a more rigorous testing, and thus development time, is required (Pressman, 2009).

- **Spiral**: a linear-iterative framework that combines the features of the prototyping and the waterfall model (Boehm, 1988). The method resembles a spiral in it's continuous structuring: at the beginning of the "spiral", the objectives, alternatives, and constraints on the new iteration of the model are determined. Alternatives and risks are evaluated and, from this outcome, the model is further developed and the next iteration is planed. This is a good approach for large-scale projects. (Boehm, 1988)

- **Rapid Application Development**: an iterative framework that merges various structured techniques with prototyping techniques (software re-use that leads to a fast succession of prototypes) to accelerate software systems development (Whitten et al., 2004). The large number of prototypes, however, can be overwhelming and suffer the risk of not being fully understood before moving to the next prototype (Whitten et al., 2004).

- **Agile Software Development**: an iterative framework which break tasks into small increments with minimal planning (Highsmith & Cockburn, 2001). The developing takes place in short intervals, resulting in "miniature software projects" and resulting in a release version (Beck et al., 2001). Agile modelling requires continuous feedback and is best applied to small projects. One of the main difficulties with this kind of modelling is defining when the final model has been reached (Highsmith & Cockburn, 2001).

All of these imply the use of iterated evaluation to help improve the model. However, it is vital that models are not designed to produce the results that the researcher is hoping to see. This is one advantage of the bottom up methodology, where the overall behaviour is not predetermined.

Nevertheless, it is still possible for a researcher to iteratively modify a model until it produces the desired result. Ideally then the model should be verified and validated on one set of data, which is separate and distinct from the data used to evaluate the results from final experiments by the model (Wodarz & Komarova, 2005).

For this kind of biological modelling, maintaining the balance between biological accuracy and computational feasibility from the early designs is key (Wodarz & Komarova, 2005). Because of the availability of the expertise of wet-lab biologists and experienced computer scientists (a crucial part of the design and implementation process); the time and computational limitations; and the necessity to understand each version of the model before adding more functionality, it was decided that incremental modelling is the appropriate methodology for this project.

3.3. Model Assessment Methodology

As we have seen in the previous chapter, there are many ways to model key aspects of cancer. Different questions require different approaches but a consistent methodology on how can we model cancer has not been established. There is, however, another important problem in mathematical modelling: model checking. Given a model of a system, how to check whether this model meets a given specification? There are 2 main approaches for model checking: 1) The model is translated into a formal mathematical framework and its correctness is demonstrated formally; 2) The model is refined and iterated until the correct answer is verified and validated. Both approaches (and combinations of them) have been used with varying degrees of success.

In order to check a computational model formally, both the model of the system and the specification could be formulated in a formal system: a well-defined mathematical system of

abstract and logic thought. Amongst the most rigorous of these formal systems there are two kinds of calculus: λ-calculus and π- calculus. λ-calculus is the core language of functional computation; representing all of the elements of the system as functions and all the interactions as the application of functions. π- calculus, on the other hand, is the core calculus of concurrent systems; representing the elements as processes and all the computation carried out as causal connections or communication exchanges on flow channels. In the context of biological models, a process could be a biological state for which changes in time, or events, are observed. These changes in time are described in terms of information flow between processes. Although both calculi would be impractical to use as a source-level programming language (their minimalism and the lack of primitives such as numbers, data structures, variables, functions, booleans, etc., prevents us from writing programs in the normal sense), they could be used as formal frameworks for a modelling language. The λ-calculus formalism can be applied as a modelling language for sequential processes, and π-calculus formalism for concurrent processes.

Although λ-calculus has yet to be embraced by the biological modelling community (Szathmary, 1995), π-calculus has been used with success to validate models for bio-molecular processes such as the signalling RTK-MAPK transduction pathway (Regev et al., 2001), theoretical protein interactions (Danos & Laneve, 2004), and models of theoretical molecular processes that account for spatial distribution (John et al., 2008). A special instance of π-calculus, Spaceπ, has also been used to fully model and validate theoretical biological processes within cellular compartments (Regev et al., 2004). Criticism of this kind of formal methods model checking comes from the fact that they can only be applied within the context of finite-state systems; otherwise, there is a risk of reaching indeterminacy in computation due to the mathematical limitations of logic programming and possible lack of biological information (Wing, 2002). Bentley has recently proposed systemic computation as a tool that provides a formal language for

describing hypothetical biological phenomenon (P. J. Bentley, 2009). Although this approach could make formal biological modelling more intuitive than Pi calculus, it may be slow when carrying out conventional serial computations (P. J. Bentley, 2009).

An alternative to this is model checking by verification and validation. This kind of model checking includes having the model checked by experts and examining the model output for reasonableness under a variety of settings of the input parameters. Under this second paradigm, Machado, from the field of systems biology, argues for a unifying formalism able to assess the representations of the structure and functionality of cellular processes. He proposes that a biological model could be assessed in five areas:

1. Integrative: The number of different biological processes being incorporated.
2. Intuitive: A metric for the clarity of the model based on the number of people from different disciplines that can interpret the model.
3. Scalable: The number of different levels that are abstracted.
4. Qualitative: An abstract measure of the structural and topological properties that may emerge from the complex system.
5. Quantitative: A measurement of key properties of cellular behaviour and the response of the system to perturbations.

Machado's formalism focuses on systems biology models, which describe biological processes as an integrative whole (Machado, et al., 2009). This methodology however comes with some caveats. More metrics are needed in order to obtain results that can be applied to cancer research: How does the model deal with incomplete information? How general is the model? How can we measure the usefulness of the extracted information from the model? How well does it represent or can predict behaviour? An attempt to answer some of these questions comes from Webb in the field of complex systems. Webb suggests seven dimensions on which models can be assessed (B. Webb, 2002):

1. Relevance: Whether the model is applicable to biology by addressing new hypotheses or describing observed behaviour.
2. Level: The elemental units of the model in the hierarchy. Deciding the appropriate levels to represent is problem specific and could range from subatomic particles to cells to societies. It is assessed in terms of whether the detail in the elemental units provides a better understanding of the system.
3. Generality: The range of biological systems the model can represent.
4. Abstraction: A gauging of the amount of detail that is included and needed in the model, from very specific instantiations to very abstract concepts.
5. Accuracy: An estimation of the accuracy of the actual biological mechanism represented that give rise to the observed behaviour.
6. Match: An appraisal of the extent in which the behaviour obtained in the model matches the real biological behaviour.
7. Medium: An assessment of the constraints that the actual build-up of the model has on its implementation.

This kind of model checking approach can be useful for keeping models in development on the right direction, with the caveats that the verification between each refinement and iteration comes from experts in the field and the validation with predictions made with the model (Wodarz & Komarova, 2005).

The hypothesis of this work deals with a non-finite state system, which evolves in time. Because of this, and because there is incomplete information on the system, it may be impractical to assess the model formally. Also, because of the availability of a close collaboration with molecular biologists and cancer researchers, who can contribute with their expertise and experimental validation, the second kind of model checking is more adequate for this project. From the verification and validation of model checking paradigms reviewed, the structure proposed by Webb is more specific about the abstraction and accuracy of the biological properties. Also,

because of the availability of experts in the relevant biological phenomenon studied, Webb's methodology is the most appropriate to assess this work.

3.4. Biological Modelling Methodology

In this work the modelling will be performed using the incremental approach, and it will be assessed by verification and validation using Webb's methodology. However perhaps the most significant element of the methodology of this work is to determine which aspects of biology should be modelled (B. Webb, 2009). This key step usually begins with a study of the biological system under consideration. A literature review is the most straightforward way to learn about the biological system (P. J. Bentley, 2009). However, it is important to keep in mind that everything in biology is in flux: as new information comes into light, old theories have to be reassessed or revised. Sometimes not much work has been done on a given topic because of experimental limitations (or lack of interest) (Cimini & Degrassi, 2005). For this project, both experimental limitations and the controversy surrounding the biological phenomenon make the determination of the key elements especially challenging.

For the construction of this model, it is necessary to isolate the data that are most relevant to the biological process. Then, the quality of the data needs to be assessed. Once this is done, the next step in the construction of the model should be to decide which elements and interactions are modelled (which parameters should be altered) and which become context or environment (fixed parameters or constants that have their values set through reference to literature) and not the focus of the investigation. Finally, an investigation needs to be carried out to determine at what level of abstraction the modelling should be focused upon.

In order to determine which elements should be modelled, and at what level of abstraction the model should take place, we can analyse the availability of data for each level. At the time of writing, new molecular techniques are bringing new data that challenges current biological thinking on the role of chromosome missegregation in cancer. Such biological results come mainly from new methods for probing tumour samples and the silencing of genes in *in vitro* experiments. The problems with these new data however is that they sometimes yield seemingly contradictory conclusions (Sen, 2000). From this we can conclude that, for the model to be flexible enough to be able to account for the disparity of results in the literature, it is necessary to extract from the literature the first principles surrounding the phenomenon at different levels.

Table 3.1- The different kinds of data needed to model at a given level of abstraction, and the availability of such data. With the help and input of cellular biologists at the Baum Lab (UCL LMCB), first principles can be extracted from the literature, assessed and incorporated into a working model.

Biological level	Data Needed	Data Available
Molecular level	Gene mutations (related to aneuploidy)	Available MAD2 Over expression (Sotillo et al., 2010), Mitotic Checkpoint Aberrations (Swanton et al., 2006), Bub1 dysregulation (Ricke et al., 2011), Defective P53 Mechanisms (Thompson & Compton, 2010), Moesin dysregulation (Kunda et al., 2008)
	Aberrations in chromosomes	Partial Chromosomal translocations (Greaves et al., 2003) and (Marcucci et al., 2011), Chromosomal breaks (Crasta et al., 2012), Karyotype analysis (Baker et al., 2009)
	Molecular mechanisms of Chromosome Missegretation	Partial Merotellic defects(Cimini, 2008), excess centrosomes (Ganem et al., 2009)
Cellular level	Genotype/ Phenotype	Available *In vitro* cell line studies (Swanton et al., 2009) and (McClelland et al., 2009), *In vivo* animal studies (B.

		Weaver & Cleveland, 2007) and (Dobles et al., 2000)
	Aneuploidy Levels	Available on cell lines Cell Line Ploidy (Phillips et al., 2001) , (A. J. X. Lee et al., 2011), (Swanton et al., 2006)
	Temporal Evolutionary Changes	Poor Changes in intratumoral Ploidy in gastric cancer (Furuya et al., 2000)
Tissue level	Tissue Structure	Available on animal models, MAD2 induced aneuploidy in lung tumour (Sotillo et al., 2010), epithelial carcinogenesis (Chang et al., 2001),
	Local Aneuploidy	Not available at the time of writing
	Global Aneuploidy	Partial Leukaemia in Twins Studies (Jacobs et al., 2006), Ploidy levels in cell lines (A. J. X. Lee et al., 2011), Ploidy levels in Barrett's oesophagus (C. Maley & Reid, 2005), Histological markers in Breast Cancer Patients (Juul et al., 2010), Somatic rearrangements in Breast Cancer (Stephens et al., 2009)
	Temporal Histological Progression	Not available at the time of writing

The general consensus amongst biologists at the time of writing is that cancer is the disruption of homeostasis in tissues caused by genetic changes. **Homeostasis**, in the context of tissues, is the natural balance between cellular death and proliferation that keeps the number of cells a constant number (Basanta et al., 2008), and helps tissues to remain structurally similar for a prolonged period of time (a liver needs to remain structurally stable for a prolonged period of time to be able to contribute functionally to the organism). While normal cells strive to keep this balance, mutant cancer cells may compromise organ function and integrity. As it was explained in Chapter 2, cancer cells have suffered alterations in their genetic code, amongst which are the activation of **Oncogenes** and the loss of **Tumour-Suppressor Genes** (Hanahan & Weinberg, 2011). Oncogenes, when activated through several mechanisms (most of which are subject of on-going research), drive cell proliferation forward. Cells with mutated oncogenes may ignore the

homeostatic limits and, without taking into account the welfare of the rest of the organism, continue overproliferating. Usually proliferative behaviour is kept in check by tumour-suppressor genes, which impose limits to **overproliferation** by mechanisms that promote apoptosis (cellular death). Key among those are the mechanism of contact inhibition (Zeng & Hong, 2008), where cells enter in mitotic arrest when in contact with other cells, and the regulation of chromosomal integrity by mitotic checkpoints (Pérez de Castro et al., 2007). When the physical structure of the organ is compromised, these cellular mechanisms kick into action by promoting cellular death and maintaining a constant number of cells. Recent work in the literature also offered some insights into the actual mechanisms that give rise to the biological phenomenon of chromosome missegregation: the malfunction of genes that ensure the fidelity during replication (Kunda et al., 2008).

The precise mechanisms of **chromosome missegregation**, and how it affects cancer are still subject of current research. A key part of this on-going research is being carried out at the UCL LMCB by the Baum Lab, integral collaborators in this project. Some of their most recent work suggests that there are "segregation genes" that ensure the proper duplication, division and segregation of genetic material during cellular division (Kunda et al., 2008). When these genes are in turn disregulated, chromosome missegregation, leading to chromosomal aberrations and the development of novel genotypes, may occur. These novel genotypes may have phenotypes that disrupt homeostasis and may have cancer-like behaviour (Dobles et al., 2000). From the Baum Lab's group work, it was suggested that chromosome missegregation, the mechanism that our hypothesis is addressing, is regulated genetically from within the cell. This mechanism could be responsible for the creation of aneuploid cells that contribute to carcinogenesis and help shape its subsequent evolution.

From this analysis, we determine that the key components and mechanisms such as cancer genes, chromosome missegregation and the disruption of homeostasis are distributed across the genetic, cellular and tissue level (Table 3.1). When determining the level at which the system should be modelled, it is clear that if we focused on modelling the genes that regulate chromosome missegregation, we would require data regarding the structure (where they are) and the mechanisms (how they act), which is not readily available. However, if we were to focus on modelling on a tissue level, we would neglect the evolutionary dynamics that happen at a genetic and cellular level. These kinds of dynamics are important because there is data available that could be used to calibrate and validate the model and where new theories can be tested.

From the literature, most of the data available regarding aneuploidy is at the cellular and genetic level. This data however is disjointed; it mostly consists of snapshots in time of different aspects of the biological phenomenon that cannot be used to establish a clear evolutionary pathway (Table 3.1). The most useful impact this model can have is on joining the different observed aspects of aneuploidy in a clear pathway. Consequently the focus of the model should be at the cellular level, because of the data available, with some genetic and tissue level components.

3.5. Selecting the Modelling Paradigm

As we saw in the previous section, the behaviour that the hypothesis addresses is multi-scale: There are genetic, cellular and tissue-level components (Table 3.1). A model that successfully tackles the biological hypothesis must include the disruption of homeostasis, the interplay between genes and how time affects the evolution of the system. After the review of the literature on cancer models, several approaches can be considered.

Differential Equations and Stochastic Models are best used to tackle specific questions of one unknown function given knowledge of the variables and their rate of change (Materi & Wishart, 2007). Differential equations could have been extracted from the relationships between the components to make a systems biology model. The main problem with this approach is that the precise biological relationships of aneuploidy and cancer are still unknown (Cimini, 2008). Because there are few measurements regarding the relationship between the variables (such as the level with which genes affect homeostasis), a differential equation that could accurately represent this behaviour would be very difficult to devise and validate (Wodarz & Komarova, 2005). Stochastic models could offer a more realistic description of the relationships, given that there is more flexibility when defining the mathematical equations. A stochastic model or a Markov chain could have been devised that describes a relationship between genes and cell behaviour, joining another set that relates cell behaviour with homeostasis (Wodarz & Komarova, 2005). These coupled differential equations would soon become too complicated to solve (Gerisch & Chaplain, 2008) and furthermore, they would be very difficult to adapt when an extension of the model was required (Gerisch & Chaplain, 2008).

Table 3.2- The advantages and disadvantages of the different modelling paradigms

Modelling Paradigm	Advantages	Disadvantages
Differential Equations	Rigorous established framework (Macklin & Lowengrub, 2007)	Relies on detailed knowledge of the system (Materi & Wishart, 2007)
Stochastic Models	Flexible established framework (Wette, Katz, & Rodin, 1974a)	Difficult to modify and expand (Gerisch & Chaplain, 2008)
Markov Chain	Helps dealing with random variables (Alarcón & Page, 2006)	Equations become too complicated to solve (Gerisch & Chaplain, 2008)
Boolean Network	Does not require knowledge of interconnectivity (Machado et al., 2009)	Focuses on one level of abstractions and has many possible states (Kampis et al., 2011)
Bayesian Network	Can cope with hierarchical structures (Barabasi, 2005)	Requires knowledge of interconnectivity (Shmulevich et al., 2002)

Cellular Automata	Can generate emergent behaviour (Wolfram, 2002)	Requires the translation of biological behaviour into update rules (Ribba et al., 2004)
Agent-Based Model	Reduces the reliance on flawed or contradictory data (Spencer et al., 2006)	Computationally Intensive (Gerlee & Anderson, 2007)

A focus on genetic or cellular networks modelled with a Boolean or Bayesian Network paradigm was considered. Nodes that accounted for the level of expression of genes or cellular connectivity between different kinds of cells could have been implemented in a computational model (Machado et al., 2009). However, these kinds of models would require many assumptions regarding the interconnectivity of the genes (Shmulevich et al., 2002). Because of lack of experimental data, the gene network could have been set up in many different ways, considering all the possible network motifs and signalling cascades (Kampis et al., 2011). Although an interesting option, it would be difficult to mine the kind of information that we could obtain without a microarray data to corroborate (Ideker & Lauffenburger, 2003).

A high-level modelling paradigm offers the possibility of generalizing behaviour, reducing the reliance on potentially flawed or contradictory data. This kind of modelling, although computationally intensive, can also help in measuring properties in different levels (both above and below the cellular level) and offer insights into the evolution of this kind of phenomena. Using this paradigm, an Artificial Life-style model with the purpose of evolving genes or genetic mechanisms from scratch could have been devised. Cellular automata, for example, could have been employed as an abstraction of cells (Wolfram, 2002). The main problem with this paradigm is that the implementation of update rules would have required the translation of relevant biological questions into rules for artificially evolved automata and vice versa (Basanta & Deutsch, 2008).

Within the high-level paradigm of cancer modelling, an agent-based method that focussed primarily on genes instead of cells was also considered. This gene-centred approach could have resulted in a very detailed account of the interactions between different kinds of genes and, in principle, could have been expandable from the genetic to the cellular and to the tissue level (Abbott et al., 2006). This approach allows for the study of the actual mechanisms that are involved in the biological phenomenon of chromosome missegregation. This bottom-up approach however required more detailed knowledge of the genetic composition of cells than the one available in the literature. The results of such a model would also incorporate a level of detail that surpasses the one required for the hypothesis presented in this thesis. This detail would be associated with a computational cost: simulations would require the use of a computer cluster and would consume a lot of computer time (Schnell, Grima, & Maini, 2007); otherwise, the scale of the simulation would be biologically implausible . A variation of this idea, using the cell as the mid-point between genetic and tissue level is both computationally practical and biologically relevant. Although highly abstract, such a model could allow for the number simulations of a stochastic biological phenomenon under diverse genetic conditions. Because of the abstract quality of the results, key insights might be gained from using the first principles of the mechanism under study to evolve without any other phenomenon. This is an important aspect that *in silico* simulations bring to the biological sciences, and which can yield predictions that can be tested in the wet lab (Wodarz & Komarova, 2005). Also, this approach allows for abstraction and implementation of cellular-centred medical procedures, such as surgery (the removal of cell mass from a tumour) and chemotherapy (the selective killing of rapidly dividing cells), which can be compared to results published in the literature.

3.6. Summary

Although biological modelling is still lacking a rigorous methodology, usually the first step is to learn as much about the system as possible through a literature review. After analysing the current literature and having used feedback from biologists in the field, it was determined that the key things that the model needs to account for are not only at the cellular level, but also the levels above (the tissue level and the disruption of homeostasis) and below (the genetic level and the disruption of the balance between oncogenes and tumour-suppressor genes). For this reason we decided to focus at the cellular level, with genetic structures that will become our independent variables, and tissue level components. It was also decided that incremental modelling and the assessment methodology proposed by Webb would be adequate to implement and evaluate a model of the role of chromosome missegregation in cancer (B. Webb, 2002).

From the cancer literature, it becomes apparent that we do not have enough information to specify every component and interaction by means of differential equations, stochastic models or Markov chains. Although we suspect that many of the interactions are intertwined, the use of Boolean or Bayesian networks is not appropriate because those methods would not allow for the explicit incorporation of the mechanisms of chromosome missegregation, nor would give us a traceable pathway on the effects that chromosome missegregation has on the origins and subsequent evolution of malignant genotypes with enough detail to identify key evolutionary transitions. During the review of the different cancer modelling methods, it became apparent that the more abstract, high-level end of the spectrum is more versatile in these respects. Cellular automata models, however, are still highly constrained to spatial modelling, where the distances between neighbours and the cell-to-cell interactions are more important than the individual genetic details

within the cells themselves. Agent based modelling, although not inherently spatially explicit, is able to give each cell a more versatile identity, which can evolve and be tractable through time.

To tackle the phenomenon of chromosomes missegregation, gain insights into the evolutionary pathways and key transitions, it was necessary to focus on the cellular level, while at the same time being able to interrogate the system at the genetic and population levels. Although computationally intensive, the paradigm that gave us the flexibility and specificity necessary for these constraints (such as internal memory in cells) was agent-based modelling. For this reasons the model has been designed using an agent-based approach, in a bottom-up paradigm.

4. The Model

The creation and development of the model followed a series of steps informed by the structure and availability of data of the biological system being modelled and the relevant questions that needed answering. The creation of the model is an iterative process in which refinements and improvements are made, evaluated and re-assessed. For this purpose, two goals were determined, abstracted and implemented for each iteration. Each iteration of the model was assessed using Webb's methodology, as discussed in the previous chapter. These incremental refinements allow for the streamlining of the model until the version of the model more adequate for our purposes is reached. This final version of the model will be explained in detail at the end of the chapter.

4.1. Introduction

From the analysis of the current cancer modelling methodologies in Chapter 3, it was determined that an agent-based model would be appropriate to study the phenomenon of chromosome missegregation. One of the main advantages offered by agent-based modelling over the other modelling techniques reviewed in the previous chapter is the ability to study explicitly the emergent behaviour that arises from defined interactions between elements of a complex system (Abbott et al., 2006). If the biological concepts are correctly abstracted, an agent-based implementation may help in understanding emergent behaviours and the key transitions that give rise to them (K. Bentley et al., 2008). With this in mind, the process begins by identifying the key biological concepts that will be modelled.

4.2. Abstracting biological concepts

One of the key goals of a cancer model is to make it as biologically accurate as possible with the right abstractions; while at the same time computationally feasible with the correct implementation (Fisher, Henzinger, Mateescu, & Piterman, 2008). From section 3.3, we know that there are key components and behaviours that are known to give rise to cancer, and some mechanisms that are suspected to contribute significantly to this. In order to create an agent-based model of the phenomenon of aneuploidy, an abstraction of reality to incorporate in an algorithm with which we can make predictions needs to be created.

4.2.1. Concepts Modelled

From the literature review in Chapter 2, from wet lab experiments carried out and with the help of molecular biologists from the Baum group at the UCL LMCB (Kunda et al., 2008), (Marinari et al., 2012), it was determined that key amongst these components and behaviours are:

- **Genes:** A gene is a molecular unit of heredity, which codes for key proteins that have a function in the organisms. The activation of genes (when genes express their product proteins) can be part of a signalling pathway (one protein acts as a regulator for another protein), or affect many other key genes simultaneously and cause a specific change in the behaviour of the cell. There is evidence to support the idea that the number of copies of a given gene correlates with the activity of that gene: the more copies of a gene there are, the higher the gene expression and vice versa (Pavelka et al., 2010b).

- **Apoptosis:** Apoptosis is the mechanism of programmed cellular death triggered by biochemical changes. Apoptosis may be induced to prevent excess growth and preserve homeostatic conditions. Conditions such as tissue crowding lead to a corresponding increase in the rate of delamination and cell death within an epithelium to maintain homeostasis. Defects in apoptotic mechanisms have been implicated in carcinogenesis and tumour progression. Tumour suppressor genes that regulate apoptosis, such as the Retinoblastoma protein (pRb) (R. A. Weinberg, 2007) and PTEN (Guo et al., 2008), are usually found mutated, inactivated or missing in cancer cells.

97

- **Contact Inhibition:** When the cell behaviour drives the tissue outside of homeostasis, such as the case of cell overcrowding due to over proliferation, genes that regulate contact inhibition act to promote cellular death (Seluanov et al., 2009).

- **Cellular Division:** Cellular division is the biological process, including cell growth and cell cycle progression, by which a parent cell divides into two daughter cells. During mitosis, a cell enters a series of phases through which one cell creates another copy of itself. In this process, each cell duplicates its entire DNA. The original DNA and its copy are segregated: ideally divided into two sets of diploid genomes and then separated to polar opposites of the dividing cell. The cell then bisects in the middle, physically separating the two genomes. Once this process finishes, two genetically identical "daughter" cells replace the old "parent" cell. Proto-oncogenes such as Ras (M. Wu, Pastor-Pareja, & Xu, 2010), Myc (Soucek et al., 2008) and PI3K (Brunet, 2009) act to promote cell growth and cell cycle progression. In many cancers, these genes are found hyper-activated because of key mutations or due to a higher copy number of these genes in the cancer genome, as discussed in chapter 2.

- **Genome Stability:** Mechanisms that stabilize and repair DNA damage are key in the maintenance of basic cellular functions. When the DNA of cells is compromised, stability genes act to assess the viability of the damaged cell. If the cell has too much damage, such that key cellular functions are severely compromised, genes that safeguard the cellular integrity promote cellular death. Such genes are found throughout the genome, the most studied being T*P53* (Matlashewski et al., 1984).

- **Chromosome Segregation:** During mitosis, the two sets of identical DNA are re-distributed amongst the newly divided cells. This re-distribution of DNA is known as chromosome segregation. Because of mistakes in the process of chromosome segregation, the redistribution may not always be symmetrical: one daughter cell may get more chromosomes than it should, while another receives less. Although work is ongoing to discover the precise mechanisms and the gene defects responsible for this behaviour, genes such as BUB1 (Ricke et al., 2011) and MAD1/MAD2 (Sotillo et al., 2007) are known to regulate the likelihood of chromosome missegregation during cell division. These kinds of genes may reduce fidelity if missing or inactive. Also, there are genes such as actin regulators (ERM proteins) that, by controlling cell shape, may help the cell divide under conditions in which they would normally have problems (Kunda et al., 2008). Both of the above are important because all normal cells have a low level of

defects in DNA (mutation e.g. from UV light) and segregation errors (e.g. from physical events in a tissue, like wounding nearby). These are normally repaired or are identified and used to trigger cell death, e.g. via p53. In addition, there are intrinsic factors that can lead to DNA damage and to chromosome segregation errors (e.g. genes that make the process one of high fidelity).This work explores the idea that that the missegregation of chromosomes because of disregulation in these kind of genes is a significant mechanism behind the generation of aneuploid genomes and may have a key role in cancer initiation and evolution.

- **Homeostasis:** Homeostasis is the property of biological systems to regulate their constituents to maintain stable, constant conditions such as tissue structure and cell number (Basanta et al., 2008). This process includes an ability to restore an equilibrium following perturbation, i.e. it is a state of dynamic equilibrium. Homeostasis restores equilibrium from overproliferation or from a cell number below the equilibrium. Loss of the ability to maintain cell number is associated with ageing and cancer (Meza et al., 2008). Complex signalling between genes in response to environmental cues during development give rise to the determining of this equilibrium and, according to the specific function of individual tissues, determine cell differentiation and their place within the tissue´s architecture. This complex phenomenon is an emergent property of an organized cellular system. By evolving a balance between apoptosis and cell division, it becomes possible to maintain a stable, constant population of cells and thus give rise to complex organisms. When homeostasis is no longer conserved, the cellular system becomes unbalanced and may veer towards two extremes: death of the system or cellular over proliferation – one of the key hallmarks of oncogenesis.

- **Euploidy:** Ploidy is the number of sets of chromosomes in the nucleus of cells. A diploid genome is a genome composed of two sets of identical chromosomes; the structures that contain the genes that regulate cell behaviour (Furuya et al., 2000). This is a result of sexual reproduction: normally sex cells (such as sperms or eggs) carry a full set of chromosomes, including a single copy of each gene. The haploid number (n) of chromosomes is the number of chromosomes in a sex cell. Two sex cells form a diploid zygote with twice this number (2n) and two copies of each gene. Cells can be described according to the number of sets present: haploid (n), diploid (2n, the normal state of cells in our bodies), triploid (3n), tetraploid (4n, the state of normal cells before undergoing cell division), etc.

- **Aneuploidy:** Aneuploidy is the state of not having euploidy. This abnormal chromosome number can be due to gains or losses of entire chromosomes during cell division (Rajagopalan & Lengauer, 2004). In humans, viable stable aneuploid genotypes include having a single extra chromosome (such as Down syndrome), or missing a chromosome (such as Turner syndrome). Most cancers are aneuploid (Stratton et al., 2009). Chromosome structural changes, such as translocations in chromosome regions, will not be discussed here.

4.2.2. Concepts not Modelled

The abstractions discussed, implemented in an agent-based model, could allow us to test the role of chromosome missegregation in cancer evolution. However it was decided that certain aspects of cancer would not be modelled. Amongst these, point mutations, which are considered an important feature of cancer systems, were not explicitly modelled. To isolate the effects of chromosome missegregation in cancer progression, it was important that the effects of mutations were not convoluted with those of chromosome missegregation. Furthermore, the rate of point mutation compared to that of chromosome missegregation has been determined to be orders of magnitude lower than that of chromosome missegregation (Wodarz & Komarova, 2005). A special mutation that allows cells to reproduce indefinitely by means of telomerase re-expression (bypassing the Hayflick limit) (Hanahan & Weinberg, 2011), was also considered to be outside of the scope of the project because of the limited timescale that is being considered.

Reproductive penalties due to a high number of chromosomes were initially considered. On-going research suggests that a larger number of chromosomes may be associated with reproductive penalties such as a slower rate of division (Torres, Williams, & Amon, 2008). It is also currently thought that losing key checkpoints, a possible result of losing chromosomes, may result in a rapid progression throughout the cell cycle (M. A. Nowak et al., 2002). If modelled,

these considerations could affect the rate at which the genetic subpopulations in the system evolve. If modelled, these concepts, although interesting, would synergise with the missegregation of chromosomes, resulting in rich behaviour that would be difficult to analyse; thus making the investigation of this complex phenomenon a more difficult task. However, as more research is carried out on these behaviours, and the behaviour obtained from chromosome missegregation alone is understood, these concepts not modelled could be incorporated in a future version of the model as a feature, or even as a variable for research.

4.3. Model Version 1: Homeostasis and Aneuploidy

Using the research methodology previously discussed, the modelling process is broken down into incremental goals. Goals are then implemented and assessed through Webb's methodology (B. Webb, 2009). After each assessment, a new iteration of the model is created. As previously discussed, the first step in building the model is to establish the framework in which the rest of the abstractions will be laid down on. Beginning with an agent-based model framework in C++, it is key to establish the agents and their ideal behaviour: homeostasis. Once this is accomplished, the basic mechanism of change, aneuploidy, should be modelled. This will function as a foundation for the rest of the abstractions.

4.3.1. Goal 1.1: Model for Homeostasis

To address whether chromosome missegregation plays a causal role in the course of a cancer, a model of tissue homeostasis in which to study cancer evolution needs to be developed. From the research carried out, and as explained in the previous chapter, it was decided to model individual cells and their interactions as agents in a computational model. Each cell is to be an agent equipped with a genome that regulates cell behaviour, while the interaction with other cells

should give rise to a homeostatic behaviour; the result of balanced rates of cell proliferation and cell death. The homeostatic constraints in the model are abstracted from real biological systems, where the overall goal of homeostasis is to maintain the tissue's relative constant size and shape. The homeostatic size of the tissue is established for each experiment through an allocated space parameter. Although measurements for homeostasis are based on global cell counts, this model is not spatially explicit. The implementation of the agent-based model considers that the rate of death and the rate of division are the same for every cell.

The components are cells modelled as agents, following for each time step the algorithm:

1. An initial population of 100 cells is created. The normal carrying capacity of the tissue is fixed at 200 cells.
2. For each time step, the total number of cells is measured and is not updated until the next time step. Each cell is examined every time step (like a cellular automaton model).
3. If a measurement of the total number of cells is greater than the tissue's carrying capacity, then the cell dies with probability of apoptosis, p_{ap}; a pre-determined parameter.
4. If the cell has not died, it has a chance to divide. The probability of division, N_{div}, is a pre-determined parameter of the rate of division. This process is independent of whether the total number of cells exceeds the carrying capacity.

So, in the first abstraction of homeostasis, a tissue composed of hundreds of cells is simulated with the following initial parameters:

For biological behaviour:

- Intrinsic rate of apoptosis: N_{ap} =0.5
- Intrinsic rate of division: N_{div} =0.5

For computational feasibility:

- Initial population: 100 cells
- Homeostatic size of the tissue: 200 cells
- Simulation end time: when reaching 100 time steps

A program was devised to capture key information of the evolution of the model. In each time step, the algorithm stores the complete simulation: each cell, its age and its genetic contents. This information is the queried through an algorithm written in the program Mathematica to extract useful information such as the total cell number (as seen in Figure 4.1 and Figure 4.5) and the number of cells that share a genotype (as seen in Figure 4.9). This information can then be reformatted and plotted in any graphics software such as Excel.
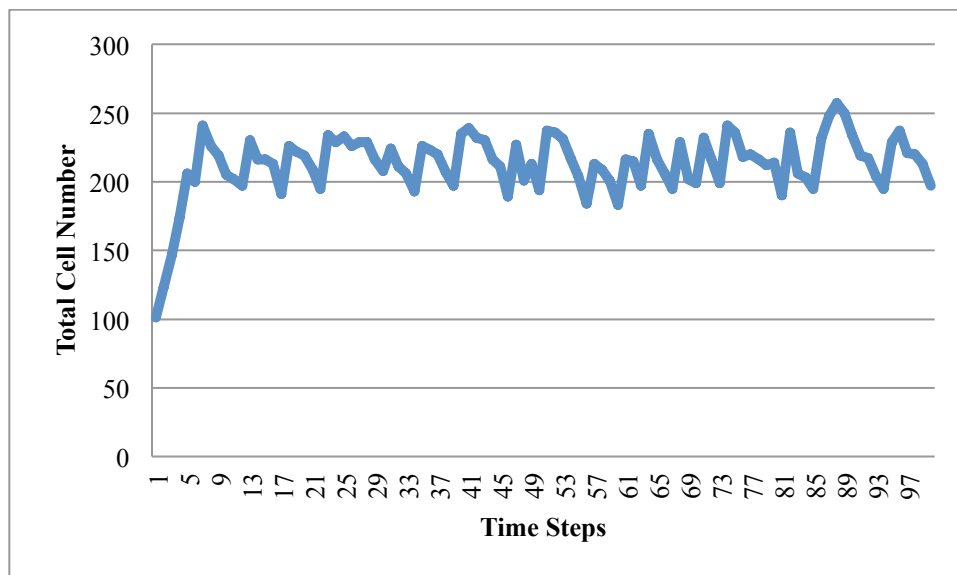
Figure 4.1 Simulation of homeostasis in Model Version 1. Starting at 100 cells, the population grows until the homeostatic 200 cells. The populations then oscillates around this homeostatic state.

In this first implementation, the rate of cellular death is equal to the rate of cellular division. This results in approximately half of the cells dying, but being replaced by the other dividing half such that a homeostatic balance is achieved, as seen in Figure 4.1. The next step in establishing the framework is defining the concept of aneuploidy in terms of an agent-based model.

4.3.2. Goal 1.2: Model of Aneuploidy

Through this goal, we want to extend the previous model, by incorporating a model of aneuploidy. The first step in modelling aneuploidy is to model the basic units of the aneuploid mechanisms: chromosomes. For this, it was decided that each cell of the initial population should have two sets of identical chromosomes. Once this structure was determined, 2 copies of each gene (a diploid genome) were distributed amongst the two sets of chromosomes. Since genes can arranged in different configurations (one example seen in Figure 4.2), the distribution may have important repercussions through the evolution of the system. The key mechanism that is incorporated occurs during cell division. As in real biological systems, when dividing, the genome of each cell needs to be duplicated and the two sets of chromosomes then segregated into two daughter cells.

This could have been modelled as one variable per chromosome, such as initially having chromosome1 = 2 and chromosome2 = 2; and in the case of an asymmetric cell division, the result could have been stored as, for example, chromosome1 = 3 (three copies) and chromosome2 = 1 (one copy). However, it was decided that the best way to create a more general framework that would allow in the future for expansion was to have two variables for each chromosome type in a cell. Each cell is then equipped with variables that store individual information for each chromosome: Chromosome1a, Chromosome1b, Chromosome2a and Chromosome2b. This provides individuality for chromosomes and could, in a future version of the model, be extended to incorporate point mutations and other individual aberrations within the chromosome.
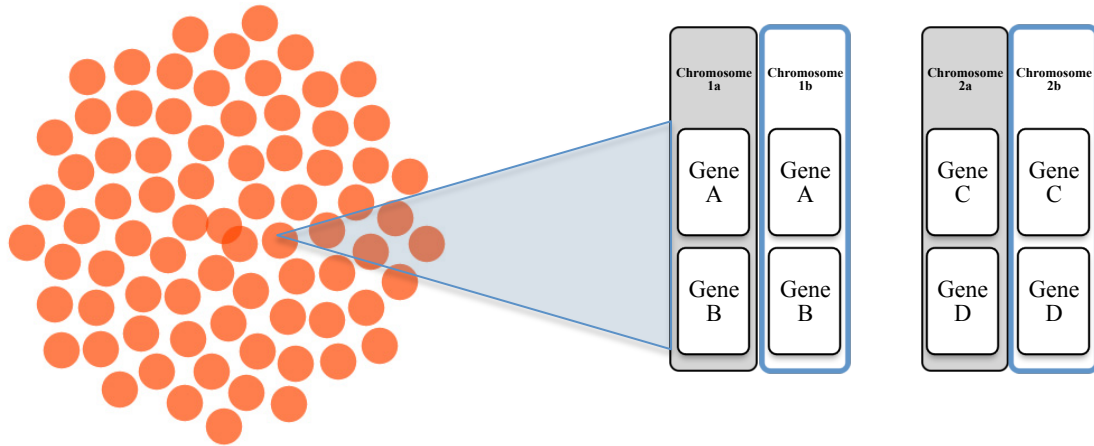
Figure 4.2 The core of the model is the abstraction of individual cells and their genomes. Each simulated genome is composed of genes in diploid chromosomes (pairs of chromosomes, the chromosomes of each pair having identical genes) as the normal state within cells. The collection of individual cells comprises a simulated tissue, whose population size is determined for each experiment through an allocated space parameter. The dynamics are determined by the gene expression of the individual cells across time.

In the previous iteration of the model, each cell had the chance to be deleted, modified or have its contents copied into a new cell every time step. Biologically, it is during the stage of cell division that chromosome missegregation events can occur, resulting in asymmetric cell division: one daughter cell with an extra chromosome, and one lacking the same chromosome. For this, the model needs to contain a finite rate of chromosome missegregation during cell division, which would generate variation amongst the cell population. The algorithm can be modified, as shown in bold, as follows:

1. An initial population of 100 cells is created. The normal carrying capacity of the tissue is fixed at 200 cells.
2. For each time step, the total number of cells is measured and is not updated until the next time step. Each cell is examined every time step.
3. If a measurement of the total number of cells is greater than the tissue's carrying capacity, then the cell dies with probability of apoptosis, p_{ap}; a pre-determined parameter.
4. If the cell has not died, it has a chance to divide. The probability of division, N_{div}, is a pre-determined parameter of the rate of division. This process is independent of whether the total number of cells exceeds the carrying capacity.

5. **If dividing, the probability of chromosome missegregation is calculated. If there is a chromosome missegregation event, one of the 4 types of chromosomes is chosen at random. If the cell still has copies of this chromosome, it is asymmetrically distributed during cell division leading to the creation of two aneuploid cells. Otherwise, the genome is duplicated and copied with fidelity, thus generating two identical daughter cells. The probability of chromosome missegregation, p_{msg}, is a fixed parameter for the rate of chromosome missegregation.**

Cells within the body have different rates of division, differentiation and death. Cells that are believed to be at risk of cancer are those that are being continuously renewed, such as the lining of the gut, colon and the prostrate (R. A. Weinberg, 2007). However, to create a model that takes all of these into account, the parameter calibration is based on the average growth rates of established cell lines, which normally proliferate with a rate of less than 10% daily (A. J. X. Lee et al., 2011). Highly malignant cancer cells such as HeLa (Hutchings & Sato, 1978) and RPE1 (Uetake & Sluder, 2004), however, are able to double their population within 2 days (up to 50% daily). For this project a rate of 9% of cell division rate per time step was estimated from experimental lab work with RPE1 cells. In the model, time is not explicitly modelled, but following these rates obtained from *in vitro* experiments, one time step is considered to be one day. This is scalable in the same way that *in vitro* rates can be scaled to *in vivo* rates by considering other key factors (outside of the scope of this project). If homeostasis is to be preserved, the rate of apoptosis needs to be equal to that of division, as observed biologically in epithelial tissues (L. Zhang et al., 2010). The rate of chromosome missegregation however has not been accurately measured as of the time of writing. In an initial attempt to capture qualitatively such a gene-regulated phenomenon (Kunda et al., 2008), it was decided that the process should be high, so that we can speed up the process. For this purpose, it will be considered that the rate of chromosome missegregation of the same as other gene-regulated

mechanisms. The philosophy behind this modelling decision, which will be carried out throughout the rest of this work, is that we are interested in the study of relative times between control and experimental populations.

As initial temporal parameterization assumes that the rate of chromosomal instability is high (for instance, the simulation could be taking place late in an organism's lifespan, or in a dish of cultured pre-cancer cells), time becomes a scalable quantity that could represent days (in the case of cultured cells), or months (in the case of organisms). The computational resources available, however, restrict the cell numbers that can be comfortably modelled. Despite of this, the development of the model allows for scalability: cell numbers can be considered an accurate model of an *in vitro* experiment (in the days time scale), while the genotypic population could in principle be extrapolated to assess an *in vitro* experiment (in the month time scale). As biological techniques allow for the further proving of this mechanism, and more computational power becomes available, a more realistic parameterization of this phenomenon can be implemented. Following this reasoning, in the first abstraction of homeostasis, a tissue composed of hundreds of cells is simulated with the following initial parameters:

For biological behaviour:

- Intrinsic rate of division: $N_{div} = 0.09$
- Intrinsic rate of apoptosis: $N_{ap} = 0.09$
- Intrinsic rate of missegregation: $N_{div} = 0.09$

For computational feasibility:

- Initial population: 100 cells
- Homeostatic size of the tissue: 200 cells
- Simulation end time: when reaching 100 time steps

```
[Session started at 2009-09-17 15:59:53 +0100.]

(Head) XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXA (Tail)

102 total cells

(Head)
XXXXXAAXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXA (Tail)
generation0

26 new cells

0 dead cells

128 total cells

(Head)
XXXXXXAAAXXXXXXXAAXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXA (Tail)
generation1

23 new cells

0 dead cells

151 total cells

(Head)
XXXXXXXXXAAAXXXXXXXXXXXAAXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXAAXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXA (Tail)
generation2

29 new cells

0 dead cells

180 total cells

(Head)
XXXXXXXXXXAAAAXXXXXXXXXXAAXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXAAXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXAAXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXA (Tail)
generation3

36 new cells

0 dead cells

216 total cells
```

Figure 4.3 A character-based output of a simulation with the Model Version 1. Homeostatic behaviour, the conservation of relative number of cells through time was modelled, and then the mechanism of chromosome missegregation was implemented. Through chromosome missegregation, aneuploid cells (displayed by the console as "A") can be distinguished from diploid cells (displayed as "X").

A simple character-based output to the console was sufficient at this stage to record the internal state of the cells. For this, a print of either X, A, O or @ is obtained, depending on the internal state of each cell's Chromosomes 1a and 1b as follows:

If the cell

- Contains both copies of Chromosome1a and Chromosome1b, display X.
- Contains no copies of Chromosome1a or Chromosome1b display O.
- Contains one copy of Chromosome1a or Chromosome1b display A.

- Contains more than 2 copies of either Chromosome1a or Chromosome 1b display @.

A character-based output of a tissue containing 100 cells through 100 time steps with a 50-50% chance of dividing/dying is shown in Figure 4.3. When including the mechanism of chromosome missegregation, aneuploid cells (Displayed as "A" or "@") are obtained.

The algorithm devised serves as a general model for homeostasis. The model incorporates chromosome missegregation as the mechanism that leads to aneuploidy. However, since aneuploid cells are regulated though external parameters, which are the same for diploid cells, their behaviour is the same: homeostatic in the vicinity of the carrying capacity of 200 cells, with cells dying and being replaced at each time step on similar numbers.

4.3.3. Model Version 1 Evaluation

In order to provide evidence to support the hypothesis, it is necessary to gauge the effectiveness of our modelling methodology. This work uses a seven-dimension analysis, as proposed by Webb (Webb, 2002), to specify the position that our methodology takes and justification for that position. It is important to remember that, although Webb's framework provides guidelines for the construction of good models, there is currently no single agreed methodology to follow (Byrne, Alarcon, Owen, Webb, & Maini, 2006). Webb's methodology, however, presents a list of statements that determines what a good model is. This is helpful in guiding the model throughout the process of construction, assessment and refinement. We assessed our progress through Webb's seven dimensions as:

- **Relevance**: The model is an abstraction of a key biological process, homeostasis, which is found disturbed in cancer as seen in Figure 4.1. This is a solid framework, but lacks relevance as a cancer model. It cannot be used to examine complex interactions as is because no cancer-like features can be observed.

- **Level**: The abstractions in the model are made at the cellular level and chromosome level, generating behaviour at the tissue level, as seen in Figure 4.2. Genetic components, a key part of this investigation, should be modelled in the next version. However, this iteration is the first step towards exploring hierarchical events.

- **Generality**: This is a very general model for homeostasis, simulating only the key features of balanced death and division, as set in Goal 1.1. This can be adapted to any system that deals with homeostasis. It is not, however, a general model of aneuploidy, because although the mechanism has been implemented in Goal 1.2, its biological role needs to be defined in the next version.

- **Abstraction**: Although this is a direct abstraction of a biological phenomenon, the regulation is artificial, as established in Goal 1.1. Only homeostatic behaviour can be obtained, as seen in Figure 4.1.

- **Accuracy**: After consulting biological experts and the literature, it is deemed an adequate but incomplete implementation of homeostasis. True homeostasis is regulated genetically (Nelson & Bissell, 2006). This is something that needs to be addressed in the next iteration of the Model. Also, although the rates of division, apoptosis and chromosome missegregation are biologically inspired, the abstract nature of the model requires a re-scaling of these rates for computational feasibility. These rates should be re-evaluated in the next iteration.

- **Match**: The behaviour obtained in simulations matches the biological phenomenon of homeostasis (Nelson & Bissell, 2006), as seen in Figure 4.1. However, the behaviour of aneuploidy as implemented on Goal 1.2 does not match that of aneuploidy in cancer.

- **Medium**: The model was implemented as an agent-based model on C++. The limitations of this medium are dictated by the hardware employed for simulations such as memory and processing speed issues that limited the number of cells that could be simulated. Our simulations were run on a 2.4 GHz Intel Core 2 Duo processor with 4 Gb of memory. With this hardware, the simulations took less than a second to be computed.

Table 4.1 Assessment of the Model Version 1 with respect to Webb's seven dimensions. The metric ranges from 0 to 5 stars, where more stars reflect an improvement over the model. 5 stars is the target goal for this model.

Dimension	Relevance	Level	Generality	Abstraction	Accuracy	Match	Medium
Assessment	**	*	***	**	**	**	***

4.4. Model Version 2: Regulation of Behaviour Through Genes

The creation of a new version of the model is motivated by the deficiencies found in the previous evaluation. In biological systems, many features of cell biology depend on the organized action of several genes. Some genes function as regulators for other genes and, through complex protein interactions and signalling, cells are able to respond to external and internal stimuli. Genetic regulation, and how the number of copies of genes affects it, is currently a subject of study. For the purpose of our model of aneuploidy, it was determined to abstract a collection of similarly acting genes as a single, representative gene regulating a specific cellular behaviour. Another key abstraction was also determined to be that the regulation of biological behaviour by each gene could be proportional to the number of copies of a given gene found in the genome of each cell. This kind of regulation has been suggested by recent studies on the effects of differences in gene number on gene expression in biological systems (Huettel, Kreil, Matzke, & Matzke, 2008), correlating an elevated copy number with higher gene expression and a decrease in copy number with lower gene expression. The genes that should be abstracted are oncogenes and tumour suppressor genes that regulate homeostasis (such as apoptosis, contact inhibition and cell division regulatory genes), and those that affect variation (genetic stability and chromosome segregation genes). As more genetic research is carried out, the identity of the genes and their actual locations across the genome could be implemented in a future version of the model, and may result in testable predictions.

4.4.1. Goal 2.1: Apoptosis and Cell Division Regulatory Genes

A homeostatic framework for an agent based model of homeostasis and aneuploidy was established in section 4.3. Each cell is equipped with chromosomes that may contain genes and could be missegregated during cell division. With this in mind, each cell in the system can be equipped with a genome composed of abstracted genes within their chromosomes. The first kind of genes that need to be modelled, are those responsible for maintaining a balance between death and division. Modelling genes that regulate cell death, **Apoptosis Regulatory Genes** are an abstraction of tumour suppressor genes such as pRb (Amato, Lentini, Schillaci, Iovino, & Di Leonardo, 2009) that regulate the probability of cell death. To balance cellular death, **cell Division Regulatory Genes** provide an abstraction of proto-oncogenes such as Ras (M. Wu et al., 2010), Myc (Soucek et al., 2008) and p110 PI3K (Brunet, 2009) and act to promote cell growth and cell cycle progression.

A balance between cellular death and division is the key to maintaining homeostasis. For this, it was decided to model the dynamical balance based on the mechanisms of contact inhibition. In real tissues, due to contact inhibition, cells enter in mitotic arrest when in contact with other cells, thus maintaining structural integrity of the organism (Zeng & Hong, 2008). Genes that regulate tissue overcrowding, or **Contact Inhibition Genes**, model the mechanism of contact inhibition; effectively limiting cell over proliferation in crowded tissues (Zeng & Hong, 2008). If the number of cells exceeds the homeostatic limit, contact inhibition genes inhibit proliferation and raise the probability of cell death, thus maintaining a constant population of cells close to the homeostatic capacity of the simulated tissue. This also enables us to model the fact that tissue overcrowding leads to a corresponding increase in the rate of delamination and cell death within an epithelium to maintain homeostasis (Marinari et al., 2012).

With changes in bold, the algorithm is now:

1. An initial population of 100 cells is created, each with diploid chromosomes. **Each initial genome was equipped with 2 copies of each type of gene, grouped into chromosomes.** The normal carrying capacity of the tissue is fixed at 200 cells.

2. For each time step, the total number of cells is measured and is not updated until the next time step. Each cell is examined every time step.

