Mathematical Modelling Of Cell Population Dynamics In The Colonic Crypt With Application To Colorectal Cancer

Matthew D. Johnston
Trinity College
University of Oxford

A thesis submitted for the degree of
Doctor of Philosophy
Trinity Term 2008

Acknowledgements

It has been a great privilege and pleasure to work with my supervisors, Professor Philip Maini, Professor Jon Chapman, Dr Carina Edwards and Professor Sir Walter Bodmer over the last four years, and I would like to thank them for all their encouragement and guidance. To Philip for cheerful enthusiasm, and being a fantastic mentor; to Jon for inspiring me and always being able to express everything clearly and succinctly; to Carina for valuable help and patience, especially early on with my bumbling writing style; and to Sir Walter for enlightening me on biological issues. I am also extremely grateful to Professor Hans Clevers and Dr Ariel Altaba for their assistance with crypt biology and the latest experimental techniques.

This research was generously supported by the Engineering and Physical Sciences Research Council via a DTA Graduate Studentship (Award Number EP/P500397/1), and part of this work was supported by the NIH (National Institutes of Health) Virtual Tumor grant U56 CA 113004-02. I would also like to acknowledge Trinity College, Oxford, for Lectureships in Applied Mathematics, and for the Sarah and Nadine Pole Scholarship for travel funding to present a paper at a conference.

I would particularly like to thank Mr Bushell, who first inspired me in Mathematics, Mr Kiddle, and Dr Chris Prior for his advice and support since I began at Trinity.

I am grateful for the support and friendship of everyone at the Centre for Mathematical Biology (CMB). I would especially like to thank Alex and Phil for great banter, help, coffee times and tetris; Ruth for valuable advice, especially when I started; and Natasha, Steve, Charlotte, James and Helen for being a wonderful group, and many memorable CMB socials.

Also, thanks to all my friends at Trinity, who have made graduate study so enjoyable. To Louise, for all your support, humour and companionship; Thijs and 'neighbor' Katie for fun and photos; Con for weekly badminton with a smile; to Andy, Ian, Pats, Dave, Martin, Marcus and Kelly Ann; and to Mark and Oli for terrific tennis games come rain or shine. Also my thanks go to Mark and Frank for keeping me sane in holiday periods, and reminding me about the real world outside Oxford.

Lastly, and most importantly, I am indebted to my family for all their support throughout my education. I would not have reached this point without you.

Abstract

Colorectal cancer has the third highest mortality and incidence rates of all cancers worldwide, but the prognosis for long-term survival is good if diagnosed early. It is a well-characterised disease, and is initiated in colonic crypts which line the colon wall. The aim of this thesis is to use mathematical modelling to describe the heavily regulated cell renewal cycle in the crypt to determine the key features of the system kinetics, and help to explain the initiation of tumourigenesis.

The dynamics of a single colorectal crypt is considered using a compartmental approach, which accounts for populations of stem, transit-amplifying and fully-differentiated cells. A number of different model formulations are derived, and their validity and suitability are discussed.

Two mechanisms are presented that could regulate the growth of cell numbers and maintain homeostasis (equilibrium), and it is illustrated how a model can describe both regulated and unregulated growth, with cancer-driving cells deriving from stem and/or transit cells. This model is used to explain the long lag phases observed in carcinogenesis, which occur between periods of rapid tumour expansion, before unlimited growth in cell numbers ensues. Significantly, it is found that, contrary to general belief, the proportion of cancer-driving cells in the exponential growth phase of a tumour may vary depending on tumour type. The process of cells accumulating mutations is also examined by considering both a stochastic individual cell-based model and an analytic approach.

Finally, an ordinary differential equation model is shown to be valid by considering a simplified description of a one-dimensional spatial model, and the latter is used to consider the effect of changing the crypt shape. The suitability of this modelling approach to tracking stem cells in a niche, as well as mutant cell clones as they propagate in the crypt, is also discussed.

Preface

This work forms part of the Integrative Biology Project, which is a larger multi-disciplinary, multi-institutional collaboration modelling colorectal cancer. Models are being produced on different length and time scales to address key biological questions about the initiation and progression of colorectal cancer. An integrative approach to this problem involves using mathematical modelling and computer simulation techniques on the sub-cellular, cellular and tissue levels of the colon. This thesis considers cell population dynamics, while other models focus on the subcellular and tissue scales, and it is hoped that the integration of these models will illuminate the mechanisms behind colorectal cancer. This collaboration opens the door to a new dimension of science, in which knowledge from different disciplines may be integrated by exploiting fully the latest computational advances.

Contents

List of Figures

List of Tables

Glossary

Adenoma	A benign tumour that develops from epithelial tissue, and may or may not transform into a cancer. Also referred to as an adenomatous polyp.	2
Allele	One of the two copies of a gene that occur at the same locus. Each individual has an allele from the mother and the father.	12
Aneuploidy	An abnormal number of chromosomes within a cell that leads to gross chromosomal abnormality.	12
Angiogenesis	The process by which a tumour develops its own blood supply.	1
APC	Adenomatous polyposis coli. A tumour suppressor gene that has many functions including the regulation of intracellular β-catenin levels.	5
Apoptosis	Programmed cell death as a result of activation of an intracellular suicide programme.	3
Avascular	A tissue that lacks a blood supply.	15
β-catenin	A protein that is an integral component in the Wnt signalling pathway, and drives cell cycle divisions.	5
Benign	A non-cancerous tumour that remains localised and does not invade its surroundings, causing damage only by local pressure or obstruction.	1
Carcinoma	A malignant tumour that develops from epithelial tissue.	2
CIN	Chromosomal instability. An enhanced rate of gaining or losing whole chromosomes or large parts of chromosomes during cell division.	12
Clone	A set of cells that share a common genetic makeup owing to descent from a common ancestor.	14
Colon	A portion of the large intestine preceding the rectum whose main function is the reabsorption of water.	1
Crypt	A tubular invagination of the intestinal epithelium that maximises the area for exchange and processing of food.	1
Differentiation	The process by which cells develop specialised functions from unspecialised precursor cells.	3

DNA	Deoxyribonucleic acid.	1
Dysplastic	An abnormality in the appearance of cells indicative of an early step towards transformation into neoplasia.	10
Epithelium	The tissue that covers the internal and external surfaces of the body, and lines hollow structures inside the body.	3
FAP	Familial adenomatous polyposis. An inherited condition in which numerous polyps form on the inside walls of the colon and rectum.	10
Fission	The process by which an organism divides into two similar or identical organisms.	10
Fitness	The average contribution of a genetic makeup to future generations.	9
Fixation	When a genetic change is conferred to every cell in the population.	25
Genetic drift	Random fluctuations in gene frequency occurring in populations from one generation to the next.	9
Genome	The complete set of genetic information of an organism.	9
Genomic instability	A raised intrinsic mutation rate.	12
Goblet cell	A mucous-secreting differentiated cell in the colonic crypt.	4
Heterozygosity	The possession of two different forms of a particular gene.	12
HNPCC	Hereditary non-polyposis colon cancer (also known as Lynch syndrome). An inherited disease in which individuals have a higher-than-normal chance of developing colorectal cancer.	10
Lineage	The descendants of a common ancestor.	4
LOH	Loss of heterozygosity. The loss of one of the two alleles at one or more loci in a cell lineage due to chromosome loss, deletion or mitotic crossing-over.	12
Luminal surface	The apical side of an epithelial cell.	3
Malignant	A cancerous growth with the ability to invade and destroy the surrounding tissue.	1
Metastasis	The process by which cancer spreads from one part of the body to another.	1
Methylation	The addition of a methyl group to a molecule. It can be involved in the regulation of gene expression and of protein function.	7

MIN	Microsatellite instability. A condition causing abnormally long or short repeated sequences of DNA due to defects in the normal DNA repair process (MMR deficiency). Not to be confused with the Min mouse which has a mutation in the mouse *APC* gene.	12
MMR	Mismatch repair. A DNA-repair mechanism that corrects nucleotide sequence errors made during DNA replication.	12
Morphogenesis	The development of shape and structure.	29
Mucosa	The mucous membrane that lines the gastrointestinal tract, which consists of several distinct layers of tissue including the epithelium.	3
Mutation	A change in a cell's DNA and chromosomes that alters its characteristics. It may arise spontaneously from errors in cell division, or be induced by carcinogens or radiation.	1
Mutator mutation	A mutation of DNA polymerase that increases the overall mutation rate.	27
Mutator phenotype	An increased tendency to specific types of mutation, caused by mutations at loci such as those involved in DNA replication or repair.	30
Necrosis	Unprogrammed death of cells as a result of external trauma such as physical damage or lack of nutrients.	39
Neoplasm	A new growth or tumour resulting from excessive cell division, which could be benign or malignant.	2
Neutral mutation	A mutation that does not effect survival, reproductive capacity or fitness.	12
PCD	Programmed cell death. Another name for apoptosis.	12
Phenotype	The observable (or measurable) physical and biochemical characteristics of an organism, resulting from the interaction of the genetic makeup and the environment.	7
Polyp	A small, mushroom-shaped growth in the epithelium which is usually benign but may become malignant. Also known as an adenoma.	10
Progenitor	An originator of a line of descendants.	3
Progeny	The offspring of a cell.	3
Proliferation	The process of reproduction or division of cells.	4
Quiescence	The state of a cell when it is neither differentiating nor undergoing cell division.	28
Ras	A protein, or small family of proteins, which is key to signalling processes for proliferation, cell adhesion, apoptosis and cell migration. Often deregulated in cancers leading to increased invasion and metastasis.	11

Recessive	The presence of two abnormal versions of a gene.	12
Selection	The differential overall reproductive rate of one organism or cell in comparison with another.	13
Somatic	All of the cells in the body except the reproductive cells.	12
Sporadic	The occurrence of a single individual in a family or pedigree, whether due to inheritance or not.	11
Stem cell	An undifferentiated cell that can divide indefinitely and give rise to highly specialised cells for each tissue.	3
Transcription	The process by which a DNA sequence is copied into messenger RNA.	11
TSG	Tumour suppressor gene. A gene involved in regulating cell behaviour which acts recessively, but has some effect in the heterozygote.	12
Tumourigenesis	The growth and development of a tumour.	2
Wild type	The most common phenotype of an organism, gene or characteristic in the natural population.	26

Chapter 1

Introduction

Colorectal cancer, or bowel cancer, is the most common cause of non-smoking related cancer deaths in the western world, and killed over 16,000 people in the UK in 2005 according to latest figures [Cancer Research UK, 2008]. Early diagnosis leads to more than 80% of patients surviving five years, whereas fewer than 10% of patients survive that long if diagnosis occurs after the cancer has spread to other organs in the body, and therefore much research targets understanding the early stages of the disease. Current clinical treatment usually involves removal of the colon and rectum, which is moderately successful, but interferes considerably with a patient's quality of life [Giles *et al.*, 2003]. Colorectal cancer is a good model system to study mathematically because there is much clinical data available for both healthy and cancerous colons.

Cancer is mostly a disease of old age [Bodmer, 1999], and the whole process from initiation to progression and diagnosis can take between 20 and 40 years [Rajagopalan *et al.*, 2003]. The formation of cancer is a multistage process [Bodmer, 1999; Renehan *et al.*, 2002], and these steps are thought to represent genetic changes in the DNA of the cells [Nowak *et al.*, 2002]. Hanahan and Weinberg [2000] proposed that there are six capabilities that are shared by most, and perhaps all, types of human tumours. These are self-sufficiency of growth signals, insensitivity to growth-inhibitory signals, evasion of cell death, limitless replicative potential, growth of a blood supply (angiogenesis), and tissue invasion and tumour growth in other organs (metastasis). The order in which these capabilities are acquired seems likely to be quite variable between different cancers, but these characteristics are common to all types of tumours. There are four distinct clinical stages of colorectal cancer progression, which are an abnormal growth in the colon wall, small benign tumours, malignant tumours invading surrounding tissues and metastatic cancer [Michor *et al.*, 2004b]. In this thesis we concentrate only on the first stage.

The gastrointestinal tract is one of the most frequent sites of carcinogenesis, possibly due in part to its continual self renewal, and the resultant large number of daily cell divisions in this tissue [Brittan and Wright, 2004b], although the small intestine turns over just as frequently as the colon but develops far fewer cancers. It is widely believed that the origins of colorectal cancer lie in the crypts, which are folds in the lining of the colon, as we show in Figure 1.1. A single layer of cells lines the crypt wall, and it is thought that mutations in these cells initiate the process of tumourigenesis. Consequently much work has been directed towards understanding the mechanisms involved in the dynamics of the cells in healthy and neoplastic (abnormal) crypts. Genetic mutations can enable cells to have independence, proliferate at a greater rate and form an adenoma (a benign

This image cannot be made available in this online version of the thesis because of copyright.

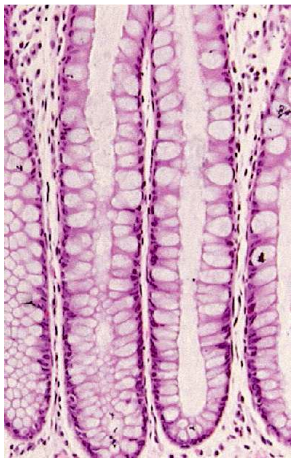

Figure 1.1: Pictures of the colon and crypts. Left: A cross-section of the colon, with millions of small invaginations (crypts) lining the inside wall. Reproduced from the University of Delaware website www.udel.edu/biology/Wags/histopage/ with permission. Right: Two normal human colonic crypts, reproduced with permission from Shih *et al.* [2001].

tumour), undergo angiogenesis, grow into a carcinoma (a malignant tumour) and spread to other organs by metastasis [Hanahan and Weinberg, 2000]. Adenomas may persist for years before the cancer is detected [Tsao *et al.*, 2000], and long lag phases may occur at each stage of this adenoma–carcinoma sequence [Tomlinson and Bodmer, 1999].

The aims of this thesis are to develop a mathematical model that will reproduce the experimentally observed dynamics in the crypt, including cell turnover and spatial dynamics, and can be used to predict the mechanisms that cause tumourigenesis. In Chapter 2 we review a number of current models of the crypt and colorectal cancer, and our modelling is presented in Chapters 3–6, before we review the work and discuss extensions in Chapter 7. Throughout the rest of this chapter we summarise the main biological features of the cell dynamics in a healthy crypt in Section 1.1, and give an outline of the effects of colorectal cancer on these processes in Section 1.2. The main mathematical and biological questions that we address are stated in Section 1.3.

1.1 The biology of a healthy colonic crypt

The colon is a tubular structure that comprises the main part of the large intestine, and runs from the small intestine to the rectum. In the colon there are tiny pits in the surface called crypts of Lieberkühn [Giles *et al.*, 2003], which resemble a series of tightly packed and regularly spaced test tubes [Araki *et al.*, 1995]. These folds in the lining of the colon generate a large surface area, allowing efficient absorption of nutritional compounds from the intestine. Between the crypts, muscle cells enclose the intestinal mucosa and contribute to the stability of the crypt shape [Drasdo and Loeffler, 2001].

We concentrate on the cells in the lining of the crypt, known as the epithelium, because this is where cancers originate. The epithelium is observed to consist of a single layer of cells which forms a barrier between the body's interior and exterior environment [Galle *et al.*, 2005], and regulates water and nutrient absorption and glandular secretion [Blanpain *et al.*, 2007].

1.1.1 Observed cell dynamics

In this section we discuss the cell dynamics in a single crypt from a healthy colon.

Stem cells are believed to reside near the bottom of the crypt [Potten and Loeffler, 1990; Preston *et al.*, 2003], and these are capable of producing a variety of cell types that are required for tissue renewal and regeneration after injury [Ro and Rannala, 2001]. The stem cells divide to produce transit-amplifying cells that migrate up the crypt wall towards the luminal surface. As the cells proceed up the crypt wall they undergo the process of differentiation which ultimately transforms them into fully-functional cells. Differentiated cells perform many tasks, such as the lubrication of the colon, but do not undergo cell division [Loeffler *et al.*, 1986]. Once at the top, the cells either undergo apoptosis or are shed into the lumen and transported away by the bulk fluid that passes along the top of the crypt [Giles *et al.*, 2003; Nowak *et al.*, 2002]. Apoptosis results in the shrivelled corpse of the cell being engulfed by nearby cells in the tissue, and this material disappears, typically within a day [Hanahan and Weinberg, 2000].

Now we shall examine the cell types in the crypt in more detail, beginning at the bottom and proceeding upwards. The stem cells are characterised by their ability to proliferate and to produce functional progeny while maintaining their own numbers [Potten and Loeffler, 1990; Roeder and Loeffler, 2002]. Since they are defined by their functional capacities, rather than their physical characteristics, it is difficult to identify stem cells without following their progress over time, and the testing procedure inevitably alters their characteristics [Potten and Loeffler, 1990]. Consequently it is unclear how many stem cells exist in each crypt. Recent work by the Clevers group [Barker *et al.*, 2007] has identified a marker gene Lgr5 (or GPR49) that can label stem cells, and it is hoped that this will lead to a clearer indication of the exact number and dynamics of this cell type in the near future. Although the stem cells are believed to comprise a very small proportion of the crypt and their number is observed to remain roughly constant, it is these cells that replenish the whole crypt [Loeffler *et al.*, 1993].

Above the stem cells are the transit-amplifying cells which are in the process of differentiating into functional cells, and these possess some properties of both stem and differentiated cells. It is hypothesised that early transit cells, or progenitors, may be able to act as stem cells if required, while later generations are committed to differentiation [Renehan *et al.*, 2002; Simon and Frisén, 2007]. This is illustrated in Figure 1.2, where the level of 'stemness' is denoted by the colour red, which fades with each division as the cells move up the crypt wall. The process of differentiation is discussed further in Section 1.1.3. It is believed that transit cell migration is predominantly caused by the mitotic force of cell division, although cell movement in the absence of mitosis has been observed to occur to a limited extent [Batlle *et al.*, 2002; Paulus *et al.*, 1993] at 20% of its original value [Paulus *et al.*, 1992], possibly due to active migration linked to EphB/EphrinB gradients [Clevers and Batlle, 2006] or other factors produced by myofibroblasts. It is unclear what proportion of the crypt contains proliferating cells. Some researchers have observed that two thirds of crypt cells are proliferative [Meineke *et al.*, 2001; Potten and Loeffler, 1990], while others report that proliferation is restricted to the lowest third [Katz *et al.*, 2002; van de Wetering *et al.*, 2002] or even the lowest fifth of the crypt [Rozen *et al.*, 2002].

Differentiated cells tend to occupy the uppermost portion of the crypt. The three main lineages of differentiated cell are colonocytes (the principal epithelial cell lineage), enteroendocrine cells (peptide hormone-secreting cells) and Goblet cells (mucous-secreting cells) [Brittan and Wright, 2002, 2004b; Halm and Halm, 2000]. Colonocytes and their pro-

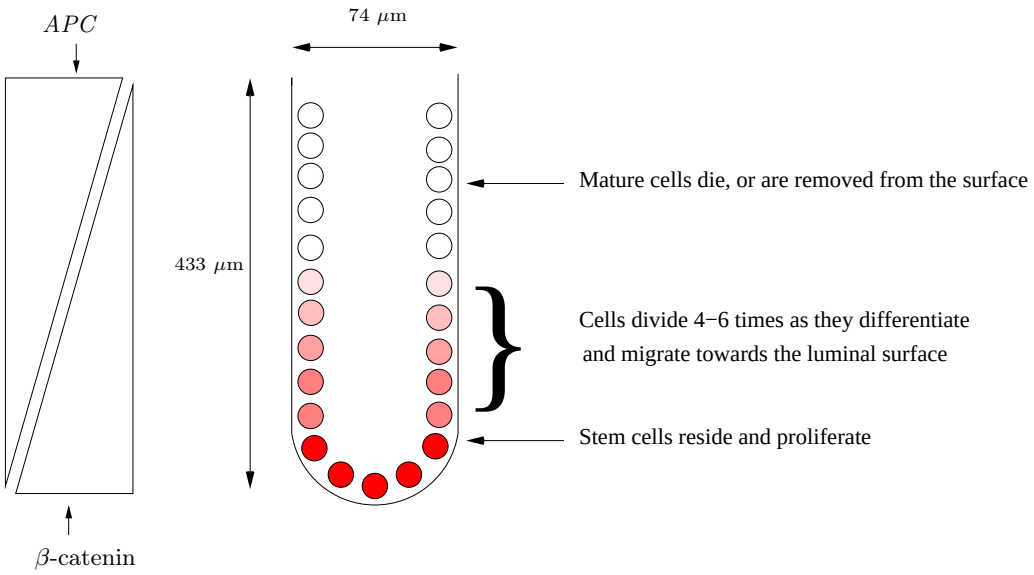

Figure 1.2: Schematic of a human colonic crypt, where the dimensions are given according to Halm and Halm [2000]. The left-hand side indicates a decreasing (increasing) gradient of β-catenin (APC) moving up the crypt wall.

genitors comprise 80% of the epithelium and have both absorptive and secretory functions, while enteroendocrine cells represent less than 1% of crypt cells and secrete hormones [Katz *et al.*, 2002]. Goblet cells secrete mucous to lubricate the bowel and aid the smooth passage of material down the intestine [Ro and Rannala, 2001]. Spatially the Goblet cells are often clustered, and are mixed with columnar cells [Paulus *et al.*, 1993]. An interesting question posed in the literature concerns whether each of the cell lineages have a particular stem cell that forms them. Studies have shown that all the lineages of differentiated cells in the crypt can be derived from one stem cell [Barker *et al.*, 2007; Kirkland, 1988], demonstrating the full potential of stem cells. It is expected that colonocytes and Goblet cells will share an early transit-amplifying cell ancestor, but that enteroendocrine cells do not share any other non-stem cell parent (W. F. Bodmer, pers. comm.). For simplicity we do not incorporate the different types of differentiated cells into our models in Chapters 3–6.

1.1.2 Experimental data: a tale of mice and men

The vast majority of data illuminating crypt cell dynamics tends to describe one of two mammalian systems. Experiments are generally performed on mice and focus on the small intestine, whereas human data are obtained in general from adenomas in the large intestine (colon). Crypt dynamics is observed to be similar in the small intestine and the colon, the main differences being that the small intestine additionally contains villi that extend into the lumen, and small intestinal crypts contain an extra differentiated cell line called Paneth cells which are located at the very bottom below the stem cells [Potten and Loeffler, 1990]. Experimental data for the murine small intestine and the human large intestine are shown in Appendix A.

The murine small intestine has been extensively studied, and detailed data have been produced by the Potten group [Bach *et al.*, 2000; Potten and Loeffler, 1990], amongst others. There are approximately 1×10^6 crypts in the small intestine, each containing

approximately 250 cells [Li *et al.*, 1994; Paulus *et al.*, 1993; Potten and Loeffler, 1990], of which it is estimated that 4–6 [Barker *et al.*, 2007; Bjerknes and Cheng, 2006; Nowak *et al.*, 2002], 4–8 [Paulus *et al.*, 1993], 4–16 [Li *et al.*, 1994; Potten and Loeffler, 1990] or 6 ± 2 (H. Clevers, pers. comm.) are stem cells, and between 150 and 160 are proliferating at any point in time [Gerike *et al.*, 1998; Potten and Loeffler, 1990]. These numbers vary considerably due to the complexities of labelling and tracking cells, particularly since stem cell markers have only recently been identified [Barker *et al.*, 2007]. Crypts contain approximately 16 cells in circumference and have a height of 24 cells [Paulus *et al.*, 1993]. The stem cells are assumed to have a cycle time of between 12 and 32 hours with an average of 24 [Li *et al.*, 1994; Potten and Loeffler, 1990]. The transit cell population has a cycle time of about 11–12 hours, thus cycling approximately twice as fast as the stem cells [Giles *et al.*, 2003; Paulus *et al.*, 1993]. The journey from the crypt base to apex is assumed to take 2–3 days [Okamoto and Watanabe, 2004] or 3–5 days [Barker *et al.*, 2007], all the cells apart from the stem cells will be replaced in this period [Nowak *et al.*, 2002; Ro and Rannala, 2001], and the cells that reach the top of the crypt will have undergone 4–6 divisions [Potten and Loeffler, 1990].

In a human colon there are about 1×10^7 crypts [Nowak *et al.*, 2002; Rajagopalan *et al.*, 2003], and each day a total of 1×10^{10} cells are shed from the colon [Rajagopalan *et al.*, 2003]. The ratio of stem to transit to fully-differentiated cells might be of the order of 40:400:1000 for a human colonic crypt [Bodmer and Tomlinson, 1996]. For modelling purposes it has been assumed that a crypt contains 1000–4000 cells [Nowak *et al.*, 2002] with up to 40 stem cells [Yatabe *et al.*, 2001]. It is believed that the cycling time for columnar and Goblet transit cells is 6 days [Ross *et al.*, 2003], and the stem cell cycle time is 168 hours, although these are calculated based on very limited studies and approximations [Potten *et al.*, 2003].

For the purpose of parameterising our models in Chapters 3–6, we use data from the murine small intestine since this system has been the best characterised.

1.1.3 Differentiation and proliferation

In this section we discuss the processes of differentiation and proliferation in more detail, and explain how they might be linked. Compartment models that describe these processes are reviewed in Chapter 2.

It is clear that the differentiated cells at the top of the crypt have undergone several divisions, which suggests that cells might become more differentiated each time they divide. Potten and Loeffler [1990], however, explained that while differentiation and proliferation might be linked for transit-amplifying cells, they are in general independent and may not affect each other, since stem cells divide but do not mature, while differentiated cells continue to mature when proliferation has ceased. The development in maturity of cells is poorly understood, and could operate via chemical signalling from outside the crypt, or by a chemical gradient of intercellular factors [Potten and Loeffler, 1990]. For example, the Notch signalling pathway is thought to control the terminal differentiation of transit cells, with specific target genes skewing the fate of transit cells towards a particular type of differentiated cell [van Es *et al.*, 2005]. EphB and EphrinB are believed to control cell positioning in the crypt [Blanpain *et al.*, 2007], in particular the ordering of differentiated cells [Batlle *et al.*, 2002; Clevers and Batlle, 2006].

Concentration gradients of the proteins *APC* and β-catenin are thought to influence crypt dynamics. *APC*, or *adenomatous polyposis coli*, is a large protein which appears to control the turnover of β-catenin (which controls epithelial cell-cell adhesion and sig-

nalling to the nucleus for growth stimulation) as it binds to and down-regulates β-catenin. Other functions of APC include cytoskeletal organisation, chromosomal segregation and co-ordinated regulation of cell motility [Fodde, 2003].

It is observed that cells in the proliferative compartment in the lower crypt accumulate nuclear β-catenin, whereas β-catenin is down-regulated in the upper portion of the crypt, which is thought to lead to cell cycle arrest and differentiation [Giles *et al.*, 2003]. In contrast, APC is markedly increased in the upper portions of the crypt, suggesting an increased level of expression with maturation, whereas it is virtually absent in the crypt region where cells are actively dividing. The concentration profiles of these proteins are included for reference in Figure 1.2. This pattern of expression is in agreement with the role of β-catenin signalling in maintaining stem cell properties and controlling differentiation in the intestine [Sancho *et al.*, 2003; van de Wetering *et al.*, 2002]. APC and β-catenin are involved in the Wnt signalling pathway, and it is mutations in this pathway, and a breakdown of these control mechanisms, that lead to tumourigenesis, which we explain in Section 1.2.2.

1.1.4 Cell divisions and the niche hypothesis

Division of stem and transit cells through mitosis is observed to occur either symmetrically or asymmetrically [Potten and Loeffler, 1990]. A symmetric cell division produces two identical daughter cells, either both of the same type as the parent cell or both more differentiated, while an asymmetric division produces one cell that is the same type as the parent and one that is more differentiated [Clevers, 2005; Morrison and Kimble, 2006]. It may be very difficult in practice to distinguish experimentally between the types of division that have occurred [Potten and Loeffler, 1990], since both daughter cells from the division are found to be the same size [Marshman *et al.*, 2002].

It is observed that the stem cell number is approximately maintained at a steady state, and two contrasting theories have been proposed for how this is achieved. The first theory is of 'asymmetric cell division', where all stem cell divisions produce one stem cell which ensures a fixed number of stem cells at any one time. However, this might also be achieved through a combination of asymmetric and symmetric divisions such that there are equal proportions of both types of symmetric divisions. It has been estimated that 95% of divisions are asymmetric and the remaining symmetric divisions maintain the equilibrium by countering any loss due to apoptosis or erroneous extra stem cells produced [Yatabe *et al.*, 2001].

Alternatively, the 'niche hypothesis' suggests that there is a region at the bottom of the crypt in which the stem cells reside, called the stem cell niche, and cells that leave this region become transit cells. The niche provides and maintains an optimal micro-environment for stem cell function [Brittan and Wright, 2004b], and instructs the cells within it to act as stem cells [Kim and Shibata, 2002; Moore and Lemischka, 2006]. The stem cell niche is believed to be formed and maintained by the underlying cells of the mesenchymal lamina propria, where secreted basement membrane factors regulate stem cell behaviour [Modlin *et al.*, 2003]. Functionally, a niche is characterised by its persistence on removal of stem cells and, conversely, if stem cells are extracted from their niche they cease to retain their stem cell potential.

The debate surrounding the niche hypothesis brings into question the clonal origin of crypt cells. Asymmetric divisions will produce 'immortal' stem cells that will remain in the crypt and their lineages will never become extinct, whereas stem cells in niches may expand with symmetric or asymmetric divisions, and a stem cell line could die out. X-inactivation

studies (analysing cell phenotype in a patch, for example Novelli *et al.* [2003, 1996]) and methylation studies (measuring the covalent addition of a methyl group to a DNA residue, for example Yatabe *et al.* [2001]) can be used to test the clonal composition within a crypt or a tumour. It is believed that approximately every 8.2 years all stem cell lineages but one within a crypt niche become extinct [Kim and Shibata, 2002], and so the crypt is maintained by one stem cell lineage. This process is called niche succession or a bottleneck [Kim *et al.*, 2004], and the diversity of a population reflects the time since the last bottleneck. It is therefore predicted that all cells within a crypt are relatively closely related to each other [Kim and Shibata, 2004], although not all data support this claim [Novelli *et al.*, 2003].

We present a model of the clonal composition of the stem cell niche in Chapter 6 to determine the niche succession time with the aim of proving or disproving the niche hypothesis.

1.1.5 Homeostasis mechanisms

There are many complex, regulatory processes that occur in the crypt to maintain an equilibrium of cell numbers, which is referred to as homeostasis [Michor *et al.*, 2004a]. Homeostatic mechanisms may include regulation of cell division, de-differentiation and an ability of each cell to sense the age and type of the other cells around it. There is much evidence that epithelial-mesenchymal interactions aid in this process, and in the regulation of cell proliferation and differentiation [Brittan and Wright, 2004b]. We discuss how homeostasis can be described mathematically in Chapter 4.

Irradiation studies, and examinations after the application of cytotoxic drugs, performed on the murine small intestine illuminate some of the homeostatic mechanisms associated with the stem cells [Paulus *et al.*, 1992]. The earliest observed effect of this treatment is that about 6% of crypt cells die and these are at the crypt base, implying that the irradiation only kills the stem cells. During the first day after irradiation it is observed that cells stop dividing, the number of cells in the crypt reduces by half and the cell migration velocity drops to about 20% of its original value, but cell migration never ceases, even while mitotic activity is suppressed [Loeffler *et al.*, 1988].

Data show that the response from the crypt is rapid, and some recovery is observed within a day after irradiation [Paulus *et al.*, 1992]. Early cell proliferative activity is seen specifically at the bottom of the crypt, where the cell cycle time is shortened dramatically. Cycle times for cells in the middle of the crypt are reduced from 12 to 8 hours, and times for cells at the bottom of the crypt fall from 24 hours to 8 hours [Loeffler *et al.*, 1997]. This results in changes in the self-maintenance of stem cells, and the production of cells for the transit cell population is decreased. This leads to a concept of negative feedback control of proliferation in the crypt. Two days after irradiation the cell velocity is nearly back to normal, and after four days the number of stem cells returns to its pre-irradiation size.

These results suggest that the stem cell population is self-regulating. Feedback is assumed to act on cell cycle times and/or stem cell self-maintenance. Stem cells that are normally immortal, only dividing asymmetrically, may expand through symmetric division to repair the damage induced by mutagenesis [Kim and Shibata, 2002]. It is possible that a single multi-potent surviving stem cell can regenerate an entire crypt [Brittan and Wright, 2004b]. Despite the observed auto-regulation of stem cells, it is unknown how they actually detect that they are present in the correct number. This could be achieved by a short-lived diffusible factor associated with stem cells, and the overall concentration of this factor could indicate the number of stem cells present [Marshman *et al.*, 2002].

There is some limited evidence that transit cells can also de-differentiate and decrease in maturity [Potten and Loeffler, 1990]. At high doses of radiation some crypts will contain no surviving stem cells. If all the stem cells are removed from the crypt then it is proposed that the transit cell population can regenerate the epithelium by returning to a stem cell phenotype [Potten and Loeffler, 1990].

Another example of regulation is in the control of mitosis. DNA studies demonstrate that during asymmetric division of stem cells the newly synthesised DNA, which is more prone to replication-induced mutation, is inherited by the daughter cell and the parent retains the template copy [Marshman *et al.*, 2002]. In this way any errors are passed to the daughter cells which are lost from the tissue within a week, thus protecting the integrity of the genome [Brittan and Wright, 2004b; Marshman *et al.*, 2002]. Also if DNA damage from mitosis is detected, then the p53 tumour suppressor protein elicits either cell cycle arrest to allow DNA repair to take place, or apoptosis if the damage is excessive [Hanahan and Weinberg, 2000].

A counting device for cell generations has been discovered which uses the ends of chromosomes, called telomeres [Hanahan and Weinberg, 2000]. Telomere ends are lost during cell division, and cells cease to divide when the length gets too short [Frank and Nowak, 2004]. Telomeres can be maintained at a length above a critical threshold, which permits unlimited multiplication of descendant cells, but this mechanism seems to be strongly suppressed in most normal human cells in order to deny them unlimited replicative potential.

Overall, the homeostatic processes in the crypt balance the rates of cellular death and proliferation [Brittan and Wright, 2004a]. If this homeostatic mechanism is shifted towards cellular growth, then neoplasia results and neoplasms are formed which are made up of populations of mutant cells undergoing constant turnover [Maley *et al.*, 2004; Michor *et al.*, 2004b].

1.2 Colorectal cancer

In this section we discuss colorectal cancer progression, how the regulatory mechanisms described in Section 1.1 break down, and the biological questions of mutation accumulation that we intend to address using mathematical modelling.

1.2.1 Tumour growth and progression

Tumour growth is initiated by one or more mutations that give a cell a selective growth advantage, which might take the form of increased proliferation or decreased apoptosis [Sieber *et al.*, 2003; Tomlinson *et al.*, 2002, 1996]. Cells with mutations are clonally selected on the basis of their improved fitness, and such cells expand to form a colony of mutant cells. Successive advantageous mutations are followed by waves of clonal expansion [Fearon and Vogelstein, 1990; Rajagopalan *et al.*, 2003; Sieber *et al.*, 2003]. Even if they provide no advantage, the mutations can spread through the tumour cell population by genetic drift [Merlo *et al.*, 2006; Taylor *et al.*, 2003].

There is debate in the literature as to whether mutant cells spread in a top-down or bottom-up direction, which casts doubt over the location of the stem cells in the crypt. Shih *et al.* [2001] observed that within a column of epithelial cells in mutant crypts there is an abrupt change between dysplastic cells in the upper regions and morphologically normal cells at the bottom of the crypt which were genetically unrelated to the cells above them. The authors proposed that either stem cells reside between crypt orifices on the surface of the

This image cannot be made available in this online version of the thesis because of copyright.

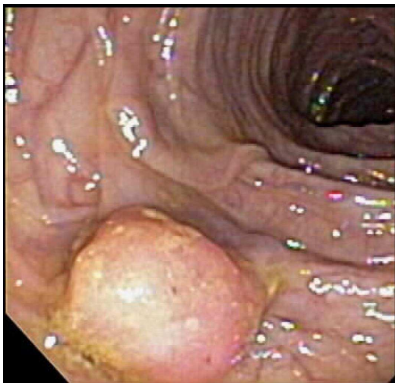

Figure 1.3: Neoplastic growth. Left: A micro-dissected crypt from a human colonic adenoma with asymmetrical branching and multiple budding, reproduced with permission from Preston *et al.* [2003]. Centre: A pre-cancerous colonic polyp, taken from the Biocomputing website www.biiuk.com/biocomputing.html. Right: A photograph of a 1.4 cm transverse colon polyp in a 64-year-old female, reproduced with permission from Summers *et al.* [2005].

colon, contrary to accepted theories, and the mutant cells spread out into the crypts in a top-down direction, or the stem cells originate at the bottom of the crypt, mutant cells migrate to the top and only there are they detected as being dysplastic, before migrating laterally and downwards into neighbouring crypts. Preston *et al.* [2003] suggested, alternatively, that early clonal expansion is through a bottom-up motion and that the downward spread of dysplastic cells is likely to be a phenomenon confined to the later stages of colorectal adenoma development.

As mutant cells accumulate, the crypt becomes dysplastic (abnormally developed). It is proposed that these adenomatous crypts expand through a process of fission [Greaves *et al.*, 2006], where the crypt divides, usually symmetrically at the base, or by budding which can be thought of as a bifurcation at some point along the column wall [Preston *et al.*, 2003]. The mutant crypts accumulate to form a polyp [Rajagopalan *et al.*, 2003], which is a small stalked outgrowth of tissue from the epithelium, often located at the top of the crypts around the intercryptal zone in the small intestine [Giles *et al.*, 2003]. Studies by Araki *et al.* [1995] showed that the majority of small polyps were round or oval in shape, and the average length of the crypt increased from 266 μm to 357 μm in a polyp. Polyps are usually benign but may become malignant, and 10–20% of the larger polyps progress to cancer by acquiring additional mutations [Rajagopalan *et al.*, 2003]. This progression to cancer is illustrated in Figure 1.3, where a budding crypt and polyps from human colonic adenomas are shown.

Some individuals have an inherited predisposition to colorectal cancers, and the two major forms of inherited syndromes in humans are familial adenomatous polyposis (*FAP*) and hereditary non-polyposis colorectal cancer (*HNPCC*) [Kinzler and Vogelstein, 1996]. *FAP* is the more common and patients have a normal mutation rate but inherit a mutation in *APC* meaning that they develop many more polyps, and consequently many more tumours [Michor *et al.*, 2004a; Sieber *et al.*, 2002]. Alternatively, *HNPCC* is a rare syndrome where patients have a raised mutation rate but are no more likely to form polyps and tumours than an individual without *HNPCC* [Sieber *et al.*, 2003; Tsao *et al.*, 2000]. In this thesis, however, we only consider cancers that have occurred sporadically in healthy patients.

Sporadic cancers appear to be controlled by a minor population of undifferentiated cells,

referred to as 'cancer-driving cells' or 'cancer stem cells' [Kern and Shibata, 2007], which are defined by uncontrolled growth, tumourigenic potential and the ability to self-renew [Dalerba *et al.*, 2007; Ricci-Vitiani *et al.*, 2007]. The identity of this population is yet to be uncovered: they could be aberrant stem cells, or even transit-amplifying or differentiated cells that when mutated can mimic stem cell behaviour [Polyak and Hahn, 2005]. Van Es *et al.* [2005] observed the remarkable similarity between transit cells and cancer cells, and the mutant transformation may simply switch off terminal differentiation [Batman *et al.*, 2007]. The possible role of cancer-driving cells in tumourigenesis is discussed further in Chapter 4.

1.2.2 Mutations and the Wnt signalling pathway

Fearon and Vogelstein [1990] proposed that mutations in four or five genes are required for malignant tumour formation, which may involve six rate-limiting steps [Tomlinson *et al.*, 1996]. Loss of APC is thought to be the initiating mutation in the crypt, ras gene mutations may be responsible for enlarging small adenomas, deletion of chromosomes 17p and loss of EphB expression may help convert an adenoma to a carcinoma [Batlle *et al.*, 2005], and loss of chromosome 18q is mainly found in late-stage adenomas and carcinomas [Fearon and Vogelstein, 1990]. However, it is the accumulation of these changes that matters, not the order [Fearon and Vogelstein, 1990], and the sequence of mutations in any given cancer is a unique genetic evolutionary pathway [Bodmer, 1999; Tsao *et al.*, 2000].

Cancer cells have defects in the regulatory circuits that govern normal cell proliferation and homeostasis [Hanahan and Weinberg, 2000]. The Wnt signalling pathway (named after its most upstream ligands, the Wnts) is believed to be one of the first regulatory mechanisms to be broken when cancers start to form. Wnt signalling is involved in the regulation of the complex balance of proliferation, migration and differentiation, and aberrant activation leads to tumour formation [Giles *et al.*, 2003; Ilyas, 2005]. In the absence of a Wnt signal, cells regulate β-catenin levels by a complex containing APC which phosphorylates β-catenin marking it for degradation [Brittan and Wright, 2004b; Clevers, 2004].

It is believed that over 90% of all colorectal cancers have an activating mutation in APC or β-catenin [Homfray *et al.*, 1998; Nowak *et al.*, 2002; Preston *et al.*, 2003; Shih *et al.*, 2001]. Most APC mutations in cancer cells lead to a truncation of the protein, which prevents its regulating function [Michor *et al.*, 2004b]. This mutation activates the Wnt signalling pathway leading to the accumulation of β-catenin in the cell nucleus [McDonald *et al.*, 2006; Sansom *et al.*, 2004]. Activation of the Wnt signalling pathway can lead to a number of outcomes including effects on transcription, cell adhesion, cell migration, and cell polarity which contribute to the invasiveness of a particular tumour. APC loss and subsequent β-catenin stabilisation result in a size increase of the proliferative compartment of the crypt. APC inactivation is thought to enable the cell to stick at the top of the crypt, evade apoptosis and initiate tumour growth [Michor *et al.*, 2004b].

Apoptosis is another key regulatory mechanism to be broken in tumourigenesis. Resistance to apoptosis can be acquired due to a mutation in the pro-apoptotic regulator p53. The normal role of p53 is to detect DNA damage in the form of breaks in the DNA strands, and to arrest cell division until these breaks have been repaired [Bodmer, 1999]. Functional inactivation of p53 is seen in more than 50% of human cancers and results in the removal of a key component of the DNA damage sensor that can induce programmed cell death (PCD) [Hanahan and Weinberg, 2000].

In Chapter 4 we discuss how the effects of mutations can be incorporated into a math-

ematical model.

1.2.3 Genetic instability

It is widely believed that cancer arises because of mutations in susceptible genes. Tumour suppressor genes (TSGs), for example *APC*, act to regulate cell behaviour but mutations can inactivate TSGs leading to loss of function [Fearon and Vogelstein, 1990; Sieber *et al.*, 2003]. A normal cell has two alleles of a TSG, and since TSGs act recessively at the cellular level, it is necessary to inactivate both maternal and paternal copies to eliminate the gene [Fearon and Vogelstein, 1990]. Inactivating the first allele is considered to be a neutral mutation, but inactivating the second provides the cell with an increased net reproductive rate [Nowak *et al.*, 2004]. It is thought that a stem cell must acquire two rate-limiting mutations in *APC*, often referred to as 'hits', for clonal expansion to occur [Knudson, 1996, 2001; Rozen *et al.*, 2002], but it is unclear whether genetic instability (also known as genomic instability) is necessary to promote this process.

There are two main types of genetic instability that have been identified in colorectal cancer tumours. The predominant form is chromosomal instability (CIN), which occurs in 85% of tumours, while the remaining 15% are due to microsatellite instability (MIN) [Giles *et al.*, 2003; Nowak *et al.*, 2002]. Both processes have been observed to occur early in tumourigenesis [Rajagopalan *et al.*, 2003], and result in a raised intrinsic mutation rate [Sieber *et al.*, 2003].

CIN occurs when cancer cells gain or lose whole or large fractions of chromosomes at a greatly increased rate during cell division [Nowak *et al.*, 2002; Rajagopalan *et al.*, 2003]. This results in an imbalance in the number of chromosomes per cell, known as aneuploidy, and an increased rate of loss of heterozygosity (LOH), which means that the cells no longer contain one normal and one mutated gene [Michor *et al.*, 2004a,b]. For genetically stable cells the probability of inactivating a tumour suppressor gene within 5 years is 1%, but if these cells have CIN then the probability is 80% [Nowak *et al.*, 2004].

MIN occurs when somatic mutations inactivate both mismatch repair (MMR) alleles [Tomlinson *et al.*, 1996]. In cells with MIN the point mutation rate is increased 50–1000 fold [Michor *et al.*, 2004b; Nowak *et al.*, 2004].

There is great debate in the literature about whether genetic instability occurs before or after *APC* inactivation and is a necessary component. Some suggest it is required for tumourigenesis [Loeb, 1991; Loeb *et al.*, 2008; Nowak *et al.*, 2002; Rajagopalan *et al.*, 2003], while others propose that natural selection is the driving force for tumour initiation and progression [Bodmer, 2008; Giles *et al.*, 2003; Sieber *et al.*, 2003; Tomlinson and Bodmer, 1999]. While genetic instability can aid tumourigenesis, it is not a universal feature of all tumours [Sieber *et al.*, 2003]. Raised mutation rates can occur by chance, by inheritance, or possibly by a secondary effect of a mutation primarily selected for its effects on cell proliferation or avoidance of death. The argument for genetic instability being the driving force of tumourigenesis is that tumours harbour too many mutations to be explained by any other factor [Sieber *et al.*, 2003].

Komarova and Wang [2004] produced a model for the accumulation of the two *APC* hits, with or without the inclusion of genetic instability, which is summarised in Chapter 2, and we present our own alternative model in Chapter 5.

1.3 Aims of this work

In this chapter we have explained the important biological aspects of the crypt, how homeostasis is maintained and discussed how this regulation is broken down during the onset of tumourigenesis. The aim of this thesis is to use mathematical modelling to capture the cell population dynamics in both a healthy and dysplastic crypt, and to explain how this system permits both a fixed, robust equilibrium in the case of a homeostatic crypt and unregulated growth in cell numbers when tumourigenesis occurs.

In Chapter 2 we review the models presented in the literature, and discuss their relevance to this thesis. In particular we discuss the compartment model proposed by Tomlinson and Bodmer [1995], which we use as a basis for our models in subsequent chapters.

In Chapter 3 we extend the Tomlinson and Bodmer [1995] model to capture more general cases of cell division, and we present three different approaches to modelling the same compartment structure. The discrete model uses difference equations to capture the cell populations at fixed time intervals; the age-structured model uses partial differential equations (PDEs) to follow populations on short time scales; and the continuous model uses ordinary differential equations (ODEs) to follow population changes over a time scale of many cell cycles. We compare these models and derive relations to link their respective parameters. The question we are trying to answer in this chapter is: what is the most appropriate modelling technique for capturing the cell population dynamics in the crypt?

In Chapter 4 we extend the continuous model presented in Chapter 3 to capture homeostasis by presenting two different types of population feedback. We analyse the steady state parameter space, estimate the parameter sizes and conduct a sensitivity analysis in order to ascertain to what extent the results depend on the chosen parameter values. We examine how the models can be used to explain both homeostasis and tumourigenesis, with cancer-driving cells deriving from either stem or transit cells. We also discuss alternative forms of feedback, and how these affect the results. The questions we are attempting to answer in this chapter are: what type of feedback can accurately capture homeostasis in the crypt; what are the parameter sizes and which are the most important parameters in the model; how does our choice of feedback affect the conclusions of the model; and how can a model describe both homeostasis and tumourigenesis?

The continuous approach taken to studying mutations in Chapter 4 is, in fact, inappropriate when considering the formation of mutated populations given that this is a stochastic process. In Chapter 5 we present a stochastic simulation model to describe the accumulation of two mutations in a crypt cell, and compare the results with those obtained by Komarova and Wang [2004]. We extend the Komarova and Wang model to include multiple stem cells, population feedback, death and different cell cycle times, and discuss to what extent their results are altered. The questions we aim to answer in this chapter are: in which population is it most likely that the first two mutational hits occur; and how accurate are Komarova and Wang's model results?

In Chapter 6 we draw together the results from the previous chapters adopting a different approach by considering one-dimensional spatial variations of the cell populations along the crypt axis. We model space dependence by following the cell population volume fractions at each point in the crypt, and we use the spatial model to validate the simpler continuous model from Chapter 3. We include feedback to maintain the expected cell placements, and we discuss which other mechanisms would also achieve this structure. We use this model to track mutant clones as they propagate in the crypt, and model the stem cell niche hypothesis. Throughout this chapter we aim to address the questions: what is the

probability that the progeny of one cell will take over a stem cell niche; following a mutation, how long does it take for the mutant clone to either be wiped out or take over the crypt; and is this modelling framework in one dimension sufficient to answer these questions?

Finally, in Chapter 7, we highlight the main results from this thesis, discuss future modelling opportunities, and suggest possible experiments to verify our model predictions and extend insight into crypt dynamics.

Chapter 2

A review of crypt and colorectal cancer models

Mathematical modelling can help explain both low-level phenomena and system-level behaviour. The more experimental evidence there is available on a research topic, the greater is the potential influence of modelling. A successful model can describe observed behaviour from a qualitative or quantitative analysis, and can show which aspects of the system are more or less influential [Mogilner *et al.*, 2006]. Critical parameters that emerge from modelling studies highlight issues that require further theoretical and experimental work. Mathematical models can inform interdisciplinary research, continually incorporating new information which will help guide experimental design [Gatenby and Maini, 2003]. Many studies have used mathematics to address the question of cancer development, notably Armitage and Doll [1954] and Kimmel and Axelrod [1990]. While models can be used to gain insight into specific aspects of tumourigenesis, a description of the process as a whole is currently well beyond our reach [Weinberg, 2007].