3. **If a measurement of the total number of cells is greater than the tissue's carrying capacity, then the probability of cell death through contact inhibition (or crowding induced delamination), p_{ci}, is calculated as.**

$$p_{ci} = r_{ci} N_{ci}$$

   where r_{ci} is a parameter for the sensitivity of contact inhibition and N_{ci} are the number of available copies of the contact inhibition regulatory genes within the cell´s genome. The cell is then killed with a probability of p_{ci}.

4. **If the cell has not died because of contact inhibition, the probability on natural apoptosis is calculated. This probability of death depends on the number of available copies of the apoptosis regulatory genes, N_{ap}, within each cell's genome. The probability of apoptosis, p_{ap}, is determined by:**

$$p_{ap} = r_{ap} N_{ap}$$

   where r_{ap} is a parameter for the rate of apoptosis. The cell is then killed with probability p_{ap}.

5. **If the cell has not died, it has a chance to divide. The probability of division depends on the number of available copies of the division regulatory genes, N_{div}, and a parameter that determines the rate of division, r_{div}. The probability that a cell divides, p_{div}, is:**

$$p_{div} = r_{div} N_{div}$$

6. If dividing, the probability of chromosome missegregation is calculated. If there is a chromosome missegregation event, one of the 4 types of chromosomes is chosen at random. If the cell still has copies of this chromosome, it is asymmetrically distributed during cell division leading to the creation of two aneuploid cells. Otherwise, the genome is duplicated and copied with fidelity, thus generating two identical daughter cells. The probability of chromosome missegregation, p_{msg}, is a fixed parameter for the rate of chromosome missegregation.

113

4.4.2. Goal 2.2: Chromosome Segregation Regulatory Genes

Having established this homeostatic model system, it is necessary to introduce a fourth gene abstraction that regulates the rate of chromosomes segregation. This **Chromosome Segregation Regulatory Gene** models genes that control the fidelity of cell division such as BUB1 (Ricke et al., 2011) and MAD1/MAD2 (Sotillo et al., 2007). The expression of this gene proportionally reduces the likelihood of chromosome missegregation at cell division. The parameterization of this gene is considered of the same order as that of the previous genes modelled, although experimental rates are still under investigation. In the future, when these rates have been experimentally established, they can be re adjusted in a more realistic version of the model. The decision taken in section 4.3.2 regarding modelling individual chromosomes affects in a small way the probability of missegregation when dealing with a small number of chromosomes compared to using one variable per chromosome, so that, if there is one kind of chromosome missing, the probability of missegregation proportionally smaller. However, there is evidence that shows that cells with higher number of chromosomes are more unstable and tend to missegregate more frequently than those with a lower number of chromosomes (Thompson & Compton, 2008). This shall be taken into account when analysing the results of the final version of the model.

Also, in real biological systems, cells with less than the diploid set of chromosomes may find it difficult to carry out essential life processes. To model this fact, an abstraction of **Genome Stability Genes** can represent the fact that missing entire chromosomes may translate in missing crucial machinery and thus a higher likelihood of death such as p53 (Thompson & Compton, 2010). Since the simulation is composed of cells with only four chromosomes, it was decided that cells with less than 2 chromosomes (an evolution from the initial four chromosome diploid state through chromosome missegregation) would be considered unfit for survival and removed

from the simulation with a probability proportional to the number of genome stability genes. A schematic of a representative gene distribution can be seen in Figure 4.4.
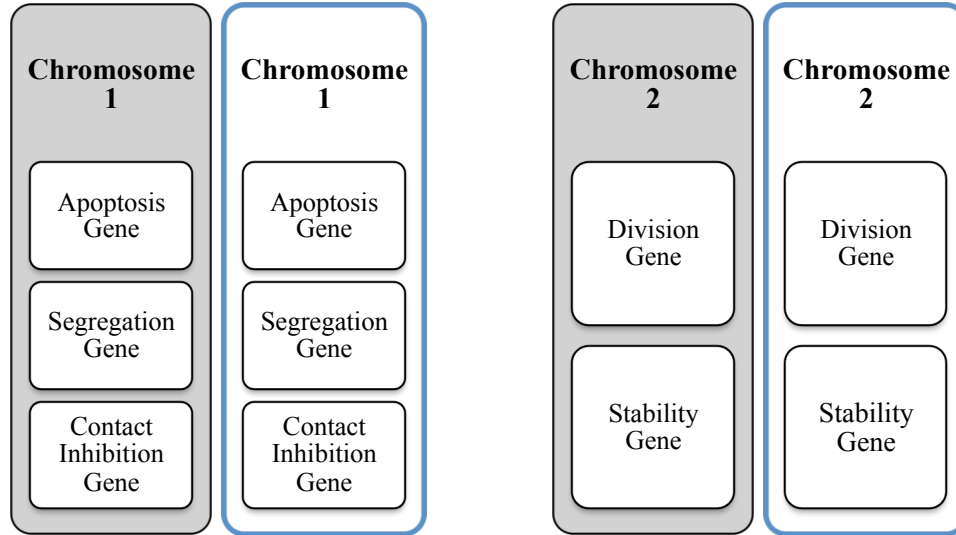
Figure 4.4- Genetic Arrangements. Abstracted genes were placed inside the chromosomes of simulated diploid cells, as in this representative configuration.

The algorithm is then modified with the statements in bold:

1. An initial population of 100 cells is created, each with diploid chromosomes. Each initial genome was equipped with 2 copies of each type of gene, grouped into chromosomes. The normal carrying capacity of the tissue is fixed at 200 cells.

2. For each time step, the total number of cells is measured and is not updated until the next time step.

3. **If a cell has 1 or less chromosomes in its genome, the cell's genetic integrity mechanisms has a probability of death through unviable genome instability, p_{gi}, is calculated as.**

$$P_{gi} = r_{gi}\, N_{gi}$$

where r_{gi} is a parameter for the sensitivity of contact inhibition and N_{gi} are the number of available copies of the genome stability genes within the cell´s genome

4. If a measurement of the total number of cells is greater than the tissue's carrying capacity, then the probability of cell death through contact inhibition (or crowding induced delamination), p_{ci}, is calculated as.

$$p_{ci} = r_{ci} N_{ci}$$

where r_{ci} is a parameter for the sensitivity of contact inhibition and N_{ci} are the number of available copies of the contact inhibition genes within the cell´s genome. The cell is then killed with a probability of p_{ci}.

5. If the cell has not died because of contact inhibition, the probability on natural apoptosis is calculated. This probability of death depends on the number of available copies of the apoptosis regulatory genes, N_{ap}, within each cell's genome. The probability of apoptosis, p_{ap}, is determined by:

$$p_{ap} = r_{ap} N_{ap}$$

where r_{ap} is a parameter for the rate of apoptosis. The cell is then killed with a probability of p_{ap}.

6. If the cell has not died, it has a chance to divide. The probability of division depends on the number of available copies of the division regulatory genes, N_{div}, and a parameter that determines the rate of division, r_{div}. The probability that a cell divides, p_{div}, is:

$$p_{div} = r_{div} N_{div}$$

7. **If dividing, the probability of chromosome missegregation is calculated. If there is a chromosome missegregation event, one of the 4 types of chromosomes is chosen at random. If the cell still has copies of this chromosome, it is asymmetrically distributed during cell division leading to the creation of two aneuploid cells. Otherwise, the genome is duplicated and copied with fidelity, thus generating two identical daughter cells. The probability of chromosome missegregation, p_{msg}, in the model is:**

$$p_{msg} = r_{msg} (4 - N_{mseg})$$

The parameters were recalibrated in order to obtain the same behaviour when incorporating a diploid set of genes as the regulators of the behaviour.

For biological plausibility, as described in section 4.3.2, are:

- Intrinsic rate of apoptosis (gene adjusted): $r_{ap} = 0.045$
- Intrinsic rate of division (gene adjusted): $r_{div} = 0.045$

- Intrinsic sensitivity to contact inhibition (gene adjusted): $r_{ci} = 0.045$

- Intrinsic action of genome stability mechanisms (gene adjusted): $r_{gs} = 0.045$

- Intrinsic rate of chromosome missegregation (gene adjusted): $r_{msg} = 0.045$

For computational feasibility:

- Initial population: 100 cells

- Homeostatic size of the tissue: 200 cells

- Simulation end time: when reaching 100 time steps

This implementation of these biological features into genes, emergent cancer-like behaviour can be obtained under certain circumstances, as seen in Figure 4.5. Because the simulations were not stopped until 100 time steps had been simulated, computational problems arose in this stage because of memory issues. When cancer-like behaviour was obtained, simulations would give rise to an exponential growth of cells. In some simulations, this rate became too high for the computer to finish the simulation on a practical amount of time. Some simulations took days to complete, while others took seconds. Some simulations became so intensive that they were aborted because of hardware failure. For computational feasibility, a cell limit needs to be set in the next iteration of the model.
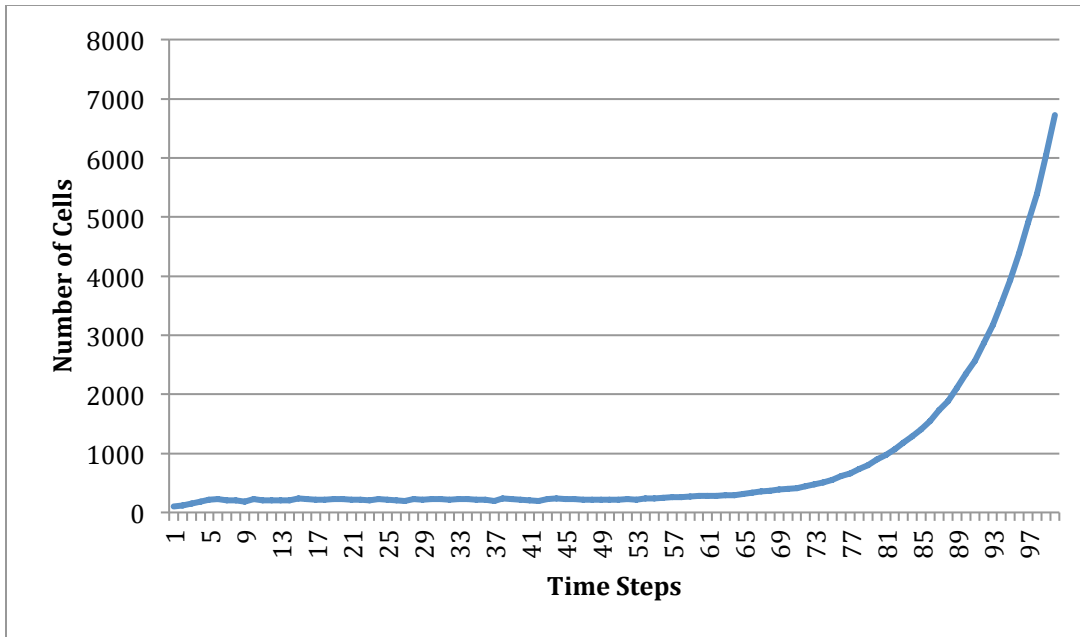
Figure 4.5- A plot of the total number of cells for each time step in the simulation. Cells maintain an homeostatic behaviour for almost half of the simulation. Then uncontrolled growth occurs. This overproliferative behaviour arises from the genetic arrangement and the phenomenon of aneuploidy.

From this version of the model simulations can be analysed in more detail, and information for the evolution of the genotype population can be extracted. Each simulation was saved as a file containing the information of number of chromosomes and age for each cell, for each time step. From this information, an analysis of ploidy, through the Mathematica extraction algorithm described in section 4.3.1, suggests that aneuploid cells are responsible for the emergent cancer-like behaviour, as seen in Figure 4.6. When plotting the average number of a given chromosome across a simulation, it is revealed that cells are consistently losing a certain chromosome: the one containing the regulator of cell death. Cells also consistently gain copies of the chromosome containing genes that promote proliferation, as seen in Figure 4.8. The effects of chromosome missegregation unbalance the rate between cell death and cell birth. While some death mechanisms still remain active, massive death can still be overrun by even greater cell birth as seen in Figure 4.7

For visualization, new tools have to be developed. A simple query for the different number of

possible genotypes, however, leads quickly to a high number of possibilities as seen in Figure 4.9.

This analysis suggests that the model needs to be simulated for longer for a more detailed analysis

of the evolutionary pathways that the genotypic populations could take.

```
540 total cells

(Head)
XA@@@@X@@@@@@@@@@@@AAAAX@@@XXXXXXXX@@@@XX@@X@@@@@@@@@@@@AAXX@@@@@@@@@@@@@XX@@@@X@@@@@@@@@@X@@@@@@@@@@@@@@@AA@@XX@XX@@@X
X@@@@@@@XX@A@AAAAAAAAAAAAAXAAA@@@@@@@@A@XXX@@@@@@XX@@@@@AAXAA@@@AXXXXXXXXXXX@@@@@@@@@XXXXXXXXXX@@@@@@@@@@@@@@@AAAXX@@@OA
@A@XX@@@X@@@@AAA@A@@@@@@@@@@@@XX@@@@@@@@@@@@@@@@@@AAAA@X@@@XXXXAAAAAAA0@@@@@@AA@@@@@AAA@@@@@XX@@@@@X@@@@A@@@@@A@@AAAAAA
AAAAAAAAAXAAAA@AAAAA@AAAAAAAA@@XXAAA@@@@@@@@@@@@@A@@@@@@AAAAAXAAAAAA@@@@@AAXAAAAA@@@@@A@AAAA@@@@@@@@@@@@@@@@@@@@@@@@@@
@@@@@X@@@@@X@@@@@AAAX@@@@XX@AAAAAAAAAAAAAAAAXXX@@AAAAAXXAAAAAAA@@@@@A@X@@@@@XX@@@X@@@@@@AAXAAAAAXXXXXXXXXXXXXXAAXXAAA
AA (Tail)

 generation: 68

111 new cells

64 dead cells

587 total cells

(Head)
A@@@@@@X@@@@@@@@@@@@@@AAAX@@@XXXAAXXX@@@XXX@@@X@@@@@@@@@@@@AAA@@@@@@@@@@@@@XXX@@@X@@@@@@@X@@@@@X@@@@@@@@AA@@X@X@@@XXX@@
@@@@@@@XXXX@A@@AAAAAAAAAAAAAAXAA@@@@@@AA@@XXX@@@X@@XX@@@@@AAAXA@@AXXXXXXXXXXX@@@X@A@@@@XXXX@@@@@@@@@@@@@@@@@AAAA@@@@OA
@@A@XXX@@@@XX@@@@AAXAA@@@@@@@@@@@@X@X@@@@@@@@@@@@@@X@@@@@@@@AAAAA@X@@@XXXXAAAAAA0@@@@@@AA@@@@@@AA@@@@@@@@@@@A@X@@@@@AA@
@X@@@@AAAAAAAAAAAAAAAAA@AAAAA@AAAAAAAAAAAAA@@@XXXAA@@@X@@@@@@@@@@@@@A@@@@@AAAAAAAAAXAAAAAA@@@@@@@@@XAXAAA@@@@@@AAAAA@
@@@@@@@@@@A@@@@@@@@@@@@@@@@@X@@@A@@@@X@@@@@AAAAX@@@@X@AAAAAAAAAAAAAAAAXXXX@@@AAAXXAAAAAAA@@@@@@A@X@@@@@XX@@@@X@
@@@@@@@@AXXAAAAAAXXXXXXXXXXXXXXXXXXXAAAAAAA (Tail)

 generation: 69

118 new cells

80 dead cells

625 total cells

(Head)
@@@@@@@X@@@@@@@@@@@@@@@AAAX@@@XXAAXXX@@@XX@@@@X@@@X@@@@@@@@@@@@AAAAA@@@@@@@@@@@@@X@XX@@@X@@@@@@@@@X@@@@@@@@@@@@@@@@@@AA@@@
@@@@@XXX@@@@@@@@@@@@@XXXA@@@AAAAAAAAAAA@@@@@@@AA@@XXX@@X@@@XX@@@@@@AAA@@XXXXXXXX@@@@X@@AA@@@@@@XXXXXX@@@@@@@@@@@@@@@@@
@AAA@@@@0@@@XXX@@@@@XXX@X@@@AAXAA@@@@@@A@@@@@X@@@@@@@@@@@@X@@@@AX@@@X@@@@@@@AAAAA@@@@XXXXAAAAAAA0@@@@@@@@@AAA@@@
@@@@A@A@@@@@@@@A@X@@@@@AAA@@X@@@AXAAAAAAAAAAAAAAAA@XAAA@AAAAAAAAAAAAAA@XXAA@@@@@@@@@@@@@@@@@@@@@@AAAAAAAAAXAAAAAA
AA@X@@@@@@@@@XAXAAAAAA@@@@AAAAA@@A@@@@@@A@@@@@A@@@@@@@@@@@@@@@@@@@@@@@@@@X@@@@@@@@@X@@@@AAAA@@@@@@X@AAAAAAAAAAAAAAAX
XX@@@A@AXAAAAAAAAA@@@@@@@A@X@@@@@XX@@@@X@A@@@@@@@AAAAAAXXXXXXXXXXXXXXXXXXXXAAAAAAA (Tail)

 generation: 70

133 new cells

90 dead cells

668 total cells
```

Figure 4.6- Overproliferation. When chromosome missegregation chromosomes, the disruption of homeostasis ensues under certain circumstances. The text- based output indicates that Aneuploid cells (denoted as "A" or "@") replace diploid cells (denoted as "X"), and have broken through the homeostatic carrying capacity of the tissue (200 cells).

Also, because of the high variability, statistical comparison across experiments cannot be made.

A contributing factor to this exponential increase in genotype diversity is that the genes can be

119

shuffled in 15 different combinations (such as Apoptosis and Division in Chromosome1 and Segregation, Stability and Contact inhibition in Chromosome 2; or Apoptosis and Stability in Chromosome 1 and Division, Segregation and Contact inhibition in Chromosome 2) and the genotypes that result from that reach the dozens by the time cancer-like behaviour initially arises. This suggests that, in order for a more direct comparison, the effects of chromosome missegregation should be dampened and the number of genotypic combinations should be reduced. This will be reflected in the next iteration of the model.
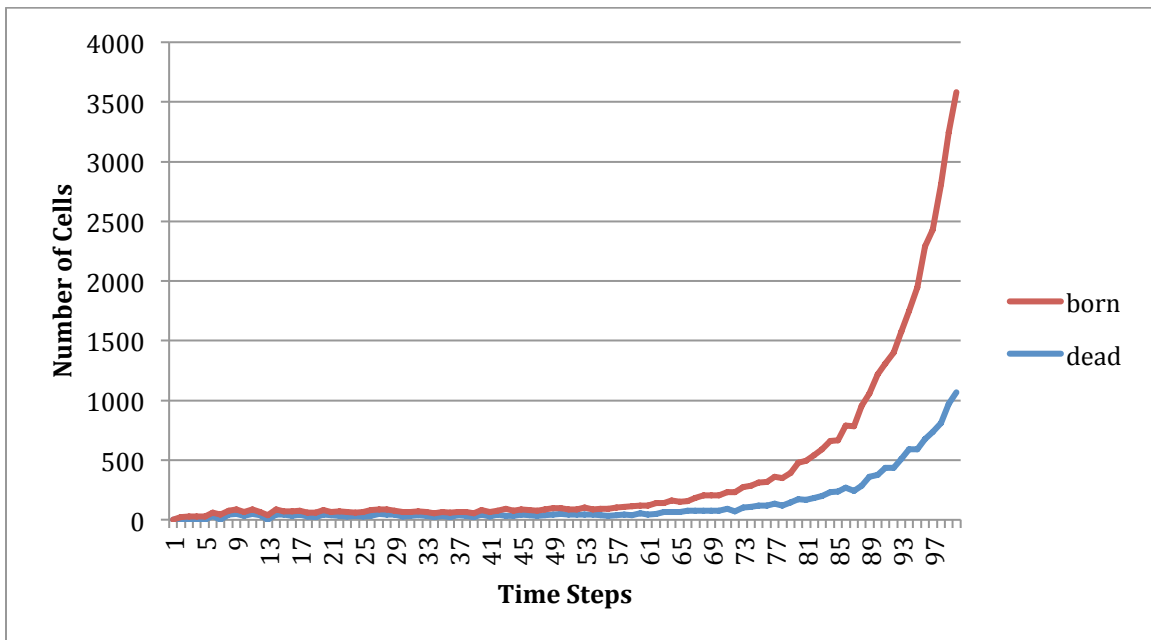
Figure 4.7- A plot of the number of dead cells and born cells for each time step. The effects of chromosome missegregation unbalance the rate between cell death and cell birth. While some death mechanisms still remain active, massive death can still be overrun by even greater cell birth. This leads to emergent overproliferation.
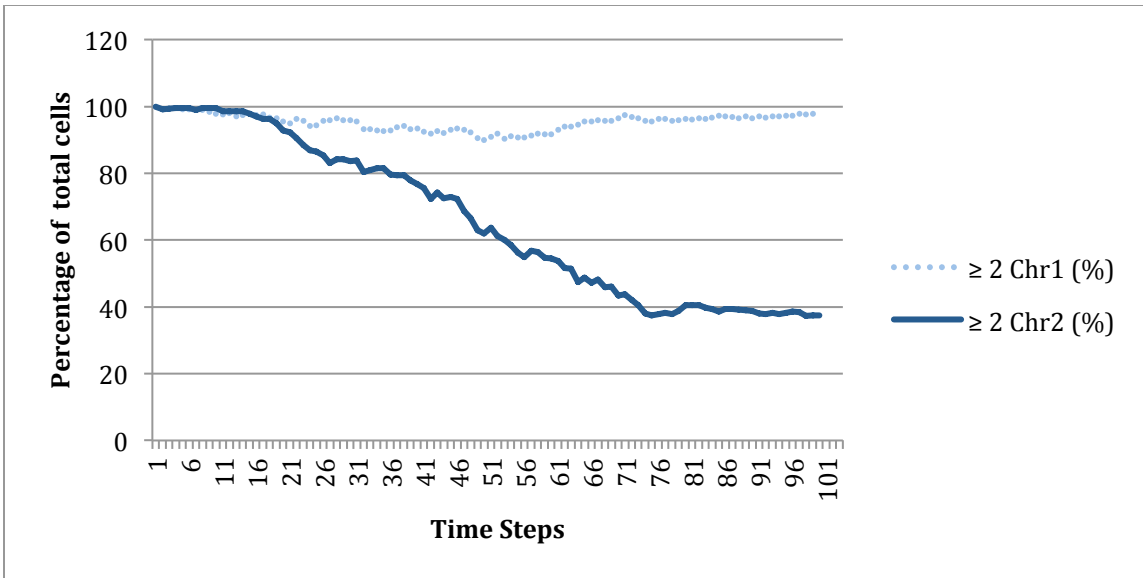
Figure 4.8- A plot of the average number of chromosomes per time step throughout an over proliferative simulation. Depending on the genetic arrangement, cells lose or gain copies of a given chromosome, depending on the advantaged that it brings. Losing chromosomes with genes that regulate apoptosis and gaining chromosomes that contain copies of genes that promote division is a trend observed in overproliferative simulations.
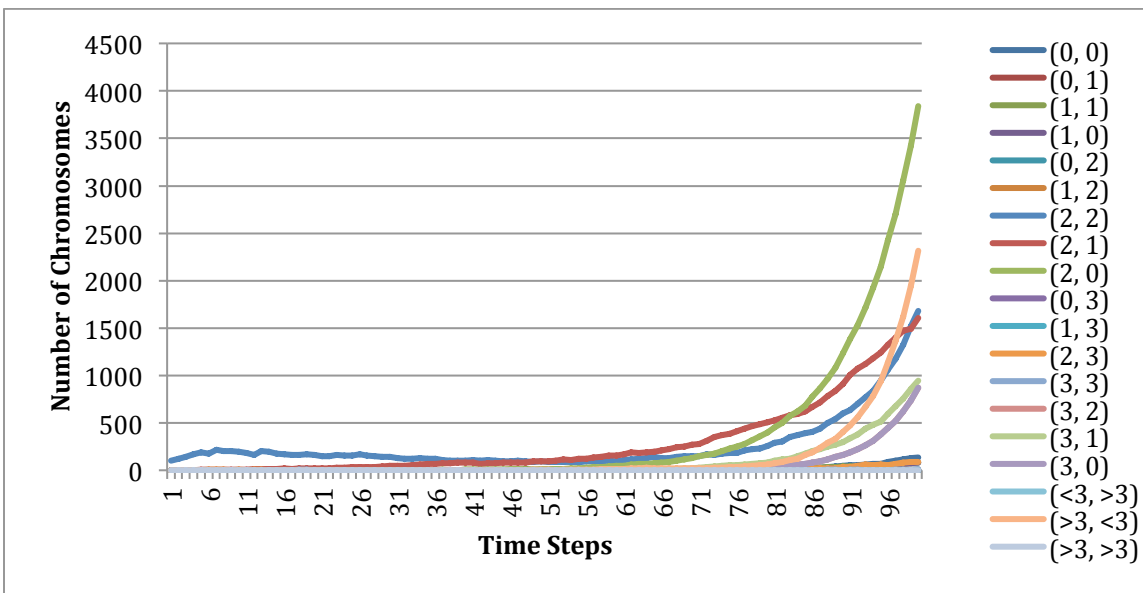
Figure 4.9- An illustrative plot of the number of cells with a given genotype. The numbers in the parenthesis denote the number of copies of Chromosome 1 and Chromosome 2 respectively. Different genotypes emerge though an overproliferative simulation. When a successful overproliferative genotype is reached, cells are able to grow from the hundreds to the thousands in a few dozen time steps. Depending on the genetic configuration, some cells consistently lose a chromosome and gain another. However, this analysis tool needs to be refined and more time steps are needed in the simulation to understand the population dynamics.

121

4.4.3. Model Version 2 Evaluation

As a new iteration of the model is reached, a new assessment is to be made through Web's methodology. It is important to remember that many Webb's criteria are qualitative and cannot be quantified. The effort of the collaborating biologists comes as a crucial part of this assessment. This iteration of the model stands in Webb's seven dimensions as follows:

- **Relevance**: The modelling of genes within chromosomes makes this a complete, dynamical model of homeostasis and aneuploidy. This is biologically relevant, since important questions can now be addressed, such as, which genetic combinations lead to the overproliferative behaviour observed in Figure 4.5.

- **Level**: The genetic component compliments the cellular components and brings out emergent behaviour at the tissue level, as seen in Figure 4.2.

- **Generality**: This is a model more specific to the phenomenon of aneuploidy than the previous general model of homeostasis. Because the model considers chromosome missegregation as the only form of variation, this model cannot be applied as is to other kinds of aberrations. However, the general nature of the framework is general and can be easily adapted to represent diverse genetic arrangements, as seen in Figure 4.4.

- **Abstraction**: The abstraction of genes was a key part of the modelling process. The abstractions regulate a direct analogy of the biological features of death, division and segregation. At this stage, discussion with molecular biologists resulted in the recommendation that this model is overcomplex and difficult to analyse.

- **Accuracy**: The representations of genes in the model accurately reflect the dynamic mechanism of chromosome missegregation and the disruption of homeostasis. However, the rate of chromosome missegregation damped by segregation genes is assumed to be equal to the effect of the other genes, as current research suggests (Kunda et al., 2008). However, this high rate of chromosome missegregation leads to high variation across experiments. A lower rate of chromosome missegregation could be considered in the next iteration, which may lead to more reproducible results.

- **Match**: The behaviour now matches the literature and current biological theories on what is observed experimentally such as the existence of cells with excess chromosomes due to chromosome missegregation (Rajagopalan & Lengauer, 2004), imbalances in gene

122

expressions (Torres et al., 2008); and clinically such as the emergence of various degrees of aneuploidy and chromosomal instability (Swanton et al., 2006), and the association of higher chromosomal instability with a worse prognosis (Walther, Houlston, & Tomlinson, 2008). However, this is at the expense of computational feasibility. Results can hardly be compared amongst themselves due to the high variability in the data. For the next iteration, the growth rates of cells should be proportionately reduced to make the model more computationally feasible. Care must be taken so that this re-scaling does not affect the overall dynamics of the system.

- **Medium**: The computational model provides complete transparency of the evolutionary pathways of each cell in each simulation. However, this evolutionary tractability of the genotypes proves to be challenging because of the large amount of data that is being calculated and the number of variables that are in play. Also, simulations are highly variable. Some take minutes, while other take hours or even days. In some simulations, the limits of the hardware were reached and, by necessity, were aborted after days of simulation. A restriction on the number of cells that the hardware can comfortably model needs to be established.

A quantitative assessment of the changes with respect to the previous version of the model can be appreciated in Table 4.2.

Table 4.2- Assessment of the Model Version 2 with respect to Webb's seven dimensions, from 0 to 5 stars. The metric ranges from 0 to 5 stars, where more stars reflect an improvement over the model. Stars in black are an improvement over the previous model rating in grey.

Dimension	Relevance	Level	Generality	Abstraction	Accuracy	Match	Medium
Assessment	✳✳✳★	✳★★	✳✳✳★	✳✳★	✳★★★	✳✳★★	✳✳✳

4.5. Model Refinement and Final Implementation

The model, as evaluated in the previous section (Table 4.2), successfully recreates the ideal experimental conditions in which the key questions of this investigation can be addressed. This model has been verified against the literature and approved by molecular biologists at the Baum

Lab as accurate and relevant. However, the data obtained by exploring the new cell population that emerges in time, quickly reveals a high degree of complexity arising from the interaction of the few constituents. This complexity is such that it becomes unclear which are the key elements and interactions that give rise to the emergent cancer-like behaviour. To address this problem, further refinements and simplifications are required.

4.5.1. Goal 3.1: Simplifying the Model

Further abstracting and re-implementing the crucial functions that give rise to homeostasis can reduce the number of genes used. The key motivation for these refinements is to have the simplest model possible that allows for the traceability of key transitions (such as the kind of chromosome missegregation event that leads to the emergence of an overproliferative genotype or the loss of tumour suppression in tandem with varying degrees of chromosome stability), while at the same accounting for the biological behaviours that are observed experimentally. Aspects that were considered important to reframe or exclude via this process are:

- **Genes that regulate apoptosis via contact inhibition**. These genes were merged with the genes that regulate stochastic apoptosis since their function was the same: the negative regulation of proliferation.

- **Genes that maintain genetic stability by inducing apoptosis when there are insufficient chromosomes**. Although, in reality, cells that have too few chromosomes usually die, this biological feature is not relevant for the process being investigated. Also, this feature was considered by the molecular biologists at the Baum Lab to be overabstracted, since the number of chromosomes was also highly abstracted. Because of this, this gene is to be removed.

4.5.2. Goal 3.2: Model Calibration and Final Implementation

To obtain more detail throughout the simulations while maximizing processing time, it was decided that a simulation would end when 7000 cells had been reached or after 200 time steps. This modelling decision is based on calculating a balance between the highest number of cells that can be currently simulated, and the need to explore the divergence of genotypic trajectories that the initial and subsequent chromosome missegregation create. This would model in real life a square area of $2x10^{-4}$ m of a constantly renewing epithelial tissue, or a typical 96-well wet lab experiment. Also, the rate of chromosome missegregation has been proportionately reduced to make the model more computationally feasible. Care was taken so that this re-scaling did not affect the overall dynamics of the system. These new limits allow a better comparison across the simulations.

The final genes simulated and their functions, described below, are:

- **Apoptosis regulatory genes**, as depicted in Figure 4.10.A, are an abstraction of tumour suppressor genes such as tRb (Amato et al., 2009) and p53 (Matlashewski et al., 1984) that regulate cell death, and enables us to model the fact that tissue crowding leads to a corresponding increase in the rate of delamination and cell death within an epithelium to maintain homeostasis (Zeng & Hong, 2008) (Marinari et al., 2012).

- **Cell division regulatory genes**, as shown in Figure 4.10.B, provide an abstraction of proto-oncogenes such as Ras (M. Wu et al., 2010), Myc (Soucek et al., 2008) and p110 PI3K (Brunet, 2009) and act to promote cell growth and cell cycle progression. Again the action of these genes is sensitive to the "homeostatic capacity" of the tissue in order to model the process know as contact inhibition (Carmona-Fontaine et al., 2008) that limits cell proliferation in crowded tissues.

- **Chromosome segregation regulatory genes**, as depicted in Figure 4.10.C, model genes controlling the fidelity of cell division such as BUB1 (Ricke et al., 2011) and MAD1/MAD2 (Sotillo et al., 2007) that reduce the likelihood of chromosome missegregation at cell division. The ERM gene, whose expressed proteins crosslink actin filaments with plasma membranes and regulate key cell functions in mitosis, will also be considered, based on experiments described in Chapter 7.

A.   B.   C.
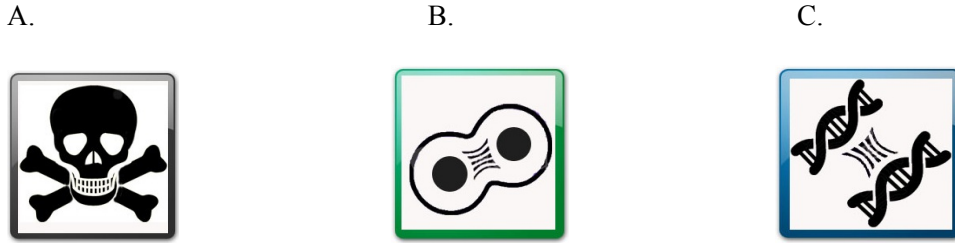
Figure 4.10- Gene abstractions in the final model. A. Apoptosis Regulatory Genes that control cell death via contact inhibition. B. Cell division regulatory genes promote cellular division and wound healing. C. Chromosome segregation regulatory genes ensure fidelity during replication.

Through an incremental approach, a final iteration of the model was reached. In this implementation, each cell has a simulated genome composed of three kinds of genes, as depicted in Figure 4.10. Each of the three genes code for corresponding actions at a cellular level, inspired by biological systems. Inspired by the processes in biological cellular behaviour through which homeostasis is maintained in organisms, the algorithm (with changes in bold) is as follows:

1. An initial population of 100 cells is created, each with diploid chromosomes. Each initial genome was equipped with 2 copies of each type of gene. The normal carrying capacity of the tissue is fixed at 200 cells.

2. For each time step, the total number of cells is measured and is not updated until the next time step. Each cell is examined every time step.

3. **If a measurement of the total number of cells is greater than the tissue's carrying capacity, then the probability of cell death is calculated**. The probability of death depends on the number of available copies of the apoptosis regulatory genes, N_{ap}, within each cell's genome. The probability of apoptosis, p_{ap}, is determined by:

$$p_{ap} = r_{ap} N_{ap}$$

where r_{ap} is a parameter for the rate of apoptosis. The cell is then killed with a probability of p_{ap}.

4. If the cell has not died, it has a chance to divide. The probability of division depends on the number of available copies of the division regulatory genes, N_{div}, and a parameter that determines the rate of division, r_{div}. The probability that a cell divides, p_{div}, is:

$$p_{div} = r_{div} N_{div}$$

5. If dividing, the probability of chromosome missegregation is calculated. If there is a chromosome missegregation event, one of the 4 types of chromosomes is chosen at random. If the cell still has copies of this chromosome, it is asymmetrically distributed during cell division leading to the creation of two aneuploid cells. Otherwise, the genome is duplicated and copied with fidelity, thus generating two identical daughter cells. The probability of chromosome missegregation, p_{msg}, in the model is:

$$p_{msg} = r_{msg} \ (4\text{-}N_{mseg})$$

where N_{mseg} is the number copies of the chromosome segregation regulatory genes within the cell's genome, and r_{msg} is a parameter for the rate of chromosome missegregation.

To investigate the properties and the dynamics of the system, and specifically the role that chromosome segregation regulatory genes have, four genome distributions were considered, and will be discussed in detail in the next chapter. The parameter settings were determined through the literature and a series of preliminary experiments, in order to ensure that the behaviour of the system was both biologically plausible and computationally feasible. After calibrating the simulations with the literature available and with feedback from molecular biologists from the Baum Lab to accurately reflect what we know so far of the interactions, the following initial parameters were decided upon:

For biological plausibility, from the literature discussed in section 4.3.2:

- Intrinsic rate of apoptosis (gene adjusted): r_{ap} =0.045
- Intrinsic rate of division (gene adjusted): r_{div} =0.045
- Intrinsic rate of chromosome missegregation (gene adjusted): r_{msg} =0.02

For computational feasibility:

- Initial population: 100 cells
- Homeostatic size of the tissue: 200 cells
- Simulation end time: when reaching 7000 cells or reaching 200 time steps

These refinements conserve the same emergent properties of the previous successful iteration of the model, and allow for the effects obtained through simulations to be traced to particular events with a better clarity regarding the key players and interactions involved.

The model was tested under diverse genetic initial conditions, which will be the focus of the next chapter. After these tests, it was found that the reduction in components did not affect the emergent behaviour (Figure 4.11). The total number of combinations for the different possible genetic arrangements has been effectively reduced from 15 to 3, giving us a much clearer view of the dynamics of the genetic populations.
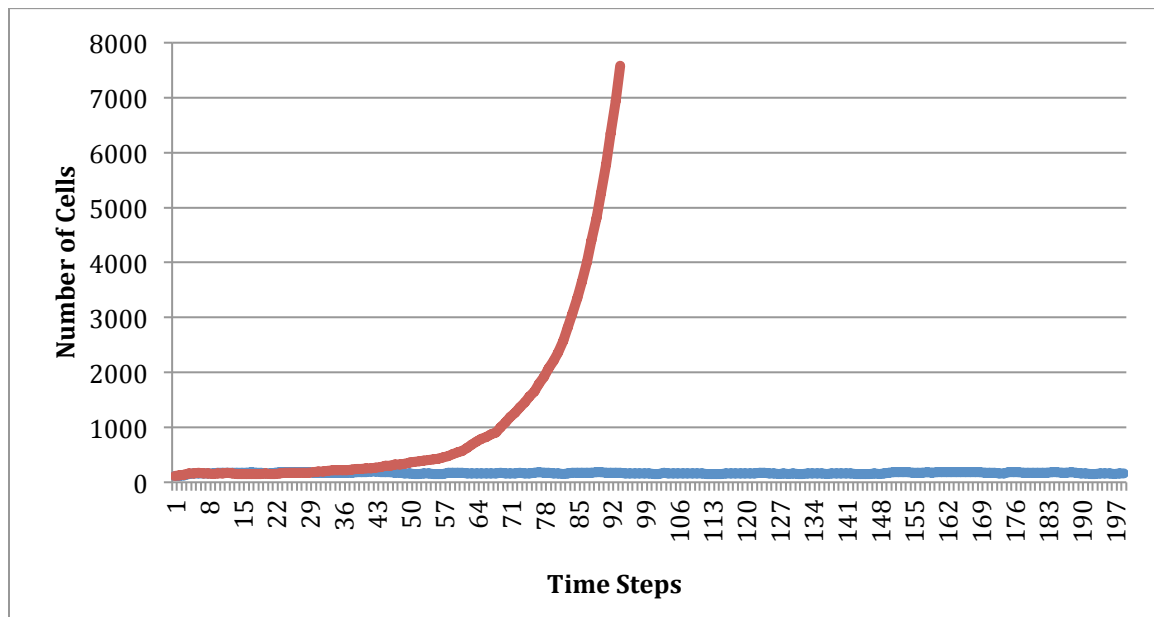
Figure 4.11- A representative plot of the total number of cells for two simulations with different genetic initial conditions. Homeostasis can be invariantly observed for one genetic configuration and Emergent overproliferative behaviour was invariantly observed for the other. Although the components were rescaled, the Model Version 3 not only conserves all of the key behaviours of the previous models, but also allows for a more transparent analysis of the behaviour, which is the focus of Chapter 5.

Using the discussed agent-based model, in silico experiments can be carried out to explore the different genetic arrangements and their properties. There is also a possibility to use this framework to simulate the interaction between chromosome missegregation and abstractions of

cancer treatments such as of surgery, the physical removal of tumour mass; and chemotherapy, a treatment where overproliferating cells are targeted and killed.

4.5.3. Model Version 3 Evaluation

As the final iteration of the model is reached, it is key to measure the model against reality through comparing it with the available literature, and with feedback from biologists (and real data when available). Following Webb's methodology, we found that the final iteration of the model stands in Webb's seven dimensions as follows:

1. **Relevance**: The model is an abstraction of current biological knowledge. It will be used to assess a new hypothesis and the behaviour that can occur.
2. **Level**: The elemental units of the model are cells. Each cell is equipped with an internal genome, which can change through time. The collection of cells makes up an organized cellular system, abstracting the formation of a homeostatic tissue. Changes in the genomes of individual cells give rise to different cell behaviours and emergent properties at the tissue level.
3. **Generality**: The system can represent organized cellular systems with diploid genomes. It has been designed with the abstractions of key factors in carcinogenesis. Although not specific to a single type of cancer, this general model can be expanded or tailored to a more specific type of cancer. Also, the time frame of this model is scalable. It could represent days in *in vitro* experiments (such as cultured cells), or months in *in vivo* studies (such as animal models) or clinical settings (in real patients).
4. **Abstraction**: A general model of carcinogenesis, the model is specifically cellular. For computational tractability, necessary simplifications in the individual genetics of each cell were made. Abstractions of cellular behaviour have a solid basis in cell biology.
5. **Accuracy**: The biological mechanisms represented in the model are a simplification of the real, mechanisms, which are too complex or still subject of intense research, to be included mechanistically. A compromise was reached by scaling down accuracy in exchange of a more abstract, but computationally feasible model. The abstractions,

however, accurately represent the triggers and key results of the mechanisms relevant to the hypothesis addressed in this work.

6. **Match**: Experiments that will be discussed in the next chapter show that the behaviour obtained in the model matches the behaviour found in real cancers. The key genetic properties, as explored in depth in the next chapter, also match qualitatively experimental results.

7. **Medium**: The simplifications made in this final implementations result in more manageable data. Through this implementation, it is possible to isolate the effects of the different kinds of genes, without losing the rich emergent behaviour obtained in the previous iteration. The computational resources (2.4 GHz Intel Core 2 Duo processor with 4 Gb of memory). were taken to the limit of their capacities and we are confident that the framework is efficiently taking advantage of all the resources available. Simulations usually take minutes to run and the output files occupy only a few gigabites of hard drive space.

After this successful assessment, it was decided that the compromise between accuracy and feasibility had been successfully met. With this implementation, key questions can be formulated in terms of the model, and confidently addressed.

Table 4.3- Assessment of the Model Version 3 with respect to Webb's seven dimensions, from 0 to 5 stars. A star in accuracy was lost in favour of reaching 5 stars in all of the other dimensions.

Dimension	Relevance	Level	Generality	Abstraction	Accuracy	Match	Medium
Assessment	*****✱	*****✱✱	*****✱	*****✱✱	✱✱✱	*****✱	*****✱✱

4.6. Summary

To address whether chromosome missegregation plays a causal role in the course of a cancer we developed a model of tissue homeostasis in which to study cancer evolution. Individual cells were modelled, each equipped with a genetically defined genome, as agents in a computational

simulation. A collection of these cells makes up a tissue that initially exhibits homeostatic behaviour, as the result of balanced rates of cell proliferation and cell death. These were modelled as stochastic processes that are regulated at a genetic level, based upon the properties of known proto-oncogenes and tumour suppressor genes (Futreal et al., 2004). We made the key abstractions of a single gene regulating a specific behaviour, and that the impact of each gene is proportional to the number of copies of a given gene found in the genome of each cell, as suggested by recent studies on the effects of differences in chromosome number on gene expression in biological systems (Huettel et al., 2008). Having established this model system, we then introduced genes that regulate the rate of division, apoptosis and chromosomes segregation. After each incremental refinement of the model, an assessment using Webb's modelling guidelines (B. Webb, 2009) was carried out. This assessment included comparing the model with the current literature, the feedback from molecular biologists at the Baum Lab and the incorporation of real data whenever possible.

After several iterations, a final version, which enables us to test the role of evolving chromosomal instability in cancer development and treatment, was reached. In this model, we can isolate the effects of chromosome instability, tumour suppressor and oncogene activity and genetic linkage on cancer progression.

5. Simulating Chromosome Missegregation

This chapter focuses on chromosome missegregation and the role that the distribution of genes across the genome have on tumourigenesis. Through this chapter, the key properties of the model will be investigated. The chapter starts with a recapitulation of the key components in the model derived in the previous chapter. Following this, key questions regarding the role of the genetic arrangement on the behaviour observed will be addressed. For this, new analysis tools had to be created, which will be explained. With these analysis tools, the model will be studied and new insights will be drawn. The chapter will end with a brief summary of the chapter.

5.1. Introduction

In the previous chapter, it was described how the relevant biological concepts were determined, abstracted and incorporated into a working model. The model aims to capture the dynamics and behaviours of the complex biological phenomenon of chromosome missegregation. In the model, each cell in the initial population has two sets of identical chromosomes: a diploid genome. Just like in real life, when dividing, the genome of each cell is duplicated and the two sets of chromosomes are then segregated into two daughter cells. It is during this stage that chromosome missegregation events can occur, resulting in asymmetric cell division: one daughter cell with an extra chromosome, and one lacking the same chromosome.

In the final version of the model, each cell in the system has a simulated genome composed of three kinds of genes:

- **Apoptosis regulatory genes**, as depicted in Figure 4.10.A, are an abstraction of tumour suppressor genes such as tRb (Amato et al., 2009) and *p53* (Matlashewski et al., 1984)

that regulate cell death, and enables us to model the fact that tissue crowding leads to a corresponding increase in the rate of delamination and cell death within an epithelium to maintain homeostasis (Zeng & Hong, 2008) (Marinari et al., 2012).

- **Cell division regulatory genes**, illustrated in Figure 4.10.B, provide an abstraction of proto-oncogenes such as Ras (M. Wu et al., 2010), Myc (Soucek et al., 2008) and p110 PI3K (Brunet, 2009) and act to promote cell growth and cell cycle progression. Again the action of these genes is sensitive to the "homeostatic capacity" of the tissue in order to model the process know as contact inhibition (Carmona-Fontaine et al., 2008) that limits cell proliferation in crowded tissues.

- **Chromosome segregation regulatory genes**, represented in Figure 4.10.C, model genes controlling the fidelity of cell division such as BUB1 (Ricke et al., 2011) and MAD1/MAD2 (Sotillo et al., 2007) that reduce the likelihood of chromosome missegregation at cell division. ERM will also be considered, based on experiments described in Chapter 7.

A.   B.   C.
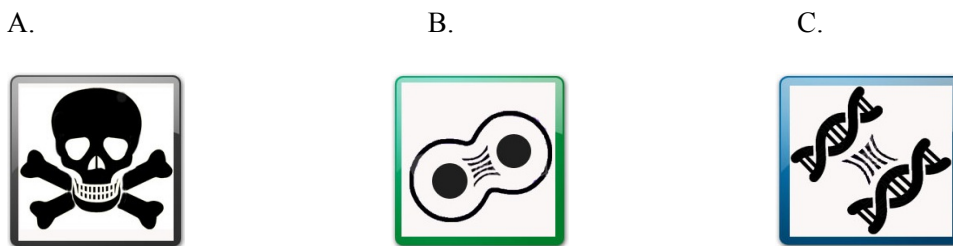
Figure 5.1- Gene abstractions in the final model. A. Apoptosis Regulatory Genes that control cell death via contact inhibition. B. Cell division regulatory genes promote cellular division and wound healing. C. Chromosome segregation regulatory genes ensure fidelity during replication.

It was determined through Webb's assessment, as discussed in the previous chapter, that the agent based model captures the essence of the biological phenomenon of chromosome missegregation.

Through the creation of the model, a set of analysis tools was created to understand the dynamics of the system. Measurements were taken from the output of the simulations with an algorithm created in Mathematica. After a parameter readjustment, as discussed in the previous chapter (Section 4.5.2), a measurement of the total cell number per time step revealed that the readjusting of the growth, division and missegregation rates have a scaling effect on the overall behaviour. However, it is the actual distribution of the genes across the chromosomes that has a direct impact on the kind of behaviour obtained. This chapter will investigate the properties that the genetic arrangements that lead to homeostasis or to over proliferation have, their initial mechanism of action and how they shape the subsequent evolution of the system.

5.2. Answering Questions with the Model

The aim of this work is to discover if the model provides an effective approach to assess the role of aneuploidy in the evolution of cancer. For this, it was decided that the most relevant question to be tackled is one that could not be easily modelled *in vitro* or *in vivo*, and which could be used to further inform experiments in whole chromosome missegregation. From the literature review, it was clear that there is a gap in the understanding of chromosome missegregation as a mechanism for aneuploidy. This is because the exact role, location and linkage of the key genes regulating cell growth, death and chromosome segregation in real human chromosomes is currently unknown (Rajagopalan & Lengauer, 2004). However with an abstract computational model of this complex system it would be possible to explore how differences in the distribution of genes on chromosomes affects the evolution of the system as a whole. The model developed in this work is in an advantageous position to address such a complex interaction with a transparent computational analysis. Results from this model could be used to inform cancer

biologists on the kinds of behaviour and their organizing principles that can be found in real life chromosome missegregation.

The first investigation using the model will aim to answer the following questions:

1. Are there any kinds of genetic arrangements through which chromosome missegregation may lead to the disruption of homeostasis?
2. If there are any overproliferative behaviours, could they be classified according to their evolution given a set of genetic constraints?
3. Are there any key genetic transitions that could result in the disturbance of homeostasis?
4. Does the distribution of genes in chromosomes have a role on the disturbance of homeostasis and over proliferation through the mechanism of chromosome missegregation?

To answer these questions, key measurements are needed of the simulations. For this, a Mathematica algorithm was created. The model, written in C++ outputs a text file containing the state of every cell (the number and kinds of chromosomes that it has, as well as its age) for every time step of the simulation. The Mathematica algorithm extracts such information and transforms it into something more statistically manageable. The Mathematica algorithm uses as input the information on individual cells and calculates the average number of copies of a given gene throughout the cell population for each time step. Also the Mathematica calculates the genetic diversity (how many different genotypes comprise the cell population) per time step of the simulation. The average rate of death, division and segregation, and the genetic diversity per time step are then plotted in data software such as Excel to give us a clear picture of the dynamics of the model.

5.3. Gene Distribution Across Chromosomes

Seeking to bridge a gap in our understanding of aneuploidy that cannot be currently addressed with biological experiments, an experiment was designed to understand the role of the distribution of genes across chromosomes in chromosome missegregation. The focus of this experiment is to change the initial genetic makeup and genetic restriction of cells and measure the effect that this has on cell death, division, segregation and genetic diversity.

To do this, it was decided to place the abstracted genes in the three different chromosomal configurations (Figure 5.2). These are distribution A, where apoptosis regulatory genes and cell-division regulatory genes are "linked" in the same chromosome; distribution B, where cell-division regulatory genes and chromosome segregation regulatory genes lie on the same chromosome; and Distribution C where genes regulating apoptosis and chromosome segregation are genetically linked. At the start of simulations each cell was then modelled as diploid, containing two copies of each chromosome.
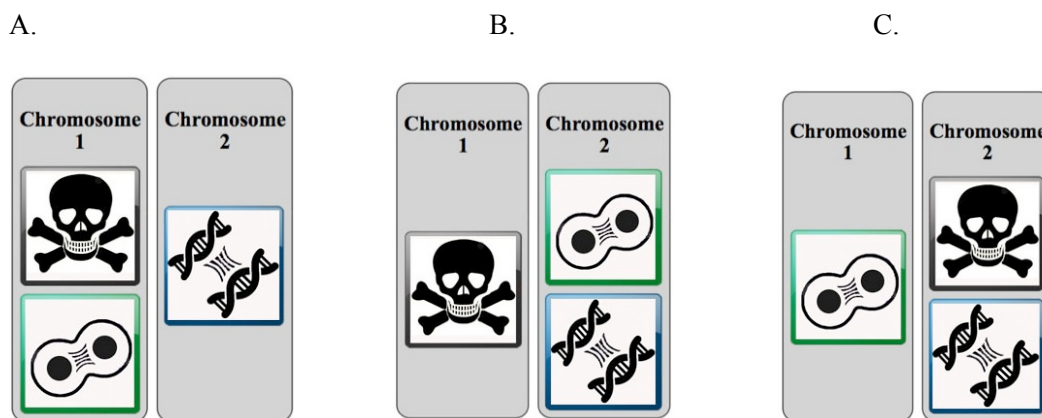
Figure 5.2- The different gene abstractions were placed into chromosomes in 3 different configurations, which led to different kinds of linkages between the genes. A. Gene Distribution A: apoptosis and division genes in Chromosome 1; segregation genes in Chromosome 2. B. Gene Distribution B: apoptosis gene in Chromosome 1; division and segregation genes in Chromosome 2. C. Gene Distribution C: division gene in Chromosome 1; apoptosis and segregation gene in Chromosome 2. The icons for each kind of gene are described in Figure 4.10.

The evolutionary dynamics in the model are then determined by the gene expression of the individual cells and the global behaviour that emerges through cell death, proliferation and missegregation over time. By quantifying the number of chromosomes that a cell has at a given time, a **genotype state** G_T is defined as:

$$G_T = (N_{div} , N_{ap}, N_{msg})$$

where N_{div} , N_{ap} and N_{msg} are the number of copies of *Cell Division Regulatory Genes, Apoptosis Regulatory Genes* and *Chromosome Segregation Regulatory Genes* respectively. The initial genotype consists of two functional copies of each chromosome: genotype state (2, 2, 2), corresponding to 2 functional copies of each gene (Division, Apoptosis and Segregation, respectively). Exploring the three different gene distributions, 100 simulations were performed for each configuration. Because instances of cell division, birth and cell death are expected to be stochastic in nature, and have been modelled as such, the behaviour of the system may be highly variable. Consistent trends may give us insights on the internal dynamics and help us extract organizing principles.

5.3.1. Experiment 1: Gene Distribution A

a) Objective and Setup

To investigate the role of the *chromosome segregation regulatory genes*, the following configuration was used:

- Chromosome 1: apoptosis regulatory genes and cell-division regulatory genes
- Chromosome 2: chromosome segregation regulatory genes
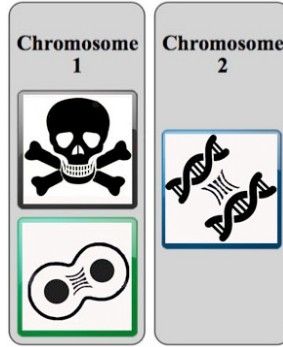
Figure 5.3- Setup for the distribution of genes into diploid chromosomes for Gene Configuration A.

This gene configuration, as seen in Figure 5.3, isolates the effects of the loss or gain of Chromosome 2 to those caused by the loss or gain of the *chromosome segregation regulatory genes*.

Figure 5.4– The graph shows the total number of cells for 100 simulations of Configuration A. Each simulation is represented as line of different colour, with the median as a thick, black line (calculated until one of the simulations came to an end). Homeostatic behaviour was always observed for Configuration A. Colours are purely used to distinguish runs and do not denote genetic distribution.

b) Results

Homeostatic behaviour can be observed in Figure 5.4. In normal conditions this kind of homeostatic behaviour provides the tissue with robustness if there were a sudden loss of cells (wound-healing capabilities), maintaining the total number of cells close to that of the carrying capacity of the tissue (200 cells).

c) Analysis

A comparison of the plot of the total number of cells across the simulations of Configuration A revealed the high variability of the simulation outcomes, as seen in Figure 5.4. Thus, it is difficult to distil meaningful information with traditional statistical methods. Despite the stochastic nature of the final cell number across experiments, an invariant qualitative behaviour can be observed for each configuration. To analyse such behaviour, measurements on the average number of genes per time step were made. These measurements were then plotted for the 100 simulations of Configuration A to generate a Broom Plot of the average number of genes per time step.

The initial genotype, genotype state (2, 2, 2), contains 2 functional copies of each gene (Division, Apoptosis and Segregation genes, respectively). Figure 5.5 shows that, on average the number of apoptosis regulatory genes is maintained around the original 2 copies per cell. The same holds true for the average number of division regulatory genes, as can be seen in Figure 5.6. It is interesting to note, however, that the chromosome segregation regulatory genes have a more pronounced behaviour, with a higher deviation across experiments towards the end of the simulations (Figure 5.7). The mean reveals that the trend is to gain more copies of the chromosome segregation regulatory genes, which makes the cell more chromosomally stable. This chromosomal stability has profound effects on genotype diversity, as can be seen in Figure

5.8, where the actual number of different genotypes increases throughout the simulation, but the rate of increase is less.

For there to be cancer-like behaviour, tumour suppressor genes need to have their function reduced and oncogenes in turn must have an increase in their expression. Because the abstracted genes that model the role of oncogenes and tumour suppressor genes are found in the same chromosome, they become auto-regulated and maintain the balance between the two. As the system evolves however, novel genotypes emerge but, because of the auto-regulation of the cancer genes, the overall behaviour generated by the new genotypes is similar to that of the original cell population, as depicted in Figure 5.4 and Figure 5.8. This leads to a micro diversity of homeostatic genotypes. The measure of diversity that will be considered in this work is the number of different genotypes coexisting at a given time point. It is of interest that the more successful genotypes naturally acquire more resistance against chromosome missegregation.

Figure 5.5- Measurement of the average number of Apoptosis Gene per time step represented in Broom Diagrams for 100 simulations. Each individual simulation is represented as line of different colour, with the median as a thick, black line (calculated until one of the simulations came to an end). Colours are purely used to distinguish runs and do not denote genetic distribution.

Configuration A
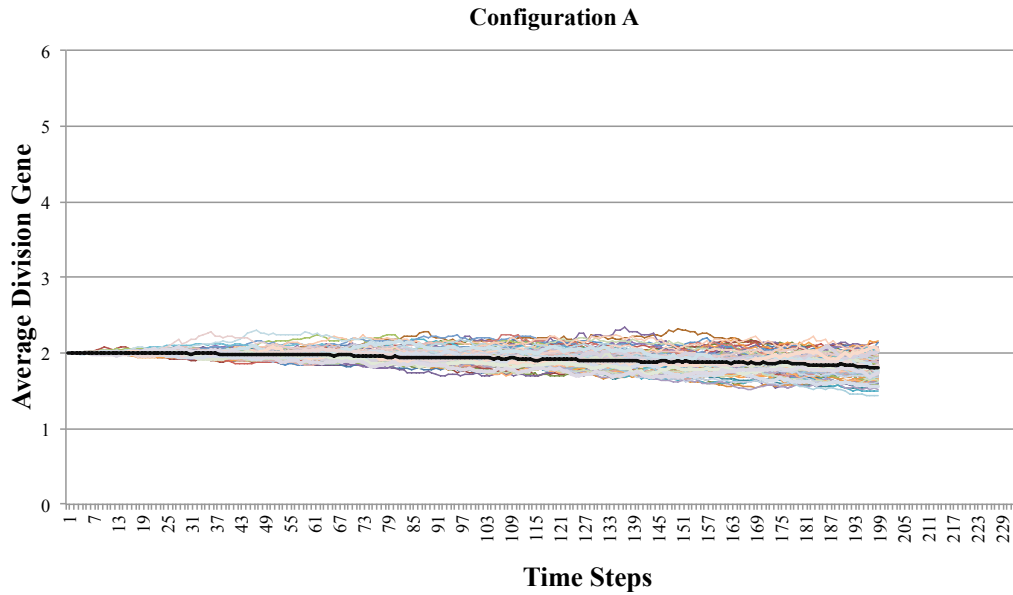
Figure 5.6- Measurement of the average number of division genes per time step represented in Broom Diagrams for 100 simulations. Each individual simulation is represented as line of different colour, with the median as a thick, black line (calculated until one of the simulations came to an end). Colours are purely used to distinguish runs and do not denote genetic distribution.

Configuration A
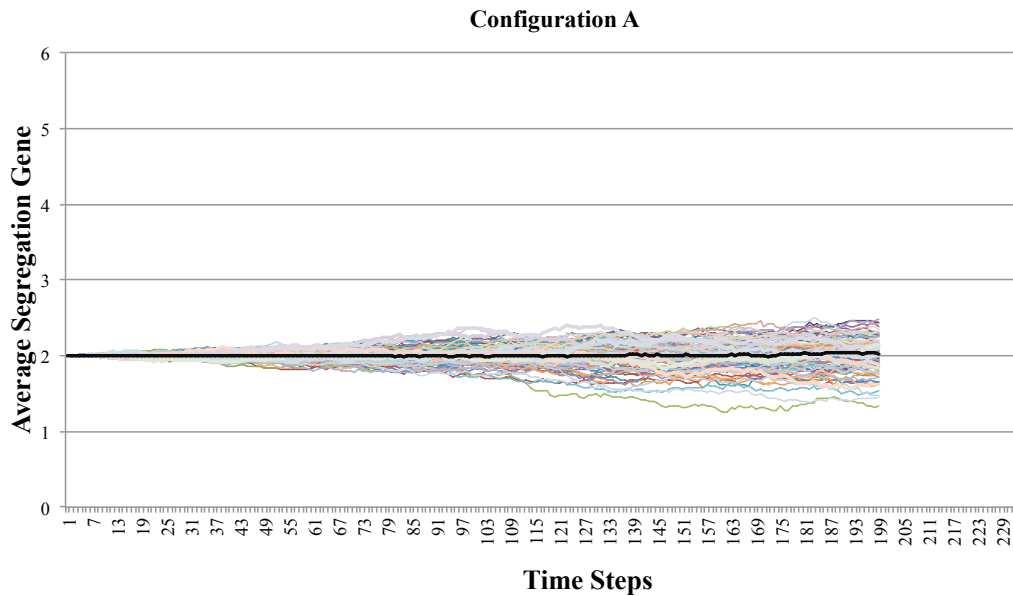
Figure 5.7- Measurement of the average number of segregation regulatory genes per time step represented in Broom Diagrams for 100 simulations. Each individual simulation is represented as line of different colour, with the median as a thick, black line (calculated until one of the simulations came to an end). Colours are purely used to distinguish runs and do not denote genetic distribution.

Configuration A
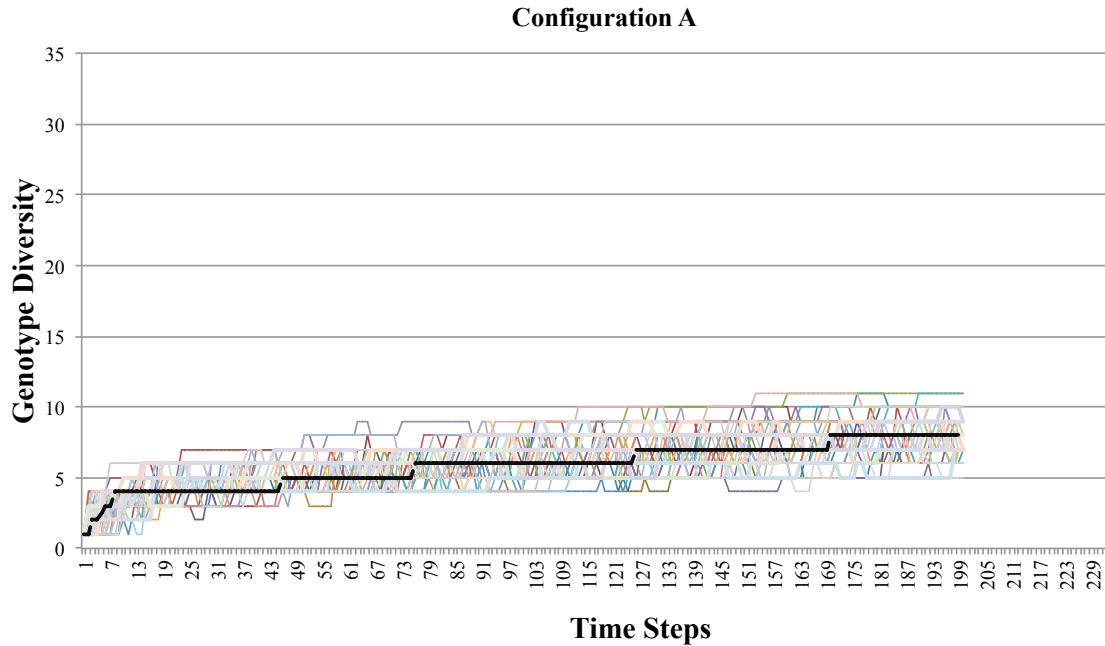
Figure 5.8 Measurement of the genotype diversity per time step represented in Broom Diagrams for 100 simulations. Each individual simulation is represented as line of different colour, with the median as a thick, black line (calculated until one of the simulations came to an end). Colours are purely used to distinguish runs. Genotype numbers stabilize towards the end of the simulations.

5.3.2. Experiment 2: Gene Distribution B

a) *Objective and Setup*

To better understand the role of the distribution of the genes in the chromosomes, the initial configuration was modified to:

- Chromosome 1: apoptosis regulatory genes
- Chromosome 2: cell-division regulatory genes and chromosome segregation regulatory genes

This gene distribution is depicted in Figure 5.9.

143
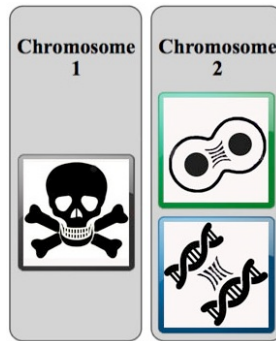
Figure 5.9- Setup for the distribution of genes into diploid chromosomes for Gene Configuration B.

Figure 5.10- The graph shows the total number of cells for 100 simulations of Configuration B. Each simulation is represented as line of different colour, with the median as a thick, black line (calculated until the time step in which any of the simulations first came to an end). Simulations were stopped when they reached 7000 cells. Cancer-like behaviour emerged through the evolution of the system.

b) *Results*

During the 100-time step experiment, a stable homeostatic behaviour can be observed for a period
of time. After that homeostatic period however, an uncontrolled proliferative behaviour follows.
The total number of cells increases exponentially, reaching the values of the order of thousands in
a very short period of time, as shown in Figure 5.10. This kind of behaviour is obtained across all
simulations.
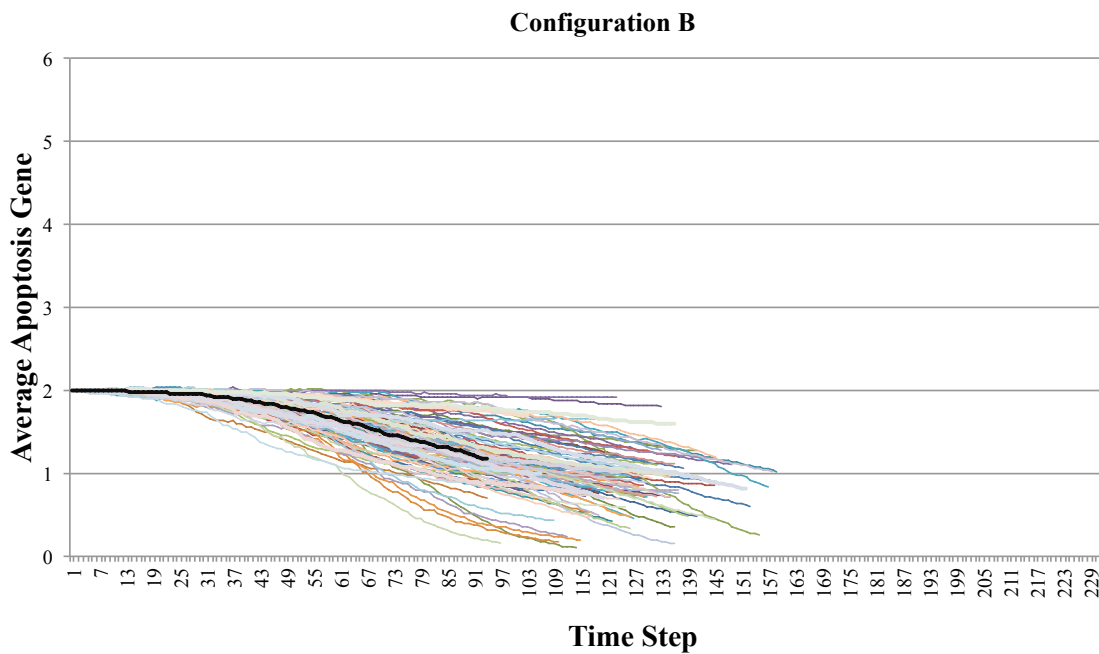
Configuration B

**Figure 5.11- Measurement of the average number of Apoptosis Gene per time step represented in
Broom Diagrams for 100 simulations for Configuration B. Each individual simulation is represented
as line of different colour, with the median as a thick, black line (calculated until the time step in
which any of the simulations first came to an end). Colours are purely used to distinguish runs and
do not denote genetic distribution. The trend is to lose copies of the genes throughout the evolution of
the system.**
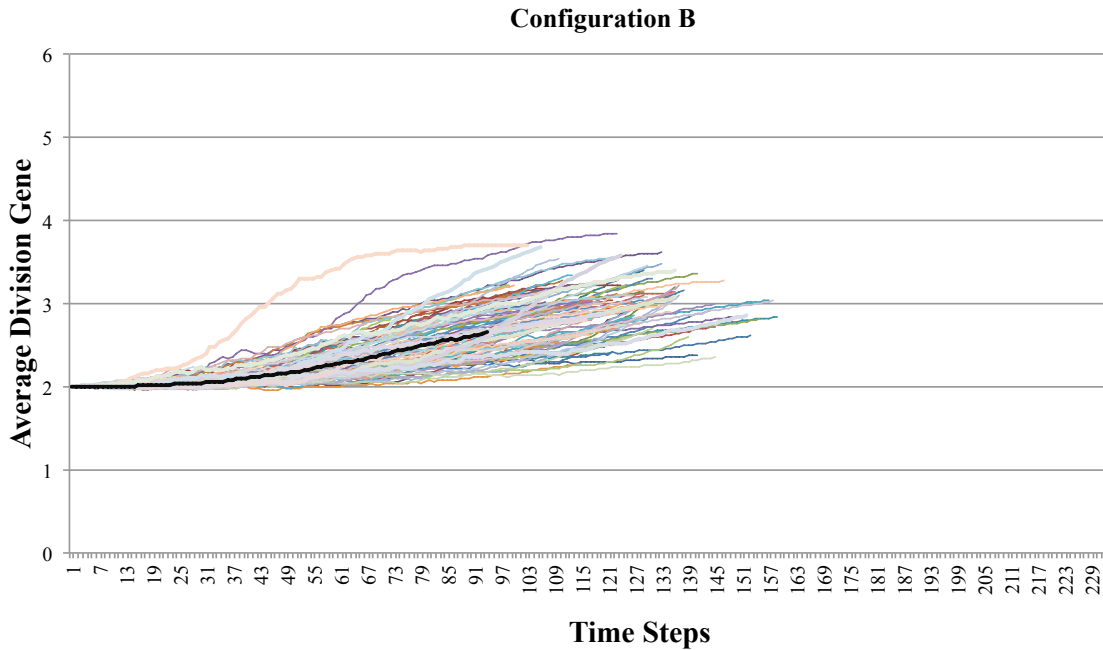
Configuration B

Figure 5.12- Measurement of the average number of Division Genes per time step represented in Broom Diagrams for 100 simulations of Configuration B. Each individual simulation is represented as line of different colour, with the median as a thick, black line (calculated until the time step in which any of the simulations first came to an end). Colours are purely used to distinguish runs and do not denote genetic distribution. The trend is to gain copies of this gene, which leads to over proliferation.

Configuration B

Figure 5.13- Measurement of the average number of Segregation Regulatory Genes per time step represented in Broom Diagrams for 100 simulations of Configuration B. Each individual simulation

146

is represented as line of different colour, with the median as a thick, black line (calculated until one of the simulations came to an end). Colours are purely used to distinguish runs and do not denote genetic distribution. The trend is to gain copies of this gene, leading to more genetically stable genotypes.
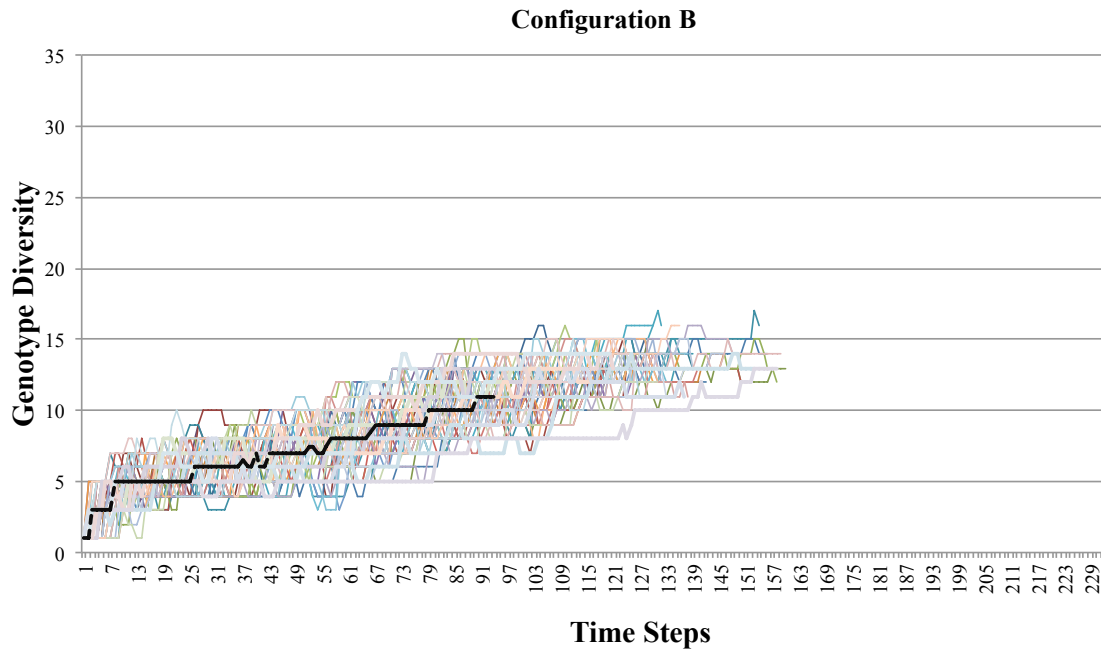
Configuration B

Figure 5.14- Measurement of the Genotype Diversity per time step represented in Broom Diagrams for 100 simulations of Configuration B. Each individual simulation is represented as line of different colour, with the median as a thick, black line (calculated until one of the simulations came to an end). Colours are purely used to distinguish runs.

c) *Analysis*

The loss of function of the tumour suppressor-inspired *Apoptosis regulatory genes* through chromosome missegregation, as seen in Figure 5.11, leads to the generation of a niche of these mutants. However, although the activation of the apoptotic mechanisms is reduced, this population remains relatively homeostatic until copies of the Division regulatory genes are gained. Also, the mean of the curves shows that there is a dampening effect on the loss of the Apoptosis genes (Figure 5.12), suggesting the stabilization of that chromosome. Because the Division genes

147

are genetically linked to the Segregation genes, the more proliferative a mutant the cell is, the more genetically stable it is (Figure 5.13). It is because of these low levels of chromosome missegregation that the genotypes evolve at a slower pace in time, as can be seen in Figure 5.14.

The evolution of the system with low levels of aneuploidy resulted in the generation of few very successful mutants that quickly dominated the entire population, suggesting a counterintuitive pathway for cancer-like behaviour with low aneuploidy. These kinds of mutations are seen in leukaemia, lymphomas and some mesenchymal tumours, where there are simple, disease-specific abnormalities (Johansson et al., 1996).

5.3.3. Experiment 3: Gene Distribution C

a) *Objective and Setup*

The next logical combination to try, as seen in Figure 5.15, is:

- Chromosome 1: cell-division regulatory genes
- Chromosome 2: and apoptosis regulatory genes and chromosome segregation regulatory genes
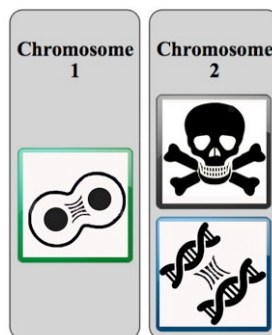
Figure 5.15- Setup for the distribution of genes into diploid chromosomes for Gene Configuration C.

b) **Results**

This configuration results in cancer-like behaviour (Figure 5.16). Although similar to the overproliferative behaviour that was obtained through the simulations with Gene Configuration B, as seen in Figure 5.10, there are significant differences.
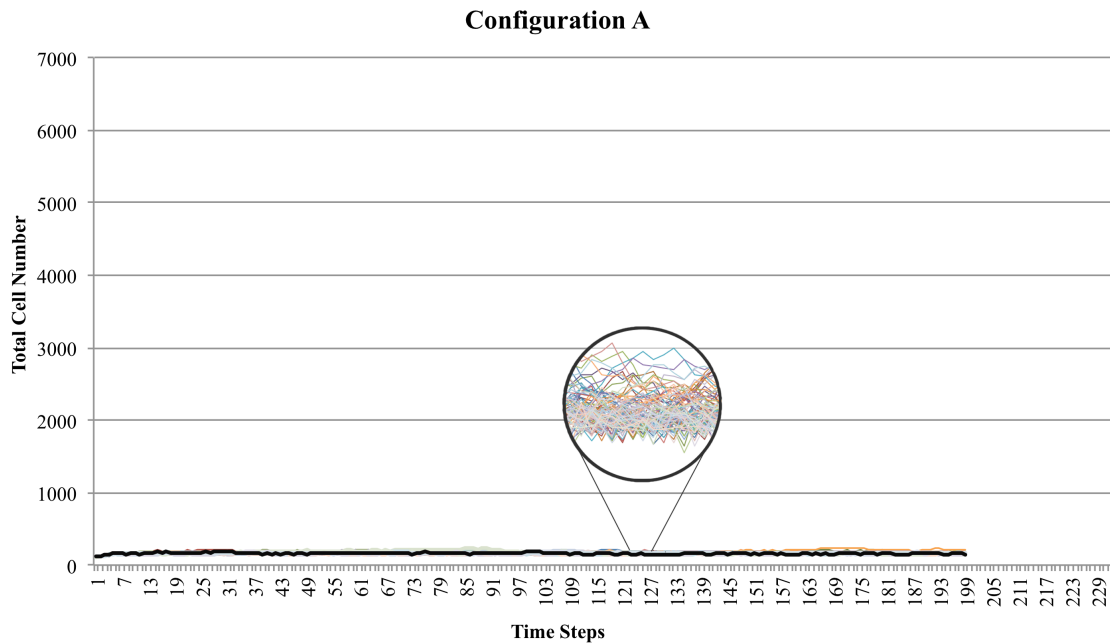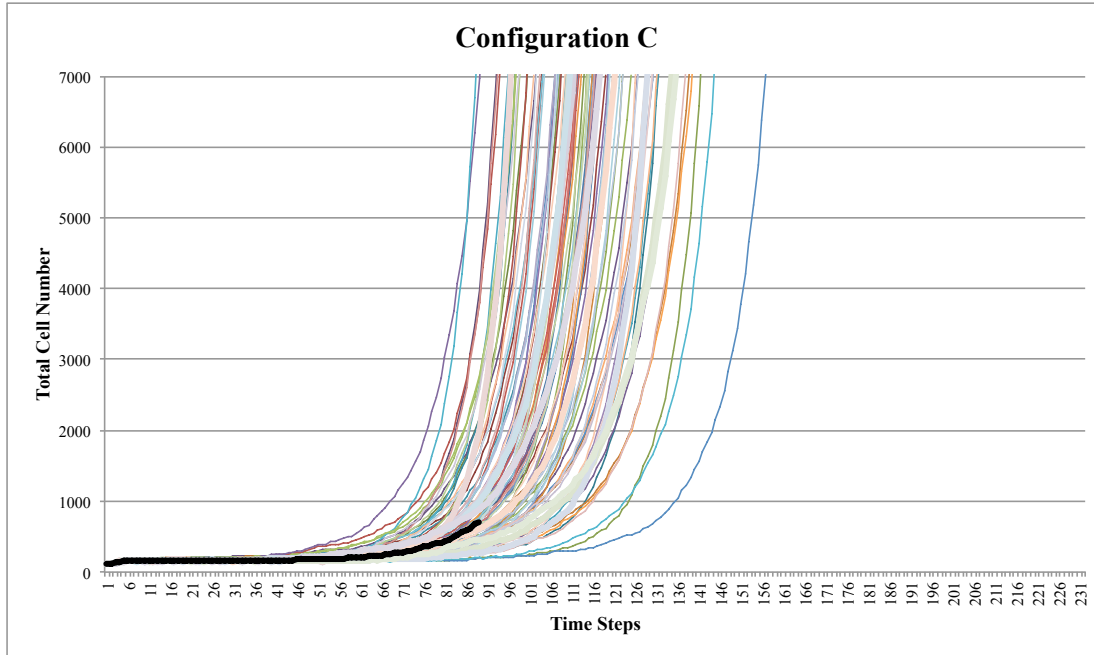

Figure 5.16- The graph shows the total number of cells for 100 simulations of Configuration C. Each simulation is represented as line of different colour, with the median as a thick, black line (calculated until one of the simulations came to an end). Simulations were stopped when they reached 7000 cells. Cancer-like behaviour emerged through the evolution of the system, but at a faster rate than that of Configuration B (Figure 5.10).

c) **Analysis**

An analysis of the loss and acquisition of genes in the evolution of the system sheds some light onto the emergence of the proliferative, cancer-like behaviour. Although the behaviour is similar to that of Gene configuration B, the emergence of a genotype that produces the cancer-like behaviour happens at an earlier stage and is faster than that observed in Configuration B. As in Configuration B, the genes that regulate apoptosis are consistently lost. This time, however, the

149

median shows a tendency to continue losing this gene at the latter stages Figure 5.17; with no dampening effect as in the case of Configuration B (Figure 5.11). This is due to two important facts. Firstly, because the Division regulatory genes start, on average, gaining copies earlier in time (Figure 5.18) than in Configuration B (Figure 5.12), which translates to increased levels of cell division early and thus higher chance of chromosome missegregation. Secondly, because the Apoptosis regulatory genes and the Segregation regulatory genes are linked, the more resistant a genotype becomes to cell death, the more genetically unstable it becomes (Figure 5.19). This genetic instability leads to an increase in diversity, as seen in Figure 5.20. Thus, this genetic arrangement makes it more likely for advantageous asymmetrical missegregations to take place either for the acquisition of oncogenes as well as the loss of tumour suppressors.
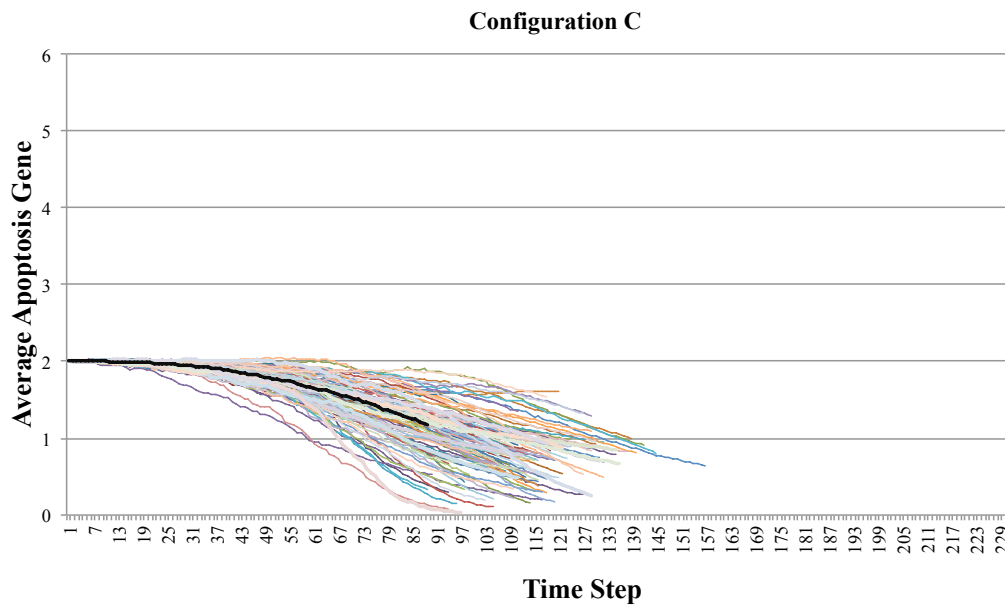
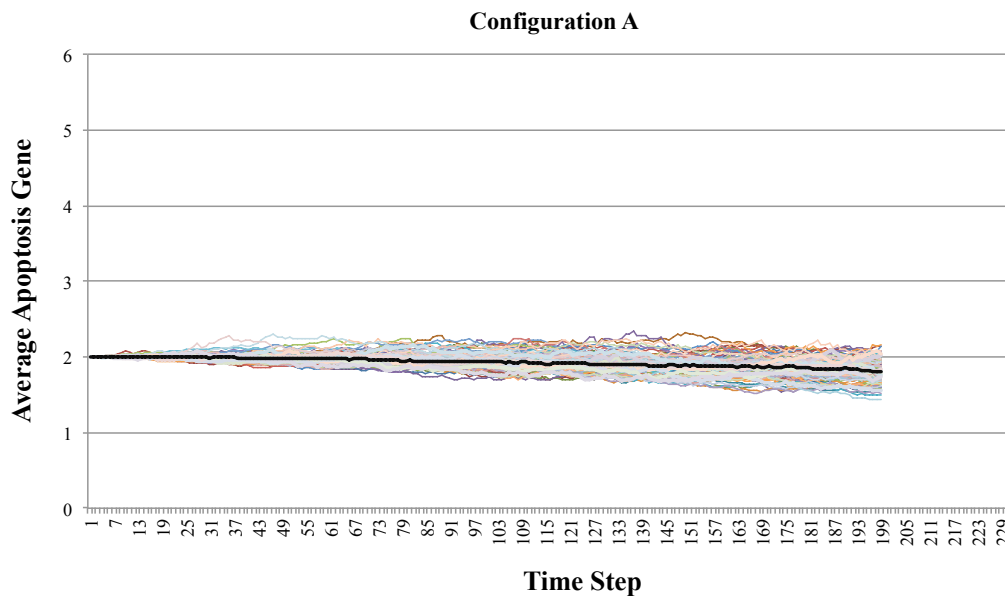**Figure 5.17- Measurement of the average number of Apoptosis Genes per time step represented in Broom Diagrams for 100 simulations of Configuration C. Each individual simulation is represented as line of different colour, with the median as a thick, black line (calculated until one of the simulations came to an end). Colours are purely used to distinguish runs and do not denote genetic distribution. The trend is to lose copies of this gene.**
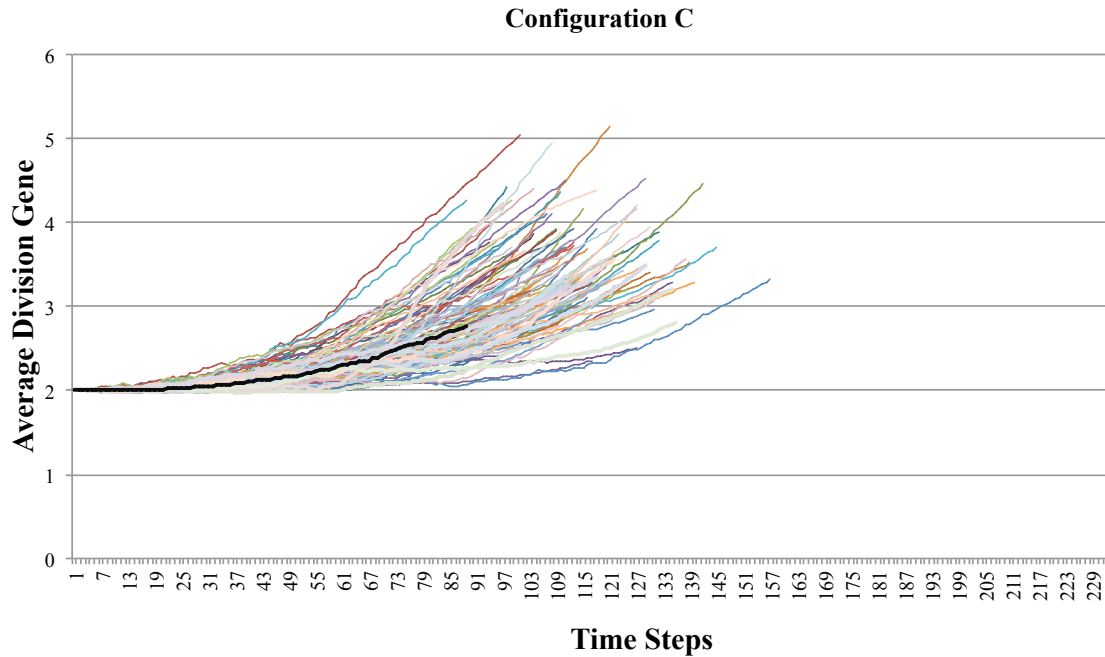
Configuration C

Figure 5.18- Measurement of the average number of Division Genes per time step represented in Broom Diagrams for 100 simulations of Configuration C. Each individual simulation is represented as line of different colour, with the median as a thick, black line (calculated until one of the simulations came to an end). Colours are purely used to distinguish runs and do not denote genetic distribution. The trend is to gain copies of this gene, which leads to over proliferation. This acquisition is on average considerably faster than in Configuration B (Figure 5.12)

The pathway discovered through this Gene Configuration may help shed some light on the reports of increasing levels of chromosome instability during premalignant neoplastic progression (Lai et al., 2007) and the development of tumours characterized by multiple and nonspecific aberrations, similar to most epithelial tumour types (Johansson et al., 1996).
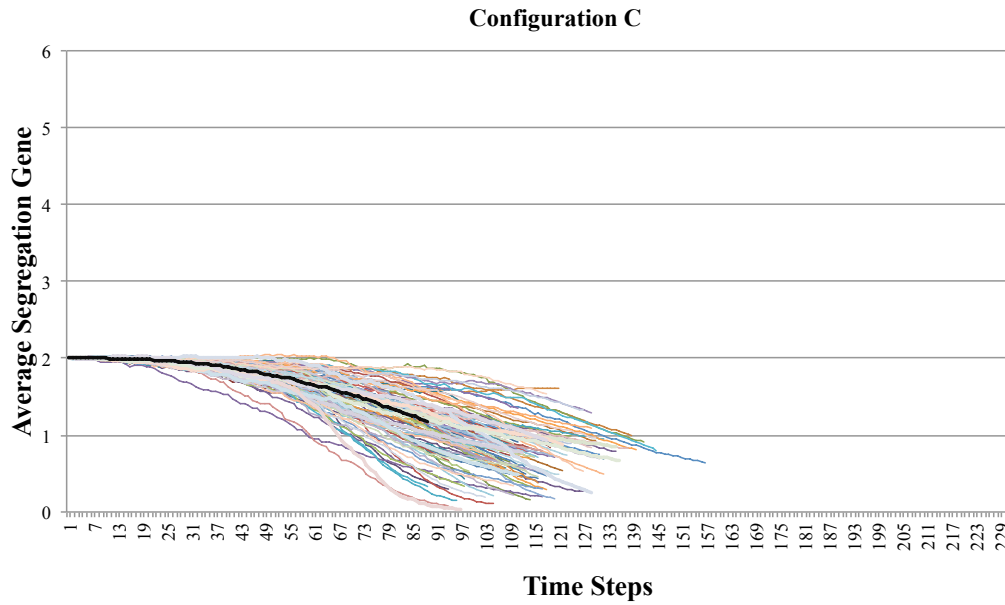
Configuration C

Figure 5.19- Measurement of the average number of Segregation regulatory genes per time step represented in Broom Diagrams for 100 simulations of Configuration C. Each individual simulation is represented as line of different colour, with the median as a thick, black line (calculated until one of the simulations came to an end). Colours are purely used to distinguish runs and do not denote genetic distribution. The trend is to lose copies of this gene, which leads to chromosomal instability.
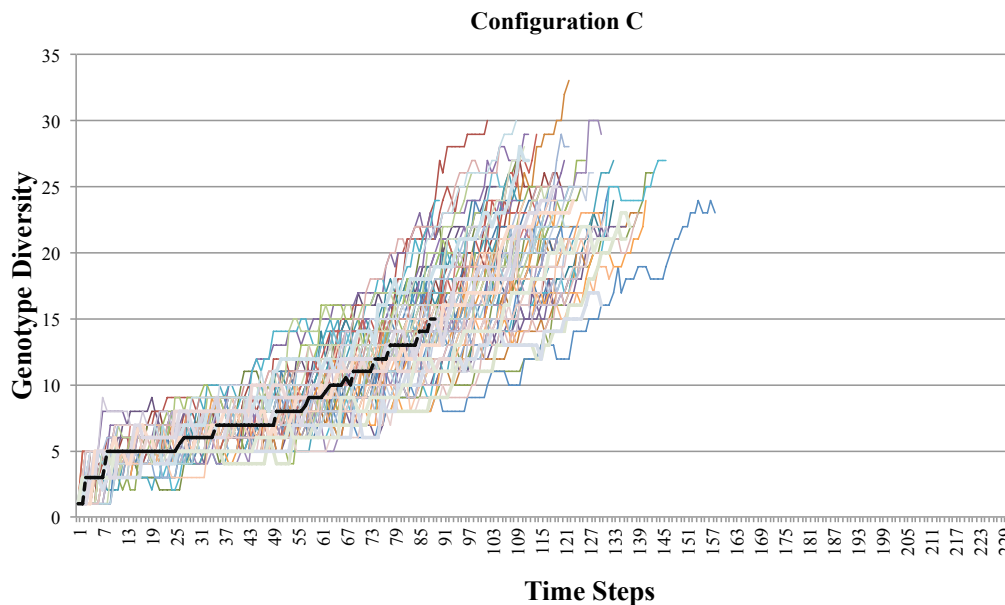
Configuration C

Figure 5.20- Measurement of the Genotype Diversity per time step represented in Broom Diagrams for 100 simulations of Configuration C. Each individual simulation is represented as line of different colour, with the median as a thick, black line (calculated until one of the simulations came to an end).

Colours are purely used to distinguish runs. The number of different genotypes increases dramatically towards the end.

5.3.4. Control Experiment: Gene Distribution D

a) *Objective and Setup*

To generate a control experiment, it was necessary that each gene was unlinked. This brought a complication, since an extra chromosome would be needed. It was decided that, although the precise numbers may not match the previous experiments due to the genetic rescaling, it was important to confirm the results previously obtained. The genetic arrangement, as seen in Figure 5.21 is as follows:

- Chromosome 1: cell-division regulatory genes
- Chromosome 2: apoptosis regulatory genes
- Chromosome 3: chromosome segregation regulatory genes
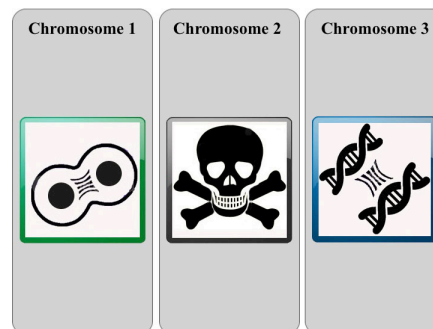
Figure 5.21- Setup for the unlinked distribution of genes into diploid chromosomes.

b) *Results*

Results indicate that all three previously obtained behaviours can be obtained in this unlinked configuration. However, given enough time, overproliferation will ensue as seen in Figure 5.22. This overproliferative behaviour takes, on average, longer to occur than the previous two overproliferative configurations: Gene distribution B and Gene distribution C.
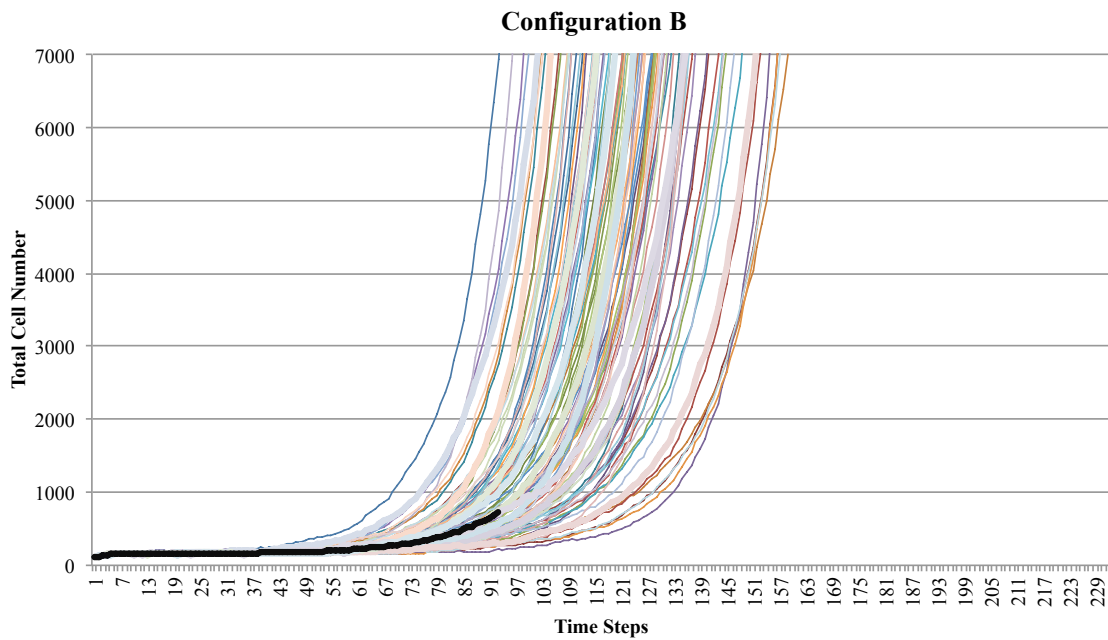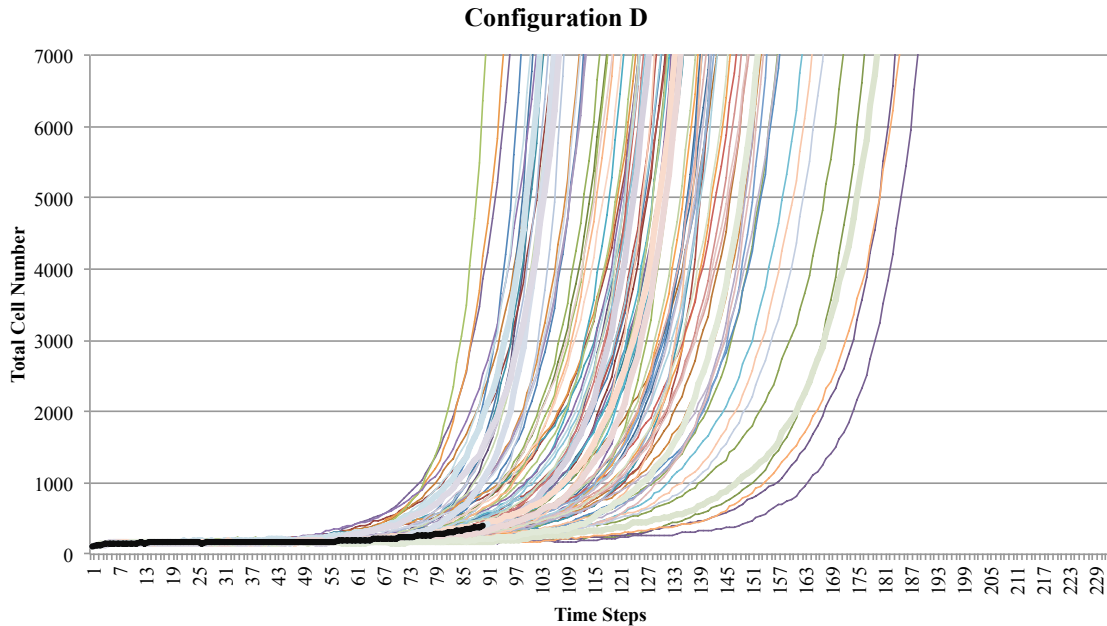
Configuration D

Figure 5.22- **The graph shows the total number of cells for 100 simulations of Configuration D. Each simulation is represented as line of different colour, with the median as a thick, black line (calculated until one of the simulations came to an end). Simulations were stopped when they reached 7000 cells. Cancer-like behaviour emerged through the evolution of the system sometimes until very late in the simulation.**

c) Analysis

The linking of genes in chromosomes imposes restrictions on the behaviour that can be observed. Without such linkage, stochasticity takes over and organizing principles cannot be easily extracted. As is the case with the previous experiments, there is a tendency to lose the gene that regulates death (Figure 5.23), and a trend to acquire genes that regulate division (Figure 5.24). However, the effects of stochasticity are most apparent on the large deviation across experiments: some simulations quickly acquire overproliferative behaviours while others remain homeostatic for considerably longer; a reflection on having no linkage between the genes.

154

A most interesting case is the plot of the evolution of the Chromosome segregation regulatory genes on Figure 5.25. The plot reveals that there is not a clear trend on either the loss or the acquisition or this gene. This measurement suggests that when linked with other genes, this mutation helps shape the pathway to the tumour initiation and subsequent evolution. On its own, however, it has a diminished role in the fate of the system.