In this chapter we review mathematical models in the literature that have been used to consider both normal crypt dynamics and colorectal cancer progression. Many modelling studies have tried to capture the tightly regulated crypt system, and three common approaches are to use deterministic, stochastic or spatial simulation models, which we discuss in Sections 2.1, 2.2 and 2.3, respectively. In particular, we focus on four specific compartment models in detail which motivate our work in Chapter 3. In Section 2.4 we present more general colorectal cancer models, and in Section 2.5 we discuss how all of these models address the biological questions posed in Chapter 1. For a more in-depth review of the techniques of mathematical modelling applied to crypt dynamics and colorectal cancer, see van Leeuwen *et al.* [2006], and more general mathematical models of avascular tumour growth are presented in Roose *et al.* [2007].

2.1 Deterministic compartment models

Population dynamics has been studied across many areas of mathematical biology, modelling wide ranges of populations from biomolecules, genes and cells to insects and mammals, across areas including cancer biology, ecology, epidemiology and genetics [Murray, 2002]. In this section we discuss four deterministic compartment models that have been used to describe cell population dynamics in the crypt, and we review these models in some detail. In particular, the model by Tomlinson and Bodmer [1995] forms the basis for our study in

the next four chapters.

2.1.1 Tomlinson and Bodmer (1995)

One of the earliest and most influential models of cell population dynamics in the crypt is the compartment model by Tomlinson and Bodmer [1995]. The authors allocate the cells in the crypt epithelium into different compartments and use a difference equation model to predict how altering different parameters can initiate unbounded cell population growth.

Tomlinson and Bodmer assign the cells in the crypt to be either stem cells, semi-differentiated cells (also termed transit-amplifying cells) or fully-differentiated cells, denoted by N_0, N_1 and N_2, respectively, and dead cells are denoted by N_3. These compartments are shown schematically in Figure 2.1. At the end of each cell cycle the stem cells will die, differentiate into transit cells or renew with constant probabilities a_1, a_2 and a_3 respectively, whereas the semi-differentiated cells will die, differentiate into fully-differentiated cells or renew with constant probabilities b_1, b_2 and b_3, respectively, where

$$a_1 + a_2 + a_3 = 1 \qquad \text{and} \qquad b_1 + b_2 + b_3 = 1. \tag{2.1}$$

It is possible that some cells will stagnate and not divide during any one cycle, which could be denoted by a probability a_4, where $a_1 + a_2 + a_3 + a_4 = 1$. However, Tomlinson and Bodmer's model formulation assumes that all the cells will divide or die every cell cycle, and so $a_4 = 0$.

The cell cycle times for stem and semi-differentiated cells are denoted by t_0 and t_1 respectively, and are measured in hours. The fully-differentiated cells do not divide and are assumed to be removed from the system (through death or shedding into the lumen) with probability c, and the authors define a reference time scale t_2 against which N_2 removal can be recorded. The probabilities can also be thought of as the proportions of the cell populations that undergo each process, and it is clear that $0 \leq a_1, a_2, a_3, b_1, b_2, b_3, c \leq 1$.

Assuming that the cells cannot stagnate during the cell cycle, let us denote the probabilities of the divisions $N_0 \longrightarrow 2N_0$, $N_0 \longrightarrow 2N_1$ and $N_0 \longrightarrow N_0 + N_1$ by p_1, p_2 and p_3, respectively, where $p_1 + p_2 + p_3 = 1$. Then $2p_1 + p_3$ corresponds to $2a_3$ and $2p_2 + p_3$ corresponds to $2a_2$ in the Tomlinson and Bodmer [1995] model. The authors assume that all cell divisions are symmetric, which is a reasonable assumption since asymmetric divisions can be thought of as half a symmetric renewal and half a symmetric differentiation.

Tomlinson and Bodmer formulated equations for the cell populations $N_i(G)$ after each stem cell division, where G is the number of stem cell divisions undergone and $i = 0, 1, 2$. Their system of difference equations was given by

$$N_0(G+1) = 2a_3 N_0(G), \tag{2.2}$$

$$N_1(G+1) = 2b_3 N_1(G)\frac{t_0}{t_1} + 2a_2 N_0(G), \tag{2.3}$$

$$N_2(G+1) = 2b_2 N_1(G)\frac{t_0}{t_1} + \left(1 - c\frac{t_0}{t_2}\right) N_2(G). \tag{2.4}$$

For example, (2.2) can be derived from $N_0(G+1) = N_0(G) + a_3 N_0(G) - a_1 N_0(G) - a_2 N_0(G) = 2a_3 N_0(G)$, using (2.1). This means that the number of stem cells at the next generation is the same as the number at the previous generation plus an extra cell for each renewing stem cell, minus the cells that are dying or differentiating. These equations are structurally unstable and neglect the asynchronicity induced if t_0/t_1 is not an integer, as

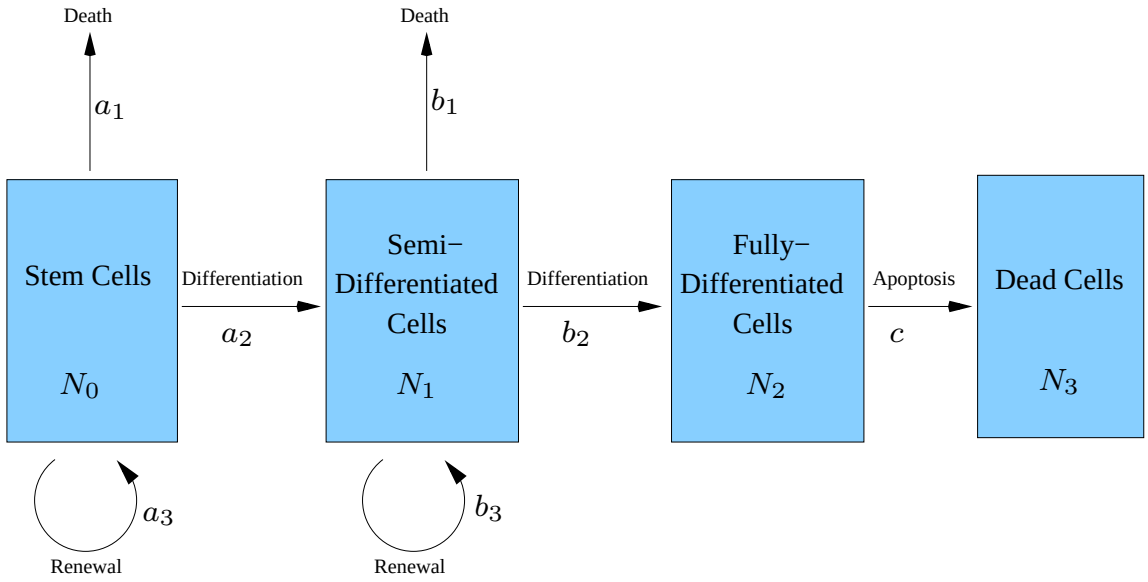

Figure 2.1: The Tomlinson and Bodmer [1995] model compartment structure. The stem cells differentiate into semi-differentiated cells, which in turn differentiate into fully-differentiated cells, which then undergo apoptosis and become dead cells.

well as the compounding effect of semi-differentiated cells cycling more frequently than stem cells. These issues are discussed in more detail in Chapter 3. Despite this, Tomlinson and Bodmer were able to make predictions about tumour growth and development, and these are summarised in the next section.

2.1.1.1 Model predictions

From (2.2), it is clear that the stem cell population can only be in steady state (denoted by N_0^*) if $a_3 = 1/2$, and in this case the steady state is given by the initial number of stem cells. The steady states for the semi- and fully-differentiated cell populations are found in terms of the stem cell steady state to be, respectively,

$$N_1^* = \frac{2a_2}{1 - 2b_3 t_0/t_1} N_0^* \qquad \text{and} \qquad N_2^* = \frac{2b_2}{c}\frac{t_2}{t_1} N_1^*. \tag{2.5}$$

Therefore, for an equilibrium to exist in the semi-differentiated cell population, the stem cells must be in equilibrium and $2b_3 t_0/t_1 < 1$. The stability of the fully-differentiated cell population mirrors that of the semi-differentiated cell population.

By perturbing each parameter individually, the authors found that changes to a_1, b_1, a_2 and b_2 could lead to exponential growth in at least one of the cell populations, while alterations to c lead to a new equilibrium at a higher value than before. However, since the probabilities (or proportions) are always linked by relations (2.1), varying one probability will always result in changes to at least one other, and so changes in any of a_1, a_2, a_3, b_1, b_2 or b_3 could either initiate exponential growth or change the equilibrium state. Tomlinson and Bodmer concluded that failure of apoptosis, or of differentiation, could lead either to exponential growth, or to a new equilibrium at higher cell numbers which is equivalent to a benign growth [Bodmer and Tomlinson, 1996], and that failure of these processes was sometimes sufficient but is not necessary for tumourigenesis, as this could also be achieved

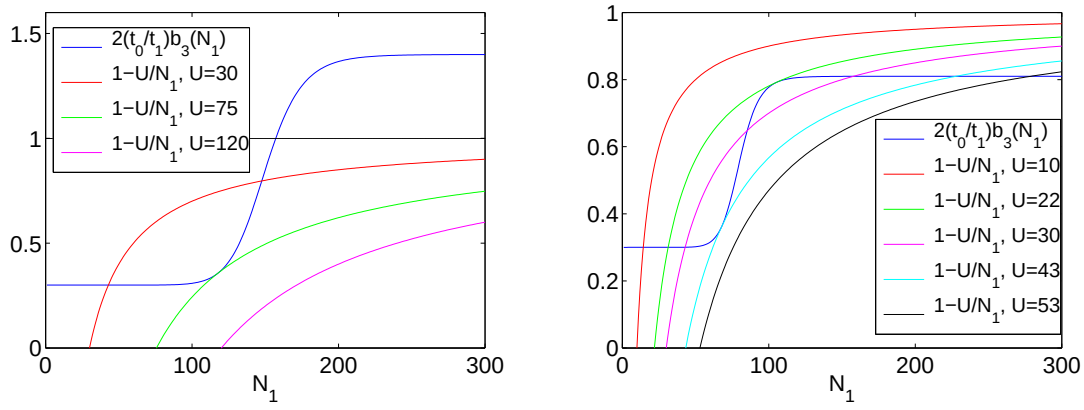

Figure 2.2: Determining the transit cell population steady state using the d'Onofrio and Tomlinson [2007] model by plotting the left- and right-hand sides of (2.7). Left: The case where $2b_3(+\infty)t_0/t_1 > 1$, where we use a Hill function form $b_3(N_1) = 0.075 + 0.275N_1^{12}/(150^{12} + N_1^{12})$ and $t_0 = 2t_1$. Three different values of U (see text) are chosen: 30, 75 and 120. Right: The case where $2b_3(+\infty)t_0/t_1 < 1$, where we use a Hill function form $b_3(N_1) = 0.075 + 0.1275N_1^{12}/(80^{12} + N_1^{12})$ and $t_0 = 2t_1$. Five different values of U are chosen: 10, 22, 30, 43 and 53.

by a proliferative advantage. These observations can be used to explain early neoplastic growths, and the stepwise growth of tumours that occurs between long lag phases.

In Chapter 3 we critically analyse and extend the Tomlinson and Bodmer [1995] model to capture more general cases of cell division, and in Chapter 4 we consider density-dependent feedback that can be used to stabilise the equilibrium to small parameter perturbations.

2.1.2 D'Onofrio and Tomlinson (2007)

D'Onofrio and Tomlinson [2007] extended the Tomlinson and Bodmer [1995] model by characterising the bifurcation between an increase in cell numbers to stable equilibrium or exponential growth. Fluctuations in the model parameters were incorporated, and a nonlinearity was introduced by assuming that the proportion parameters depend on the numbers of cells in each compartment. We note that despite the additions that this makes to the Tomlinson and Bodmer model, the same restrictions on the cell cycle times still apply.

The authors extended the system (2.2)–(2.4), where $b_2(N_1)$ and $b_3(N_1)$ now depend on the transit cell population size and $a_2(G)$ and $a_3(G)$ vary stochastically with each generation G. They assumed that $b_1(N_1)$ is a decreasing function of N_1, $b_2(N_1)$ is a constant or slowly increasing function of N_1 and that $b_3(N_1)$ is an increasing function of N_1, which is approximately constant for small N_1. The only explicit function form prescribed in the paper was

$$b_3(N_1) = b_{3_{min}} + 0.5b_{3_{max}}\left(1 + \frac{2}{\pi}A\tan[A(N_1 - N_{1m})]\right), \tag{2.6}$$

where $A \gg 1$.

Firstly, d'Onofrio and Tomlinson [2007] assumed that $a_3 = 1/2$ and $a_2 =$ constant, meaning that $U = 2a_2N_0(G)$ could be used as a fixed parameter. From (2.3), a semi-

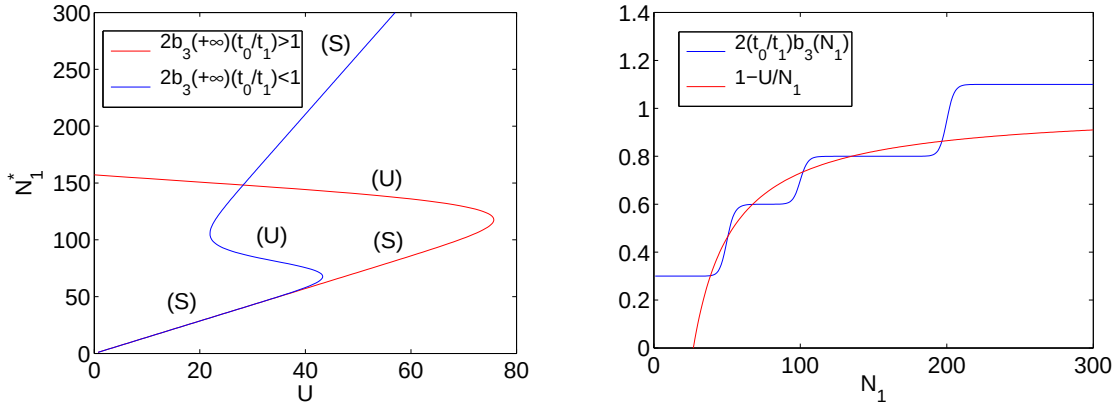

Figure 2.3: Left: A bifurcation diagram for the cases $2b_3(+\infty)t_0/t_1 > 1$ and $2b_3(+\infty)t_0/t_1 < 1$ described in Figure 2.2 using the parameter U. (S) and (U) denote, respectively, the stable and unstable branches of the curve. Right: Determining the semi-differentiated cell population steady states by plotting the left- and right-hand sides of (2.7), for a scenario where multiple steady states exist. We use a sum of Hill functions $b_3(N_1) = 0.075 + 0.075N_1^{20}/(50^{20} + N_1^{20}) + 0.05N_1^{40}/(100^{40} + N_1^{40}) + 0.075N_1^{80}/(200^{80} + N_1^{80})$, along with values $t_0 = 2t_1$ and $U = 27$.

differentiated cell population steady state N_1^* is given by

$$2\frac{t_0}{t_1}b_3(N_1^*) = 1 - \frac{U}{N_1^*}. \tag{2.7}$$

The model dynamics is different depending on whether $2b_3(+\infty)t_0/t_1$ is greater than or less than 1. If $2b_3(+\infty)t_0/t_1 > 1$ then there is a saddle node bifurcation with at most two steady states, but if $2b_3(+\infty)t_0/t_1 < 1$ then there is a hysteresis bifurcation with up to three steady states, depending on the value of U. This is illustrated in Figure 2.2, where we plot both the left- and right-hand sides of (2.7) for different values of U, and a bifurcation diagram for both cases is shown in Figure 2.3.

Secondly, the bifurcations may be driven by variation in the number of stem cells. The authors assumed that $a_2(G)$ and $a_3(G)$ were Gaussian random variables, where $\mathbb{E}[a_3(G)] = 0.5$. Results from simulations suggested that unbounded growth may be driven by noise, and therefore the time at which growth starts may not be calculated deterministically. The authors showed that noise was sufficient to switch between different stable steady states, and in general makes exponential growth more likely if the fluctuations are large enough for a sufficient period of time.

D'Onofrio and Tomlinson [2007] also postulated that more complex behaviour might be possible, with multiple stable equilibria. In particular, if $b_3(N_1)$ is an increasing function of N_1 with $m \geq 2$ inflection points, and $2b_3(+\infty)t_0/t_1 > 1$, then there may be $2n$ equilibria (alternating in stability) of which n are stable, with $n-1$ hysteresis bifurcations followed by a saddle node bifurcation. In this case it might be possible to explain the transition between different steady states by parameter fluctuations alone, leading to the long lag phases of tumour growth. Multiple inflection points might be achieved using a sum of Hill functions, which is illustrated in the right panel of Figure 2.3, and is discussed further in Chapter 4.

Alternatively, the authors tested other forms of feedback where the parameters were dependent on functions of $N_1 + N_2$, N_1/N_0 or $N_1/(N_1 + N_0)$. In all cases qualitatively similar bifurcation curves were obtained, and simulations showed the same behaviour as with the original feedback.

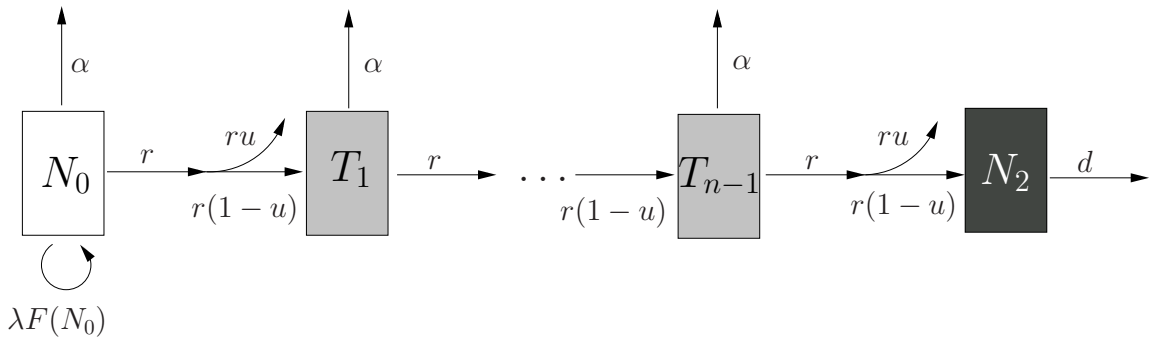

Figure 2.4: The Wodarz [2007] model compartment structure. This is similar to the Tomlinson and Bodmer [1995] model, but the transit cell compartment is split into $n-1$ sub-compartments denoted by $T_1, \ldots, T_{n-1}$. Stem cells renew symmetrically at a rate $\lambda F(N_0)$ and asymmetrically at a rate r. The transit cells divide symmetrically at a rate r, and the fully-differentiated cells die at a rate d. Cells are mutated at cell division (at a rate u) or independent of cell division (at a rate α). As before, N_0 denotes the stem cell compartment and N_2 the fully-differentiated cell compartment.

2.1.3 Wodarz (2007)

Wodarz [2007] proposed a similar compartment model to Tomlinson and Bodmer [1995] to investigate how different cellular turnover rates influence the ability of genetic alterations to induce symptoms of ageing and contribute to the development of cancer. The key differences between the models are that Wodarz includes multiple transit cell sub-populations, feedback in the stem cell population, and mutations at or after cell division.

It is assumed that n divisions are required for the progeny of a stem cell to terminally differentiate, and so the transit cell population is split up into $n-1$ sub-populations denoted by $T_1, \ldots, T_{n-1}$, where the i^{th} population denotes the cells that have undergone i divisions since leaving the stem cell compartment. This compartment structure is illustrated in Figure 2.4.

The author assumes that the stem cells undergo asymmetric division at a rate r and symmetric division at a density dependent rate $\lambda F(N_0)$, where λ is the intrinsic rate of stem cell division, for some function F. The transit cells divide symmetrically at a rate r, and the fully-differentiated cells die at a rate d. A central component of the model is that cells can undergo mutations, which either occur at cell division (at a rate u) or are independent of cell division (at a rate α), and might result from exposure to radiation or chemicals, for example. This model does not account for different cell population cycle times, and there is no cell death other than after mutations. An ODE model was adopted, where the $n+1$

equations are given by

$$\frac{\mathrm{d}N_0}{\mathrm{d}t} = N_0\left[\lambda(1-u)F(N_0) - (ru+\alpha)\right], \tag{2.8}$$

$$\frac{\mathrm{d}T_1}{\mathrm{d}t} = r(1-u)N_0 - (r+\alpha)T_1, \tag{2.9}$$

$$\frac{\mathrm{d}T_2}{\mathrm{d}t} = 2r(1-u)T_1 - (r+\alpha)T_2, \tag{2.10}$$

$$\vdots \qquad \vdots \tag{2.11}$$

$$\frac{\mathrm{d}T_{n-1}}{\mathrm{d}t} = 2r(1-u)T_{n-2} - (r+\alpha)T_{n-1}, \tag{2.12}$$

$$\frac{\mathrm{d}N_2}{\mathrm{d}t} = 2r(1-u)T_{n-1} - (d+\alpha)N_2. \tag{2.13}$$

The density-dependent feedback was chosen in the form of logistic growth, $F(N_0) = 1 - N_0/k$, where k is the stem-cell carrying capacity. Logistic growth is a common choice, and is also assumed in a similar model by Hardy and Stark [2002].

Wodarz found that if mutations occurred independently from cell division, a trade-off is observed between reducing the occurrence of tissue ageing and the development of cancer. While the degree of ageing is reduced by a high cellular turnover rate, the development of cancer is promoted. Therefore, there must be an intermediate cellular turnover rate which optimises this trade-off and maximises the life span of an organism. Alternatively, if mutations occur upon cell division such a trade-off does not exist, and a higher cellular turnover rate results both in higher degrees of tissue ageing and a higher chance of developing cancer. Thus, the best strategy is to have a low turnover rate where stem cells divide rarely, and the differentiated cells are long-lived.

2.1.4 Boman et al. (2001)

Another model that uses feedback to maintain structural stability in the cell populations was presented by Boman *et al.* [2001]. The authors used an ODE model to simulate cellular dynamics in the colonic crypt, and deduced that tumour initiation in the colon is caused by crypt stem cell over-production. The same system of stem, semi- and fully-differentiated and dead cell populations was used as in Tomlinson and Bodmer [1995], and this was additionally coupled with cell cycle compartments for G_1, S, G_2 and M. Pathways for de-differentiation from transit cells to stem cells were also included in the model. The compartments are summarised in Figure 2.5, where $k_1, \ldots, k_{10}$ represent the rate parameters for the processes.

Two types of feedback loops are included in the model, and the authors propose that fully-differentiated and dead cell population sizes control the dynamics. A negative feedback mechanism allows the number of fully-differentiated cells to regulate the steps $N_1 \to S$ and $G_1 \to S$ by the expression $1/(1+N_2)$, and a positive feedback mechanism allows the number of dead cells to regulate the steps $G_1 \to N_1$, $N_1 \to N_2$ and $N_2 \to N_3$ by the expression $(1+N_3^2)$. The authors began with arbitrary parameter values for k_i and the initial number of stem cells ($\hat{n}_0$) in simulations, and used an iterative approach to obtain the most suitable

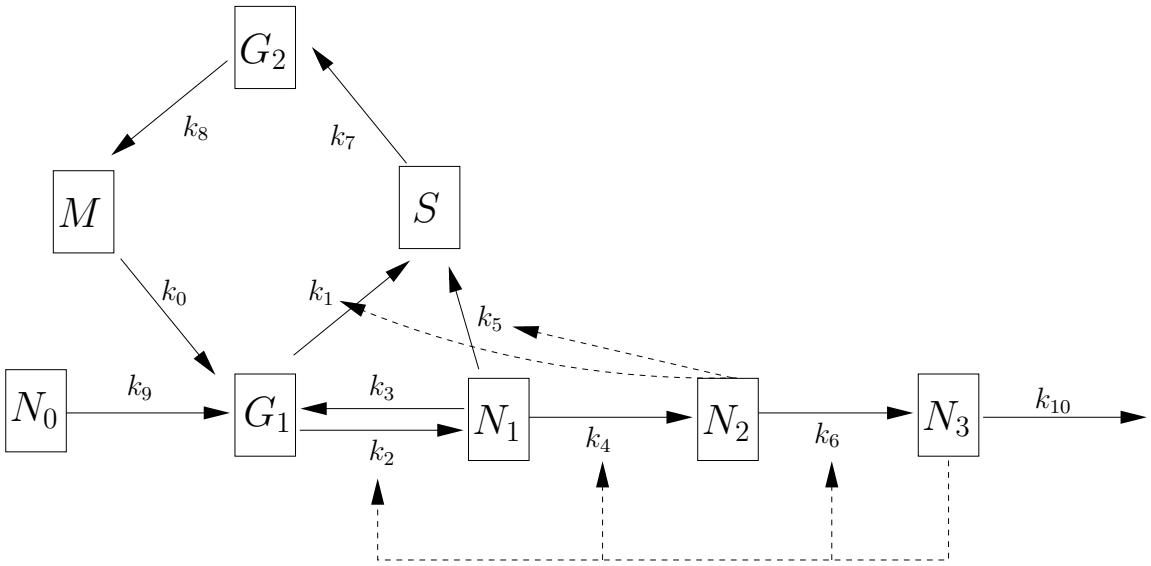

Figure 2.5: The Boman *et al.* [2001] model compartment structure. Cell population compartments are for stem cells (N_0), transit cells (N_1), differentiated cells (N_2), apoptotic cells (N_3), in addition to the four cell cycle stages G_1, S, G_2 and M. Solid arrows denote the main cellular processes of proliferation, differentiation and apoptosis, and dashed arrows denote feedback mechanisms.

parameter set. The model equations were

$$\frac{\mathrm{d}N_0}{\mathrm{d}t} = -k_9 N_0, \tag{2.14}$$

$$\frac{\mathrm{d}N_1}{\mathrm{d}t} = k_2 G_1(1+N_3^2) - k_3 N_1 - k_4 N_1(1+N_3^2) - \frac{k_5 N_1}{1+N_2}, \tag{2.15}$$

$$\frac{\mathrm{d}N_2}{\mathrm{d}t} = k_4 N_1(1+N_3^2) - k_6 N_2(1+N_3^2), \tag{2.16}$$

$$\frac{\mathrm{d}N_3}{\mathrm{d}t} = k_6 N_2(1+N_3^2) - k_{10} N_3, \tag{2.17}$$

$$\frac{\mathrm{d}M}{\mathrm{d}t} = k_8 G_2 - k_0 M, \tag{2.18}$$

$$\frac{\mathrm{d}G_1}{\mathrm{d}t} = k_9 N_0 + 2k_0 M + k_3 N_1 - \frac{k_1 G_1}{1+N_2} - k_2 G_1(1+N_3^2), \tag{2.19}$$

$$\frac{\mathrm{d}S}{\mathrm{d}t} = \frac{k_1 G_1}{1+N_2} + \frac{k_5 N_1}{1+N_2} - k_7 S, \tag{2.20}$$

$$\frac{\mathrm{d}G_2}{\mathrm{d}t} = k_7 S - k_8 G_2. \tag{2.21}$$

We note that the forms of the feedbacks are not motivated biologically, and the system will not yield analytical solutions. In addition we note from (2.14) that the stem cell population will decay unless the rate at which stem cells enter the G_1 phase is zero, which is not biologically feasible.

Perturbing each parameter in turn, Boman *et al.* [2001] found that only an increase in the initial number of stem cells (and not in the rate of cell cycle proliferation, differentiation or removal) caused simulation results to match biological data on the percentage of cells

in the S phase along the crypt axis in *FAP* patients. However, it is not clear how the spatial interpretation was deduced from the model since there is no spatial component in the model formulation. The authors concluded that an expansion of the stem cell population is sufficient to explain the observed proliferative abnormality in *FAP* patients.

2.2 Stochastic models of mutations

Deterministic models can provide an idea about global population changes and steady state conditions, but when events (such as mutations) occur randomly and affect only one cell, a stochastic model is required to capture the specific behaviour at these time and length scales.

In this section we consider some stochastic models of the crypt that have been presented in the literature. The "Nowak/Komarova/Michor" group have produced a series of models which we review in Section 2.2.1, and other relevant stochastic models are referenced in Section 2.2.2. A review of stochastic modelling techniques was presented by Erban *et al.* [2007].

2.2.1 The Nowak/Michor/Komarova group

Nowak, Michor, Komarova and co-workers have produced a set of models that consider the sequence of mutations that initiates colorectal cancer in order to explore the role of genetic instability in the early stages of tumourigenesis. It is unclear whether or not genetic instability is necessary for initiating tumourigenesis or whether it occurs as a result of *APC* inactivation, which was discussed in Chapter 1, although it was argued strongly by Bodmer [2008] that the latter scenario is not likely. The authors use probabilistic models and compare different scenarios and different sets of parameter values.

Nowak *et al.* [2002] calculated conditions for CIN to initiate tumourigenesis before the inactivation of *APC*. The model predicted that if there is a significant selective cost for CIN then the first event will be a point mutation in *APC* followed by a CIN mutation, followed by LOH in *APC*. If CIN is effectively neutral, then the first event could be either the CIN mutation or a point mutation in *APC*, followed by LOH in *APC*, although it is argued elsewhere that the first *APC* mutation is not neutral [Tomlinson and Bodmer, 1999]. The model does not distinguish between different cell types, and it assumes that the probability of fixation of a cell with two *APC* mutations is close to 1, which is not expected to be the case. The most radical interpretation of this model is that CIN mutations seem to initiate colorectal cancer.

Nowak *et al.* [2004] used the same model as Nowak *et al.* [2002] to calculate how the kinetics of TSG inactivation depends on the cell population size and the mutation rates for two hits. The authors derived a probability that at least one cell in a population has two hits at time t. Results show that in small, medium and large populations it takes respectively 2, 1 and 0 rate-limiting steps to inactivate a TSG. The model does not take into account the differences between stem and transit cells, which would significantly alter these findings. Michor *et al.* [2005] then extended this model to include four genetic alterations: inactivation of two alleles of *APC*, activation of ras and the inactivation of p53. The authors believe that the first *APC* allele loss is a neutral mutation, and that CIN accelerates the rate of loss of the second allele. The model is used to predict that in a group of 1000 70 year olds, there would be 9 cases of colorectal cancer, 8 of which would have been initiated by CIN. This model is dependent on the first *APC* mutation and the cost of CIN both

being neutral, which are not necessarily realistic since, in particular, CIN can cause an increase in the death rate. Only stem cells are assumed to become cancer cells, although this assumption is questioned by the results of the model in Section 2.2.1.1.

Michor *et al.* [2004b] studied the consequences of mutations in different cell types to calculate the rate of colorectal cancer initiation and conditions for CIN to precede *APC* inactivation, assuming a linear flow from stem cells to differentiated cells to apoptosis. Results showed that inactivation of the first *APC* allele must occur in a stem cell, otherwise the mutated cells are rapidly lost from the crypt, but other mutations could occur in stem or transit cells. The authors concluded that a small number of CIN genes is sufficient to ensure that CIN occurs before *APC* inactivation, meaning that it is very likely that a CIN mutation initiates tumourigenesis. The analysis relies on the assumption that the doubly mutated cell will stick in the crypt, although there is little evidence to support this claim.

Komarova and Cheng [2006] presented a finite branching process in order to determine the optimal number of stem cells in the crypt to minimise the likelihood of mutations. The authors found that if mutation rates are higher for stem cells than transit cells then the optimal strategy is to have as few stem cells as possible, otherwise as many stem cells as possible is the optimum, although there is limited biological data in support for this (W. F. Bodmer, pers. comm.). There may be a high optimal number of stem cells initially which may reduce over time, which is a strategy that may protect the young from cancer but will be detrimental in old age. The authors concluded that the optimal number of stem cells is determined by the balance between the stem and transit cell mutation rates, and may vary between the young and old. However, there is little justification for different mutation rates between stem and transit cells. The calculations additionally do not take into account different cycling times for the different cell types.

Komarova [2006] extended the previous models by considering a one-dimensional spatial generalisation of a constant population continuous-time discrete state-space birth-death process. The results showed that fixation of a one-hit mutant will occur faster in a space-free model if the mutant is either advantageous or disadvantageous. For a neutral mutation, the fixation rate is the same with or without space. However, for a two-hit model, the generation of a double-hit mutant occurs faster in the spatial model than in the non-spatial model. Hence, previous results from non-spatial models give an under-estimation for the rates of cancer initiation when the first event is the inactivation of a TSG. One simplification of the one-dimensional spatial model is that the number of cells is fixed, and cell division only occurs to replace a dead cell, which is unrealistic.

2.2.1.1 Komarova and Wang (2004)

There is much debate in the literature about where the mutational hits occur in the crypt, and in which cell population they arise. It is widely believed that genetic changes can only persist in stem cells, since these are the only permanent residents of the crypt, and that mutations in transit cells will be shed from the crypt within a few days unless the mutation has other affects that prevent this [Potten and Loeffler, 1990].

Komarova and Wang [2004] presented a model to determine where the two *APC* hits are most likely to occur. The authors assume that the first hit confers no selective advantage to the cell, and once the second *APC* mutation has occurred the cell remains in the crypt and is immune from being sloughed. Three scenarios are considered for how the mutations could accumulate. Firstly, a mutation could occur in a stem cell, and then, after a few divisions, the entire crypt will consist of mutated cells. At some point, a second mutation occurs

in a stem cell which results in the entire crypt containing double mutants. This pathway predicts that the crypt will be monoclonal with respect to double mutations. Secondly, the first mutation could occur in a stem cell which then spreads throughout the crypt, but the second mutation occurs in the transit cell compartment. This mutant divides and its progeny spread up the crypt axis, meaning that the lower part of the crypt is monoclonal with respect to one *APC* mutation and the upper part is monoclonal with respect to two mutations. Thirdly, a mutation could occur in one of the transit cells, and as its progeny moves up the crypt, but before the mutant cell undergoes apoptosis, one of the mutant daughter cells receives a second hit. In this scenario the lower part of the crypt consists of wild type cells, which are not mutated, and the upper part is composed of monoclonal double-mutants.

The results from the model formulae and simulations suggest that at least one of the mutations may occur in the transit cells, and there are scenarios where both hits are likely to occur in transit cells if the rate of loss of the second allele is increased by CIN. This means that the possible role of transit cells in the mutation process cannot be discarded simply because they are short-lived [Komarova, 2005]. Since a transit cell is the most likely target of the second hit, cells below it will be phenotypically different.

Komarova and Wang [2004] considered one stem cell and its progeny, and assumed that a full crypt could be modelled by summing multiple sets of these clonal populations, but this neglects the idea of niche succession where the stem cell pool could become monoclonal with respect to the first mutation. We consider the Komarova and Wang [2004] model in more detail in Chapter 5, where we relax some of the assumptions and test the predictions.

2.2.2 Other stochastic models

Other groups have also used stochastic and/or probabilistic models to capture crypt behaviour.

Luebeck and Moolgavkar [2002] used a probabilistic multistage model to fit to data on incidence rates of colorectal cancers. Their results are consistent with two rare events, interpreted as the loss of *APC*, followed by a high-frequency event in the conversion of a normal stem cell into a mutant that clonally expands to give rise to an adenomatous polyp. Following this, the model predicts that only one more rare event is necessary for malignant transformation. These results agree with Tomlinson and Bodmer's predictions that genomic instability is not necessary to explain colorectal cancer incidence rates in human populations, although genomic instability could still occur later in the process.

Little and Li [2007] compared the models by Nowak *et al.* [2002] and Luebeck and Moolgavkar [2002] and found equivalently good fits using data from colorectal cancer incidence rates, which suggests that models that do not assume a role for genomic instability are equally as plausible as those that do. New methods, for example incorporating quantitative information on exposure to various mutagenic agents, might be necessary to discriminate between these models.

Beckman and Loeb [2006] presented a deterministic model to calculate the probability of cancer arising at a fixed time within a single cell lineage with or without the help of a mutator mutation which raises the mutation rate and is an extra mutation to be achieved. The authors determined conditions on the parameters for mutator mutations to be the more likely pathway. Beckman and Loeb predicted that mutator mutations are important in most carcinogenic scenarios, and are of the greatest importance when they are the initial step in carcinogenesis, although this model does not take account of the possibility of selective

advantage in early mutations.

Other probabilistic models have been used to investigate steady state phenomena that are not captured by deterministic models. Clayton *et al.* [2007] used a probabilistic model of epidermal homeostasis to show that the full range of differentiated cells can be reproduced using just the stem cell compartment, without the need for including transit cells. Hotton and Colvin [2007] produced a discrete stochastic model for populations of stem and differentiated cells which assumed fixed probabilities for renewal and differentiation. The authors noted that allowing for a small amount of de-differentiation in the model can have a large effect on stem cell population growth. It is believed that de-differentiation can be an exceedingly rare event, which might explain the difficulty of observing it *in vivo*, but it could still serve as an important role in stem cell replenishment.

2.3 Spatial simulation models

Computer simulation models are often used to capture spatial effects in the crypt. Two approaches are to use a cellular automaton with a fixed two-dimensional grid to represent each cell position [von Neumann, 1966], or to use a lattice-free model which allows cells to move freely. The grid models often assume that the crypt can be modelled as a cylinder and unwrapped into a rectangular sheet of cells, which can lead to crucial errors in over-counting the stem cells at the base where the crypt should be tapered.

2.3.1 The Potten group

Potten and co-workers have used Monte Carlo stochastic simulation models based upon the behaviour of individual cells, which can incorporate experimental data, represent homeostatic crypt dynamics and predict spatial behaviour patterns of the cells [Gerike *et al.*, 1998; Loeffler *et al.*, 1993, 1997, 1988, 1986; Meineke *et al.*, 2001; Paulus *et al.*, 1993, 1992; Potten and Loeffler, 1987].

Loeffler *et al.* [1986] used a cellular automaton grid of 16×24 cells to represent a crypt, with rows one and two containing just two and four cells, respectively, to taper the crypt base. 16 stem cells were assumed, with four subsequent generations of transit cells before terminal differentiation. Results suggest that cells have an ability to sense the maturity or age of their neighbours and divide in such a way as to displace older cells to higher cell positions up the crypt.

Since the crypt shrinks in circumference after mitotic arrest but cell migration is still observed, it is proposed that a counteracting force squeezes the crypt inwards resulting in the fusion of columns, which provides the force for vertical migration [Potten and Loeffler, 1987]. The normal crypt geometry may be in dynamic equilibrium between lateral merging and mitotic expansion [Loeffler *et al.*, 1988].

Paulus *et al.* [1993] used the Loeffler *et al.* [1986] simulation model to investigate the origin and lineage development of Goblet cells. They concluded that Goblet and columnar cells share a common stem cell in the first few transit cell generations, and the irreversible decision for a cell to attain Goblet properties is stochastic, occurring two or three generations from the end of the lineage.

Paulus *et al.* [1992] included dynamic self-maintenance in the original model in order to investigate post-irradiation recovery data. Results showed that stem cell self-regulation could explain the early changes after irradiation when the damage primarily affects the

crypt, and first generation transit cells could adopt stem cell properties under some conditions. This provides an alternative to the previous assumption of villus-crypt feedback in the small intestine.

Loeffler *et al.* [1993] showed that the cellular regeneration of intestinal crypts can be explained by several indistinguishable stem cells which can replace each other using symmetric divisions. In addition, Loeffler *et al.* [1997] found that experimentally observed behaviour can be reproduced by the model allowing for fluctuations in the number of stem cells, meaning that one stem cell can be replaced by the progeny of another leading to monoclonality. The decision to divide symmetrically or asymmetrically is assumed to be stochastic in this model, and depends on the number of stem cells in the crypt. Gerike *et al.* [1998] showed that this model is consistent with experimental observations when a hypothetical intra-epithelial growth factor is included and individual cells can divide if a critical growth factor concentration is exceeded.

These models have used a rigid cell matrix which may distort cell migration, so Meineke *et al.* [2001] presented a new form of the simulation model using dynamic movement on a lattice-free cylindrical surface. Here cell movement is a result of mitotic activity and cells interact by viscoelastic forces. This supersedes the previous models because local cell sorting rules were not required to capture the experimentally-observed cell dynamics. An extension of this model is being developed by the Integrative Biology Project[1], which is discussed further in Chapter 7.

2.3.2 Other spatial models

Lobachevsky and Radford [2006] showed that a two-tiered model of stem cells that have a finite replicative life span is consistent with experimental data. The model assumes that a crypt contains a quiescent ('deep') stem cell and a few more actively cycling ('proximate') stem cells. Monte Carlo simulations of mutagenesis data are used to test the model, and suggest that the proximate stem cell cycle time is about 80 hours, assuming a replicative life span of 50–150 divisions, and that the deep stem cell divides approximately every 30 weeks.

Other spatial models have considered cancer progression. Spencer *et al.* [2004] used an ODE model describing the multi-step transformation to cancer to predict the most likely order of acquiring specific mutations. The authors tested the ability to induce angiogenesis (A), avoid death (D), acquire genetic instability (G) and possess an increased replication rate (R). Results showed that the overwhelmingly most likely pathway was $D \rightarrow R \rightarrow A \rightarrow G$. This model was extended by Spencer *et al.* [2006] by using a three-dimensional agent-based model that resembles a stochastic cellular automaton. The results predicted that early-onset cancers proceed through a different sequence of mutations than late onset cancers, and that tumour heterogeneity varies with acquisition of genetic instability and mutation pathway. The authors concluded that genetic instability is often the initial mutation in earlier onset tumours, while limitless replicative potential is often the first mutation in later onset tumours.

Maley and Forrest [2000] used an agent-based model to calculate the probability of developing cancer via evolutionarily neutral or selective mutations. The simulations suggested that there must be at least two selectively neutral mutations necessary for the development of cancer.

[1] See http://www.integrativebiology.ox.ac.uk/ for details of the Integrative Biology Project.

2.4 Other models of colorectal cancer

Other models of cancer progression have considered different spatial scales.

Bjerknes [1996] presented an ODE compartment model of normal and mutant crypts to estimate the rate of expansion of the mutant stem cell population. The model predicts that the mutant stem cell pool grows significantly faster (greater than 40 times faster) than a normal stem cell pool, which might mean that mutant stem cells are more likely to divide symmetrically to produce two stem cells.

Nagy [2004] used an ODE model to explore the role of natural selection in later-stage tumour evolution. Results predicted the possibility of a hypertumour, which is a focus of aggressively reproducing cells that invade and destroy part or all of the tumour.

2.4.1 Mechanical models and morphogenesis

Mechanical models for the colon have been used to describe morphogenesis, which is the folding of epithelial sheets of cells [Odell *et al.*, 1980, 1981]. Cell junctions are attached to arrays of microfilament bundles which can contract and change the shape of the cell, which in turn causes neighbouring cells to contract, sending contraction waves across the sheet to shape the epithelium. The initial contraction is almost certainly initiated by a chemical signal.

Edwards and Chapman [2007] adopted a continuum model where the crypt epithelium is modelled as a growing beam attached to the underlying lamina by cell bonds, which generate tension within the layer. The authors showed that buckling of the tissue may be generated by an increase in net proliferation, an enlargement of the proliferative compartment, an increase in the strength of the cellular attachment to the underlying lamina, or a change in the rate of cell growth or bonding. Crypt fission was also addressed by Drasdo and Loeffler [2001], who used a single cell based lattice-free model to explain epithelial buckling.

2.4.2 Other work by Tomlinson and Bodmer

Tomlinson and Bodmer have constructed simple arguments to calculate the number of mutations required for a cancer, and determine the role of genetic instability in cancer.

Tomlinson *et al.* [1996] believe selection is more likely than an increase in the mutation rate to be the driving force of sporadic tumourigenesis. Some cancers may acquire a mutator phenotype, probably leading to faster growth, but mutator phenotypes are not necessary for carcinogenesis.

Estimates of the number of mutations in a cancer have varied widely in different models. It has been suggested that there may be 10,000 mutations of an insertion/deletion type in the DNA in a benign or malignant colorectal tumour, which some commentators have found surprisingly high. However, Tomlinson *et al.* [2002] predicted the number of mutations to be several orders of magnitude higher: 10^{12} detectable mutations in an adenoma and 10^{14} in a carcinoma, and cancers with mutator phenotypes will have even more mutations. The number of mutations seems very high, but these occur due to the accumulation of somatic mutations by normal stem cells over time, the large number of generations for most tumours to grow, and the large number of cells comprising the tumour itself. Many cancers harbour multiple mutations, the great majority of which probably have no significant effect on tumour growth.

In order to test whether tumours grow in a cell-autonomous fashion, Tomlinson and Bodmer [1997] modelled two non-autonomous strategies that tumour cells may use. The

strategies assume that tumour cells produce firstly an angiogenesis growth factor that benefits both the cell from which it originates and neighbouring cells, and secondly growth factors to prevent apoptosis. The authors conclude that many of the mutations observed in tumours may have non-autonomous effects.

2.5 Discussion

In this chapter we have reviewed a cross-section of the mathematical models presented in the literature that address biological questions regarding healthy crypt dynamics and dysplastic growth.

In Section 2.1 we discussed deterministic compartment models of the crypt. Such models are generally used to capture qualitative steady state dynamics, and also attempt to explain the alterations required to initiate carcinogenesis. Analytic formulae, in particular, are useful for determining which processes are dominant and which affect the system to a lesser degree. In Chapter 3 we present a compartment model of the cell population dynamics in the crypt which is an extension of the model by Tomlinson and Bodmer [1995], and discuss whether a discrete, age-structured or continuous model is most appropriate. In Chapter 4 we extend this model to include feedback in order to more accurately capture both homeostasis and tumourigenesis. We discuss our proposed feedback in relation to that in the models by d'Onofrio and Tomlinson [2007], Wodarz [2007] and Boman *et al.* [2001].

In Section 2.2 we summarised some of the stochastic models in the literature. These are often used to model mutation accumulation, and in particular investigate the role of genetic instability in cancer initiation. We present our own stochastic simulation model in Chapter 5, where we relax the assumptions in the model by Komarova and Wang [2004] to investigate the occurrence of the first two mutational hits.

Models that have considered spatial effects in the crypt were reviewed in Section 2.3, in particular the two-dimensional grid and lattice-free models that are adopted by, among others, Potten and colleagues. These models are generally used to investigate steady state crypt behaviour such as migration, lineage differentiation and stem cell self-maintenance. However, most of these models assume that cell motion results from mitotic activity and cannot capture the movement that is observed when mitosis is inhibited. In Chapter 6 we extend our model presented in Chapters 3 and 4 to include spatial variation, where we restrict ourselves to considering only one dimension in order to test if this is sufficient to answer key biological questions.

Finally in Section 2.4 we briefly mentioned some other models of interest, in particular those on different modelling scales. In the next chapter we review the model by Tomlinson and Bodmer [1995] in more detail, and use this as a base from which to begin our study.

Chapter 3

Cell population dynamics in the crypt[1]

In this chapter we study the cell population dynamics of a single colonic crypt with the aim of determining the most appropriate modelling technique for capturing cell division at the population level. We begin in Section 3.1 by considering the widely-cited model by Tomlinson and Bodmer [1995], which we reviewed in Chapter 2, and we comment on the assumptions made and the results that were obtained. We compare and contrast discrete, age-structured and continuous modelling approaches, which are presented in Sections 3.2, 3.3 and 3.4, respectively, and we discuss which is the simplest and most appropriate technique for modelling crypt dynamics. In Section 3.5, we compare the solutions from the age-structured and continuous models, and derive relations linking the parameters from each model. Throughout this chapter we work with the models in dimensional variables.

3.1 The Tomlinson and Bodmer model

As a starting point for our study we consider the widely-cited compartment model by Tomlinson and Bodmer [1995], which is summarised in Figure 3.1 and was reviewed in Chapter 2. Recall that the authors separated the cells in the crypt into populations of stem cells (denoted N_0), fully-differentiated cells (denoted N_2), and a third population for all the transit-amplifying cells that are in the process of differentiating, referred to as semi-differentiated cells (denoted N_1). For the purposes of this chapter we neglect the dead cells, which are assumed to have no bearing on the crypt dynamics. At the end of each cell cycle the stem cells can die, differentiate into transit cells or renew with probabilities a_1, a_2 and a_3 respectively, whereas transit cells will die, differentiate into fully-differentiated cells or renew with probabilities b_1, b_2 and b_3 respectively. The fully-differentiated cells will be removed from the system with probability c. The probabilities can also be thought of as the proportions of the cell populations that undergo each process. The cell cycle times for the stem and transit cell populations are denoted by t_0 and t_1, respectively, and t_2 refers to a reference time scale to record N_2 death.

Recall that Tomlinson and Bodmer formulated a difference equation model which measured the cell populations at discrete time steps. This implicitly assumes that the cells in

[1]The results of this chapter form the basis of parts of the papers Johnston *et al.* [2007a] and Johnston *et al.* [2007b], where the former includes a discussion of the appropriate modelling techniques, and the latter presents a comparison between the age-structured and continuous models.

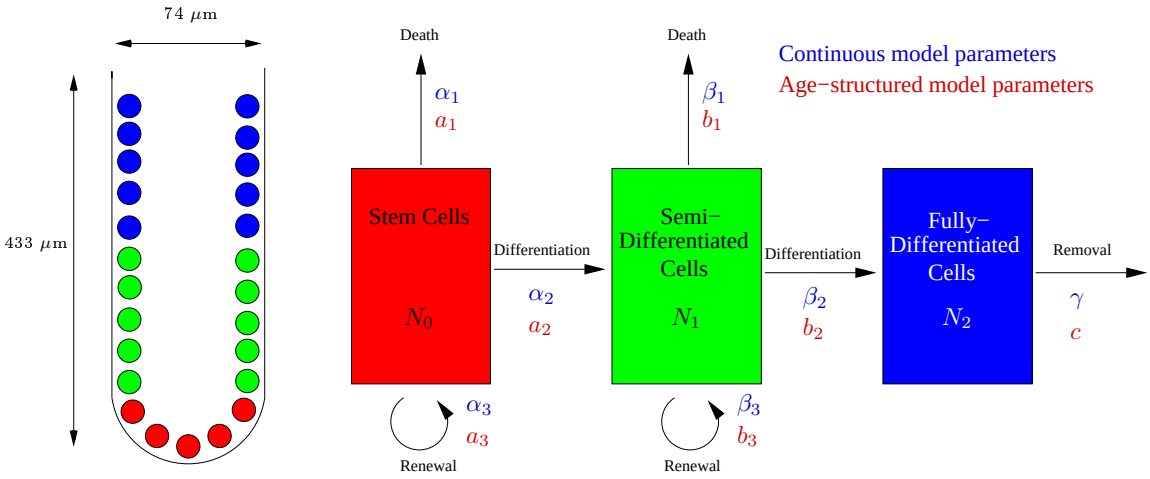

Figure 3.1: Left: A schematic diagram of a colonic crypt, with stem, semi-differentiated (transit-amplifying) and fully-differentiated cell populations. The dimensions given are for a human colonic crypt according to Halm and Halm [2000]. Right: A diagram showing the compartment structure used in the model by Tomlinson and Bodmer [1995]. The stem cells differentiate into semi-differentiated cells, which in turn differentiate into fully-differentiated cells. Each cell population can die, and the stem and semi-differentiated cells can renew. The parameters are the proportions a_i, b_i and c of the populations that are leaving the compartments. Alternatively, their Greek equivalents α_i, β_i and γ denote the rates of conversion for the continuous model, as discussed in Section 3.4.

both the semi- and fully-differentiated populations divide at the same time, which requires the stem cell cycle time to be an integer multiple of the semi-differentiated cell cycle time ($t_0/t_1 = p \in \mathbb{N}$), otherwise this synchrony condition will be broken after one generation of stem cells when the semi-differentiated cell population will contain cells at two different stages through their cycle. The system of difference equations is given by

$$N_0(G+1) = 2a_3 N_0(G), \tag{3.1}$$

$$N_1(G+1) = 2b_3 N_1(G)\frac{t_0}{t_1} + 2a_2 N_0(G), \tag{3.2}$$

$$N_2(G+1) = 2b_2 N_1(G)\frac{t_0}{t_1} + \left(1 - c\frac{t_0}{t_2}\right) N_2(G). \tag{3.3}$$

There are some implicit assumptions in this model which make it applicable only under special conditions. This model neglects the compounding effect of transit cells dividing on a different time scale. For example, when the cell cycle times are taken to be $t_0 = 1$ day and $t_1 = 1/2$ day, and we start with $\hat{n}_0$ stem cells and $\hat{n}_1$ semi-differentiated cells, after 1 day there are $2a_2\hat{n}_0 + (2b_3)^2\hat{n}_1$ semi-differentiated cells, but the Tomlinson and Bodmer [1995] model predicts $2a_2\hat{n}_0 + 4b_3\hat{n}_1$ semi-differentiated cells. Thus, the Tomlinson and Bodmer model (3.1)–(3.3) is strictly only valid when all the cell division times t_i are equal. We relax this assumption, and also the assumption that the cells divide synchronously, by considering an age-structured model in Section 3.3. The limiting case where the cell cycle times are commensurate is briefly considered in Section 3.2.

In the Tomlinson and Bodmer [1995] model it is assumed that the transition probabilities are constant, so that the resulting equations (3.1)–(3.3) are linear and the model is structurally unstable, that is, a small change in some of the parameters can lead to a qualitative

change in the solution. For example, unless exactly half the stem cells renew ($a_3 = 1/2$) the stem cell population will exhibit either exponential growth ($a_3 > 1/2$) or exponential decay ($a_3 < 1/2$). This is not physically realistic, since in order to achieve homeostasis in any physical or biological system some degree of feedback is necessary to maintain stability in the face of infinitesimal parameter changes. One possibility is to make Tomlinson and Bodmer's assumption that a_3 is a lumped parameter incorporating lower scale processes in which feedback and control ensure that $a_3 = 1/2$. In our modelling we make the feedback explicit, and we address the subject of feedback and homeostasis in Chapter 4.

3.2 Discrete model

A discrete modelling approach, such as that adopted by Tomlinson and Bodmer [1995], follows the number of cells in a population at regular time intervals, and does not take into account where the individual cells are in their cell cycle. As discussed in the previous section, a discrete model is only valid in the limiting case where the cells in each population divide at the same time, which requires the stem cell cycle time to be an integer multiple of the semi-differentiated cell cycle time, $t_0 = pt_1$, where $p \in \mathbb{N}$. Whilst this is a rare case, it is observed that possible parameter values are $t_0 = 24$ hours [Potten and Loeffler, 1990] and $t_1 = 12$ hours [Paulus *et al.*, 1993] which satisfy this constraint, and so we briefly consider commensurate cell populations in this section before relaxing the synchrony assumption and adopting an age-structured model in Section 3.3.

For our models we adopt the same notation that was used in Section 3.1 for the Tomlinson and Bodmer [1995] model, and a table of variables and parameters is shown in Appendix B. For the stem cell population there is only one relevant time scale, and so the stem cell population can be determined by

$$N_0(t) = \hat{n}_0(2a_3)^n \tag{3.4}$$

at time intervals $t = nt_0$, where $N_0 = \hat{n}_0$ initially and $n \in \mathbb{N}$. In fact, if the stem cell population is initially synchronous, then the stem cell population is known at all times to be

$$N_0(t) = \hat{n}_0(2a_3)^{\lfloor t/t_0 \rfloor}. \tag{3.5}$$

While the stem cell population can always be expressed in the form (3.4), characterising the forms of the semi-differentiated cell population solution, and in particular how differentiating cells contribute to subsequent populations, is heavily dependent on the specific choices for the cell cycle times.

3.2.1 Solutions for synchronous populations

One method of solving for the semi-differentiated cell population is to start with the initial population sizes and solve the problem iteratively, assuming that the cell cycle times are known. Iteration can provide an illustration of the solution in some simple cases. We assume, as in the Tomlinson and Bodmer [1995] model, that each cell dies, differentiates or renews at the end of the division cycle, with no changes occurring in the middle of a cycle. As before, we take the initial number of semi-differentiated cells to be $\hat{n}_1$. The differences in the cell division times affect how these cell populations interact.

Suppose that the stem cell cycle time is an integer multiple of the transit cell cycle time $t_0 = pt_1$, where $p \in \mathbb{N}$, and that all cells in each population are initially synchronous in their cell cycles. Since t_1 is the smallest time step we define $g_n := N_1(nt_1)$, where $n \in \mathbb{N}$ refers to the number of t_1 time steps undergone since the initial data were prescribed, and so $g_0 = \hat{n}_1$. Since stem cells will only differentiate every p^{th} time step, we require p difference equations of the form

$$g_{pn} = 2b_3 g_{pn-1} + 2a_2\hat{n}_0(2a_3)^{n-1}, \tag{3.6}$$

$$g_{pn+i} = 2b_3 g_{pn+i-1}, \tag{3.7}$$

where $i = 1, 2 \ldots p-1$. These can be combined into one difference equation by substituting the equation for g_{pn+j} into that for g_{pn+j+1} for each subsequent j, which yields

$$g_{pn+p-1} = (2b_3)^p g_{pn-1} + 2a_2\hat{n}_0(2a_3)^{n-1}(2b_3)^{p-1}. \tag{3.8}$$

Making the substitution $F_n = g_{pn+p-1}$ reduces this equation to

$$F_n = (2b_3)^p F_{n-1} + \frac{2a_2\hat{n}_0(2b_3)^{p-1}}{2a_3}(2a_3)^n, \tag{3.9}$$

and $F_0 = g_{p-1} = \hat{n}_1(2b_3)^{p-1}$ can be found by applying (3.7). Noting that the general solution of $F_n = aF_{n-1} + bc^n$ is

$$F_n = \begin{cases} F_0 a^n + \dfrac{bc}{c-a}\left[c^n - a^n\right] & \text{if } c \neq a, \\ (F_0 + bn)a^n & \text{if } c = a, \end{cases} \tag{3.10}$$

we can solve (3.9) to find that

$$N_1(t) = (2b_3)^i\left[\bar{A}(2a_3)^n + (\hat{n}_1 - \bar{A})(2b_3)^{pn}\right], \quad \text{where} \quad \bar{A} = \frac{2a_2\hat{n}_0}{2a_3 - (2b_3)^p} \tag{3.11}$$

at time points $t = (pn + i)t_1$, assuming that $2a_3 \neq (2b_3)^p$.

Now let us consider two special cases of the cell cycle times. Firstly, in the case where $t_0 = t_1 = 1$ day, the general solution reduces to

$$N_1(t) = \bar{A}(2a_3)^n + (\hat{n}_1 - \bar{A})(2b_3)^n, \quad \text{where} \quad \bar{A} = \frac{2a_2\hat{n}_0}{2a_3 - 2b_3} \tag{3.12}$$

when $t = nt_0$. When we set $t_0 = t_1$ in the Tomlinson and Bodmer [1995] model, the solution to (3.2) agrees with (3.12), where the generations G increase in steps of n days, which verifies that (3.2) is correct in this special case.

Secondly, in the more plausible case where $t_0 = 1$ day and $t_1 = 1/2$ day, the semi-differentiated cell population at time t is given by

$$N_1(t) = \begin{cases} \bar{A}(2a_3)^n + (\hat{n}_1 - \bar{A})(2b_3)^n, & t = 2nt_1, \\ 2b_3\left[\bar{A}(2a_3)^n + (\hat{n}_1 - \bar{A})(2b_3)^n\right], & t = (2n+1)t_1, \end{cases} \tag{3.13}$$

where

$$\bar{A} = \frac{2a_2\hat{n}_0}{2a_3 - (2b_3)^2}. \tag{3.14}$$

It is clear from the solutions (3.11), (3.12) and (3.13) that the growth behaviour of the system is the same whatever the cell cycle times are, with the solution stability dependent on whether a_3 and b_3 are greater than or less than $1/2$. The solutions merely differ in their multiplicative constant $\bar{A}$, which depends on the cell cycle times.

Note that even if the cell cycle times are such that the cell populations are synchronous, the synchrony condition will be broken unless the cell populations are initially commensurate in their cell cycle. In order to relax these assumptions we must consider an asynchronous set of populations where the stage through the division cycle of each cell is accounted for, and this is described in the next section.

3.3 Age-structured model

In this section we present a model of the cell populations in the crypt that removes the constraint of synchronicity of cell division and takes careful account of the different cell cycle times of stem and semi-differentiated cells. In order to do this we need to keep track of not just the numbers of cells in each population but also of their age distribution, that is, we need an age-structured model (see, for example, Arino and Kimmel [1993]; Chapman *et al.* [2007]; Murray [2002]; Trucco [1965a,b]). In this context, the 'age' of a cell refers to the length of time since its last division, so there will be a distribution of ages at each point in time, which is more realistic than populations where all the individuals have the same age [Paulus *et al.*, 1992].

As in the previous model, we consider populations of stem, semi- and fully-differentiated cells, and we denote by $N_i = N_i(t, a)$, for $i = 0, 1, 2$, the age distribution function for cells of population N_i at time t, so that there are $N_i \delta a$ cells in the age range $[a, a + \delta a]$. We assume that each cell is committed to death, differentiation or renewal only at the end of its cell cycle, and that this process is instantaneous. The proportions of cells undergoing each process are a_i, b_i and c as in the discrete model, which was summarised in Figure 3.1. Recall that we denote the cell cycle times of stem and semi-differentiated cells by t_0 and t_1 respectively, and since the fully-differentiated cells are not progressing through the cell cycle, we follow Tomlinson and Bodmer [1995] by introducing a corresponding reference time t_2 for fully-differentiated cells after which they may die or be shed. For population N_i the domain of possible cell ages is denoted by $A_i = \{a : 0 \leq a < t_i\}$. As in the Tomlinson and Bodmer model, we assume that all cell divisions are symmetric.

When the populations are totalled over all possible ages in the cell cycle the solutions from the age-structured model should reduce to the discrete model solutions at fixed time intervals. The total number of cells in population N_i at a time t is given by

$$\hat{N}_i(t) = \int_{a \in A_i} N_i(t, a) \, \mathrm{d}a = \int_0^{t_i} N_i(t, a) \, \mathrm{d}a. \tag{3.15}$$

A comparison between the discrete and age-structured models and their solutions is discussed in Section 3.6.

3.3.1 The model equations

In order to derive the equations for the cell populations, we assume that cell numbers are only changed at the end of a cell cycle. Therefore, conservation of cell numbers implies that

$$\frac{\partial N_i}{\partial t} + \frac{\partial N_i}{\partial a} = 0, \tag{3.16}$$

for $i = 0, 1, 2$, where we take age to increase linearly with time, which have general solutions

$$N_i(t, a) = f_i(t - a), \tag{3.17}$$

where f_i are arbitrary functions that can be determined in terms of the initial age profiles. The renewal, death or differentiation of the cells at the end of the cell cycle gives the conditions

$$N_0(t, 0) = 2a_3 N_0(t, t_0), \tag{3.18}$$
$$N_1(t, 0) = 2b_3 N_1(t, t_1) + 2a_2 N_0(t, t_0), \tag{3.19}$$
$$N_2(t, 0) = 2b_2 N_1(t, t_1) + (1 - c)N_2(t, t_2). \tag{3.20}$$

Using (3.17), these constraints are equivalent to

$$N_0(t, 0) = 2a_3 N_0(t - t_0, 0), \tag{3.21}$$
$$N_1(t, 0) = 2b_3 N_1(t - t_1, 0) + 2a_2 N_0(t - t_0, 0), \tag{3.22}$$
$$N_2(t, 0) = 2b_2 N_1(t - t_1, 0) + (1 - c)N_2(t - t_2, 0). \tag{3.23}$$

Conditions (3.21)–(3.23) can be thought of as generalisations of the Tomlinson and Bodmer difference equations (3.1)–(3.3).

We note that cell death is either due to altruistic suicide (apoptosis), for example after a cell has detected a replicative error of the DNA in mitosis, or it is due to involuntary death (necrosis), possibly as a result of nutrient deficiency [Chen *et al.*, 2001]. In this model, cell death occurs only at the end of the cell cycle and can be considered to be apoptotic. On the other hand, necrosis could occur at any point during the cell cycle and would be represented by a constant death rate term in (3.16).

As initial conditions we prescribe age profiles for the cell populations, i.e. we take $t = 0$, $a = \sigma$, $N_i(0, \sigma) = n_i(\sigma)$, where $0 \leq \sigma < t_i$, which gives $f_i(\sigma) = n_i(-\sigma)$ for $-t_i < -\sigma \leq 0$. Therefore the general solutions of (3.16), with these initial conditions, are given by

$$N_i(t, a) = n_i(a - t) \qquad \text{for } -t_i < t - a \leq 0, \tag{3.24}$$

and the initial data will be propagated along the characteristics $t - a =$ constant. In the following sections we derive solutions to the age-structured model for the stem and transit cells, and we remain in dimensional variables throughout.

3.3.2 Solving for the stem cell population

In this section we derive solutions for the stem cell population for two special cases of initial age profile. Using (3.24), with a general initial age profile $N_0(0, a) = n_0(a)$, it is clear that

$$N_0(t, a) = n_0(a - t), \tag{3.25}$$

when $-t_0 < t - a \leq 0$. Applying the cell cycle condition (3.21), we find the general solution for the stem cell population to be

$$N_0(t, a) = (2a_3)^n n_0(nt_0 + a - t), \tag{3.26}$$

where n is the unique integer such that $(n-1)t_0 < t - a \leq nt_0$, corresponding to the number of times cells of age a have been through the cell cycle at time t.

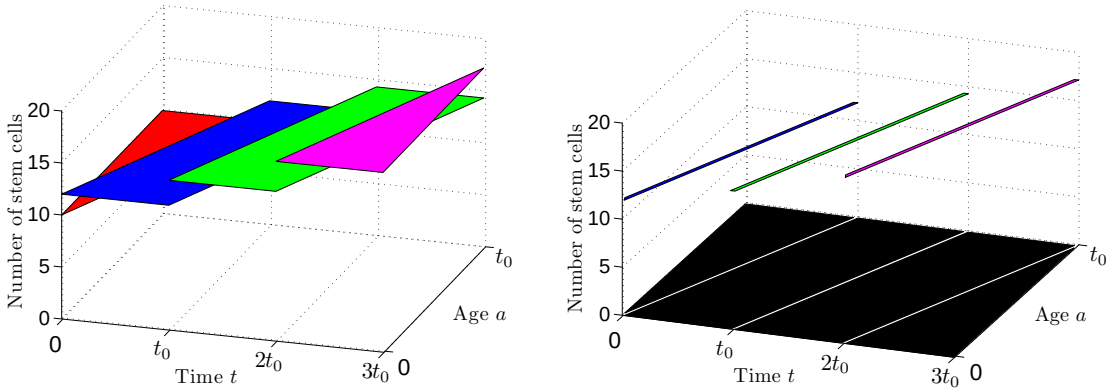

Figure 3.2: Plots of the age-structured model solutions for the stem cell population for different initial age profiles. Left: a uniform age profile from (3.27). Right: a δ-function age profile from (3.29). The parameter values are taken to be $a_3 = 0.6$, $\hat{n}_0 = 10$ and $t_0 = 1$ day. Colour is for visual clarity and represents different numbers of stem cells.

In the calculations that follow we consider two special cases of the initial age profile: $n_i(a) = \hat{n}_i/t_i$ and $n_i(a) = \hat{n}_i\delta(a)$, where $\hat{n}_i$ is constant. The former implies a uniform distribution of ages initially, while the latter implies that all the cells are aged 0. Using (3.15), the total number of cells at time $t = 0$ is $\hat{n}_i$ for both of these age profiles. We assume throughout that the same form of initial conditions are applied to each cell population. The uniform age distribution is a realistic assumption which implies asynchronous cell division, and *in vivo* we expect there to be cells at varying stages of the cell cycle. The δ-function distribution corresponds to synchronous cell division, which means that every cell in the population is at the same stage in the cell cycle. While this is less physically realistic, it is easier to obtain closed form solutions, and it allows us to compare results with the discrete model.

If the initial age profile is constant, $n_0(a) = \hat{n}_0/t_0$, then the solution is simply

$$N_0(t, a) = \frac{\hat{n}_0}{t_0}(2a_3)^n, \tag{3.27}$$

for $(n-1)t_0 < t - a \leq nt_0$. If, however, the initial age profile is concentrated at one age, where $n_0(a) = \hat{n}_0\delta(a)$, then the solution is

$$N_0(t, a) = \hat{n}_0(2a_3)^n\delta(t - nt_0 - a), \tag{3.28}$$

for $(n-1)t_0 < t - a \leq nt_0$. We note that $\hat{n}_0$ is dimensionless and $\delta(a)$ has dimensions 1/[time]. Since the δ-function is only non-zero at a specific point, we can express the full solution as an infinite sum

$$N_0(t, a) = \hat{n}_0 \sum_{n=0}^{\infty}(2a_3)^n\delta(t - nt_0 - a), \tag{3.29}$$

for $-t_0 < t - a \leq \infty$. Therefore $N_0(t, a)$ is only non-zero at points where $t - a = nt_0$. Both solutions (3.27) and (3.29) are illustrated in Figure 3.2.

3.3.3 Solving for the semi-differentiated cell population

The solution for the stem cell population could be obtained in a general form for all intervals of $a - t$, but the populations of semi- and fully-differentiated cells cannot be so neatly expressed. This is due to the stem and semi-differentiated cell populations dividing on different time scales.

From (3.17), we know that the general solution for the semi-differentiated cells is given by $N_1(t, a) = g(a - t)$, for some arbitrary function g. Combining this with the cell cycle condition (3.22), g must satisfy

$$g(t) = 2b_3 g(t - t_1) + 2a_2 N_0(t - t_0, 0). \tag{3.30}$$

As we explained in Section 3.3.1, this can be solved with a prescribed initial age profile to give $N_1(t, a) = n_1(a - t)$, for $-t_1 < t - a \leq 0$.

Again we consider the two cases of initial age profiles where $n_1(a) = \hat{n}_1/t_1$ and $n_1(a) = \hat{n}_1\delta(a)$. The cell cycle condition (3.22) includes two time scales, t_0 and t_1, and the relation between these times is critical to how the solution can be found. This is because the contributions to the N_1 population from differentiating N_0 cells will either superpose the renewing N_1 population or add a separate contribution that is not concurrent. In the next three sections we illustrate this point by considering particular ratios of the cell cycle times using both uniform and δ-function initial age profiles. For these purposes we assume that the initial conditions are of the same type and prescribed synchronously for both the stem and transit cell populations.

3.3.3.1 Illustrative solutions for a uniform age distribution

Let us consider the case where the initial semi-differentiated cell population age profile is uniformly distributed, $N_1(0, a) = \hat{n}_1/t_1$, and where the stem cell cycle time is an integer multiple of the transit cell cycle time, $t_0 = pt_1$, where $p \in \mathbb{N}$. The smallest common time step to both time scales is t_1, and so we only need to consider the solution at time points $g_n := g(nt_1)$, since the solution is constant in the interval $(n-1)t_1 < t-a \leq nt_1$. Here $n \in \mathbb{N}$ refers to the number of time intervals undergone since the initial data were prescribed.

There are p difference equations for each interval t_1 within a stem cell cycle, which are given by

$$g_{pn} = 2b_3 g_{pn-1} + \frac{2a_2\hat{n}_0}{t_0}(2a_3)^{n-1}, \tag{3.31}$$

$$g_{pn+i} = 2b_3 g_{pn+i-1} + \frac{2a_2\hat{n}_0}{t_0}(2a_3)^{n}, \tag{3.32}$$

where $i = 1, 2 \ldots p-1$. As in Section 3.2.1, these can be reduced to one difference equation by substituting the equation for g_{pn+j} into that for g_{pn+j+1}, for each j, which produces

$$g_{pn+p-1} = (2b_3)^p g_{pn-1} + \frac{2a_2\hat{n}_0}{t_0}(2a_3)^n \left[\frac{(2b_3)^{p-1}}{2a_3} + \frac{1-(2b_3)^{p-1}}{1-2b_3}\right]. \tag{3.33}$$

Making the substitution $F_n = g_{pn+p-1}$ reduces this equation to

$$F_n = (2b_3)^p F_{n-1} + \frac{2a_2\hat{n}_0}{t_0}(2a_3)^n \left[\frac{(2b_3)^{p-1}}{2a_3} + \frac{1-(2b_3)^{p-1}}{1-2b_3}\right], \tag{3.34}$$

and the general solution from (3.10) can be used to find that

$$F_n = F_0(2b_3)^{pn} + \frac{2a_2\hat{n}_0/t_0}{2a_3-(2b_3)^p}\left[(2b_3)^{p-1} + \frac{2a_3}{1-2b_3}\left(1-(2b_3)^{p-1}\right)\right]\left[(2a_3)^n - (2b_3)^{pn}\right]. \tag{3.35}$$

Therefore the full solution in each of the p intervals is

$$g_{pn+i} = \frac{\hat{n}_1}{t_1}(2b_3)^{pn+i} + \frac{2a_2\hat{n}_0}{t_0}\left[\frac{[(2b_3)^i(1-(2b_3)^{p-i}) + 2a_3(1-(2b_3)^i)](2a_3)^n - (1-(2b_3)^p)(2b_3)^{pn+i}}{(1-2b_3)(2a_3-(2b_3)^p)}\right]. \tag{3.36}$$

In particular, when $t-a = pnt_1$, (3.36) simplifies to

$$N_1(t,a) = \tilde{A}(2a_3)^n + \left(\frac{\hat{n}_1}{t_1} - \tilde{A}\right)(2b_3)^{pn}, \quad \text{where} \quad \tilde{A} = \frac{2a_2\hat{n}_0[1-(2b_3)^p]}{t_0[2a_3-(2b_3)^p](1-2b_3)}. \tag{3.37}$$

If, however, we had considered the case where $t_1 = qt_0$, then the transit cell population solution at intervals $t-a = nt_1$ can be found from the difference equation

$$g_n = 2b_3 g_{n-1} + \frac{2a_2\hat{n}_0}{t_0}(2a_3)^{qn-1}, \tag{3.38}$$

which is solved to give

$$N_1(t,a) = \breve{A}(2a_3)^{qn} + \left(\frac{\hat{n}_1}{t_1} - \breve{A}\right)(2b_3)^n, \quad \text{where} \quad \breve{A} = \frac{2a_2\hat{n}_0(2a_3)^{q-1}}{t_0[(2a_3)^q - 2b_3]}. \tag{3.39}$$

Clearly both (3.37) and (3.39) reduce to the same form if $p = q = 1$.

3.3.3.2 General solutions for a δ-function age distribution

Let us consider the case where the initial semi-differentiated cell population age profile is the δ-function $N_1(0,a) = \hat{n}_1\delta(a)$ and, as before, we make the assumption that the initial age profiles for the N_0 and N_1 populations begin at the same time. In this case we can determine the general solution for unspecified cell cycle times t_0 and t_1. We add together each contribution to the N_1 population from differentiating stem cells and renewing semi-differentiated cells. The contribution from the original, prescribed semi-differentiated cells will be

$$\hat{n}_1(2b_3)^m\delta(t-a-mt_1), \tag{3.40}$$

after m cell cycles, while each band of differentiating stem cells will produce

$$2a_2\hat{n}_0(2a_3)^{n-1}(2b_3)^m\delta(t-a-nt_0-mt_1) \tag{3.41}$$

new cells from the cells that differentiated after n stem cell cycles and then divided m times as semi-differentiated cells. Therefore, the full solution is given by

$$N_1(t,a) = \hat{n}_1\sum_{m=0}^{\infty}(2b_3)^m\delta(t-a-mt_1) + 2a_2\hat{n}_0\sum_{n=1}^{\infty}\sum_{m=0}^{\infty}(2a_3)^{n-1}(2b_3)^m\delta(t-a-nt_0-mt_1), \tag{3.42}$$

for $0 < t-a \leq \infty$. The general solution for the fully-differentiated cells can also be obtained in the same way to give

$$N_2(t,a) = \hat{n}_2 \sum_{p=0}^{\infty}(1-c)^p\delta(t-a-pt_2) + 2b_2\hat{n}_1 \sum_{m=1}^{\infty}\sum_{p=0}^{\infty}(2b_3)^{m-1}(1-c)^p\delta(t-a-mt_1-pt_2)$$
$$+ 2a_2\hat{n}_0(2b_2)\sum_{n=1}^{\infty}\sum_{m=1}^{\infty}\sum_{p=0}^{\infty}(2a_3)^{n-1}(2b_3)^{m-1}(1-c)^p\delta(t-a-nt_0-mt_1-pt_2). \quad (3.43)$$

For the δ-function initial age profile, the contributions to the N_1 population from N_0 cells will only superpose the renewing N_1 population if the ratio of time scales is rational and the births are commensurate at some point in time, otherwise the contributions from N_0 and N_1 will remain separate. In the next section we solve for the transit cell population in some more specific cases.

3.3.3.3 Illustrative solutions for a δ-function age distribution

Now let us consider the case where the initial age profile of transit cells is the δ-function $N_1(0,a) = \hat{n}_1\delta(a)$, and $t_0 = pt_1$. There will only be non-zero solutions when $t-a = nt_1$, for $n \in \mathbf{N}$, so we define g_n by

$$N_1(t,a) = g(t-a) = \sum_{n=0}^{\infty} g_n\delta(t-nt_1-a). \quad (3.44)$$

Substituting (3.44) into (3.30), and using (3.29), we find that

$$\sum_{n=0}^{\infty} g_n\delta(t-nt_1) = 2b_3\sum_{n=0}^{\infty} g_n\delta(t-(n+1)t_1) + 2a_2\hat{n}_0\sum_{n=0}^{\infty}(2a_3)^n\delta(t-p(n+1)t_1). \quad (3.45)$$

Specifying $g_0 = \hat{n}_1$, and noting that the final term will only contribute once every p time steps, (3.45) produces difference equations

$$g_{pn} = 2b_3 g_{pn-1} + 2a_2\hat{n}_0(2a_3)^{n-1}, \quad (3.46)$$
$$g_{pn+i} = 2b_3 g_{pn+i-1}, \quad (3.47)$$

where $i = 1, 2 \ldots p-1$. These are identical to (3.6) and (3.7), and produce the solution

$$N_1(t) = \sum_{n=0}^{\infty}\sum_{i=0}^{p-1}(2b_3)^i\left[\bar{A}(2a_3)^n + (\hat{n}_1 - \bar{A})(2b_3)^{pn}\right]\delta(t-a-(pn+i)t_1), \quad (3.48)$$

where $\bar{A}$ was defined in (3.11).

If, however, we had considered the case where $t_1 = qt_0$, then the transit cell population will only be non-zero when $t-a = nt_0$. We define g_n by

$$N_1(t,a) = g(t-a) = \sum_{n=0}^{\infty} g_n\delta(t-nt_0-a), \quad (3.49)$$

and the difference equations are given by

$$g_n = 2a_2\hat{n}_0(2a_3)^{n-1}, \quad \text{for} \quad 1 \leq n \leq q-1, \quad (3.50)$$
$$g_n = 2b_3 g_{n-q} + 2a_2\hat{n}_0(2a_3)^{n-1}, \quad \text{for} \quad n \geq q. \quad (3.51)$$

Starting from $g_0 = \hat{n}_1$, and solving these equations by iteration, gives

$$N_1(t,a) = \sum_{n=0}^{\infty} \left[\breve{A}(2a_3)^{qn} + (\hat{n}_1 - \breve{A})(2b_3)^n \right] \delta(t - a - nqt_0) \\ + \sum_{n=0}^{\infty} \sum_{i=1}^{q-1} \breve{A}(2a_3)^{i-q} \left[(2a_3)^{q(n+1)} - (2b_3)^{n+1} \right] \delta(t - a - (nq+i)t_0), \quad (3.52)$$

where $\breve{A}$ was defined in (3.39).

When $p = q = 1$, clearly both (3.48) and (3.52) reduce to the form

$$N_1(t,a) = \sum_{n=0}^{\infty} \left[\bar{A}(2a_3)^n + (\hat{n}_1 - \bar{A})(2b_3)^n \right] \delta(t - a - n\tau), \quad \text{where} \quad \bar{A} = \frac{2a_2\hat{n}_0}{2a_3 - 2b_3}. \quad (3.53)$$

Comparing the weights of this solution (i.e. the coefficients of the δ-functions) to the discrete model solution (3.12), we see that the models agree.

3.3.4 Results summary

In summary, we have presented an age-structured model that relaxes the assumptions of cell division synchrony and generalises the discrete model from Section 3.2. Since it is difficult to interpret the behaviour of the general solution in closed form, we solved the equations for two special cases of the cell population initial age profiles: a uniform age distribution and a δ-function distribution.

In the most plausible case where the cell cycle times satisfy $t_0 = 2t_1$, the solution for the uniform initial age profile from (3.36) reduces to

$$N_1 = \begin{cases} \dfrac{\hat{n}_1}{t_1}(2b_3)^{2n} + \dfrac{2a_2\hat{n}_0(1+2b_3)}{t_0(2a_3 - (2b_3)^2)} \left[(2a_3)^n - (2b_3)^{2n} \right], & \text{if } \ t - a = 2nt_1, \\ \dfrac{\hat{n}_1}{t_1}(2b_3)^{2n+1} + \dfrac{2a_2\hat{n}_0/t_0}{2a_3 - (2b_3)^2} \left[(2a_3 + 2b_3)(2a_3)^n - (1+2b_3)(2b_3)^{2n+1} \right], \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{if } \ t - a = (2n+1)t_1, \end{cases} \quad (3.54)$$

where $n \in \mathbb{N}$. For the case of δ-function initial conditions, using (3.48) we obtain the formula

$$N_1(t,a) = \sum_{n=0}^{\infty} \left[\delta(t - a - 2nt_1) + 2b_3\delta(t - a - (2n+1)t_1) \right] \\ \times \left[\hat{n}_1(2b_3)^{2n} + \frac{2a_2\hat{n}_0}{2a_3 - (2b_3)^2} \left[(2a_3)^n - (2b_3)^{2n} \right] \right]. \quad (3.55)$$

The similarity between expressions (3.54) and (3.55) indicates that the distribution of cell ages, while necessary for consistency when modelling on the time scale of the cell cycle, is not crucial in determining the long-time behaviour of the solution, and can make the solutions overly complicated. We therefore now develop a much simpler ordinary differential equation (ODE) model which assumes continuous cell division.

3.4 Continuous model

In this section, we adopt a third approach for modelling the cell population dynamics in the crypt which uses a continuum model to follow each cell population $N_i(t)$ varying in continuous time t. Since we are mainly interested in the development of cell populations over time scales much longer than the cell cycle time, we can assume that growth occurs continuously rather than in discrete multiples of cell cycles. We also assume that the cell populations are large enough that, to a first approximation, they vary continuously with time, rather than taking only integer values. The continuum modelling approach also has the added benefit that ODEs are much simpler to analyse than difference equations or PDEs.

Whereas in the discrete and age-structured models the cell population renewal, differentiation and death are assumed to occur only at the end of the cell cycle, in a continuous model these processes are assumed to occur continuously in time. In the discrete and age-structured models we considered the proportions of the populations that undergo each process at the end of each cell cycle, but in the continuous model we must monitor the rates at which these processes occur in (continuous) time. We denote the per-capita rate of stem (respectively semi-differentiated) cell proliferation by α_3 (respectively β_3), differentiation by α_2 (respectively β_2) and death by α_1 (respectively β_1), and the per-capita removal rate of fully-differentiated cells by γ. Thus, for example, over a small time τ we would expect a proportion $\alpha_2\tau$ of stem cells to have differentiated, on average. We first assume that the rates take constant values. These parameters are summarised in Figure 3.1 and also in the parameter table in Appendix B. Note that these rates are analogous but not equivalent to the corresponding proportions of the cell populations in the age-structured model; the relation between the two sets of parameters will be determined in Section 3.5.

The ODEs for the system are given by

$$\frac{\mathrm{d}N_0}{\mathrm{d}t} = (\alpha_3 - \alpha_1 - \alpha_2)N_0, \tag{3.56}$$

$$\frac{\mathrm{d}N_1}{\mathrm{d}t} = (\beta_3 - \beta_1 - \beta_2)N_1 + \alpha_2 N_0, \tag{3.57}$$

$$\frac{\mathrm{d}N_2}{\mathrm{d}t} = \beta_2 N_1 - \gamma N_2. \tag{3.58}$$

Given initial cell populations $N_i = \hat{n}_i$, we find that

$$N_0(t) = \hat{n}_0 e^{\alpha t}, \tag{3.59}$$

$$N_1(t) = Ae^{\alpha t} + (\hat{n}_1 - A)e^{\beta t}, \tag{3.60}$$

$$N_2(t) = Be^{\alpha t} + Ce^{\beta t} + (\hat{n}_2 - B - C)e^{-\gamma t}, \tag{3.61}$$

where $\alpha = \alpha_3 - \alpha_1 - \alpha_2$ and $\beta = \beta_3 - \beta_1 - \beta_2$ are the net stem and semi-differentiated cell population per-capita growth rates, respectively, and the constants A, B and C are given by

$$A = \frac{\alpha_2 \hat{n}_0}{\alpha - \beta}, \quad B = \frac{\beta_2 A}{\gamma + \alpha} \quad \text{and} \quad C = \frac{\beta_2(\hat{n}_1 - A)}{\gamma + \beta}. \tag{3.62}$$

If $\alpha > 0$ or $\beta > 0$, then there will be exponential growth in the cell populations, while if $\alpha < 0$ and $\beta < 0$, then all populations will decay to zero. Only in the case where $\alpha = 0$ and $\beta < 0$ will each cell population approach a non-zero steady state. The results of Tomlinson

and Bodmer [1995] essentially correspond in this model to observing that altering the value of β (or indeed α_2, β_2 or γ) while keeping $\alpha = 0$ can lead to a new steady state in the size of the populations, without necessarily leading immediately to exponential growth so long as β stays negative.

The sensitivity of the solutions (3.59)–(3.61) to the values of the rates α and β again illustrates the structural instability of this linear model. For a more biologically realistic model we introduce feedback into the system in Chapter 4. For the rest of this chapter, however, we compare and contrast the age-structured and continuous models, and discuss when each approach is more appropriate.

3.5 Comparing the age-structured and continuous models

In the age-structured model we consider proportions of the cell populations dying, differentiating or renewing at discrete time intervals, whereas in the continuous model we assume that these processes occur continuously and we work with the rates at which they occur. Both models produce the same observed dynamics, and we would expect that under certain limits these models should be equivalent. In order to compare the age-structured and continuous models it is important to relate the two sets of parameters. Also, we might expect experimental data to be prescribed either in the form of the proportions or the rates, or possibly a combination of the two, which means parameter values can be shared between the models if we know how the rates and proportions are linked. The rates are given dimensionally in units of hours^{-1}, while the proportions are percentages and will not have any units.

One way to compare the models would be to integrate the age-structured model equations over age, and average over many cell cycles. Alternatively, we could seek separable solutions of the form $e^{\sigma t} f(a)$, similar to Gaffney [2004] or Murray [2002], and compare the results with the continuous model solutions. Instead, we choose to compare the solutions to the models directly, which is explained in the next section.

3.5.1 Age-structured model solutions integrated over age

In order to compare the age-structured and continuous model solutions we must integrate the age-structured model solutions $N_i(t, a)$ over all possible ages a to obtain solutions only varying in time $\hat{N}_i(t)$. For this purpose we choose the most simple form of the age-structured model solutions to integrate which are the solutions with $\delta-$function initial conditions where the cells all start at age zero, so that $n_i(a) = \hat{n}_i\,\delta(a)$, where $\hat{n}_i$ is the initial total population of cells of type i. The solutions in this case for the stem, semi- and fully-differentiated cell populations are given by (3.29), (3.42) and (3.43), respectively.

The total age-integrated cell population densities, denoted by $\hat{N}_i(t)$ for population N_i at any time t, can be found by integrating over all possible ages of cells, as was shown in (3.15). Note that t_2 is an arbitrary time scale against which fully-differentiated cell removal is measured, and it is not an important time scale in this problem.

In Sections 3.5.1.1, 3.5.1.2 and 3.5.1.3 we derive, respectively, $\hat{N}_0(t)$, $\hat{N}_1(t)$ and $\hat{N}_2(t)$, and in Section 3.5.2 we compare the age-structured and continuous model solutions.

3.5.1.1 Stem cell population integrated over age

The general age-structured model solution for the stem cells in the case where $n_0(a) = \hat{n}_0\,\delta(a)$ is given by (3.29). Integrating this over all possible ages a, and assuming that the integration and infinite summation may be swapped, we find that

$$\hat{N}_0(t) = \hat{n}_0 \int_0^{t_0} \sum_{n=0}^{\infty} \delta(t - nt_0 - a)(2a_3)^n \,\mathrm{d}a \tag{3.63}$$

$$= \hat{n}_0 \sum_{n=0}^{\infty} (2a_3)^n \int_0^{t_0} \delta(t - nt_0 - a) \,\mathrm{d}a \tag{3.64}$$

$$= \hat{n}_0 (2a_3)^{\lfloor t/t_0 \rfloor}. \tag{3.65}$$

The last equality follows since the only δ-function which gives a non-zero integral is that satisfying $nt_0 < t < (n+1)t_0$, which picks out the single value $n = \lfloor t/t_0 \rfloor$ from the sum.

3.5.1.2 Semi-differentiated cell population integrated over age

The general age-structured model solution for the semi-differentiated cells where $n_1(a) = \hat{n}_1\,\delta(a)$ is given by (3.42). Integrating this over a, and again assuming that the integration and infinite summation may be swapped, leads to

$$\hat{N}_1(t) = \hat{n}_1 \sum_{m=0}^{\infty} (2b_3)^m \int_0^{t_1} \delta(t - a - mt_1) \,\mathrm{d}a + 2a_2\hat{n}_0 \sum_{n=1}^{\infty} \sum_{m=0}^{\infty} (2a_3)^{n-1}(2b_3)^m \int_0^{t_1} \delta(t - a - nt_0 - mt_1) \,\mathrm{d}a. \tag{3.66}$$

The index m in the first summation must satisfy $0 \leq t - mt_1 < t_1$, and so the integral of the first δ-function picks out the value $m = \lfloor t/t_1 \rfloor$. For the second term, the summation indices n and m must satisfy $0 \leq t - nt_0 - mt_1 < t_1$, which picks out the value $m = \lfloor t/t_1 - nt_0/t_1 \rfloor$, while n can run from 1 to $\lfloor t/t_0 \rfloor$ and still satisfy this inequality. Therefore

$$\hat{N}_1(t) = \hat{n}_1 (2b_3)^{\lfloor t/t_1 \rfloor} + 2a_2\hat{n}_0 \sum_{n=1}^{\lfloor t/t_0 \rfloor} (2a_3)^{n-1}(2b_3)^{\lfloor t/t_1 - nt_0/t_1 \rfloor} \tag{3.67}$$

represents the total semi-differentiated cell population at all times t.

3.5.1.3 Fully-differentiated cell population integrated over age

The general age-structured model solution for the fully-differentiated cells, where $n_2(a) = \hat{n}_2\,\delta(a)$, is given by (3.43). Integrating this over a, and assuming that the integration and

infinite summation may be swapped, gives

$$\begin{aligned}\hat{N}_2(t) = \hat{n}_2 \sum_{p=0}^{\infty}(1-c)^p \int_0^{t_2} \delta(t-a-pt_2)\,\mathrm{d}a \\ + 2b_2\hat{n}_1 \sum_{m=1}^{\infty}\sum_{p=0}^{\infty}(2b_3)^{m-1}(1-c)^p \int_0^{t_2} \delta(t-a-mt_1-pt_2)\,\mathrm{d}a \\ + 2a_2\hat{n}_0(2b_2)\sum_{n=1}^{\infty}\sum_{m=1}^{\infty}\sum_{p=0}^{\infty}(2a_3)^{n-1}(2b_3)^{m-1}(1-c)^p \int_0^{t_2} \delta(t-a-nt_0-mt_1-pt_2)\,\mathrm{d}a.\end{aligned} \tag{3.68}$$

The first δ-function picks out the value $p = \lfloor t/t_2 \rfloor$, and the second picks out the value $p = \lfloor t/t_2 - mt_1/t_2 \rfloor$, where m runs from 1 to $\lfloor t/t_1 \rfloor$, as for the semi-differentiated cells. For the third term, the summation indices must satisfy $0 \leq t - nt_0 - mt_1 - pt_2 < t_2$ which picks out the value $p = \lfloor t/t_2 - nt_0/t_2 - mt_1/t_2 \rfloor$. Hence m can run from 1 to $\lfloor t/t_1 - nt_0/t_1 \rfloor$, and n runs from 1 to $\lfloor t/t_0 - t_1/t_0 \rfloor$. Therefore

$$\begin{aligned}\hat{N}_2(t) = \hat{n}_2(1-c)^{\lfloor t/t_2 \rfloor} + 2b_2\hat{n}_1 \sum_{m=1}^{\lfloor t/t_1 \rfloor}(2b_3)^{m-1}(1-c)^{\lfloor t/t_2 - mt_1/t_2 \rfloor} \\ + 2a_2\hat{n}_0(2b_2)\sum_{n=1}^{\lfloor t/t_0 - t_1/t_0 \rfloor}\sum_{m=1}^{\lfloor t/t_1 - nt_0/t_1 \rfloor}(2a_3)^{n-1}(2b_3)^{m-1}(1-c)^{\lfloor t/t_2 - nt_0/t_2 - mt_1/t_2 \rfloor}\end{aligned} \tag{3.69}$$

represents the total fully-differentiated cell population at all times t.

3.5.2 Comparing the age-structured and continuous model solutions

In order to compare the solutions from the age-structured and continuous models we would like to obtain approximations to the closed form expressions for $\hat{N}_1(t)$ and $\hat{N}_2(t)$ from (3.67) and (3.69), respectively. In order to do this we must approximate the summations

$$\sigma_1 = \sum_{n=1}^{\lfloor t/t_0 \rfloor}(2a_3)^{n-1}(2b_3)^{\lfloor t/t_1 - nt_0/t_1 \rfloor}, \tag{3.70}$$

$$\sigma_2 = \sum_{m=1}^{\lfloor t/t_1 \rfloor}(2b_3)^{m-1}(1-c)^{\lfloor t/t_2 - mt_1/t_2 \rfloor}, \tag{3.71}$$

$$\sigma_3 = \sum_{n=1}^{\lfloor t/t_0 - t_1/t_0 \rfloor}\sum_{m=1}^{\lfloor t/t_1 - nt_0/t_1 \rfloor}(2a_3)^{n-1}(2b_3)^{m-1}(1-c)^{\lfloor t/t_2 - nt_0/t_2 - mt_1/t_2 \rfloor}. \tag{3.72}$$

Approximations for σ_1, σ_2 and σ_3 for general t_0, t_1 and t_2 are derived in Appendix C. However, the time scale t_2 is merely a reference against which fully-differentiated cell removal can be measured, and so it can be arbitrarily chosen. The most sensible choice is to set t_2 equal to t_1, since this is the time scale that feeds into N_2. In this case we can evaluate σ_2 exactly and obtain a better approximation to σ_3, which is also shown in Appendix C.

When $t_1 = t_2$, we approximate

$$\sigma_1 \approx f_1 \left(\frac{(2a_3)^{t/t_0} - (2b_3)^{t/t_1}}{(2a_3)^{t_1} - (2b_3)^{t_0}} \right), \tag{3.73}$$

$$\sigma_2 = \frac{(2b_3)^{\lfloor t/t_1 \rfloor} - (1-c)^{\lfloor t/t_1 \rfloor}}{2b_3 - (1-c)}, \tag{3.74}$$

$$\sigma_3 \approx \frac{1}{2b_3 - (1-c)} \left[f_1 \frac{(2a_3)^{t/t_0} - (2b_3)^{t/t_1}}{(2a_3)^{t_1} - (2b_3)^{t_0}} - f_6 \frac{(2a_3)^{t/t_0} - (1-c)^{t/t_1}}{(2a_3)^{t_1} - (1-c)^{t_0}} \right], \tag{3.75}$$

where

$$f_1 = \sum_{n=1}^{t_1} (2a_3)^{t_1-n} (2b_3)^{\lfloor (n-1)t_0/t_1 \rfloor} \qquad \text{and} \qquad f_6 = \sum_{n=1}^{t_1} (2a_3)^{t_1-n} (1-c)^{\lfloor (n-1)t_0/t_1 \rfloor}. \tag{3.76}$$

Using (3.73)–(3.75), and noting that for $t \gg t_i$ we can approximate $\lfloor t/t_i \rfloor \approx t/t_i$, the total age-integrated age-structured model solutions (3.65), (3.67) and (3.69) can be written in the form

$$\hat{N}_0(t) \approx \hat{n}_0 (2a_3)^{t/t_0}, \tag{3.77}$$

$$\hat{N}_1(t) \approx \hat{A}(2a_3)^{t/t_0} + (\hat{n}_1 - \hat{A})(2b_3)^{t/t_1}, \tag{3.78}$$

$$\hat{N}_2(t) \approx \hat{B}(2a_3)^{t/t_0} + \hat{C}(2b_3)^{t/t_1} + (\hat{n}_2 - \hat{B} - \hat{C})(1-c)^{t/t_2}, \tag{3.79}$$

where

$$\hat{A} = \frac{2a_2 \hat{n}_0 f_1}{(2a_3)^{t_1} - (2b_3)^{t_0}}, \tag{3.80}$$

$$\hat{B} = \frac{2a_2 \hat{n}_0 (2b_2)}{2b_3 - (1-c)} \left[\frac{f_1}{(2a_3)^{t_1} - (2b_3)^{t_0}} - \frac{f_6}{(2a_3)^{t_1} - (1-c)^{t_0}} \right], \tag{3.81}$$

$$\hat{C} = \frac{2b_2(\hat{n}_1 - \hat{A})}{2b_3 - (1-c)}. \tag{3.82}$$

For comparison, the continuous model solutions from (3.59)–(3.61) are

$$N_0(t) = \hat{n}_0 e^{\alpha t}, \tag{3.83}$$

$$N_1(t) = Ae^{\alpha t} + (\hat{n}_1 - A)e^{\beta t}, \tag{3.84}$$

$$N_2(t) = Be^{\alpha t} + Ce^{\beta t} + (\hat{n}_2 - B - C)e^{-\gamma t}, \tag{3.85}$$

where

$$A = \frac{\alpha_2 \hat{n}_0}{\alpha - \beta}, \qquad B = \frac{\beta_2 A}{\alpha + \gamma} \quad \text{and} \qquad C = \frac{\beta_2(\hat{n}_1 - A)}{\beta + \gamma}. \tag{3.86}$$

The approximated age-integrated age-structured model solutions (3.77)–(3.79) are in a similar form to the continuous model solutions (3.83)–(3.85), and so we can compare terms in order to obtain relations between the two sets of parameters.