The number of genotypes that evolve through this configuration is, on average, higher than in the previous configurations Figure 5.26. Although the effects that the acquisition and loss of segregation genes play a minor role in this, it is mostly because of the constraints from the gene linkage have been removed. Although these results are not directly comparable numerically to the previous experiments, they have helped us validate the results from the previous sections, and expanded our understanding of the key genetic transitions that occur throughout the evolution of the system.
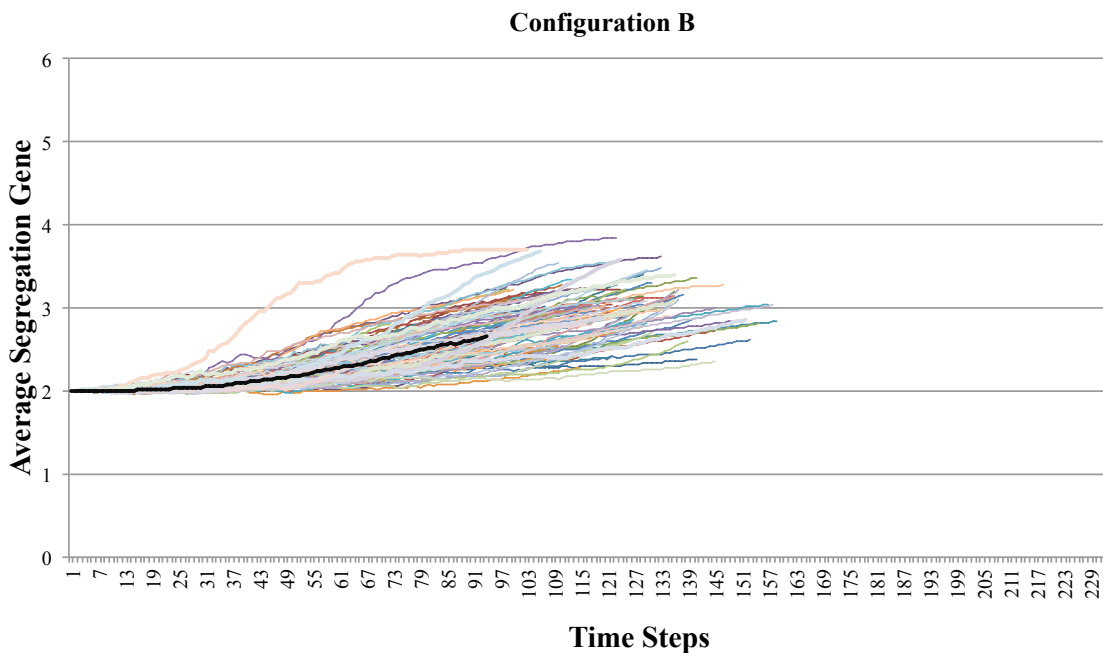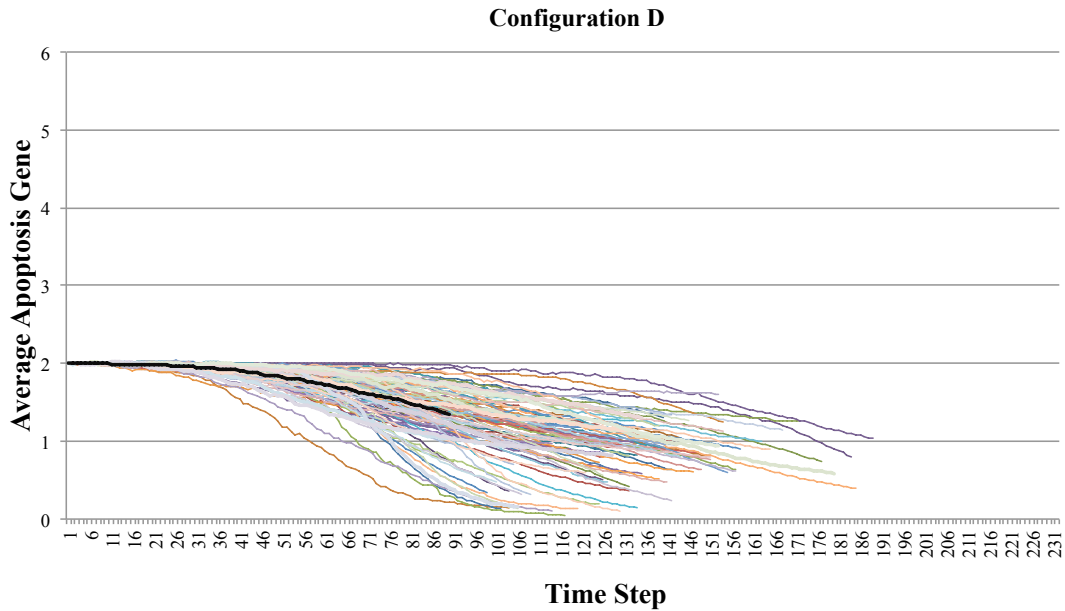
Configuration D

Figure 5.23- Measurement of the average number of Apoptosis regulatory genes per time step represented in Broom Diagrams for 100 simulations of Configuration D. Each individual simulation is represented as line of different colour, with the median as a thick, black line (calculated until one of the simulations came to an end). Unlinked from other genes, the tendency was to lose this gene.
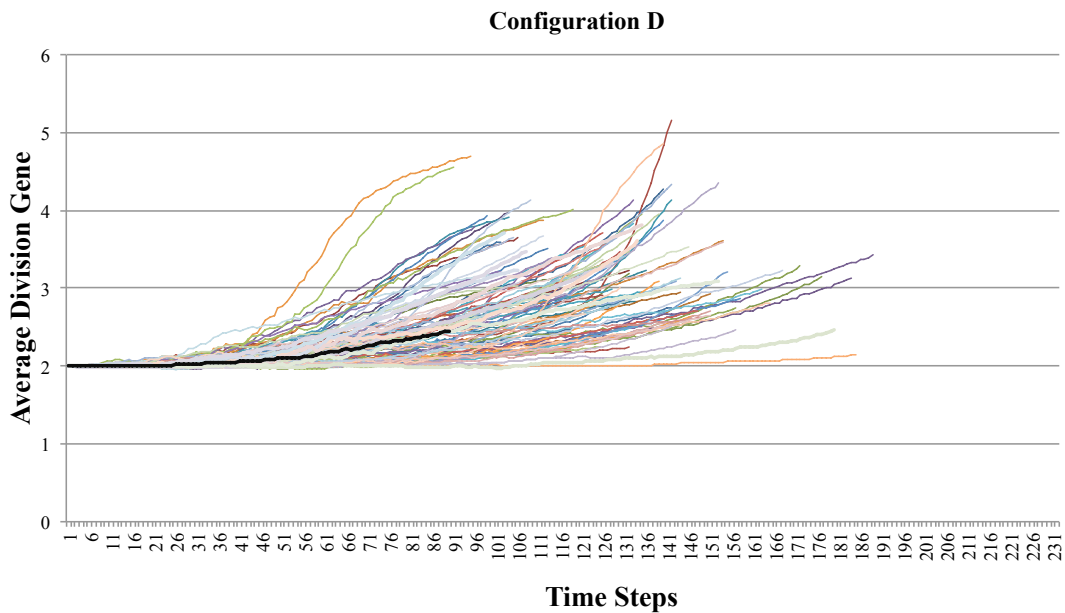
Configuration D

Figure 5.24- Measurement of the average number of Division regulatory genes per time step represented in Broom Diagrams for 100 simulations of Configuration D. Each individual simulation is represented as line of different colour, with the median as a thick, black line (calculated until one of the simulations came to an end). Unlinked from other genes, the tendency was to gain copies this gene.
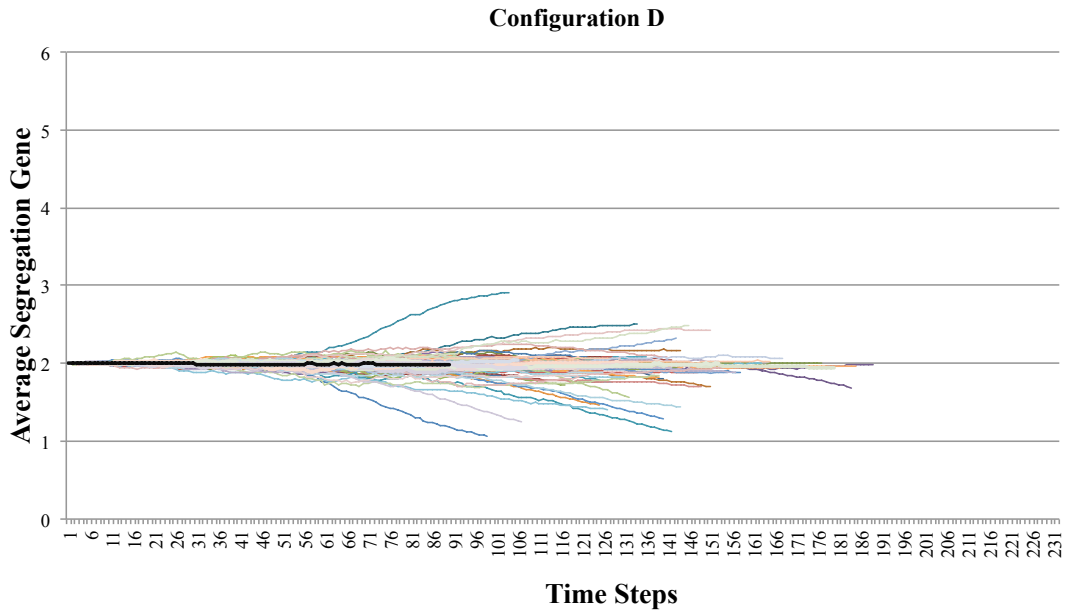
Configuration D

Figure 5.25- Measurement of the average number of Segregation regulatory genes per time step represented in Broom Diagrams for 100 simulations of Configuration D. Each individual simulation is represented as line of different colour, with the median as a thick, black line (calculated until one of the simulations came to an end). Unlinked from other genes, there is no general tendency on losing or acquiring this gene.
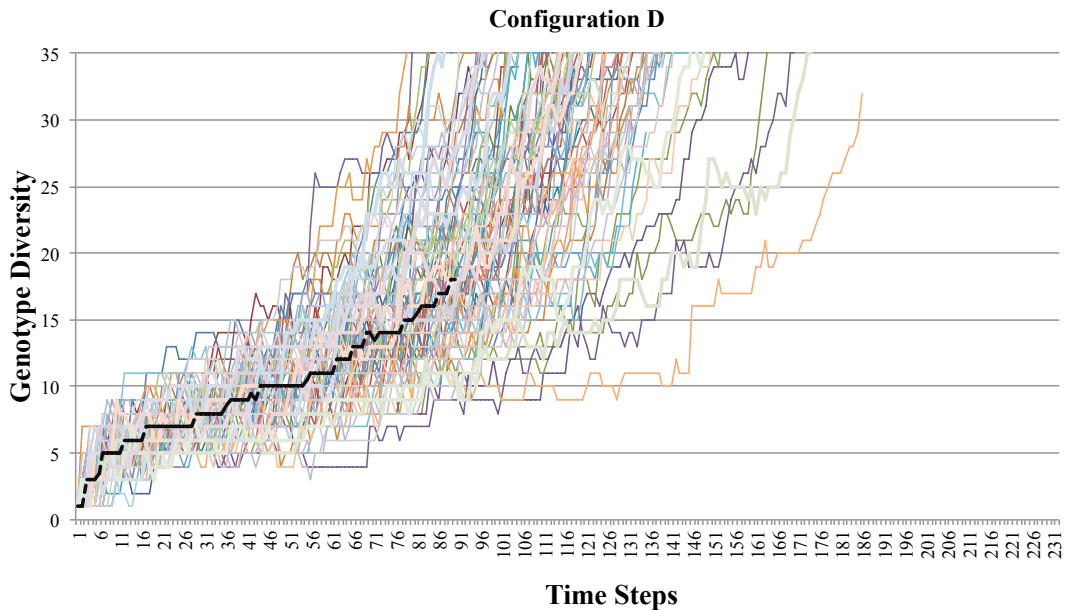
Configuration D

Figure 5.26- Measurement of the Genotype Diversity per time step represented in Broom Diagrams for 100 simulations of Configuration D. Each individual simulation is represented as line of different colour, with the median as a thick, black line (calculated until one of the simulations came to an end). Colours are purely used to distinguish runs.

157

5.4. Significance and Discussion

Simulations of Gene Distribution A resulted in homeostatic behaviour, in which the system as a whole responds to fluctuations in cell number to maintain the total number of cells close to that of the carrying capacity of the tissue (200 cells). As expected, the plot of the total number of cells across the simulations of Distribution A, as seen in Figure 5.4, revealed increasing variability in the genetic make-up of individual cells over time as the result of chromosome missegregation induced genetic drift; similar to that which might be seen in an ageing homeostatic tissue. Although this variation makes the statistical analysis challenging, an invariant behaviour can be observed for each configuration; best visualized by broom plots. In this case, because the abstracted genes that model the role of oncogenes and tumour suppressor genes were coupled by being situated on the same chromosome, the balance between death (Figure 5.5) and division (Figure 5.6) was maintained despite the generation of new genotypes (Figure 5.8) emerged through chromosome missegregation events (Figure 5.7). Significantly, some of the more successful genotypes naturally acquired more resistance against chromosome missegregation, through the acquisition of an extra copy of the chromosome segregation regulatory gene (such as genotype states (2,2,3), (3,3,3) and (1,1,3)).

For Gene Distribution B, the gradual accumulation of chromosome missegregation events leads to a breakdown in homeostatic behaviour, giving rise to uncontrolled proliferation, as seen in Figure 5.10. Once this occurred, total cell number increased exponentially, reaching the values of the order of thousands in a very short period of time. This kind of overproliferative behaviour was consistent across simulations. Emergent aneuploid genotypes evolved through Gene Distribution B, would create mutants either increasing proliferation (and stability), such as (3,2,3); or reducing the probability of Apoptosis, as in genotype (2,1,2). From these aneuploid genotypes, initially only slightly different to the original one, the population branches out to generate more malignant

158

genetically distinct variants such as (3,1,3) and (2,0,2). Different kinds of successful (and less successful) genotypes are gradually evolved. Successful genotypes have the qualities of being apoptosis-resistant (low number of apoptosis genes) and overproliferative (increased number of division genes). In this distribution, however, because the genes that regulate division are coupled to those that regulate fidelity during segregation, there is a brake applied to the subsequent generation of aneuploid genotypes with increased division rates. As a result, this population of aneuploid cells remained relatively homogeneous once cells had acquired the key genetic anomalies driving deregulated tumour growth. This kind of evolution observed across experiments suggests a possible pathway for oncogenesis that is associated with stable aneuploidy. Diseases such as leukaemia, lymphomas and some mesenchymal tumours that exhibit specific abnormalities may follow a similar path (Johansson et al., 1996).
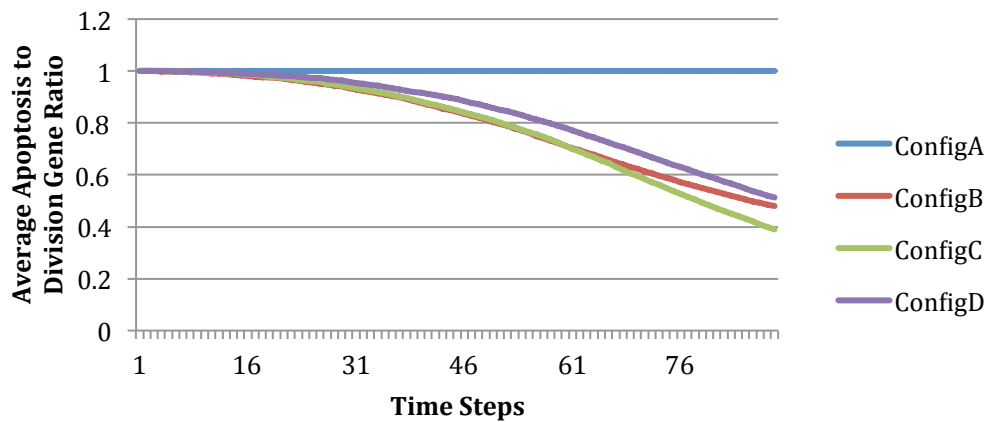
Figure 5.27- A Plot of the average of the ratios of Apoptosis Genes to Division Genes for the four configurations. Configuration A keeps the same ratio of 1 throughout the simulations. Configurations B, C and D tend to lower this ratio with different slopes, characteristic of their own characteristic internal dynamics.

Simulations of Gene Distribution C displayed overproliferative behaviour, similar to that of Gene Distribution B (Figure 5.16). On a closer inspection, however, significant differences in the dynamics of cancer evolution were observed. Because the genes that regulate death are

genetically linked to those that regulate segregation in Gene Distribution C, cancer evolution was accompanied by an increase genotypic diversity as the loss of apoptosis regulators drive an increase in aneuploidy, such as genotype (3,1,1) and then a more malignant genotype (3,0,0). This in turn drives to the emergence of ever more aggressive clones such as (4,0,0), (5,0,0) and (6,0,0), which corresponds to a 3-fold increase in the rate of cell proliferation. This serves as a model for the emergence of heterogeneous tumours, like those seen in clinical settings, for example during the neoplastic progression characteristic of epithelial tumours (Lai et al., 2007) (Johansson et al., 1996).

These simulations for Distribution B and C show how chromosome missegregation events can drive tumour evolution by inducing a change in the balance of regulators that maintain normal tissue homeostasis. Our control experiment Distribution D shows that gene linkage plays a key role in the kind of aneuploidy that is obtained through chromosome missegregation. As can be seen on the broom plot for Chromosome Segregation regulatory genes (Figure 5.25), when unlinked, the gene that regulates chromosome missegregation follows a stochastic drift, while the division and apoptosis regulatory genes continue their previous trends (Figure 5.24 and Figure 5.23, respectively). Taken together these experiments recapitulate some of the findings of Komarova, in which evolutionary pathways differ whether chromosomal instability is an early event in oncogenesis or not (Komarova et al, 2003).

Finally, a plot of the average of the ratio between Apoptosis regulatory genes and Division regulatory genes across the different configurations (Figure 5.27) reveals that, although the trends of behaviour are similar across Configurations B, C each one has a different slope ("an average slope of -0.0080 for Configuration B and -0.0116 for Configuration C) which reflects a configuration-specific mechanism that is heavily influenced by the Chromosome segregation

regulatory genes. This graph shows Configuration C as having the steepest slope, which correlates with a more aggressive behaviour. On average, simulations of Configuration C reached the simulation limit of 7000 cells 10 time steps earlier than those of Configuration B (last average time step in simulations for Configuration B is 127.78 and 117.87 for Configuration C).

5.5. Summary

In this chapter a computational model was created in order to investigate the role of chromosome missegregation in tumour evolution. By integrating the concept of chromosome missegregation in an otherwise homeostatic model, new genotypes were evolved. Those evolved genotypes that enabled cells to express high levels of cell division genes and low levels of cell death genes quickly spread through the population. Two different pathways to overproliferative behaviour were found: one chromosomally stable (Configuration B) and one chromosomally unstable (Configuration C). The questions proposed in Section 5.2 can now be answered.

Are there any kinds of genetic arrangements through which chromosome missegregation may lead to the disruption of homeostasis?

From experiments carried out, it was only Configuration A, which remained homeostatic through its entire evolution. An analysis of the genetic evolution revealed that the linkage between genes that regulate division and segregation was responsible for maintaining the balance in the event of chromosome missegregation. If these kinds of genes are uncoupled, the balance may be broken which may lead to the disruption of homeostasis.

If there are any overproliferative behaviours, could they be classified according to their evolution given a set of genetic constraints?

Two main pathways were found to occur, depending on the type of insult to the genetic integrity.

Are there any key genetic transitions that could result in the disturbance of homeostasis?

If the first hit is the loss of tumour suppression (Gene Configuration B), cells endure for longer despite stress conditions. This resistance to apoptosis makes it more likely for cells to acquire an oncogene with time and then combine the death resilience with an increase in cell proliferation, leading finally to tumourigenesis. If the first hit is the activation of an oncogene (like in Gene Configuration C), increased proliferation is counterbalanced at first by a high death rate until the activation of tumour suppression occurs. These results are consistent with the observed activation of oncogenes and loss of tumour suppressor genes in tumours (Michor, Iwasa, & Nowak, 2004).

Does the distribution of genes in chromosomes have a role on the disturbance of homeostasis and over proliferation through the mechanism of chromosome missegregation?

Although the model makes a number of assumptions, including the assumption that the number of copies of a gene has a direct effect on the up or down regulation of that gene, the interactions and results can be interpreted in terms of actual biological behaviour (i.e.; the up or down regulation of an oncogene or a tumour suppressor gene). The model suggests that through chromosome missegregation, the arrangement of genes on chromosomes has a profound effect on genetic diversity, giving rise to different kinds of cancer-like behaviours, which resemble key differences observed in real cancers (Cahill et al., 1999). These results are dependent on the modelling decision regarding contact inhibition affecting apoptosis.

The role that chromosome segregation regulatory genes play in this model is largely determined by their position with respect to the other genes in the chromosomes. The model suggests that high levels of chromosome missegregation lead to a genetic diversity that helps cells overcome

the low probability of oncogenic mutations, as shown in the analysis of Gene Configuration C. Surprisingly, low levels of chromosome missegregation may also give rise to a different kind of cancer-like behaviour, as shown in the simulations of Gene Configurations B. By maintaining a relatively uniform population, specific mutations are conserved and spread throughout the population until a cancer-like genotype is reached. To determine the precise role that chromosome segregation regulatory genes have in cancer systems, the development of appropriate tools for statistical analysis and further experiments are needed.

6. Cancer Therapies and Chromosome Missegregation

6.1. Introduction

From the experiments on the model in the previous chapter, it was found that cancer-like growth will ensue if the number of oncogenes increases and/or if tumour suppressors are lost. However, it was also revealed that there is a high level of variability across experiments. This variability is the kind of biological variability that is found in real cancer systems. Like the simulations, no two cancers are the same. This has made the process of extracting general organizing principles from *in vivo* and *in vitro* experiments very difficult. It is anticipated that the model developed in this work can help address this issue. Because the computer uses a pseudo random number generator, which uses a seed number, individual simulations are reproducible. This gives us the unique situation in which it is possible to explore the outcome of different treatments on the exact same individual simulation. With the tools developed, we set out to simulate the two most common cancer therapies: surgery and chemotherapy, and investigate the effects that chromosome missegregation has on them.

No two tumours are alike, as our computational model has highlighted in the previous chapter. Because of this, one of the main goals of cancer research is to be able to offer individualized diagnoses and treatments. In the clinic, the most common cancer treatments are surgery and chemotherapy. Surgery, the oldest form of cancer treatment, is often used to remove localized tumours. Chemotherapy uses drugs to control rapidly dividing cancer cells. Physicians often use a combination of these treatments to obtain the best results.

Usually, when a patient is diagnosed with cancer, treatment methods are determined based on the kind of cancer, its location and its stage. However, the genetic stability of the tumour is

something that is not usually considered. Although chromosome instability has been associated with poor prognosis, the precise role that chromosome missegregation has on the mechanisms of cancer a therapy is poorly, if at all, understood. Biological experiments may address some of the issues surrounding this phenomenon. Our model can help develop insights regarding the role of chromosome missegregation under diverse therapeutic scenarios. These insights may be used to help devise biological experiments, and bridge the gap between the laboratory and the clinic.

6.2. Questions in Cancer Therapies

It is currently unknown how chromosome missegregation affects cancer therapies. However, it has been recently proposed that chromosomally unstable cancers, such as the ones modelled by Configuration C, have a poorer prognosis than stable cancers, such as the ones addressed by Configuration B (Swanton, Nicke, Marani, Kelly, & Downward, 2007). With the computational tool developed in the previous chapter, it is possible to tackle some of the questions that cannot be addressed *in vitro* or *in vivo*. In this work we address these specific questions:

- Is there a genetic configuration that has a better general prognosis?
- Are there key genetic events in the evolution of tumours that make them more or less susceptible to cancer treatments?
- Are there any predictive markers for a successful therapy?(Walther et al., 2009)

To address such questions, cancer therapies must be abstracted and then incorporated in the context of the model developed.

6.3. Simulating Cancer Therapies

In patients, tumours composed of cells that are chromosomally unstable have been associated with a poor prognosis (Swanton et al., 2009). Therefore, it was decided to used Gene Distribution

165

B (Figure 6.1.A) and Gene Distribution C (Figure 6.1.B) to determine the relative efficacy of different treatment strategies in dealing with tumour evolution under conditions of low and high levels of genome instability. Because configuration D served as a null experiment, it will not be used in the experiments for this section.
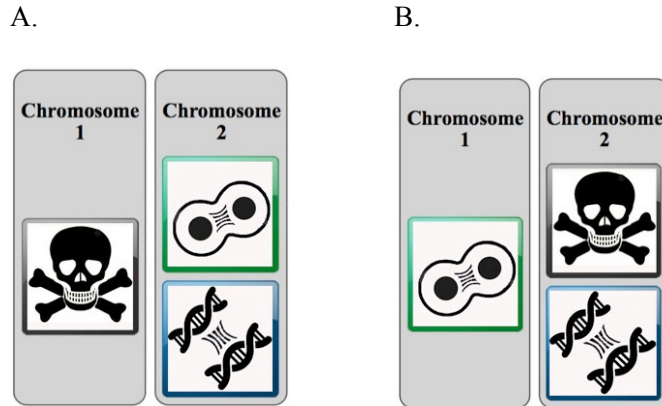
Figure 6.1- Genetic configurations used to evaluate the role of chromosome missegregation in cancer therapies.

Although in real cancer, treatment occurs late in the development of the tumour, the model developed can still address this relevant question because we have considered an initial high rate of chromosome missegregation, such as the one observed in early tumours and some cancer cells. Because of the previously discussed computational limitations (Chapter 4), we considered that tumour detection would occur when the population reached 1000 cells (10 times the initial size of the population). It will be at this step that treatment will be administered. This modelling decision was based on the numbers of cells that are considered in wet-lab experiments (see Chapter 7), for which this model could serve as a guide without any re-scaling. By the same standard, we considered that the tumour had relapsed when it again reached the 1000 cell mark after treatment. To make these comparable to the simulations obtained in the previous chapter, it will be considered that a simulation has reached the end (the death of the organism) at 7000 cells.

With these considerations, we modelled the outcome of different treatments on a single tumour (or patient, if the system is re-scaled), so that we could directly compare the outcomes in each case, despite the expected variability in the course of tumour growth between different simulations (tumours/patients).

6.4. Scenario i: Surgical Treatment

6.4.1. Objective and Setup

For these experiments, the same system as described in Section 5.3 was used. For Gene Configuration B (Section 5.3.2) and Configuration C (Section 5.3.3), the overproliferative genotypes, surgery was implemented at the time step when each simulation reached 1000 cells. The simulation of tumour removal was implemented by eliminating 90% of the cells, regardless of the cell type. The simulation is considered to have relapsed when, after the surgery, it grows back to 1000 cells and beyond. 100 simulations with each configuration, with the same seeds as those described in Section 5.4, were performed.

6.4.2. Results

Invariantly, the tumour recovers and grows back, as can be seen in Figure 6.2. Complete relapse takes place once the size of the population reaches again 1000 cells. The most distinctive difference between the two Gene Configurations, from Figure 6.2, is the average time of relapse. Configuration C (Figure 6.2.B) has, on average, a faster relapse than Configuration B (Figure 6.2.A). This difference is a reflection of the internal dynamics: the gain and loss of oncogenes and tumour suppressors, which is regulated by the genetic instability. This is best visualized by broom plots with data re-centered around the time of therapy. Unlike the broom plots shown in the previous chapter, in which all data sets are shown to start at time step 0, these plots have starting

points at different time steps. This starting offset was calculated so that all the graphs share the same start of therapy time point: the time step when each individual run initially reaches 1000 cells. With the time of start therapy as an alignment point, it is easier to understand the relationship between key genetic aspects of simulations and the therapies.

As seen in the broom plots, the consistent loss of the Apoptosis Genes (Figure 6.3) reveals the different ways surgery interacts with the dynamics of specific genetic arrangements. For Configuration B (Figure 6.3.A), surgery appears to stabilize, on average, the loss of the tumour suppressor. In contrast, Configuration C (Figure 6.3.B) seems to maintain the same rate of loss of Apoptosis Genes as before the treatment. The effect of surgery on Division regulation genes is more subtle. Configuration B (Figure 6.4.A) has a shift, on average, back towards the original diploid state. The slope of the rate of gain of this oncogene however remains the same. Configuration C (Figure 6.4.B) has a bigger shift towards diploidy: however, this gain is only temporary. The slope of the rate of gain changes after surgery in such a way that the acquisition of Division Genes is, on average, faster than before surgery.

Surgery has a noticeable effect on the Segregation Genes for the chromosomally stable Configuration B (Figure 6.5.A), than for the chromosomally unstable Configuration C (Figure 6.5.B). However, on average, the slope of the rate of acquisition (Configuration B) and loss (Configuration C) seem to be unaffected by this therapy. This is reflected on the emergence of new genotypes. Configuration B (Figure 6.6.A) continues to generate new genotypes at a linear progression after surgery, while Configuration C (Figure 6.6.B) generates increasing diversity. Removing more than 90% of the tumour may cause even more variable results. For example, removing 99.9% of the cells would result in leaving 1 cell that could be from the precancerous cell population, or one of the most aggressive clones.

168

A.
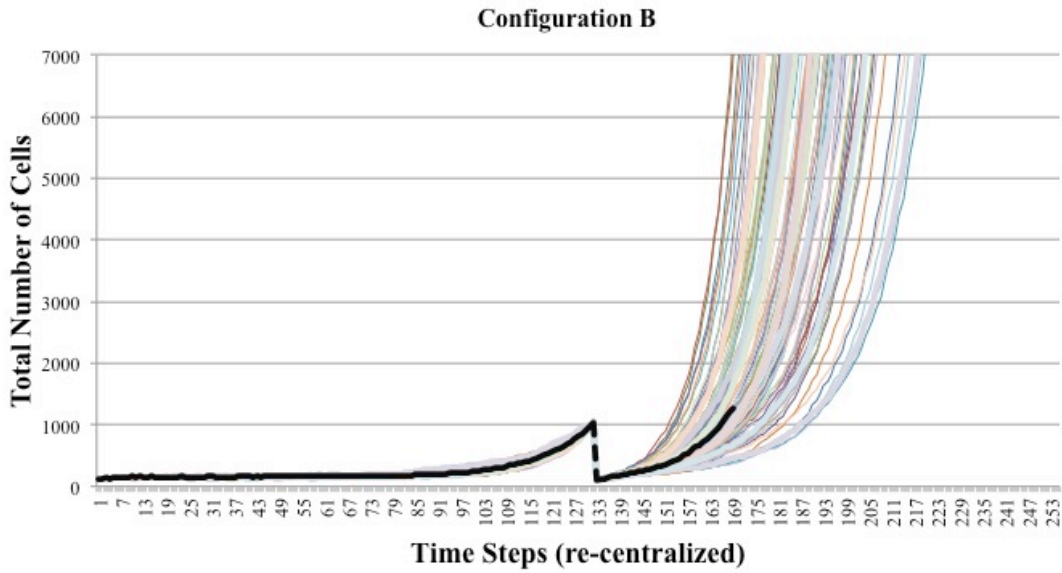
B.
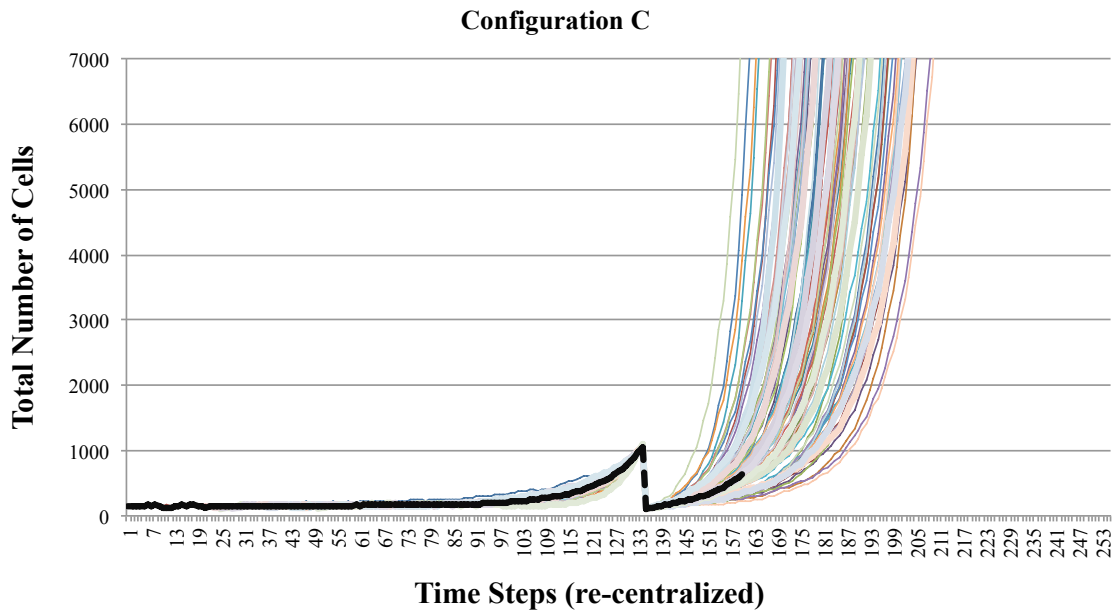
Figure 6.2 Total number of cells for simulated Surgery. A. 100 simulations for Configuration B. B. 100 simulations for Configuration C. Each simulation is represented as line of different colour, with the median as a thick, black line (calculated until one of the simulations came to an end). Simulations were re-aligned with respect to the time step in which treatment starts.

A.
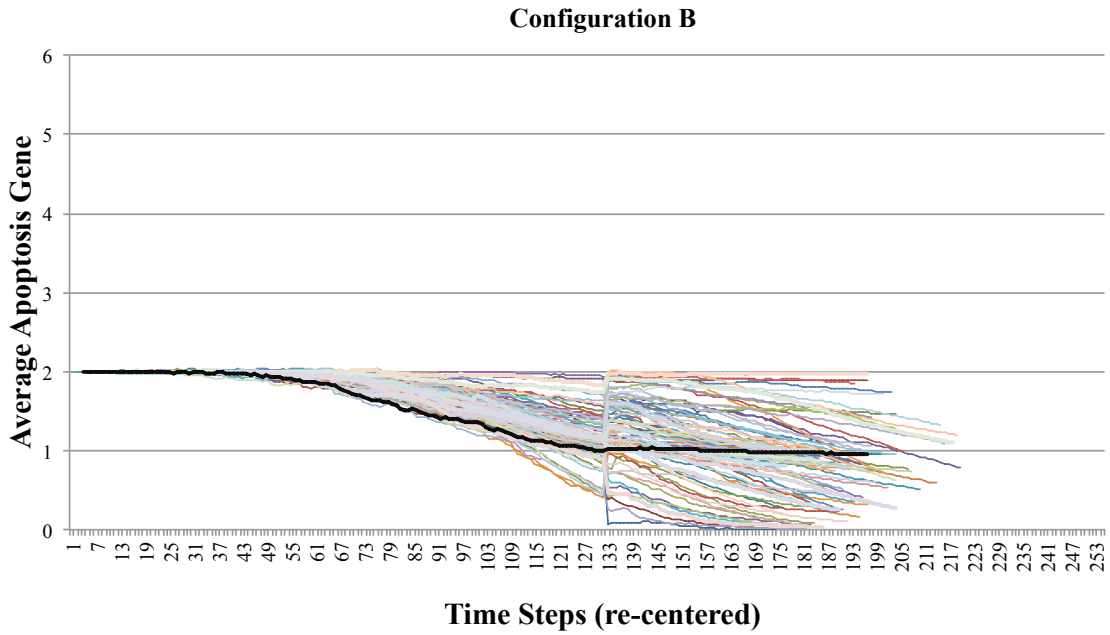
B.
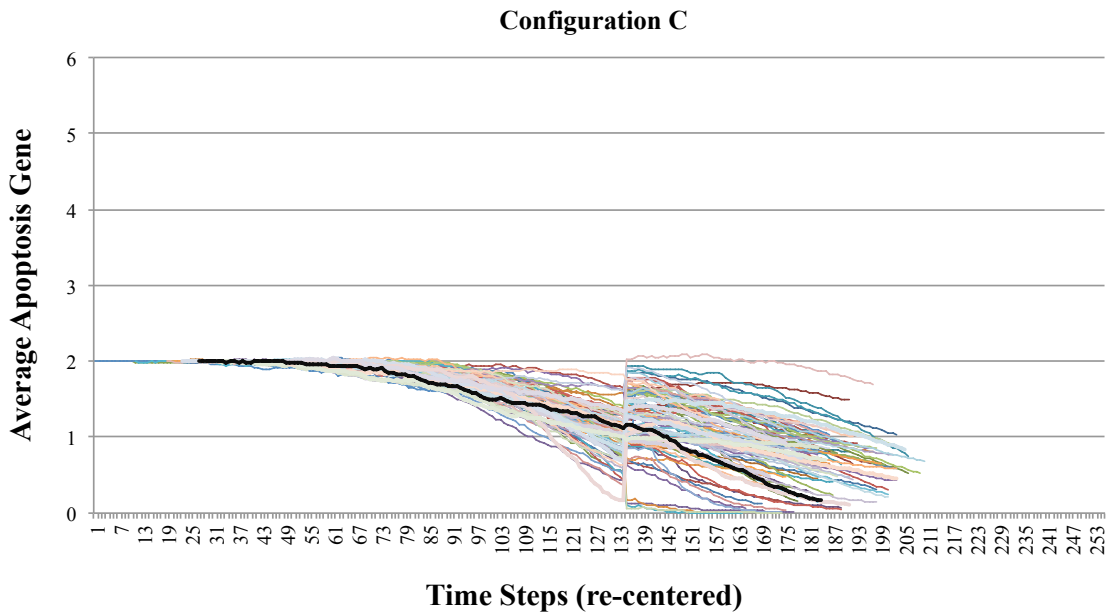
Figure 6.3- Measurement of the average number of Apoptosis Genes under Surgery scenario in Broom Diagrams A. 100 simulations of Configuration B. B. 100 simulations of Configuration C. Each individual simulation is represented as line of different colour, with the median as a thick, black line (calculated until one of the simulations came to an end). Simulations were re-aligned with respect to the time step in which treatment starts

170

A.

Configuration B

B.
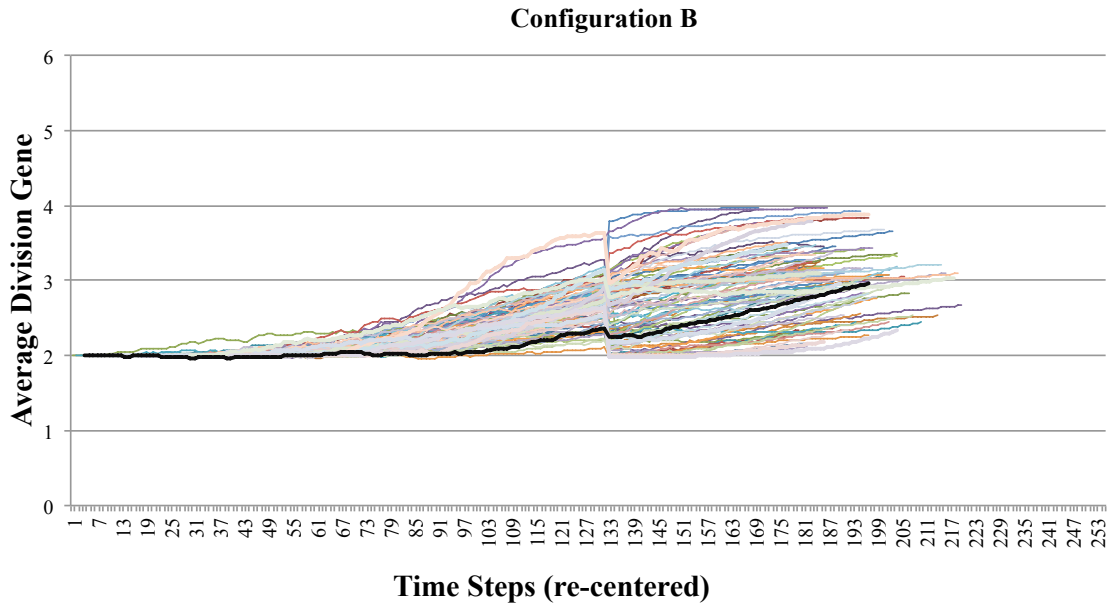
Configuration C

Figure 6.4- Measurement of the average number of Division Genes under Surgery scenario in Broom Diagrams A. 100 simulations of Configuration B. B. 100 simulations of Configuration C. Each individual simulation is represented as line of different colour, with the median as a thick, black line (calculated until one of the simulations came to an end). Simulations were re-aligned with respect to the time step in which treatment starts

171

A.

B.

Figure 6.5- Measurement of the average number of Segregation Genes under Surgery scenario in Broom Diagrams A. 100 simulations of Configuration B. B. 100 simulations of Configuration C. Each individual simulation is represented as line of different colour, with the median as a thick, black line (calculated until one of the simulations came to an end). Simulations were re-aligned with respect to the time step in which treatment starts.

A.

Configuration B

B.
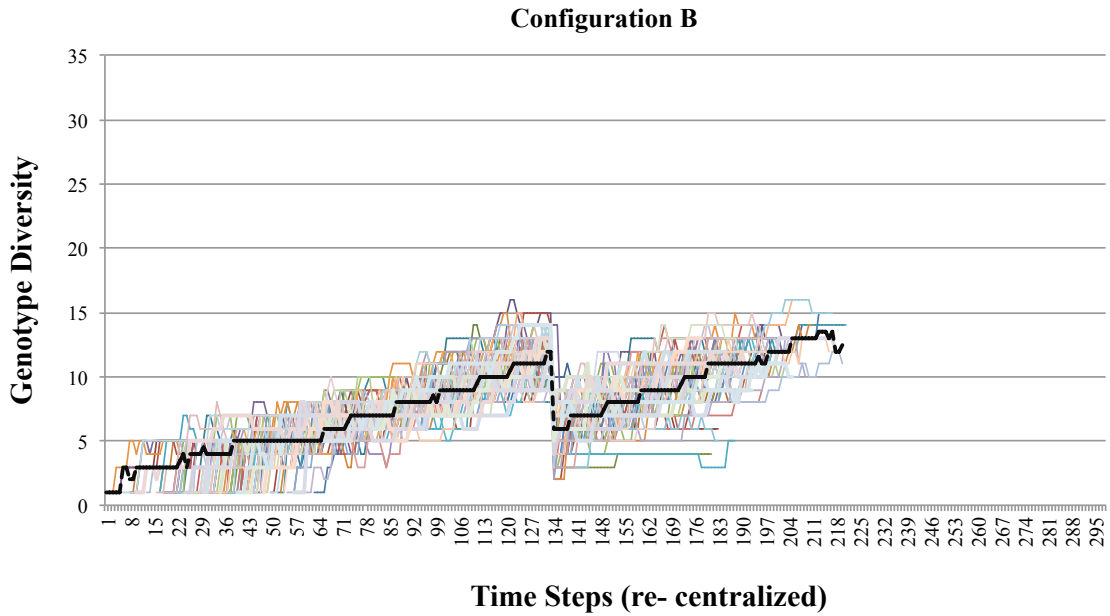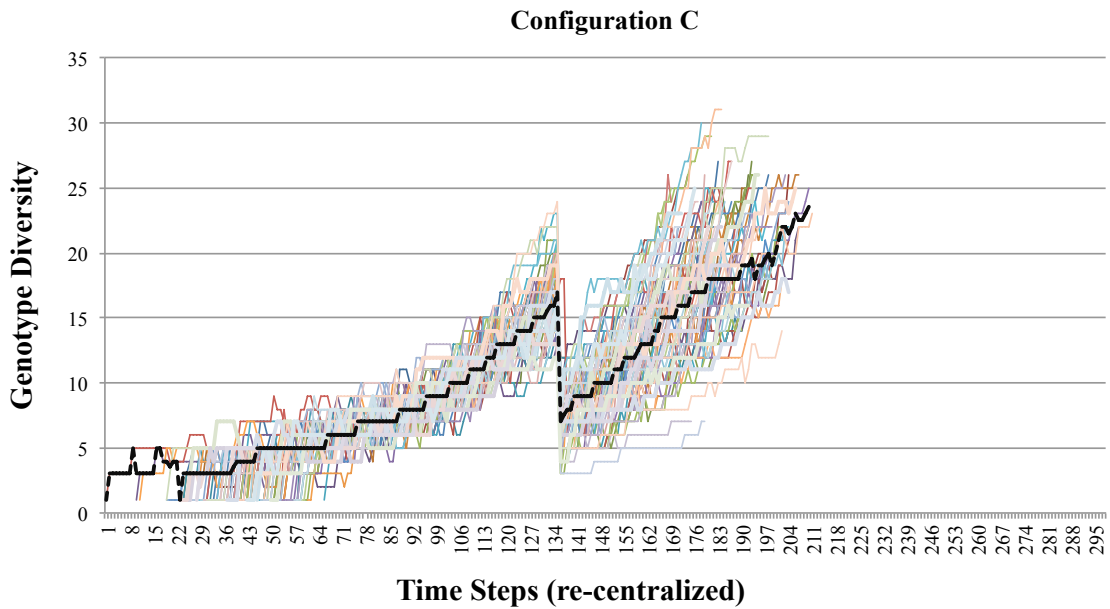
Configuration C

Figure 6.6- Measurement of the Genotype Diversity under Surgery scenario represented in Broom Diagrams A. 100 simulations of Configuration B. B. 100 simulations of Configuration C. Each individual simulation is represented as line of different colour, with the median as a thick, black line (calculated until one of the simulations came to an end). Colours are purely used to distinguish runs. Simulations were re-aligned with respect to the time step in which treatment starts.

173

6.5. Scenario ii: Chemotherapy

6.5.1. Objective and Setup

Chemotherapy, in real clinical settings, consists of drugs that selectively kill rapidly dividing cells (McClelland et al., 2009). For these simulations, the same system as described in Section 5.3 was used: Gene Configuration B (Section 5.3.2) and Configuration C (Section 5.3.3). This procedure was implemented as an algorithm that killed all the cells that attempted cell division in the nine consecutive time steps following tumour detection; when the simulation reaches 1000 cells. A simulation is considered to have relapsed when reaching again 1000 cells, and the death of the organism at 7000 cells. 100 simulations with each configuration, with the same seeds as those described in Section 5.4, were performed.

6.5.2. Results

After the therapy, simulations invariantly relapse. Like the simulation of Surgery (Scenario i), tumours seem to take a reduction of the order of 90% after the nine rounds of Chemotherapy. This reduction is short lived however, as the tumour recovers and continues to grow exponentially (Figure 6.7). As is the case for surgery, Gene Configuration B (Figure 6.7.A) has, on average, a longer relapse time than Configuration C (Figure 6.7.B).

In contrast to Scenario i (surgery), Chemotherapy has a more gradual impact on the evolution of new genotypes. Although the consistent loss of the Apoptosis Regulatory Gene for Configuration B (Figure 6.8.A) and Configuration C (Figure 6.8.B) appear to be only lightly dampened, the consistent gain of Division Genes has a noticeable offset for both Configuration B (Figure 6.9.A) and Configuration C (Figure 6.9.B). As in Scenario i, the rate of gain of Division Genes in Configuration B is maintained, whereas the rate of Configuration C increases after therapy.

174

An interesting situation occurs with respect of the Segregation Genes. Configuration B usually acquires Segregation Genes in its evolution. However, because of therapy, Configuration B (Figure 6.10.A) loses a considerable number of Segregation Genes, as it is driven back towards an average diploidy. Hence, these cells become on average more genetically unstable after therapy. In contrast, Configuration C (Figure 6.10.B) conserves the linear rate of loss of this kind of genes after therapy. The Genetic Diversity, the number of different genotypes at any time step, is on average higher for Configuration C (Figure 6.11.B), than for Configuration B (Figure 6.11.A). Also Configuration B on average responds better than Configuration C to Chemotherapy.

A.
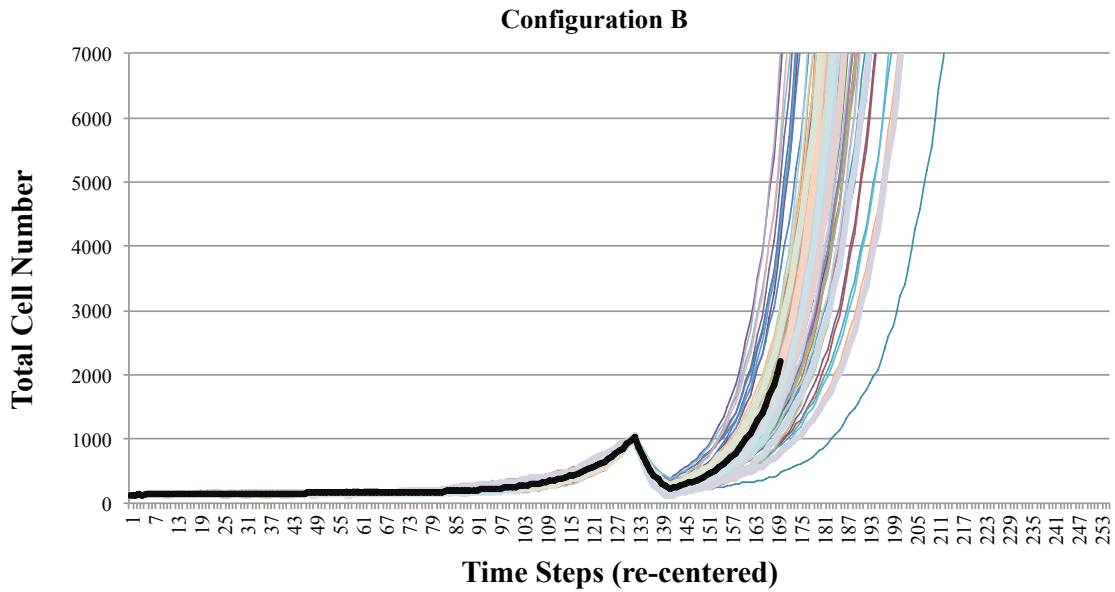
B.
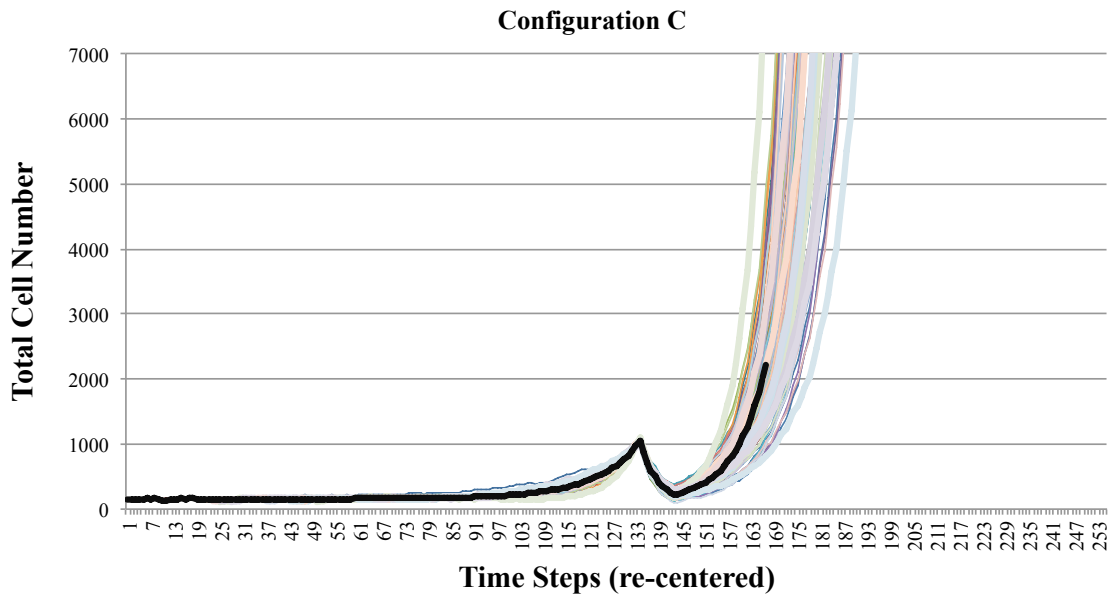
Figure 6.7- Total number of cells for simulated Chemotherapy. A. 100 simulations for Configuration B. B. 100 simulations for Configuration C. Each simulation is represented as line of different colour, with the median as a thick, black line (calculated until one of the simulations came to an end). Simulations were re-centered around the start of treatment.

A.

Configuration B

(y-axis) Average Apoptosis Gene

(x-axis) Time Steps (re-centered)

B.