3.5.3 Relating the rates and proportions

In this section we derive approximate formulae relating the rate and proportion parameters. The continuous model solutions contain five unknown rate parameters α, β, γ, α_2 and β_2, and the five unknown proportions in the age-structured model solutions are a_3, b_3, c, a_2 and b_2 (a_1 and b_1 can be determined since the a_i and b_i proportions each sum to 1). We assume that the cell cycle times t_i are known. Let us suppose that we know the rate parameters and wish to determine the proportions.

Equating $N_0 = \hat{N}_0$, $N_1 = \hat{N}_1$ and $N_2 = \hat{N}_2$ from (3.77)–(3.79) and (3.83)–(3.85), it is immediately clear that we can identify

$$a_3 = \frac{1}{2}e^{\alpha t_0}, \qquad b_3 = \frac{1}{2}e^{\beta t_1}, \qquad c = 1 - e^{-\gamma t_2}. \tag{3.87}$$

Therefore the proportion a_3 can be determined in terms of α, b_3 in terms of β and c in terms of γ. This leaves three relations $A = \hat{A}$, $B = \hat{B}$ and $C = \hat{C}$ linking the two other proportions a_2 and b_2 in terms of the five rates α, β, γ, α_2 and β_2, so this system is over-determined.

Note that if $\alpha \geq 0$ and $\beta < 0$, then $e^{\alpha t} \gg e^{\beta t}$, $e^{\alpha t} \gg e^{-\gamma t}$, and there is only exponential growth due to stem cells. Therefore we must only compare the coefficients of $e^{\alpha t}$, and so we take $A = \hat{A}$ and $B = \hat{B}$, disregarding $C = \hat{C}$. If $\alpha \leq 0$ and $\beta > 0$, then $e^{\beta t} \gg e^{\alpha t}$, $e^{\beta t} \gg e^{-\gamma t}$, and only N_1 and N_2 grow exponentially. This means we only compare the coefficients of $e^{\beta t}$, giving $\hat{n}_1 - A = \hat{n}_1 - \hat{A}$ and $C = \hat{C}$, disregarding $B = \hat{B}$. Finally if $\alpha > 0$ and $\beta > 0$, then exponential growth is due to both stem and semi-differentiated cells and we take $A = \hat{A}$, $B = \hat{B}$ and $C = \hat{C}$. We can choose any pair of these relations to determine a_2 and b_2 and there will be a consistency condition to satisfy the third relation.

Taking $A = \hat{A}$, and substituting a_3, b_3 and c from (3.87), gives

$$a_2 = \frac{\alpha_2}{2f_1}\left(\frac{e^{\alpha t_0 t_1} - e^{\beta t_0 t_1}}{\alpha - \beta}\right). \tag{3.88}$$

Likewise, taking $B = \hat{B}$ gives

$$b_2 = \frac{\beta_2 f_1 (e^{\beta t_1} - e^{\gamma t_1})}{2(\alpha+\gamma)(e^{\alpha t_0 t_1} - e^{\beta t_0 t_1})}\left[\frac{f_1}{e^{\alpha t_0 t_1} - e^{\beta t_0 t_1}} - \frac{f_6}{e^{\alpha t_0 t_1} - e^{\gamma t_0 t_1}}\right]^{-1}, \tag{3.89}$$

and taking $A = \hat{A}$ and $C = \hat{C}$ together gives

$$b_2 = \frac{\beta_2}{2}\left(\frac{e^{\beta t_1} - e^{-\gamma t_1}}{\beta + \gamma}\right). \tag{3.90}$$

Equating (3.89) and (3.90), we find the consistency condition to be

$$\frac{f_1(\beta+\gamma)}{(\alpha+\gamma)(e^{\alpha t_0 t_1} - e^{\beta t_0 t_1})} = \left[\frac{f_1}{e^{\alpha t_0 t_1} - e^{\beta t_0 t_1}} - \frac{f_6}{e^{\alpha t_0 t_1} - e^{\gamma t_0 t_1}}\right]. \tag{3.91}$$

If we assume that the floor functions can be replaced by their arguments, then we can approximate

$$f_1 \approx \frac{e^{\alpha t_0 t_1} - e^{\beta t_0 t_1}}{e^{\alpha t_0} - e^{\beta t_0}} \qquad \text{and} \qquad f_6 \approx \frac{e^{\alpha t_0 t_1} - e^{-\gamma t_0 t_1}}{e^{\alpha t_0} - e^{-\gamma t_0}}, \tag{3.92}$$

and using this approximation (3.89) reduces to

$$b_2 \approx \frac{\beta_2}{2}\left(\frac{e^{\alpha t_0}-e^{-\gamma t_0}}{\alpha+\gamma}\right)\left(\frac{e^{\beta t_1}-e^{-\gamma t_1}}{e^{\beta t_0}-e^{-\gamma t_0}}\right), \tag{3.93}$$

and the consistency condition reduces to

$$\frac{\beta+\gamma}{\alpha+\gamma} \approx \frac{e^{\beta t_0}-e^{-\gamma t_0}}{e^{\alpha t_0}-e^{-\gamma t_0}}. \tag{3.94}$$

The approximation of the age-structured model by the continuous model assumes that the cell populations are varying on a time scale much longer than that of the cell cycle, that is, that the renewal proportions are close to 1/2. Expanding the exponentials when the arguments are small, we can approximate all the proportions by

$$a_3 \approx \frac{1}{2}(1+\alpha t_0), \quad a_2 \approx \frac{1}{2}\alpha_2 t_0, \quad b_3 \approx \frac{1}{2}(1+\beta t_1), \quad b_2 \approx \frac{1}{2}\beta_2 t_1, \quad c \approx \gamma t_2, \tag{3.95}$$

and the final proportions are determined by

$$a_1 = 1 - a_2 - a_3 \qquad \text{and} \qquad b_1 = 1 - b_2 - b_3. \tag{3.96}$$

These approximations are valid as long as the rates α_i, β_i and γ are $O(1)$ because the exponents in the exponentials are small. In this case both (3.90) and (3.93) simplify to give $b_2 \approx \beta_2 t_1/2$, and the consistency condition (3.94) is always satisfied.

3.6 Discussion

In this chapter we have introduced three types of model that try to capture the behaviour of the cell populations in a healthy crypt in the colon. We have extended the discrete time model of Tomlinson and Bodmer [1995] by producing two different types of models to capture the asynchronous division of cells: the age-structured and continuous models. Each of the solutions to the discrete, age-structured or continuous model has the same overall growth behaviour. We illustrate this point in Figure 3.3 by comparing each model solution for the stem cell population, where we use an initial age profile for the age-structured model where all the cells have the same age.

The discrete model approach is valid when considering only one population, and in the case of multiple populations dividing synchronously. However, it is inappropriate for two or more populations dividing on different time scales, which is precisely the case for the crypt, and so either the age-structured or continuous models are required. If it is necessary to follow specific variations at times on the order of the cell cycle times, then the age-structured model is the better choice. The age-structured model is more detailed than the continuous model but produces more complicated solutions. Alternatively, when a population changes over a time scale of many cell cycles, the continuous model can capture the same overall behaviour with more clarity: ODEs are much easier to solve mathematically than difference equations or PDEs.

It is clear that the solutions to both the age-structured and continuous models are very sensitive to the cell population renewal proportions a_3 and b_3 or net proliferation rates α and β. A non-trivial steady state can occur only if $a_3 = 1/2$, corresponding to $\alpha = 0$. If $a_3 > 1/2$ ($\alpha > 0$) the model exhibits exponential growth, while if $a_3 < 1/2$ ($\alpha < 0$) the

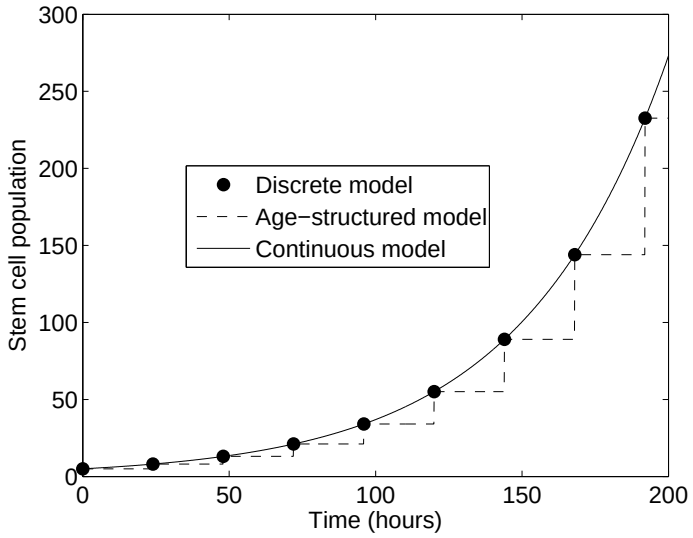

Figure 3.3: A comparison of the solutions for the stem cell population using the discrete model (3.4), the age-structured model (3.65) and the continuous model (3.59). The parameters are taken to be $\alpha = 0.02$ hours^{-1}, $t_0 = 24$ hours and $\hat{n}_0 = 5$. a_3 is chosen such that $a_3 = e^{\alpha t_0}/2$, according to (3.87).

model exhibits exponential decay. In biological systems we expect there to be noise, and so such a model with structural instability is unrealistic. There is a similar but less stringent requirement on b_3 and β. For $b_3 < 1/2$ ($\beta < 0$) there is a nontrivial steady state, while for $b_3 > 1/2$ ($\beta > 0$) there is exponential growth. The problem of structural instability can be resolved by including feedback into the system, and we address this problem in the next chapter.

Chapter 4

Homeostasis and tumourigenesis[1]

Each of the models in Chapter 3 captured the long-term behaviour of the cell populations in the crypt, but small changes to key parameters produced solutions that either grew or decayed exponentially, implying that the models are structurally unstable. In biological systems we expect there to be noise, and so such a model with structural instability will be unrealistic. This instability arose due to the linearity of the models where the rates of differentiation, death and renewal of the cell populations took constant values. Furthermore, if the number of stem cells is decreased following irradiation it has been observed that the crypt rapidly reverts to the homeostatic state [Paulus *et al.*, 1992]. This could be achieved by more stem cells dividing symmetrically, or by de-differentiation of first generation transit cells, but neither scenario can be captured by a linear model. It is not known what physical processes control the feedback in the crypt, although these could include chemical cues in the lumen or epithelial-mesenchymal interactions, which were discussed in Section 1.1.5.

This regulation of cell numbers is broken when tumourigenesis occurs. It is believed that a succession of genetic mutations or epigenetic changes can lead to homeostasis in the crypt being overcome, with subsequent unbounded growth leading to colorectal cancer [Fearon and Vogelstein, 1990].

In this chapter we begin in Section 4.1 by briefly explaining the need to add structural stability to the continuous model. We examine the possible forms that the feedback could take, presenting two specific types, and consider how the results depend on our choices. We show how a model can be developed with the dual aim of modelling homeostasis in the crypt and also explaining how this regulation can be broken down in colorectal cancer, and that instabilities in the model can also be generated through feedback delay. In Section 4.2, we analyse the model parameter space to determine feasible parameter ranges, and we examine how sensitive the results are to parameter changes. Finally, in Section 4.3, we discuss the effect that mutations might have and possible pathways to tumourigenesis via cancer-driving cells originating from stem and/or transit cells. Throughout this chapter we again work with the model in dimensional variables.

4.1 Feedback in the continuous model

In this section we consider the qualitative effects of different feedback models which can be used to describe homeostasis in the crypt, and we analyse the linear stability of two specific types. We discuss which model is more suitable for capturing both homeostasis and

[1]The results of this chapter form the basis of parts of the papers by Johnston *et al.* [2007a,b].

unbounded cell population growth in the crypt, and then we examine the most sensible set of parameters for this model in Section 4.2. The continuous (in time) model, as described in Chapter 3, is used here since we are only interested in the evolution of the cell populations over long time scales.

4.1.1 Illustrating structural instability in the linear model

We begin by illustrating the instability in the linear models introduced in Chapter 3. Recall that for the continuous model the system is

$$\frac{\mathrm{d}N_0}{\mathrm{d}t} = (\alpha_3 - \alpha_1 - \alpha_2)N_0, \tag{4.1}$$

$$\frac{\mathrm{d}N_1}{\mathrm{d}t} = (\beta_3 - \beta_1 - \beta_2)N_1 + \alpha_2 N_0, \tag{4.2}$$

$$\frac{\mathrm{d}N_2}{\mathrm{d}t} = \beta_2 N_1 - \gamma N_2. \tag{4.3}$$

We defined the net per-capita growth rates for the stem and semi-differentiated cell populations to be $\alpha = \alpha_3 - \alpha_1 - \alpha_2$ and $\beta = \beta_3 - \beta_1 - \beta_2$, and the solutions to these equations were given in (3.59)–(3.61) in Chapter 3. It can be clearly seen that if $\alpha > 0$ then there is exponential growth in all populations, and if $\beta > 0$ then there is exponential growth in the semi- and fully-differentiated cell populations. If $\alpha < 0$ then the stem cells decay to zero. Only in the case where $\alpha = 0$ and $\beta < 0$ (and of course $\gamma > 0$) do the cell populations approach a non-zero steady state given by

$$N_0^* = \hat{n}_0, \quad N_1^* = -\frac{\alpha_2 \hat{n}_0}{\beta}, \quad N_2^* = -\frac{\alpha_2 \beta_2 \hat{n}_0}{\beta\gamma}, \tag{4.4}$$

where the stem cell population is equal to its (undetermined) initial value $\hat{n}_0$, and the other populations are determined in terms of this. To reduce this critical dependence on the parameters α and β, we must introduce feedback into the model.

4.1.2 General feedback functions for the rates of conversion

What form might the feedback take in order to maintain crypt homeostasis? In a discrete or age-structured model it is possible that the cell cycle times or population renewing proportions might vary dependent on the number of cells in each population. For our purposes we use the continuous model, and so we make the assumption that the feedback acts through the rate parameters dependent on the cell population sizes. It is possible that the rate parameters are linked to the concentration of a biochemical species (e.g. Wnt, *APC* or β-catenin) whereby if the number of cells changes the threshold level of the substance required for proliferation/differentiation/death changes accordingly [Arino and Kimmel, 1991]. In this section we derive general formulae which all forms of this feedback must satisfy, and then we consider two specific types in Sections 4.1.3 and 4.1.4.

If we allow the rates of conversion to depend on the population densities, then a gener-

alisation of the continuous model equations (4.1)–(4.3) is

$$\frac{\mathrm{d}N_0}{\mathrm{d}t} = [\alpha_3(N_0, N_1, N_2) - \alpha_1(N_0, N_1, N_2) - \alpha_2(N_0, N_1, N_2)]N_0, \tag{4.5}$$

$$\frac{\mathrm{d}N_1}{\mathrm{d}t} = [\beta_3(N_0, N_1, N_2) - \beta_1(N_0, N_1, N_2) - \beta_2(N_0, N_1, N_2)]N_1 + \alpha_2(N_0, N_1, N_2)N_0, \tag{4.6}$$

$$\frac{\mathrm{d}N_2}{\mathrm{d}t} = -\gamma(N_0, N_1, N_2)N_2 + \beta_2(N_0, N_1, N_2)N_1, \tag{4.7}$$

where each of the rates α_i, β_i and γ may vary depending on any of the cell population densities N_0, N_1 or N_2.

It is believed that feedback is more likely to occur within a cell population than between populations (W. F. Bodmer, pers. comm.).[2] We will use this assumption, and we note that we would proceed as below with the same methodology for the more general case, but the results from the reduced system will be easier to interpret and the model will be simpler, with the N_0 equation decoupling. Therefore, we consider a reduced version of the system (4.5)–(4.7), where the cell population rates of conversion only depend on the density of that particular population, and so $\alpha_i = \alpha_i(N_0)$, $\beta_i = \beta_i(N_1)$ and $\gamma = \gamma(N_2)$. We define general functions $f(N_0) = \alpha_3(N_0) - \alpha_1(N_0) - \alpha_2(N_0)$ and $g(N_1) = \beta_3(N_1) - \beta_1(N_1) - \beta_2(N_1)$ to represent the net per-capita growth rates of the stem and semi-differentiated cell populations, respectively. Therefore the reduced system can be written as

$$\frac{\mathrm{d}N_0}{\mathrm{d}t} = f(N_0)N_0, \tag{4.8}$$

$$\frac{\mathrm{d}N_1}{\mathrm{d}t} = g(N_1)N_1 + \alpha_2(N_0)N_0, \tag{4.9}$$

$$\frac{\mathrm{d}N_2}{\mathrm{d}t} = -\gamma(N_2)N_2 + \beta_2(N_1)N_1. \tag{4.10}$$

In this case the equation for N_2 decouples and can be considered separately. We denote all positive roots of the equations $f(x_1) = 0$, $g(x_2) = 0$ and $g(x_3)x_3 + \alpha_2(x_1)x_1 = 0$ by the values x_1, x_2 and x_3, respectively. The steady states of the system (4.8)–(4.9) are

$$(N_0^*, N_1^*) = (0, 0), \quad (0, x_2) \quad \text{and} \quad (x_1, x_3). \tag{4.11}$$

We wish to establish general conditions that must hold for $f(N_0)$ and $g(N_1)$ in order for there to be a biologically feasible steady state in the system. The Jacobian matrix of (4.8)–(4.9) is given by

$$J(N_0^*, N_1^*) = \begin{pmatrix} f(N_0^*) + N_0^* f'(N_0^*) & 0 \\ \alpha_2(N_0^*) + N_0^* \alpha_2'(N_0^*) & g(N_1^*) + N_1^* g'(N_1^*) \end{pmatrix},$$

and so the eigenvalues for linear stability are

$$\lambda_1 = f(N_0^*) + N_0^* f'(N_0^*) \qquad \text{and} \qquad \lambda_2 = g(N_1^*) + N_1^* g'(N_1^*). \tag{4.12}$$

The stability of the steady states (4.11) is summarised in Table 4.1. If we also want a population just of stem cells to exist on its own, we require $f(0) > 0$ and $f'(x_1) < 0$.

[2] Whilst this might be too stringent in practice, feedback is more likely to occur between neighbouring cells and, in particular, the stem cells are likely to be autonomous and only regulated by their own number.

(N_0^*, N_1^*)	Type/Stability	Parameter range
$(0,0)$	Stable node	$f(0) < 0$ and $g(0) < 0$
	Unstable node	$f(0) > 0$ and $g(0) > 0$
	Unstable saddle	$f(0) \gtrless 0$ and $g(0) \lessgtr 0$
$(0, x_2)$	Stable node	$f(0) < 0$ and $g'(x_2) < 0$
	Unstable node	$f(0) > 0$ and $g'(x_2) > 0$
	Unstable saddle	$f(0) \gtrless 0$ and $g'(x_2) \lessgtr 0$
(x_1, x_3)	Stable node	$f'(x_1) < 0$ and $g(x_3) + x_3 g'(x_3) < 0$
	Unstable node	$f'(x_1) > 0$ and $g(x_3) + x_3 g'(x_3) > 0$
	Unstable saddle	$f'(x_1) \gtrless 0$ and $g(x_3) + x_3 g'(x_3) \lessgtr 0$

Table 4.1: Linear stability of the stem and transit cell population steady states (4.11), where x_1, x_2 and x_3 denote any positive roots of the equations $f(x_1) = 0$, $g(x_2) = 0$ and $g(x_3)x_3 + \alpha_2(x_1)x_1 = 0$.

Combining these with the restrictions for normal crypt dynamics from Table 4.1, we require the following conditions:

$$f(0) > 0, \tag{4.13}$$

$$f'(x_1) < 0, \tag{4.14}$$

$$g(x_3) + x_3 g'(x_3) < 0, \tag{4.15}$$

where $f(x_1) = 0$ and $g(x_3)x_3 + \alpha_2(x_1)x_1 = 0$. Condition (4.13) is a sufficient condition for the steady states $(0,0)$ and $(0, x_2)$ to be unstable, and conditions (4.14) and (4.15) ensure that the non-zero steady state (x_1, x_3) is linearly stable. We note that there might also be further requirements to ensure feasible solutions for x_1 and x_3.

Conditions (4.13)–(4.15) must hold for all steady states of N_0 and N_1 to be biologically feasible. There may exist multiple values of x_1 and x_3, although the system might then have complex dynamics, and this is considered in Section 4.3.4. We note that when N_1 is in a stable steady state, N_2 will also evolve to a stable steady state. In the next two sections we consider more specific forms of feedback.

4.1.3 Feedback model 1: linear feedback

One way in which the cells could act to maintain homeostasis is by altering the rate of cell differentiation in response to changes in the cell population sizes. Indeed Arino and Kimmel [1991] noted that the differentiation process is probably directly coupled with population regulation.

We assume that when the populations of stem and semi-differentiated cells increase, the (per-capita) rate at which they differentiate increases in proportion. Thus we replace α_2 and β_2 in (4.1)–(4.3) by, respectively, $\alpha_2 + k_0 N_0$ and $\beta_2 + k_1 N_1$, where $k_i > 0$ are constants. Since the rate at which cells enter the fully-differentiated cell population is regulated by the feedback controlling the transit cell differentiation rate, there is no requirement for feedback within the fully-differentiated cell population, and so the N_2 removal rate γ is assumed to

be constant. This produces the system of equations

$$\frac{\mathrm{d}N_0}{\mathrm{d}t} = \alpha N_0 - k_0 N_0^2, \tag{4.16}$$

$$\frac{\mathrm{d}N_1}{\mathrm{d}t} = \beta N_1 - k_1 N_1^2 + N_0(\alpha_2 + k_0 N_0), \tag{4.17}$$

$$\frac{\mathrm{d}N_2}{\mathrm{d}t} = -\gamma N_2 + N_1(\beta_2 + k_1 N_1), \tag{4.18}$$

where, as before, $\alpha = \alpha_3 - \alpha_1 - \alpha_2$ and $\beta = \beta_3 - \beta_1 - \beta_2$. The stem cell equation (4.16) can be solved directly, with $N_0(0) = \hat{n}_0$, to give

$$N_0(t) = \frac{\alpha}{k_0 + (\alpha/\hat{n}_0 - k_0)\, e^{-\alpha t}}. \tag{4.19}$$

The stem cells exhibit standard logistic growth and the solution evolves to a steady state given by

$$N_0^* = \frac{\alpha}{k_0}, \tag{4.20}$$

which is the stem cell carrying capacity. If $\alpha > 0$, all solutions of (4.16) are attracted to this stable steady state; if $\alpha < 0$, all solutions are attracted to $N_0 = 0$. We see that there are no values of the parameters (providing $k_0 > 0$) that allow unbounded growth in N_0.

For the semi-differentiated cells, when $\alpha < 0$ there are two steady states with all solutions attracted to $N_1 = 0$ if $\beta < 0$ and to $N_1 = \beta/k_1$ if $\beta > 0$. Therefore for biologically realistic solutions we impose $\beta < 0$, otherwise the transit cells will reach a non-zero steady state without any stem cells. When $\alpha > 0$ there is one positive, linearly-stable steady state given by

$$N_1^* = \frac{1}{2k_1}\left[\beta + \sqrt{\beta^2 + 4k_1 N_0^*(\alpha_2 + k_0 N_0^*)}\right], \tag{4.21}$$

which exists for all values of β. As with the stem cells, no values of the parameters lead to exponential growth in the transit cell population. Consequently there is also one positive, stable, fully-differentiated cell population steady state given by

$$N_2^* = \frac{N_1^*}{\gamma}(\beta_2 + k_1 N_1^*). \tag{4.22}$$

We note that the steady state values for the linear feedback model will, in general, not be the same as those for the model without feedback in (4.4) which depend on the initial stem cell value $\hat{n}_0$, whereas (4.20)–(4.22) do not. Exponential growth in the cell populations is no longer possible with linear feedback and can only be achieved if the feedback is removed by setting $k_0 = 0 = k_1$.

In terms of the general functions in (4.8)–(4.9), $f(N_0) = \alpha - k_0 N_0$ and $g(N_1) = \beta - k_1 N_1$, and conditions (4.13)–(4.15) are satisfied if $\alpha > 0$ for all values of β, and the only values for x_1 and x_3 are given by (4.20) and (4.21), respectively.

4.1.3.1 Analysis of the effect of noise

In reality, local factors in the crypt may cause the model parameters to alter continually (noise). One simple method for adding noise into the model is to perturb a rate parameter about a critical value at regular intervals. Suppose we perturb the rate of stem cell

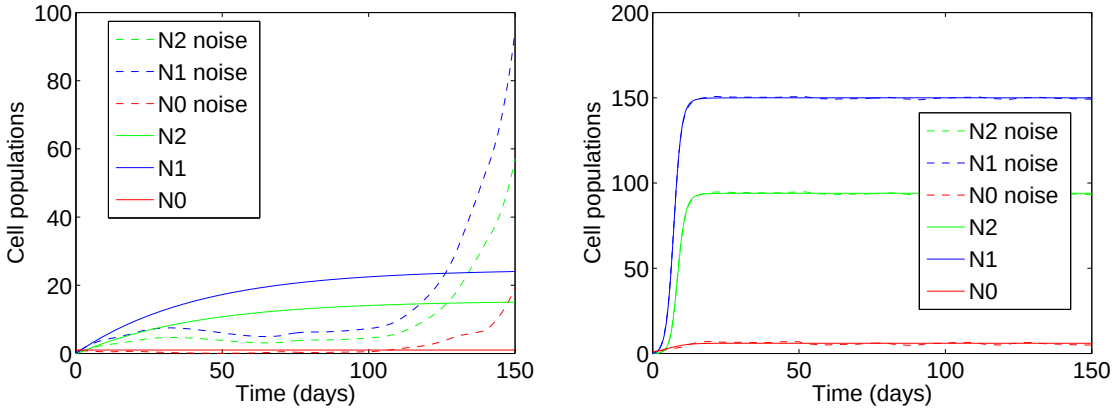

Figure 4.1: Illustrating the effects of noise on the continuous model with and without linear feedback. Left: A comparison of solutions for the model without feedback (4.1)–(4.3) with and without noise. The parameter α_2 is perturbed every 10 days by setting $\alpha_2 = \alpha_{2crit}(1 + R)$, where $R \sim U(-0.25, 0.25)$ is a uniformly distributed random number. The other parameters are taken to be $\alpha_1 = 0.100$, $\alpha_{2crit} = 0.593$, $\alpha_3 = 0.693$, $\beta_1 = 0.263$, $\beta_2 = 1.147$, $\beta_3 = 1.386$, $\gamma = 1.830$, with initial populations $\hat{n}_0 = 1$, $\hat{n}_1 = 0$ and $\hat{n}_2 = 0$. Solutions for the system without noise ($R = 0$) are included for comparison. Right: The same comparison is conducted using the linear feedback model (4.16)–(4.18), where $\alpha_{2crit} = 0.293$, $\beta_2 = 0.397$, $k_0 = 0.05$ and $k_1 = 0.005$.

differentiation (α_2) by setting $\alpha_2 = \alpha_{2crit}(1 + R)$, where $R \sim U(-0.25, 0.25)$ is a uniformly distributed random number.

In the model without feedback we choose α_{2crit} such that $\alpha = \alpha_3 - \alpha_1 - \alpha_{2crit} = 0$, which produces a stable steady state when $R = 0$. Clearly perturbations about this parameter value will lead to the solution fluctuating between exponential growth and approach to a steady state. In Figure 4.1 we compare the model solutions with and without noise for both the model without feedback and the linear feedback model, and we assume that the noise results in α_2 being perturbed every 10 days. If, however, α_{2crit} was chosen to make α positive at all points, regardless of the noise, then exponential growth would occur in the linear model regardless of noise, but the linear feedback model is unaltered, and a steady state is still achieved with or without noise.

4.1.4 Feedback model 2: saturating feedback

As illustrated in the previous section, the linear feedback model solutions remain stable for all parameter values, and their stability is unaffected by parameter perturbations. This accurately captures the observed population dynamics in the crypt when the populations remain roughly constant over time, but unbounded population growth can only be captured when the feedback is switched off. Ideally, we would like to be able to achieve unbounded growth through parameter changes alone and not by changing the structure of the model. In this section we present an alternative form of feedback for the continuous model which can allow unlimited cell population growth in a certain parameter regime.

As before, we assume that when the population of stem or semi-differentiated cells increases the rate at which they differentiate increases, but instead of assuming a linear dependence of the per-capita rate on population size we now assume that there is a maximum per-capita rate of differentiation. As indicated in Section 1.1.5, this could be because of a

short-lived diffusible factor released by cells to help count their number, or due to contact inhibition, which prevents a cell population exceeding a critical size. Thus we replace α_2 and β_2 in (4.1)–(4.3) by, respectively, $\alpha_2 + k_0N_0/(1 + m_0N_0)$ and $\beta_2 + k_1N_1/(1 + m_1N_1)$, where $m_i > 0$ are constants. The k_i parameters represent the rate of population response to change, while the m_i parameters represent feedback saturation. This gives

$$\frac{\mathrm{d}N_0}{\mathrm{d}t} = \alpha N_0 - \frac{k_0N_0^2}{1 + m_0N_0}, \tag{4.23}$$

$$\frac{\mathrm{d}N_1}{\mathrm{d}t} = \beta N_1 - \frac{k_1N_1^2}{1 + m_1N_1} + N_0\left(\alpha_2 + \frac{k_0N_0}{1 + m_0N_0}\right), \tag{4.24}$$

$$\frac{\mathrm{d}N_2}{\mathrm{d}t} = -\gamma N_2 + N_1\left(\beta_2 + \frac{k_1N_1}{1 + m_1N_1}\right), \tag{4.25}$$

where, again, $\alpha = \alpha_3 - \alpha_1 - \alpha_2$ and $\beta = \beta_3 - \beta_1 - \beta_2$. This feedback may be considered as a more general form of the linear feedback model. Note that by setting $m_0 = 0 = m_1$ we return to the linear feedback model equations, (4.16)–(4.18), and by setting $k_0 = 0 = k_1$ we return to the model without feedback, (4.1)–(4.3).

The stem cell equation (4.23) can be solved, with $N_0(0) = \hat{n}_0$, to give

$$\frac{\alpha + (\alpha m_0 - k_0)N_0}{\alpha + (\alpha m_0 - k_0)\hat{n}_0} = \left(\frac{N_0e^{-\alpha t}}{\hat{n}_0}\right)^{1-\alpha m_0/k_0}. \tag{4.26}$$

We find that the extinct state $N_0 = 0$ is unstable for $\alpha > 0$ and globally attracting for $\alpha < 0$; thus we require $\alpha > 0$ for the crypt to be viable. A further steady state of stem cells, corresponding to homeostasis, exists when α lies in the range $0 < \alpha < k_0/m_0$, and is given by

$$N_0^* = \frac{\alpha}{k_0 - m_0\alpha}. \tag{4.27}$$

If $\alpha > k_0/m_0$, this steady state is biologically unfeasible and the stem cell population grows unboundedly.

When the stem cell population is in steady state, the behaviour of the semi- and fully-differentiated cell populations depends on whether β is greater or less than the critical value k_1/m_1. If $\beta < k_1/m_1$, then there is a single, real, positive steady state for the semi- and fully-differentiated cell populations given by

$$N_1^* = \frac{1}{2(k_1 - \beta m_1)}\left[\beta + m_1D + \sqrt{(\beta - m_1D)^2 + 4Dk_1}\right], \tag{4.28}$$

$$N_2^* = \frac{N_1^*}{\gamma}\left(\beta_2 + \frac{k_1N_1^*}{1 + m_1N_1^*}\right), \tag{4.29}$$

where $D = \alpha_2N_0^* + k_0N_0^{*2}/(1 + m_0N_0^*) = (\alpha_3 - \alpha_1)N_0^*$ is the stem-cell differentiation rate. If $\beta > k_1/m_1$, then the semi- and fully-differentiated cell populations exhibit unbounded growth. As with the linear feedback model, the saturating feedback satisfies the equations (4.8)–(4.9), where

$$f(N_0) = \alpha - \frac{k_0N_0}{1 + m_0N_0} \qquad \text{and} \qquad g(N_1) = \beta - \frac{k_1N_1}{1 + m_1N_1}. \tag{4.30}$$

The conditions (4.13)–(4.15) are again satisfied if $\alpha > 0$ for all values of β, and the only values for x_1 and x_3 are given by (4.27) and (4.28), respectively.

In summary, this model permits unbounded growth via a change of parameters or by removing the feedback, which makes this more flexible than the linear feedback model.

4.1.5 Alternative forms of feedback

The linear and saturating feedbacks represent two forms of regulation that could maintain homeostasis in the crypt, but do these span all possible behaviours? In this section we investigate to what extent our choice of feedback has dictated the system dynamics, and whether other forms of feedback would produce qualitatively different behaviour.

4.1.5.1 Saturating feedback in death or renewal

So far we have assumed that the feedback would act through the rate of cell differentiation by attaching a term of the form $k_i N_i^2/(1+m_i N_i)$ to the differentiation rates, but would it change the biological predictions of the model if this term was associated with the death or renewal instead?

Suppose that we replace the death rates α_1 and β_1 in (4.1)–(4.3) by, respectively, $\alpha_1 + k_0 N_0/(1+m_0 N_0)$ and $\beta_1 + k_1 N_1/(1+m_1 N_1)$, where $k_i, m_i > 0$ are constants. Alternatively, if the saturating feedback was included in the renewal rates, then we replace the rates α_3 and β_3 in (4.1)–(4.3) by, respectively, $\alpha_3 - k_0 N_0/(1+m_0 N_0)$ and $\beta_3 - k_1 N_1/(1+m_1 N_1)$, where $k_i, m_i > 0$ are constants. Here we require the saturating renewal rates $\alpha_3 - k_0/m_0$ and $\beta_3 - k_1/m_1$ to be positive in order to be biologically feasible. The resultant equations are of the same form for both these cases, and are given by

$$\frac{\mathrm{d}N_0}{\mathrm{d}t} = \alpha N_0 - \frac{k_0 N_0^2}{1+m_0 N_0}, \tag{4.31}$$

$$\frac{\mathrm{d}N_1}{\mathrm{d}t} = \beta N_1 - \frac{k_1 N_1^2}{1+m_1 N_1} + \alpha_2 N_0, \tag{4.32}$$

$$\frac{\mathrm{d}N_2}{\mathrm{d}t} = -\gamma N_2 + \beta_2 N_1, \tag{4.33}$$

where, as before, $\alpha = \alpha_3 - \alpha_1 - \alpha_2$ and $\beta = \beta_3 - \beta_1 - \beta_2$. In this case the stem cell steady states and conclusions are unaltered, with the non-zero steady state given by (4.27). The functional forms for the steady states of semi- and fully-differentiated cells are altered slightly, given by

$$N_1^* = \frac{1}{2(k_1 - \beta m_1)}\left[\beta + m_1\alpha_2 N_0^* + \sqrt{(\beta - m_1\alpha_2 N_0^*)^2 + 4\alpha_2 k_1 N_0^*}\right], \tag{4.34}$$

$$N_2^* = \frac{\beta_2 N_1^*}{\gamma}. \tag{4.35}$$

The only real change here is that the differentiation rate is now constant and not changing. In conclusion, the steady state sizes will be altered, but the overall stability behaviour is unchanged.

4.1.5.2 Hill function feedback

Alternatively, a different form of feedback function can be chosen altogether. In Sections 4.1.3 and 4.1.4 we assumed that the stem cell differentiation rate was a monotonic increasing function of the stem cell population size, but the functional form could take many different shapes. The differentiation rate increased linearly with linear feedback, and with a Michaelis-Menten type functional form for saturating feedback, but one alternative is a sigmoidal shape. If we take a Hill function form for the differentiation rate, we replace the

differentiation rates α_2 and β_2 in (4.1)–(4.3) by, respectively, $\alpha_2 + k_0 N_0^n/(m_0^n + N_0^n)$ and $\beta_2 + k_1 N_1^n/(m_1^n + N_1^n)$, where $k_i, m_i > 0$ are constants and $n \in \mathbb{N}$ is a predetermined index. Here we have chosen the same coefficients in the Hill functions for both the stem and transit cells for simplicity, but they could equally be represented by different coefficients. Now the system becomes

$$\frac{\mathrm{d}N_0}{\mathrm{d}t} = \alpha N_0 - \frac{k_0 N_0^{n+1}}{m_0^n + N_0^n}, \tag{4.36}$$

$$\frac{\mathrm{d}N_1}{\mathrm{d}t} = \beta N_1 - \frac{k_1 N_1^{n+1}}{m_1^n + N_1^n} + N_0 \left[\alpha_2 + \frac{k_0 N_0^n}{m_0^n + N_0^n}\right], \tag{4.37}$$

$$\frac{\mathrm{d}N_2}{\mathrm{d}t} = -\gamma N_2 + N_1 \left[\beta_2 + \frac{k_1 N_1^n}{m_1^n + N_1^n}\right]. \tag{4.38}$$

The non-zero stem cell steady state is given by

$$N_0^* = m_0 \left(\frac{\alpha}{k_0 - \alpha}\right)^{1/n}. \tag{4.39}$$

The semi-differentiated cell population steady state N_1^* is determined by

$$(\beta - k_1)(N_1^*)^{n+1} + D(N_1^*)^n + \beta m_1^n N_1^* + D m_1^n = 0, \tag{4.40}$$

where $D = N_0^*[\alpha_2 + k_0 (N_0^*)^n/(m_0^n + (N_0^*)^n)]$ is the stem cell differentiation rate. The second and fourth terms in (4.40) are always positive, and the first and third could be either positive or negative. We apply Descartes' Rule of Signs to count the number of sign changes, which is equivalent to the number of positive roots of the equation. In the case where $\beta > k_1$, each term in (4.40) is positive, which means that there are no positive steady states. If $0 < \beta < k_1$, there is one positive steady state, whereas if $\beta < 0$ there could either be one or three steady states. This offers the possibility of bistability, but it is questionable whether this is biologically realistic. We discuss this further in Section 4.3.4.

4.1.5.3 Comparison with other feedback models

The requirement for feedback has been recognised in other models of the crypt, as we reviewed in Chapter 2.

Wodarz [2007] used logistic growth of stem cells, which is equivalent to our linear feedback presented in Section 4.1.3. Clearly this model captures homeostasis but cannot explain unbounded cell population growth.

Boman *et al.* [2001] assumed that feedback from the fully-differentiated cell population regulates the transit cells entering the cell cycle by the expression $1/(1 + N_2)$, and feedback from the dead cells controls the semi- and fully-differentiated cell population sizes by the expression $(1 + N_3^2)$. These forms are not biologically motivated and it is not clear why dead cells should influence the dynamics. In addition, the stem cell equation (2.14) is structurally unstable and not regulated which is not biologically realistic.

D'Onofrio and Tomlinson [2007] included feedback in the transit cell parameters, with functional forms that satisfied $b_1'(N_1) < 0$, $b_2'(N_1) \geq 0$ and $b_3'(N_1) > 0$, which is in contrast to our choices of either $\beta_1'(N_1) > 0$, $\beta_2'(N_1) > 0$ or $\beta_3'(N_1) < 0$, since it is expected that the net growth rate decreases as the cell population size increases [Calabresi *et al.*, 1985]. The

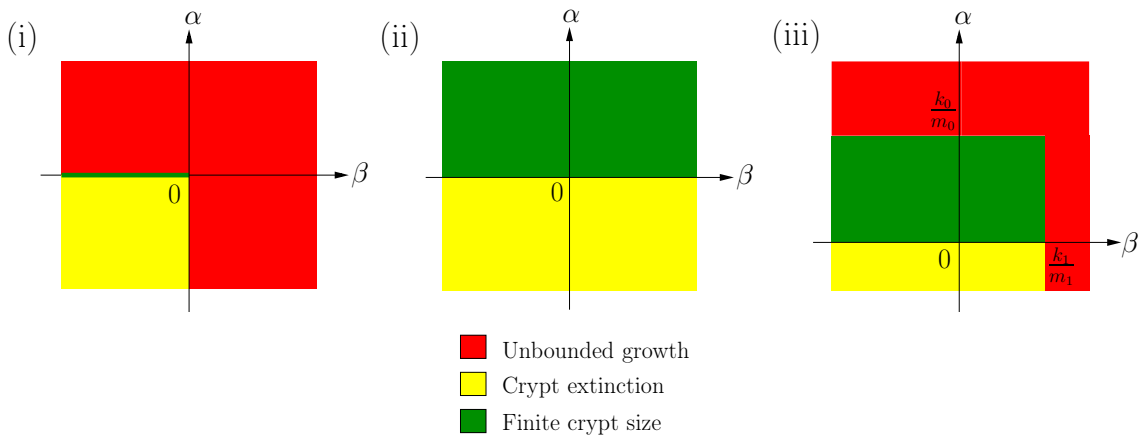

Figure 4.2: Illustrating the regions of linear stability for the cell population models with (i) no feedback; (ii) linear feedback and (iii) saturating feedback. In the case of no feedback, (i), there are only linearly stable solutions on the line $\alpha = 0$, $\beta < 0$; otherwise there is either extinction or unbounded growth. In the case of the linear feedback, (ii), there is a bounded steady state if $\alpha > 0$ and extinction if $\alpha < 0$. For the saturating feedback, (iii), there is a strip of the parameter space that permits linearly stable solutions, when $0 < \alpha < k_0/m_0$ and $\beta < k_1/m_1$, and outside this region there is either extinction or unbounded growth.

authors also tested several other types of feedback: they assumed that $b_i = b_i(N_1, N_2)$ for $i = 1, 2, 3$, where

$$\frac{\partial b_1}{\partial N_1} < 0, \quad \frac{\partial b_1}{\partial N_2} < 0, \quad \frac{\partial b_2}{\partial N_1} \geq 0, \quad \frac{\partial b_2}{\partial N_2} \geq 0, \quad \frac{\partial b_3}{\partial N_1} > 0, \quad \frac{\partial b_3}{\partial N_2} > 0, \tag{4.41}$$

in particular when $b_i = b_i(N_1 + N_2)$. In addition they also tested functions of N_1/N_0 and $N_1/(N_1 + N_0)$. In each case the same qualitative behaviour was observed. It is possible that feedback could depend on the total number of cells in the crypt, $N_0 + N_1 + N_2$, or on the total number of the proliferating cells $N_0 + N_1$.

There is no biological evidence for any particular feedbacks, and so we chose the simplest form that can capture the required behaviour of homeostasis and unbounded growth.

4.1.6 Stability of the different models

In this section we summarise the behaviour of the models from Sections 4.1.1, 4.1.3 and 4.1.4 by comparing their stability in different regions of the (α, β)-parameter space, which is illustrated in Figure 4.2.

For the model without feedback, the line $\alpha = 0$, $\beta < 0$ corresponds to finite crypt size, the region $\alpha < 0$ and $\beta < 0$ corresponds to crypt extinction, and the regions $\alpha > 0$ and $\beta > 0$ correspond to unbounded growth. For the linear feedback model, the region $\alpha > 0$ corresponds to finite crypt size, and the region $\alpha < 0$ corresponds to crypt extinction. Finally, for the saturating feedback model, the region $\alpha < 0$ corresponds to crypt extinction, the region $0 < \alpha < k_0/m_0$ and $\beta < k_1/m_1$ corresponds to finite crypt size, and the regions $\alpha > k_0/m_0$ and $\beta > k_1/m_1$ correspond to unbounded growth.

These parameter regimes can be explained biologically in the following way. For the model without feedback, a linearly stable steady state is achieved only if the stem cell renewal rate is exactly equal to the sum of the death and differentiation rates $(\alpha_1 + \alpha_2)$.

Otherwise, if the renewal rate is greater than this value exponential growth will occur, and if it is less than that value the stem cell population will decrease to zero.

For the linear feedback model, providing the renewal rate of stem cells, α_3, is bigger than some critical size $(\alpha_1+\alpha_2)$, the stem cell population is able to sustain itself and reaches a linearly stable non-zero steady state, otherwise the stem cell population decays to zero. Small changes in any of the parameters will change the size of the steady state but will never lead to unregulated growth.

For the saturating feedback model, if the stem cell renewal rate is less than some lower critical value $(\alpha_1+\alpha_2)$ then the stem cell population cannot maintain itself and will reduce to zero. If the renewal rate is greater than this value, but less than an upper critical value $(\alpha_1+\alpha_2+k_0/m_0)$, then a stem cell steady state is found. If the renewal rate is above this upper value, then no steady state can be found and exponential growth results. Small parameter changes will alter the steady state size unless the change results in the renewal being greater than the upper value, in which case a steady state is no longer possible. The same is true for the semi-differentiated cells if their renewal rate is above/below $\beta_1+\beta_2+k_1/m_1$. Therefore, in this saturating feedback model, unbounded cell population growth is achievable if the cell renewal rates exceed a critical value.

In summary, a change in the parameters will lead to a change in the size of the steady state using either feedback model. However, with the saturating feedback model, if the parameter change is sufficiently large to push the critical parameters α and β above a critical threshold, then the system will change from a steady state to uncontrolled cell population growth, and the feedback is not able to overcome this growth. No amount of parameter change could create this instability with linear feedback, since that feedback always overcomes the tendency for the cell numbers to deviate from their expected size, and unbounded growth can only occur if the feedback is removed. In Section 4.3 we discuss which of these models is most appropriate for modelling colorectal tumourigenesis.

4.1.7 Instability due to feedback delay

Another explanation for the occurrence of instability in cell population numbers could be if the feedback acts with a time delay. This theory has been much studied, notably by Mackey and Glass [1977] and Mackey [1978].

Let us suppose that the saturating feedback acts after a time delay τ_0 in the stem cell population and a delay τ_1 in the transit cell population. Thus we replace α_2 and β_2 in (4.1)–(4.3) by, respectively, $\alpha_2+k_0N_0(t-\tau_0)/(1+m_0N_0(t-\tau_0))$ and $\beta_2+k_1N_1(t-\tau_1)/(1+m_1N_1(t-\tau_1))$. Hence

$$\frac{\mathrm{d}N_0(t)}{\mathrm{d}t} = \alpha N_0(t) - \frac{k_0N_0(t-\tau_0)N_0(t)}{1+m_0N_0(t-\tau_0)}, \tag{4.42}$$

$$\frac{\mathrm{d}N_1(t)}{\mathrm{d}t} = \beta N_1(t) - \frac{k_1N_1(t-\tau_1)N_1(t)}{1+m_1N_1(t-\tau_1)} + N_0(t)\left[\alpha_2 + \frac{k_0N_0(t-\tau_0)}{1+m_0N_0(t-\tau_0)}\right]. \tag{4.43}$$

This system produces the same steady states as before, given by (4.27)–(4.29). We perform a linear stability analysis of the steady states by setting $N_0(t) = N_0^* + \epsilon u(t)$ and $N_1(t) =$

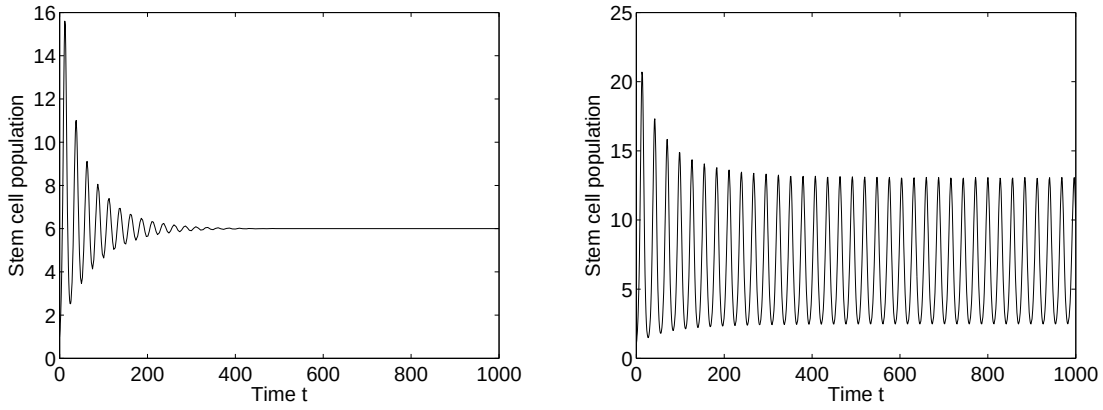

Figure 4.3: Illustrating the effect of feedback delay on the stem cell population from (4.42). The parameters are taken to be $\alpha_1 = 0.100$, $\alpha_2 = 0.218$, $\alpha_3 = 0.693$, $k_0 = 0.1$ and $m_0 = 0.1$, which produces a steady state value of $N_0^* = 6$ and a critical time delay $\tau_{0crit} = 6.7$ using (4.47). Left: the case where $\tau_0 = 6$. Right: the case where $\tau_0 = 7$.