Configuration C

(y-axis) Average Apoptosis Gene

(x-axis) Time Steps (re-centered)
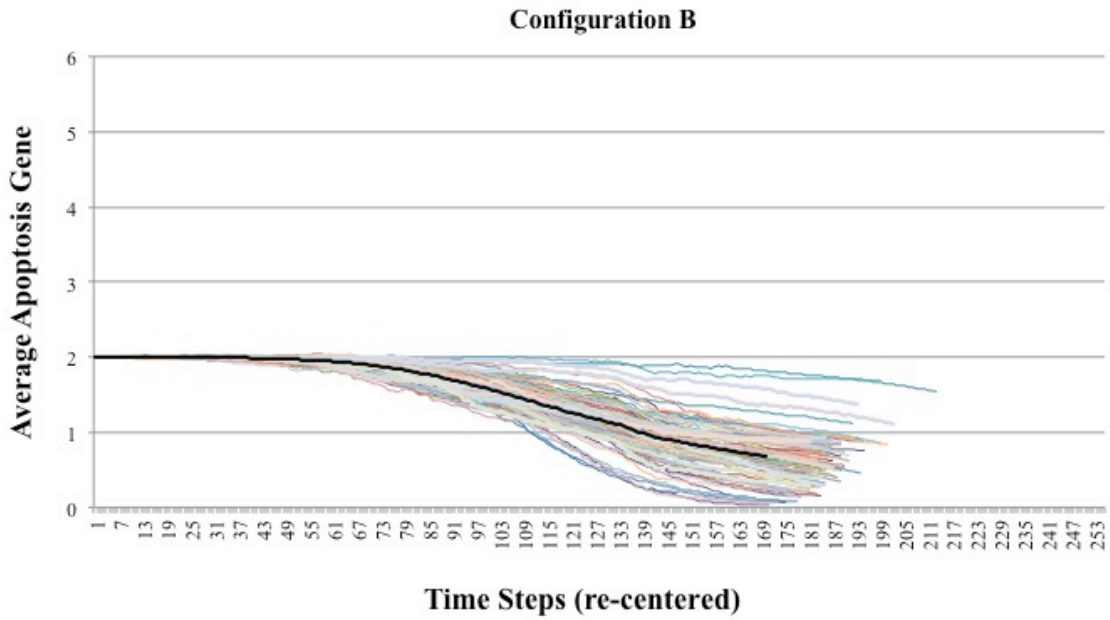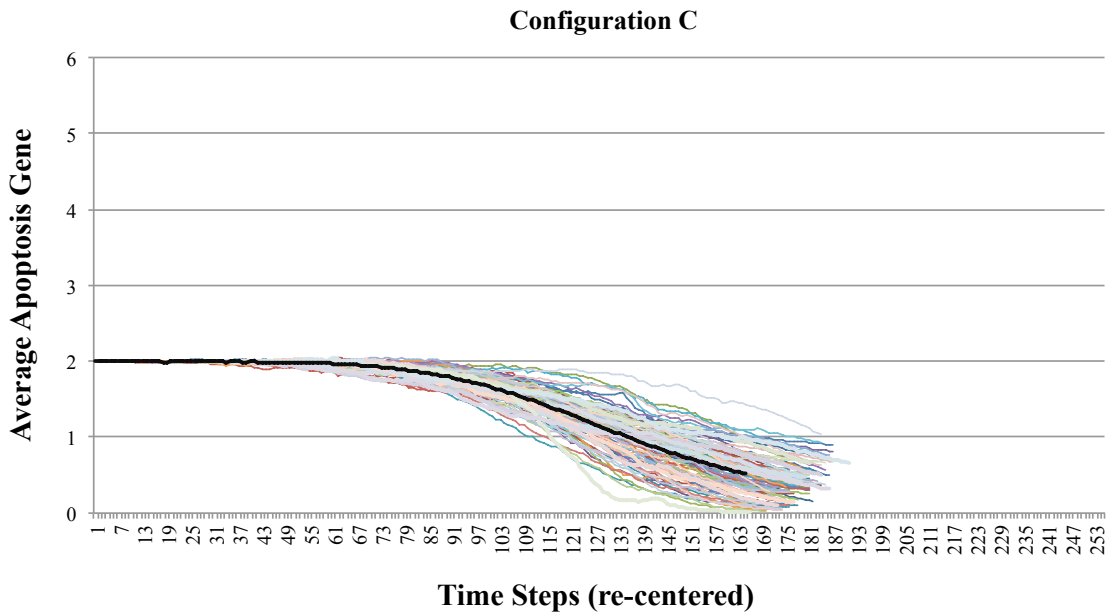
Figure 6.8- Measurement of the average number of Apoptosis Genes under Chemotherapy scenario in Broom Diagrams A. 100 simulations of Configuration B. B. 100 simulations of Configuration C. Each individual simulation is represented as line of different colour, with the median as a thick, black line (calculated until one of the simulations came to an end). Simulations were re-aligned with respect to the time step in which treatment starts.

177

A.

Configuration B

B.
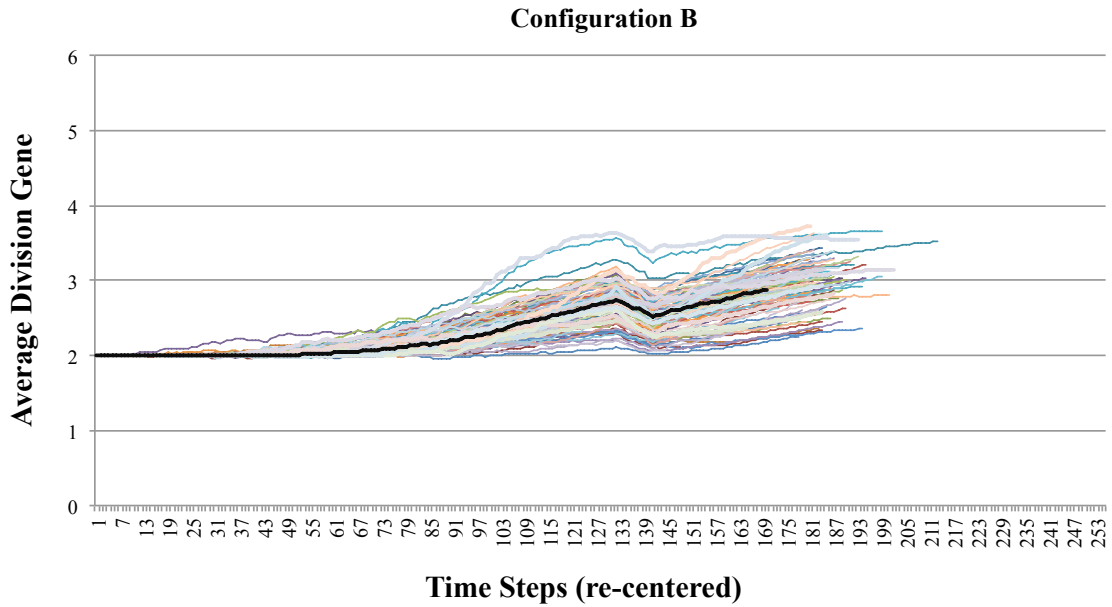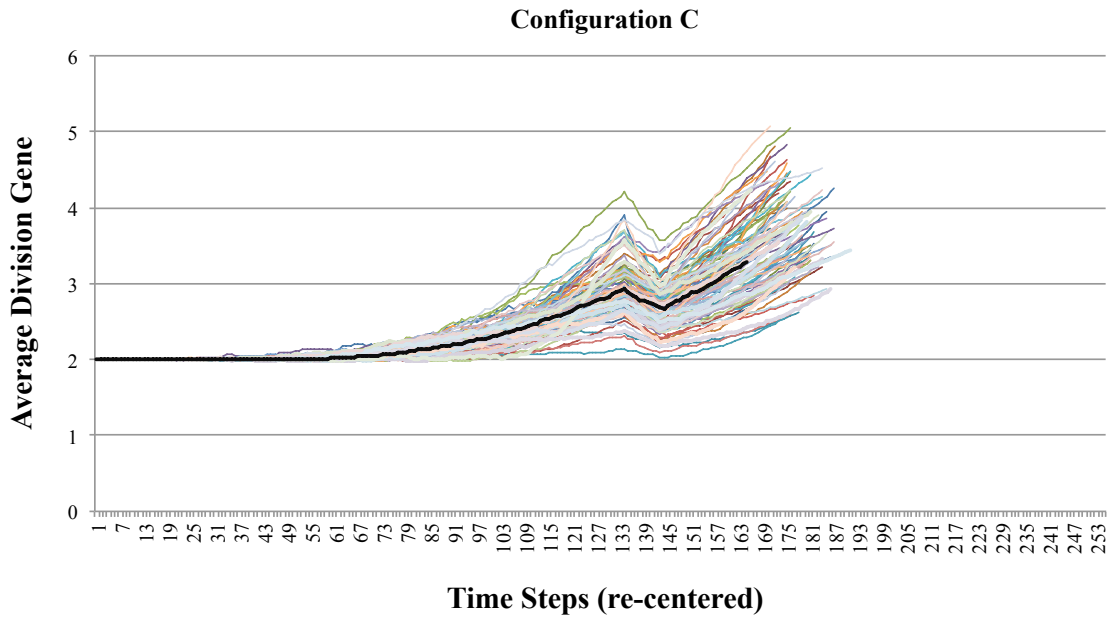
Configuration C

Figure 6.9- Measurement of the average number of Division Genes under Chemotherapy scenario in Broom Diagrams A. 100 simulations of Configuration B. B. 100 simulations of Configuration C. Each individual simulation is represented as line of different colour, with the median as a thick, black line (calculated until one of the simulations came to an end). Simulations were re-aligned with respect to the time step in which treatment starts.

178

A.
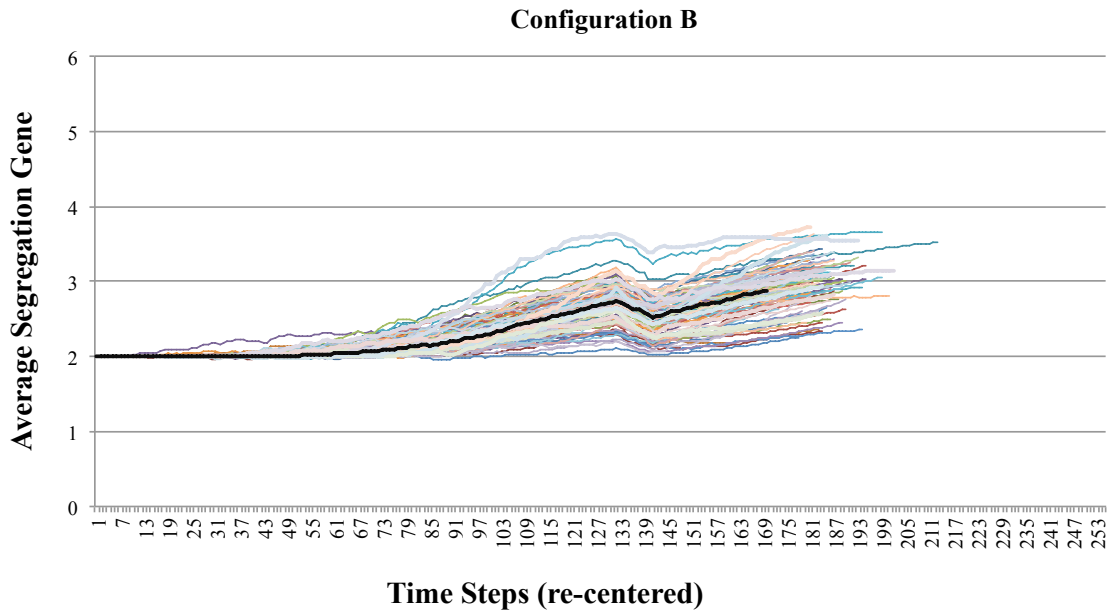
B.
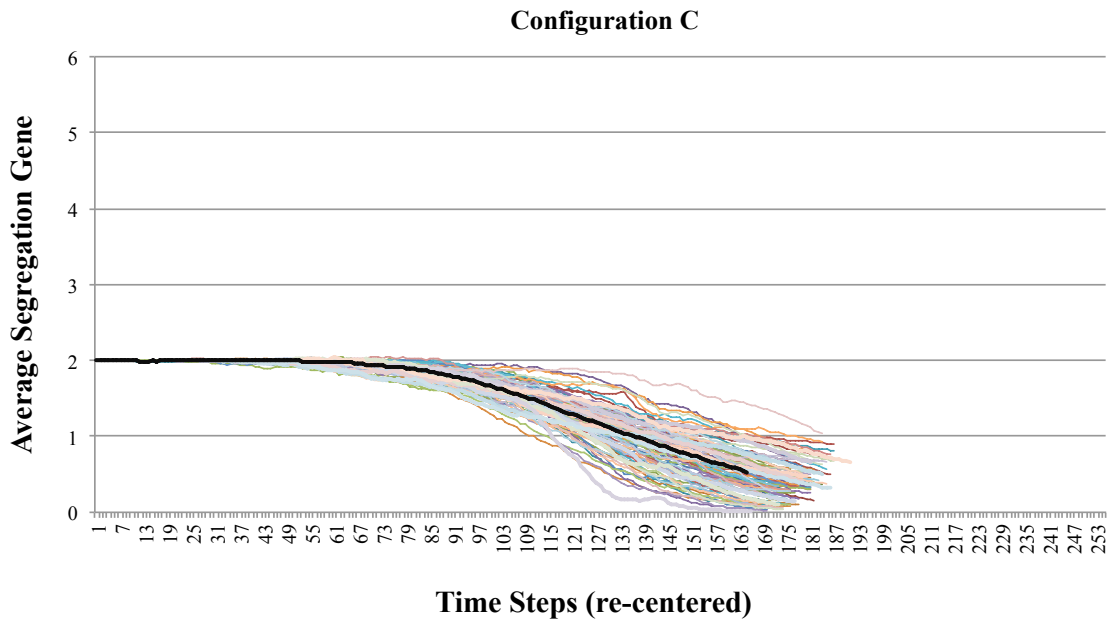
Figure 6.10- Measurement of the average number of Segregation Genes under Chemotherapy scenario in Broom Diagrams A. 100 simulations of Configuration B. B. 100 simulations of Configuration C. Each individual simulation is represented as line of different colour, with the median as a thick, black line (calculated until one of the simulations came to an end). Simulations were re-aligned with respect to the time step in which treatment starts.

A.

Configuration B

B.
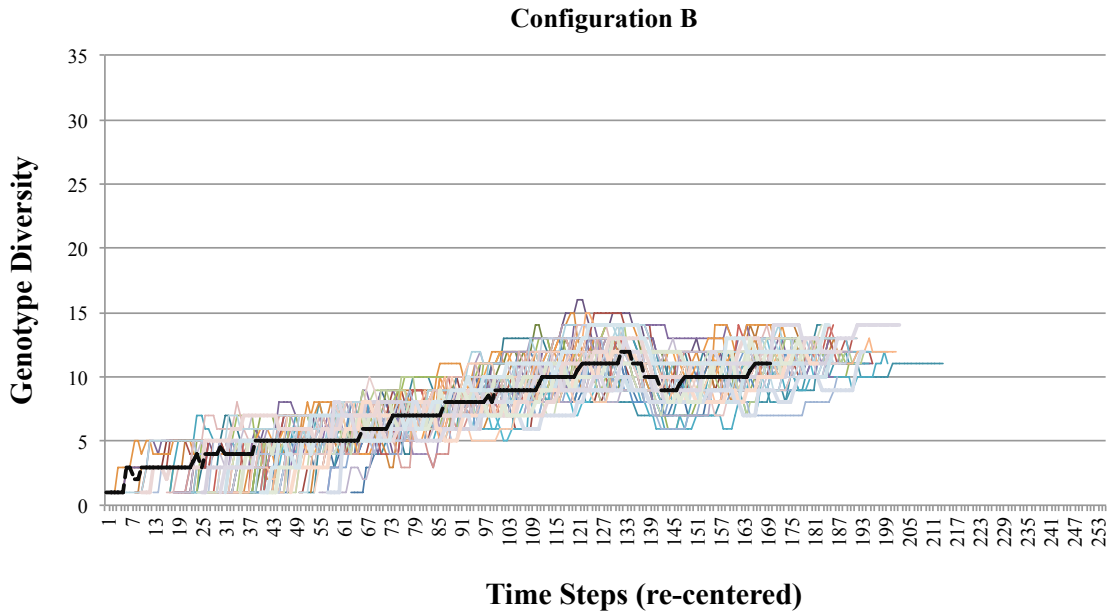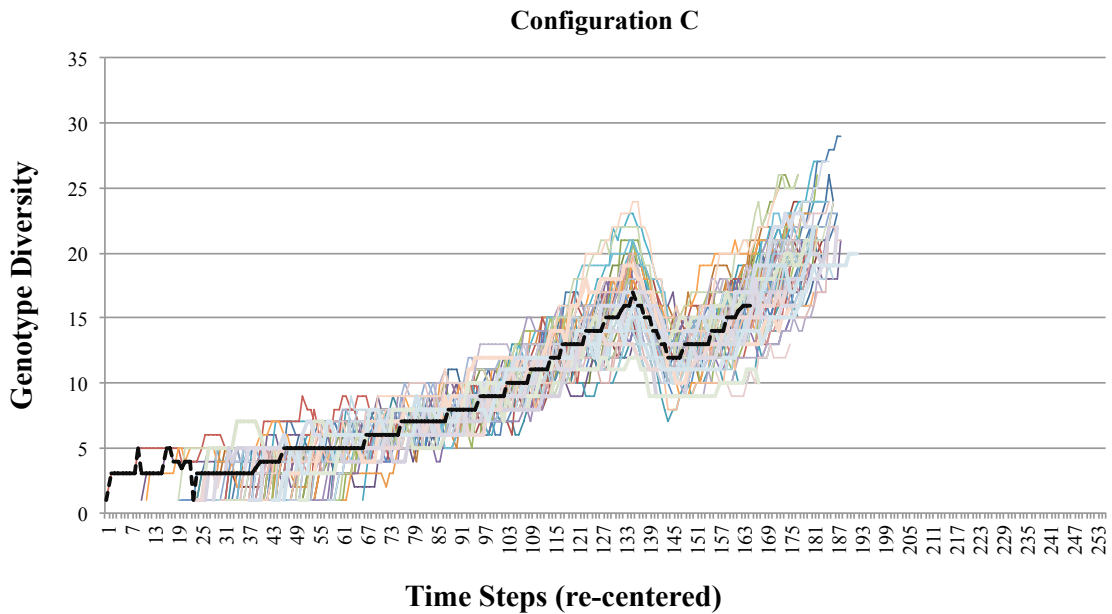
Configuration C

Figure 6.11- Measurement of the Genotype Diversity under Chemotherapy scenario represented in Broom Diagrams A. 100 simulations of Configuration B. B. 100 simulations of Configuration C. Each individual simulation is represented as line of different colour, with the median as a thick, black line (calculated until one of the simulations came to an end). Colours are purely used to distinguish runs. Simulations were re-aligned with respect to the time step in which treatment starts.

180

6.6. Scenario iii: Combination Therapy

6.6.1. Objective and Setup

As in common in the clinic, due to a significant improvement in overall survival, it was decided to model surgery combined with postoperative adjuvant chemotherapy (X. Wu et al., 2011). This combination was modelled by implementing surgery, the removal of 90% of tumour mass as described in Scenario 1; followed by 9 rounds of chemotherapy, the systematic killing of cells that enter mitosis during the 9 therapy time steps. For these experiments, the same system as described in Section 5.3 was used. 100 simulations for Gene Configuration B (Section 5.3.2) and Configuration C (Section 5.3.3), with the same seeds as those described in Section 5.4, were performed.

6.6.2. Results

As in the clinic, the combination of both treatments yield more noticeable results towards longer survival times of both configurations, as can be seen in a plot for the total number of cells for Configuration B (Figure 6.12.A) and Configuration C (Figure 6.12.B). Broom Plots for the loss of Apoptosis Regulatory Genes for Configuration B (Figure 6.13.A) and Configuration C (Figure 6.13.B) reveal that the initial cell removal creates an offset for the average number of apoptosis genes, bringing the average back to the original diploid state. After this, chemotherapy maintains this gene stable. However, at the end of therapy, there is a linear relapse for both configurations.

During this Combination Therapy, the number of Division Genes continues decreasing after the initial offset caused by surgery. Configuration B (Figure 6.14.A) has a larger rate of decrease compared to Configuration C (Figure 6.14.B). Afterwards, both configurations have a linear gain

of this kind of oncogene, with Configuration C having a higher rate of acquisition than Configuration B.

Segregation Genes in Configuration B (Figure 6.15.A) have a similar response as in Scenario ii: The gene has a linear rate of loss towards diploidy, making it less genetically stable for the duration of therapy. Afterwards, the rate of gain of this gene is reinstated. For Configuration C (Figure 6.15.B), There is an initial offset that increases the average number of this gene. Synergizing with this initial reduction, chemotherapy maintains this gene stable for the duration of the therapy, making cells are more genetically stable.

During treatments, the number of different genetic populations in the simulation is reduced in both configurations. After treatments, Configuration B has stabilized its genotype diversity (Figure 6.16.A), while that in Configuration C continues to increase linearly (Figure 6.16.B).

A.
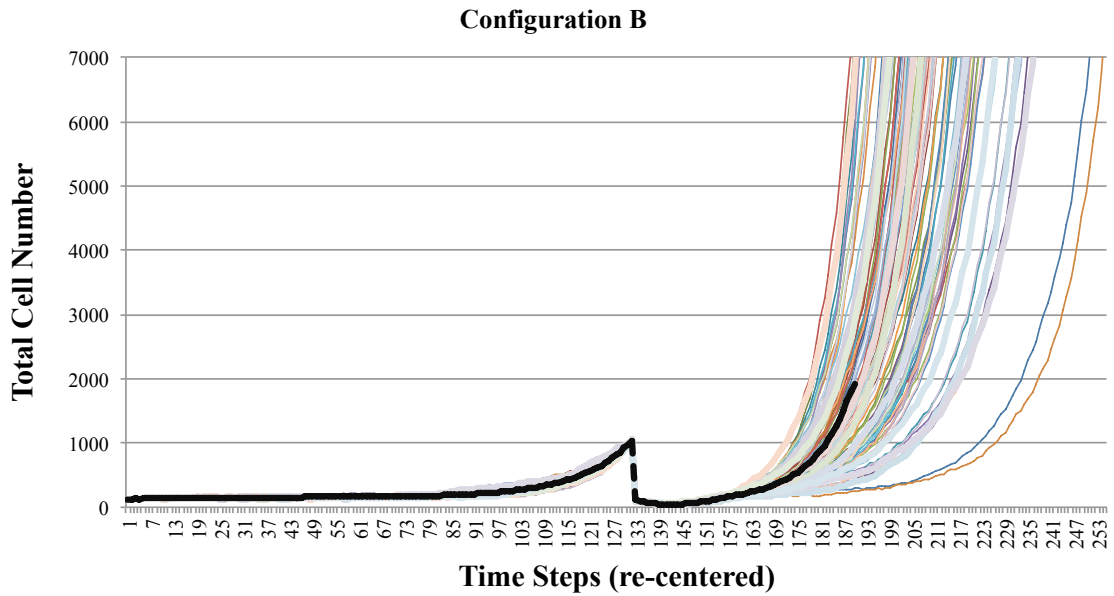
B.
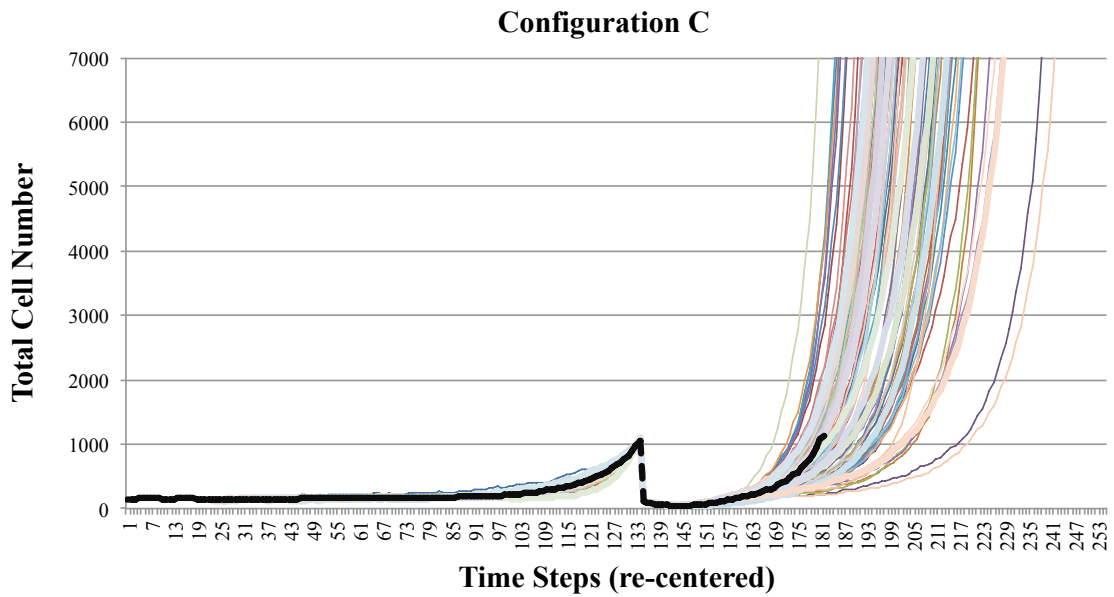
Figure 6.12- Total number of cells for the simulation of Combined Treatments. A. 100 simulations for Configuration B. B. 100 simulations for Configuration C. Each simulation is represented as line of different colour, with the median as a thick, black line (calculated until one of the simulations came to an end). Simulations were re-aligned with respect to the time step in which treatment starts.
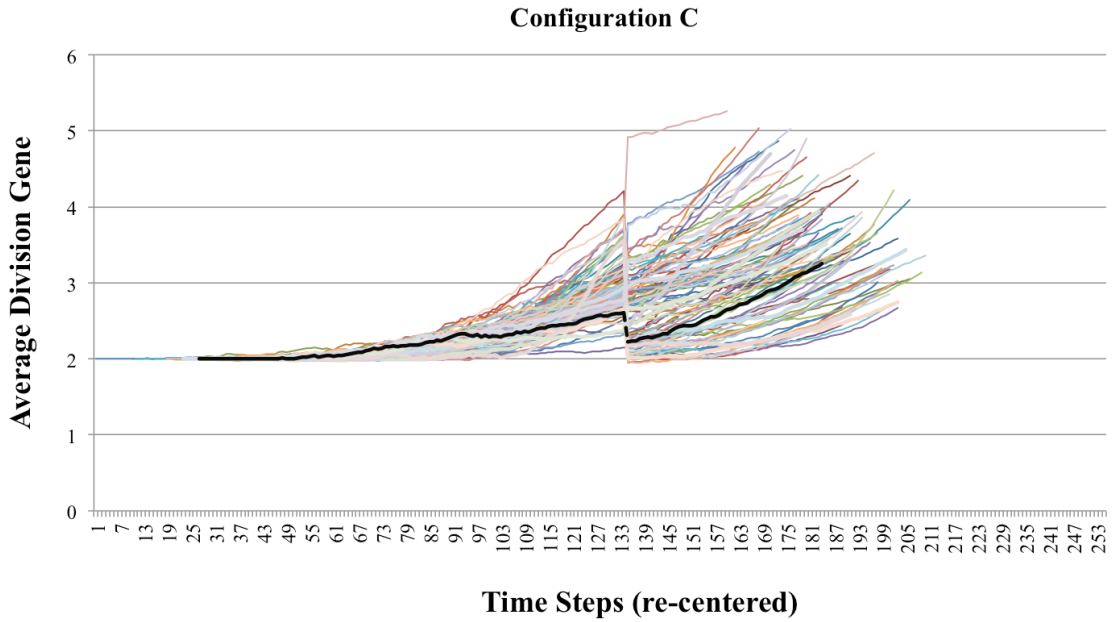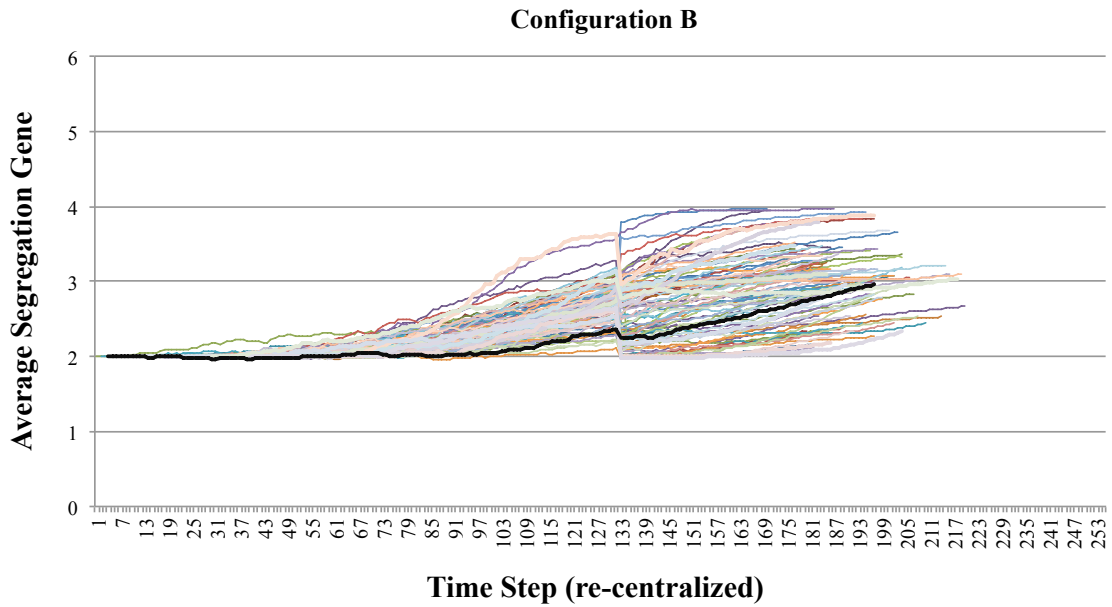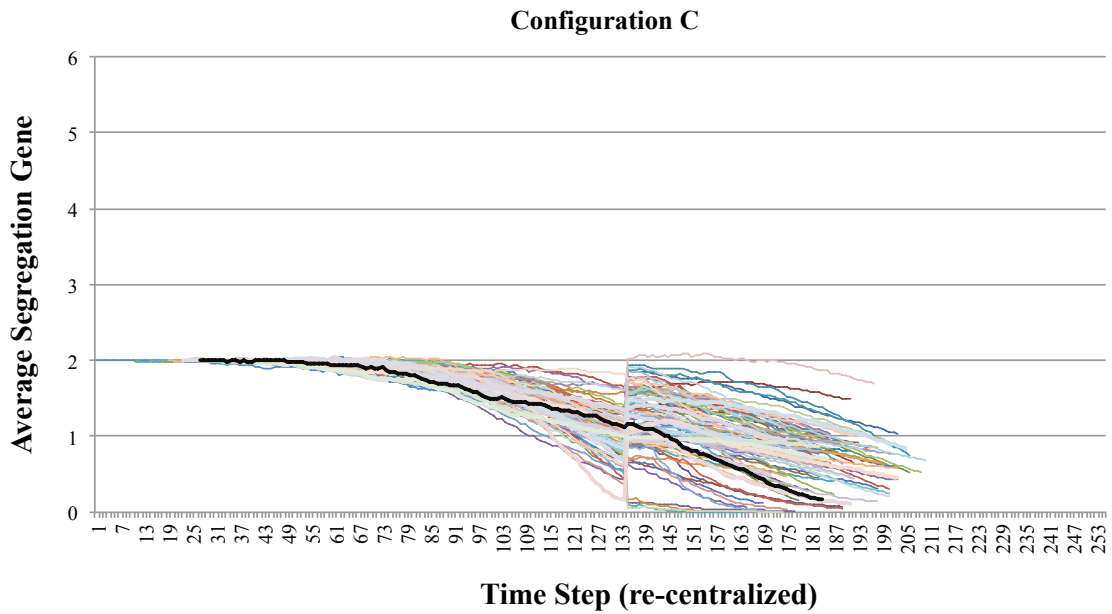
A.

Configuration B

B.
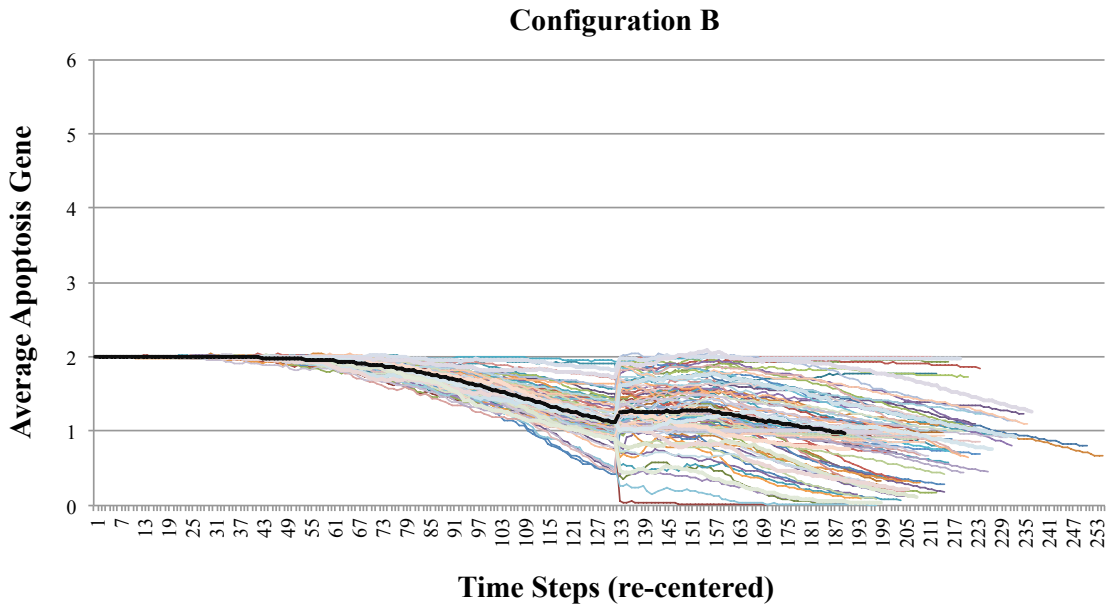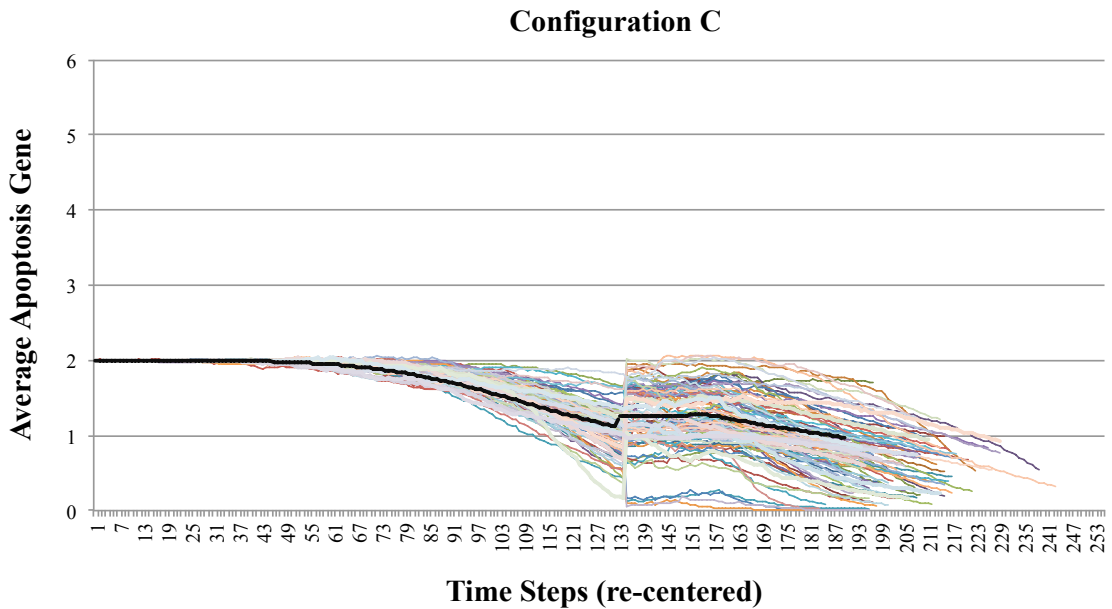
Configuration C

Figure 6.13- Measurement of the average number of Apoptosis Genes for a scenario of Combined Treatments in Broom Diagrams A. 100 simulations of Configuration B. B. 100 simulations of Configuration C. Each individual simulation is represented as line of different colour, with the median as a thick, black line (calculated until one of the simulations came to an end). Simulations were re-aligned with respect to the time step in which treatment starts.

184

A.
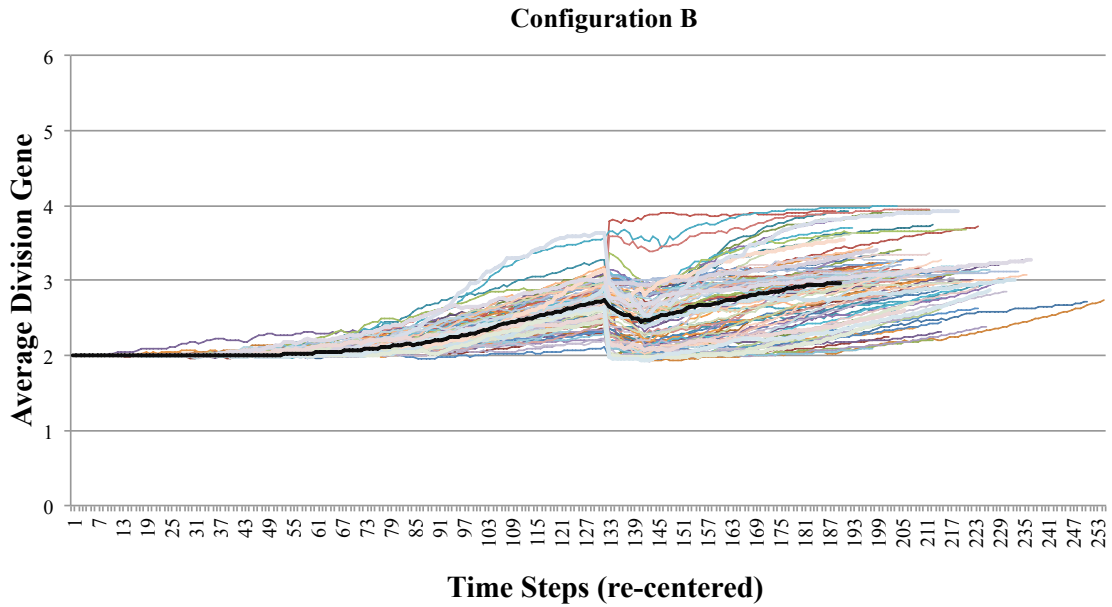
B.
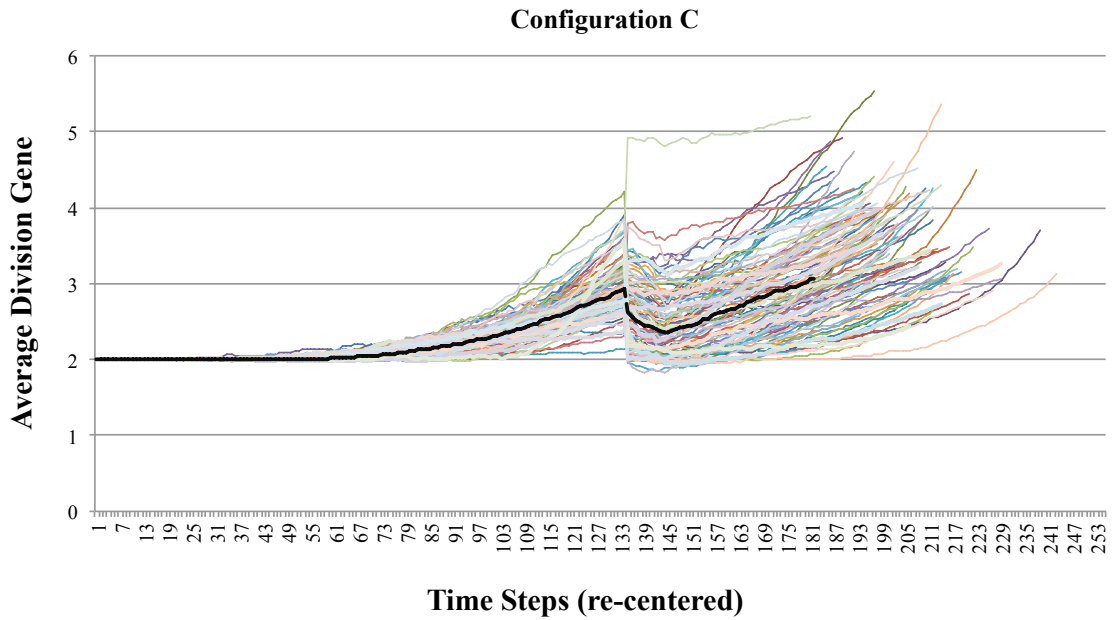
Figure 6.14- Measurement of the average number of Division Genes for a scenario of Combined Treatments in Broom Diagrams A. 100 simulations of Configuration B. B. 100 simulations of Configuration C. Each individual simulation is represented as line of different colour, with the median as a thick, black line (calculated until one of the simulations came to an end). Simulations were re-aligned with respect to the time step in which treatment starts.

A.

Configuration B

B.
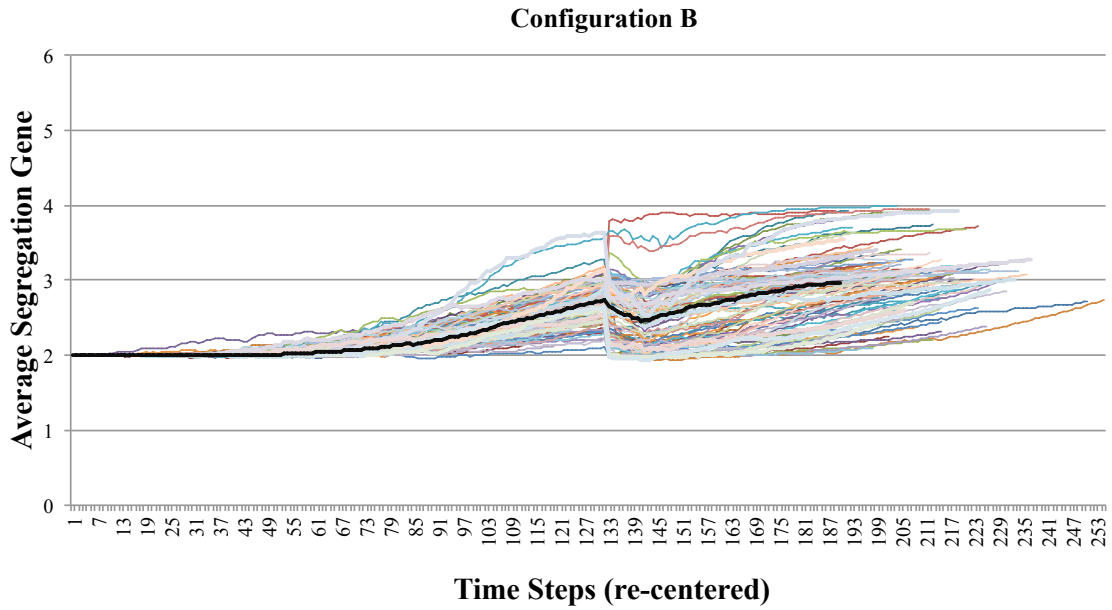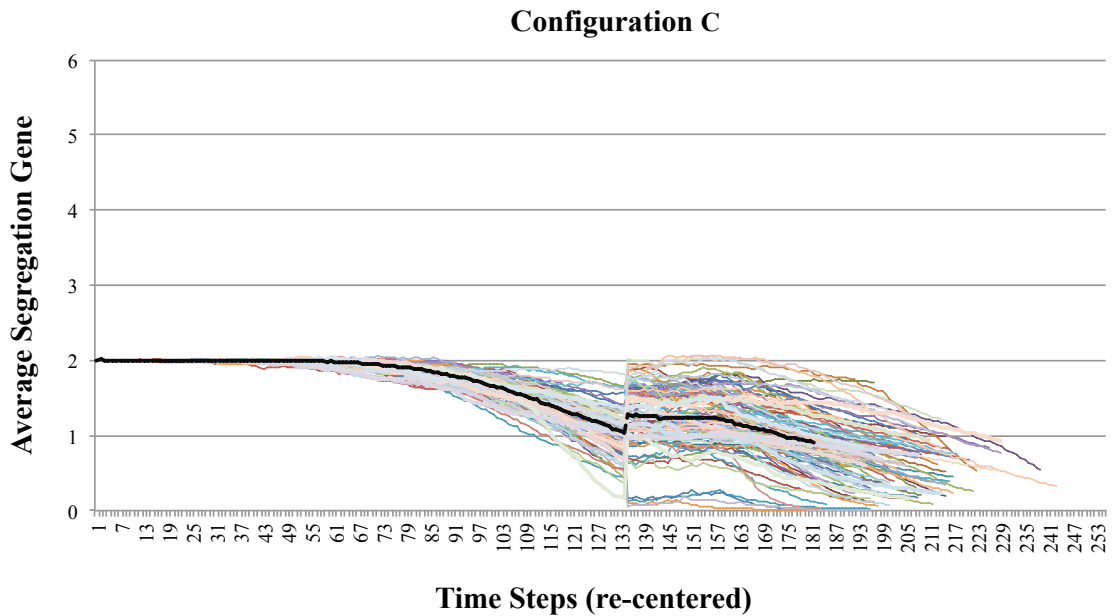
Configuration C

Figure 6.15- Measurement of the average number of Segregation Genes for a scenario of Combined Treatments in Broom Diagrams A. 100 simulations of Configuration B. B. 100 simulations of Configuration C. Each individual simulation is represented as line of different colour, with the median as a thick, black line (calculated until one of the simulations came to an end). Simulations were re-aligned with respect to the time step in which treatment starts.

186

A.

Configuration B

B.
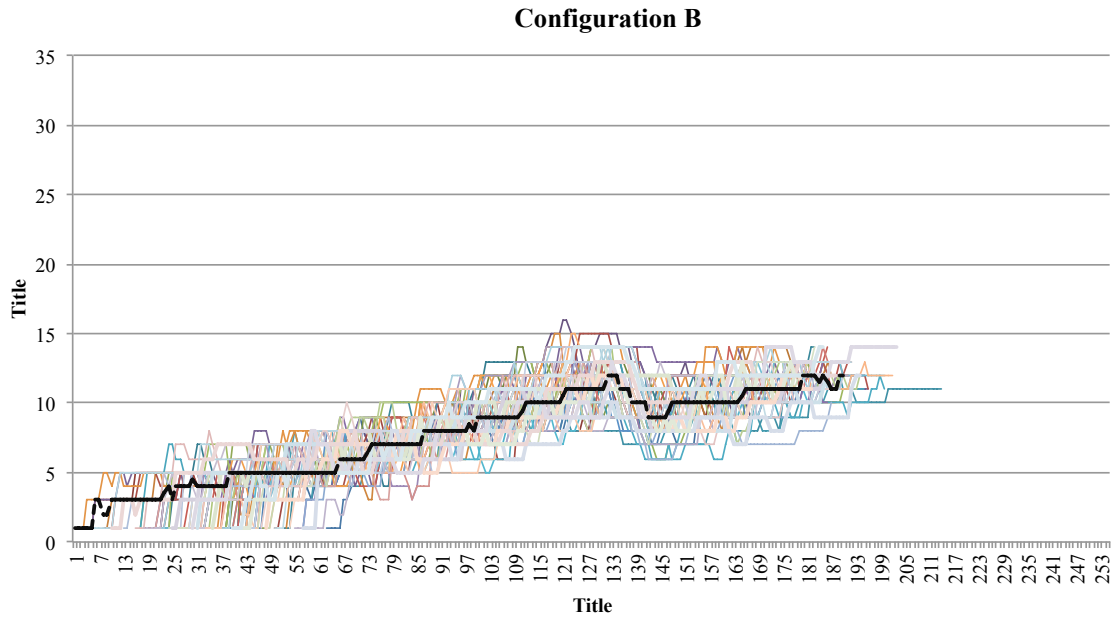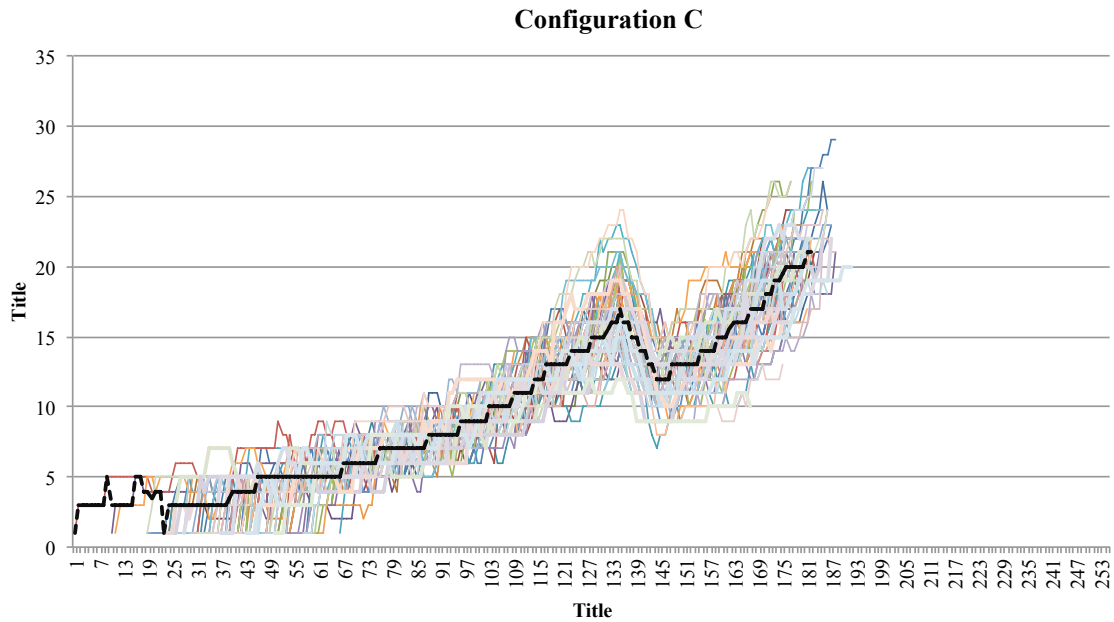
Configuration C

Figure 6.16- Measurement of the Genotype Diversity under Combined Treatment scenario represented in Broom Diagrams A. 100 simulations of Configuration B. B. 100 simulations of Configuration C. Each individual simulation is represented as line of different colour, with the median as a thick, black line (calculated until one of the simulations came to an end). Colours are purely used to distinguish runs. Simulations were re-aligned with respect to the time step in which treatment starts.

6.7. Analysis

For the analysis of such complex interactions, a set of new tools was developed. These tools will be explained and then referred to through the analysis section.

6.7.1. Analysis Tools

a)   *Heat Maps*

Because we tested the therapies with simulations that have the same integral state, it is very useful to compare the outcome of therapies for those individual simulations across the different scenarios. For this, we developed Heat Maps as a way to visually analyse the relapse time for individual simulations across experiments. This was accomplished by writing Mathematica code that reads the text-based output of the C++ simulation, as described in Section 5.2, in which the entire state of the simulation, time step by time step, is stored. As the Mathematica code reads the entire set of simulations, it generates a record of the time of relapse for each simulation; when cells grow back to 1000 after treatment. The Mathematica output is then fed into graphics software, in this case Excel, to produce the visual output that is required. This is done by using a Visual Basic script to assign a colour to each configuration based on its time of relapse. This produces a Heat Map in which the best outcomes (the ones with the most delayed relapse time) across simulations were coloured blue, and the ones with the worst outcomes were coloured red, as seen in Figure 6.17.
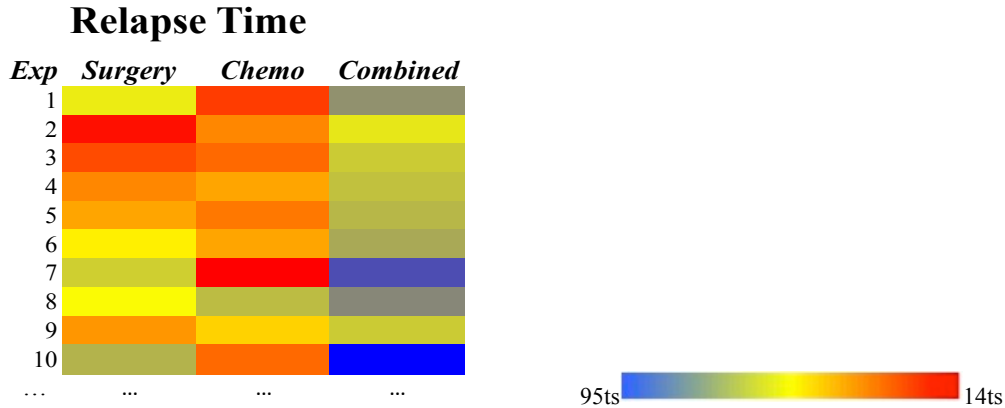
Relapse Time

Figure 6.17- Sample of a Heat Map for relapse of a Genetic Configuration. Columns are the different scenarios, while rows are the individual simulations. The Heat Map scale goes from better prognosis (the longest relapse time was 95 time steps, and was assigned blue at one end of the spectrum) to worse prognosis (the shortest relapse time was 14 time steps, and was assigned red at the other end of the spectrum) according to the scale shown. The number of time steps was determined through a Mathematica code, as described in Section 6.7.1.a.

b)   Marble Diagrams

To better understand the key genetic alterations and cellular processes that affect the treatments, interesting simulations can be individually analysed. This can be accomplished by generating a Heat Map for the entire set of simulations (Figure 6.17) and then selecting those that, based on the results obtained in the previous section, could give us a representative view into the evolution of the genotypes. From the Heatmap, two representative simulations of each configuration, one with the best general prognosis and one with the worst general prognosis can give us a better insight on specific genetic events. These can then be plotted as Marble Diagrams, where the stacked percentage of Genetic Diversity across time can be visualized, as seen in Figure 6.18.

189
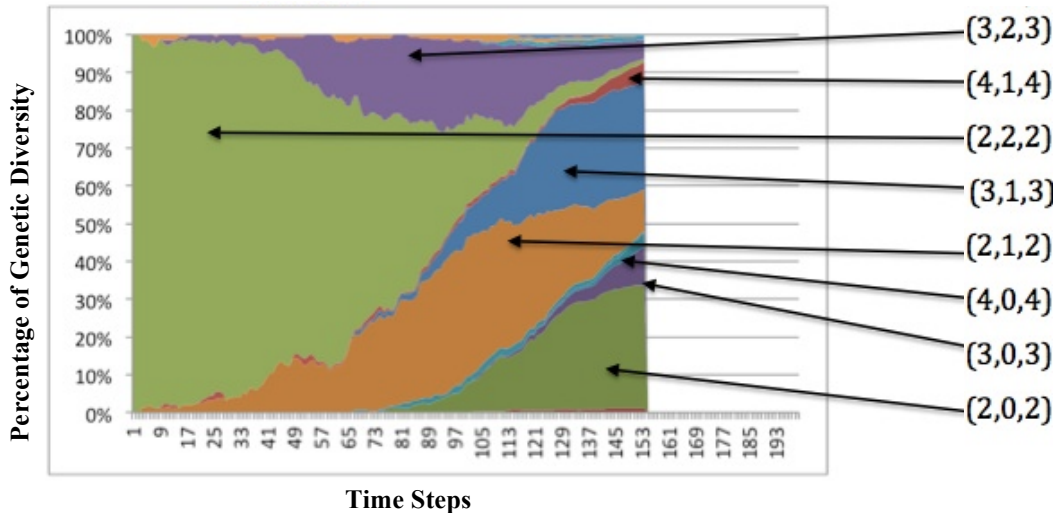
Figure 6.18- Sample of a Marble Diagram. The percentage that a given genetic population, such as the initial genotype (2,2,2), occupies in the genotype population is calculated across time. Each genotype is assigned a colour and plotted as a stacked percentage within the population. The dominant populations change across time, and the dynamics of the diminishing and emergent genotypes are visualized. In this example the proportion of the population with genotype (2,2,2) green reduces as new genotypes (2,1,2) orange and (3,2,3) purple increase by time step 75. By the end of the simulation these are now reducing as genotypes (2,0,2) dark green and (3,0,3) dark purple take over, with yet more genotypes (4,1,4) red and (4,0,4) cyan starting to emerge.

c) *RGB Diagrams*

It is of interest to understand the key similarities that the two simulations with the best prognosis have. We determined that a visual comparison between the two was required. Fort this, an RGB scale was implemented. Colours in the RGB model are defined by three components (Figure 6.23). Because of this, a three-dimensional volume is described by treating the component values as ordinary Cartesian coordinates in a Euclidean space. For the RGB model, this is represented by a cube using non-negative values within a 0–1 range, assigning black to the origin at the vertex (0, 0, 0), and with increasing intensity values running along the three axes up to white at the vertex (1, 1, 1), diagonally opposite black. An RGB triplet (red, green, blue) represents the three-dimensional coordinate of the point of the given colour within an RGB colour cube, or its faces or along its edges in a simplified version. This approach allows computations of the colour similarity of two given RGB colours by simply calculating the distance between them: the shorter

the distance, the higher the similarity. We have taken advantage of this to describe the different genotypes that evolve in our system by normalizing the maximum observed Genotype State (The triplet used for gene quantification, as described in Chapter 5.3). We have assigned a colour to each of the abstracted genes: Red for division, green for death and blue for segregation; or in triplet notation (Division=Red, Apoptosis=Green, Segregation=Blue), as seen in Figure 6.19. By comparing the similarity of the colours assigned to a given genotype, we are able to tell visually the proportions in which the genes are distributed, with intensity values corresponding to the number of genes: (0,0,0) being black, the initial genotype (2, 2, 2) being dark grey and the maximum observed genotype of interest (5, 5, 5) being white.
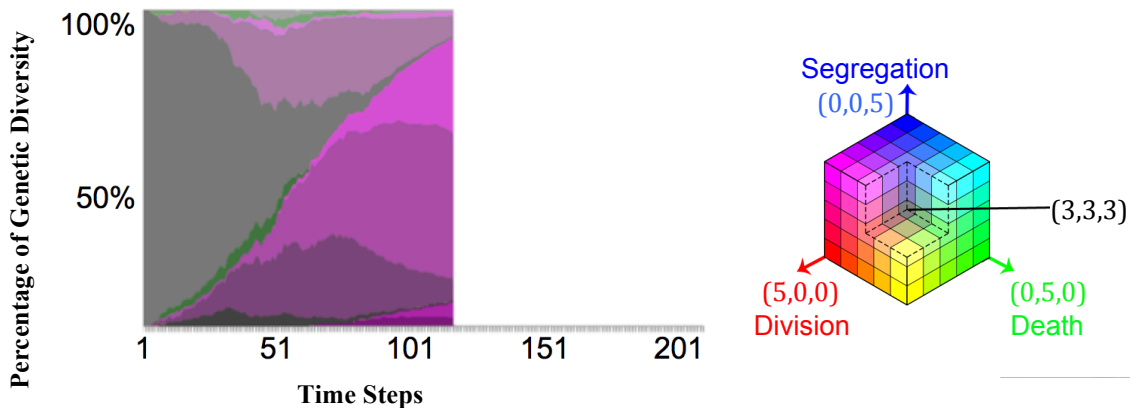
Figure 6.19- A sample RGB Diagram. We have used the RGB colour model, as seen to the left, to visually describe the different genotypes that evolve in the system by normalizing the maximum observed Genotype State (Division, Apoptosis, Segregation). We have assigned a colour to each of the abstracted genes: red for division, green for death and blue for segregation. By comparing via an RGB system the colours assigned to a given genotype, we are able to tell visually the proportions in which the genes are distributed, with intensity values corresponding to the number of genes: (0,0,0) being black, the initial genotype (2, 2, 2) being dark grey and the maximum observed genotype of interest (5, 5, 5) being white. In this example, we start with grey (2,2,2). The proportion of the population with genotype (2,2,2) grey reduces as new genotypes (3,2,3) grey-pink and (2,1,2) purple increase by time step 75. By the end of the simulation these are now reducing as genotypes (3,1,3) pink and (2,0,2) purple take over. Purple represents a mixture of a genotype with Segregation and Division genes, with no Apoptosis Genes. Green genotypes, those still containing copies of the Apoptosis Gene disappear at the end of the simulation.

d)	***Statistical tests***

For the statistical tests we used an unpaired t-test (Rochon, Gondan, & Kieser, 2012) to determine if the means of the results in our two sets of experiments (Configuration B and C) are significantly different in key aspects of simulated treatments. Our null hypothesis is that the observed response of the two configurations to treatments is due to chance. The alternative hypothesis is that the observed response to treatments depends on the configuration. For the tests, we have assumed a two-tailed distribution and equal variance.

6.7.2. Analysis

Sections 6.4, 6.5 and 6.6 investigated the dynamics of each gene across the simluations; how therapies affected these dynamics and the impact on the number of different genotypes in the cell populations. We now analyse the results from those experiments using the new analysis tools. It was decided to search for those simulations with the best and worse prognosis. Heat Maps were used to search for this information. It is important to note that all results depend on the settings used and may not be valid for other values.

When the simulation of tumour removal surgery was implemented by eliminating all but 10% of cells in the tissue (Scenario i), results were highly variable and depended on the kind of tumour cells that survived (Figure 6.2). Though the actual evolutionary pathways exhibit a high degree of variation across simulations, a pair of representative Marble Diagrams for each gene distribution captured the kinds of evolutionary pathway that most of the simulations followed, as shown in Figure 6.21 for Configuration B and Figure 6.22 for Configuration C.
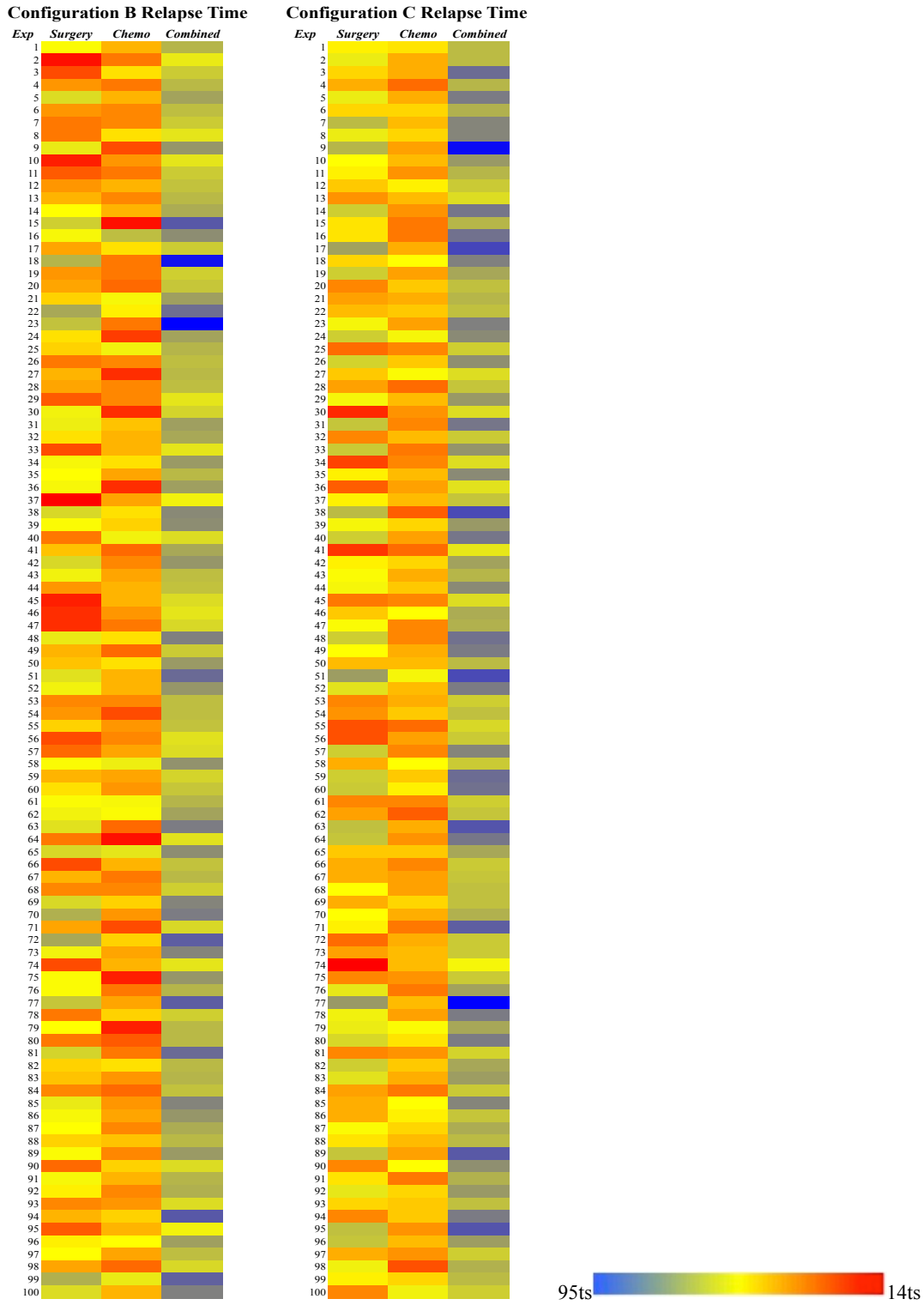
Figure 6.20- Heat Map for relapse of Configuration B (left) and Configuration C (right). Columns are the different scenarios, while rows are the individual simulations. The Heat Map scale goes from better prognosis (95 time steps, blue) to worse prognosis (14 time steps, red) according to the scale shown.

Marble Diagrams for Configuration B (Figure 6.21) reveal that treatments effectively reduce the size of some genetic populations, while leaving room for a small number of populations to dominate the relapsed tumour. From these diagrams, it can be appreciated how the recovery of genotypes with better tumour suppression, as denoted in the triplet Genotype Notation (Division, Apoptosis, Segregation), lead to a better prognosis.

On the other hand Marble Diagrams for Configuration C (Figure 6.22) reveal that Genetic Instability plays a key role in determining the prognosis of a tumour. Simulations with better prognosis (Figure 6.22.i) have recovered some tumour suppression in the form of Division Genes, and some stability in the form of Segregation Genes. However, a more chromosomally unstable simulation (Figure 6.22.ii) reveals that once tumour suppression is lost, a few, highly aggressive genotypes will take over the population even in the presence of therapies. To directly compare the two overproliferative configurations, it was decided to use a colour-coding key to represent the information as the RGB Diagrams.

A more direct comparison across the two best outcomes through RGB Diagrams, that of Configuration B (Figure 6.23.i) and Configuration C (Figure 6.23.ii), across all the scenarios reveals that the tumour strives to gain copies of oncogenes (acquiring red colour) and lose copies of the Tumour Suppressor (shedding green colour). However, Configuration B is modulated by the acquisition of Segregation Genes (a blue shade that mixes with the oncogene red to create purple), while Configuration D lacks this Gene (as the resulting genotypes are pure red).
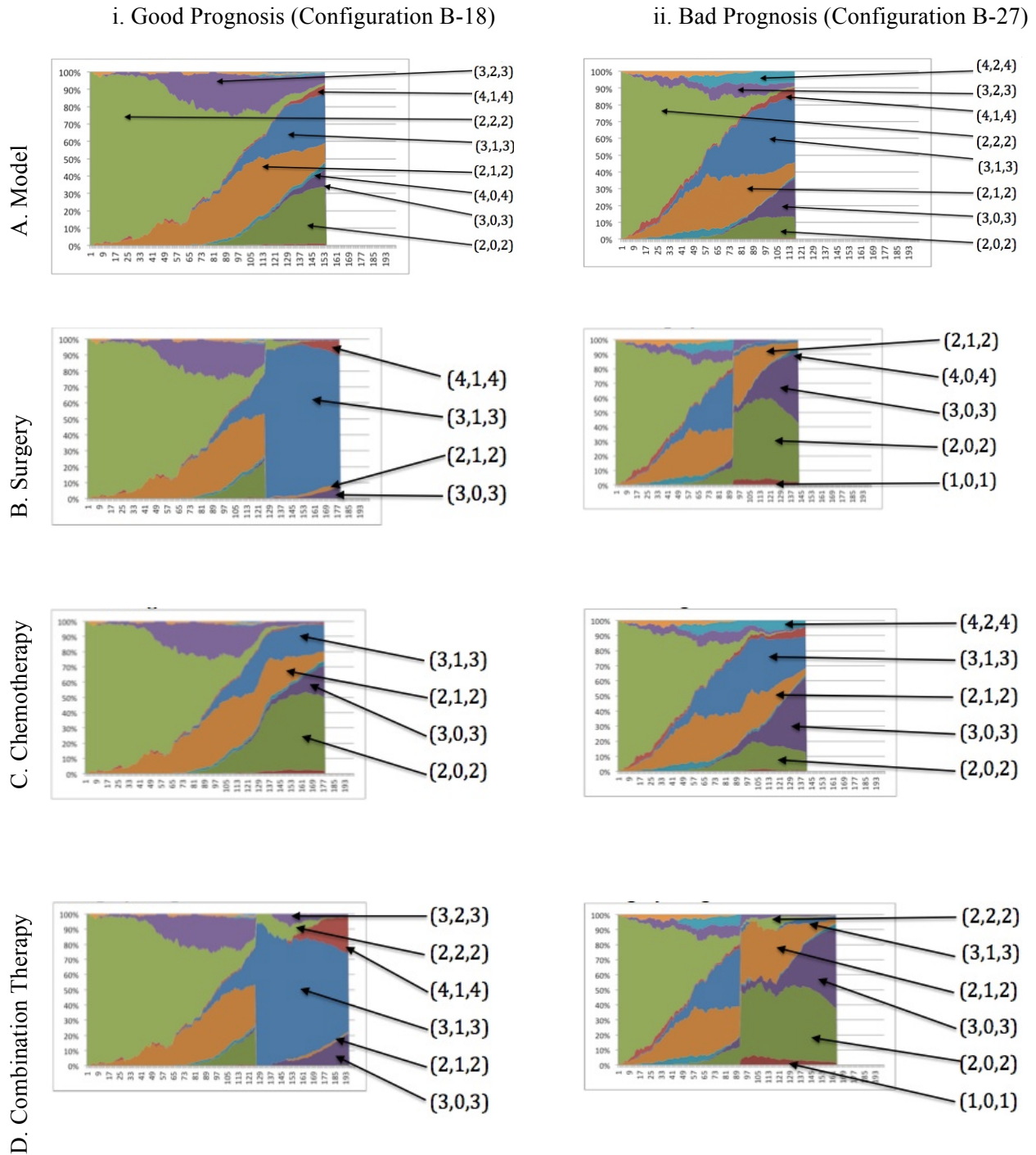
Figure 6.21- Marble Diagrams. These diagrams display the stacked percentage of Genetic Diversity across time for two representative simulations of Gene Configuration B (Column i. Better Prognosis; Column ii. Worse Prognosis) across different scenarios: Row A. The Model; row B. Surgery (scenario i); row C. Chemotherapy (scenario ii); row D. Combination Therapy (scenario iii).
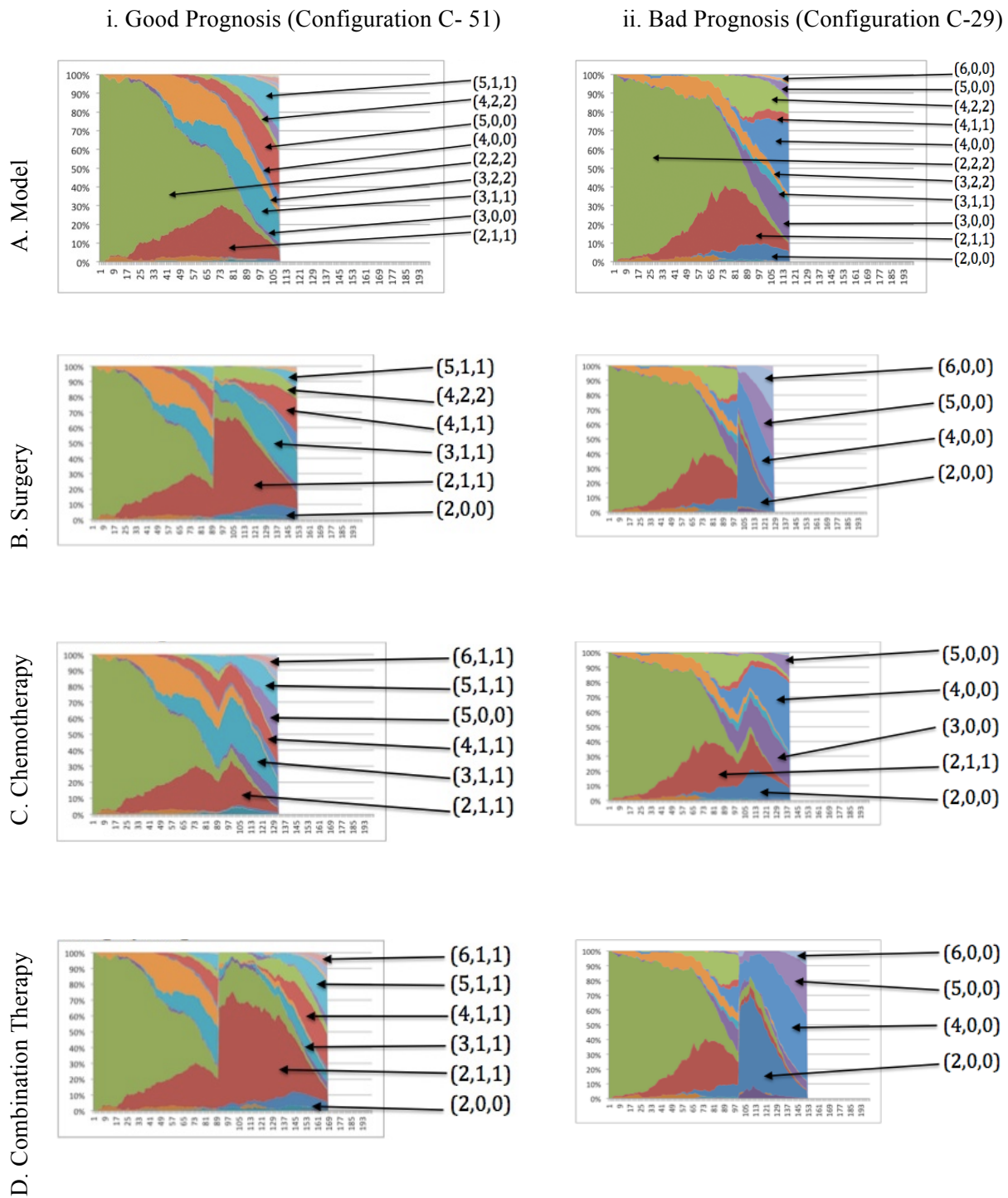
Figure 6.22- Marble Diagrams. These diagrams display the stacked percentage of Genetic Diversity across time for two representative simulations of Gene Configuration C (Column i. Better Prognosis; Column ii. Worse Prognosis) across different scenarios: Row A. The Model; row B. Surgery (scenario i); row C. Chemotherapy (scenario ii); row D. Combination Therapy (scenario iii).

i. Configuration B

ii. Configuration C

Segregation
(0,0,5)

(3,3,3)

(5,0,0)
Division

(0,5,0)
Death

A. Model

B. Surgery

C. Chemotherapy

D. Combination

Time Steps

Time Steps
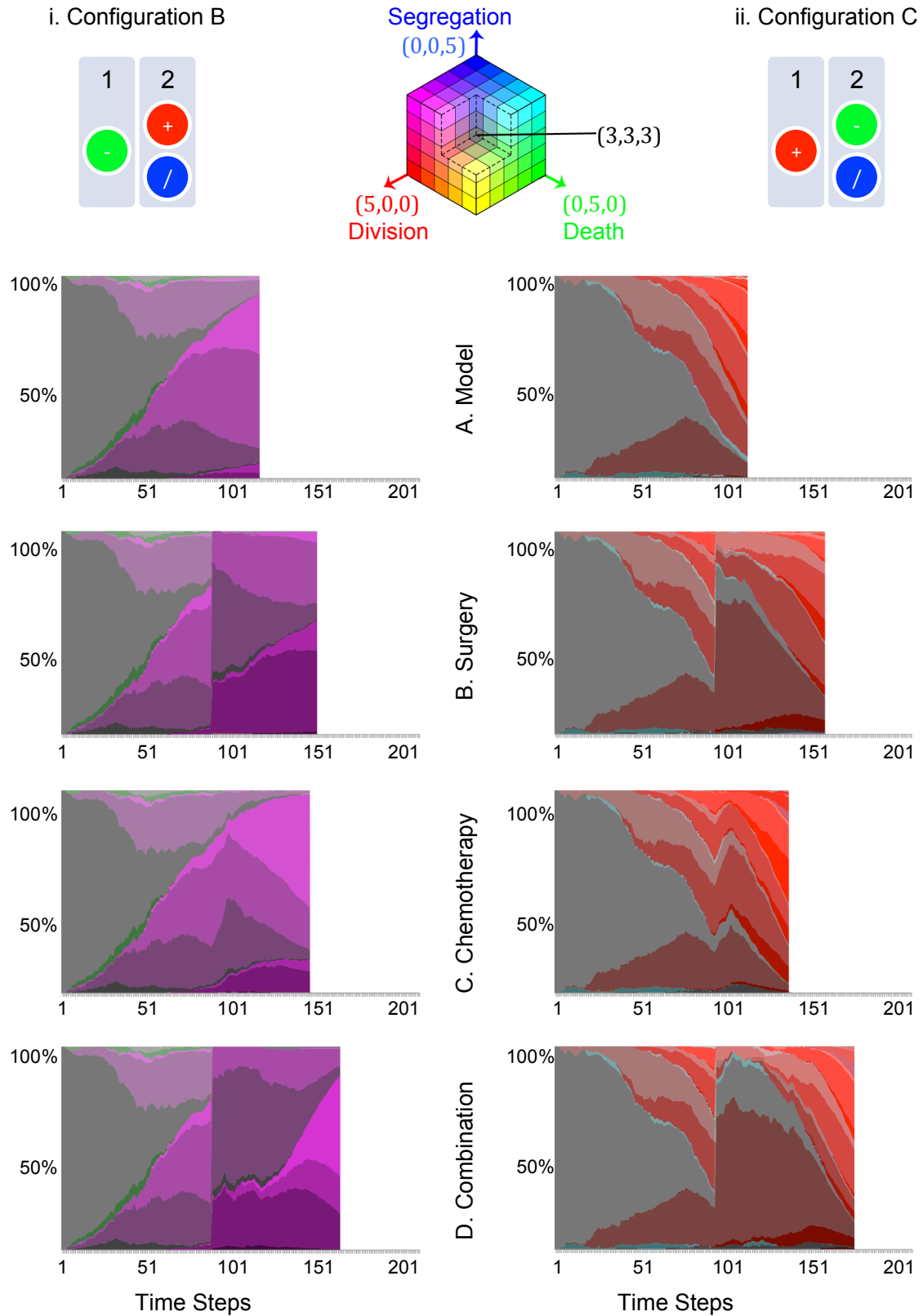
Figure 6.23- The two overproliferative genetic arrangements, in simulated diploid chromosomes, and the RGB key in the middle. These diagrams display the stacked percentage of Genetic Diversity

across time for a representative simulation of Gene Configurations B and C across different scenarios. The beginning of therapies (when reaching 1000 cells) are marked with a black vertical line, while relapse times (when reaching again 1000 cells) are marked using a dashed line. A. Representative Marble Diagram for a simulation with the Model. B. Representative Marble Diagram for a Simulation of Surgery. C. Representative Marble Diagram for a Simulation of Chemotherapy. D. Representative Marble Diagram for a therapy combination of Surgery followed by Chemotherapy.

After surgery an average of 105 cells were left (std. 4.50) for distribution B and 106 cells (std. 5.13) for distribution C. However, over 100 simulations the prognosis was significantly better (p=0.0499) for tumours with Gene Distribution B, which exhibit relatively low levels of chromosome missegregation (relapse time was an average of 35.22 time steps and a standard deviation of 8.33), compared to those with Gene Distribution C and high levels of chromosome missegregation (with an average of 32.84 and a standard deviation of 8.70), as seen in Figure 6.24.A. This behaviour was due in part to the greater likelihood of a relatively normal population of cells remaining after surgery from a population with low genetic heterogeneity in comparison to that from a highly heterogeneous population.
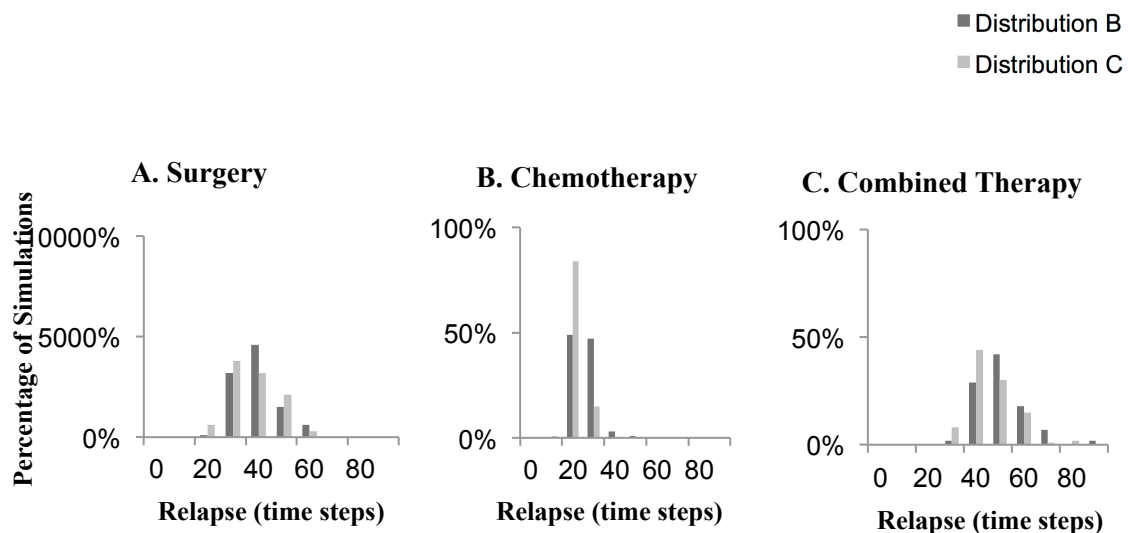
Figure 6.24- Distribution of the response to treatments under different scenarios. The histograms correspond to a measure of the distribution of the relapse times (the time it took each simulation to grow back to 1000 cells after treatment) for 100 simulations of each gene configuration (Configuration B- dark grey, Configuration C- light grey) under three different therapy scenarios: A. Surgery Scenario, B. Chemotherapy Scenario and C. Combination of both treatments (Surgery followed by Chemotherapy).

Simulation to simulation variability was determined in part by the kind of initial aberrations present in the population. Thus, cells that suffered a loss of tumour suppressors did not overproliferate until they underwent additional missegregation events, delaying the period of time until relapse. In contrast, for simulations in which overproliferative genotypes are the first to emerge, cells that remain after surgery quickly re-grow into a full-sized tumour. Thus, the relapse time in simulations is determined, primarily, by the oncogenic load, which is higher in the chromosomally unstable populations. As a measure of the difference in the two types of lesion driving tumour growth, we compared the ratio of the average number of Apoptosis Genes to the average number of Division Genes, as seen in Figure 6.25.A for Surgery. When this was analysed in the 25 time steps after surgery, it was clear that Distribution C has a reproducibly higher rate of loss of Tumour Suppression and Oncogene acquisition. This can be seen most clearly by comparing changes in the rate of the ratio of the average number of Apoptosis Genes to the average number of Division Genes following treatment (Figure 6.25.A), which has an near linear slope of -0.0067 (std. 0.0037) for Distribution C, which is significantly steeper (p=0.005E-1) than the average slope for Distribution B (slope -0.0049, std. 0.0030). This reflects the greater generation of new more malignant genotypes in type C simulations, where chromosomal instability is high, compared to simulations for Distribution B, where aneuploidy is relatively stable. This in turn correlates with a worse prognosis for the genetically unstable tumours.

For simulations of chemotherapy (Scenario ii), the outcome for both distributions was generally worse, reflecting in part the differences in the overall rate of killing induced by the treatment. On average, 226.17 cells (std. 53.12) were left after therapy for Distribution B and 231.88 (std. 50.06) cells for Distribution C. More significantly, while Gene Distribution B relapsed in average at 21.95 time steps (std. 4.89), the relapse time seen in Gene Distribution C was again faster, occurring at an average of 18.30 time steps (std. 3.42) as can be seen in Figure 6.24.B. This

significant difference (p=0.003E-5) in relapse time could again be attributed to differences in genetic diversity. Moreover, when we measured the rate of acquisition of new variants that have increased oncogenic load and a reduced number of tumour suppressor genes (the ratio of the average number of Apoptosis Genes to the average number of Division Genes) there was a marked and significant difference (p=0.004E-7) between simulations over 25 time steps after chemotherapy - an average slope of -0.0048 (std. 0.0016) for Distribution B, and -0.0068, (std. 0.0019) for Distribution C (as seen in Figure 6.25.B). In addition, the spread of behaviour across individual simulations (quantified as a difference in the standard deviation in the two cases) was significantly different between the recovery of tumours following chemotherapy and following surgery.

A combination of the two therapies (Scenario iii) yielded an overall better prognosis for both gene distributions. After this combined therapy there were on average 36.09 (std. 8.56) cells left for Distribution B and 36.29 (std. 7.99) cells for Distribution C. Again, the results indicate that Gene Distribution B still has a significantly better prognosis (p=0.008) than Gene Distribution C: Gene Distribution B had an average relapse of 46.55 (std. 10.06), while Gene Distribution C had an average relapse of 43.09 (std. 9.44). These results can be compared across scenarios in the form of histograms in Figure 6.24.C. Again, the overall impact of genetic linkage on the evolution of the tumour after treatment can be most easily visualized by comparing the average slope of the ratio between Apoptosis and Division Genes (Figure 6.25.C). When we considered the 25 time steps after therapy, this shifted significantly (p=0.005E-2): -0.0036 (std. 0.0025) for Distribution B and -0.0052 (std. 0.0034) for Distribution C.

i. Configuration B　　　　**ii. Configuration C**

A. Surgery

B. Chemotherapy

C. Combination Therapy
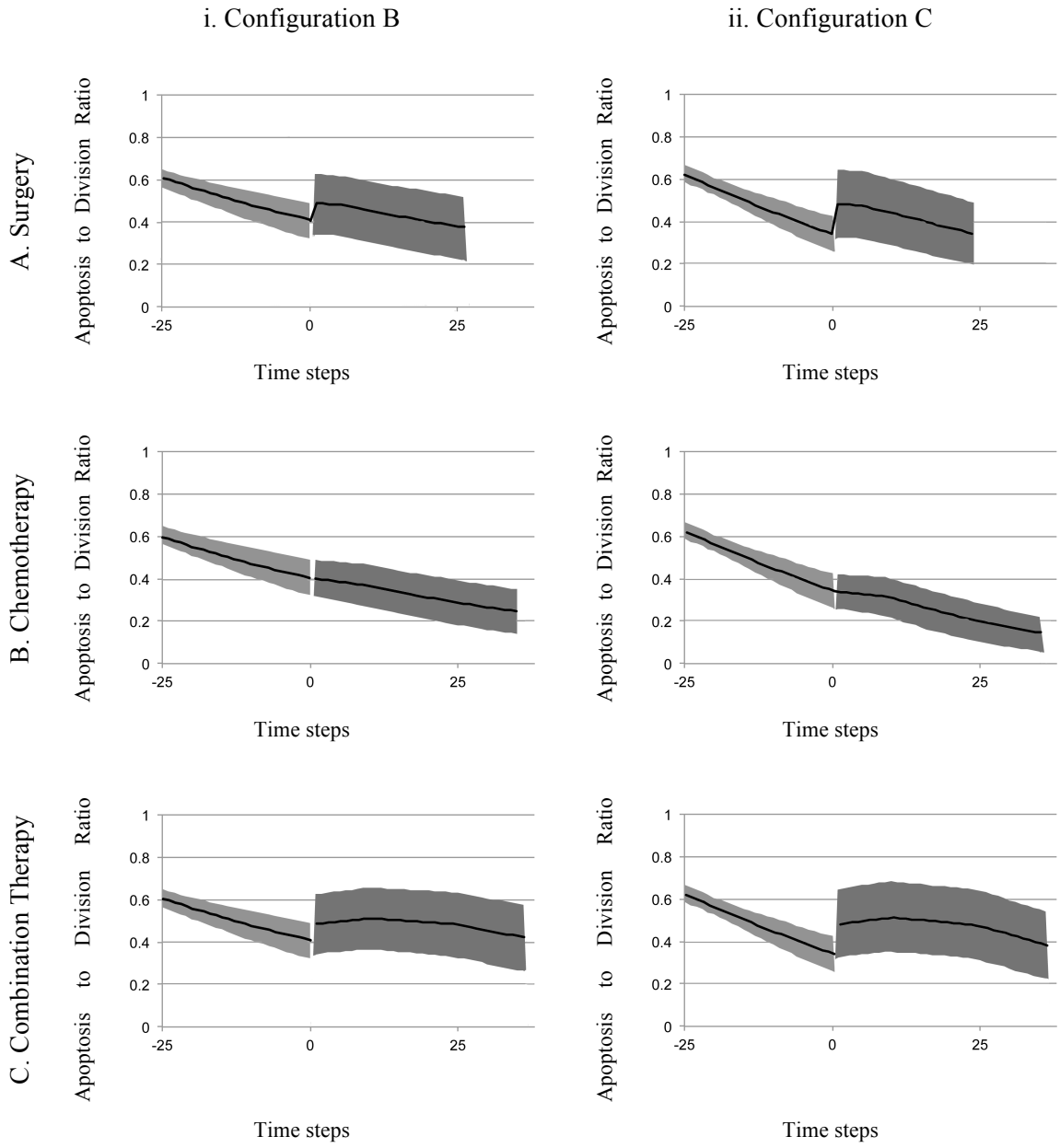
Figure 6.25- The average ratio of apoptosis to division genes. These graphs show the tendency of reducing the number of apoptosis genes and increasing the number of division genes with respect to time across different scenarios: A. Surgery Scenario, B. Chemotherapy Scenario and C. Combination of both treatments (Surgery followed by Chemotherapy). The dark line is the median of the samples and the shadowed area represents the variance. Interventions were carried out at time step zero. The reported slopes were measured taking into account 25 time steps after each therapy.

6.4. Summary

In our model we observe two distinct pathways for evolution towards oncogenesis that have a direct impact on the tumour's response to treatments. In the first case, dominant proliferating clones within the tumour exhibit a relatively stable state of aneuploidy, as seen in leukaemia, lymphomas and some mesenchymal tumours (Johansson et al., 1996) . In the second, selection for the loss of the aneuploidy gene results in tumours that continually generate increasing levels of heterogeneity and ever-more malignant subclones, such as those observed in the neoplastic progression characteristic of epithelial tumours (Lai et al., 2007) (Johansson et al., 1996). In this chapter, we explored the effects of 3 different types of simulated treatment in each case. This revealed interactions between the treatments and the gene distribution in each case. We are now ready to answer the question proposed at the beginning of the Chapter.

Is there a genetic configuration that has a better general prognosis?

Generally, the chromosomally stable configuration (Configuration B) had an overall better prognosis than the genetically unstable configuration (Configuration C), as can be appreciated in the histograms on Figure 6.24.

Are there key genetic events in the evolution of tumours that make them more or less susceptible to cancer treatments?

For this analysis, as a measure of the types of lesion driving tumour formation and relapse, we compared the ratio of the average number of Apoptosis Genes to the average number of Division Genes in simulations. When this was analysed in the 25 time steps after surgery, it was clear that Gene Configuration C has a reproducibly higher rate of loss of Tumour Suppression and Oncogene acquisition than Configuration B. This can be seen most clearly by comparing changes in the rate of the ratio of the average number of Apoptosis Genes to the average number of

Division Genes following treatment , which has an near linear slope of -0.0067 (std. 0.0037) for Configuration C, which is significantly steeper (p=0.005E-1) than the average slope for Configuration B (slope -0.0049, std. 0.0030). This reflects the greater generation of more malignant novel genotypes in Configuration C simulations, where chromosomal instability is high, compared to simulations for Configuration B, where aneuploidy is relatively stable. This in turn correlates with a worse prognosis for the genetically unstable tumours. Thus, in our simulations, surgery acts as a hit-or-miss therapy because it leaves cells that are related to each other. We observed a marked and significant difference (p=0.004E-7) between simulations over 25 time steps after chemotherapy - an average slope of -0.0048 (std. 0.0016) for Configuration B, and -0.0068, (std. 0.0019) for Configuration C. This reflects the presence of higher numbers of cells poised in a pre-cancerous state following treatment in Distribution C.. However there is a high degree of variability across experiments.

Adding to this, treatments seem to reduce the number of populations that make up most of the tumour; especially in the case of chemotherapy. The reduced spread in gene distributions seen following chemotherapy is due to the selective killing of overproliferating cells, compared to surgery, which does not discriminate based upon genetic makeup. In both cases, however, the number of tumour suppressor genes remaining after treatment appeared to be a critical factor in determining the course of the relapse. As can be appreciated from Figure 6.21.C and Figure 6.22.C, once the chemotherapy was applied and the number of cells with overproliferative genotypes reduced, the likelihood that the remaining cells are able to resist overproliferation by means of contact inhibition is increased if they retain an intact apoptosis regulator. If the remaining cells lack an intact tumour suppression gene at the time of therapy, the intervention can act as an evolutionary bottleneck to leave a population dominated by malignant cells. In addition, in simulations in which chromosomal instability is high, cells that survive the treatment as the

result of a slow rate of cell proliferation rapidly acquire extra copies of the oncogenes to reinitiate tumour formation.

Are there any predictive markers for a successful therapy?

Our simulations suggests that if there are a number of genotypes that retain functional tumour suppressors at the time of therapeutic intervention, it is possible to recover less aggressive genotypes, leading to a better prognosis. If tumour suppressor function is compromised prior to treatment, the intervention can lead to an evolutionary bottleneck that selects for malignancy.

Why does surgery appear to give rise to the best outcomes in simulations? Even though it seems intuitively likely that chemotherapy, by selectively targeting actively proliferating cells, would be more a effective treatment than surgery, in reality there are two kinds of dangerous cells. There are cells that have lost tumour suppression and the ones that have acquired oncogenes. While surgery acts against both kinds of cells, chemotherapy will miss cells that divide slowly but which are no longer subject to apoptosis-mediated tumour suppression. Thus, while both therapies lead to recoveries that are characterised by a similar average change in the ratio of Apoptosis to Division Genes; there was a larger variation in the response to surgery. This data spread, as seen in Figure 6.25, translates into chemotherapy being a consistent therapy, while surgery is a "hit or miss" therapy that can on occasion cure the tumour. Surgery, when successful, was accompanied by the recovery of genotypes with active tumour suppression, leading to a better outcome than chemotherapy. From these simulations, it is clear that the best outcomes are mainly due to the recovery of tumour suppression. If tumour suppression is not recovered, the treatment fails.

Thus, targeting chromosomally unstable cells may be an important part of future cancer therapies. It has been suggested that chromosomal instability may also play a role later in generating the

genetic diversity required for cancer cells to survive the trials of invasion and metastasis. It is important to note that heterogeneity is often seen as being detrimental in the clinical setting (Gerlinger,et al., 2012). Only through the tracing of clear evolutionary pathways will it be possible to understand the different roles that these complex mutations have throughout the process of carcinogenesis and thus help us to develop better treatments.

In sum, our model in exploring the evolutionary pathway of cancer clones in tumour development provides new insights into the interplay between aneuploidy and tumour therapies. Future work will need to build on such models to bring them closer to reality; to study the role of aneuploidy on more advanced kinds of tumours, and to simulate other kinds of cancer treatments. Future work will be needed to assess how scrambling of the genome may combine with missegregation events to drive the evolution of chromosomes that have specific complements of genes.

7. Preliminary Experimental Data

7.1. Introduction

The hypothesis tested in this work is that an integrated computational model of chromosome missegregation during cell division provides an effective approach to assess the role of aneuploidy in the initiation and evolution of cancer. To test this theory, a computational model has been developed and implemented with chromosome missegregation as the main drive for the generation of novel genotypes in the absence of chromosome recombination. For the creation of this model, experimental evidence was initially gathered. Once complete, *in silico* experiments with the computational model were carried out, suggesting that chromosome missegregation has a key role in shaping the development of cancer. The computational model suggests the existence of two kinds of tumour development: one that is chromosomally stable (Gene Configuration B) and one that is chromosomally unstable (Gene Configuration C). From the analysis of the previous chapters, it was found that the chromosomally unstable genetic configuration is more aggressive and has a generally worse prognosis under surgery and chemotherapy than the chromosomally stable one.

Although the main method used to measure the effectiveness of the tool developed in this work is to compare with published results of independent experiments, this can be supplemented with biological experiments performed specifically for the computational model. It was decided to carry out biological experiments, informed by the computational model, with real cancer cell lines and tissues. The objective was to provide proof of principle of how this computational model can help biologists design experiments and how data from those experiments can be fed back into the computational model.

7.2. Experiments for the Analysis of Ploidy

Chromosome missegregation may be a key moulding force in real cancer systems (Chin et al., 2004); however, to use the computational model presented in this work in biological settings, there is an experimental gap that needs to be addressed: The development of tools for measuring ploidy. The computational model developed in this work can be used to formulate specific questions about ploidy, the number of homologous sets of chromosomes in a biological cell. In humans, like in our computational model, cells are usually diploid, containing two copies of the same chromosome. Because the quantity of nuclear DNA can change during an organism's lifetime by means of replication, poly-ploidization, gain or deletion of bases, such changes may lead to an abnormal DNA chromosome content that could lead to cancer (Holland & Cleveland, 2009). *In vitro* experiments require tools for a quantitative analysis of local and global ploidy-related characteristics that are just beginning to be developed. During this work, biological experiments that both informed and were informed by the computational model were carried out. Some of the data and principles obtained from these experiments were incorporated in the computational model in the process of model refinement.

7.3. Immunohistochemical Staining of ERM

7.3.1. Objective

During the initial development of the computational model, a gene that regulates chromosome segregation was incorporated. This Chromosome Segregation Gene was based on recent evidence suggesting that the ERM gene family; consisting of the EZRIN, RADIXIN and MOESIN genes, might have such a behaviour in the context of real tumours. In a ground-breaking publication, Patricia Kunda, a Postdoctoral researcher at the Baum Lab at the time when these experiments were carried out, discovered that one of the ERM genes, MOESIN, controls cortical rigidity, cell

rounding, and spindle morphogenesis during mitosis (Kunda et al., 2008), The more this gene is expressed, the more likely it is to help carry out a faithful cell division segregation (Hunter, 2004). If this gene is missing, however, aberrations during cell division may occur (Curto & McClatchey, 2004).

The hypothesis derived from Dr. Kunda's work is that tumours that have low levels of ERM proteins may be linked to chromosomal instability and aneuploidy. To test this theory it was decided that a quantification of ERM expression in tumour samples would provide a starting point for the validation of this theory. These initial results would then be incorporated in our computational model in as the dynamics of the Chromosome Segregation Regulatory Gene.

7.3.2. Setup

In an initial test of the method proposed, selected human cancer tissues was collected. We used the technique of immunohistochemical staining on human graded cancer tissues for the purpose of testing this hypothesis. Breast cancer samples were obtained from the Wolfson Institute for Biomedical Research, University College London, on collaboration with Marco Loddo (a PhD student at the time of writing). We stained six samples, from tumours whose ploidy had been previously analysed, in order to detect the level of ERM proteins.

The immunostaining protocol was developed through a series of preliminary experiments, in which we found the ideal concentrations for our novel P-ERM antibody. The procedure is as follows:

1. Fix and embed the tissue.
2. Cut and mount a tissue section.

3. Remove Paraffin and rehydrate the tissue section.

4. Wash away the lipids with TRIS and Triton X-100 for 5 minutes.

5. Blocking Step: Block in specific sites with 5% BSA mixed in PBS.

6. Incubate Primary Antibody (P-ERM 30 µl with a dilution of 1/100 TBS with 1% BSA).

7. Incubate overnight at 4°C.

8. Wash 3 times with TRIS and Triton X-100 for 5 minutes each.

9. Incubate Secondary Antibody (P.HST3 30 µl with a dilution of 1/200 TBS with 1% BSA).

10. Wash 2 times was TRIS and Triton X-100 for 5 minutes each.

11. Incubate overnight at 4°C.

12. Wash with TRIS and Triton X-100 for 5 minutes.

13. Fix with formaldehyde.

14. Dehydrate and stabilize with mounting medium.

15. Fix sample in a 3mm slide and label.

16. View the staining under a microscope.

For the experiments, we used Custom Anti-Peptide Polyclonal Antibodies from Eurogentec, rabbit variation (http://www.eurogentec.com/).