$N_1^* + \epsilon v(t)$, where $\epsilon \ll 1$, which gives

$$\frac{\mathrm{d}u(t)}{\mathrm{d}t} = \frac{\alpha m_0}{k_0}\left(\alpha - \frac{k_0}{m_0}\right) u(t-\tau_0) + O(\epsilon), \tag{4.44}$$

$$\frac{\mathrm{d}v(t)}{\mathrm{d}t} = v(t)\left[\beta - \frac{k_1 N_1^*}{1+m_1 N_1^*}\right] - \frac{k_1 N_1^* v(t-\tau_1)}{(1+m_1 N_1^*)^2} + u(t)\left[\alpha_2 + \frac{k_0 N_0^*}{1+m_0 N_0^*}\right] + \frac{k_0 N_0^* u(t-\tau_0)}{(1+m_0 N_0^*)^2} + O(\epsilon). \tag{4.45}$$

Firstly, let us consider the stem cell equation (4.44), and non-dimensionalise time with the stem cell delay, $t \sim \tau_0$. We seek leading order solutions of the form $u(t) = Ae^{\sigma t}$, which produces the transcendental equation

$$\sigma = -Ke^{-\sigma}, \quad \text{where} \quad K = \frac{\tau_0 \alpha m_0}{k_0}\left(\frac{k_0}{m_0} - \alpha\right). \tag{4.46}$$

Instability occurs if $\Re(\sigma) > 0$, and the instability is oscillatory if $\Im(\sigma) \neq 0$. If $\sigma = i\Omega$, then $\cos\Omega = 0$ and $\Omega = K\sin\Omega$, which means that the first occurrence of instability is when $K = \pi/2$. This means there is a Hopf bifurcation if $K > \pi/2$, or, in dimensional terms, if

$$\tau_0 > \tau_{0crit} = \frac{\pi}{2\alpha(1-\alpha m_0/k_0)}. \tag{4.47}$$

Therefore, oscillations in the stem cell population are possible in the saturating feedback model if the delay is large enough ($\tau_0 > \tau_{0crit}$), and in the linear feedback model $\tau_{0crit} = \pi/(2\alpha)$. Since $\sigma \approx i\pi/2$ near the bifurcation, in dimensional terms $\Re(u(t)) = A\cos(2\pi t/(4\tau_0))$, and so the time period for oscillations is approximately $4\tau_0$. We illustrate this instability in Figure 4.3, where two values of τ_0 are chosen to show the oscillatory steady state as both stable and unstable.

If we write $\sigma = \sigma_1 + i\sigma_2$, then $\sigma_1 = -Ke^{-\sigma_1}\cos\sigma_2$ and $\sigma_2 = Ke^{-\sigma_1}\sin\sigma_2$ parametrically define the roots σ of (4.46) in the complex plane in terms of the parameter K. This is plotted in Figure 4.4. In particular, we note that σ is real if $K < 1/e$.

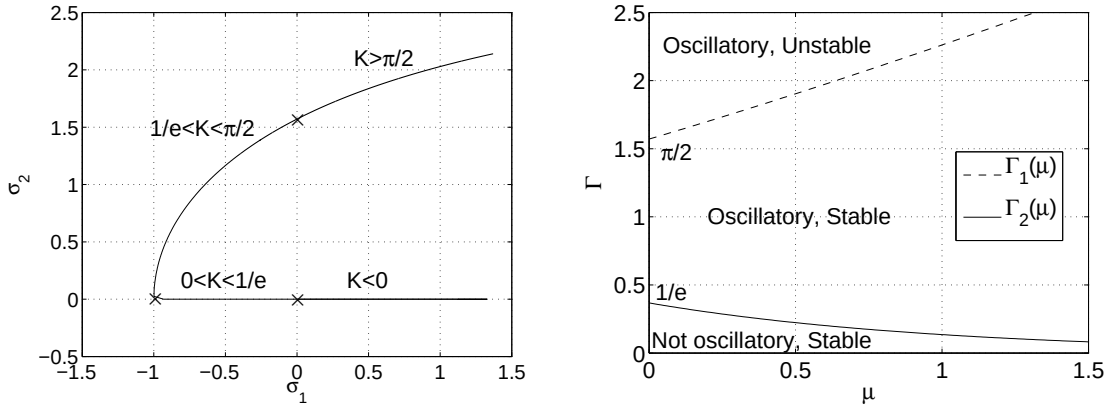

Figure 4.4: Steady state stability in the stem and transit cell populations with a time delay. Left: The complex plane $\sigma = \sigma_1 + i\sigma_2$ for the stem cell population from (4.46) in terms of the parameter K. Right: The (μ, Γ)-space for the transit cell population from (4.49), where $\Gamma_1(\mu)$ and $\Gamma_2(\mu)$ are defined by, respectively, $\cos^{-1}(\mu/\Gamma_1) = \sqrt{\Gamma_1^2 - \mu^2}$ and $\Gamma_2 = e^{-(\mu+1)}$.

Secondly, we consider the transit cell equation (4.45), where leading order solutions $u(t) = Ae^{\sigma t}$ and $v(t) = Be^{\sigma t}$ satisfy

$$\sigma B = B\left[\beta - \frac{k_1 N_1^*}{1 + m_1 N_1^*}\right] - \frac{k_1 N_1^* B e^{-\tau_1 \sigma}}{(1 + m_1 N_1^*)^2} + A\left[\alpha_2 + \frac{k_0 N_0^*}{1 + m_0 N_0^*}\right] + \frac{k_0 N_0^* A e^{-\tau_0 \sigma}}{(1 + m_0 N_0^*)^2}. \quad (4.48)$$

We can find σ from (4.46), and (4.48) can be used to find B as a function of A, which is a measure of the impact of the N_0 time delay on the N_1 population.

If we assume that the stem cell population is in steady state, setting $A = 0$, and non-dimensionalise time with the transit cell delay $t \sim \tau_1$, then (4.48) reduces to

$$\sigma = -\mu - \Gamma e^{-\sigma}, \quad \text{where} \quad \mu = \frac{D\tau_1}{N_1^*}, \quad \Gamma = \frac{\tau_1 (D + \beta N_1^*)^2}{k_1 (N_1^*)^3}. \quad (4.49)$$

We note that $\mu, \Gamma > 0$. A Hopf bifurcation occurs when $\sigma = i\Omega$, which gives $\mu = -\Gamma \cos\Omega$ and $\Omega = \Gamma \sin\Omega$. Therefore the transition curve to instability in the (μ, Γ)-space is found by eliminating Ω to be $\Gamma = \Gamma_1(\mu)$, where $\cos^{-1}(\mu/\Gamma_1) = \sqrt{\Gamma_1^2 - \mu^2}$. In dimensional variables, the condition for instability is

$$\tau_1 > \tau_{1crit} = \frac{k_1 (N_1^*)^3}{\sqrt{(D + \beta N_1^*)^4 - D^2 k_1^2 (N_1^*)^4}} \cos^{-1}\left(-\frac{D k_1 (N_1^*)^2}{(D + \beta N_1^*)^2}\right). \quad (4.50)$$

Writing $\sigma = \sigma_1 + i\sigma_2$, we equate real and imaginary parts and eliminate σ_1 to give $\Gamma = (\sigma_2 / \sin\sigma_2) \exp(-\mu - \sigma_2 / \tan\sigma_2)$. The solutions are not oscillatory as $\sigma_2 \to 0$ when $\Gamma \to e^{-(\mu+1)}$. Therefore, the transition to oscillations in the (μ, Γ)-space occurs when $\Gamma = \Gamma_2(\mu)$, where $\Gamma_2 = e^{-(\mu+1)}$. This is illustrated in Figure 4.4.

If we take the parameters to be $\alpha_1 = 0.100$, $\alpha_2 = 0.218$, $\alpha_3 = 0.693$, $\beta_1 = 0.263$, $\beta_2 = 0.547$, $\beta_3 = 1.386$, $\gamma = 1.830$, $k_0 = 0.1$, $k_1 = 0.01$, $m_0 = 0.1$ and $m_1 = 0.01$, the critical delays are $\tau_{0crit} = 6.7$ and $\tau_{1crit} = 7.0$. The origin of this set of parameters is explained in the next section. This suggests that a delay in N_0 is slightly more destabilising than a delay in N_1. In this section we have shown that instability can also occur through time delays as well as through parameter changes.

4.2 Parameter space analysis in steady state

In this section we analyse the steady state of the saturating feedback model to obtain parameter estimations and predict feasible parameter regimes for the rate parameters, and we determine the key parameters in the linear and saturating feedback models in Section 4.2.3. We wish to find approximate values for the 11 parameters α_1, α_2, α_3, β_1, β_2, β_3, γ, k_0, k_1, m_0 and m_1, in terms of known data relating to the crypt, and we work with dimensional parameters throughout.

4.2.1 Incorporating measurable biological data

Firstly, we know estimates for the cell population steady states, denoted by N_0^*, N_1^* and N_2^*, and so using (4.23)–(4.25) these are linked by the equations

$$\alpha_3 - \alpha_1 = \alpha_2 + \frac{k_0 N_0^*}{1 + m_0 N_0^*}, \tag{4.51}$$

$$N_1^* \left(\beta_3 - \beta_1 - \beta_2 - \frac{k_1 N_1^*}{1 + m_1 N_1^*} \right) + N_0^* \left(\alpha_2 + \frac{k_0 N_0^*}{1 + m_0 N_0^*} \right) = 0, \tag{4.52}$$

$$-\gamma N_2^* + N_1^* \left(\beta_2 + \frac{k_1 N_1^*}{1 + m_1 N_1^*} \right) = 0. \tag{4.53}$$

Secondly, we know estimates for the cell cycle times t_0 and t_1. For the purposes of this section, we define the stem cell renewal rate, α_3, to be the total rate of cell production by the stem cells and so, in the absence of death and differentiation, we have

$$\alpha_3 = \frac{\log 2}{t_0}. \tag{4.54}$$

Similarly, for the semi-differentiated cells, we obtain

$$\beta_3 = \frac{\log 2}{t_1}. \tag{4.55}$$

In addition, the crypt 'turnover time' could be measured; that is, it is expected that roughly 1000 cells will leave the crypt in 3 days (W. F. Bodmer, pers. comm.). This could either be interpreted as the time it takes for a cell to move from the bottom of the crypt to the top, or that there is a flux of cells out of the top of the crypt that consists of the total number of cells in the crypt (minus stem cells) over that time period. In theory, experiments could be designed in order to determine both these parameters (although we can only determine one of them from the turnover time). For the purposes of these calculations we assume that both parameters are known, and then we determine feasible ranges of their values in Section 4.2.2.1. In order to model the sloughing of cells at the top of the crypt, and the crypt turnover time, in this continuous model without a space component, we will use an ad hoc approach to gain an estimate for these parameters. We are aware that a spatial model is required to accurately describe sloughing, but this is a simple attempt to use that available data with this model.

The flux of cells being removed at the top of the crypt, V_T, is given by

$$V_T = \gamma N_2^*. \tag{4.56}$$

We also define the residence time T_r to be the length of time taken by a cell leaving the stem cell population to be ejected from the top of the crypt. We assume that the cell

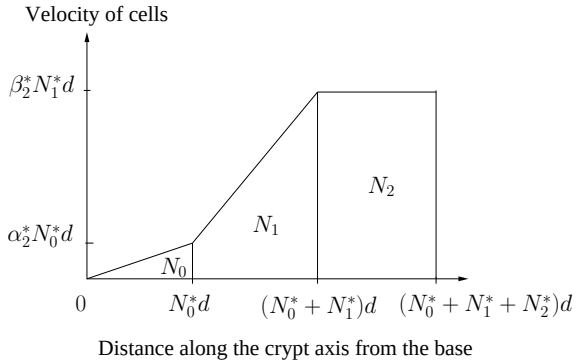

Figure 4.5: Schematic of the velocity of cells along the crypt axis according to (4.57). The velocity is taken to be linear within the cell populations in the crypt, and the steady state differentiation rates α_2^* and β_2^* are defined in (4.60).

velocity increases linearly within the N_0 and N_1 compartments, which is shown in Figure 4.5, and we use this to calculate the residence time. We take the diameter of an average cell (assumed fixed for all cell populations) to be d, denote the distance up the crypt wall from the base by x, and assume that there is no mixing at boundaries between cell populations.

The velocity at the $N_0 : N_1$ boundary (when $x = N_0^* d$) will be $N_0^* d[\alpha_2 + k_0 N_0^*/(1 + m_0 N_0^*)]$, which is the flow of differentiating cells out of the stem cell compartment. The velocity at the $N_1 : N_2$ boundary (when $x = (N_0^* + N_1^*)d$) will be $N_1^* d[\beta_2 + k_1 N_1^*/(1 + m_1 N_1^*)]$, and this will also be the velocity at the top of the crypt (when $x = (N_0^* + N_1^* + N_2^*)d$), since there is no cell production or loss in this compartment. Therefore the velocity profile $v(x)$ is given by

$$v(x) = \begin{cases} \left[\left(\beta_2 + \dfrac{k_1 N_1^*}{1 + m_1 N_1^*}\right) - \dfrac{N_0^*}{N_1^*}\left(\alpha_2 + \dfrac{k_0 N_0^*}{1 + m_0 N_0^*}\right)\right](x - N_0^* d) \\ \qquad + \left(\alpha_2 + \dfrac{k_0 N_0^*}{1 + m_0 N_0^*}\right) N_0^* d, \qquad \text{if} \quad N_0^* d \le x \le (N_0^* + N_1^*)d, \\ \left(\beta_2 + \dfrac{k_1 N_1^*}{1 + m_1 N_1^*}\right) N_1^* d, \qquad \text{if} \quad (N_0^* + N_1^*)d \le x \le (N_0^* + N_1^* + N_2^*)d. \end{cases} \tag{4.57}$$

Since $\dot{x} = v(x)$, the crypt residence time T_r is found by integrating the reciprocal of the velocity of cells from the edge of the stem cell compartment to the top of the crypt,

$$T_r = \int_{N_0^* d}^{(N_0^* + N_1^* + N_2^*)d} \frac{\mathrm{d}x}{v(x)}, \tag{4.58}$$

which evaluates to

$$T_r = \frac{N_1^*}{N_1^*[\beta_2 + k_1 N_1^*/(1 + m_1 N_1^*)] - N_0^*[\alpha_2 + k_0 N_0^*/(1 + m_0 N_0^*)]} \times \log\left(\frac{N_1^*[\beta_2 + k_1 N_1^*/(1 + m_1 N_1^*)]}{N_0^*[\alpha_2 + k_0 N_0^*/(1 + m_0 N_0^*)]}\right) + \frac{N_2^*}{N_1^*(\beta_2 + k_1 N_1^*/(1 + m_1 N_1^*))}. \tag{4.59}$$

We have amassed 7 equations (4.51), (4.52), (4.53), (4.54), (4.55), (4.56) and (4.59) relating the 11 parameters α_1, α_2, α_3, β_1, β_2, β_3, γ, k_0, k_1, m_0 and m_1 to the 7 experimentally determinable parameters N_0^*, N_1^*, N_2^*, t_0, t_1, V_T and T_r. We would need 4 more

relations to determine this system of 11 parameters uniquely, but we note that in these equations the quantities α_2, β_2, k_0, k_1, m_0 and m_1 only appear in the fixed combinations $\alpha_2 + k_0 N_0^*/(1 + m_0 N_0^*)$ and $\beta_2 + k_1 N_1^*/(1 + m_1 N_1^*)$, which represent the steady state differentiation rates for stem and semi-differentiated cells, respectively. If we introduce

$$\alpha_2^* = \alpha_2 + \frac{k_0 N_0^*}{1 + m_0 N_0^*} \qquad \text{and} \qquad \beta_2^* = \beta_2 + \frac{k_1 N_1^*}{1 + m_1 N_1^*}, \tag{4.60}$$

and rearrange the equations, the system becomes

$$\alpha_3 = \frac{\log 2}{t_0}, \quad \beta_3 = \frac{\log 2}{t_1}, \quad \beta_2^* = \frac{V_T}{N_1^*}, \quad \gamma = \frac{V_T}{N_2^*}, \tag{4.61}$$

$$T_r = \frac{N_1^*}{V_T - \alpha_2^* N_0^*} \ln\left(\frac{V_T}{\alpha_2^* N_0^*}\right) + \frac{N_2^*}{V_T}, \tag{4.62}$$

$$\alpha_1 = \frac{\log 2}{t_0} - \alpha_2^*, \tag{4.63}$$

$$\beta_1 = \frac{\log 2}{t_1} - \frac{V_T}{N_1^*} + \frac{N_0^*}{N_1^*}\alpha_2^*. \tag{4.64}$$

There are now 7 equations for the 7 unknowns α_1, α_2^*, α_3, β_1, β_2^*, β_3 and γ, which can be solved uniquely in terms of the 7 'known' quantities N_0^*, N_1^*, N_2^*, t_0, t_1, V_T and T_r. The rates α_3, β_3, β_2^* and γ can all be found directly from (4.61), whereas α_2^* is determined implicitly by (4.62), and then α_1 and β_1 and be found in terms of α_2^* from (4.63) and (4.64), respectively. Plots of the parameters are shown in Figure 4.6, where T_r, V_T, t_0 and t_1 are varied. From this figure it appears that rough parameter sizes are given by $\gamma > \beta_3 > \beta_2^* > \alpha_3 \approx \alpha_2^* > \alpha_1 \approx \beta_1$.

There are, however, no fixed values for the parameters N_0^*, N_1^*, N_2^*, t_0, t_1, V_T or T_r, with a wide range of possible values suggested in the literature, which are summarised in Appendix A. We therefore wish to predict ranges of possible values for the rate parameters based on the results of this model.

4.2.2 Parameter estimation

In this section we analyse the system (4.61)–(4.64) to find feasible parameter ranges for the 7 parameters α_1, α_2^*, α_3, β_1, β_2^*, β_3 and γ.

4.2.2.1 Parameter restrictions

We assume that t_0, t_1, N_0^*, N_1^* and N_2^* are known parameters and we find ranges for the 7 rate parameters α_1, α_2^*, α_3, β_1, β_2^*, β_3 and γ when T_r and V_T vary. Since t_0, t_1, N_0^*, N_1^*, N_2^*, T_r, $V_T > 0$, the four rates α_3, β_2^*, β_3 and γ will always be positive, from (4.61). In Figure 4.7 we plot the contours and surfaces of α_1, α_2^* and β_1 as functions of T_r and V_T. Positivity of the rates α_1 and α_2^* from (4.63) implies that

$$\alpha_1 \leq \frac{\log 2}{t_0}, \qquad \alpha_2^* \leq \frac{\log 2}{t_0}, \tag{4.65}$$

and applying this to (4.64) gives

$$\beta_1 \leq \frac{\log 2}{t_1} - \frac{V_T}{N_1^*} + \frac{N_0^*}{N_1^*}\frac{\log 2}{t_0}. \tag{4.66}$$

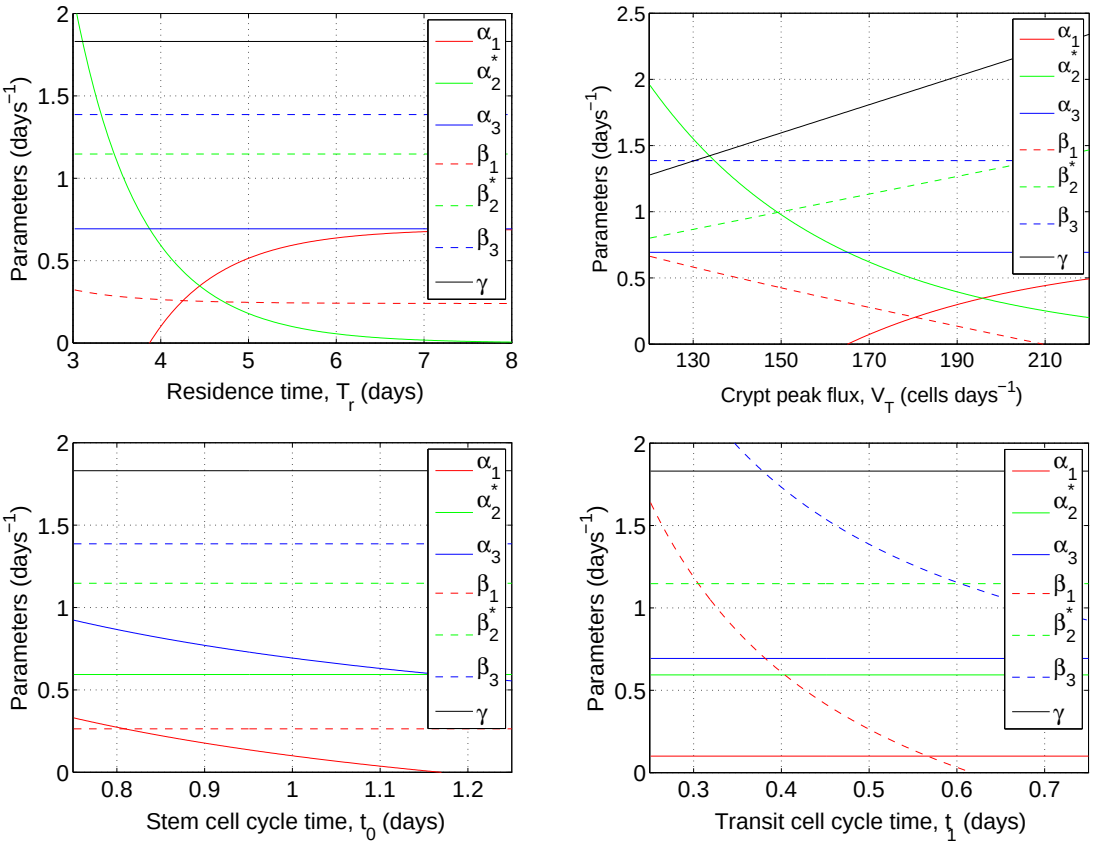

Figure 4.6: Plots of the rate parameters from (4.61)–(4.64) when the residence time T_r, the flux at the top of the crypt V_T and the cell cycle times t_0 and t_1 are varied. The parameter values, when not being varied, are taken to be $N_0^* = 6$, $N_1^* = 150$, $N_2^* = 94$, $t_0 = 1$ day, $t_1 = 0.5$ days, $V_T = 172$ cells/day and $T_r = 4$ days.

Given that $V_T \geq 0$, this gives an upper bound on β_1 to be

$$\beta_1 \leq \frac{\log 2}{N_1^*}\left(\frac{N_0^*}{t_0} + \frac{N_1^*}{t_1}\right). \tag{4.67}$$

Using (4.66), the restriction that β_1 must be non-negative creates an upper bound for the flux

$$V_T \leq \log 2\left(\frac{N_0^*}{t_0} + \frac{N_1^*}{t_1}\right). \tag{4.68}$$

Using the bound (4.65) on α_2^* and applying it to (4.62), where we note that T_r is monotonically decreasing in α_2^*, we find that

$$T_r \geq \frac{N_1^* t_0}{t_0 V_T - N_0^* \log 2}\log\left(\frac{t_0 V_T}{N_0^* \log 2}\right) + \frac{N_2^*}{V_T}. \tag{4.69}$$

Using (4.68), and noting that T_r is monotonically decreasing in V_T, the lower bound on the residence time is

$$T_r \geq \frac{t_1}{\log 2}\log\left(1 + \frac{N_1^* t_0}{N_0^* t_1}\right) + \frac{N_2^*}{\log 2\,(N_1^*/t_1 + N_0^*/t_0)}. \tag{4.70}$$

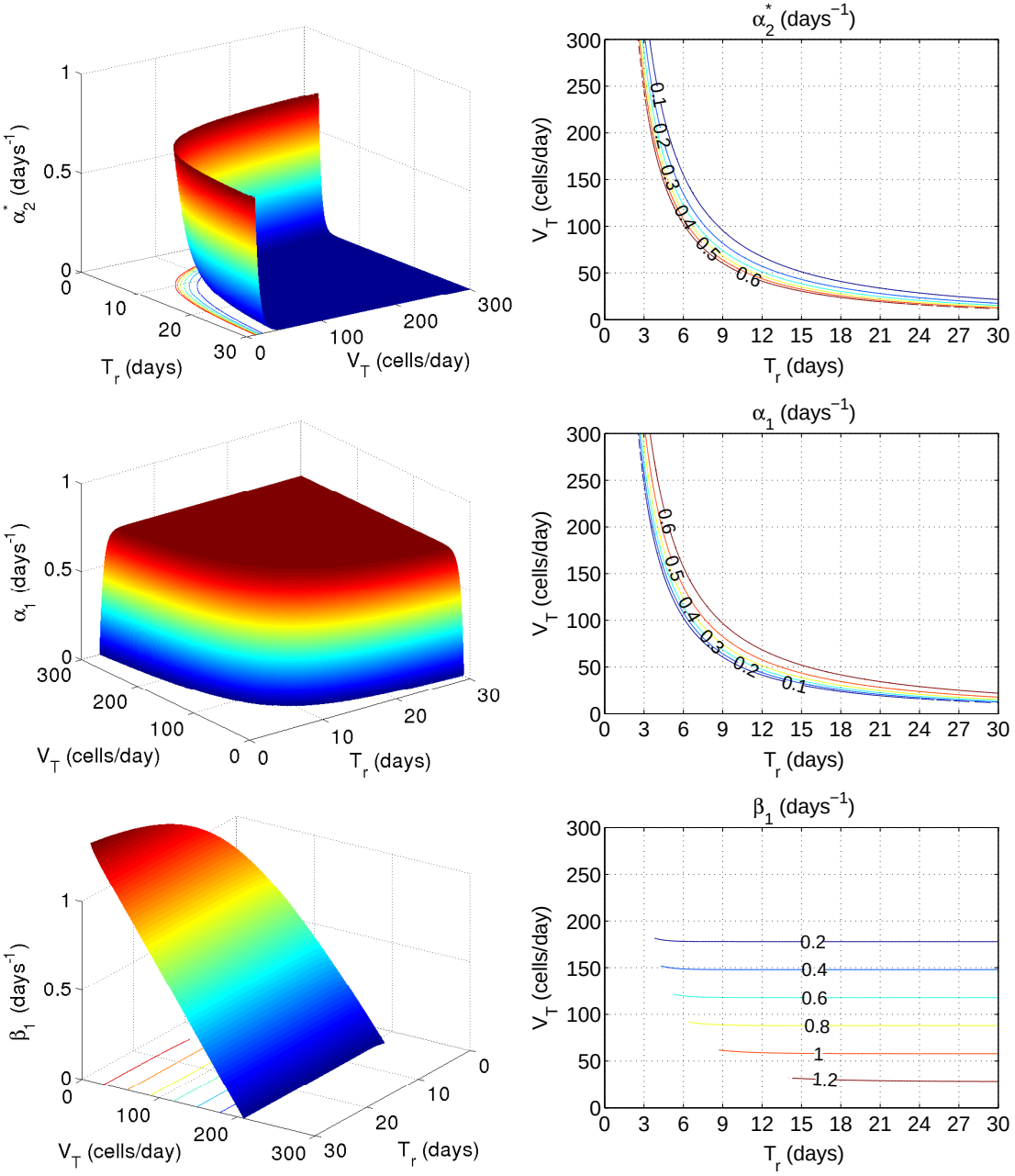

Figure 4.7: Surface (left) and contour (right) plots of the rate parameters α_2^* (top), α_1 (middle) and β_1 (bottom) against T_r and V_T from (4.62)–(4.64). The other parameters are fixed at $N_0^* = 6$, $N_1^* = 150$ and $N_2^* = 94$ cells, $t_0 = 1$ day and $t_1 = 1/2$ day. For these curves, only values of $\alpha_2^* < 0.693$ have been displayed because $\alpha_2^* > 0.693$ will result in α_1 being negative.

The combination of the restrictions (4.68) and (4.69) defines the region of the (T_r, V_T) parameter space for feasible solutions, which is approximately the intersection of the marked regions in the contour plots of α_1 and β_1 in Figure 4.7. The upper right section of the (T_r, V_T) plane in the contour plot, where $\alpha_2^* \to 0$, corresponds to the scenario where the stem cells decouple, with death and renewal balancing. In steady state, however, we feasibly expect death to be small compared with renewal and differentiation (W. F. Bodmer, pers. comm.), which should approximately balance. Therefore, we may further restrict our range of acceptable parameter values by assuming that $\alpha_1 < \alpha_2^*$, which corresponds to the strip of α_1 and α_2^* contours in Figure 4.7.

4.2.2.2 Feasible parameter values

In order to evaluate the constraints derived in the last section, we use data from Potten and Loeffler [1990] for the murine small intestine, where there about 250 cells in a crypt of which 4–16 are stem cells, and about two thirds (or 150–160) are proliferative; stem cells cycle every 12–32 hours with an average of 24; and transit cells cycle every 11–12 hours [Potten and Loeffler, 1987]. Therefore, as a first approximation, we choose the steady state parameter values to be $N_0^* = 6$, $N_1^* = 150$, $N_2^* = 94$, $t_0 = 1$ day and $t_1 = 1/2$ day. Using these parameter values, we find that

$$\alpha_3 = 0.693, \quad \beta_3 = 1.386, \tag{4.71}$$

while using (4.65), (4.67), (4.68) and (4.70), the other parameters must lie in the ranges

$$0 \le \alpha_1 \le 0.693, \quad 0 \le \alpha_2^* \le 0.693, \quad 0 \le \beta_1 \le 1.414, \quad 0 \le V_T \le 212.1, \quad T_r \ge 3.279. \tag{4.72}$$

Finally, using this bound on V_T, ranges for β_2^* and γ from (4.61) can be found to be

$$0 \le \beta_2^* \le 1.414, \quad 0 \le \gamma \le 2.256, \tag{4.73}$$

where the rates are measured in days^{-1}, the time is in days and the flux is in cells/day. One particular point in the feasible parameter space gives a parameter set of

$$\begin{aligned} N_0^* = 6 \text{ cells}, \quad N_1^* = 150 \text{ cells}, \quad N_2^* = 94 \text{ cells}, \quad t_0 = 1 \text{ day}, \quad t_1 = 1/2 \text{ day}, \\ \alpha_1 = 0.100 \text{ days}^{-1}, \quad \alpha_2^* = 0.593 \text{ days}^{-1}, \quad \alpha_3 = 0.693 \text{ days}^{-1}, \\ \beta_1 = 0.263 \text{ days}^{-1}, \quad \beta_2^* = 1.147 \text{ days}^{-1}, \quad \beta_3 = 1.386 \text{ days}^{-1}, \\ \gamma = 1.830 \text{ days}^{-1}, \quad V_T = 172 \text{ cells/day} \quad \text{and} \quad T_r = 4 \text{ days}. \end{aligned} \tag{4.74}$$

We use this set of parameters for subsequent calculations, and other sets of values give qualitatively similar results.

Given suitable data, we have so far been able to determine estimates for the rate parameters α_1, α_2^*, α_3, β_1, β_2^*, β_3 and γ, but the six parameters α_2, β_2, k_0, k_1, m_0 and m_1 are linked by the two relations $\alpha_2^* = \alpha_2 + k_0 N_0^*/(1 + m_0 N_0^*)$ and $\beta_2^* = \beta_2 + k_1 N_1^*/(1 + m_1 N_1^*)$, and therefore there are multiple values possible for the remaining parameters. For the purposes of calculations in the next section, we choose a set of values for the undetermined parameters. For the saturating feedback model we choose $\alpha_2 = 0.218$, $\beta_2 = 0.547$, $k_0 = 0.1$, $k_1 = 0.01$, $m_0 = 0.1$ and $m_1 = 0.01$, and for the linear feedback model we choose $\alpha_2 = 0.293$, $\beta_2 = 0.397$, $k_0 = 0.05$ and $k_1 = 0.005$.

An alternative method to the one presented in this section is to assume that T_r and V_T are fixed parameters and t_0 and t_1 vary. By performing a similar calculation we note that a sensitivity analysis shows that T_r and V_T are considerably more sensitive to changes in t_1 than t_0 (results not shown).

4.2.3 Sensitivity analysis of the parameters

Which parameters are the most influential on the system? This is an important question that can be addressed by using a sensitivity analysis. If it can be shown that some parameters have a strong influence while others have little or no effect, then experiments need only focus (at least initially) on the most influential processes. In this section, we discuss sensitivity coefficients, and how they can be used to ascertain key parameters.

By way of example, let us consider the effect that changes in a parameter p have on a cell population steady state N. Suppose we change p by a small amount Δp, and assume that this induces a small change in N of ΔN. The *sensitivity coefficient* $S_{N,p}$ of N with respect to p is defined to be the relative change in N divided by the relative change in p

$$S_{N,p} = \frac{\Delta N}{N} \Big/ \frac{\Delta p}{p} = \frac{p}{N}\frac{\Delta N}{\Delta p}. \tag{4.75}$$

In the limit as Δp tends to zero, this can be expressed in terms of the derivative

$$S_{N,p} = \frac{\partial(\ln N)}{\partial(\ln p)} = \frac{p}{N}\frac{\partial N}{\partial p}. \tag{4.76}$$

If $S_{N,p}$ is independent of p then we effectively find $S_{N,p}$ such that $N \propto p^{S_{N,p}}$, but in general $S_{N,p}$ will be a function of all the parameters in the system, so this relation is only true at the particular point in the parameter space at which $S_{N,p}$ has been calculated. If $|S_{N,p}| > 1$, a percentage change in p will produce a larger percentage change in N, while if $0 < |S_{N,p}| < 1$, then a percentage change in p produces a smaller percentage change in N. If $S_{N,p} > 0$ ($S_{N,p} < 0$), then increasing p will cause an increase (decrease) in N.

Applying formula (4.76) to the stem cell steady state with saturating feedback (4.27), and eliminating α and k_0, the sensitivity coefficient of N_0^* with respect to α is $S_{N_0^*,\alpha} = 1 + m_0 N_0^*$. Since both N_0^* and m_0 are positive, $S_{N_0^*,\alpha} > 1$ which means that an increase in α will produce a greater percentage change in N_0^*. Evaluating the sensitivity coefficient using parameter values $\alpha = 0.375$ and $k_0 = m_0 = 0.1$ produces $S_{N_0^*,\alpha} = 1.6$. Alternatively, if $\alpha = 0.8$, then $S_{N_0^*,\alpha} = 5$ and the dependence of N_0^* on α is much stronger. This illustrates how the dependence on α can vary at different points in the parameter space. Now let us consider the sensitivity coefficient of the stem cell steady state to the parameter m_0, which is given by $S_{N_0^*,m_0} = S_{N_0^*,\alpha} - 1 = m_0 N_0^* > 0$. Since the sensitivity coefficient $S_{N_0^*,m_0}$ will always be less than $S_{N_0^*,\alpha}$, we can say that the stem cell steady state is more sensitive to α than it is to m_0.

The sensitivity coefficients for the steady states N_0^*, N_1^* and N_2^* with respect to each of the parameters are summarised in Tables 4.2 and 4.3 for the linear and saturating feedback models, respectively. For the stem cell steady state, we vary α_1, α_2 and α_3 in combination since it is only their net size that affects the steady state. For the semi-differentiated cells α_2 appears separately and we consider the combination $\alpha_3 - \alpha_1$. For the fully-differentiated cells the same combinations will occur in the β parameters. We have chosen the particular parameter sets to give a sense of the order of magnitude for each sensitivity coefficient.

	$\alpha = \alpha_3 - \alpha_1 - \alpha_2$	k_0	$\beta = \beta_3 - \beta_1 - \beta_2$	k_1	γ
N_0^*	1	-1	0	0	0

	α_2	$\alpha_3 - \alpha_1$	k_0	$\beta = \beta_3 - \beta_1 - \beta_2$	k_1	γ
N_1^*	-0.030	0.091	-0.031	0.94	-0.97	0

	α_2	$\alpha_3 - \alpha_1$	k_0	β_2	$\beta_3 - \beta_1$	k_1	γ
N_2^*	-0.050	0.15	-0.051	-0.50	0.85	-0.95	-1

Table 4.2: Sensitivity coefficients for the cell population steady states (4.20)–(4.22) in the linear feedback model in terms of the different parameters. The parameters are taken to be $\alpha_1 = 0.100$, $\alpha_2 = 0.293$, $\alpha_3 = 0.693$, $\beta_1 = 0.263$, $\beta_2 = 0.397$, $\beta_3 = 1.386$, $\gamma = 1.830$, $k_0 = 0.05$ and $k_1 = 0.005$, which produces steady state values of $N_0^* = 6$, $N_1^* = 150$ and $N_2^* = 94$.

	$\alpha = \alpha_3 - \alpha_1 - \alpha_2$	k_0	m_0	$\beta = \beta_3 - \beta_1 - \beta_2$	k_1	m_1	γ
N_0^*	1.6	-1.6	0.60	0	0	0	0

	α_2	$\alpha_3 - \alpha_1$	k_0	m_0	$\beta = \beta_3 - \beta_1 - \beta_2$	k_1	m_1	γ
N_1^*	-0.084	0.32	-0.14	0.054	2.2	-2.3	1.4	0

	α_2	$\alpha_3 - \alpha_1$	k_0	m_0	β_2	$\beta_3 - \beta_1$	k_1	m_1	γ
N_2^*	-0.10	0.38	-0.17	0.065	-2.0	5.1	-2.2	1.3	-1

Table 4.3: Sensitivity coefficients for the cell population steady states (4.27)–(4.29) in the saturating feedback model in terms of the different parameters. The parameters are taken to be $\alpha_1 = 0.100$, $\alpha_2 = 0.218$, $\alpha_3 = 0.693$, $\beta_1 = 0.263$, $\beta_2 = 0.547$, $\beta_3 = 1.386$, $\gamma = 1.830$, $k_0 = 0.1$, $k_1 = 0.01$, $m_0 = 0.1$ and $m_1 = 0.01$, which produces steady state values of $N_0^* = 6$, $N_1^* = 150$ and $N_2^* = 94$.

These results show that changes in the parameters α or k_0 have the largest effect on the stem cell number, changes in β or k_1 have the greatest effect on the semi-differentiated cell population, and changes in β, γ, k_1 or m_1 cause the biggest changes to the fully-differentiated cells. Changes in m_0 produce a less significant change in the steady states. However, we must remember that m_0 is crucial in determining the stability of the stem cell steady state (4.27). In the next section we discuss how these parameters might be affected by mutations in the early stages of adenomatous growth.

4.3 Mutations in the continuous model

Mutations are known to occur in the division process leading to cell phenotypes with a selective growth advantage that can cause the cell to evade death, proliferate faster and/or out-compete other cell lines in the crypt [Sieber *et al.*, 2003]. Successive mutations provide incremental growth advantages that are followed by waves of clonal expansion, and it is believed that even if they provide no advantage mutations can spread by genetic drift [Rajagopalan *et al.*, 2003; Tomlinson *et al.*, 1996]. It is observed that mutations lead to a succession of clonal overgrowths at higher cell numbers, and then expansion plateaus until

the next mutation strikes and another growth spurt occurs [Calabresi *et al.*, 1985]. There are several long lag phases between periods of rapid growth, and mutations gradually accumulate over time until the combinations sufficient for visible clonal expansion are acquired and exponential growth ensues [Tsao *et al.*, 2000]. It is predicted that six rate-limiting steps are required for malignant tumour formation [Tomlinson *et al.*, 1996].

Canonical and non-canonical pathways of mutations have been suggested (for example, Fearon and Vogelstein [1990]), and while details of some of these mutations are known, the details of others are disputed. This makes modelling the accumulation of mutations difficult to predict [Giles *et al.*, 2003]. In Section 4.3.1 we discuss how mutations can be included in our model. Following this in Section 4.3.2 we present a simple competition model between mutant and healthy populations to derive conditions for the mutants to take over the crypt using the saturating feedback model from Section 4.1.4. Finally, in Section 4.3.3, we combine this information to use our model to predict how tumour progression might occur.

4.3.1 What effect could mutations have?

In this section we discuss the possible effects mutations might have, and how this can be incorporated into the model.

Firstly, a mutation might alter a cell's proliferation, differentiation or death rate. Although each of these is separate, the net result will be a change in the net per-capita growth rates α and/or β. Secondly, a mutation could affect the response to feedback, either by slowing down the response by reducing k_0 or k_1, or by turning on or off the saturation parameters m_0 and m_1. Thirdly, it might be the case that the mutation removes the feedback completely, which will result in exponential growth if $\alpha > 0$ or $\beta > 0$.

An alternative explanation is that the mutation switches off terminal differentiation of transit cells ($\beta_2 = 0$) resulting in an accumulation of transit cells at the expense of fully-differentiated cells, and that the mutated cells do not proliferate more (H. Clevers, pers. comm.). Fully-differentiated cells will disappear and the transit cell number will grow without bound, and this could mimic the exponential growth observed.

In this chapter we restrict ourselves to considering mutations that individually affect only one model parameter. In reality different mutations might affect more than one parameter, for instance p53 is known to be involved in cell cycle inhibition, apoptosis and genetic stability, in particular. However, modelling multiple interactions would require knowledge of which combinations are possible and their probabilities of occurring, but these are still scarcely understood [Spencer *et al.*, 2006]. It is expected that mutations will mainly occur in stem cells, but often the growth advantage might only be expressed in transit cells (W. F. Bodmer, pers. comm.), and this can be described by altering the parameters in this model. Additionally, mutations may affect neighbouring cells directly or indirectly, but relatively little is known about this [Tomlinson and Bodmer, 1997].

4.3.2 Analytic model for competing populations

If a mutation occurs, possibly changing the parameters related to the cell, will the mutant population coexist with the healthy cells or drive the healthy population to extinction?

In this section we use the continuous model with saturating feedback to consider two parallel compartment structures of healthy and mutant cells, and we discuss the parameter restrictions on the mutants taking over the crypt. We are aware that a stochastic model is

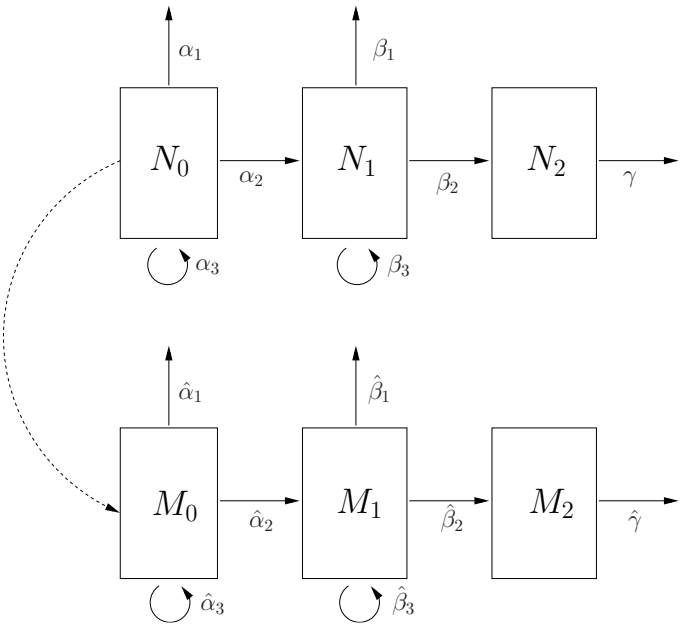

Figure 4.8: A compartment structure for parallel normal and mutant cell populations in the crypt, denoted by N_i and M_i, respectively. The parameters for the healthy cells are the same as those described in Section 4.1.4, and the parameters for mutant cells are denoted by the equivalent parameters with hats.

necessary to capture the variations that might occur, especially if the mutant cell dies out. However, since our main interest is in the competition between the normal and mutant cells, we use the ODE model because it is the most tractable, and it will give us an insight into the dynamics. The small number of mutations that become established and then die out in the ODE model can be thought of as altering the mutation rate. We consider a stochastic model of mutations in Chapter 5.

We denote the two parallel compartment structures of normal and mutant populations by N_i and M_i, respectively; the parameters for the normal healthy population by α_i, β_i for $i = 1, 2, 3$, k_j and m_j for $j = 0, 1$ and γ; and the parameters for the mutant population by $\hat{\alpha}_i$, $\hat{\beta}_i$ for $i = 1, 2, 3$, $\hat{k}_j$ and $\hat{m}_j$ for $j = 0, 1$ and $\hat{\gamma}$. The compartment structure is shown in Figure 4.8.

We use the saturating feedback model described in Section 4.1.4 to allow the possibility of both homeostasis and unregulated growth in different parameter regimes. The mutant cells are really just a subset of the normal cell populations that could take different parameter values if they were altered by the mutation. Therefore, we assume that the feedback in the differentiation rates depends on the sum of the healthy and mutant populations, $\alpha_2 =$

$\alpha_2(N_0+M_0)$ and $\beta_2=\beta_2(N_1+M_1)$. This creates a system of the form

$$\frac{\mathrm{d}N_0}{\mathrm{d}t}=\alpha N_0-\frac{k_0N_0(N_0+M_0)}{1+m_0(N_0+M_0)}, \tag{4.77}$$

$$\frac{\mathrm{d}M_0}{\mathrm{d}t}=\hat{\alpha}M_0-\frac{\hat{k}_0M_0(N_0+M_0)}{1+\hat{m}_0(N_0+M_0)}, \tag{4.78}$$

$$\frac{\mathrm{d}N_1}{\mathrm{d}t}=\beta N_1-\frac{k_1N_1(N_1+M_1)}{1+m_1(N_1+M_1)}+\alpha_2N_0+\frac{k_0N_0(N_0+M_0)}{1+m_0(N_0+M_0)}, \tag{4.79}$$

$$\frac{\mathrm{d}M_1}{\mathrm{d}t}=\hat{\beta}M_1-\frac{\hat{k}_1M_1(N_1+M_1)}{1+\hat{m}_1(N_1+M_1)}+\hat{\alpha}_2M_0+\frac{\hat{k}_0M_0(N_0+M_0)}{1+\hat{m}_0(N_0+M_0)}, \tag{4.80}$$

$$\frac{\mathrm{d}N_2}{\mathrm{d}t}=-\gamma N_2+\beta_2N_1+\frac{k_1N_1(N_1+M_1)}{1+m_1(N_1+M_1)}, \tag{4.81}$$

$$\frac{\mathrm{d}M_2}{\mathrm{d}t}=-\hat{\gamma}M_2+\hat{\beta}_2M_1+\frac{\hat{k}_1M_1(N_1+M_1)}{1+\hat{m}_1(N_1+M_1)}, \tag{4.82}$$

where $\alpha=\alpha_3-\alpha_1-\alpha_2$, $\hat{\alpha}=\hat{\alpha}_3-\hat{\alpha}_1-\hat{\alpha}_2$, $\beta=\beta_3-\beta_1-\beta_2$ and $\hat{\beta}=\hat{\beta}_3-\hat{\beta}_1-\hat{\beta}_2$. The purpose of this model is to determine the parameter restrictions on the mutant populations taking over the crypt, and as such we are only interested in what happens after a mutant is created. The initial conditions will therefore be $(N_0,N_1,N_2)=(N_0^*,N_1^*,N_2^*)$, and then $(M_0,M_1,M_2)=(1,0,0)$ or $(M_0,M_1,M_2)=(0,1,0)$ depending on if the mutant created is a stem or transit cell.

(4.77)–(4.78) decouple to produce a closed system for N_0 and M_0, which is a standard phase place exhibiting classic competition behaviour (see, for example, predator-prey models such as the Lotka-Volterra system in Murray [2002]). We use a steady state analysis to determine parameter restrictions on the effects of each mutation. The healthy and mutant stem cell steady states (denoted by N_0^* and M_0^*) are given by

$$N_0^*\left[p-(N_0^*+M_0^*)\right]=0=M_0^*\left[\hat{p}-(N_0^*+M_0^*)\right], \tag{4.83}$$

where

$$p=\frac{\alpha}{k_0-\alpha m_0}\qquad\text{and}\qquad\hat{p}=\frac{\hat{\alpha}}{\hat{k}_0-\hat{\alpha}\hat{m}_0}, \tag{4.84}$$

which produces three steady states $(N_0^*,M_0^*)=(0,0)$, $(p,0)$ and $(0,\hat{p})$, and also a line of steady states given by $N_0^*+M_0^*=p$ if $p=\hat{p}$.

The linear stability of the steady states depends on the key parameters p and $\hat{p}$, which represent the stem cell steady state sizes for the respective populations. If a mutation occurs, then the mutant population will wipe out the healthy stem cells if the mutation raises the key parameter $\hat{p}$ such that $\hat{p}>p$, and the mutants will die out if the mutation results in $\hat{p}<p$. The only possibility for the healthy and mutant stem cell populations to coexist is if the mutation leaves $\hat{p}=p$ unchanged (a neutral mutation), and in that case the relative population sizes will depend on the initial conditions. The zero steady state for both populations will only occur if both stem cell population growth rates are negative. These regions of stability are summarised in Figure 4.9.

For the semi-differentiated cells, the steady states depend on the healthy and mutant

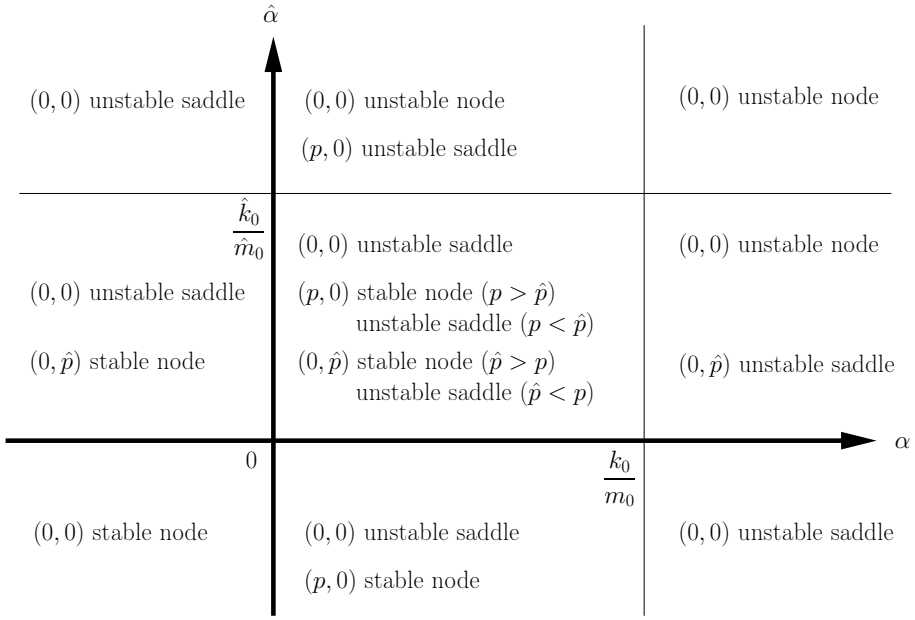

Figure 4.9: Regions of stability for the normal and mutant stem cell steady states in the $(\alpha, \hat{\alpha})$ phase-space from the system (4.77)–(4.78). Additionally, there is a line of steady states $N_0^* + M_0^* = p$ if $p = \hat{p}$ which is not shown.

stem cell differentiation rates given by D and $\hat{D}$, respectively, where

$$D(N_0^*, M_0^*) = \alpha_2 N_0^* + \frac{k_0 N_0^*(N_0^* + M_0^*)}{1 + m_0(N_0^* + M_0^*)}, \tag{4.85}$$

$$\hat{D}(N_0^*, M_0^*) = \hat{\alpha}_2 M_0^* + \frac{\hat{k}_0 M_0^*(N_0^* + M_0^*)}{1 + \hat{m}_0(N_0^* + M_0^*)}. \tag{4.86}$$

The forms of the steady states will vary dependent on the stem cell steady states (N_0^*, M_0^*).