A
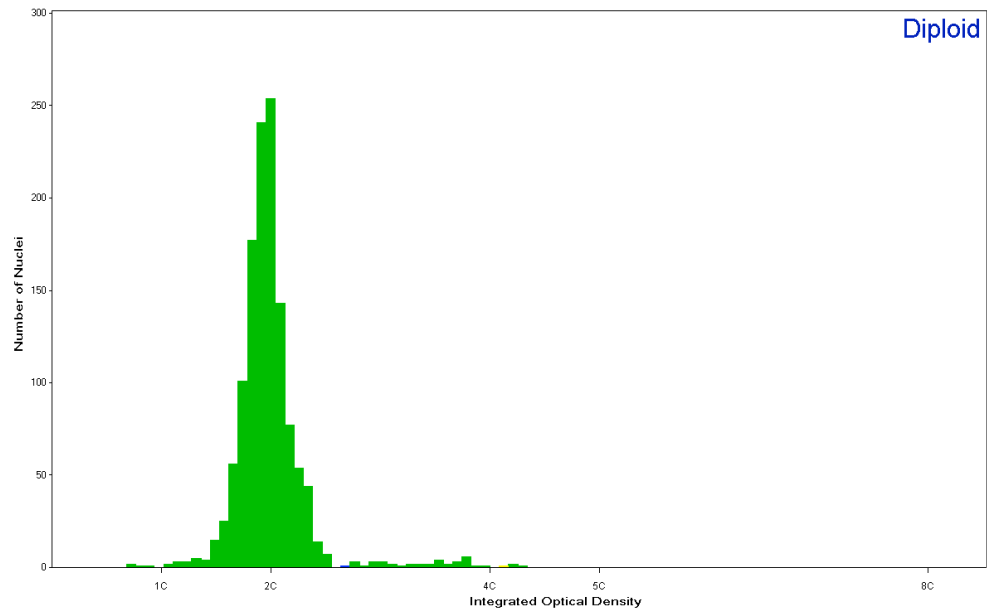
B.
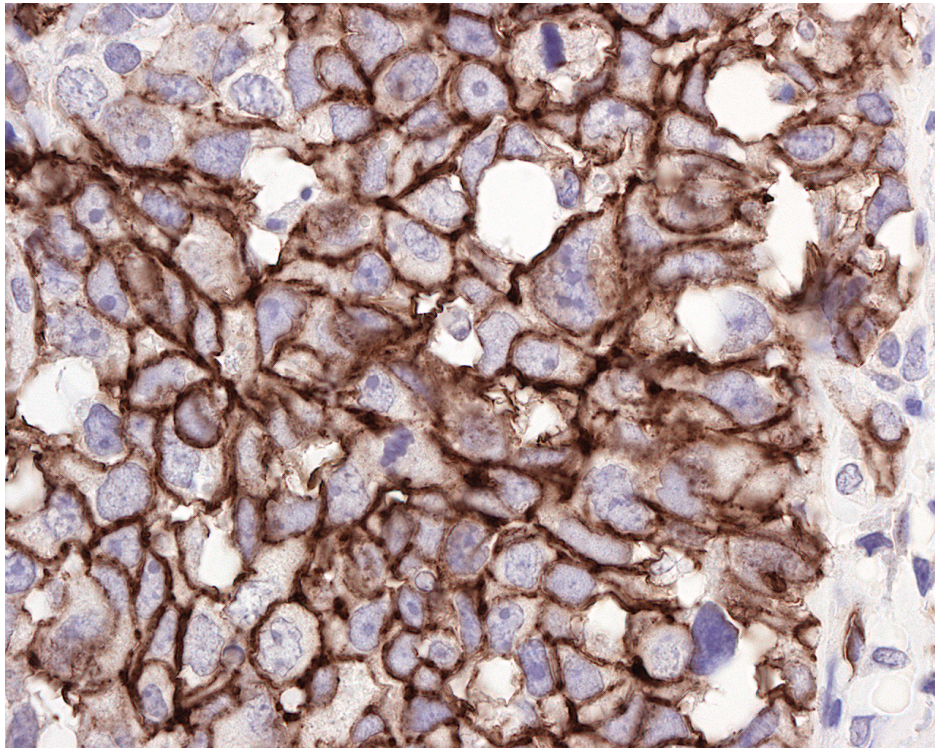
Figure 7.1- Immunohistochemical staining on a diploid human breast cancer tissue. A. The histogram describes the sample as Diploid according to measurements correlating the Integrated Optical Density with ploidy. B. The dark staining in this initial experiment suggests a high level of ERM protein expression, which may be implicated in faithful chromosome segregation.

A.
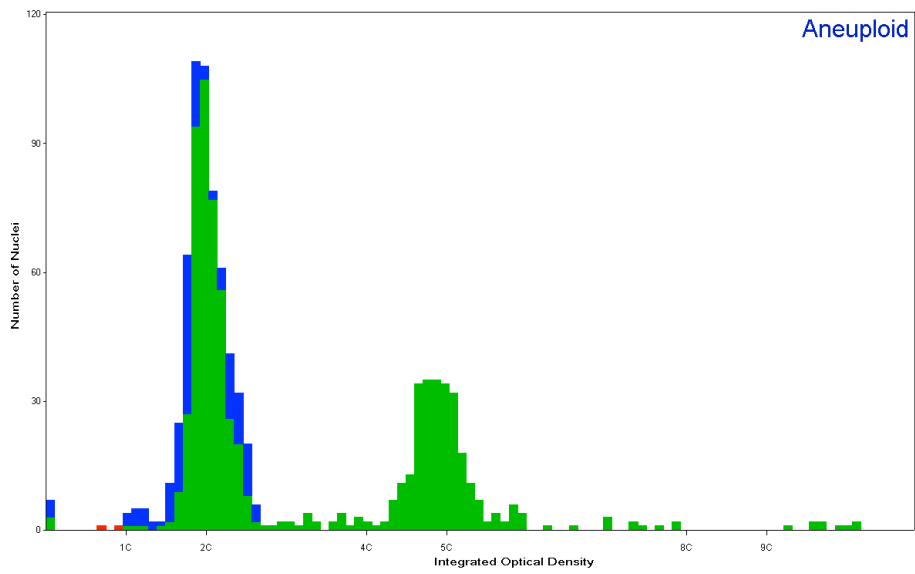
B.
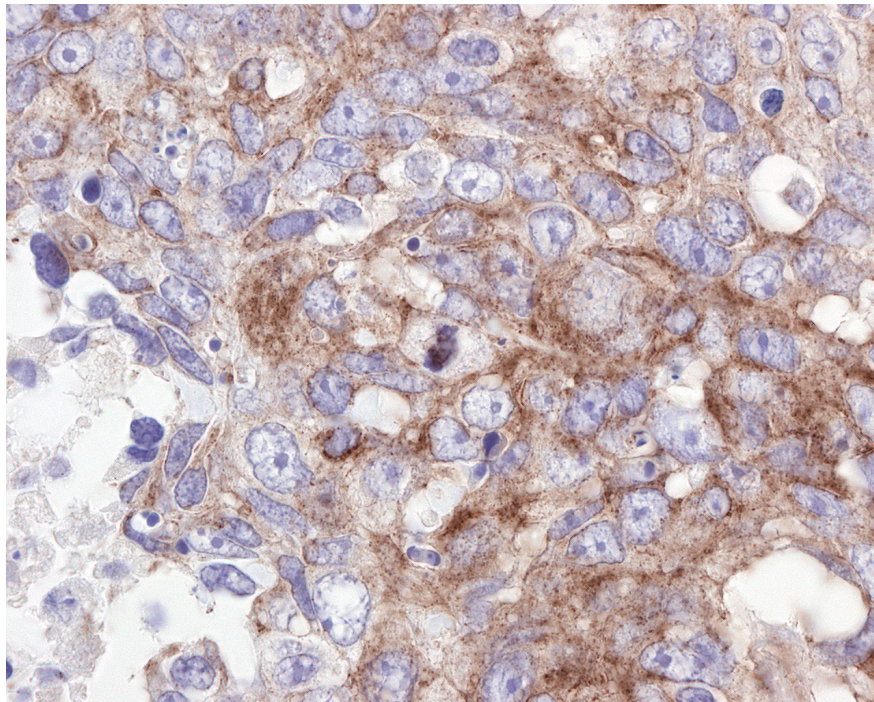
Figure 7.2- Immunohistochemical staining on a diploid human breast cancer tissue. A. The histogram describes the sample as Aneuploid according to measurements correlating the Integrated Optical Density with ploidy. B. The light staining in this initial experiment suggests a low level of ERM protein expression, compared to that of Figure 7.1.

7.3.3. Results

A considerable number of samples need to be analysed in order to obtain a quantitative measure of the real effect that this gene might have on chromosome missegregation. However, this initial test suggests an important result: high levels of ERM expression may be associated with diploidy (Figure 7.1), while low levels of may be associated with aneuploidy (Figure 7.2). These dynamics were incorporated into our model as the behaviour of Chromosome Segregation Regulation Genes, such that a lower expression would translate in less fidelity during chromosome segregation and a higher expression in higher fidelity (Chapter 4.4.2). It is important to note that, in some samples, relatively high and low levels of ERM were found localized in certain parts of the tumour in an uneven distribution. This may be a result of tumour heterogeneity, as results from our computational model highlight. Current research carried out by Gigna Patel, at the Baum lab, continues this investigation.

7.4. DNA Ploidy Tests in HeLa Cell Line

7.4.1. Objective

The computational model makes testable predictions regarding the relationship between the ploidy of a tumour in chromosomally stable/unstable configurations and the aggressiveness of the tumour. The computational model suggests that the more chromosomally unstable the tumour, the more aggressive it could be. It also suggests that the ploidy changes with time. To test these results, the creation of tools for the reliable measurements of ploidy is needed. It is possible to measure ploidy with Flow Cytometry techniques. Because cell cytometry works by suspending cells in a stream of fluid and passing them one by one through a laser-based detection apparatus, cells that initially were part of a specific colony (possibly clonal), are mixed with cells from other

colonies. The use of this procedure can give an accurate ploidy assessment, but all the spatial information regarding heterogeneity across the sample (such as clone colonies and regions of high diversity) is lost. This is because taking high-resolution images of the sample can preserve spatial information; and then it is possible to use image cytometry techniques to investigate the sample's ploidy. Although ploidy cannot be quantified directly by image cytometry techniques, such as correlating the Optical Integrated Density within the nuclei of the cells in the sample with ploidy, it is possible to qualitatively identify DNA outside the euploid regions as abnormal (or aneuploid) (Mendelsohn et al., 1969). An initial test using the software CellProfiler was carried out in order to determine the ploidy of a control laboratory sample of HeLa cancer cells. Once we have obtained a proof of principle for this technique, it is possible to then apply this analysis to more specific ploidy experiments.

7.4.2. Setup

High-resolution images were acquired from HeLa cell line samples. Once categorized, the images were then processed using CellProfiler, a free, open-source cell image analysis software package (Vokes & Carpenter, 2008). CellProfiler was used for the detection, segmentation and quantification of an image of scattered HeLa cells and their stained nuclei (Carpenter et al., 2006). The following CellProfiler segmentation values, determined through preliminary experiments, were used:

- Threshold of the image: 0.171
- It was calculated that, for our samples, 12.7% of the image consist of nuclei (one important parameter that affects cell profiler's accuracy. A parameter that will have to be adjusted depending on how sparse cells are plated).
- Smoothing Filter size: 4.7
- Maxima suppression size: 5

- Don't merge identified close objects.
- The typical HeLa nucleus is approximately 23 pixels wide.

Once the image has been segmented (nuclei and cells outlined), the intensity of the nuclei against the background can be plotted in a histogram. It is then possible to estimate the levels of Ploidy on a sample through image Cytometry techniques (Mendelsohn et al., 1969). After obtaining a measure of the "Integrated intensity" per cell, per image, the information is stored in the form of a MySQL database. This vast amount of information can then be sorted, classified and represented as a histogram of the distribution of Integrated Intensity by means of the data exploration software CPAnalyst, a complement of the CellProfiler software that is able to handle biological data in a MySQL database format.

Before a data histogram can be interpreted as the DNA content of the sample, a "c scale" should be determined. A "c scale" is a rescaling of the Integrated Density Values obtained through CellProfiler. This can be achieved by measuring the Integrated Density Values of cells with known DNA content. The Integrated Density Values of these normal, diploid cells are then transformed to a reference unit scale based on the number of chromosomes in cells, or "c" scale, which peaks at a value of "2c" in a histogram of a sample of diploid cells. The value "2c" obtained in this histogram then represents the mean nuclear DNA content of cells from a diploid population in G0/G1 cell cycle phase ("4c" for G2 Mitotic cells) (Haroskea et al., 1998).

7.4.3. Results

Once the "c" scale was introduced by identifying the diploid peak on the histogram (Figure 7.3.A) of a diploid sample (Figure 7.3.B), the rest of the data obtained was then assessed as either having a ploidy multiple of c, or as aneuploid if more peaks are observed (Girouda et al., 1998), as can be seen in Figure 7.4.A. for an aneuploid HeLa sample (Figure 7.4.B). To implement this

214

technique to sort cells according to their stage in cell cycle in high throughput data, machine-learning methods for the correct identification of phenotypes could be implemented through the same CPAnalyst software.

This technique can also be corroborated by means of a Flow Cytometer analysis: a technique for counting and examining microscopic particles, such as cells and chromosomes, by suspending them in a stream of fluid and passing them by an laser detection apparatus. It is important to use Image Cytometry together with Flow Cytometry analysis to obtain a reliable global measure of aneuploidy, as well as local measures that may give some clues as to the identity of the viable karyotypes of aneuploid cells. Together these methods can shed light on the evolutionary dynamics of aneuploidy and chromosome missegregation at different stages of the cancer evolution, and will be addressed as Future Work in the next chapter.

A.
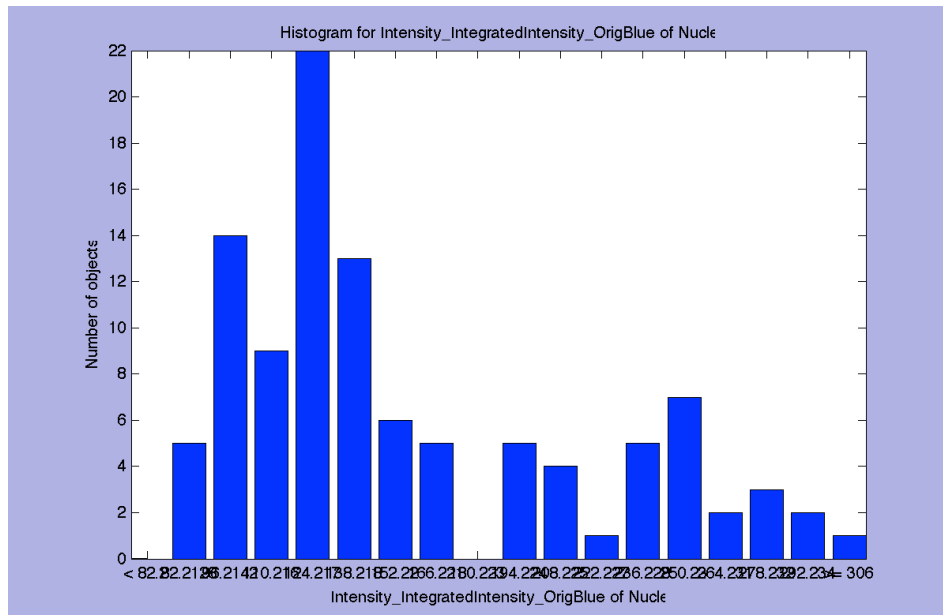
B.
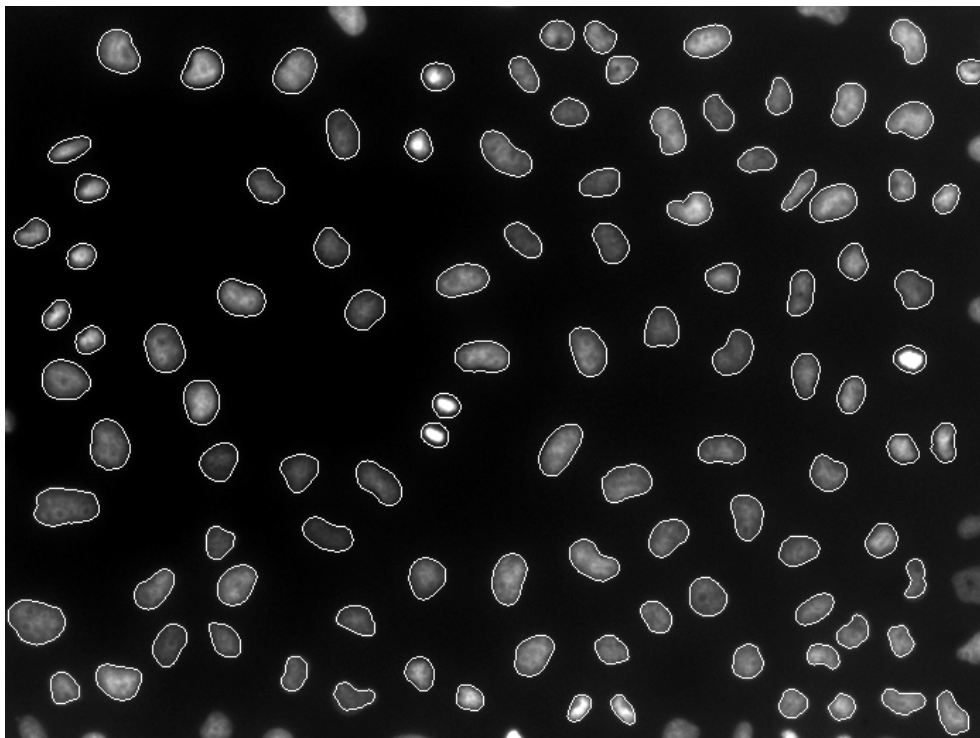
Figure 7.3- Ploidy Analysis of a diploid sample of cancer HeLa cells using the software Cell Profiler. A. The histogram correlates Optical Integrated Intensity with Ploidy. DNA content seems to be divided into two separate groups, hinting at a high level of diploid cells and a lower level of tetraploid cells (presumably cells in mitosis). B. The sample, segmented by Cell Profiler.

216

A.
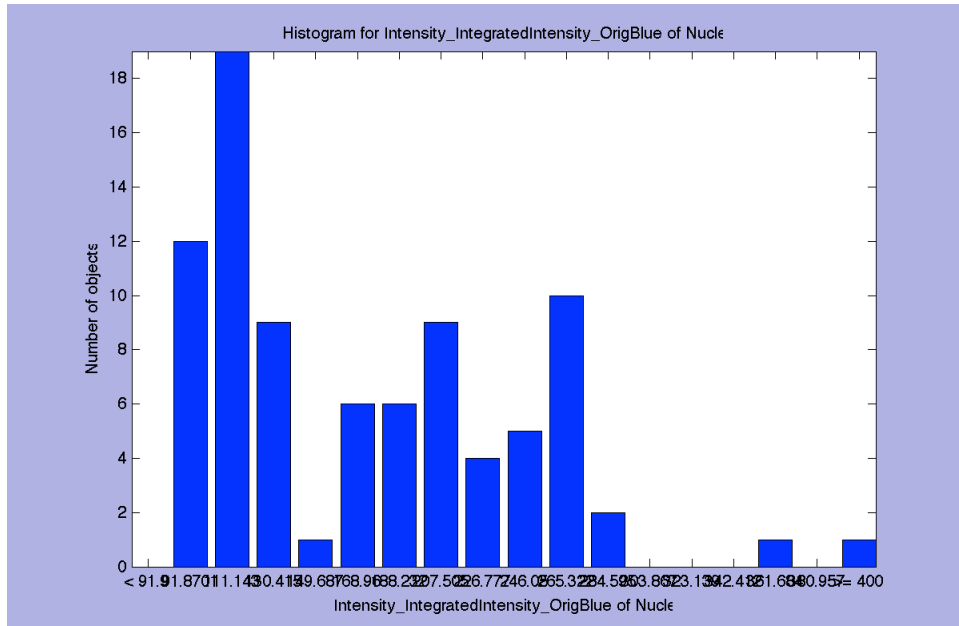
B.
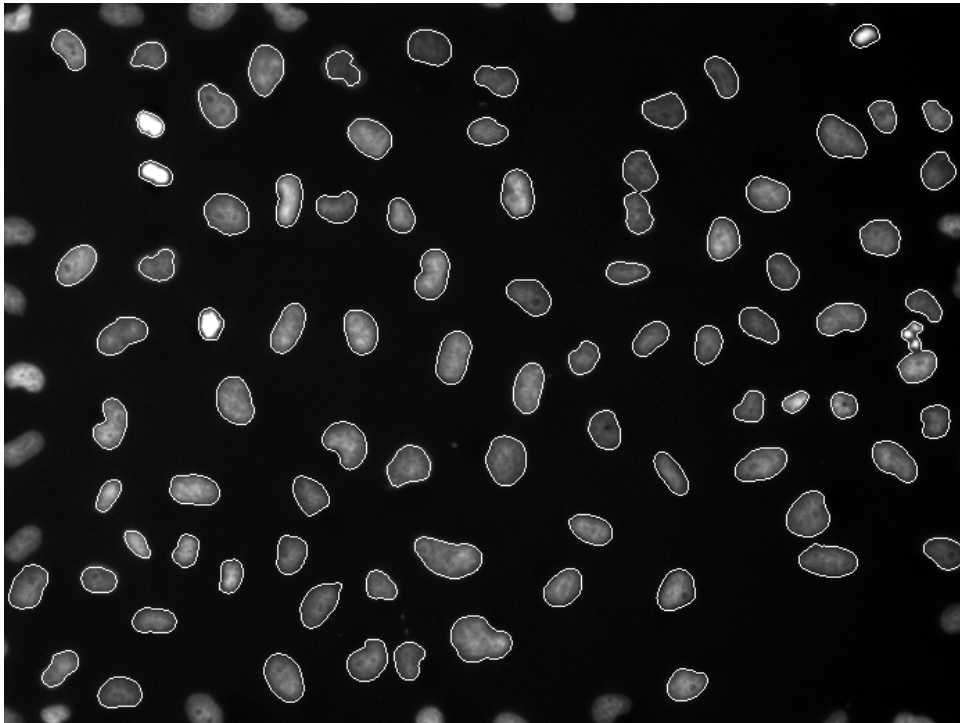
Figure 7.4- Ploidy Analysis of a previously measured diploid sample of cancer HeLa cells using the software Cell Profiler. A. The histogram correlates Optical Integrated Intensity with Ploidy. In the histogram, the occurrence of several peaks in the distribution is an indication of aneuploidy in the sample. B. The sample, segmented by Cell Profiler.

7.5. Inducing Aneuploidy in Cell Lines

7.5.1. Objective

To bring the computational model closer to reality, measurements of the properties of aneuploid cells are needed. It is currently unknown, for instance, whether aneuploidy is detrimental, neutral or beneficial for division/death rate. For these kinds of measurements, the controlled creation of aneuploid cells is needed. It may be possible to use a drug to increase the likelihood of generating aneuploid cells. In the literature one such drug exists for the inhibition of Aurora Kinases (Scharer et al., 2008). Aurora kinases regulate cell cycle transit from G2 through cytokinesis. A new drug, Aurora Inhibitor, has been proven to scatter the mitotic spindles of dividing cells. In this experiment, we seek to test whether this drug can be used to de-stabilize the mitotic apparatus of cells and generate defects in cell division.

7.5.2. Setup

In this experiment, we seek to induce aneuploidy in two cell lines: HeLa and RPE1. It was decided to initially test for best time for Aurora Inhibitor to induce aneuploidy. Cells were exposed to the drug for 0, 5, 8 and 24 hours. Then the cells were stained for DNA (DAPI, blue), cytoskeleton (TRITC, red) and cellular cortex (FITC, green).

A low cell concentration (50 cells per well) was used, in order to isolate aneuploid cells that could be used later in a follow up experiment (ideally allowing them to grow and measuring the growth rate of low DNA vs high DNA aneuploidy). This number was determined through a simple plating experiment with different cell concentrations.

The protocol used were developed with the help of Marina Fedorova, the lab assistant at the Baum Lab at the time of the experiments; based on a previous protocols used in the Baum Lab. The protocol for the Aurora Inhibitor was a novel design, and was tested on a series of previous experiments to determine the optimal drug concentration for the generation of aneuploidy. For this experiment the following protocols, in order of execution, were used:

a)   Thawing Cells

1. Pre-heat medium to 37ºC (Do not use drug containing media for the first few days after thawing).
2. Remove cells from liquid nitrogen and defrost.
3. Place cells in a 15ml tube and add 10ml of warm medium.
4. Centrifuge at 1000 rpm for 4 minutes.
5. Remove supernatant and re-suspend in 10ml medium
6. Plate cells in a culture dish.

b)   Splitting Cells

1. Pre-heat media and trypsin to 37ºC
2. Carefully aspirate the old medium from cells.
3. Add 10ml PBS, rinse briefly and aspirate to remove.
4. Add 1.5ml of trypsin and incubate at 37ºC for 2-3 minutes.
5. Harvest cells by adding 8.5ml of fresh medium.
6. Pipette up and down to remove clumps (but avoid bubbles).
7. Dilute cells appropriately in a new dish with fresh medium. (e.g. for a 1:20 split add 0.5ml cells to 9.5ml medium)

c)   Changing The Cell Media

1. Carefully aspirate the old media from the cells

2. Add 10 ml of PBS.

3. Rinse briefly.

4. Aspirate to remove PBS.

5. Add 10ml of fresh medium to the cells.

d)   Create new media for cells

For Hela Cells (Complete DMEM)

	500 ml	250 ml
1 x DMEM (+4.5% G/L/Glucose)	445	222.5
10% Fetal Bovine Serum	50	25
1% Penicillin-Streptomycin	5	2.5

For RPE-1 Cells (DMEM F-12)

	500 ml	250 ml
1 x DMEM/F12 + Glutamax	428	214
10% Fetal Bovine Serum	50	25
1% Penicillin-Streptomycin	5	2.5
7.5% Sodium Bicarbonate	17	8.5

e)   Diluting Cells

1. Pre-heat media and trypsin to 37ºC.

2. Carefully aspirate the old medium from cells.

3. Add 10ml PBS, rinse briefly and aspirate to remove.

4. Add 1ml trypsin and incubate at 37ºC for 3 minutes.

5. Harvest cells by adding 9ml fresh medium –pipette up and down to remove clumps (but avoid bubbles)

6. Count cells and dilute accordingly (50 cells per well)

f)   *Prepare Dishes for Cell Plating*

1. Add 700μl of Fibronectin 10μg/ml per well (100 times dilution in PBS from 1mg/ml stock)
2. Place dishes in incubator at 37º for 1 hour
3. Suck out Fibronectin
4. Plate cells

g)   *Adding Aurora Inhibitor Drug*

During this time, a novel protocol was designed and tested through previous experiments to obtain the adequate concentration of the drug for our samples. The protocol for the addition of this novel drug, developed with the help of Marina Fedorova, is as follows:

1. Prepare Aurora Inhibitor 10mM (from 2M stock)
2. Aspirate Media.
3. Add fresh Media with drug diluted to 2μM (1:5000).
4. Incubate for 6 hours (or the time it requires for treatment).
5. Aspirate Media and add 4% Formaldehyde (fresh).
6. Incubate for 10 to 20 minutes at room temperature.
7. Aspirate Formaldehyde.
8. Wash 3 times with PBS.
9. Incubate or Para-film dish and store in a refrigerator.

h)   *Cell Fixing*

1. Prepare 4% Formaldehyde (4 times dilution from 16% solution) + PBS.
2. Aspirate media.
3. Add (200μl) of a mix of 4% Formaldehyde diluted in PBS.

4. Leave for 20 minutes at room temperature.

5. Aspire Formaldehyde mix.

6. Wash wells 3 times with PBS.

7. If necessary, permeabilize, block and add antibodies.

8. Cover with Para-film and Foil.

i)   *Permeabilization and Blocking*

1. Suck media from wells

2. Add 20µl of 0.2% Triton 100 in PBS (as detergent mix) per well

3. Leave for 5 minutes at room temperature.

4. Aspirate Triton mix and add blocking Solution (5% BSA in PBS) (or 5% BSA in PBS + 3% FBS in PBS).

5. Leave 30 minutes to 1 hour at room temperature.

6. Aspirate blocking solution and add antibodies (this blocks non-specific sites).

j)   *Create Staining Antibody Solution*

1. Prepare antibodies as follows:

2. Create 1% BSA from 5% BSA solution (dilute 5 times in PBS).

     i.   DAPI-   1/1000 concentration from 1mg/ml

     ii. PH-TRITC   1/1600 concentration from 0.2mg/ml

     iii. TUBULIN- FITC   1/400 concentration

3. Add antibodies and stain for 1 hour at room temperature using (20µl) per well.

4. Wash 3 times with PBS

5. Store in (70µl) of Triton 0.2% in PBS (to prevent bacteria from growing) or in PBS 0.1%NaN$_3$

k)   *Staining in Cover Slip for analysis on a Confocal Microscope*

1. Add antibodies and stain for 1 hour at room temperature using 50-100µl per well.

2. Wash 3 times with PBS.

3. Put 300μl PBS in each well.

4. Remove off plastic from the glass (make sure there is no glue left).

5. Use FluourSave reagent (deliver 10μl by drops).

6. Place Cover Slip on top and cover in foil.

7. Store in a refrigerator overnight or in an incubator at 37º for 1 hour.

8. View in Confocal Microscope.

l)   *Freezing Cells for further experiments*

1. Prepare freezing Medium: <u>10ml</u>

     a.   Complete DMEM   7ml

     b.   + 20% FBS   2ml

     c.   + 10% DMSO   1ml

2. Label freezing vials with cell line, passage number, name and date.

3. Harvest cells into a 15ml tube.

4. Centrifuge at 1000 rpm for 4 minutes.

5. Remove supernatant and re-suspend in the freezing medium.

6. Quickly aliquot 1ml into each vial.

7. Quickly place in a cryobox and then in a -80ºC freezer (checking that there is enough iso-propanol in the cryobox).

8. Once cells are frozen, cells can be transferred to a liquid nitrogen tank and stored.
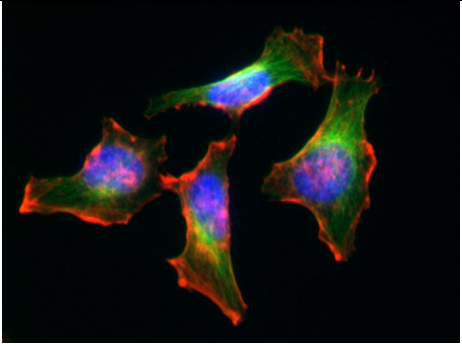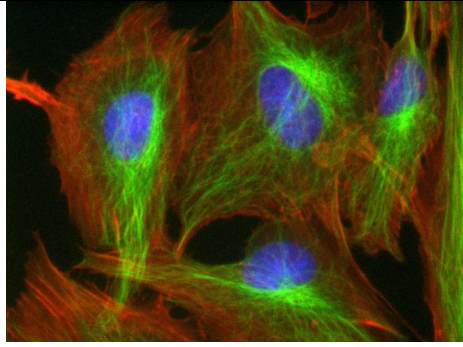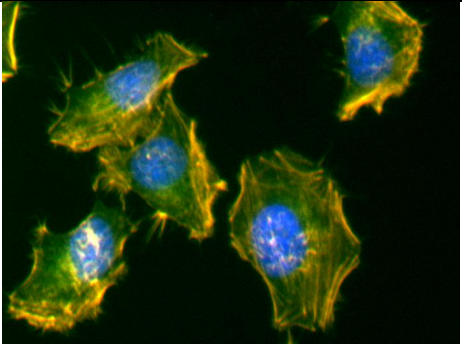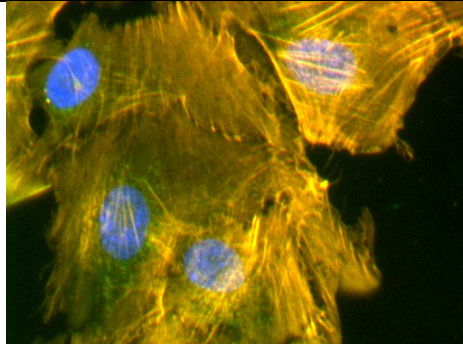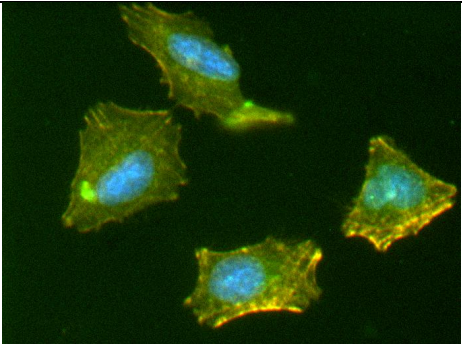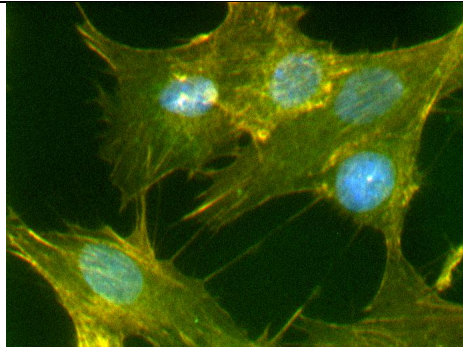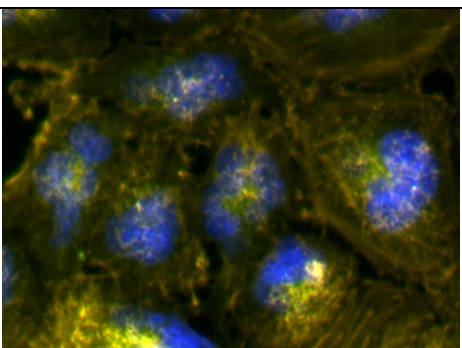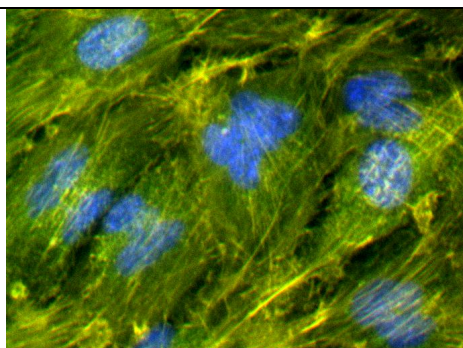
Drug	i. HeLa	ii. RPE1
A. No drug		
B. 5 hours		
C. 8 hours		
D. 24 hours		

Figure 7.5- Aurora Inhibitor was applied to HeLa and RPE1 cells for the duration of 0 (no drug), 5, 8 and 24 hours. Representative images from each experiment are shown. A. Without drugs, cells have their cortex (red) intact. B. After 5 hours of drug, the cells begin to have problems with their cortex

and cytoskeleton (green). C. After 8 hours of drug, the DNA content (blue) of some of the cells increases, preparing for cell division but failing to do so. After 24 hours of the drug, the ploidy of both cell lines has changed, and chromosome missegregation can be observed (as in the centre of D.i.).

7.5.3. Results

The experiments reveal a progression towards the cortical and mitotic destabilization for both cell lines, as seen in Figure 7.5. Progressively, cells that enter mitosis under Aurora inhibition either suffer from mitotic frustration (Figure 7.5.C) and become tetraploid, or start having asymmetrical divisions (Figure 7.5.D). This experiment suggests that Aurora Inhibitor, as used in the protocol presented in this work, can be used to generate aneuploid cells. These aneuploid cells can be used to validate some of the model predictions, such as the interplay between of aneuploidy with and cancer treatments (as described in Chapter 6), and make model refinements, as will be proposed in Future Work (Chapter 7).

7.6. Conclusions

In this chapter, key biological experiments that both inform and are informed by our computational model were carried out. In the first experiment, immunostaining experiments with the ERM gene product suggest that over expression of this gene family correlates with diploidy in cancer tissues, while the lack of activity may correlate with aneuploid samples. These initial experiments became the basis for the dynamics of our theorised Segregation Regulatory Gene (Chapter 4.4.2). Further experiments are being carried out at the Baum lab to quantify the behaviour of this gene. Those results can then be incorporated into the computational model, to bring it closer to reality.

In section 7.4, we noted the lack of tools to measure the kind of results that experiments informed by our computational model would yield. To bridge this gap, a technique for measuring the

Ploidy via Image cytometry techniques was proposed. The principles were tested on a sample of HeLa cells, resulting in the detection of diploid and aneuploid samples. With this method developed, it is possible to preserve important information regarding the space, such as the kind of colonies that are formed in a sample (such as clonal with low missegregation or heterogeneous with high missegregation). However, for an accurate measure of ploidy, it is recommended to use this technique in tandem with a flow cytometer since a scale for DNA comparison has to be first established.

Finally, we noted that aneuploid cells need to be created in order to carry out the kind of experiments that could be informed by our computational model. For this, we developed a protocol for generating aneuploid cells in controlled conditions. We used Aurora Inhibitor on the cell lines HeLa and RPE1 to successfully induce chromosome missegregation and generate aneuploid cells that can be used in future work. All of these experiments taken together are the first step towards bridging the gap between theories and experiments regarding the role of chromosome missegregation of cancer. These experiments have provided valuable information that has been incorporated in our computational model, and they have been in turn been informed by results from the computational model.

8. Conclusions and Future Work

8.1. Summary and Objectives Revisited

The aim of this work has been to investigate whether an integrated computational model of chromosome missegregation during cell division provides an effective approach to assess the role of aneuploidy in the initiation and evolution of cancer. The first chapter of this work presented a brief introduction to the biological phenomenon of aneuploidy. The question of the role of chromosome missegregation in cancer was asked, and a computational model as an effective mean to address this question was proposed. In this chapter, the objectives and work plan that were carried out throughout this work were set. The main objectives, as presented in Chapter 1, were met as in the following chapters, as will be described.

Chapter 2 provided a review in the literature of the current ideas regarding the phenomenon of chromosome missegregation in the context of cancer. In the literature review, Objective 1, the investigation of possible pathways for cancer to originate through chromosome missegregation, or for the subsequent evolution to be affected by the phenomenon of aneuploidy was carried out. Following this, a review of some of the most representative cancer modelling techniques used in cancer research was also presented in Chapter 2. Through this, Objective 2, an investigation on which modelling paradigms would be better suited to model the biological phenomenon of aneuploidy, was met.

In Chapter 3 the research methodology for this work was presented.. From all the modelling techniques reviewed, it was determined that an agent-based model would be ideal to address this kind of biological question. The modelling paradigm that was used for the development of the

model, an incremental approach to software development, was also selected. Here, objective 3, an investigation on how computational models could be validated or corroborated was performed. We selected Webb's methodology for the assessment of each version of the model, as described in Chapter 3 and used in Chapter 4. Each new version of the model, assessed through Webb's methodology, built on what was learned in the previous iteration.

In Chapter 5, the first principles of chromosome missegregation were summarized and the relevant biological concepts were abstracted for computational implementation (Objective 4). The first version of the computational model addressed the modelling of homeostasis and aneuploidy. Once this model was assessed, a second version of the model incorporated models of genes and the behaviour they regulate. The final version of the model included a streamlined algorithm and more abstracted biology in favour of computational feasibility and relevance.

Chapter 6 presents an exploration of the key properties of the model through a series of *in silico* experiments. The model presented in this work suggests that, though evolution, novel genotypes that promote unregulated cellular proliferation can be reached. *In silico* experiments were designed to explore key hypothesis of aneuploidy during the progression of cancer. These *in silico* experiments revealed that the location of genes plays an important role in determining the system's behaviour and also suggest that chromosome missegregation plays a key role in shaping two types of evolution that match those reported in the literature: chromosomally stable and chromosomally unstable tumours (Swanton & Caldas, 2009).

Results from the model (Objective 6) confirm that:

- Chromosome missegregation, **under certain genetic conditions**, can lead to further genetic instability, which may be advantageous to cancer progression (B. A. A. Weaver et al., 2007).
- The levels of aneuploidy change at different stages of cancer progression, **being moulded by the distribution of genes across chromosomes. These genetic restrictions heavily influence** the subsequent tumour evolution (Jefford & Irminger-Finger, 2006).

Analysis tools for the extraction of information from the model were designed and presented in Chapter 6. These tools were used to identify two key emergent tumour initiation pathways (Objective 5): initiation through oncogene activation and activation through tumour suppression.

Chapter 6 introduced the simulation of cancer treatments using the model developed (Objective 7). It was decided to model the two more common cancer treatments: surgery and chemotherapy; and a combination of both. Results suggest that a good marker for a successful cancer therapy is whether cells with some tumour suppression capabilities are rescued.

Chapter 7 describes the development of biological tools and *in vitro* experiments carried out that are complementary to this work (Objective 8). An immunostaining technique used for the assessment of a candidate Chromosome Segregation Gene, and the initial results were described. A computational method using the software CellProfiler for the assessment of aneuploidy by image cytometry was also designed and tested. Finally, an *in vitro* experiment informed by the computational model presented in this work was carried out. A protocol was created to induce aneuploidy, and results are shown using two cell lines: HeLa and RPE1; providing proof of principle for *in vitro* experiments that could be carried out informed by results from the model.

8.2. Discussion of this Work in the Context of Cancer Research

In order to test the hypothesis presented in this work, a computational model was created. We developed a model of tissue homeostasis in which to study cancer evolution. Individual cells were modelled, each equipped with a genetically defined genome, as agents in a computational simulation. A collection of these cells makes up a tissue that initially exhibits homeostatic behaviour, as the result of balanced rates of cell proliferation and cell death. These were modelled as stochastic processes that are regulated at a genetic level, based upon the properties of known proto-oncogenes and tumour suppressor genes (Futreal et al., 2004). We made the key abstractions of a single gene regulating a specific behaviour, and that the impact of each gene is proportional to the number of copies of a given gene found in the genome of each cell.

Having established this model system, we then introduced genes that regulate the rate of division, apoptosis and chromosomes segregation. Each incremental refinement of the model was assessed using Webb's modelling guidelines (B. Webb, 2009). This assessment included comparing the model with the current literature, the feedback from molecular biologists at the Baum Lab and the incorporation of real data whenever possible. After several iterations, a final version, which enables us to test the role of evolving chromosomal instability in cancer development and treatment, was reached. In this model, we were able to isolate the effects of chromosome instability, tumour suppressor and oncogene activity and genetic linkage on cancer progression. Following Webb's methodology, in order to present evidence that the model is effective and useful, the key discoveries made with this computational model will be discussed and assessed in the context of cancer research.

8.2.1. The Organizing Principles of Chromosome Missegregation

Tumours have been recognised as aneuploid for over a century (Holland & Cleveland, 2009). In addition, recent genomic sequencing studies have revealed enormous heterogeneity within single tumours (Meyerson et al., 2010). Nevertheless, the role of chromosome missegregation in cancer development is still debated, and experiments testing the effects of perturbing rates of chromosome segregation on tumour formation in mice have yielded contradictory results (Schvartzman et al., 2010). Therefore, to better understand the roles of chromosomal instability in the evolution of tumours, we have decided to take a theoretical approach and have developed an agent-based model of whole chromosome missegregation during cell division, in which we can test the hypothesis that chromosome missegregation can help drive cancer initiation.

For this purpose, we focused on modelling individual cells and their genomes in a homeostatic tissue whose behaviour is determined by a balance of cell death and cell proliferation within the context of a constrained environment. When events of chromosome missegregation are introduced, however, the dynamics of the system change in such a way that new, interesting complex behaviours emerge, which can be used to shed some light on the basic principles of aneuploidy in tumourigenesis. In simulations, chromosome missegregation events generate novel genotypes that promote unregulated cellular proliferation and impaired cellular death can be reached, driving cancer development. Importantly, this analysis also revealed that the location of these genes across chromosomes plays a key role in determining the system's behaviour and in shaping the genetic structure of the tumour population (Schneider & Grosschedl, 2007). This is driven by the fact that the copy number of genes, which regulate the fidelity of chromosome segregation, can alter as the result of the missegregation of their host chromosome at cell division. This leads to differences in the rates of missegregation that evolve during the course of tumour development in a way that depend on genetic linkage with oncogenes and tumour suppressors. So,

for example, in the absence of direct selection for chromosomes based upon the presence of genes promoting or inhibiting cell proliferation, we observed a reproducible increase in the number of clones with a decreased rate of chromosome missegregation.

In our model we observe two distinct pathways for evolution towards oncogenesis that have a direct impact on the tumour's response to treatments. In the first case, dominant proliferating clones within the tumour exhibit a relatively stable state of aneuploidy. In the second, selection for the loss of the aneuploidy gene results in tumours that continually generate increasing levels of heterogeneity and ever-more malignant subclones. Future work will be needed to assess how scrambling of the genome may combine with missegregation events to drive the evolution of chromosomes that have specific complements of genes.

8.2.2. The Role of Chromosome Missegregation in Cancer Therapies

In this analysis we also explored the effects of three different types of simulated treatment in the model. This revealed interactions between the treatments and the gene distribution in each case. The reduced spread in gene distributions seen following chemotherapy was due to the selective killing of overproliferating cells, compared to surgery, which does not discriminate based upon genetic makeup. In both cases, however, the number of tumour suppressor genes remaining after treatment are a critical factor in determining the course of the relapse. Once the chemotherapy was applied and the number of cells with overproliferative genotypes reduced, the likelihood that the remaining cells are able to resist overproliferation by means of contact inhibition is increased if they retain an intact apoptosis regulator. If the remaining cells lack an intact tumour suppression gene at the time of therapy, the intervention can act as an evolutionary bottleneck to leave a population dominated by malignant cells. In addition, in simulations in which

chromosomal instability is high, cells that survive the treatment as the result of a slow rate of cell proliferation rapidly acquire extra copies of the oncogenes to reinitiate tumour formation.

If there are a number of genotypes that retain functional tumour suppressors at the time of therapeutic intervention, it is possible to recover less aggressive genotypes, leading to a better prognosis. If tumour suppressor function is compromised prior to treatment, the intervention can lead to an evolutionary bottleneck that selects for malignancy. From our model, it is clear that the best outcomes in treatment simulations are mainly due to the recovery of tumour suppression. If tumour suppression is not recovered, the treatment fails.

Targeting chromosomally unstable cells may be an important part of future cancer therapies. It has been suggested that chromosomal instability may also play a role later in generating the genetic diversity required for cancer cells to survive the trials of invasion and metastasis. Only through the tracing of clear evolutionary pathways will it be possible to understand the different roles that these complex mutations have throughout the process of carcinogenesis and thus help us to develop better treatments.

8.2.3. Assessment Methodology

Following Webb's methodology, we found that the model stands in Webb's seven dimensions as follows:

1. **Relevance**: The model is an abstraction of current biological knowledge. It was successfully be used to assess a new hypothesis and the behaviour that can occur under therapeutic scenarios.

2. **Level**: Changes in the genomes of individual cells give rise to different cell behaviours and emergent properties at the tissue level. This was important when analysing the genetic makeups of tumour populations and linking that to the outcome of therapies.

3. **Generality**: The time frame of this model is scalable. It could represent days in *in vitro* experiments (such as cultured cells), or months in *in vivo* studies (such as animal models) or clinical settings (in real patients). The tumours simulated, like the clinical ones were proven to be highly variable. As such, this model can be adapted to study many kinds of individual genetics and dynamics.

4. **Abstraction**: A general model of carcinogenesis, the model is specifically cellular with a solid basis on cell biology. Therapies were abstracted and implemented in such a way that results are biologically and clinically meaningful.

5. **Accuracy**: The abstractions of cancer therapies accurately represent the mechanisms relevant for biological and clinical settings.

6. **Match**: *In silico* experiments show that the behaviour obtained in the model matches the behaviour found in real cancers. The simulation of therapies shed light on clinical settings that put in context individual observations in the literature.

7. **Medium**: Because the computer uses a pseudo random number generator, which uses a seed number, individual simulations are reproducible. This gives us the unique situation in which it is possible to explore the outcome of different treatments on the exact same individual simulation. With the tools developed, it was possible to prove the simulation in ways that are not possible *in vitro*, and that can be used as a guide when devising biological experiments.

8.3. Self Evaluation

From the literature review in Chapter 2, the modelling methodologies discussed in Chapter 3 and the development of the model in Chapter 4, the number of things that could have been done differently is apparent. With the benefit of hindsight, there are small things that could have helped the model be faster, such as representing the number of chromosomes as an integer number within cells (instead of being variables for each kind of chromosome). The implementation of a different way of traversing the linked list where all the information is stored could have been a

more effective approach (and will be considered in a future implementation of the model). However, the overall selection of the modelling paradigm, the evaluation methodology and the implementation of the model worked well.

There are however big issues that can be improved. From the point of view of modelling, space could be explicitly modelled. Using a cellular automata-like display could be useful to identify the location of the colonies of clones, and regions of diversity within the tumour. While computationally intensive, that kind of model could be a source of the kind of information that pathologists and cell biologist would find easier to understand and integrate into experiments. It is also of interest to incorporate a realistic version of the environment into the model, since microenvironment selection could also cooperate with aneuploidy to promote tumour progression (Anderson, Math Med Biol, 2005) (Anderson, Weaver, Cummings, & Quaranta, 2006),

A problem that persisted through the modelling process and the phase of in silico experimentation was the lack of metrics for such a complex process. This led to biological features such as point mutations, chromosome breakings or fusions and fitness penalties or benefits due to the loss of key genetic machinery being left outside of the scope of this work. Although such biological features could be modelled, researchers should take care that they maintain a reasonable level of complexity so that the behaviour can still be analysed in a transparent way. Finally, the issues faced with computational power available and the limitations faced when designing and measuring biological experiments are the subject of current research.

8.4. Future Work

We constructed our model to be flexible enough so that it can be adapted to incorporate more realistic biological features and consider diverse scenarios. From this body of work, it became clear that there are different kinds of aneuploidy. Aneuploid cells can develop different genetic properties, such as chromosomal instability, accelerated proliferation and resistance to apoptosis. However, it is still unknown whether gaining or losing chromosomes affect the fitness of aneuploid cells. This question should be investigated and can be readily implemented in our model.

In our model, we can consider different kinds of cell fitness with respect to ploidy, such as:

- Fitness proportional to chromosome number (Figure 8.1.A).
- Fitness inversely proportional to chromosome number (Figure 8.1.B).
- A general increase in fitness for non-diploid cells (Figure 8.1.C).
- A general decrease in fitness for non-diploid cells (Figure 8.1.D).

Implementing these fitness scenarios in the model, results that reflect changes in key properties such as growth and genetic instability can be obtained. These results can then be validated in an *in vitro* experiment.

With the experimental the techniques developed in Chapter 7, it is possible to test for the growth rate associated with a kind of ploidy. For this experiment, Aurora inhibitor could be used for 24 hours on sparsely plated cells (as described in section 7.5). Then, taking measurements of the different ploidy levels and cell numbers with the method proposed in section 7.4, it is possible to measure the difference in growth of diploid against non-diploid cells. This can be done across different kinds of aneuploid cells, if cells are sparsely plated: 100 cells per well, would directly correlate with the model. Cells with different levels of ploidy can be left to grow their own

colonies (clonal or genetically heterogeneous, depending on the level of chromosome segregation regulation). If samples of different ploidies are left to grow and then are fixed at different time points (1 day, 2 days, 3 days and 4 days), the growth rate of the individual colonies can be measured.
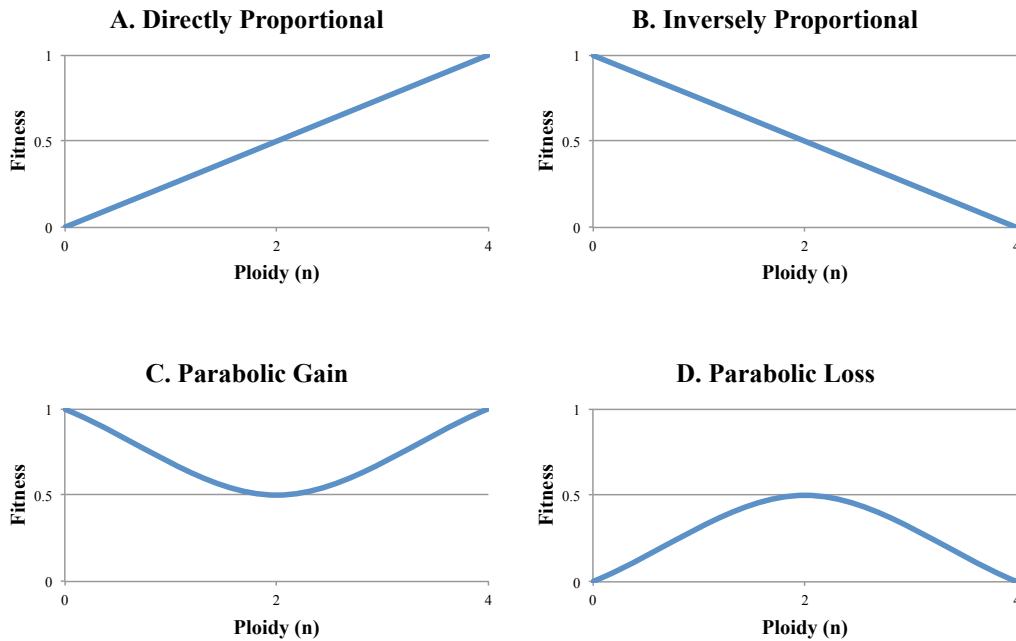
Figure 8.1- Possible fitness scenarios for Aneuploid cells. Depending on the level of Ploidy, n (see Chapter 7), different scenarios can be modelled. A. Fitness proportional to ploidy; B. Fitness inversely proportional to ploidy; C. A general increase in fitness for non-diploid cells; D. A general decrease in fitness for non-diploid cells.