If $(N_0^*, M_0^*) = (0,0)$, then $D = 0 = \hat{D}$ and so the semi-differentiated cell population steady state equations reduce to exactly the same form as the stem cell equations, and so there are steady states $(N_1^*, M_1^*) = (0,0)$, $(q,0)$ and $(0,\hat{q})$, and the line of solutions $N_1^* + M_1^* = q$, where

$$q = \frac{\beta}{k_1 - \beta m_1} \qquad \text{and} \qquad \hat{q} = \frac{\hat{\beta}}{\hat{k}_1 - \hat{\beta}\hat{m}_1}, \tag{4.87}$$

with exactly the same form of linear stability. In general we would expect the transit cell population to die out if the stem cell population was zero, so we impose $\beta, \hat{\beta} < 0$ which means that $q, \hat{q} < 0$ and only $(N_1^*, M_1^*) = (0,0)$ exists.

If $(N_0^*, M_0^*) = (p,0)$, then $\hat{D} = 0$ and $D \neq 0$, and so we either get $(N_1^*, 0)$, where

$$N_1^* = \frac{1}{2(k_1 - \beta m_1)}\left[\beta + m_1 D + \sqrt{(\beta - m_1 D)^2 + 4Dk_1}\right], \tag{4.88}$$

as in the un-mutated form from (4.28), which is stable when $0 < \beta < k_1/m_1$, or alternatively

$$(N_1^*, M_1^*) = \left(\frac{Dq(1 + m_1\hat{q})}{\beta(\hat{q} - q)}, \hat{q} - \frac{Dq(1 + m_1\hat{q})}{\beta(\hat{q} - q)}\right), \tag{4.89}$$

which is unfeasible, since $\beta, \hat{\beta}, q, \hat{q} < 0$ means that either $N_1^* < 0$ or $M_1^* < 0$ (or both). By symmetry the same occurs for $(N_0^*, M_0^*) = (0, \hat{p})$, only with the parameters for N_1 and M_1 reversed.

For the purposes of this study, we restrict ourselves to considering only advantageous mutations. The line of stem cell steady states, $N_0^* + M_0^* = p$, can exist only when the mutation in the stem cells is neutral conferring no advantage to the mutated population, and so we therefore do not explore these steady states further.

Finally, for the fully-differentiated cells, there is a stable node in N_2 and M_2 for every steady state in N_1 and M_1, given by

$$N_2^* = \frac{1}{\gamma}\left[\beta_2 N_1^* + \frac{k_1 N_1^*(N_1^* + M_1^*)}{1 + m_1(N_1^* + M_1^*)}\right], \tag{4.90}$$

$$M_2^* = \frac{1}{\hat{\gamma}}\left[\hat{\beta}_2 M_1^* + \frac{\hat{k}_1 M_1^*(N_1^* + M_1^*)}{1 + \hat{m}_1(N_1^* + M_1^*)}\right]. \tag{4.91}$$

In summary, we have found that either the healthy or mutant cell populations will die out unless the mutation is neutral (keeping the key parameter $p = \hat{p}$ fixed) and the mutant has no growth advantage. Therefore an advantageous mutation of any kind will lead to fixation of the mutant, according to the continuous model with saturating feedback. Since we are interested in explaining tumour development, which requires a certain number of rate-limiting mutations, we consider only mutations that confer some form of growth advantage. Consequently, we will henceforth assume that all mutations reach fixation in the crypt. It should be noted, however, that these conclusions are likely to be dependent on our choice of kinetics, and other behaviour might be possible with a different system.

4.3.3 Explaining both homeostasis and tumourigenesis

In this section we use the feedback models presented in Section 4.1 to explain colorectal cancer progression. The aim is to explain the observed step-wise tumour growth, separated by long lag phases, before unregulated tumour expansion ensues.

Both the linear and saturating feedbacks capture homeostatic regulation of stem cells and produce models that are structurally stable. However, in a system with the linear form of feedback, unbounded growth can only be achieved by removing the feedback completely (by setting the rates of feedback response k_i to zero), but with the saturating form unbounded growth can also be achieved by overwhelming the feedback. Therefore, only the saturating form can describe both homeostasis and unbounded growth of cell numbers without removing the feedback.

In light of the results from the last section, we propose that an advantageous mutation (corresponding to a genetic hit) will change the parameters for one cell and these changes will be conferred to all cells in the crypt in a short time period. This change in the model parameters will change the value of the steady state cell populations and cause an increase in compartment size which could be interpreted as the first stage in the process transforming a normal crypt via an adenomatous polyp to a carcinoma. Thus we see the possibility of successive genetic hits increasing either α or β (either by increasing the proliferation rate, or decreasing the differentiation or death rates) moving the crypt through increasing steady state cell populations, until finally α or β exceed a critical value in the saturating feedback model and unbounded growth occurs. This process will mimic the step-wise tumour growth and long lag phases observed between growth spurts before the tumour grows exponentially.

The same scenario would occur with the linear feedback model, but unbounded cancer growth can only be achieved if a mutation removes the feedback control, and no amount of parameter changes could lead to exponential growth.

An alternative explanation for the rapid expansion phase of tumour development might not involve unbounded growth. If a parameter change resulted in a large increase in steady state size, then the growth towards this steady state might mimic unbounded growth. As we showed in Section 4.2.3, the cell population steady states are very sensitive to the parameters k_0 and k_1. Therefore changes in α or β might increase the steady states early in tumour growth, but the mutation that initiates the fast cancer progression (mimicking unbounded growth) might affect k_0 or k_1.

In Sections 4.1.3 and 4.1.4 we have assumed that both the stem cells and the semi-differentiated cells are subject to the same form of feedback. However, it is possible to 'mix and match' different feedbacks for different cell populations, while still maintaining some feedback saturation. There are two possibilities for how the saturating feedback might occur. The stem and transit cell populations might both be subject to saturating feedback, and consequently both populations would be able to initiate tumourigenesis. Alternatively, the stem cells may be subject to linear feedback which maintains a permanent equilibrium, with the transit cells subject to saturating feedback and possible tumourigenesis. This is postulated to be the more likely scenario in the crypt, where it is not expected that the stem and transit cells will be subject to the same type of feedback (W. F. Bodmer, pers. comm.). In the next two sections we discuss both possibilities and present examples of each.

4.3.3.1 Unbounded growth via stem or transit cells

Firstly, we consider the case where both the stem and transit cell populations are subject to saturating feedback, with $m_0, m_1 > 0$. The stability of this system in the $(\alpha, \beta)-$parameter space is summarised in the right panel of Figure 4.10. If $\alpha < 0$ and $\beta < k_1/m_1$ (Region I), the stem cells cannot sustain their number and consequently the crypt dies out. If $0 < \alpha < k_0/m_0$ and $\beta < k_1/m_1$ (Region II), the crypt will be in a normal equilibrium. If either $\alpha > k_0/m_0$ or $\beta > k_1/m_1$ (or both), there is unregulated cell population growth in the system. For $\alpha > k_0/m_0$ and $\beta < k_1/m_1$ (Region III), the cancer-driving cells promoting tumourigenesis are the stem cells, whereas for $\alpha < k_0/m_0$ and $\beta > k_1/m_1$ (Region IV), the cancer-driving cells are derived from transit cells and the stem cell population either remains fixed ($\alpha > 0$) or dies out ($\alpha < 0$). Finally, if both $\alpha > k_0/m_0$ and $\beta > k_1/m_1$ (Region V), unbounded growth is through cancer-driving cells derived from both stem cells and transit cells.

We demonstrate the multi-stage mutational process described in the previous section with the following example. For our purposes we assume that once a mutation occurs its selective advantage is instantly conferred to all the other cells, which we justify by using the results from Section 4.3.2. Consider the initial parameter set $\alpha_1 = 0.100$, $\alpha_2 = 0.218$, $\alpha_3 = 0.693$, $\beta_1 = 0.263$, $\beta_2 = 0.547$, $\beta_3 = 1.386$, $\gamma = 1.830$, $k_0 = 0.1$, $k_1 = 0.01$, $m_0 = 0.1$ and $m_1 = 0.01$. This produces $\alpha = 0.375$, $\beta = 0.576$ and critical threshold values $\alpha = k_0/m_0 = 1$ and $\beta = k_1/m_1 = 1$; the population is therefore stable, with $N_0^* = 6$, $N_1^* = 150$, $N_2^* = 94$ in Region II of the parameter space in Figure 4.10. Suppose a mutation (in either β_1 or β_3) raises β to 0.7; then N_0^* is unchanged, but $N_1^* = 250$, $N_2^* = 172$. Now suppose a second mutation makes $\alpha = 0.8$, which produces a new steady state of $N_0^* = 40$, $N_1^* = 403$, $N_2^* = 297$. A third mutation making $\beta = 0.8$ produces a steady state of $N_0^* = 40$, $N_1^* = 636$, $N_2^* = 490$. Finally, a fourth mutation causes $\beta = 1.1$, which is

larger than the critical threshold and lies in Region IV of the parameter space in Figure 4.10. Consequently there can no longer be a steady state, resulting in unregulated growth in the semi- and fully-differentiated cell populations. The effect of this series of mutations is illustrated in Figure 4.10.

4.3.3.2 Unbounded growth via transit cells only

Secondly, we consider the case where the stem cells are now subject to linear feedback and the transit cells are again regulated by saturating feedback. The stability of the steady states is again dependent on the two net per-capita growth rates α and β, which is summarised in Figure 4.11. If $\alpha < 0$ and $\beta < k_1/m_1$ (Region I), the stem cells cannot sustain their number and consequently the crypt dies out. If $\alpha > 0$ and $\beta < k_1/m_1$ (Region II), the crypt will be in a healthy equilibrium. If $\beta > k_1/m_1$ (Region III), the stem cells will either remain fixed ($\alpha > 0$) or die out ($\alpha < 0$), and unbounded growth is initiated by a cancer-driving cell derived from the transit cell population. In this model, unbounded growth can only be driven by transit cells and not stem cells.

For a numerical example of this sequence of mutations, we take the initial parameters to be $\alpha_1 = 0.100$, $\alpha_2 = 0.293$, $\alpha_3 = 0.693$, $\beta_1 = 0.263$, $\beta_2 = 0.547$, $\beta_3 = 1.386$, $\gamma = 1.830$, $k_0 = 0.05$, $k_1 = 0.01$ and $m_1 = 0.01$, which gives $\alpha = 0.3$, $\beta = 0.576$ and produces the critical saturation threshold $\beta = k_1/m_1 = 1$. Therefore the population is stable with $N_0^* = 6$, $N_1^* = 150$ and $N_2^* = 94$ in Region II of the parameter space in Figure 4.11. If a mutation occurs (in either β_1 or β_3) that increases β to 0.9, then the stem cells remain at $N_0^* = 6$ and the other population steady states increase to $N_1^* = 939$ and $N_2^* = 745$. Now suppose a second mutation raises α to 1.2, which further increases the steady states to $N_0^* = 24$, $N_1^* = 1286$ and $N_2^* = 1037$. Finally, a third mutation raises β to 1.4 which is greater than the critical threshold and lies in Region III of Figure 4.11. Consequently there can no longer be a steady state, resulting in unregulated growth in the cell populations.

4.3.4 Cancer development via multiple steady states

An alternative explanation for step-wise tumour growth with plateau phases was presented by d'Onofrio and Tomlinson [2007]. The authors, using their discrete model, postulated that if the transit cell renewal probability, $b_3(N_1)$, was an increasing function of N_1 with $m \geq 2$ inflexion points, then there may be $n \leq m$ stable steady states alternating in stability. In this case it might be possible to explain the transition between different steady states by parameter fluctuations alone.

In order to adapt this to our continuous model approach, we extend the saturating feedback model to include multiple inflection points in the transit cell growth rate by using a sum of Hill functions, with the aim of explaining all the plateau phases of tumour growth as different predetermined steady states in the model. Supposing there are n_m different mutations required for carcinogenesis to occur, the differentiation rate might take the form

$$\beta_2(N_1) = \beta_2 + \sum_{i=1}^{n_m} \frac{k_i N_1^{e_i}}{1 + (m_i N_1)^{e_i}}, \tag{4.92}$$

where k_i and m_i are constants and exponents e_i are integers. If $n_m = 1$, then this reduces to the saturating feedback described in Section 4.1.4. If we assume linear feedback in the

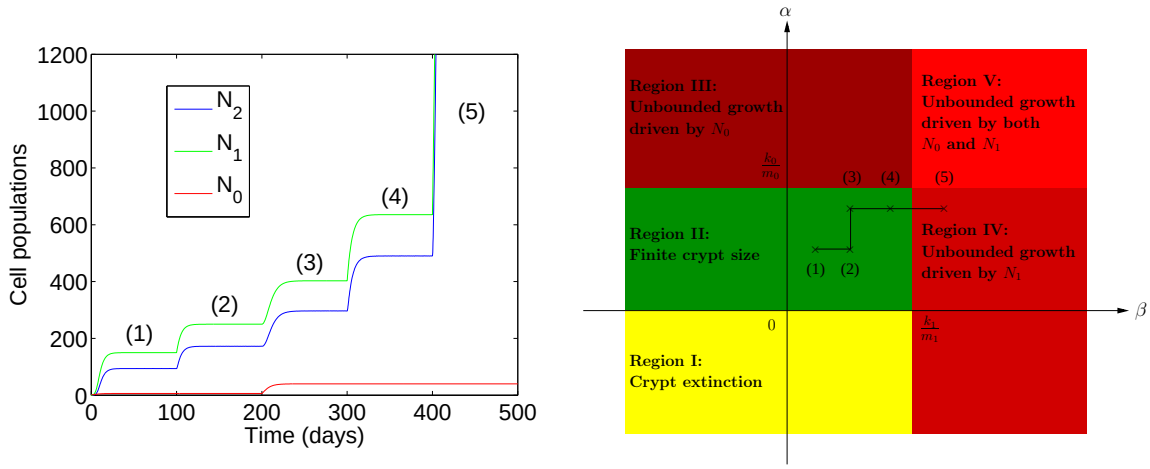

Figure 4.10: A mutational sequence in a system with saturating feedback in both stem and transit cells from (4.23)–(4.25). Left: An illustrative sequence of mutations (occurring every 100 days), where the initial parameters are taken to be $\alpha_1 = 0.100$, $\alpha_2 = 0.218$, $\alpha_3 = 0.693$, $\beta_1 = 0.263$, $\beta_2 = 0.547$, $\beta_3 = 1.386$, $\gamma = 1.830$, $k_0 = 0.1$, $k_1 = 0.01$, $m_0 = 0.1$ and $m_1 = 0.01$, which produces $\alpha = 0.375$ and $\beta = 0.576$. The mutations cause, successively, $\beta = 0.7$, $\alpha = 0.8$, $\beta = 0.8$ and $\beta = 1.1$. After the last mutation $\beta > k_1/m_1$, which means there is no steady state and the cell populations grow without bound. The numbers (1)–(5) correspond to the different points in the $(\alpha, \beta)-$parameter space. Right: Regions of stability in the $(\alpha, \beta)-$parameter space for the same system.

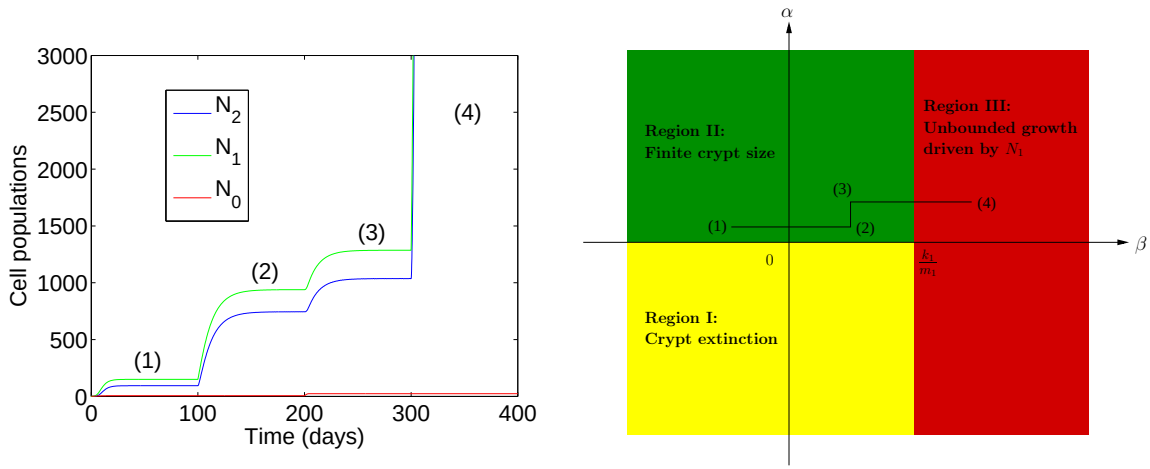

Figure 4.11: A mutational sequence in a system with linear feedback in stem cells and saturating feedback in transit cells from (4.23)–(4.25) with $m_0 = 0$. Left: An illustrative sequence of mutations (occurring every 100 days), where the initial parameters are taken to be $\alpha_1 = 0.100$, $\alpha_2 = 0.293$, $\alpha_3 = 0.693$, $\beta_1 = 0.263$, $\beta_2 = 0.547$, $\beta_3 = 1.386$, $\gamma = 1.830$, $k_0 = 0.05$, $k_1 = 0.01$ and $m_1 = 0.01$, which produces $\alpha = 0.3$ and $\beta = 0.576$. The mutations cause, successively, $\beta = 0.9$, $\alpha = 1.2$ and $\beta = 1.4$. After the last mutation $\beta > k_1/m_1$, which means there is no steady state and the cell populations grow without bound. The numbers (1)–(4) correspond to the different points in the $(\alpha, \beta)-$parameter space. Right: Regions of stability in the $(\alpha, \beta)-$parameter space for the same system.

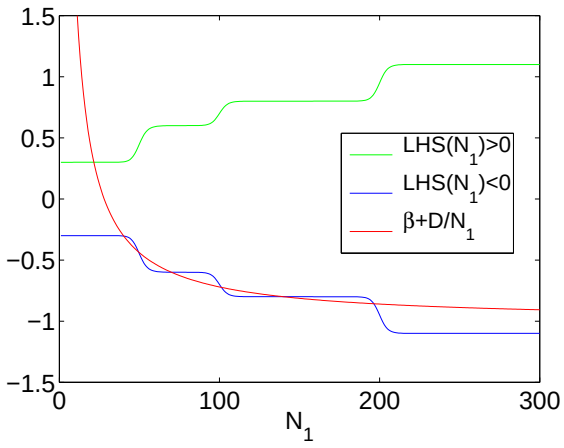

Figure 4.12: Plots of the left- and right-hand sides of equation (4.94) in the two cases where the parameters k_i are positive or negative. We use a sum of Hill functions ($n_m = 3$) for the left-hand side $LHS(N_1) = \pm[0.3 + 0.3N_1^{20}/(50^{20} + N_1^{20}) + 0.2N_1^{40}/(100^{40} + N_1^{40}) + 0.3N_1^{80}/(200^{80} + N_1^{80})]$, depending on the signs of k_i, $\beta = -1$ and $D = 28$.

stem cells, then the steady state equation for the semi-differentiated cells is given by

$$0 = \beta N_1^* - \sum_{i=1}^{n_m} \frac{k_i (N_1^*)^{e_i+1}}{1 + (m_i N_1^*)^{e_i}} + N_0^*(\alpha_2 + k_0 N_0^*). \tag{4.93}$$

We can use the stem cell differentiation rate $D = N_0^*(\alpha_2 + k_0 N_0^*)$ as a bifurcation parameter, as d'Onofrio and Tomlinson [2007] did, and rearrange to get

$$\sum_{i=1}^{n_m} \frac{k_i (N_1^*)^{e_i}}{1 + (m_i N_1^*)^{e_i}} = \beta + \frac{D}{N_1^*}. \tag{4.94}$$

Denote the left-hand side by $LHS(N_1)$. In the saturating feedback model we imposed $k_i > 0$ for the differentiation rate to be an increasing function of population size, but d'Onofrio and Tomlinson [2007] chose the opposite with $k_i < 0$. We illustrate these two choices in Figure 4.12, where we plot the left- and right-hand sides of (4.94) when the left-hand side is, respectively, positive and negative corresponding to k_i being positive and negative, and we choose $n_m = 3$ to produce three stable steady states. We see that with our assumptions of the differentiation rate being an increasing function of N_1 there is only one steady state, whereas if the differentiation rate is a decreasing function of N_1 then there can be multiple steady states (six in this case), depending on the parameters. This feedback model permits instability through altering initial conditions in addition to parameter changes, and random parameter fluctuations can lead to the crypt moving between different steady states. Therefore, it is possible to explain the transition between multiple stable steady states by parameter fluctuations alone if we assume that the differentiation rate is a decreasing function of cell population size.

It is important to determine, however, whether this is a plausible assumption, and whether this functional form is biologically realistic. If a small fluctuation in the parameters can be induced, then experiments can be designed to test between these two different hypotheses for plateau phase tumour growth. If the fluctuation resulted in the steady state

cell populations increasing or decreasing, then d'Onofrio and Tomlinson's multiple steady state hypothesis would be verified, whereas if the cell populations returned to their original size then our hypothesis of steady state increases due to parameter changes would be supported (assuming that the fluctuation did not result in unbounded growth in the saturating feedback model). At the moment experimental techniques are not able to distinguish between these two propositions in the colon, but in other tissues it is thought that feedback is the key mechanism.

4.4 Discussion

In this chapter we have discussed the need to extend the models described in Chapter 3 to maintain stability in the face of infinitesimal parameter changes. One way to capture the homeostatic regulation of the cell populations is to introduce feedback into the system by making the parameter values dependent on the population sizes.

We introduced two different types of feedback to make the continuous model structurally stable. The linear feedback model controls the population growth so that there is always a stable equilibrium and exponential growth in cell numbers cannot occur unless mutations are able to knock out the feedback. The saturating feedback model allows stable equilibria for a certain range of parameters, but also permits uncontrolled growth in the cell populations if the growth rates are pushed above a critical threshold. We also showed that instability could occur through delay in the feedback. The algebraic complexity is such that we could not solve the most general cases of feedback, and so we have only examined these two specific forms, although other forms are possible, such as those suggested by d'Onofrio and Tomlinson [2007]. Different forms could produce different conclusions, although we showed that the results do not change qualitatively if feedback occurs in death, renewal or differentiation.

We analysed the parameter space for the saturating feedback model and used this to predict bounds on the residence time T_r and flux of cells at the top of the crypt V_T to produce feasible parameter values. We can predict the parameter values α_1, α_2^*, α_3, β_1, β_2^*, β_3 and γ if we are given known data N_0^*, N_1^*, N_2^*, t_0, t_1, T_r and V_T. In fact, if we are given any 7 of the 14 unknown parameters, then we can find the other 7 parameters, as long as these do not include all the unknowns in any of the equations (4.61)–(4.64). However, very few of the parameters have experimentally determined values, so we must obtain more known parameter bounds before we can use our model to make predictions about precise values for the other parameters. A sensitivity analysis demonstrated that the key parameters in the model are the net per-capita growth rates of the stem and transit cell populations, which represent renewal minus death and differentiation, the rate of response of stem and transit cells to changes in their number and the rate at which fully-differentiated cells are removed/die. Throughout this chapter we have taken one particular set of parameter values for illustrative purposes, but the results are typical and the same qualitative behaviour is observed using different parameter sets in Johnston *et al.* [2007a,b].

Our results suggest that an advantageous mutation will always lead to that mutant reaching fixation in the crypt, while neutral mutations will result in mutants coexisting with healthy cells. The saturating feedback model could help explain the observed lag phases after advantageous mutations occur, and thus the existence of benign tumours or adenomas before carcinogenesis takes over. Early mutations in the adenoma-carcinoma sequence could raise the net per-capita growth rates, but keep them below their critical

values, which will create new, higher steady states. Later stage mutations could cause the net per-capita growth rate of stem or transit cells to exceed a critical value, ensuring that unregulated cell population growth occurs. However, if no genetic changes occur that take a tumour out of the stable parameter range corresponding to finite crypt size, then it remains benign. Though the evidence supports the view that essentially all colorectal cancers go through an adenoma, or benign phase, by no means do all adenomas develop into carcinomas.

These results show that failure of programmed cell death or differentiation could lead to tumour growth, as concluded from the model by Tomlinson and Bodmer [1995], as well as an increased proliferation rate. However, mutations in any of the key parameters (death, differentiation or renewal rates for the stem or transit cells) can initiate tumourigenesis, and eventually exponential growth in cell numbers leading to a carcinoma, but it is only changes in the net per-capita growth rates that are important, and then only with a suitable feedback model. We have shown how by incorporating the more stringent linear feedback into the stem, but not the transit, cell compartment, it is then only the transit cells that can become the cancer-driving cells. This may well be the best reflection of the actual situation in the large colorectal crypt [Johnston *et al.*, 2007a].

Another key observation is that the ratio of the semi- to fully-differentiated cells in the exponential growth phase using the saturating feedback model is a constant, given by $N_1/N_2 = (\beta - k_1/m_1 + \gamma)/(\beta_2 + k_1/m_1)$ as $t \to \infty$, assuming that the stem cell population is fixed. It is believed that the proportion of cancer-driving cells, thought to be derived from transit cells, in a tumour is small but this formula shows that the ratio of driving to differentiated cells in a cancer can take any value depending on β_2, β, k_1, m_1 and γ. Therefore the proportion of driving cells might be larger than expected in adenomas and carcinomas, and this might determine how aggressive the tumour is (W. F. Bodmer, pers. comm.).

Throughout this chapter we have made the simplifying assumption that the selective advantage gained by a mutation will be instantly conferred to the whole population, but a more detailed model might track the progress of healthy and mutant cells separately. This is considered in more detail in the next chapter, where we develop a stochastic simulation model of individual cells in the crypt in order to model the accumulation of mutations.

Chapter 5

Mutation accumulation in the crypt

In the previous chapter we used the continuous model presented in Chapter 3 to discuss how the cell populations in the crypt might accumulate multiple mutations during the onset of tumourigenesis. In reality, stochastic models are required to describe the mutational process which cannot be adequately captured by deterministic models alone, and so here we present a stochastic simulation model for that purpose.

In this chapter we discuss the question of how crypt cells achieve the two mutational hits that are required to inactivate the *APC* gene, and consider the relative likelihood of different scenarios. It had previously been assumed that mutations would always occur in stem cells since they are the only long-term residents of the crypt and, in particular, that any first mutations in transit cells are inconsequential because they would be swept out of the crypt within a few cell divisions. However, despite the transit cells being short-lived, they vastly outnumber their stem cell ancestors and cycle at twice the frequency. Hence, intuitively, one might expect the first hit to be in a stem cell but then the second to occur in a transit cell descendant.

Komarova and Wang [2004] proposed a model that investigates if mutations in transit cells need to be considered or not, and determines which is the most likely pathway to two hits for different parameter regimes. The three pathways considered are SS (both hits occur in a stem cell), SD (the first hit occurs in a stem cell and the second occurs in a differentiating transit cell) and DD (both hits occur in transit cells).

Firstly, we introduce the model by Komarova and Wang [2004] in Section 5.1, and discuss their assumptions and the validity of their approximations. In Section 5.2 we present an individual cell-based model which we use to verify the results from Komarova and Wang, and to test their assumptions for different scenarios. We derive analytical formulae to support the simulation results in Section 5.3, and we summarise our findings in Section 5.4. Throughout this chapter, we only consider the occurrence of the first two mutational hits, and we assume that achieving these mutations allows the cell to evade apoptosis, stick to the basal lamina, and remain in the crypt.

5.1 The Komarova and Wang '2 hit' model

Komarova and Wang [2004] considered a simplified version of a crypt with only stem and transit cells, neglecting fully-differentiated cells because they do not divide and are not

Parameter	Description
L	Number of cell divisions (1 asymmetric, $L-1$ symmetric) ($L \sim 10$)
N	Number of cells in each clone including the stem cell ($N = 2^L \sim 10^3$)
p_1	Probability per division of the 1st hit ($p_1 \sim 10^{-7}$)
p_2	Probability per division of the 2nd hit ($p_2 \sim 10^{-7}$, or 10^{-2} with CIN)
x_0	Probability that the whole crypt is wild type (state X_0)
x_1	Probability that the whole crypt has the 1st mutation (state X_1)
x_2	Probability that at least 1 cell has 2 mutations (state X_2)
P_{SS}	Probability of getting a double mutant from path SS
P_{SD}	Probability of getting a double mutant from path SD
P_{DD}	Probability of getting a double mutant from path DD
R	Rate at which 2 hits occur in transit cells before the 1st hit is lost

Table 5.1: Parameters and variables in the model by Komarova and Wang [2004].

relevant to the calculations. Consequently, when the authors quote a crypt size this really corresponds to the size of the proliferating compartment. Only one stem cell and its clone are considered, by which the authors mean all transit cells that have derived from that ancestral stem cell. A full crypt is assumed to be the sum of several independent clones. The key model assumptions are:

- There are 2000 proliferating cells in the crypt, and between 1 and 32 of them are stem cells;
- Stem and transit cells are assumed to have the same cycling time, and all cell divisions are synchronous;
- Stem cells only divide asymmetrically and transit cells always divide symmetrically;
- Transit cells divide $L-1$ times before they differentiate, and so each clone contains $N = 2^L$ cells, including the stem cell;
- There is assumed to be no cell death, other than cells being shed from the crypt;
- Mutations are assumed to be equally likely to occur in either stem or transit cells;
- The first hit is assumed to be neutral (conferring no selective advantage which results in no parameter changes), and the second hit will either create a cancer-driving cell or allow the cell to evade apoptosis and remain in the crypt.

The probabilities of mutation per division are p_1 and p_2 for the first and second hits, respectively, which allows for the model to include chromosomal instability (CIN) which would raise the probability of the second hit by several orders of magnitude. x_0, x_1 and x_2 (X_0, X_1 and X_2) denote, respectively, the probability (state) that the crypt is wild type, all first mutants, or at least one cell is a double mutant. The probability of the two hits occurring by routes SS, SD and DD are denoted by P_{SS}, P_{SD} and P_{DD}, respectively.

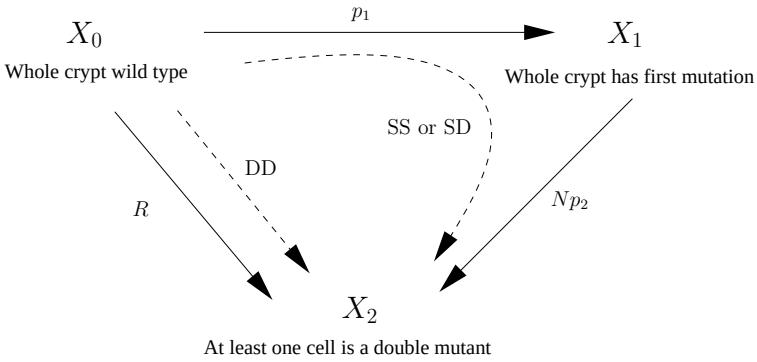

Figure 5.1: Schematic illustrating the mutational process used in the model by Komarova and Wang [2004], converting a wild-type crypt (state X_0), via the whole crypt containing a first mutant (state X_1), to the state where at least one cell has a second hit (state X_2). A double mutant may also be achieved without the first hit being conferred to all cells, bypassing state X_1. The ODEs relating to these processes are given in (5.1)–(5.3).

Finally, R denotes the rate at which double mutants are created by both hits occurring in the transit cell compartment before the first hit is lost. The model parameters and variables are summarised in Table 5.1.

The frequency at which new mutants are created is given by Np_1, and only $1/N$ of those mutations will occur in a stem cell (since the authors assume that the cell cycle times are equal and mutations are equally likely to occur in stem or transit cells. We relax this assumption in Section 5.3). The authors assume that $p_1 \ll 1/N$, which means there will be no mutants most of the time. This is typically the case, with mutation rates for the first hit estimated to be of the order of 10^{-7} [Michor *et al.*, 2004b].

In Sections 5.2 and 5.3 we will introduce a slightly different model to test the validity of Komarova and Wang's assumptions by considering multiple stem cells per crypt and using feedback to keep the stem cell compartment size as we expect; by allowing asynchronous cell division with different cell cycle times for stem and transit cells; and by allowing cells to die other than at the end of their natural lifetime in the crypt. For now, we examine the analysis in the Komarova and Wang [2004] model in the next section, and discuss their results in Section 5.1.4.

5.1.1 Mutational process equations

For each clone, Komarova and Wang wrote down a differential equation model for the mutational processes creating cells with two hits, they evaluated and approximated the rate of two transit cell hits, R, and used this to derive a condition for determining whether $P_{SD} \gtrless P_{DD}$.

There are two possible routes to achieving a double mutant. One route is that the whole crypt starts with wild-type cells and a mutation occurs in a stem cell (at a rate p_1) which reaches fixation, and then a second hit occurs (in either a stem or transit cell) at a rate Np_2. This route is $X_0 \to X_1 \to X_2$ and combines the SS and SD pathways. Alternatively, the crypt may convert directly from being wild type to having a double mutant at a rate R without fixation of a first hit. This route is $X_0 \to X_2$ and is the DD pathway. This mutational process is illustrated schematically in Figure 5.1. Hence, the authors wrote

down equations for these processes given by

$$\dot{x}_0 = -p_1 x_0 - R x_0, \tag{5.1}$$
$$\dot{x}_1 = p_1 x_0 - N p_2 x_1, \tag{5.2}$$
$$\dot{x}_2 = N p_2 x_1 + R x_0. \tag{5.3}$$

Taking initial conditions of $x_0 = 1$ and $x_1 = 0 = x_2$, we can clearly see that $x_0 + x_1 + x_2 = 1$ throughout. The solutions of these equations are

$$x_0 = e^{-(p_1+R)t}, \qquad x_1 = \frac{p_1}{N p_2 - p_1 - R}\left[e^{-(p_1+R)t} - e^{-N p_2 t}\right] \tag{5.4}$$
$$\text{and} \qquad x_2 = 1 - e^{-(p_1+R)t} - \frac{p_1}{N p_2 - p_1 - R}\left[e^{-(p_1+R)t} - e^{-N p_2 t}\right]. \tag{5.5}$$

Since p_1, p_2, N, $R > 0$, it is clear that $0 \leq x_0 \leq 1$ for $t \geq 0$. Also $(e^{-\alpha t} - e^{-\beta t})/(\beta - \alpha)$ will always be positive $\forall\ \alpha, \beta$, and since $\dot{x}_2 \geq 0$ with $x_2 = 0$ initially, using $x_0 + x_1 + x_2 = 1$ it is clear that $0 \leq x_1, x_2 \leq 1$ for $t \geq 0$.

The probability of getting a double mutant, x_2, is simply the sum of the probabilities from each different pathway, and so

$$x_2 = P_{SS} + P_{SD} + P_{DD}. \tag{5.6}$$

We can split up (5.3) to give

$$\dot{P}_{SS} + \dot{P}_{SD} = N p_2 x_1 \qquad \text{and} \qquad \dot{P}_{DD} = R x_0, \tag{5.7}$$

and hence

$$P_{DD} = \frac{R}{p_1 + R}\left(1 - e^{-(p_1+R)t}\right) \to \frac{R}{p_1 + R} \quad \text{as } t \to \infty, \tag{5.8}$$

and

$$P_{SS} + P_{SD} = \frac{p_1}{p_1 + R} - \frac{N p_1 p_2 e^{-(p_1+R)t}}{(p_1 + R)(N p_2 - p_1 - R)} + \frac{p_1 e^{-N p_2 t}}{N p_2 - p_1 - R} \tag{5.9}$$
$$\to \frac{p_1}{p_1 + R} \quad \text{as } t \to \infty. \tag{5.10}$$

Since $P_{SS} + P_{SD} + P_{DD} \to 1$ as $t \to \infty$, it is certain that, given time, a double mutant cell will be created by one of the pathways. The authors assume that $P_{SD} \approx N P_{SS}$ (neglecting different cell cycle times for stem and transit cells) since there are N times as many transit cells as stem cells that can achieve the second hit. By this assumption, the vast majority of double mutants with a first hit in a stem cell will achieve their second hit in the transit cell compartment, and therefore

$$P_{SD} \approx \frac{N}{N+1} \frac{p_1\left[N p_2 (e^{-(p_1+R)t} - 1) - (p_1 + R)(e^{-N p_2 t} - 1)\right]}{(p_1 + R)(p_1 + R - N p_2)}. \tag{5.11}$$

5.1.2 Finding the rate of two transit cell hits

In this section we explain how Komarova and Wang derived the rate R at which two hits occur in transit cells without fixation of the first hit, and discuss their approximations.

In transit cell generation i there are 2^{i-1} descendants of any stem cell, and so $r_i = p_1 2^{i-1}$ is the probability per division of the first hit occurring in the i^{th} generation. If the i^{th} division of a transit cell creates a mutant with the first hit, then that cell has $L-i$ divisions remaining to achieve a second hit, and so $2^{L-i+1}-2$ descendants will be created before that clone terminally differentiates. The probability that at least one cell gets a second hit is

$$\begin{aligned} P_i &= \mathbb{P}(2^{\text{nd}} \text{ hit occurs given the } 1^{\text{st}} \text{ hit occurred in transit generation } i) \\ &= 1-(1-p_2)^{2^{L-i+1}-2}. \end{aligned} \tag{5.12}$$

Therefore, R can be written as

$$R = \sum_{i=1}^{L} r_i P_i = \sum_{i=1}^{L} p_1 2^{i-1} \left[1-(1-p_2)^{2^{L-i+1}-2}\right]. \tag{5.13}$$

As a first approximation, the authors assume that

$$R \approx \sum_{i=1}^{L} p_1 2^{i-1} \left[1-e^{-p_2\left(2^{L-i+1}-2\right)}\right], \tag{5.14}$$

to order $O(p_2^2)$, which is a good approximation since $p_2 \ll 1$. The presence of i in the exponent of 2 within the exponential term means that a simple closed form solution cannot be obtained for this summation. Instead, in order to simplify this, the authors replace the summation by an integral to give

$$R \approx p_1 \int_1^L 2^{x-1} \left[1-e^{-p_2\left(2^{L-x+1}-2\right)}\right] \, \mathrm{d}x. \tag{5.15}$$

Setting $y = p_2 2^{L-x+1}$, the integral becomes

$$R \approx \frac{N p_1}{\log 2}\left[\frac{1}{2} - \frac{1}{N} - p_2 e^{2p_2} \int_{2p_2}^{Np_2} \frac{e^{-y}}{y^2} \, \mathrm{d}y\right], \tag{5.16}$$

and integrating by parts produces

$$R \approx \frac{p_1}{\log 2}\left[-1+e^{-p_2(N-2)} + N p_2 e^{2p_2} \int_{2p_2}^{Np_2} \frac{e^{-x}}{x} \, \mathrm{d}x\right]. \tag{5.17}$$

The authors then approximate the last term in (5.17) using a formula for the exponential integral which is given by

$$Ei(x) = \int_x^{\infty} \frac{e^{-t}}{t} \, \mathrm{d}t = -\gamma - \log x + \sum_{k=1}^{\infty} \frac{(-1)^{k+1} x^k}{k k!}, \tag{5.18}$$

where $\gamma \approx 0.577$ is the Euler constant. It is assumed both that Np_2 is sufficiently small that $e^{-p_2(N-2)} = 1 + O(Np_2)$, but also that Np_2 is sufficiently large that the upper limit of

the integral can be replaced by ∞, both of which overestimate R. The summation terms are neglected because they are $O(p_2)$. Therefore, Komarova and Wang obtain

$$R \approx \frac{Np_1p_2}{\log 2}\left(|\log p_2| - \log 2 - \gamma\right). \tag{5.19}$$

Both these approximations introduced significant errors. Instead, assuming just that Np_2 is small, we may replace the integral by $Ei(2p_2) - Ei(Np_2)$, which produces a better approximation

$$R \approx \frac{Np_1p_2}{\log 2}(\log N - \log 2 - 1). \tag{5.20}$$

5.1.3 Finding the most likely pathway

According to (5.8), if $R > p_1$ then the most likely pathway for the creation of a double mutant is with both mutations occurring in transit cells. Using this condition, and R from (5.19), Komarova and Wang derived a condition for the probability of the DD pathway being more likely than the SD pathway given by

$$p_2|\log p_2| > \frac{1}{N}. \tag{5.21}$$

This was obtained since the authors ignore the $-\log 2 - \gamma$ and the factor of $1/\log 2 \approx 1.44$, which reduces the value of R to compensate for the overestimation in deriving (5.19). Note that this condition is independent of p_1 and only ever dependent on p_2 and N (or L, where $N = 2^L$). The authors arbitrarily choose 2000 cells in a crypt, and so the parameter N is given by $N = 2000/N_0$, where the number N_0 of stem cells per crypt varies between 1 and 32. Therefore, the division line between the SD and DD pathways in the (p_2, N_0) parameter space, using Komarova and Wang's approximations, is given by

$$N_0 = Cp_2|\log p_2|, \tag{5.22}$$

where $C = 2000$ is the size of the proliferating compartment. Several approximations have been made to achieve this formula. If the authors had not made any approximations to (5.13), the result would have been

$$\sum_{i=1}^{L} 2^{i-1}\left[1 - (1-p_2)^{2^{L-i+1}-2}\right] = 1, \qquad \text{where} \qquad L = \frac{\log(C/N_0)}{\log 2}. \tag{5.23}$$

If the better approximation (5.20) for R was used, the boundary would be given by

$$p_2 = \frac{N_0 \log 2}{C\left[\log\left(C/(2N_0)\right) - 1\right]}. \tag{5.24}$$

In Figure 5.2, we plot the dividing line between the most likely pathways for the two hits being SD or DD in the (p_2, N_0) parameter space using (5.22), (5.23) and (5.24). Below the line DD is more likely, and SD is more likely above it. Komarova and Wang gave no justification for their approximations, but the graph shows that these approximations hold for low orders of p_2, but as p_2 increases to higher orders the approximations break down. Their results table is reproduced in Table 5.2, where the authors use (5.21) to predict the most likely pathway to cancer for different values of p_2 and N_0. We present our simulation results from Section 5.2.2 in tables of the same form in order to compare with Komarova and Wang.

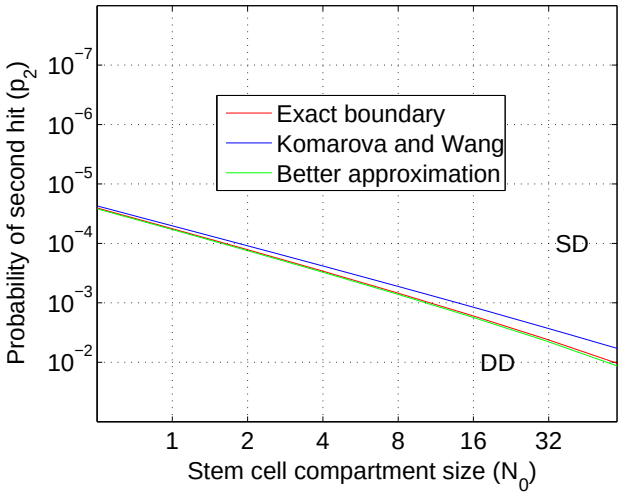

Figure 5.2: Testing the approximations made by Komarova and Wang [2004] for the SD:DD boundary. We plot the exact boundary (5.23), Komarova and Wang's approximation (5.22) and a better approximation given by (5.24), with $C = 2000$.

		N_0					
		1	2	4	8	16	32
p_2	10^{-7}	SD	SD	SD	SD	SD	SD
	10^{-5}	SD	SD	SD	SD	SD	SD
	10^{-4}	DD	SD	SD	SD	SD	SD
	10^{-3}	DD	DD	DD	DD	SD	SD
	10^{-2}	DD	DD	DD	DD	DD	DD

Table 5.2: Pathways to cancer predicted by Komarova and Wang [2004] for variable numbers of stem cells (N_0) and probability of the second hit (p_2) using (5.22).

5.1.4 Results from the model

To check their analytic results, Komarova and Wang used numerical simulations to follow the progeny of one stem cell until a second hit occurred, and they recorded which pathway was taken. They repeated the simulation 10^3 and 10^7 times and the same trends were observed, suggesting convergence for the results.

Using their simulation results, and their analytic formula, Komarova and Wang deduced that the probability of the SS pathway is negligible, meaning that it is very unlikely that both hits occur in stem cells. The relative importance of the SD and DD pathways depends on the probability of the second hit and the number of cells in the clone. The authors concluded that at least one hit is likely to occur in a transit cell, and if the mutation rate of the second hit is sufficiently raised (due to chromosomal instability), then DD becomes the most likely pathway. This is consistent with observations that below the dysplastic layer in early adenomatous crypts the cells are phenotypically normal [Shih *et al.*, 2001]. This model challenges previously-held beliefs that only stem cells are the likely targets for cancer initiation, and suggests that a possible role of the transit cells in tumourigenesis cannot be discarded.

Komarova and Wang's assumptions		Our assumptions	
KW1	Stem and transit cells have the same cell cycle time	J1	The stem and transit cell cycle times can differ
KW2	There is no cell death	J2	Cell death is included
KW3	There are 2000 proliferating cells in a crypt, 1–32 of which are stem cells	J3	There are 156 proliferating cells in a crypt, 1–16 of which are stem cells
KW4	1 stem cell and its clone are considered	J4	All proliferating cells are considered
KW5	Stem cells divide asymmetrically and their number remains fixed	J5	Stem cells divide both symmetrically and asymmetrically and their number fluctuates
KW6	All cells divide synchronously	J6	Asynchronous cell division occurs

Table 5.3: Key assumptions made by Komarova and Wang [2004], and our corresponding assumptions, that are tested by our simulation model in Section 5.2 and our analytic model in Section 5.3. We assume up to 16 stem cells and 156 proliferating cells based on data from Potten and Loeffler [1990].

5.2 Individual cell-based model

For the rest of this chapter we aim to reproduce the results in Komarova and Wang [2004], which we described in Section 5.1, and to test the assumptions they made. In this section we use an individual cell-based simulation model, and in Section 5.3 we verify these results using an analytic approach.

The key assumptions made by Komarova and Wang (KW1–KW6), and our corresponding assumptions (J1–J6), are listed in Table 5.3. Specifically we aim to:

- permit different cell cycle times for the stem and transit cells (J1);
- include cell death (J2);
- use a more realistic proliferating compartment size (J3);
- include multiple stem cells in the crypt (J4);
- allow the number of stem cells to vary through symmetric divisions (J5);
- allow asynchronous cell divisions (J6).

Our individual cell-based simulation model in this section incorporates all the assumptions J1–J6. The simulation allows for random division of cells in order to try to capture the stochastic nature of crypt homeostasis. We use a model where the transit cells divide a fixed number of times before terminally differentiating, which is a common assumption [Loeffler *et al.*, 1997], and this model is explained further in Section 5.3. For the purposes of our simulation and analytic models we ignore the presence of the fully-differentiated cells, since these do not divide and are not necessary for our purposes, and we assume that a first mutation has already occurred. We are interested in the population and mutational path

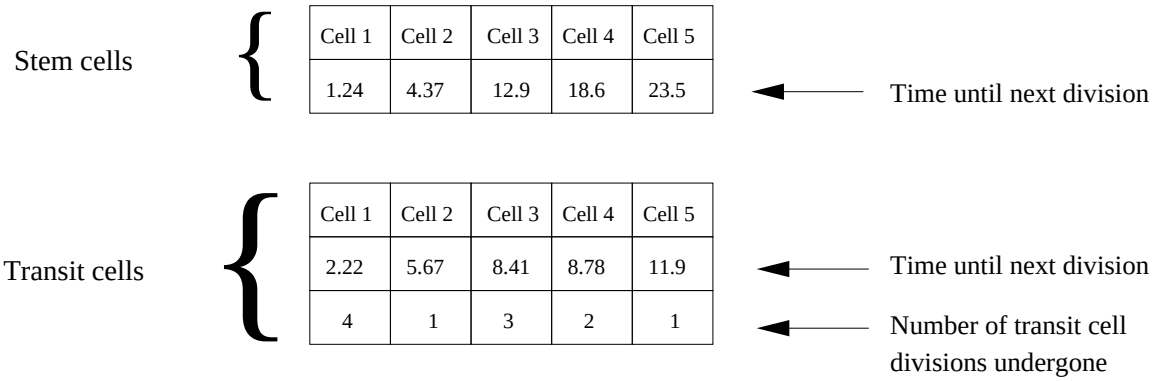

Figure 5.3: Schematic of cell storage in the individual cell-based simulation process described in Section 5.2.1. For each stem and transit cell the length of time until the next division is stored, and for the transit cells the number of divisions undergone is also stored.

of the cell that has achieved a second hit, and we are not concerned whether the second hit reaches fixation. In the next section we describe the simulation process that we use.

5.2.1 The simulation process

The crypt is divided into populations of stem and transit cells, both of which are represented by a matrix. Each column in the matrix corresponds to a particular cell, and the value represents the length of time until that cell next divides, stored in ascending order. The transit cell matrix also records the number of divisions undergone whilst in the transit cell compartment. This cell storage process is illustrated in the schematic in Figure 5.3.

In order to consider the mutants in the crypt, each stem (respectively transit) cell matrix is subdivided into healthy and mutant cells denoted by N_0 and M_0 (respectively N_1 and M_1). Komarova and Wang assumed that the proliferating compartment size C is fixed, and so we achieve this at each time step by removing the oldest transit cells. Komarova and Wang chose $C = 2000$, but a more appropriate value for the murine small intestinal crypt, according to data from Potten and Loeffler [1990], is $C = 156$: we use both values in our simulations. In order to model asynchronous cell division, each cell is allocated a cell cycle time upon division which is drawn from a normal distribution with expected values t_0 and t_1 for stem and transit cells, respectively.

There are several parameters that are fixed in these simulations: the total number of proliferating cells in a crypt ($N_0 + N_1 + M_0 + M_1 = C$); the probability of death at each division (d); the expected number of stem cells ($\bar{N}_0$); and the expected cell cycle times ($t_0 = 24$ and $t_1 = 12$ hours). The cell cycle times are varied using a normal distribution with mean values t_0 and t_1, and a standard deviation of 0.1.

The simulation algorithm is as follows:

1. Search through the stem and transit cell populations to find the shortest time until the next division. Advance to that point in time by subtracting that value from each stem and transit cell entry.

2. Search for all the cells with zero remaining time until the next division, and then work out how they will divide:

(a) If a stem cell is to divide, pick randomly whether it will die (probability $a_1 = 0.1$) or not. If it does not die, homeostatic regulation determines the result of the division. If the stem cell population size is less than expected ($N_0 + M_0 < \bar{N}_0$), then the stem cell will divide symmetrically to produce two stem cells; if the stem cell population is greater than expected ($N_0 + M_0 > \bar{N}_0$), then the stem cell will divide symmetrically to produce two transit cells; finally, if the stem cell number is as expected ($N_0 + M_0 = \bar{N}_0$), then the cell will divide asymmetrically to produce one stem cell and one transit cell;

(b) Transit cell divisions are symmetric and produce two transit cells with the number-of-divisions counter increased by one from their parent;

(c) When creating a stem (transit) cell, the matrix entry is drawn from a normal distribution about the cell cycle time t_0 (t_1).

3. After each division the total number of proliferating cells in the crypt is maintained at a fixed size by deleting the oldest residents of the transit cell compartment if the fixed size C is exceeded.

4. This loop ends and we return to step 1 for the next cell division.

We also assume that the first mutation will occur randomly in either stem or transit cells, because there is no evidence to support different mutation rates for different populations. Therefore, for each run, we begin the simulation with a random initial configuration of cells for each population, with transit cells at different stages of differentiation, and we pick a cell at random to be mutated. This choice is weighted with the relative cell cycle times, because we expect that since the transit cells cycle approximately twice as fast as the stem cells there will be twice as many divisions that could mutate. Hence, the probabilities of a stem or transit cell being mutated are given by, respectively,

$$\mathbb{P}(1^{\text{st}} \text{ hit in an } N_0) = \frac{N_0/t_0}{N_0/t_0 + N_1/t_1} \quad \text{and} \quad \mathbb{P}(1^{\text{st}} \text{ hit in an } N_1) = \frac{N_1/t_1}{N_0/t_0 + N_1/t_1}. \tag{5.25}$$

Komarova and Wang [2004] effectively assumed that $N_0/(N_0 + N_1)$ of the first mutations will occur in stem cells, but our calculation predicts close to half this value.

We assume that an N_0 is mutated to an M_0 and an N_1 is mutated to an M_1. Once a cell is chosen for mutation, the system is run until either that mutant is lost (because it is a transit cell that has terminally differentiated, or because it is a stem cell that has been chosen for death) or a doubly-mutated cell has been created. If all mutants have died out, the simulation returns to the initial configuration and another cell is picked to receive the first hit. If a double mutant was created, then its pathway (SS, SD or DD) is recorded and the simulation is stopped. A record is kept of the number of times that the simulation returned to the initial configuration before the two hits occurred, which can give an insight into the time scale for achieving two hits. Simulations were coded in C++, and the results are described in the next section.

5.2.2 Simulation results

For each set of the simulations we fixed the parameters C, d, t_0 and t_1, and simulated each combination of $N_0 \in \{1, 2, 4, 8, 16, 32\}$ and $p_2 \in \{10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}\}$, where p_2 is the probability of the second mutation. For each set of parameter values the simulation was run

		First Set				Second Set				Third Set			
p_2	N_0	SS	SD	DD	M	SS	SD	DD	M	SS	SD	DD	M
10^{-2}	1	0	1	99	27	0	0	100	29	0	4	96	29
	2	1	3	96	32	0	2	98	40	1	14	85	26
	4	0	3	97	26	0	3	97	27	1	21	78	22
	8	0	8	92	24	0	7	93	25	1	39	60	20
	16	1	15	84	23	2	8	90	24	5	67	28	20
	32	0	24	76	22	0	11	89	28	-	-	-	-
10^{-3}	1	1	7	92	142	0	2	98	158	0	37	63	122
	2	0	15	85	183	0	9	91	175	1	35	64	114
	4	1	27	72	148	0	7	93	157	3	60	37	105
	8	0	38	62	104	1	21	78	147	6	68	26	68
	16	0	60	40	86	0	32	68	113	4	92	4	49
	32	1	68	31	79	0	51	49	110	-	-	-	-
10^{-4}	1	0	39	61	810	0	20	80	1159	0	66	34	245
	2	0	52	48	804	0	36	64	1184	0	87	13	250
	4	0	68	32	539	0	55	45	744	1	88	11	249
	8	0	79	21	419	0	69	31	640	5	92	3	179
	16	1	95	4	248	1	68	31	447	6	92	2	101
	32	1	93	6	183	0	86	14	316	-	-	-	-
10^{-5}	1	0	72	28	1623	0	63	37	2770	0	64	36	255
	2	0	86	14	1601	0	79	21	2536	1	97	2	276
	4	0	97	3	502	0	81	19	2409	2	97	1	328
	8	0	99	1	247	0	89	11	1811	4	95	1	225
	16	0	100	0	118	0	94	6	1082	9	90	1	221
	32	3	97	0	63	1	96	3	770	-	-	-	-

Table 5.4: Results from the simulations described in Section 5.2.2 for varying p_2 and N_0 at three sets of parameter values C, d, t_0 and t_1. For each parameter set, from 100 runs we record the number of pathways that were SS, SD or DD, and M denotes the number of mutations that were lost before one led to a second hit. For the first set of simulations the parameters were $C = 2000$, $d = 0$ and $t_0 = t_1$; for the second set we used $C = 2000$, $d = 0.1$ and $t_0 = 2t_1$; and for the third set we chose $C = 156$, $d = 0.1$ and $t_0 = 2t_1$. Since Potten and Loeffler [1990] assumed up to 16 stem cells, we did not perform the third set of simulations when $N_0 = 32$.

100 times in order to obtain an estimate for the probability of each pathway occurring. For varying C, d, t_0 and t_1, we can test assumptions KW1–KW3, and each of our simulations use our assumptions J4–J6. The results are summarised in Table 5.4 which shows, for each combination of the parameters, how many of the 100 runs ended in the SS, SD and DD pathways, and how many mutations were lost before one led to a second hit. Summaries of these results for simulation sets 1, 2 and 3 are shown in Tables 5.5, 5.6 and 5.7, respectively, in the same format as Table 5.2 in order to ease comparison with the results from Komarova and Wang [2004]. Each entry corresponds to the most frequent pathway out of the 100 runs, and when the result differs from that in Table 5.2, we highlight the pathway in red.

For the first set of simulations we aimed to reproduce the Komarova and Wang [2004] results using our model, and so we used assumptions KW1–KW3, and tested our assumptions J4–J6 against KW4–KW6. The parameter values were $t_0 = t_1$, $d = 0$ and $C = 2000$, and the results are summarised in Table 5.5. We conclude that our simulation results agree

		N_0					
		1	2	4	8	16	32
p_2	10^{-5}	SD	SD	SD	SD	SD	SD
	10^{-4}	DD	SD	SD	SD	SD	SD
	10^{-3}	DD	DD	DD	DD	SD	SD
	10^{-2}	DD	DD	DD	DD	DD	DD

Table 5.5: Summary of the pathways to cancer for the first set of simulations presented in Table 5.4, when $C = 2000$, $d = 0$ and $t_0 = t_1$.

		N_0					
		1	2	4	8	16	32
p_2	10^{-5}	SD	SD	SD	SD	SD	SD
	10^{-4}	DD	DD	SD	SD	SD	SD
	10^{-3}	DD	DD	DD	DD	DD	SD
	10^{-2}	DD	DD	DD	DD	DD	DD

Table 5.6: Summary of the pathways to cancer for the second set of simulations presented in Table 5.4, when $C = 2000$, $d = 0.1$ and $t_0 = 2t_1$. Red denotes differences compared with Table 5.2.

		N_0				
		1	2	4	8	16
p_2	10^{-5}	SD	SD	SD	SD	SD
	10^{-4}	SD	SD	SD	SD	SD
	10^{-3}	DD	DD	SD	SD	SD
	10^{-2}	DD	DD	DD	DD	SD

Table 5.7: Summary of the pathways to cancer for the third set of simulations presented in Table 5.4, when $C = 156$, $d = 0.1$ and $t_0 = 2t_1$. Red denotes differences compared with Table 5.2.

with those presented by Komarova and Wang, and so the inclusion of multiple stem cells, stem cell feedback and asynchronous cell division, whilst being more realistic, does not significantly alter their results. Clones may also be considered separately because they do not influence each other. Komarova and Wang began their simulations with a healthy crypt, whereas we started each simulation by introducing a first hit to reduce the computation time, and since the results agree we also conclude that our assumption was acceptable.

For the second set of simulations, we aimed to test whether cell death and different cell cycle times could be neglected, and so we used assumptions J1–J2, KW3 and J4–J6. The parameter values were $C = 2000$, $t_0 = 2t_1$ and $d = 0.1$, and the results are summarised in Table 5.6. We conclude that the inclusion of death and different cell cycle times increases the likelihood of the DD pathway to higher mutation rates, and permits the threshold for the SD:DD boundary to occur at slightly lower values of p_2. This is as we would expect, since the second hit will occur more readily in the transit cells when they cycle faster than the stem cells.

For the third set of simulations, we aimed to test whether varying the proliferating compartment size altered the results, and so we used assumptions J1–J6. The parameter

values were $t_0 = 2t_1$, $d = 0.1$ and $C = 156$, which is the expected proliferating compartment size according to Potten and Loeffler [1990], and we only vary the number of stem cells between 1 and 16. The results are summarised in Table 5.7, and indicate that reducing the proliferating compartment size C shifts the SD:DD boundary by an order of magnitude towards higher values of p_2. Therefore, with a smaller proliferating compartment, the probability of both hits occurring in a transit cell is markedly decreased.

We have assumed throughout that the transit cells divide a fixed number of times before terminally differentiating, and in order to test this assumption we also considered a different model of transit cell differentiation. We assumed that transit cells could terminally-differentiate at any division with a fixed probability, which was chosen such that the total proliferating compartment size is C on average. Simulations were run for the same parameters as in the first set of simulations above and the results broadly agreed, suggesting that the model of differentiation does not affect our conclusions (details not shown).

5.3 Analytic model

In this section we aim to derive an analytic formula to determine under which parameter regimes the different pathways of SS, SD and DD are most likely, and to verify our simulation results from Section 5.2. We test our assumptions J1–J3 but revert to the Komarova and Wang [2004] assumptions KW4–KW6, because we use a steady state in order to simplify the calculations. Komarova and Wang assumed that the transit cells divide a fixed number of times before differentiating, and we do the same. Again, we assume that a first mutation has already occurred, and we calculate the probability of a second hit occurring through each pathway, whether or not this mutant reaches fixation.

In the next section we write down a mathematical model where there is a fixed number of transit cell divisions, which was described in Section 5.2, and we use this to calculate the probabilities of the pathways in the subsequent sections.

5.3.1 Model for a fixed number of transit cell divisions

We assume that the transit cells divide a pre-determined number of times (n) before they terminally differentiate, and each division is considered as a 'sub-compartment' of the transit cell population. We denote the n sub-compartments by T_i, $i = 1, \ldots, n$, where i denotes the number of divisions undergone since the cell left the stem cell compartment. For our calculations here, we use the age-structured model with the same proportion parameters as in Chapter 3 for the stem and fully-differentiated cell populations. We assume that the probabilities of death and maturation to the next level are the same for each transit cell sub-population, and are denoted by d and $m = 1 - d$, respectively. The compartments are summarised in Figure 5.4.

The cell populations will satisfy the same age-structured equations as we derived in Chapter 3, and so from (3.17) we see that $N_i = f_i(t - a)$ and $T_i = g_i(t - a)$ for arbitrary functions f and g, and the cell division conditions will now be

$$N_0(t, 0) = 2a_3 N_0(t, t_0), \tag{5.26}$$

$$T_1(t, 0) = 2a_2 N_0(t, t_0), \tag{5.27}$$

$$T_i(t, 0) = 2mT_{i-1}(t, t_1), \quad i = 2, \ldots, n, \tag{5.28}$$

$$N_2(t, 0) = 2mT_n(t, t_1). \tag{5.29}$$

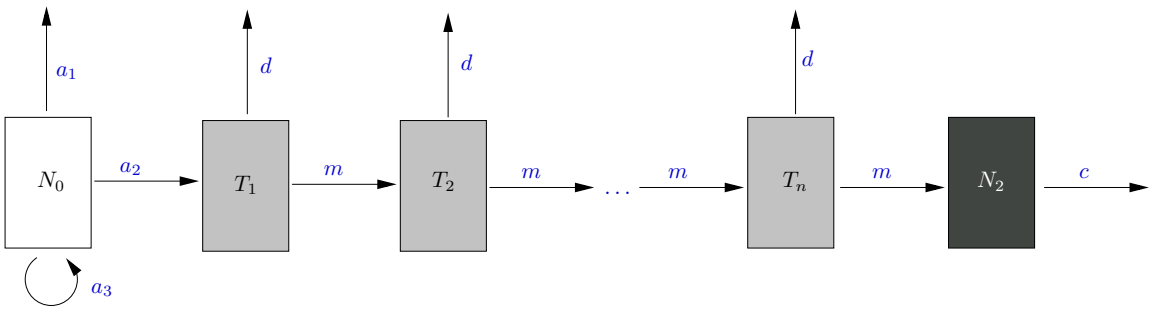

Figure 5.4: The compartment structure for the model where the transit cells divide a fixed number of times. This has the same stem and fully-differentiated cell compartments as described in Chapter 3, but now the transit cell compartment (N_1) is split up into n sub-compartments $T_1 \ldots T_n$. The proportions of transit cells that die and differentiate in each compartment are denoted by, respectively, d and m.

For the purposes of our calculations in the next section, we only consider this model in steady state. Therefore we let all the cell population sizes be constant and, in particular, the stem cells will be maintained by exact renewal, which means that $a_3 = 1/2$. If we take a uniform initial age profile for the stem cells, then the stem cell population will be constant throughout with $N_0(t,a) = \hat{n}_0/t_0, \ \forall\, t, a$. This means there is a fixed number of stem cells ($\hat{n}_0$) at any point in time.

Clearly the steady state T_1 population will be $T_1(t,a) = 2a_2\hat{n}_0/t_0$, and so the total age-integrated steady-state population $\hat{T}_1$ in the i^{th} transit cell compartment is

$$\hat{T}_i = 2a_2\hat{n}_0\frac{t_1}{t_0}(2m)^{i-1}. \tag{5.30}$$

Therefore, the total N_1 population is given by

$$N_1 = \sum_{i=1}^{n}\hat{T}_i(t) = 2a_2\hat{n}_0\frac{t_1}{t_0}\left(\frac{(2m)^n - 1}{2m-1}\right). \tag{5.31}$$

We assume that the probability of death is the same for stem and transit cells, $a_1 = d$, which means that $a_2 = 1/2 - d$. Therefore we require $d < 1/2$ for the probability of stem cell differentiation to be positive. If $C = N_0 + N_1$ is the total proliferating compartment size, then after rearranging and substituting d for m, and writing $\hat{n}_0 = N_0$, we find that

$$(2-2d)^n = 1 + \frac{t_0}{t_1}\left(\frac{C}{N_0} - 1\right) \tag{5.32}$$

$$\Rightarrow \qquad n = \frac{\log\left[1 + \frac{t_0}{t_1}\left(\frac{C}{N_0} - 1\right)\right]}{\log(2-2d)}. \tag{5.33}$$

According to Potten and Loeffler [1990], it is likely that $4 \leq N_0 \leq 16$, C is between 50% and 80% of 250 cells, $12 \leq t_0 \leq 32$ and $11 \leq t_1 \leq 12$, and so $3.5 < n < 8.4$ when $d = 0.1$. In particular, if $N_0 = 6$, $C = 156$, $t_0 = 24$ and $t_1 = 12$, then $n = 6.7$.

5.3.2 Probabilities of each pathway

In order to derive probabilities for the different pathways SS, SD and DD, we consider the Partition Theorem and condition on the probability of the first hit being either a stem or

transit cell. Therefore

$$P_{SS} = \mathbb{P}(SS|H_1 = S)\mathbb{P}(H_1 = S) + \mathbb{P}(SS|H_1 = D)\mathbb{P}(H_1 = D), \tag{5.34}$$

where P_{**} denotes the probability that the next double mutant is created via the $**$ pathway, $\mathbb{P}(H_i = *)$ denotes the probability that the i^{th} hit occurs in population $*$, and $\mathbb{P}(**|H_1 = *)$ denotes the probability that the next double mutant is created via the $**$ pathway, given that a first hit has occurred in population $*$. If the first hit occurs in a stem cell, then in order to get a double hit via the SS pathway either a second hit will occur in that stem cell or it will die out and another stem cell will get a first hit, resulting in SS. Hence

$$\begin{aligned}\mathbb{P}(SS|H_1 = S) &= \mathbb{P}(H_2 = S|H_1 = S) + \mathbb{P}[(H_1 = S \to 0) \cap (SS \text{ occurs next})] &(5.35)\\ &= \mathbb{P}(H_2 = S|H_1 = S) + \mathbb{P}(H_1 = S \to 0) \times P_{SS}, &(5.36)\end{aligned}$$

where $\mathbb{P}(H_1 = * \to 0)$ denotes the probability that after the first hit occurs in population $*$ the mutants die out without achieving a second hit, and we return to the initial configuration, whereby the probability of the SS pathway is independent of previous events. This allows for the possibility that multiple first hits are created and then lost, as we found in the simulations, before one mutated cell achieves a second hit. In particular, the mutated stem cell itself can die, and we return to the initial configuration. Finally, if the SS pathway is to create the next double mutant, but the first hit occurs in a transit cell, that mutant cell line must die out before another first hit can occur in a stem cell, and so

$$\mathbb{P}(SS|H_1 = D) = \mathbb{P}[(H_1 = D \to 0) \cap (SS \text{ occurs next})] = \mathbb{P}(H_1 = D \to 0) \times P_{SS}. \tag{5.37}$$

Combining (5.34), (5.36) and (5.37), and rearranging for P_{SS}, we find that

$$P_{SS} = \frac{\mathbb{P}(H_1 = S)\mathbb{P}(H_2 = S|H_1 = S)}{1 - \mathbb{P}(H_1 = S)\mathbb{P}(H_1 = S \to 0) - \mathbb{P}(H_1 = D)\mathbb{P}(H_1 = D \to 0)}. \tag{5.38}$$

Similarly, for the SD and DD pathways we derive

$$P_{SD} = \frac{\mathbb{P}(H_1 = S)\mathbb{P}(H_2 = D|H_1 = S)}{1 - \mathbb{P}(H_1 = S)\mathbb{P}(H_1 = S \to 0) - \mathbb{P}(H_1 = D)\mathbb{P}(H_1 = D \to 0)}, \tag{5.39}$$

$$P_{DD} = \frac{\mathbb{P}(H_1 = D)\mathbb{P}(H_2 = D|H_1 = D)}{1 - \mathbb{P}(H_1 = S)\mathbb{P}(H_1 = S \to 0) - \mathbb{P}(H_1 = D)\mathbb{P}(H_1 = D \to 0)}. \tag{5.40}$$

We assume that the probability of more than one distinct first mutation existing at any one time is vanishingly small, since $Cp_1 \ll 1$, and the expected time for the mutated cell line either to be mutated again or die out is much shorter than the expected time until another first hit.

Our goal is to derive a parameter boundary for the most likely pathway being between the SS, SD and DD pathways, which can be achieved by equating (5.38), (5.39) and (5.40). We note that the denominators of P_{SS}, P_{SD} and P_{DD} are all the same and will cancel, meaning that we only need to find the numerators. The denominator can be found using the condition that $P_{SS} + P_{SD} + P_{DD} = 1$. For the SD:DD boundary we set $P_{SD} = P_{DD}$ to get

$$\mathbb{P}(H_1 = S)\mathbb{P}(H_2 = D|H_1 = S) = \mathbb{P}(H_1 = D)\mathbb{P}(H_2 = D|H_1 = D), \tag{5.41}$$

for the SS:SD boundary we equate $P_{SS} = P_{SD}$ to find

$$\mathbb{P}(H_2 = S|H_1 = S) = \mathbb{P}(H_2 = D|H_1 = S), \tag{5.42}$$

and for the SS:DD boundary we equate $P_{SS} = P_{DD}$ which gives

$$\mathbb{P}(H_1 = S)\mathbb{P}(H_2 = S|H_1 = S) = \mathbb{P}(H_1 = D)\mathbb{P}(H_2 = D|H_1 = D). \tag{5.43}$$

Hence, to derive these three boundaries we simply need to find the five probabilities $\mathbb{P}(H_1 = S)$, $\mathbb{P}(H_1 = D)$, $\mathbb{P}(H_2 = S|H_1 = S)$, $\mathbb{P}(H_2 = D|H_1 = S)$ and $\mathbb{P}(H_2 = D|H_1 = D)$.

5.3.3 Deriving the probabilities

The probabilities of the first hit occurring in a stem or transit cell were given in (5.25). These can be rearranged using (5.33) to give

$$\mathbb{P}(H_1 = S) = \frac{N_0}{Ct_0/t_1 - N_0(t_0/t_1 - 1)} = \frac{1}{(2-2d)^n} \tag{5.44}$$

and

$$\mathbb{P}(H_1 = D) = \frac{t_0}{t_1}\left(\frac{C - N_0}{Ct_0/t_1 - N_0(t_0/t_1 - 1)}\right) = \frac{(2-2d)^n - 1}{(2-2d)^n}. \tag{5.45}$$

Note that since we always introduce the first hit, the probability of a first hit occurring at each division (p_1) will never be required in the model.

Next we calculate $\mathbb{P}(H_2 = S|H_1 = S)$, the probability that a stem cell gets a second mutation before any transit cell does and before all the mutants die out, given that the first hit has occurred in a stem cell. Let d be the probability that the cell dies at each division, p_2 be the probability that the cell survives and mutates, and $1 - d - p_2$ be the probability that the cell survives and does not mutate.

Since the number of cells in the i^{th} transit cell steady state, using (5.30), is given by

$$\hat{T_i} = (1-2d)N_0\frac{t_1}{t_0}(2-2d)^{i-1}, \tag{5.46}$$

the expected total number of transit cells in the crypt at any one time that are descendants of a single stem cell is given by

$$\sum_{i=1}^{n}(1-2d)\frac{t_1}{t_0}(2-2d)^{i-1} = \frac{t_1}{t_0}\left[(2-2d)^n - 1\right]. \tag{5.47}$$

Between each stem cell division there will be

$$\frac{t_1}{t_0}\left[(2-2d)^n - 1\right] \times \frac{t_0}{t_1} = (2-2d)^n - 1 \tag{5.48}$$

divisions of the mutant transit cells.

Since the time taken for the progeny of the mutant stem cell to divide n times to create a mutant clone is negligible compared to the time until the second hit, we start with the full mutant clone for the purposes of the calculation, which has the effect of slightly overestimating the effect of SD in comparison with DD. If $t = 0$ is the time of fixation of the

first hit, then by time $t = jt_0$, $j \in \mathbb{N}$, there will have been j divisions of the mutant stem cell and $j[(2-2d)^n - 1]$ divisions of the mutant transit cells. Therefore, the probability that the second hit occurs in a stem cell before any transit cell is given by

$$\mathbb{P}(H_2 = S|H_1 = S) = \sum_{j=0}^{\infty} \mathbb{P}(\text{second hit occurs at time } t = jt_0) \tag{5.49}$$

$$= \sum_{j=0}^{\infty} p_2(1-d-p_2)^j(1-p_2)^{j[(2-2d)^n-1]} \tag{5.50}$$

$$= \frac{p_2}{1-(1-d-p_2)(1-p_2)^{(2-2d)^n-1}}. \tag{5.51}$$

To find $\mathbb{P}(H_2 = D|H_1 = S)$, we condition on the next stem cell division to get

$$\begin{aligned}\mathbb{P}(H_2 = D|H_1 = S) = d \times \mathbb{P}(&\text{an } N_1 \text{ gets a hit before this batch of mutant cells leave})\\ &+ p_2 \times \mathbb{P}(\text{an } N_1 \text{ gets hit before } N_0|N_0 \text{ gets hit now})\\ &+ (1-d-p_2) \times \mathbb{P}(H_2 = D \text{ before or after the next } N_0 \text{ division}).\end{aligned} \tag{5.52}$$

Denote the three unknown probabilities on the right-hand side by $Pr1$, $Pr2$ and $Pr3$. Clearly

$$Pr2 = \mathbb{P}(\text{an } N_1 \text{ gets hit before } N_0|N_0 \text{ gets hit now}) = 0, \tag{5.53}$$

and conditioning on the second hit occurring in a transit cell before or after the next stem cell division, and using (5.48) we get

$$Pr3 = \mathbb{P}(H_2 = D \text{ before or after the next } N_0 \text{ division}) \tag{5.54}$$

$$= 1 \times \left[1-(1-p_2)^{(2-2d)^n-1}\right] + \mathbb{P}(H_2 = D|H_1 = S) \times (1-p_2)^{(2-2d)^n-1}. \tag{5.55}$$

Also, if the mutant stem cell dies, then we must calculate

$$Pr1 = \mathbb{P}(\text{an } N_1 \text{ gets a hit before this batch of mutant cells leave}). \tag{5.56}$$

The number of transit cells in each transit generation steady state within the descendants of each stem cell, using (5.46), is given by $\hat{T}_i/N_0$. For transit cells in generation i, the number of progeny in each subsequent generation before the cells leave is given by $(2-2d)^j$, $j = 1, 2 \ldots n-i$, since there are $n-i$ generations remaining. Therefore, each cell in generation i will produce

$$\sum_{j=1}^{n-i}(2-2d)^j = \frac{(2-2d)^{n-i+1} - (2-2d)}{1-2d} \tag{5.57}$$

remaining cells before they terminally differentiate. Hence, the total number of cells that could be mutant is given by

$$\sum_{i=1}^{n-1} \frac{\hat{T}_i}{N_0}\left[\frac{(2-2d)^{n-i+1} - (2-2d)}{1-2d}\right] = \frac{t_1}{t_0}\left[n(2-2d)^n + \left(\frac{2-2d}{1-2d}\right)\left[1-(2-2d)^n\right]\right], \tag{5.58}$$

and so

$$Pr1 = 1 - (1-p_2)^{\frac{t_1}{t_0}\left[n(2-2d)^n + \left(\frac{2-2d}{1-2d}\right)[1-(2-2d)^n]\right]}. \tag{5.59}$$

Hence, combining (5.52), (5.53), (5.55) and (5.59), we get

$$\begin{aligned}&\mathbb{P}(H_2 = D|H_1 = S)\\ &= \frac{d\left[1-(1-p_2)^{\frac{t_1}{t_0}\left[n(2-2d)^n + \left(\frac{2-2d}{1-2d}\right)[1-(2-2d)^n]\right]}\right] + (1-d-p_2)\left[1-(1-p_2)^{(2-2d)^n-1}\right]}{1-(1-d-p_2)(1-p_2)^{(2-2d)^n-1}}.\end{aligned} \tag{5.60}$$

Finally we consider $\mathbb{P}(H_2 = D|H_1 = D)$, the probability that the second hit occurs in a transit cell after the first hit has also occurred in a transit cell before all the mutants are lost. Using (5.46), the probability that the first hit occurs in transit generation i is given by

$$\hat{T}_i \Big/ \sum_{i=1}^{n} \hat{T}_i = \frac{(1-2d)(2-2d)^{i-1}}{(2-2d)^n - 1}. \tag{5.61}$$

Combining this with (5.57), the probability that a transit cell suffers two hits is given by

$$\begin{aligned}&\mathbb{P}(H_2 = D|H_1 = D)\\ &\qquad = \frac{(1-2d)}{(2-2d)^n - 1}\sum_{i=1}^{n}(2-2d)^{i-1}\left[1-(1-p_2)^{[(2-2d)^{n-i+1}-(2-2d)]/(1-2d)}\right].\end{aligned} \tag{5.62}$$

Hence, using (5.44), (5.45), (5.51), (5.60) and (5.62), the numerators of P_{SS}, P_{SD} and P_{DD} from (5.38)–(5.40) are given by

$$\text{Numerator of } P_{SS} = \frac{1}{(2-2d)^n} \times \frac{p_2}{1-(1-d-p_2)(1-p_2)^{(2-2d)^n-1}}, \tag{5.63}$$

$$\begin{aligned}&\text{Numerator of } P_{SD} = \frac{1}{(2-2d)^n}\times\\ &\frac{d\left[1-(1-p_2)^{\frac{t_1}{t_0}\left[n(2-2d)^n + \left(\frac{2-2d}{1-2d}\right)[1-(2-2d)^n]\right]}\right] + (1-d-p_2)\left[1-(1-p_2)^{(2-2d)^n-1}\right]}{1-(1-d-p_2)(1-p_2)^{(2-2d)^n-1}},\end{aligned} \tag{5.64}$$

$$\text{Numerator of } P_{DD} = \frac{(1-2d)}{(2-2d)^n}\sum_{i=1}^{n}(2-2d)^{i-1}\left[1-(1-p_2)^{[(2-2d)^{n-i+1}-(2-2d)]/(1-2d)}\right]. \tag{5.65}$$

From (5.41), (5.42) and (5.43), the SD:DD, SS:SD and SS:DD boundaries can be obtained by equating, respectively, (5.64) and (5.65); (5.63) and (5.64); and (5.63) and (5.65). In particular, we note that our relation for the SD:DD boundary gives good agreement with the exact formula (5.23) in the model by Komarova and Wang [2004], which we illustrate in Figure 5.5. Any small differences are due to the different model formulations: Komarova and Wang use a time-dependent model which begins with no hits, whereas we consider a steady state model and impose the first hit.

In Figure 5.6 we illustrate the regions of the (p_2, N_0) space where each of (5.63), (5.64) and (5.65) are the largest. This shows that the SS pathway is only probable when the number of stem cells is very large, typically bigger than 2^{10}, which is biologically unfeasible.

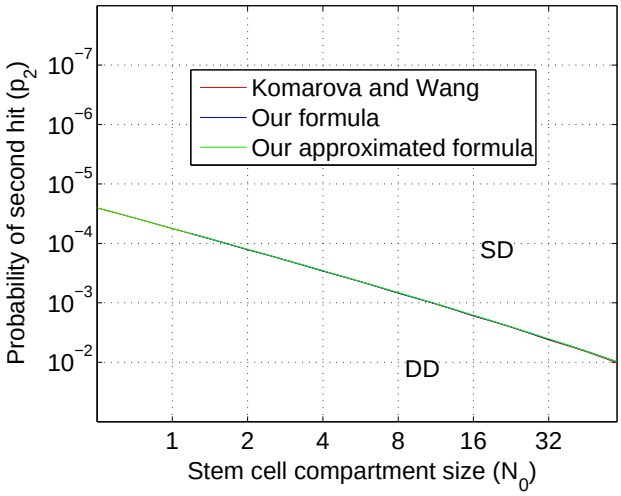

Figure 5.5: A comparison of formulae for the SD:DD boundary using the exact boundary (5.23) in the Komarova and Wang [2004] model, our result by equating (5.64) and (5.65), and our approximation (5.70) to this boundary, where the parameters are $C = 2000$, $d = 0$ and $t_0 = t_1$. Both of our boundaries show good agreement with the Komarova and Wang result.

5.3.4 Approximating the boundaries

In this section we aim to produce a closed form solution to the SD:DD boundary found by equating (5.64) and (5.65). We assume that $p_2 \ll 1$, since Komarova and Wang [2004] note that even chromosomal instability would only increase the mutation rate to at most $p_2 = 10^{-2}$ which is still small.

Firstly, considering the numerator of P_{SD} from (5.64), when $p_2 \ll 1$ we can approximate $(1-p_2)^K \approx e^{-p_2 K}$, and also assuming that $(2-2d)^n \gg 1$, produces

$$\text{Numerator of } P_{SD} \approx \frac{d\left[1 - e^{-p_2(2-2d)^n \frac{t_1}{t_0}\left[n - \left(\frac{2-2d}{1-2d}\right)\right]}\right] + (1-d)\left[1 - e^{-p_2(2-2d)^n}\right]}{(2-2d)^n\left[1-(1-d)e^{-p_2(2-2d)^n}\right]}. \quad (5.66)$$

Secondly, when approximating (5.65) for $p_2 \ll 1$, we can expand the last term in the summation in powers of p_2 to give

$$\text{Numerator of } P_{DD} = \frac{(1-2d)}{(2-2d)^n}\sum_{i=1}^{n}(2-2d)^{i-1}\left[p_2\frac{(2-2d)^{n-i+1}-(2-2d)}{1-2d}\right.$$
$$\left.-\frac{p_2^2}{2}\left(\frac{(2-2d)^{n-i+1}-(2-2d)}{1-2d}\right)\left(\frac{(2-2d)^{n-i+1}-(2-2d)}{1-2d}-1\right)+O(p_2^3)\right]. \quad (5.67)$$

Along the SD:DD boundary, between $N_0 = 1$ and $N_0 = 32$ we find that $0.11 < p_2(2-2d)^n < 0.26$ when $d = 0$, $t_0/t_1 = 1$ and $C = 2000$; that $0.08 < p_2(2-2d)^n < 0.16$ when $d = 0.1$, $t_0/t_1 = 2$ and $C = 2000$; and that $0.12 < p_2(2-2d)^n < 0.35$ between $N_0 = 1$ and $N_0 = 16$ when $d = 0.1$, $t_0/t_1 = 2$ and $C = 156$. We therefore assume that $p_2(2-2d)^n$ is a small parameter, which is most accurate when N_0 is near 1, and discard terms of $O(p_2^2)$ in the summation (5.67), which reduces to

$$\text{Numerator of } P_{DD} \approx \frac{p_2[n(1-2d)(2-2d)^n - (2-2d)^{n+1} + (2-2d)]}{(1-2d)(2-2d)^n}. \quad (5.68)$$

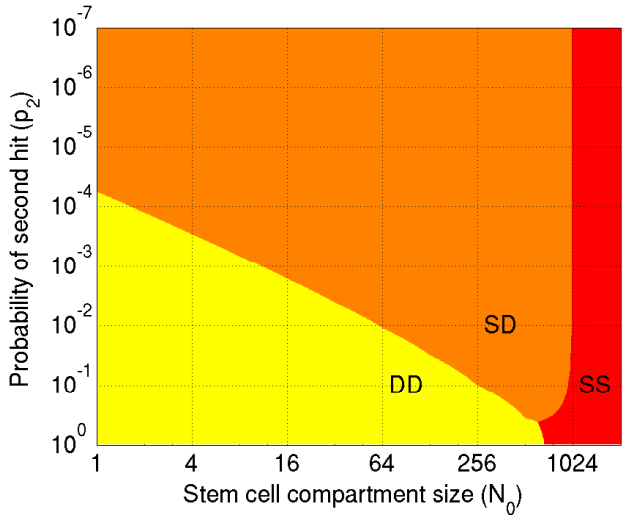

Figure 5.6: The different regions of the (N_0, p_2) parameter space corresponding to the most likely pathway to two hits being SS, SD and DD, using (5.63), (5.64) and (5.65), respectively. The parameter values are taken to be $C = 2000$, $d = 0$ and $t_0 = t_1$.

Assuming again that $(2-2d)^n \gg 1$, this reduces to

$$\text{Numerator of } P_{DD} \approx p_2 \left(n - \frac{2-2d}{1-2d} \right). \tag{5.69}$$

Equating (5.66) and (5.69) produces

$$(1 - A\epsilon)\left[1 - (1-d)e^{-\epsilon}\right] = de^{-\frac{t_1}{t_0}A\epsilon}, \quad \text{where} \quad \epsilon = p_2(2-2d)^n, \quad A = n - \frac{2-2d}{1-2d}. \tag{5.70}$$

(5.70) gives an approximate relation for the SD:DD boundary between p_2 and N_0 in terms of d, C and t_0/t_1, where n is given by (5.33). This is in good agreement with the exact boundary, found by equating (5.64) and (5.65), and we illustrate this in Figure 5.5.

The SS:SD boundary is found by equating (5.63) and (5.64), which gives

$$p_2 = d\left[1 - (1-p_2)^{\frac{t_1}{t_0}\left[(2-2d)^n\left(n - \frac{2-2d}{1-2d}\right) + \frac{2-2d}{1-2d}\right]}\right] + (1-d-p_2)\left[1 - (1-p_2)^{(2-2d)^n - 1}\right]. \tag{5.71}$$

Using (5.71), the asymptote of N_0 as $p_2 \to 0$, which is seen in Figure 5.6, can be found from

$$(2-2d)^n\left[d\frac{t_1}{t_0}\left(n - \frac{2-2d}{1-2d}\right) + 1 - d\right] = 2 - d - d\frac{t_1}{t_0}\left(\frac{2-2d}{1-2d}\right). \tag{5.72}$$

In particular, when $d = 0$ this is given by

$$N_0 = \frac{C}{1 + t_1/t_0}. \tag{5.73}$$

When $t_0 = 2t_1$ this gives $N_0 = 2C/3$. Therefore, for the SS pathway to be the most likely route to two hits, it is necessary that the stem cells must comprise at least two thirds of the proliferating compartment. Since this is not expected biologically, we conclude that there can be no feasible parameter values for the SS pathway to be the most likely route to two hits.

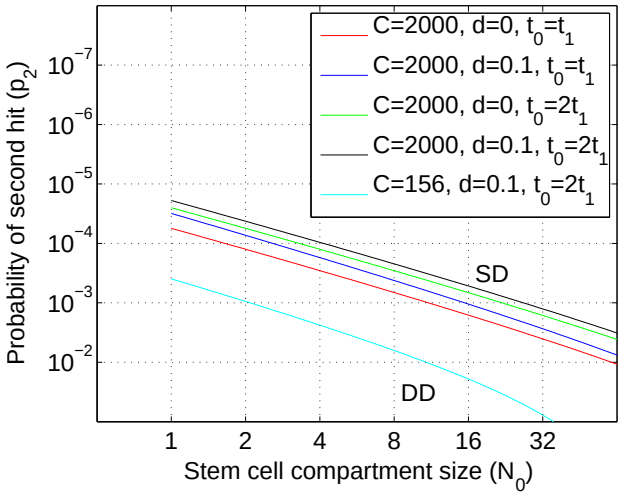

Figure 5.7: A comparison of the effects of death, different cell cycle times, and varying proliferating compartment size on the SD:DD boundary given by (5.70).

t_0/t_1	C	d	$\log_{10} p_2$	Difference
1	2000	0	-4.25	0
1	2000	0.1	-4.50	-0.25
2	2000	0	-4.60	-0.35
2	2000	0.1	-4.72	-0.47
2	156	0.1	-3.40	+0.85

Table 5.8: Quantifying the effects of including death (d), different cell cycle times (t_0 and t_1), and a varying proliferating compartment size (C) on the SD:DD boundary given by (5.70), which is illustrated in Figure 5.7. The value of $\log_{10} p_2$ when $N_0 = 1$ is displayed, along with the difference between this value and the result from Komarova and Wang [2004].

5.3.5 Interpreting the results

In this section we aim to specifically test the effect of relaxing Komarova and Wang's assumptions KW1, KW2 and KW3 from Table 5.3: namely, permitting different stem and transit cell cycle times, cell death and a more realistic proliferating compartment size of 156 cells. We test these assumptions by plotting the SD:DD boundary from (5.70) for different combinations of the parameters, with either $C = 2000$, or $C = 156$ as suggested by Potten and Loeffler [1990]; $t_0/t_1 = 1$ or $t_0/t_1 = 2$; and $d = 0$ or $d = 0.1$. These lines are plotted in Figure 5.7.

These curves are approximately parallel straight lines, and the parameter changes essentially translate the line up or down the p_2 axis. Therefore, to quantify the comparative effect of these changes, we consider the changes in the value of p_2 for a fixed number of stem cells. In Table 5.8 we show $\log_{10} p_2$ when $N_0 = 1$ for the five lines in Figure 5.7.

The inclusion of death can be seen to shift the line by -0.25, the inclusion of different cell cycle times has a greater effect shifting the line by -0.35, and including both shifts the line by -0.47. Therefore, these results suggest that the value of p_2 need not even be as high as expected by Komarova and Wang [2004] (by about half an order of magnitude in p_2) for the DD pathway to be the most likely, which means that the DD pathway may be significantly

more likely than predicted by the authors. However, this is assuming that Komarova and Wang's compartment size of 2000 proliferating cells is a reasonable estimate.

Using the Potten and Loeffler value of $C = 156$ for a murine small intestinal crypt, in addition to including death and different cell cycle times, the line shifts by +1.32 from the same line with $C = 2000$, emphasising the great effect changes in C will have. The net shift from the Komarova and Wang line is +0.85, and so the net likelihood of the DD pathway is reduced by nearly an order of magnitude. This suggests that Komarova and Wang have significantly overestimated the importance of the first mutational hit occurring in a transit cell. Although these values are for the small intestine, they represent realistic, experimentally-determined parameters and give an indication of similar effects in the colon.

In conclusion, the effect of including different cell cycle times and cell death have a small effect on the SD:DD boundary, and the effect of including a more realistic proliferating compartment size has a larger effect. We note that these results agree exactly with those from the simulation model in Tables 5.5, 5.6 and 5.7.

5.4 Discussion

In this chapter we have examined the model by Komarova and Wang [2004], which calculated the probability of different scenarios leading to two mutational hits in the *APC* gene in a crypt cell. We tested the authors' assumptions using an individual cell-based simulation and compared these results to an analytic model.

The assumption that multiple stem cells in a crypt can be achieved by combining individual clones neglects the idea of niche succession and bottlenecks. Van Leeuwen *et al.* [2006] used a stochastic simulation model to test whether two stem cell hits would occur more quickly in a crypt with or without the possibility of symmetric divisions. The authors found that allowing for symmetric divisions decreased the niche succession time, and increased the time until *APC* inactivation in a stem cell since single mutant stem cells have a greater chance of becoming extinct than of reaching fixation. Our simulation results, however, have shown that including symmetric divisions did not increase the relative likelihood of any pathway occurring.

Our results have shown that the assumptions of cell division synchrony, a fixed number of stem cells, and only considering one stem cell and its progeny did not affect the results. Including cell death and different cell cycle times led to a greater likelihood that both hits occur in transit cells. However, changing the size of the proliferative compartment had the greatest effect and, in particular, using a Potten and Loeffler [1990] compartment size for the murine small intestine dramatically reduced the likelihood that both hits occur in transit cells. Therefore, it is fundamentally important to make a good estimate of the proliferating compartment size in order to accurately predict the likelihood of mutations occurring in each cell population. We note that it will be difficult to test these results experimentally due to the problems with identifying the original double mutant cell.

Both Komarova and Wang [2004] and our model are based on a number of assumptions that need further critical analysis from a biological point of view. One questionable assumption is neglecting the possibility of the first hit conferring a selective advantage, which might make a crucial difference. A further development of this model would allow for different growth rates for mutant and healthy cells, for example. Also, it is assumed that two *APC* hits in a cell are enough to form a tumour, but this could only occur if the second hit gives unlimited growth to the cell or makes it a cancer-driving cell, but all the evidence suggests

that the second *APC* mutation occurs at the adenoma stage (W. F. Bodmer, pers. comm.). However, this model may also be used to consider the wider question of whether the critical mutation that takes the adenoma to a cancer, and creates a cancer-driving cell, occurs in a stem or transit cell.

Chapter 6

Spatial models of the crypt

In the models that have been presented in Chapters 3, 4 and 5 it has been assumed that all the cell populations are well-mixed within the crypt. There are, however, many spatial variations, not least that the stem cells are expected to reside near the bottom, the fully-differentiated cells near the top and the semi-differentiated cells in the middle. Chemical cues vary along the crypt axis, and the velocity of the cells is assumed to increase as the cells proceed up the crypt [Potten and Loeffler, 1990].

In this chapter we aim to address biological questions that can only be considered by a model that incorporates spatial dependencies. We begin by presenting a population density model in Section 6.1, where the volume fractions of each cell population are tracked along the crypt axis. We discuss what mechanisms might be in place to maintain the cells in their expected locations in the crypt, in particular with the rate parameters varying at different heights along the crypt axis, and we also show how chemotaxis or varying velocities can have a similar effect. This model is simply a spatial extension of the continuous model presented in Chapter 3, and we use the spatial model to verify the validity of the continuous model in Section 6.2. The model in Section 6.1 assumes a one-dimensional crypt; in Section 6.3 we consider a transformation to a more realistic crypt shape and discuss how this alters the cell population profiles.

The spatial model can be used to address the question of competition between different cell populations, which is discussed in Section 6.4. We illustrate how mutant cell clones propagate in the crypt, and we also extend the model to include individual stem cells in a niche in order to address the question of niche succession and a time scale for monoclonality, which was discussed in Chapter 1. Finally, we sum up these results in Section 6.5. Throughout this chapter we will consider the general theme of whether this modelling framework is sufficient to answer the questions that we pose, or whether a more detailed model, possibly in two or three dimensions, is required.

6.1 Population density model

In this section we present an extension to the continuous model from Chapter 3, where the cell populations have varying densities at each point in space. The crypt epithelium is viewed as a continuum of stem, semi- and fully-differentiated cells, and the volume fraction of each cell population is denoted by N_0, N_1 and N_2, respectively. Let $N_i = N_i(\boldsymbol{x}, t)$ depend on time t and the spatial position $\boldsymbol{x}$ in the crypt, for $i = 0, 1, 2$, and denote the local velocity of cells in compartment N_i by $\boldsymbol{v_i}(\boldsymbol{x}, t)$.

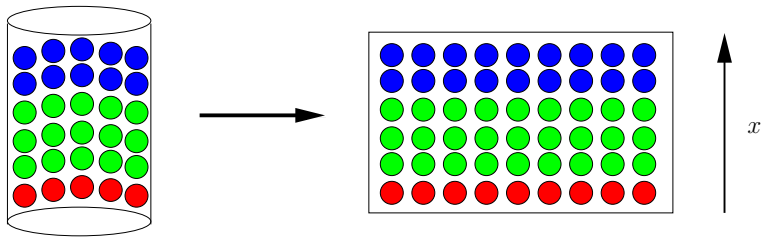

Figure 6.1: Schematic of a crypt represented by a cylinder being 'unwrapped' into a rectangular sheet of cells, so that there is a density of cells at each height along the crypt axis.

The epithelium of the crypt consists of a single layer of cells, with only one cell at each point $\boldsymbol{x}$, although we assume there is a density of cells at each point in space. We consider, as a first approximation, that the crypt is cylindrical and can be 'unwrapped', see for example Figure 6.1, and we discuss this assumption further in Section 6.3. Thus each point in a one-dimensional crypt really corresponds to a circle, and so the assumption of a density of each cell type at each point is plausible.

6.1.1 Model development

The population density model is developed from the same principles as the models by Ward and King [1997, 1999a,b], where the authors use a continuum approach to model avascular tumour growth. We assume that all the cells are equal in size, and that when they die they burst releasing fluid into the crypt, with death and dissipation occurring instantaneously. We assume that a dead cell has no volume, and so these are neglected in the model formulation. Following the same compartment assumptions as in Figure 3.1, the volume fractions for the cell populations satisfy

$$\frac{\partial N_0}{\partial t} + \boldsymbol{\nabla} \cdot (\boldsymbol{v_0} N_0) = (\alpha_3 - \alpha_1 - \alpha_2) N_0, \tag{6.1}$$

$$\frac{\partial N_1}{\partial t} + \boldsymbol{\nabla} \cdot (\boldsymbol{v_1} N_1) = \alpha_2 N_0 + (\beta_3 - \beta_1 - \beta_2) N_1, \tag{6.2}$$

$$\frac{\partial N_2}{\partial t} + \boldsymbol{\nabla} \cdot (\boldsymbol{v_2} N_2) = \beta_2 N_1 - \gamma N_2, \tag{6.3}$$

where $\alpha_i = \alpha_i(N_j, \boldsymbol{x}, t), \beta_i = \beta_i(N_j, \boldsymbol{x}, t)$ and $\gamma = \gamma(N_j, \boldsymbol{x}, t)$, for $i = 1, 2, 3$ and $j = 0, 1, 2$; that is, the rate parameters could vary depending on the population densities, space or time. We discuss possible dependences the rate parameters could take in Section 6.1.3. The second terms on the left-hand sides of these equations, $\boldsymbol{\nabla} \cdot (\boldsymbol{v_i} N_i)$, is composed of the advective effects of the cells upstream, $(\boldsymbol{v_i} \cdot \boldsymbol{\nabla}) N_i$, and the local volume change due to cell growth and division, $N_i(\boldsymbol{\nabla} \cdot \boldsymbol{v_i})$.