These results can then be compared against model predictions and the results would be two:

1. It will be determined which kind of fitness curve (or curves) aneuploid cells have, as seen in Figure 8.1.
2. Realistic growth rates can then be directly incorporated in the model, thus bringing the model closer to reality.

Also, it is of interest to consider the real locations of known cancer genes across the genome. As more and more genes and the linkage between them are being mapped and understood through projects like the cancer genome atlas (http://cancergenome.nih.gov/), it will be probably feasible in the not too distant future to incorporate this information in a future version of the model. Future implementations could also make use of the continuously growing bio-informatics data; specifically information on where genes are on the chromosomes. These refinements could yield more realistic behaviour and may from a better bridge between basic research and the clinic.

8.5. Final Remarks

This work has described how and why an integrated computational model of chromosome missegregation during cell division does provide an effective approach to assess the role of aneuploidy in the initiation and evolution of cancer. Our model explores the evolutionary pathway of cancer clones in tumour development, and shows the interplay between aneuploidy and tumour therapies. With the analysis tools developed, and the *in silico* and *in vitro* experiments carried out, we have shed some light on the mechanism through which chromosome missegregation generates variation and how it could shape tumour development. We have met all the objectives that we set out to do and contributed with original research that, through publications (as seen in Chapter 1) has had an impact on the computer science community and the cancer research communities.

This interdisciplinary research has provided a tool that has the potential to help biologists and clinicians to ask questions that would have been near impossible to answer were it not for the transparency of the computational simulations. Future work will need to build on bringing the

model closer to reality; to study the role of aneuploidy on more advanced kinds of tumours, and to simulate other kinds of cancer treatments.

References

Abbott, R. G., Forrest, S., & Pienta, K. J. (2006). Simulating the hallmarks of cancer. *Artificial Life*, *12*(4), 617–634. doi:10.1162/artl.2006.12.4.617

Alarcón, T., & Page, K. M. (2006). Mathematical models of the VEGF receptor and its role in cancer therapy. *Journal of the Royal Society Interface*, *4*(13), 283–304. doi:10.1098/rsif.2006.0170

Amato, A., Lentini, L., Schillaci, T., Iovino, F., & Di Leonardo, A. (2009). RNAi mediated acute depletion of retinoblastoma protein (pRb) promotes aneuploidy in human primary cells via micronuclei formation. *BMC Cell Biology*, *10*, 79. doi:10.1186/1471-2121-10-79

Anand, P., Kunnumakara, A. B., Sundaram, C., Harikumar, K. B., Tharakan, S. T., Lai, O. S., et al. (2008). Cancer is a preventable disease that requires major lifestyle changes. *Pharmaceutical Research*, *25*(9), 2097–2116. doi:10.1007/s11095-008-9661-9

Anderson, A. R. A. (2005). A hybrid mathematical model of solid tumour invasion: the importance of cell adhesion. *Mathematical Medicine and Biology*, *22*(2), 163–186. doi:10.1093/imammb/dqi005

Anderson, A. R. A., & Chaplain, M. A. J. (2012). Continuous and discrete mathematical models of tumor-induced angiogenesis. *Bulletin of Mathematical Biology*, *60*(5), 857–899.

Anderson, A. R. A., Weaver, A. M., Cummings, P. T., & Quaranta, V. (2006). Tumor morphology and phenotypic evolution driven by selective pressure from the microenvironment. *Cell*, *127*(5), 905–915. doi:10.1016/j.cell.2006.09.042

Araujo, R. P., & McElwain, D. L. S. (2004). A history of the study of solid tumour growth: the contribution of mathematical modelling. *Bulletin of Mathematical Biology*, *66*(5), 1039–1091. doi:10.1016/j.bulm.2003.11.002

Austin, K., Gupta, M., Coats, S., Tulpule, A., Mostoslavsky, G., Balazs, A., et al. (2008). Mitotic spindle destabilization and genomic instability in Shwachman-Diamond syndrome. *Journal of Clinical Investigation*, *118*(4), 1511–1518.

Baker, D. J., Chen, J., & van Deursen, J. M. A. (2005). The mitotic checkpoint in cancer and aging: what have mice taught us? *Current Opinion in Cell Biology*, *17*(6), 583–589. doi:10.1016/j.ceb.2005.09.011

Baker, D. J., Jin, F., Jeganathan, K. B., & Deursen, J. M. V. (2009). Whole chromosome instability caused by bub1 insufficiency drives tumorigenesis through tumor suppressor gene loss of heterozygosity. *Cancer Cell*, *16*(6), 475–486. doi:10.1016/j.ccr.2009.10.023

Barabasi, A. (2005). Taming complexity. *Nature Physics*, *1*(2), 68–70.

Barabási, A.-L. (2003). *Linked*. Plume Books.

Basanta, D., & Deutsch, A. (2008). A game theoretical perspective on the somatic evolution of cancer. *Selected Topics in Cancer Modeling*, 1–16.

Basanta, D., Miodownik, M., & Baum, B. (2008). The evolution of robust development and homeostasis in artificial organisms. *PLoS Computational Biology*, *4*(3), e1000030. doi:10.1371/journal.pcbi.1000030

Bates, S. E., Amiri-Kordestani, L., & Giaccone, G. (2012). Drug development: portals of discovery. *Clinical Cancer Research*, *18*(1), 23–32. doi:10.1158/1078-0432.CCR-11-1001

Beck, K., Beedle, M., Van Bennekum, A., Cockburn, A., Cunningham, W., Fowler, M., et al. (2001). Manifesto for agile software development. *The Agile Alliance*, 2002–2004.

Bentley, K., Gerhardt, H., & Bates, P. A. (2008). Agent-based simulation of notch-mediated tip cell selection in angiogenic sprout initialisation. *Journal of Theoretical Biology*, *250*(1), 25–

36. doi:10.1016/j.jtbi.2007.09.015

Bentley, P. J. (2009). Methods for improving simulations of biological systems: systemic computation and fractal proteins. *Journal of the Royal Society Interface*, *6*(Suppl_4), S451–S466. doi:10.1098/rsif.2008.0505.focus

Berman, D. W. (2011). Apples to Apples: The Origin and Magnitude of Differences in Asbestos Cancer Risk Estimates Derived Using Varying Protocols. *Risk Analysis*, *31*(8), 1308–1326. doi:10.1111/j.1539-6924.2010.01581.x

Bissell, M., Rizki, A., & Mian, I. (2003). Tissue architecture: the ultimate regulator of breast epithelial function. *Current Opinion in Cell Biology*, *15*(6), 753–762.

Boehm, B. W. (1988). A spiral model of software development and enhancement. *Computer*, *21*(5), 61–72.

Brooks, R. A. (1990). Elephants don't play chess. *Robotics and Autonomous Systems*, *6*(1), 3–15.

Brunet, A. (2009). Cancer: When restriction is good. *Nature*, *458*(7239), 713–714.

Burton, A. C. (1966). Rate of growth of solid tumours as a problem of diffusion. *Growth*, *30*(2), 157–176.

Busenberg, S., & Mahaffy, J. (1985). Interaction of spatial diffusion and delays in models of genetic control by repression. *Journal of Mathematical Biology*, *22*(3), 313–333.

Byrne, H. M., Alarcon, T., Owen, M. R., Webb, S. D., & Maini, P. K. (2006). Modelling aspects of cancer dynamics: a review. *Philosophical Transactions of the Royal Society a: Mathematical, Physical and Engineering Sciences*, *364*(1843), 1563–1578. doi:10.1098/rsta.2006.1786

Cahill, D., Kinzler, K., Vogelstein, B., & Lengauer, C. (1999). Genetic instability and darwinian selection in tumours. *Trends in Biochemical Sciences*, *24*(12), M57–M60.

Carmona-Fontaine, C., Matthews, H., Kuriyama, S., Moreno, M., Dunn, G., Parsons, M., et al. (2008). Contact inhibition of locomotion *in vivo* controls neural crest directional migration.

Nature. 456.7224 (2008): 957-961.

Carpenter, A., Jones, T., Lamprecht, M., Clarke, C., Kang, I., Friman, O., et al. (2006). CellProfiler: image analysis software for identifying and quantifying cell phenotypes. *Genome Biology*, *7*(10), R100.

Chang, S., Khoo, C., & DePinho, R. A. (2001). Modeling chromosomal instability and epithelial carcinogenesis in the telomerase-deficient mouse. *Seminars in Cancer Biology*, *11*(3), 227–239. doi:10.1006/scbi.2000.0374

Chen, C., Nagl, S., & Clack, C. (2007). Specifying, detecting and analysing emergent behaviours in multi-level agent-based simulations. *Proceedings of the 2007 Summer Computer Simulation Conference*, 969–976.

Chen, C., Nagl, S., & Clack, C. (2008). A method for validating and discovering associations between multi-level emergent behaviours in agent-based simulations. *Agent and Multi-Agent Systems: Technologies and Applications*, 1–10.

Chin, K., De Solorzano, C. O., Knowles, D., Jones, A., Chou, W., Rodriguez, E. G., et al. (2004). In situ analyses of genome instability in breast cancer. *Nature Genetics*, *36*(9), 984–988. doi:10.1038/ng1409

Cimini, D. (2008). Merotelic kinetochore orientation, aneuploidy, and cancer. *Biochimica Et Biophysica Acta*, *1786*(1), 32–40. doi:10.1016/j.bbcan.2008.05.003

Cimini, D., & Degrassi, F. (2005). Aneuploidy: a matter of bad connections. *Trends in Cell Biology*, *15*(8), 442–451.

Collins, M. J., Napoli, I., Ribeiro, S., Roberts, S., & Lloyd, A. C. (2012). Loss of Rb cooperates with Ras to drive oncogenic growth in mammalian cells. *Current Biology*, *22*(19), 1765–1773. doi:10.1016/j.cub.2012.07.040

Crasta, K., Ganem, N. J., Dagher, R., Lantermann, A. B., Ivanova, E. V., Pan, Y., et al. (2012). DNA breaks and chromosome pulverization from errors in mitosis. *Nature*, *482*(7383), 53–

58. doi:10.1038/nature10802

Curto, M., & McClatchey, A. I. (2004). Ezrin...a metastatic detERMinant? *Cancer Cell*, *5*(2), 113–114.

Danos, V., & Laneve, C. (2004). Formal molecular biology. *Theoretical Computer Science*, *325*(1), 69–110. doi:10.1016/j.tcs.2004.03.065

Deakin, A. S. (1975). Model for the growth of a solid *in vitro* tumor. *Growth*, *39*(1), 159–165.

Dean, J. L., Mcclendon, A. K., Stengel, K. R., & Knudsen, E. S. (2010). Modeling the effect of the RB tumor suppressor on disease progression: dependence on oncogene network and cellular context. *Oncogene*, *29*(1), 68–80. doi:10.1038/onc.2009.313

Deisboeck T., Wang Z. , Macklin P. & Cristini V. (2011). Multiscale cancer modeling. Annual Review of Biomedical Engineering, Vol. 13: 127 -155

Denko, N. C., Giaccia, A. J., Stringer, J. R., & Stambrook, P. J. (1994). The human Ha-ras oncogene induces genomic instability in murine fibroblasts within one cell cycle. Proceedings of the National Academy of Sciences, 91(11), 5124-5128.

Ding, L., Getz, G., Wheeler, D., Mardis, E., McLellan, M., Cibulskis, K., et al. (2008). Somatic mutations affect key pathways in lung adenocarcinoma. *Nature*, *455*(7216), 1069–1075.

Dobles, M., Liberal, V., Scott, M. L., Benezra, R., & Sorger, P. K. (2000). Chromosome missegregation and apoptosis in mice lacking the mitotic checkpoint protein Mad2. *Cell*, *101*(6), 635–645.

Duesberg, P., & Rasnick, D. (2000). Aneuploidy, the somatic mutation that makes cancer a species of its own. *Cell Motility and the Cytoskeleton*, *47*(2), 81–107.

Düchting, W., & Dehl, G. (1980). A computer model for study of spatial and temporal tumor growth. *Proceedings of the Annual Symposium on Computer Application in Medical Care, 2*, 713.

Elzagheid, A., Korkeila, E., Bendardaf, R., Buhmeida, A., Heikkilä, S., Vaheri, A., et al. (2008).

Intense cytoplasmic ezrin immunoreactivity predicts poor survival in colorectal cancer. *Human Pathology*, *39*(12), 1737–1743.

Endy, D., & Brent, R. (2001). Modelling cellular behaviour. *Nature*, *409*(6818), 391–395.

Evan, G. I. (2006). Can't kick that oncogene habit. *Cancer Cell*, *10*(5), 345–347. doi:10.1016/j.ccr.2006.10.013

Feldser, D. M., Hackett, J. A., & Greider, C. W. (2003). Telomere dysfunction and the initiation of genome instability. *Nature Reviews Cancer*, *3*(8), 623–627. doi:10.1038/nrc1142

Felsher, D., & Bishop, J. (1999). Transient excess of MYC activity can elicit genomic instability and tumorigenesis. *Proceedings of the National Academy of Sciences of the United States of America*, *96*(7), 3940.

Fisher, J., Henzinger, T., Mateescu, M., & Piterman, N. (2008). Bounded asynchrony: Concurrency for modeling cell-cell interactions. *Formal Methods in Systems Biology*, 17–32.

Frumkin, D., Wasserstrom, A., Itzkovitz, S., Stern, T., Harmelin, A., Eilam, R., et al. (2008). Cell lineage analysis of a mouse tumor. *Cancer Research*, *68*(14), 5924–5931. doi:10.1158/0008-5472.CAN-07-6216

Furney, S., Higgins, D., Ouzounis, C., & López-Bigas, N. (2006). Structural and functional properties of genes involved in human cancer. *BMC Genomics*, *7*(1), 3.

Furuya, T., Uchiyama, T., Murakami, T., Adachi, A., Kawauchi, S., Oga, A., ... & Sasaki, K. (2000). Relationship between chromosomal instability and intratumoral regional DNA ploidy heterogeneity in primary gastric cancers. Clinical cancer research, 6(7), 2815-2820.

Futreal, P., Coin, L., Marshall, M., Down, T., Hubbard, T., Wooster, R., et al. (2004). A census of human cancer genes. *Nature Reviews Cancer*, *4*(3), 177–183. doi:10.1038/nrc1299

Gallaher, J., & Anderson, A. R. (2013). Evolution of intratumoral phenotypic heterogeneity: the role of trait inheritance. arXiv preprint arXiv:1305.0524.

Ganem, N. J., Godinho, S. A., & Pellman, D. (2009). A mechanism linking extra centrosomes to

chromosomal instability. *Nature*, *460*(7252), 278–282. doi:10.1038/nature08136

Gatenby, R. (2006). Commentary: Carcinogenesis as Darwinian evolution? Do the math! *International Journal of Epidemiology*, *35*(5), 1165.

Gatenby, R. A., Smallbone, K., Maini, P. K., Rose, F., Averill, J., Nagle, R. B., et al. (2007). Cellular adaptations to hypoxia and acidosis during somatic evolution of breast cancer. *British Journal of Cancer*, *97*(5), 646–653. doi:10.1038/sj.bjc.6603922

Gavrilov, L., & Gavrilova, N. (2002). Evolutionary theories of aging and longevity. *TheScientificWorldJOURNAL*, *2*, 339–356.

Gerisch, A., & Chaplain, M. A. J. (2008). Mathematical modelling of cancer cell invasion of tissue: Local and non-local models and the effect of adhesion. *Journal of Theoretical Biology*, *250*(4), 684–704. doi:10.1016/j.jtbi.2007.10.026

Gerlee, P., & Anderson, A. R. A. (2007). An evolutionary hybrid cellular automaton model of solid tumour growth. *Journal of Theoretical Biology*, *246*(4), 583–603. doi:10.1016/j.jtbi.2007.01.027

Gerlinger, M., Rowan, A. J., Horswell, S., Larkin, J., Endesfelder, D., Gronroos, E., et al. (2012). Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *The New England Journal of Medicine*, *366*(10), 883–892. doi:10.1056/NEJMoa1113205

Gevertz, J., Gillies, G., & Torquato, S. (2008). Simulating tumor growth in confined heterogeneous environments. *Physical Biology*, *5*(036010), 036010.

Ghaemi, M., & Shahrokhi, A. (2006). Combination of the cellular Potts model and lattice gas cellular automata for simulating the avascular cancer growth. *Lecture Notes in Computer Science*, *4173*, 297.

Gibbs, W. (2003). Untangling the roots of cancer. *Scientific American*, *289*(1), 56–65.

Gierer, A. (1981). Generation of biological patterns and form: Some physical, mathematical, and logical aspects. *Prog. Biophys. Mol. Biol*, *37*, 1–47.

Girouda, F., Haroskeb, G., Reithc, A., & Böckingd, A. (1998). Part II: Specific recommendations for quality assurance. *Analytical Cellular Pathology*, *17*(4), 201–208.

Greaves, M., Maia, A., Wiemels, J., & Ford, A. (2003). Leukemia in twins: lessons in natural history. *Blood*, *102*(7), 2321.

Greenspan, H. (1972). Models for the growth of a solid tumor by diffusion. *Stud. Appl. Math*, *51*(4), 317–340.

Grimm, T. (1998). The Human Condition: A justification for rapid prototyping. *Time Compression Technologies*, *3* (3), 1–6.

Guo, W., Lasky, J. L., Chang, C.-J., Mosessian, S., Lewis, X., Xiao, Y., et al. (2008). Multi-genetic events collaboratively contribute to Pten-null leukaemia stem-cell formation. *Nature*, *453*(7194), 529–533. doi:10.1038/nature06933

Ha, M., Freedman, M., Hauptmann, M., Sigurdson, A., Doody, M., Mabuchi, K., & Linet, M. (2004). Cigarette smoking at early age and breast cancer in the U.S. radiologic technologist health study. *Annals of Epidemiology*, *14*(8), 595. doi:10.1016/j.annepidem.2004.07.009

Hahn, W., Counter, C., Lundberg, A., Beijersbergen, R., Brooks, M., & Weinberg, R. (1999). Creation of human tumour cells with defined genetic elements. *Nature*, *400*(6743), 464–468.

Hajdu, S. I. (2004). Greco-Roman thought about cancer. *Cancer*, *100*(10), 2048–2051. doi:10.1002/cncr.20198

Hajdu, S. I. (2010). A note from history: Landmarks in history of cancer, part 1. *Cancer*, *117*(5), 1097–1102. doi:10.1002/cncr.25553

Hanahan, D., & Weinberg, R. (2000). The hallmarks of cancer. *Cell*, *100*(1), 57–70.

Hanahan, D., & Weinberg, R. A. (2011). Hallmarks of cancer: the next generation. *Cell*, *144*(5), 646–674. doi:10.1016/j.cell.2011.02.013

Hanks, S., & Rahman, N. (2005). Aneuploidy-cancer predisposition syndromes - A new link between the mitotic spindle checkpoint and cancer. *Cell Cycle (Georgetown, Tex)*, *4*(2), 225–

227.

Hanks, S., Coleman, K., Reid, S., Plaja, A., Firth, H., FitzPatrick, D., et al. (2004). Constitutional aneuploidy and cancer predisposition caused by biallelic mutations in BUB1B. *Nature Genetics*, *36*(11), 1159–1161.

Haroskea, G., Giroudb, F., Reithc, A., & Böckingd, A. (1998). Part I: Basic considerations and recommendations for preparation, measurement and interpretation. *Analytical Cellular Pathology*, *17*(4), 189–200.

Hede, K. (2005). Which came first? Studies clarify role of aneuploidy in cancer. *JNCI Journal of the National Cancer Institute*, *97*(2), 87.

Heidtke, K., & Schulze-Kremer, S. (1998). Design and implementation of a qualitative simulation model of lambda phage infection. *Bioinformatics*, *14*(1), 81–91.

Hertzberg, L., Betts, D. R., Raimondi, S. C., Schäfer, B. W., Notterman, D. A., Domany, E., & Izraeli, S. (2007). Prediction of chromosomal aneuploidy from gene expression data. *Genes Chromosomes & Cancer*, *46*(1), 75–86. doi:10.1002/gcc.20391

Highsmith, J., & Cockburn, A. (2001). Agile software development: The business of innovation. *Computer*, *34*(9), 120–127.

Hogeweg, P. (2002). Computing an organism: on the interface between informatic and dynamic processes. *Biosystems*, *64*(1-3), 97–109.

Holland, A. J., & Cleveland, D. W. (2009). Boveri revisited: chromosomal instability, aneuploidy and tumorigenesis. *Nature Reviews Molecular Cell Biology*, *10*(7), 478–487. doi:10.1038/nrm2718

Huang, S., & Ingber, D. (2007). A non-genetic basis for cancer progression and metastasis: self-organizing attractors in cell regulatory networks. *Breast Disease*, *26*(1), 27–54.

Hudson, T. J., Anderson, W., Aretz, A., Barker, A. D., Bell, C., Bernabé, R. R., ... & Lichter, P. (2010). International network of cancer genome projects. Nature, 464(7291), 993-998.

Huettel, B., Kreil, D., Matzke, M., & Matzke, A. (2008). Effects of aneuploidy on genome structure, expression, and interphase organization in Arabidopsis thaliana. *PLoS Genetics*, *4*(10), e1000226.

Hunter, K. (2004). Ezrin, a key component in tumor metastasis. *Trends in Molecular Medicine*, *10*(5), 201–204. doi:10.1016/j.molmed.2004.03.001

Hutchings, S. E., & Sato, G. H. (1978). Growth and maintenance of HeLa cells in serum-free medium supplemented with hormones. *Proceedings of the National Academy of Sciences of the United States of America*, *75*(2), 901–904.

Ideker, T., & Lauffenburger, D. (2003). Building with a scaffold: emerging strategies for high- to low-level cellular modeling. *Trends in Biotechnology*, *21*(6), 255–262. doi:10.1016/S0167-7799(03)00115-X

Jacobs, E., Luce, J., & Cailleau, R. (2006). Chromosome abnormalities in human cancer. Report of a patient with chronic myelocytic leukemia and his nonleukemic monozygotic twin. *CA a Cancer Journal for Clinicians*, *19*(6), 869–876.

Jefford, C., & Irminger-Finger, I. (2006). Mechanisms of chromosome instability in cancers. *Critical Reviews in Oncology and Hematology*, *59*(1), 1–14.

Jemal, A., Siegel, R., Ward, E., Hao, Y., Xu, J., Murray, T., & Thun, M. J. (2008). Cancer Statistics, 2008. *CA a Cancer Journal for Clinicians*, *58*(2), 71–96. doi:10.3322/CA.2007.0010

Jeon, J., Quaranta, V., & Cummings, P. T. (2010). An off-lattice hybrid discrete-continuum model of tumor growth and invasion. Biophysical journal, 98(1), 37-47.

Johansson, B., Mertens, F., & Mitelman, F. (1996). Primary vs, secondary neoplasia-associated chromosomal abnormalities - Balanced rearrangements vs, genomic imbalances? *Genes Chromosomes & Cancer*, *16*(3), 155–163.

John, M., Ewald, R., & Uhrmacher, A. M. (2008). A Spatial Extension to the π Calculus.

Electronic notes in theoretical computer science, *194*(3), 133–148. doi:10.1016/j.entcs.2007.12.010

Juul, N., Szallasi, Z., Eklund, A. C., Li, Q., Burrell, R. A., Gerlinger, M., et al. (2010). Assessment of an RNA interference screen-derived mitotic and ceramide pathway metagene as a predictor of response to neoadjuvant paclitaxel for primary triple-negative breast cancer: a retrospective analysis of five clinical trials. *The Lancet Oncology*, *11*(4), 358–365. doi:10.1016/S1470-2045(10)70018-8

Kampis, G., Karsai, I., & Szathmáry, E. (2011). Advances in Artificial Life. Darwin Meets von Neumann - 10th European Conference, ECAL 2009, Budapest, Hungary, September 13-16, 2009, Revised Selected Papers, Part I. *Ecal 2011*. doi:10.1007/978-3-642-21283-3

Kauffman, S. (1969). Metabolic stability and epigenesis in randomly constructed genetic nets. *Journal of Theoretical Biology*, *22*(3), 437.

Kimura, J., Nguyen, S. T., Liu, H., Taira, N., Miki, Y., & Yoshida, K. (2008). A functional genome-wide RNAi screen identifies TAF1 as a regulator for apoptosis in response to genotoxic stress. *Nucleic Acids Research*, *36*(16), 5250–5259. doi:10.1093/nar/gkn506

Kittler, R., Pelletier, L., Heninger, A., Slabicki, M., Theis, M., Miroslaw, L., et al. (2007). Genome-scale RNAi profiling of cell division in human tissue culture cells. *Nature Cell Biology*, *9*(12), 1401–1412.

Komarova, N. L., Sengupta, A., & Nowak, M. A. (2003). Mutation–selection networks of cancer initiation: tumor suppressor genes and chromosomal instability. Journal of theoretical biology, 223(4), 433-450.

Kontopoulos, E. V., Gualtieri, M., & Quintero, R. A. (2012). Successful in utero treatment of an oral teratoma via operative fetoscopy: case report and review of the literature. *Ymob*, *207*(1), e12–e15. doi:10.1016/j.ajog.2012.04.008

Kunda, P., Pelling, A., Liu, T., & Baum, B. (2008). Moesin controls cortical rigidity, cell

rounding, and spindle morphogenesis during mitosis. *Current Biology*, *18*(2), 91–101.

Kwon, M., Godinho, S. A., Chandhok, N. S., Ganem, N. J., Azioune, A., Thery, M., & Pellman, D. (2008). Mechanisms to suppress multipolar divisions in cancer cells with extra centrosomes. *Genes & Development*, *22*(16), 2189–2203. doi:10.1101/gad.1700908

Laan, M.-H., Soubieux, D., Mérat, L., Bouret, D., Luneau, G., Dambrine, G., & Thoraval, P. (2004). Genetic analysis of a divergent selection for resistance to Rous sarcomas in chickens. *Genetics Selection Evolution*, *36*(1), 65. doi:10.1186/1297-9686-36-1-65

Lai, L. A., Paulson, T. G., Li, X., SancheZ, C. A., Maley, C., Ocize, R. D., et al. (2007). Increasing genomic instability during premalignant neoplastic progression revealed through high resolution array-CGH. *Genes Chromosomes & Cancer*, *46*(6), 532–542. doi:10.1002/gcc.20435

Le Martelot, E., & Bentley, P. J. (2009). Modelling biological processes naturally using systemic computation: Genetic algorithms, neural networks, and artificial immune systems. *Nature-Inspired Informatics for Intelligent Applications and Knowledge Discovery: Implications in Business, Science and Engineering, R. Choing, Ed. IGI Global*.

Lee, A. J. X., Endesfelder, D., Rowan, A. J., Walther, A., Birkbak, N. J., Futreal, P. A., et al. (2011). Chromosomal instability confers intrinsic multidrug resistance. *Cancer Research*, *71*(5), 1858–1870. doi:10.1158/0008-5472.CAN-10-3604

Lee, S., Huang, H., Niu, Y., Tommasino, M., Lenoir, G., & Sylla, B. S. (2007). Dok1 expression and mutation in Burkitt's lymphoma cell lines. *Cancer Letters*, *245*(1-2), 44–50. doi:10.1016/j.canlet.2005.10.045

Li, X., Schimenti, J., & Tye, B. (2009). Aneuploidy and improved growth are coincident but not causal in a yeast cancer model. *PLoS Biology*, *7*, e1000161.

Liotta, L. A., Saidel, G. M., & Kleinerman, J. (1976). Stochastic model of metastases formation. *Biometrics*, *32*(3), 535–550.

Macal, C., & North, M. (2005). Tutorial on agent-based modeling and simulation. *Proceedings of the 37th Conference on Winter Simulation*, 2–15.

Machado, D., Costa, R., Rocha, M., Rocha, I., Tidor, B., & Ferreira, E. (2009). A critical review on modelling formalisms and simulation tools in computational biosystems. *Distributed Computing, Artificial Intelligence, Bioinformatics, Soft Computing, and Ambient Assisted Living*, 1063–1070.

Macklin, P., & Lowengrub, J. (2006). An improved geometry-aware curvature discretization for level set methods: application to tumor growth. *Journal of Computational Physics*, *215*(2), 392–401.

Macklin, P., & Lowengrub, J. (2007). Nonlinear simulation of the effect of microenvironment on tumor growth. *Journal of Theoretical Biology*, *245*(4), 677–704.

Mahoney, A., Smith, B., Flann, N., & Podgorski, G. (2008). Discovering Novel Cancer Therapies: A computational modeling and search approach. *IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology, 2008. CIBCB'08*, 233–240.

Maley, C., & Reid, B. (2005). Natural selection in neoplastic progression of Barrett's esophagus. *Seminars in Cancer Biology*, *15*(6), 474–483.

Manchester, K. (1995). Theodor Boveri and the origin of malignant tumours. *Trends in Cell Biology*, *5*(10), 384–387.

March, N., & Plaskett, J. (1956). The relation between the wentzel-kramers-brillouin and the thomas-fermi approximations. *Proceedings of the Royal Society of London. Series a. Mathematical and Physical Sciences*, *235*(1202), 419–431.

Marcucci, G., Haferlach, T., & Döhner, H. (2011). Molecular genetics of adult acute myeloid leukemia: prognostic and therapeutic implications. *Journal of Clinical Oncology*, *29*(5), 475–486. doi:10.1200/JCO.2010.30.2554

Marinari, E., Mehonic, A., Curran, S., Gale, J., Duke, T., & Baum, B. (2012). Live-cell delamination counterbalances epithelial growth to limit tissue overcrowding. *Nature*, *484*(7395), 542–545. doi:10.1038/nature10984

Materi, W., & Wishart, D. (2007). Computational systems biology in cancer: modeling methods and applications. *Gene Regulation and Systems Biology*, *1*, 91.

Matlashewski, G., Lamb, P., Pim, D., Peacock, J., Crawford, L., & Benchimol, S. (1984). Isolation and characterization of a human p53 cDNA clone: expression of the human p53 gene. *The EMBO Journal*, *3*(13), 3257.

McClelland, S. E., Burrell, R. A., & Swanton, C. (2009). Chromosomal instability: a composite phenotype that influences sensitivity to chemotherapy. *Cell Cycle (Georgetown, Tex)*, *8*(20), 3262–3266.

McDougall, S. R., Anderson, A. R. A., Chaplain, M. A. J., & Sherratt, J. A. (2002). Mathematical modelling of flow through vascular networks: implications for tumour-induced angiogenesis and chemotherapy strategies. *Bulletin of Mathematical Biology*, *64*(4), 673–702. doi:10.1006/bulm.2002.0293

Mendelsohn, M., Hungerford, D., Mayall, B., Perry, B., Conway, T., Judith, M., & Prewitt, S. (1969). Computer-oriented analysis of human chromosomes. ii. integrated optical density as a single parameter for karyotype analysis. *Annals of the New York Academy of Sciences*, *157*(1 Data Extraction and Processing of Optical Images in the Medical and Biological Sciences), 376–392.

Merks, R., & Glazier, J. (2005). A cell-centered approach to developmental biology. *Physica a-Statistical Mechanics and Its Applications*, *352*(1), 113–130. doi:10.1016/j.physa.2004.12.028

Merks, R., & Glazier, J. (2006). Dynamic mechanisms of blood vessel growth. *Nonlinearity*, *19*(1), C1–C10. doi:10.1088/0951-7715/19/1/000

Merlo, L. M. F., Pepper, J. W., Reid, B. J., & Maley, C. C. (2006). Cancer as an evolutionary and ecological process. *Nature Reviews Cancer*, *6*(12), 924–935. doi:10.1038/nrc2013

Meyerson, M., Gabriel, S., & Getz, G. (2010). Advances in understanding cancer genomes through second-generation sequencing. *Nature Publishing Group*, *11*(10), 685–696. doi:10.1038/nrg2841

Meza, R., Jeon, J., Moolgavkar, S., & Luebeck, E. (2008). Age-specific incidence of cancer: Phases, transitions, and biological implications. *Proceedings of the National Academy of Sciences*, *105*(42), 16284.

Michor, F., Iwasa, Y., & Nowak, M. (2004). Dynamics of cancer progression. *Nature Reviews Cancer*, *4*(3), 197–205.

Michor, F. (2005). Chromosomal instability and human cancer. Philosophical Transactions of the Royal Society B: Biological Sciences, 360(1455), 631-635.

Minsky, M. (1961). Steps toward artificial intelligence. *Proceedings of the IRE*, *49*(1), 8–30.

Minsky, M. (1974). A framework for representing knowledge. http://hdl.handle.net/1721.1/6089

Muzny, D. M., Bainbridge, M. N., Chang, K., Dinh, H. H., Drummond, J. A., Fowler, G., et al. (2012). Comprehensive molecular characterization of human colon and rectal cancer. *Nature*, *487*, 330–337. doi:10.1038/nature11252

Nagl, S. (2006). A path to knowledge: from data to complex systems models of cancer. *Cancer Bioinformatics*, 3–27.

Nagl, S., Williams, M., & Williamson, J. (2007). Objective Bayesian nets for systems modelling and prognosis in breast cancer. *Innovations in Bayesian Networks: Theory and Applications. Springer*.

Nagl, S., Williams, M., El-Mehidi, N., Patkar, V., & Williamson, J. (2006). Objective Bayesian nets for integrating cancer knowledge: a systems biology approach. *Proceedings of the Workshop on Probabilistic Modelling and Machine Learning in Structural and Systems*

Biology, Tuusula, June17-18, 44–49.

Navin, N., Kendall, J., Troge, J., Andrews, P., Rodgers, L., Mcindoo, J., et al. (2011). Tumour evolution inferred by single-cell sequencing. *Nature*. doi:10.1038/nature09807

Nelson, C. M., & Bissell, M. J. (2006). Of extracellular matrix, scaffolds, and signaling: Tissue architecture regulates development, homeostasis, and cancer. *Annual Review of Cell and Developmental Biology*, *22*, 287–309. doi:10.1146/annurev.cellbio.22.010305.104315

Nigg, E. A. (2002). Centrosome aberrations: cause or consequence of cancer progression? *Nature Reviews Cancer*, *2*(11), 815–825. doi:10.1038/nrc924

Nowak, M. A., Komarova, N. L., Sengupta, A., Jallepalli, P. V., Shih, I.-M., Vogelstein, B., & Lengauer, C. (2002). The role of chromosomal instability in tumor initiation. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(25), 16226–16231. doi:10.1073/pnas.202617399

Pavelka, N., Rancati, G., & Li, R. (2010a). Dr Jekyll and Mr Hyde: role of aneuploidy in cellular adaptation and cancer. *Current Opinion in Cell Biology*, *22*(6), 809–815. doi:10.1016/j.ceb.2010.06.003

Pavelka, N., Rancati, G., Zhu, J., Bradford, W. D., Saraf, A., Florens, L., et al. (2010b). Aneuploidy confers quantitative proteome changes and phenotypic variation in budding yeast. *Nature*, *468*(7321), 321–325. doi:10.1038/nature09529

Pérez de Castro, I., de Cárcer, G., & Malumbres, M. (2007). A census of mitotic cancer genes: new insights into tumor cell biology and cancer therapy. *Carcinogenesis*, *28*(5), 899–912. doi:10.1093/carcin/bgm019

Phillips, J., Hayward, S., Wang, Y., Vasselli, J., Pavlovich, C., Padilla-Nash, H., et al. (2001). The consequences of chromosomal aneuploidy on gene expression profiles in a cell line model for prostate carcinogenesis. *Cancer Research*, *61*(22), 8143–8149.

Pressman, R. (2009). Software Engineering: A Practitioner's Approach, 7 edition. *Software*

Engineering: a Practitioner's Approach, 7 Edition.

Rabiner, L. R. (1999). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, *77*(2), 257–286.

Rajagopalan, H., & Lengauer, C. (2004). Aneuploidy and cancer. *Nature*, *432*(7015), 338–341. doi:10.1038/nature03099

Regev, A., Panina, E. M., Silverman, W., Cardelli, L., & Shapiro, E. (2004). BioAmbients: an abstraction for biological compartments. *Theoretical Computer Science*, *325*(1), 141–167. doi:10.1016/j.tcs.2004.03.061

Regev, A., Silverman, W., & Shapiro, E. (2001). Representation and simulation of biochemical processes using the pi-calculus process algebra. *Pacific Symposium on Biocomputing*, *6*, 459–470.

Ribba, B., Alarcon, T., Marron, K., Maini, P., & Agur, Z. (2004). The use of hybrid cellular automaton models for improving cancer therapy. *Lecture Notes in Computer Science*, 444–453.

Ricke, R. M., Jeganathan, K. B., & Deursen, J. M. V. (2011). Bub1 overexpression induces aneuploidy and tumor formation through Aurora B kinase hyperactivation. *The Journal of Cell Biology*, *193*(6), 1049–1064. doi:10.1083/jcb.201012035

Rochon, J., Gondan, M., & Kieser, M. (2012). To test or not to test: Preliminary assessment of normality when comparing two independent samples. *BMC Medical Research Methodology*, *12*(1), 81.

Roose, T., Chapman, S., & Maini, P. (2007). Mathematical models of avascular tumor growth. *SIAM Review*, *49*(2), 179.

Royce, W. W. (1970). Managing the development of large software systems. *Proceedings of IEEE WESCON*, *26*(8).

Saez-Rodriguez, J., Alexopoulos, L. G., Epperlein, J., Samaga, R., Lauffenburger, D. A., Klamt,

S., & Sorger, P. K. (2009). Discrete logic modelling as a means to link protein signalling networks with functional analysis of mammalian signal transduction. *Molecular Systems Biology*, *5*, 1–19. doi:10.1038/msb.2009.87

Sahin, Ö., Fröhlich, H., Löbke, C., Korf, U., Burmester, S., Majety, M., et al. (2009). Modeling ERBB receptor-regulated G1/S transition to find novel targets for de novo trastuzumab resistance. *BMC Systems Biology*, *3*(1), 1. doi:10.1186/1752-0509-3-1

Sakorafas, G., & Safioleas, M. (2009). Breast cancer surgery: an historical narrative. Part I. From prehistoric times to Renaissance. *European Journal of Cancer Care*, *18*(6), 530–544. doi:10.1111/j.1365-2354.2008.01059.x

Sakorafas, G., & Safioleas, M. (2010). Breast cancer surgery: an historical narrative. Part II. 18th and 19th centuries. *European Journal of Cancer Care*, *19*(1), 6–29. doi:10.1111/j.1365-2354.2008.01060.x

Sakorafas, G., Safioleas, M., & Safioleas, M. (2010). Breast cancer surgery: an historical narrative. Part III. From the sunset of the 19th to the dawn of the 21st century. *European Journal of Cancer Care*, *19*(2), 145–166. doi:10.1111/j.1365-2354.2008.01061.x

Saul, J. M., & Schwartz, L. (2007). Cancer as a consequence of the rising level of oxygen in the Late Precambrian. *Lethaia*, *40*(3), 211–220. doi:10.1111/j.1502-3931.2007.00014.x

Scharer, C. D., Laycock, N., Osunkoya, A. O., Logani, S., McDonald, J. F., Benigno, B. B., & Moreno, C. S. (2008). Aurora kinase inhibitors synergize with paclitaxel to induce apoptosis in ovarian cancer cells. *Journal of Translational Medicine*, *6*(1), 79. doi:10.1186/1479-5876-6-79

Schneider, R., & Grosschedl, R. (2007). Dynamics and interplay of nuclear architecture, genome organization, and gene expression. *Genes & Development*, *21*(23), 3027–3043. doi:10.1101/gad.1604607

Schnell, S., Grima, R., & Maini, P. (2007). Multiscale modeling in biology. *American Scientist*,

95(1), 134–142.

Schvartzman, J.-M., Sotillo, R., & Benezra, R. (2010). Mitotic chromosomal instability and cancer: mouse modelling of the human disease. *Nature Reviews Cancer*, *10*(2), 102–115. doi:10.1038/nrc2781

Selmecki, A. M., Dulmage, K., Cowen, L. E., Anderson, J. B., & Berman, J. (2009). Acquisition of aneuploidy provides increased fitness during the evolution of antifungal drug resistance. *PLoS Genetics*, *5*(10), e1000705. doi:10.1371/journal.pgen.1000705

Seluanov, A., Hine, C., Azpurua, J., Feigenson, M., Bozzella, M., Mao, Z., et al. (2009). Hypersensitivity to contact inhibition provides a clue to cancer resistance of naked mole-rat. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(46), 19352–19357.

Sen, S. (2000). Aneuploidy and cancer. *Current Opinion in Oncology*, *12*(1), 82–88.

Shackleton, M., Quintana, E., & Fearon, E. (2009). Heterogeneity in cancer: cancer stem cells versus clonal evolution. *Cell*. 138.5 (2009): 822-829.

Shmulevich, I., Dougherty, E. R., Kim, S., & Zhang, W. (2002). Probabilistic Boolean networks: a rule-based uncertainty model for gene regulatory networks. *Bioinformatics*, *18*(2), 261–274.

Silverston, L., Inmon, W. H., & Graziano, K. (1997). The Data Model Resource Book: A Library of Logical Data Models and Data Warehouse Designs.

Solé, R. V., & Deisboeck, T. S. (2004). An error catastrophe in cancer?. Journal of theoretical biology, 228(1), 47-54.

Sotillo, R., Hernando, E., Diaz-Rodriguez, E., Teruya-Feldstein, J., Cordon-Cardo, C., Lowe, S. W., & Benezra, R. (2007). Mad2 overexpression promotes aneuploidy and tumorigenesis in mice. *Cancer Cell*, *11*(1), 9–23. doi:10.1016/j.ccr.2006.10.019

Sotillo, R., Schvartzman, J.-M., Socci, N. D., & Benezra, R. (2010). Mad2-induced chromosome

instability leads to lung tumour relapse after oncogene withdrawal. *Nature*, *464*(7287), 436–440. doi:10.1038/nature08803

Soucek, L., Whitfield, J., Martins, C. P., Finch, A. J., Murphy, D. J., Sodir, N. M., et al. (2008). Modelling Myc inhibition as a cancer therapy. *Nature*, *455*(7213), 679–683. doi:10.1038/nature07260

Spencer, S. L., Gerety, R. A., Pienta, K. J., & Forrest, S. (2006). Modeling somatic evolution in tumorigenesis. *PLoS Computational Biology*, *2*(8), e108. doi:10.1371/journal.pcbi.0020108

Stephens, P. J., Mcbride, D. J., Lin, M.-L., Varela, I., Pleasance, E. D., Simpson, J. T., et al. (2009). Complex landscapes of somatic rearrangement in human breast cancer genomes. *Nature*, *462*(7276), 1005–1010. doi:10.1038/nature08645

Stephens, P., Hunter, C., Bignell, G., Edkins, S., Davies, H., Teague, J., et al. (2004). Lung cancer: intragenic ERBB2 kinase mutations in tumours. *Nature*, *431*(7008), 525–526.

Stokes, C. L., & Lauffenburger, D. A. (1991). Analysis of the roles of microvessel endothelial cell random motility and chemotaxis in angiogenesis. *Journal of Theoretical Biology*, *152*(3), 377–403. doi:10.1016/S0022-5193(05)80201-2

Strachan, T., Abitbol, M., Davidson, D., & Beckmann, J. S. (1997). A new dimension for the human genome project: towards comprehensive expression maps. *Nature Genetics*, *16*(2), 126–132.

Stratton, M. R., Campbell, P. J., & Futreal, P. A. (2009). The cancer genome. *Nature*, *458*(7239), 719–724. doi:10.1038/nature07943

Swanton, C., & Caldas, C. (2009). Molecular classification of solid tumours: towards pathway-driven therapeutics. *British Journal of Cancer*, *100*(10), 1517–1522. doi:10.1038/sj.bjc.6605031

Swanton, C., Nicke, B., Marani, M., Kelly, G., & Downward, J. (2007). Initiation of high frequency multi-drug resistance following kinase targeting by siRNAs. *Cell Cycle*

(Georgetown, Tex), *6*(16), 2001–2004.

Swanton, C., Nicke, B., Schuett, M., Eklund, A. C., Ng, C., Li, Q., et al. (2009). Chromosomal instability determines taxane response. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(21), 8671–8676. doi:10.1073/pnas.0811835106

Swanton, C., Tomlinson, I., & Downward, J. (2006). Chromosomal instability, colorectal cancer and taxane resistance. *Cell Cycle (Georgetown, Tex)*, *5*(8), 818–823.

Szathmary, E. O. (1995). A classification of replicators and lambda-calculus models of biological organization. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *260*(1359), 279–286.

Thompson, S. L., & Compton, D. A. (2008). Examining the link between chromosomal instability and aneuploidy in human cells. *The Journal of Cell Biology*, *180*(4), 665–672. doi:10.1083/jcb.200712029

Thompson, S. L., & Compton, D. A. (2010). Proliferation of aneuploid human cells is limited by a p53-dependent mechanism. *The Journal of Cell Biology*, *188*(3), 369–381. doi:10.1083/jcb.200905057

Thompson, S. L., Bakhoum, S. F., & Compton, D. A. (2010). Mechanisms of chromosomal instability. *Current Biology : CB*, *20*(6), R285–95. doi:10.1016/j.cub.2010.01.034

Torres, E. M., Sokolsky, T., Tucker, C. M., Chan, L. Y., Boselli, M., Dunham, M. J., & Amon, A. (2007). Effects of aneuploidy on cellular physiology and cell division in haploid yeast. *Science*, *317*(5840), 916–924. doi:10.1126/science.1142210

Torres, E. M., Williams, B. R., & Amon, A. (2008). Aneuploidy: cells losing their balance. *Genetics*, *179*(2), 737–746. doi:10.1534/genetics.108.090878

Tzenov, Y. R., Andrews, P. G., Voisey, K., Popadiuk, P., Xiong, J., Popadiuk, C., & Kao, K. R. (2013). Human Papilloma Virus (HPV) E7-mediated attenuation of Retinoblastoma (Rb) induces hPygopus2 expression via Elf-1 in cervical cancer. *Molecular Cancer Research*.

doi:10.1158/1541-7786.MCR-12-0510

Uetake, Y., & Sluder, G. (2004). Cell cycle progression after cleavage failure: mammalian somatic cells do not possess a "tetraploidy checkpoint." *The Journal of Cell Biology*, *165*(5), 609–615. doi:10.1083/jcb.200403014

Vokes, M., & Carpenter, A. (2008). Using CellProfiler for automatic identification and measurement of biological objects in images. *Current Protocols in Molecular Biology*, *14*, 1–14.17.

Wagner, K. U. (2004). Models of breast cancer: quo vadis, animal modeling? *Breast Cancer Research*, *6*(1), 31–38. doi:10.1186/bcr723

Walther, A., Houlston, R., & Tomlinson, I. (2008). Association between chromosomal instability and prognosis in colorectal cancer: a meta-analysis. *Gut*, *57*(7), 941–950. doi:10.1136/gut.2007.135004

Walther, A., Johnstone, E., Swanton, C., Midgley, R., Tomlinson, I., & Kerr, D. (2009). Genetic prognostic and predictive markers in colorectal cancer. *Nature Reviews Cancer*, *9*(7), 489–499. doi:10.1038/nrc2645

Wang, L. E., Xiong, P., Strom, S. S., Goldberg, L. H., Lee, J. E., Ross, M. I., et al. (2005). *In Vitro* sensitivity to ultraviolet B light and skin cancer risk: a case-control analysis. *JNCI Journal of the National Cancer Institute*, *97*(24), 1822–1831. doi:10.1093/jnci/dji429

Weaver, B. A. A., Silk, A. D., Montagna, C., Verdier-Pinard, P., & Cleveland, D. W. (2007). Aneuploidy acts both oncogenically and as a tumor suppressor. *Cancer Cell*, *11*(1), 25–36. doi:10.1016/j.ccr.2006.12.003

Weaver, B., & Cleveland, D. (2006). Does aneuploidy cause cancer? *Current Opinion in Cell Biology*, *18*(6), 658–667.

Weaver, B., & Cleveland, D. (2007). Aneuploidy: instigator and inhibitor of tumorigenesis. *Cancer Research*, *67*(21), 10103.

Webb, B. (2002). Can robots make good models of biological behaviour? *Behavioral and Brain Sciences*, *24*(06). doi:10.1017/S0140525X01000127

Webb, B. (2009). Animals versus animats: Or why not model the real iguana? *Adaptive Behavior*, *17*(4), 269–286.

Weinberg, R. A. (2007). *The biology of cancer*. Garland Pub.

Weinstein, I. B., & Joe, A. (2008). Oncogene addiction. *Cancer Research*, *68*(9), 3077–80; discussion 3080. doi:10.1158/0008-5472.CAN-07-3293

Wette, R., Katz, I., & Rodin, E. (1974a). Stochastic processes for solid tumor kinetics I. Surface-regulated growth. *Mathematical Biosciences*, *19*(3-4), 231–255.

Wette, R., Katz, I., & Rodin, E. (1974b). Stochastic processes for solid tumor kinetics II. Diffusion-regulated growth. *Mathematical Biosciences*, *21*(3-4), 311–338.

Whitten, J. L., Bentley, L. D., & Dittman, K. C. (2004). *Systems analysis and design methods*. McGraw-Hill/Irwin.

Wing, J. M. (2002). FAQ on π-Calculus. *Microsoft Research*, 1–8. Retrieved from http://www.cs.tufts.edu/~nr/cs257/archive/jeannette-wing/pi.pdf

Wodarz, D., & Komarova, N. L. (2005). Computational biology of cancer: lecture notes and mathematical modeling, 250.

Wolfram, S. (2002). *A new kind of science*. Wolfram Media Inc.

Wu, M., Pastor-Pareja, J. C., & Xu, T. (2010). Interaction between Ras(V12) and scribbled clones induces tumour growth and invasion. *Nature*, *463*(7280), 545–U165. doi:10.1038/nature08702

Wu, X., Zhang, J., He, X., Wang, C., Lian, L., Liu, H., et al. (2011). Postoperative adjuvant chemotherapy for stage ii colorectal cancer: a systematic review of 12 randomized controlled trials. *Journal of Gastrointestinal Surgery*, *16*(3), 646–655. doi:10.1007/s11605-011-1682-8

Zeng, Q., & Hong, W. (2008). The emerging role of the hippo pathway in cell contact inhibition,

organ size control, and cancer development in mammals. *Cancer Cell*, *13*(3), 188–192. doi:10.1016/j.ccr.2008.02.011

Zhang, L., Ren, X., Alt, E., Bai, X., Huang, S., Xu, Z., et al. (2010). Chemoprevention of colorectal cancer by targeting APC-deficient cells for apoptosis. *Nature*, *464*(7291), 1058–1061. doi:10.1038/nature08871