We assume that the rate of volume change drives the movement, although other features such as contact guidance, contact inhibition, haptotaxis (the directed movement of cells up an adhesive gradient) and galvanotaxis (movement generated by electric potentials) can also be considered [Murray, 2002]. Chemotaxis is considered in Section 6.1.4. We also assume a no-voids condition, following Ward and King [1997], where the cells are fully packed in the crypt with no spaces, and so the volume fractions satisfy

$$N_0 + N_1 + N_2 = 1. \tag{6.4}$$

As a simplification, we assume that each cell population has the same velocity, $\boldsymbol{v_i} = \boldsymbol{v}$. In this case, summing (6.1)–(6.3) and applying (6.4), the velocity satisfies

$$\boldsymbol{\nabla} \cdot \boldsymbol{v} = (\alpha_3 - \alpha_1)N_0 + (\beta_3 - \beta_1)N_1 - \gamma N_2. \tag{6.5}$$

Using (6.4) we can eliminate N_2 from (6.5), and so the model can be reduced to a system of three equations for the unknowns N_0, N_1 and $\boldsymbol{v}$, given by

$$\frac{\partial N_0}{\partial t} + \boldsymbol{\nabla} \cdot (\boldsymbol{v} N_0) = (\alpha_3 - \alpha_1 - \alpha_2)N_0, \tag{6.6}$$

$$\frac{\partial N_1}{\partial t} + \boldsymbol{\nabla} \cdot (\boldsymbol{v} N_1) = \alpha_2 N_0 + (\beta_3 - \beta_1 - \beta_2)N_1, \tag{6.7}$$

$$\boldsymbol{\nabla} \cdot \boldsymbol{v} = (\alpha_3 - \alpha_1)N_0 + (\beta_3 - \beta_1)N_1 - \gamma(1 - N_0 - N_1). \tag{6.8}$$

6.1.1.1 Reduction to one dimension

In order to obtain simple model solutions for (6.6)–(6.8), we consider only one spatial dimension along the crypt axis, where $\boldsymbol{x} = x \in [0, L]$, $\boldsymbol{v} = v$, and L is the crypt apex. In one dimension we have a closed system of equations for the variables N_0, N_1 and v, and (6.6)–(6.8) become

$$\frac{\partial N_0}{\partial t} + \frac{\partial}{\partial x}(v N_0) = \alpha(x) N_0, \tag{6.9}$$

$$\frac{\partial N_1}{\partial t} + \frac{\partial}{\partial x}(v N_1) = \beta(x) N_1 + \alpha_2(x) N_0, \tag{6.10}$$

$$\frac{\partial v}{\partial x} = [\alpha(x) + \alpha_2(x)]N_0 + [\beta(x) + \beta_2(x)]N_1 - \gamma(x)\left(1 - N_0 - N_1\right), \tag{6.11}$$

where $\alpha(x) = \alpha_3(x) - \alpha_1(x) - \alpha_2(x)$, $\beta(x) = \beta_3(x) - \beta_1(x) - \beta_2(x)$ and $\gamma = \gamma(x)$ may depend on x. We prescribe the initial spatial profiles of the cell populations to be Hill functions of the form

$$N_0(x,0) = \frac{1}{1 + \left(\frac{x/L}{x_s}\right)^p}, \qquad N_1(x,0) = \frac{1}{1 + \left(\frac{2x/L - x_t - x_s}{x_t - x_s}\right)^p}, \tag{6.12}$$

where $x_s L$ and $x_t L$ are, respectively, the $N_0 : N_1$ and $N_1 : N_2$ expected boundaries, with $0 < x_s < x_t < 1$, and x_s, x_t and p are constants. The initial spatial profiles (6.12) are illustrated in Figure 6.2, where $x_s = 0.2$ and $x_t = 0.5$ are chosen for illustrative purposes since we are only interested in the qualitative behaviour of the system.

We impose conditions at the bottom of the crypt for $t > 0$, and then the volume fractions at the top of the crypt can be determined by solving the system. At the crypt base, we impose that the velocity is zero and we could either choose the volume fractions or impose zero flux for N_0 and N_1. These conditions, along with (6.12), for (6.9)–(6.11) will fully determine N_0, N_1 and v, and N_2 can be found using (6.4).

6.1.1.2 Non-dimensionalisation

In order to reduce the number of parameters, and allow relative orders of magnitude to be considered, we non-dimensionalise the model. We choose

$$N_i = \bar{N}_i, \quad x = L\bar{x}, \quad t = t_1\bar{t} \quad \text{and} \quad v = \frac{L}{t_1}\bar{v}, \tag{6.13}$$

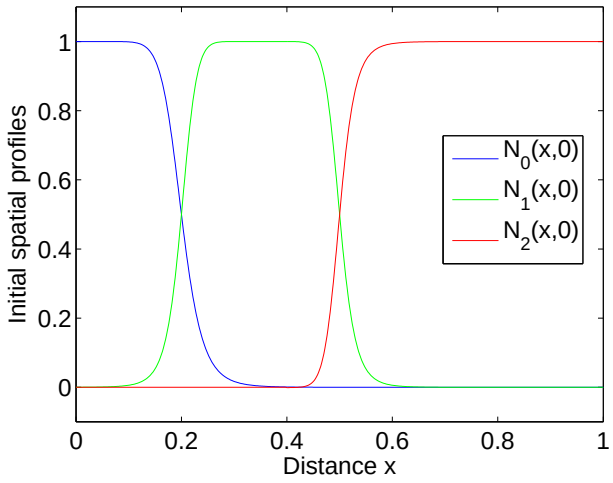

Figure 6.2: Initial spatial profiles for the cell population volume fractions according to (6.12), where $x_s = 0.2$, $x_t = 0.5$, $L = 1$ and $p = 10$.

where L is the crypt length, and t_1 is the semi-differentiated cell cycle time; typically $t_1 = 1/2$ day. We also rescale the rates with t_1,

$$\bar{\alpha}_i = t_1\alpha_i, \quad \bar{\beta}_i = t_1\beta_i, \quad \bar{\gamma} = t_1\gamma. \tag{6.14}$$

Dropping the bars, we get the dimensionless system

$$\frac{\partial N_0}{\partial t} + \frac{\partial}{\partial x}(vN_0) = f(N_0, x), \tag{6.15}$$

$$\frac{\partial N_1}{\partial t} + \frac{\partial}{\partial x}(vN_1) = g(N_0, N_1, x), \tag{6.16}$$

$$\frac{\partial v}{\partial x} = h(N_0, N_1, x), \tag{6.17}$$

where

$$f(N_0, x) = \alpha(x)N_0, \tag{6.18}$$

$$g(N_0, N_1, x) = \beta(x)N_1 + \alpha_2(x)N_0, \tag{6.19}$$

$$h(N_0, N_1, x) = [\alpha(x) + \alpha_2(x)]N_0 + [\beta(x) + \beta_2(x)]N_1 - \gamma(x)\,(1 - N_0 - N_1)\,, \tag{6.20}$$

and $x \in [0, 1]$. The initial spatial profiles of the cell populations are given by

$$N_0(x, 0) = \frac{1}{1 + (x/x_s)^p}, \qquad N_1(x, 0) = \frac{1}{1 + [(2x - x_t - x_s)/(x_t - x_s)]^p}, \tag{6.21}$$

where $x_s = 0.2$ and $x_t = 0.5$. For illustrative purposes, we use the boundary conditions at $x = 0$ to ensure a stem cell niche at the crypt base,

$$N_0(0, t) = 1, \qquad N_1(0, t) = 0 \qquad \text{and} \qquad v(0, t) = 0, \tag{6.22}$$

and not zero flux conditions. Throughout the rest of this chapter we will work with the dimensionless system (6.15)–(6.17) with initial profiles (6.21) (or their equivalent step function profiles) and boundary conditions (6.22).

6.1.2 The method of characteristics

In order to analyse the system (6.15)–(6.17), the equations can be written in the form

$$\mathbf{A}\frac{\partial \boldsymbol{u}}{\partial t} + \mathbf{B}\frac{\partial \boldsymbol{u}}{\partial x} = \mathbf{c}, \tag{6.23}$$

where

$$\boldsymbol{u} = \begin{pmatrix} N_0 \\ N_1 \\ v \end{pmatrix}, \ \mathbf{A} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \ \mathbf{B} = \begin{pmatrix} v & 0 & N_0 \\ 0 & v & N_1 \\ 0 & 0 & 1 \end{pmatrix}, \ \mathbf{c} = \begin{pmatrix} f(N_0, x) \\ g(N_0, N_1, x) \\ h(N_0, N_1, x) \end{pmatrix}. \tag{6.24}$$

Solving $|\mathbf{B} - \lambda\mathbf{A}| = 0$, where $\lambda = \mathrm{d}x/\mathrm{d}t$, the eigenvalues λ_i are found to be

$$\lambda_1 = \lambda_2 = \frac{\mathrm{d}x}{\mathrm{d}t} = v(x,t), \quad \text{and} \quad \lambda_3 = \infty. \tag{6.25}$$

Therefore the system is degenerate hyperbolic and the characteristics are given by $t =$ constant, and by solving $\dot{x} = v(x,t)$. The left eigenvectors for eigenvalues λ_i $(i = 1, 2, 3)$ are, respectively,

$$\boldsymbol{e}_1 = \begin{pmatrix} 1 \\ 0 \\ -N_0 \end{pmatrix}, \quad \boldsymbol{e}_2 = \begin{pmatrix} 0 \\ 1 \\ -N_1 \end{pmatrix} \quad \text{and} \quad \boldsymbol{e}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}. \tag{6.26}$$

Writing down the Pfaffian equations

$$\boldsymbol{e}_i^T \mathbf{A}\frac{\mathrm{d}\boldsymbol{u}}{\mathrm{d}t} = \boldsymbol{e}_i^T \mathbf{c} \tag{6.27}$$

produces the system of equations

$$\frac{\mathrm{d}N_0}{\mathrm{d}t} = f(N_0, x) - N_0 h(N_0, N_1, x), \tag{6.28}$$

$$\frac{\mathrm{d}N_1}{\mathrm{d}t} = g(N_0, N_1, x) - N_1 h(N_0, N_1, x), \tag{6.29}$$

where $\dot{x} = v(x,t)$. This is, of course, equivalent to writing (6.15)–(6.16) in Lagrangian coordinates. Unfortunately, since v is unknown, (6.28)–(6.29) are not especially useful. The Pfaffian for $\lambda_3 = \infty$ corresponds to $\mathrm{d}v/\mathrm{d}t = \infty$ on the characteristic $t =$ constant.

6.1.3 Solution using the method of lines

Given that $t =$ constant is a characteristic of equations (6.15)–(6.17), we can solve the full system computationally by using a method of lines technique, where the time derivatives are discretised and the resultant equations are solved as ODEs in the space variable x. We approximate the time derivatives by

$$\frac{\partial N_i}{\partial t}(x,t) \approx \frac{N_i(x,t) - N_i(x, t-\Delta t)}{\Delta t} + O(\Delta t). \tag{6.30}$$

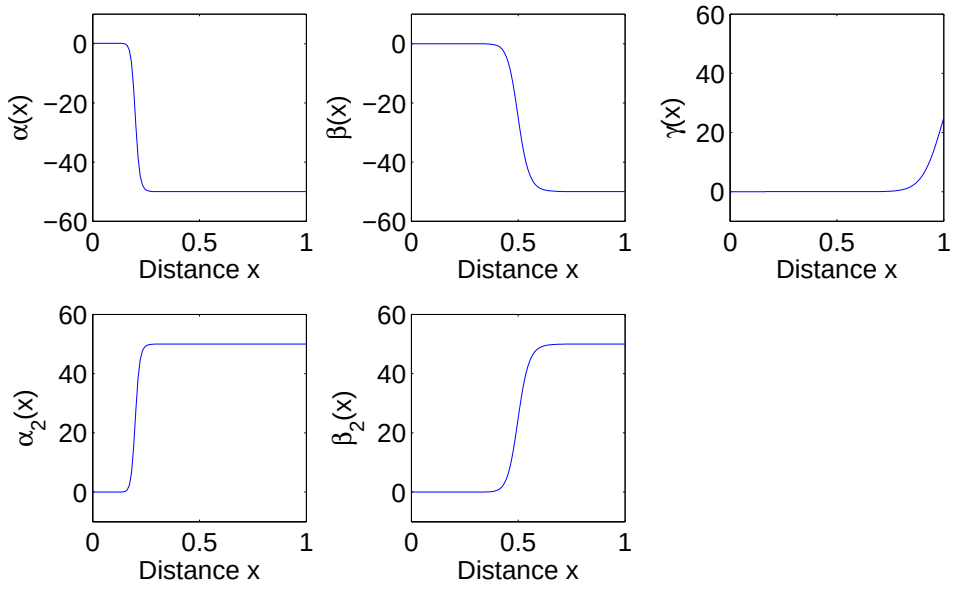

Figure 6.3: Profiles of the spatial dependence of the rate parameters in the population density model, where specific forms for $\alpha(x)$, $\beta(x)$, $\gamma(x)$, $\alpha_2(x)$ and $\beta_2(x)$ are given in (6.34)–(6.35). Stem/transit cells will only be expected to proliferate in the bottom/middle of the crypt, and to differentiate in the middle/top. Fully-differentiated cells should only be removed at the top of the crypt. The parameters were taken to be $\hat{\alpha} = 0.126$, $\hat{\beta} = -0.02$, $m = 50$, $p = 20$, $x_s = 0.2$ and $x_t = 0.5$.

Denoting $N_i(x, t + n\Delta t)$ by N_i^n and $v(x, t + n\Delta t)$ by v^n, (6.15)–(6.17) can be formulated as ODEs

$$\frac{\mathrm{d}N_0^n}{\mathrm{d}x} = \left[f(N_0^n, x) - N_0^n h(N_0^n, N_1^n, x) - \frac{N_0^n - N_0^{n-1}}{\Delta t} \right] \Big/ v^n, \tag{6.31}$$

$$\frac{\mathrm{d}N_1^n}{\mathrm{d}x} = \left[g(N_0^n, N_1^n, x) - N_1^n h(N_0^n, N_1^n, x) - \frac{N_1^n - N_1^{n-1}}{\Delta t} \right] \Big/ v^n, \tag{6.32}$$

$$\frac{\mathrm{d}v^n}{\mathrm{d}x} = h(N_0^n, N_1^n, x). \tag{6.33}$$

The system (6.31)–(6.33) was computed using the Matlab ODE solver ode45, along with boundary conditions (6.22) and initial conditions that were prescribed as step functions, although using (6.21) produced the same outcome. It was necessary to take $v(0, t) = \epsilon \ll 1$ for numerical solution purposes, since v^0 is the denominator in (6.31) and (6.32), and the solutions converged for arbitrarily small ϵ. The rate parameters $\alpha(x)$, $\beta(x)$, $\gamma(x)$, $\alpha_2(x)$ and $\beta_2(x)$ were assumed to vary along the crypt axis according to

$$\alpha(x) = \frac{\hat{\alpha} x_s^p - m x^p}{x^p + x_s^p}, \quad \beta(x) = \frac{\hat{\beta} x_t^p - m x^p}{x^p + x_t^p}, \tag{6.34}$$

$$\alpha_2(x) = \frac{m x^p}{x^p + x_s^p}, \quad \beta_2(x) = \frac{m x^p}{x^p + x_t^p} \quad \text{and} \quad \gamma(x) = \frac{m x^p}{x^p + 1}, \tag{6.35}$$

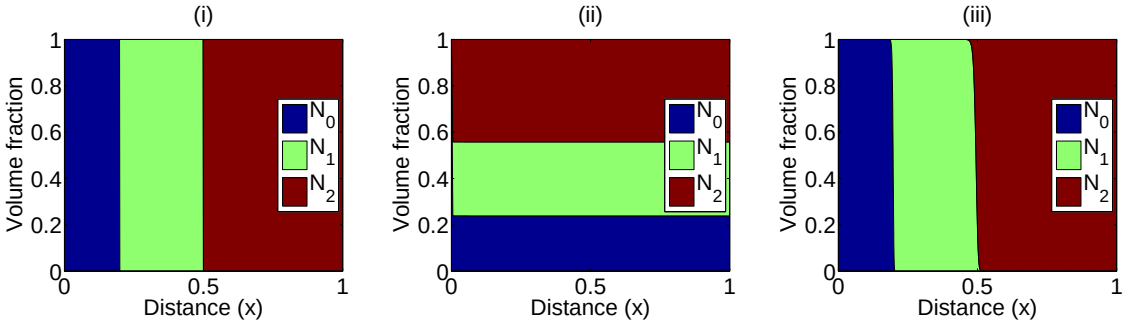

Figure 6.4: Volume fractions of the cell populations along the crypt axis. (i) The initial spatial profiles given by step functions at $x_s = 0.2$ and $x_t = 0.5$. The cell population volume fractions at $t = 100$ days found by solving the system (6.31)–(6.33) are shown when (ii) the rate parameters are kept constant, with values $\alpha = 0.126$, $\beta = -0.02$, $\alpha_2 = 0.195$, $\beta_2 = 0.5$ and $\gamma = 0.235$, and when (iii) the rate parameters vary according to (6.34)–(6.35), where $\hat{\alpha} = 0.126$, $\hat{\beta} = 0.252$, $m = 50$, $p = 100$, $x_s = 0.2$ and $x_t = 0.5$.

where x_s, x_t. $\hat{\alpha}$, $\hat{\beta}$ and m are constants and the exponent $p \gg 1$. The rate parameters (6.34)–(6.35) are plotted in Figure 6.3.

6.1.3.1 Results for constant and varying rate parameters

To illustrate the system behaviour we solved (6.31)–(6.33) numerically for two cases with or without the rate parameter varying along the crypt axis, and the results are shown in Figure 6.4. In panel (i) of Figure 6.4 we show the initial spatial profiles for the cell population volume fractions.

Firstly, we consider the case where each of the rate parameters $\alpha(x)$, $\beta(x)$, $\gamma(x)$, $\alpha_2(x)$ and $\beta_2(x)$ are assumed to be constant in x, and the steady-state cell population volume fractions are plotted in panel (ii) of Figure 6.4. In this case there are no specific positional cues to the cells, and so they spread out across the crypt axis so that there are some stem, semi- and fully-differentiated cells at all positions in the crypt.

Secondly, we consider the case where the rate parameters vary along the crypt axis according to (6.34)–(6.35). This is one mechanism for linking in the chemical cues, such as Wnt signalling, that occur in the crypt, and these rates can be considered to be varying dependent on chemical concentrations [Giles *et al.*, 2003]. Alternatively, these parameters could be made explicit functions of Wnt, *APC* or β-catenin concentrations. We would not expect there to be any stem cells in positions other than the basal region, and so outside the stem cell compartment α is set to be very large and negative so that any stem cells will not remain there, and similarly for β outside the semi-differentiated cell compartment. This has the effect that any misplaced cells will either die or differentiate such that the cell populations will occupy the expected regions. The effect of this spatial feedback is to maintain the initial population structure, and the cells remain in their expected positions along the crypt axis, as we show in panel (iii) of Figure 6.4.

In the next section we discuss other possible mechanisms that could be used to maintain these expected cell population positions in the crypt.

6.1.4 Different feedbacks

In Section 6.1.3 we allowed the rate parameters to vary with spatial position in order to fix the cell populations in the crypt. However, other approaches might be used and in this section we consider two such mechanisms. Chemotaxis might drive the stem cells to the bottom, or the cell populations might each have a different velocity. Chemotaxis would be included in the model by writing

$$\frac{\partial N_0}{\partial t} + \frac{\partial}{\partial x}(vN_0) = f - \frac{\partial}{\partial x}\left(N_0 \frac{\partial C}{\partial x}\right), \tag{6.36}$$

where $\partial C/\partial x$ is the gradient of some chemical concentration (for example Wnt, APC or β-catenin) along the crypt axis, whereas an additional velocity v_0 would be included as

$$\frac{\partial N_0}{\partial t} + \frac{\partial}{\partial x}[(v + v_0)N_0] = f. \tag{6.37}$$

Both (6.36) and (6.37) are in the same form if we think of $v_0 = \partial C/\partial x$. Therefore, this form of feedback could be thought of as either a modified velocity or a chemotaxis term.

We consider the new feedback as extensions of the original system (6.1)–(6.3) in one dimension, which we write into the form

$$\frac{\partial N_0}{\partial t} + \frac{\partial}{\partial x}[(v + v_0)N_0] = \alpha N_0 \qquad = f, \tag{6.38}$$

$$\frac{\partial N_1}{\partial t} + \frac{\partial}{\partial x}(vN_1) = \beta N_1 + \alpha_2 N_0 \quad = g, \tag{6.39}$$

$$\frac{\partial N_2}{\partial t} + \frac{\partial}{\partial x}(vN_2) = -\gamma N_2 + \beta_2 N_1 = h, \tag{6.40}$$

$$N_0 + N_1 + N_2 = 1, \tag{6.41}$$

where v_0 is imposed. We can eliminate N_2 and use the method of lines, as in the previous section, to discretise the time derivatives, and hence write the model as a system of ODEs of the form

$$\frac{\mathrm{d}N_0}{\mathrm{d}x} = \frac{F - N_0 H}{v + v_0(1 - N_0)}, \tag{6.42}$$

$$\frac{\mathrm{d}N_1}{\mathrm{d}x} = \frac{G[v + v_0(1 - N_0)] - HN_1(v + v_0) + Fv_0N_1}{v[v + v_0(1 - N_0)]}, \tag{6.43}$$

$$\frac{\mathrm{d}v}{\mathrm{d}x} = \frac{H(v + v_0) - Fv_0}{v + v_0(1 - N_0)}, \tag{6.44}$$

where

$$F = f - \frac{N_0(x,t) - N_0(x,t - \Delta t)}{\Delta t} - N_0 \frac{\partial v_0}{\partial x}, \tag{6.45}$$

$$G = g - \frac{N_1(x,t) - N_1(x,t - \Delta t)}{\Delta t}, \tag{6.46}$$

$$H = f + g + h - N_0 \frac{\partial v_0}{\partial x}. \tag{6.47}$$

Firstly, when we set $v_0 = 0$ in (6.42)–(6.44), we return to the original equations (6.31)–(6.33). Secondly, we chose $v_0 = -k_1 \tanh(k_2 x)$ to illustrate a velocity profile which directs

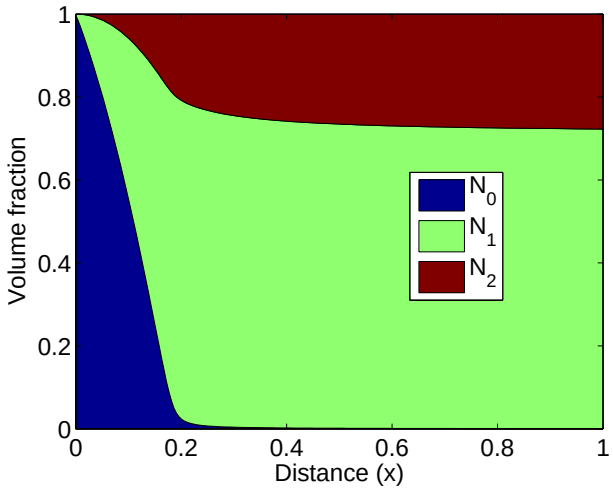

Figure 6.5: Volume fractions of the cell populations along the crypt axis using a model with chemotaxis given by (6.38)–(6.41). The parameters are taken to be $\alpha = 0.01$, $\beta = 2$, $\alpha_2 = 1$, $\beta_2 = 1$ and $\gamma = 1$. Chemotaxis is included using a term $v_0 = -k_1 \tanh(k_2 x)$, where $k_1 = 0.3$ and $k_2 = 500$.

the stem cells to the bottom of the crypt. In this case we see that the stem cells only occupy the base of the crypt, and the semi- and fully-differentiated cells move towards the top, as we show in Figure 6.5. There is no spatial structure to the transit and fully-differentiated cell populations because there are no cues affecting these cells. To achieve that structure, chemotaxis could also be included for the transit cells, or the transit cell rate parameters could be chosen to change at the expected $N_1 : N_2$ boundary, as in Section 6.1.3. A more rigorous model, however, would keep track of both space and age for transit cells, with the division rate as a function of the number of divisions undergone.

We conclude that the configuration of stem cells in the crypt can be maintained by velocity variation as well as proliferation variation.

6.2 Relating the spatial and non-spatial models

Given the results of the population density model in Section 6.1, it is important to verify the relation between the spatial model of the crypt and the (non-spatial) continuous model presented in Chapter 3.

The spatial model (6.1)–(6.3) follows the volume fractions of the cell populations $N_i(\boldsymbol{x}, t)$ as they vary in time (t) and along the crypt axis ($\boldsymbol{x}$). If we integrate out the space-dependence then we should be able to compare these equations with those from the continuous model. We denote the total volume fraction for the cell populations at time t by

$$\hat{N}_i(t) = \int_0^L N_i(x,t)\ \mathrm{d}x, \tag{6.48}$$

and therefore integrating the no-voids condition (6.4) yields

$$\hat{N}_0 + \hat{N}_1 + \hat{N}_2 = L. \tag{6.49}$$

Using a one-dimensional version of the system (6.1)–(6.3), and assuming that the rate parameters are constant, we get

$$\frac{\partial N_0}{\partial t} + \frac{\partial}{\partial x}(vN_0) = (\alpha_3 - \alpha_1 - \alpha_2)N_0, \tag{6.50}$$

$$\frac{\partial N_1}{\partial t} + \frac{\partial}{\partial x}(vN_1) = (\beta_3 - \beta_1 - \beta_2)N_1 + \alpha_2 N_0, \tag{6.51}$$

$$\frac{\partial N_2}{\partial t} + \frac{\partial}{\partial x}(vN_2) = -\gamma N_2 + \beta_2 N_1. \tag{6.52}$$

Integrating these equations with the boundary conditions (6.22) produces

$$\frac{\mathrm{d}\hat{N}_0}{\mathrm{d}t} = (\alpha_3 - \alpha_1 - \alpha_2)\hat{N}_0 - \underbrace{(vN_0)\Big|_{x=L}}_{T_1}, \tag{6.53}$$

$$\frac{\mathrm{d}\hat{N}_1}{\mathrm{d}t} = (\beta_3 - \beta_1 - \beta_2)\hat{N}_1 + \alpha_2\hat{N}_0 - \underbrace{(vN_1)\Big|_{x=L}}_{T_2}, \tag{6.54}$$

$$\frac{\mathrm{d}\hat{N}_2}{\mathrm{d}t} = \beta_2\hat{N}_1 - \gamma\hat{N}_2 - \underbrace{(vN_2)\Big|_{x=L}}_{T_3}. \tag{6.55}$$

Comparing this system with the continuous model equations (3.56)–(3.58) from Chapter 3 given by

$$\frac{\mathrm{d}N_0}{\mathrm{d}t} = (\alpha_3 - \alpha_1 - \alpha_2)N_0, \tag{6.56}$$

$$\frac{\mathrm{d}N_1}{\mathrm{d}t} = (\beta_3 - \beta_1 - \beta_2)N_1 + \alpha_2 N_0, \tag{6.57}$$

$$\frac{\mathrm{d}N_2}{\mathrm{d}t} = \beta_2 N_1 - \gamma N_2, \tag{6.58}$$

we note that the only differences are with the additional terms T_1, T_2 and T_3 which are contributions for the velocity and cell populations at the top of the crypt. The precise forms for T_i can only be determined by solving the system (6.4) and (6.50)–(6.52) directly. The T_i terms represent cell sloughing into the lumen at the top of the crypt. The death rate in the ODE model represents apoptosis plus sloughing, and assumes that both are proportional to the population size, whereas in the spatial model apoptosis and sloughing are separate and the sloughing rate is not proportional to the population size. This formulation means that the spatial model will always maintain the same crypt size.

In summary, apart from the sloughing disparities, the continuous ODE model is a good approximation to the population density model.

6.3 Transformation to a more realistic geometry

Until now in this chapter we have considered spatial models of the crypt in one dimension along the crypt axis. In this section we test this assumption by considering a more realistic crypt geometry, and note how this affects our earlier results. In Section 6.1, we assumed that the crypt can be unwrapped so that there is a density of cells at each crypt height, but

in reality unwrapping would not produce a rectangular sheet due to the tapered end at the bottom of the crypt. Since the stem cells reside in this tapered region, it becomes important to choose a realistic geometry so that the stem cell population is not over-estimated.

Suppose that we transform the system (6.1)–(6.3) into cylindrical polar coordinates $\boldsymbol{r} = \boldsymbol{r}(r, \theta, z)$ given by $\boldsymbol{r} = (r(z)\cos\theta, r(z)\sin\theta, z)$, where $r = r(z)$ defines the shape of the crypt, for $r \geq 0$, $0 \leq \theta < 2\pi$ and $0 \leq z < 1$. Therefore we transform from three-dimensional Cartesian coordinates (x, y, z) to the two-dimensional coordinate system (θ, z) on the surface of the crypt, and the orthogonal coordinate axes are given by

$$\boldsymbol{e_\theta} = (-\sin\theta, \cos\theta, 0) \quad \text{and} \quad \boldsymbol{e_z} = \frac{1}{\sqrt{1 + [r'(z)]^2}} \left(r'(z)\cos\theta, r'(z)\sin\theta, 1\right), \tag{6.59}$$

where $\boldsymbol{e_z}$ is normal to the surface of the crypt. Assuming radial symmetry, and that the tangential components of velocity for each cell population are equal to v, the system (6.1)–(6.3) transforms into

$$\frac{\partial N_0}{\partial t} + \frac{1}{r\sqrt{1 + (r')^2}} \frac{\partial}{\partial z}(rvN_0) = f(N_0, s), \tag{6.60}$$

$$\frac{\partial N_1}{\partial t} + \frac{1}{r\sqrt{1 + (r')^2}} \frac{\partial}{\partial z}(rvN_1) = g(N_0, N_1, s), \tag{6.61}$$

$$\frac{\partial N_2}{\partial t} + \frac{1}{r\sqrt{1 + (r')^2}} \frac{\partial}{\partial z}(rvN_2) = h(N_1, N_2, s), \tag{6.62}$$

in addition to the no-voids condition (6.4), where f, g and h are given by (6.18), (6.19) and (6.20), respectively. The same initial volume fraction profiles (6.21) and boundary conditions (6.22) apply, although now x is replaced by z. We also assume that the rate space-dependence conditions (6.34)–(6.35) apply to the arc length s of the curve and not x.

Let us consider three different crypt shapes. Firstly, if we assume a cylindrical shape with the crypt mapped onto a straight line, as in Section 6.1, then $r(z) =$ constant, $\boldsymbol{e_z} = (0, 0, 1)$ and we return to the original one-dimensional equations (6.50)–(6.52). This is equivalent to a cylindrical crypt with no base surface. Secondly, we suppose that $z(r) = r^2/(1 - r^2)$, or equivalently $r(z) = \sqrt{z/(1 + z)}$. This is similar to the original cylinder but now the crypt base is tapered and resembles a test tube. Finally, we also consider a parabolic crypt where $z(r) = r^2$ (equivalently $r(z) = z^{1/2}$) which spreads out as the crypt increases in height. These crypt shapes are illustrated in Figure 6.6.

The solutions of the system (6.60)–(6.62) in the three different cases of crypt geometry are plotted in Figure 6.6. These results show that the stem and transit cells do not extend as far up the vertical crypt axis with the test tube and parabola shapes as they do with the cylinder, when the chemical cues remain at the same arc lengths. Therefore, the cylindrical geometry over-estimates the number of stem cells compared to the test tube or parabola shapes. The results for the test tube and parabola are similar, suggesting that it is the tapered base that is important, and not the shape near the top of the crypt.

In conclusion, it is clear that the geometry of the crypt is crucial in determining realistic cell population dynamics, where the simple cylinder approximation over-estimates the number of cells at the bottom of the crypt.

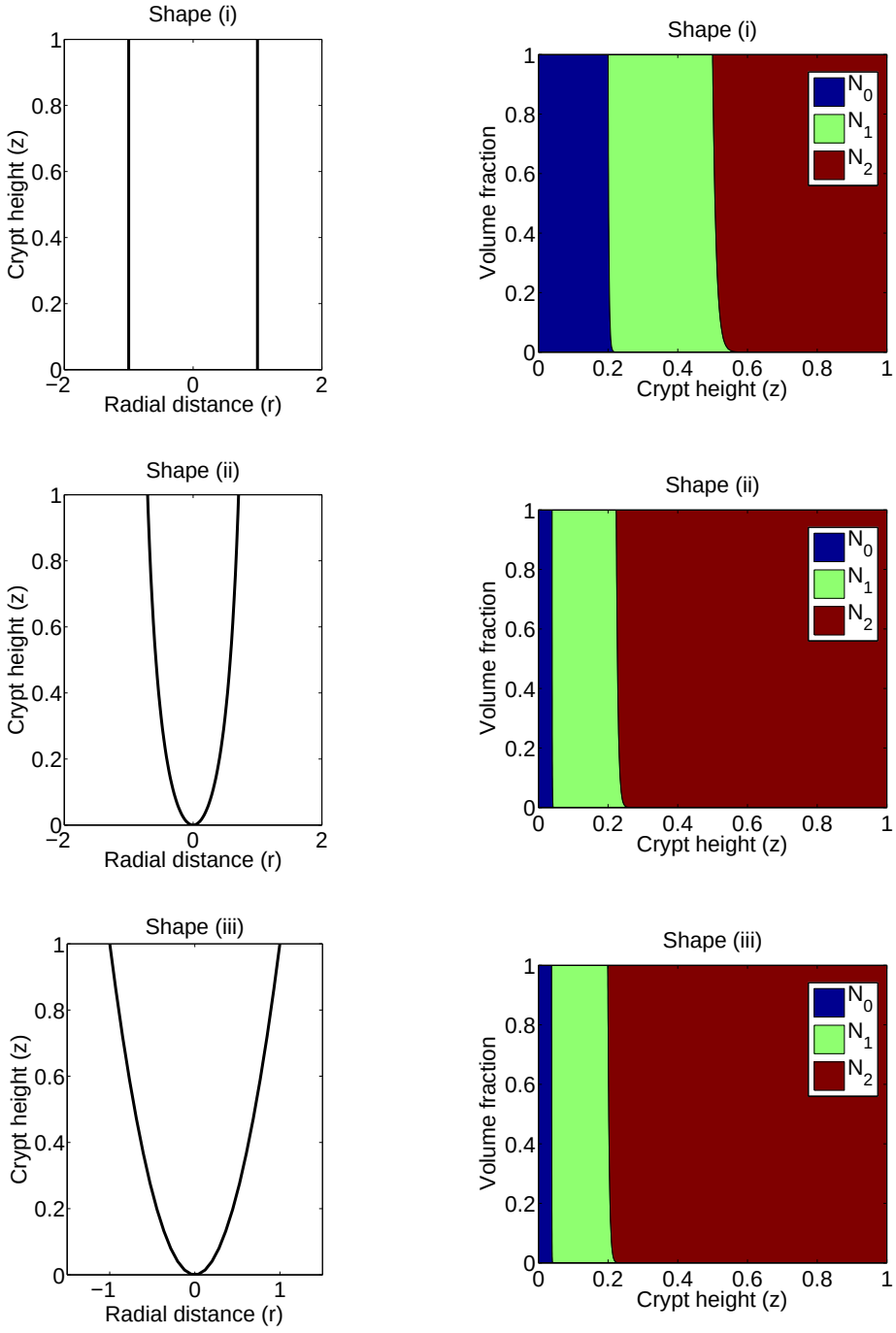

Figure 6.6: The spatial model with different crypt shapes. Left: Illustrations of the different crypt shapes (i) $r(z) =$ constant, (ii) $r(z) = \sqrt{z/(1+z)}$ and (iii) $r(z) = \sqrt{z}$. Right: Solutions of the spatial model (6.60)–(6.62) for the cases (i), (ii) and (iii) using (6.4) with initial volume fraction profiles (6.21) and boundary conditions (6.22), where now x is replaced by z, and rate space-dependence (6.34)–(6.35), where now x is replaced by arc length s (see text). Parameter values as in Figure 6.4 (iii).

6.4 Cell population competition models

Finally, in this section we use the spatial model to track the competition between different cell populations as they compete for positions in space. In Section 6.4.1 we follow cells with a mutation as they spread in order to determine whether the mutants take over the crypt, die out, or coexist with healthy cells. In Section 6.4.2 we present a model of the stem cell niche in order to calculate a time scale for clonal domination of the niche.

6.4.1 Tracking mutant cell clones in the crypt

Suppose that we introduce a mutation into the population density model described in Section 6.1 in order to follow how a mutant population spreads. We assume, as a first approximation, that the mutant cells have identical properties to the healthy cells after a single mutation (which is thought to be the case for the first mutation in APC), and are subject to the same spatial cues as the healthy cells through the rate parameters. Therefore we will have two distinct systems, each of three equations, for the stem, semi- and fully-differentiated cell populations for both healthy and mutant cells, which are linked by a no-voids condition for the sum of all the populations. Denote the healthy cell populations by N_i, as before, and their mutant counterparts by M_i. The coupled system of equations is

$$\frac{\partial N_0}{\partial t} + \frac{\partial}{\partial x}(vN_0) = F(N_0, x), \tag{6.63}$$
$$\frac{\partial N_1}{\partial t} + \frac{\partial}{\partial x}(vN_1) = G(N_0, N_1, x), \tag{6.64}$$
$$\frac{\partial N_2}{\partial t} + \frac{\partial}{\partial x}(vN_2) = H(N_1, N_2, x), \tag{6.65}$$
$$\frac{\partial M_0}{\partial t} + \frac{\partial}{\partial x}(vM_0) = F(M_0, x), \tag{6.66}$$
$$\frac{\partial M_1}{\partial t} + \frac{\partial}{\partial x}(vM_1) = G(M_0, M_1, x), \tag{6.67}$$
$$\frac{\partial M_2}{\partial t} + \frac{\partial}{\partial x}(vM_2) = H(M_1, M_2, x), \tag{6.68}$$

where

$$F(U, x) = \alpha(x)U, \tag{6.69}$$
$$G(U, V, x) = \beta(x)V + \alpha_2(x)U, \tag{6.70}$$
$$H(U, V, x) = \beta_2(x)U - \gamma(x)V, \tag{6.71}$$

and the no-voids condition is

$$N_0 + N_1 + N_2 + M_0 + M_1 + M_2 = 1. \tag{6.72}$$

In this example we assume that a mutation occurs in the stem cell region, but not at the crypt base. We take boundary conditions at $x = 0$ to be

$$N_0(0,t) = 1, \quad N_1(0,t) = N_2(0,t) = M_0(0,t) = M_1(0,t) = M_2(0,t) = 0, \quad v(0,t) = 0, \tag{6.73}$$

and initial spatial profiles of the cell populations to be

$$N_0(x,0) = \begin{cases} 1, & 0 < x < x_0, \\ 0, & \text{otherwise}, \end{cases} \qquad M_0(x,0) = \begin{cases} 1, & x_0 < x < x_s, \\ 0, & \text{otherwise}, \end{cases} \tag{6.74}$$

$$N_1(x,0) = \begin{cases} 1, & x_s < x < x_t, \\ 0, & \text{otherwise}, \end{cases} \qquad M_1(x,0) = 0, \quad \forall x, \tag{6.75}$$

$$N_2(x,0) = \begin{cases} 1, & x_t < x < 1, \\ 0, & \text{otherwise}, \end{cases} \qquad M_2(x,0) = 0, \quad \forall x, \tag{6.76}$$

where $0 \leq x_0 \leq x_s \leq x_t \leq 1$.

This model was solved by the same method which was described in Section 6.1, and two cases of initial and boundary conditions were chosen. Firstly, for the conditions (6.73) and (6.74)–(6.76), the solution is plotted in Figure 6.7. This shows that the mutants are wiped out within a few cell cycles, and this result is repeated if we move the location of the mutant stem cell or alternatively choose the mutation to occur in a transit cell. It is only possible for mutant cells to reach fixation in the crypt if the mutation occurs in the stem cell at the bottom, which requires swapping the initial and boundary conditions for N_0 and M_0 to give

$$N_0(0,t) = 0, \quad M_0(0,t) = 1, \tag{6.77}$$

$$N_0(x,0) = \begin{cases} 1, & x_0 < x < x_s, \\ 0, & \text{otherwise}, \end{cases} \qquad M_0(x,0) = \begin{cases} 1, & 0 < x < x_0, \\ 0, & \text{otherwise}, \end{cases} \tag{6.78}$$

and the other initial and boundary conditions from (6.73)–(6.76) remain the same. In this case the mutant cells completely take over the crypt, and the solution is plotted in Figure 6.8.

In conclusion, this one-dimensional spatial model of the crypt is insufficient for modelling the spread of mutant clones, because the only possibility that mutants will reach fixation is if the mutation occurs in the bottom cell in the crypt, and in this case fixation will occur within a few cell divisions. We have assumed that the mutation does not confer a selective advantage to the cell, but this does not affect the outcome of whether a mutant will take over the crypt or not. The result would only be altered if the mutant cell was subject to a strong chemoattractant signal directing it towards the crypt base. A more sophisticated model is required to capture the complicated interactions of cells, most probably in two or more dimensions, and almost certainly with a stochastic element attached to each cell division.

6.4.2 Niche model of competing stem cells

In this section we attempt to use the spatial model from Section 6.1 as a basis for addressing questions regarding the clonality of a stem cell niche, which was discussed in Chapter 1. We recall that the stem cell niche hypothesis implies that a cell is given signals instructing it to be of a certain type dependent on its position in the crypt. We intend to determine both the likelihood of one cell line becoming fixated in the stem cell compartment (which leads to the whole crypt becoming monoclonal) and, if this does occur, the time scale for monoclonality.

We shall use a reduced version of the model from Section 6.1, where we only consider the stem cell compartment. Since there only a small number of cells in the niche, we consider

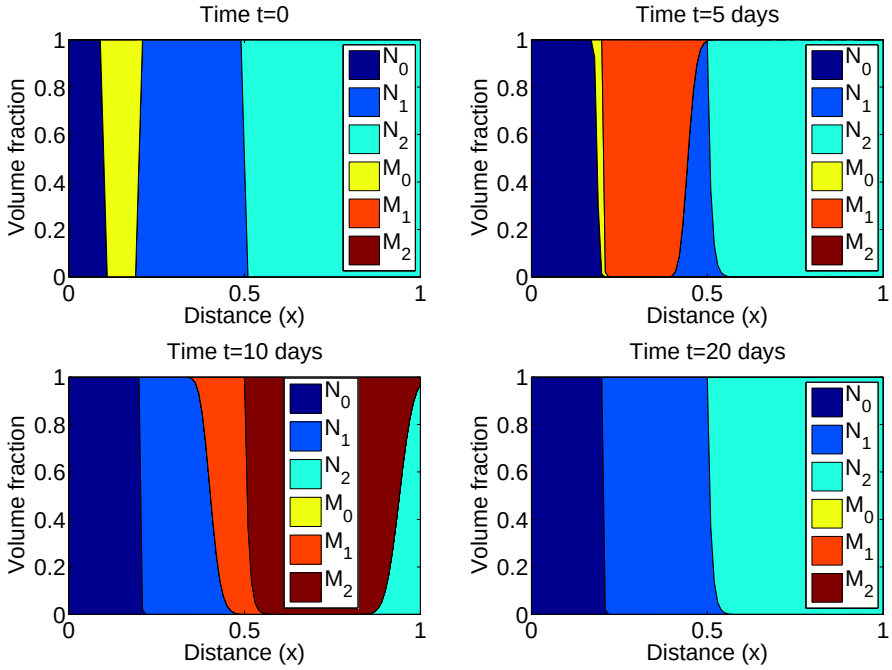

Figure 6.7: A mutation that does not reach fixation in the crypt. This is found by solving the system (6.63)–(6.68) and (6.72)–(6.76), where $x_0 = 0.1$, $x_s = 0.2$ and $x_t = 0.5$ and the other parameters are the same as in Figure 6.4 (iii).

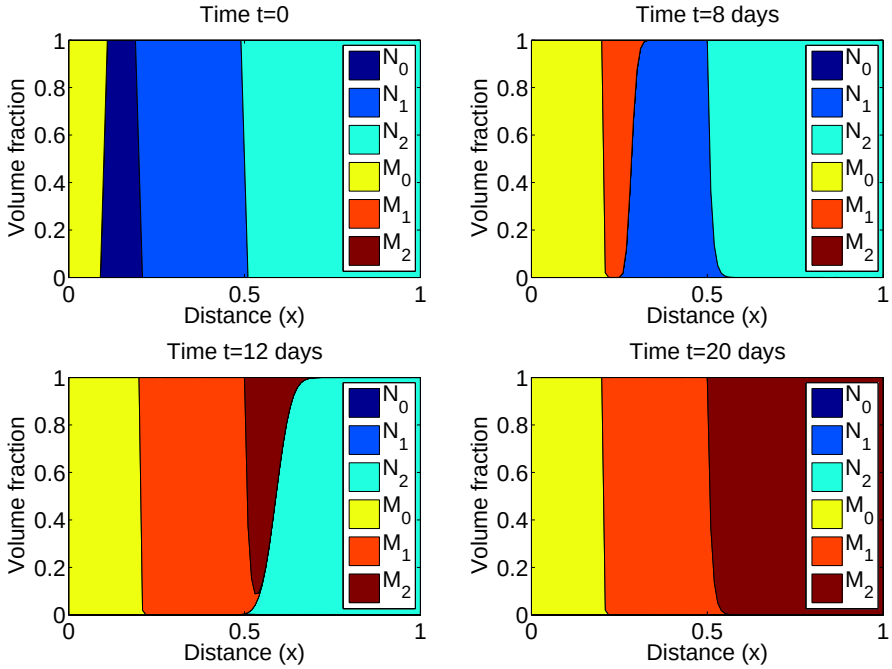

Figure 6.8: A mutation that reaches fixation in the crypt. This is found by solving the system (6.63)–(6.68), (6.72) and (6.77)–(6.78), where $x_0 = 0.1$, $x_s = 0.2$ and $x_t = 0.5$ and the other parameters are the same as in Figure 6.4 (iii).

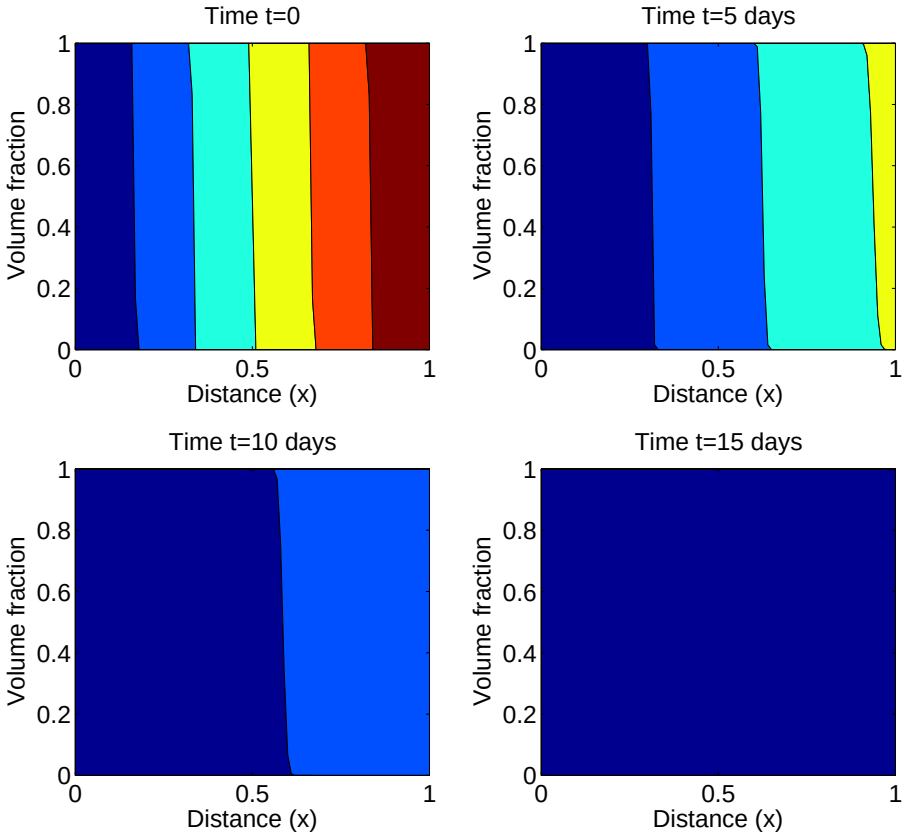

Figure 6.9: Volume fractions of stem cells in the niche, obtained by solving (6.79)–(6.82) using $\alpha = 0.126$, after $t = 0, 5, 10$ and 15 days. Each colour represents a different cell and its progeny.

the probability of a particular cell line occupying a point in space, instead of using volume fractions at each position. Previously we assumed that the feedback resulted in the rate parameters being roughly constant throughout the expected stem cell region, so for the purposes of this section we take the parameter α to be constant.

Let us suppose that there are six stem cells in the stem cell niche of a murine small intestinal crypt (ignoring the presence of Paneth cells), and we consider the crypt in one dimension, as before. Suppose that $S_i(x,t)$ denotes the probability of cell i, and its descendants, occupying position x at time t in the dimensionless region $x \in [0,1]$. If we assume that cells divide proportionate to their probabilities, then each $S_i(x,t)$ satisfies the equations

$$\frac{\partial S_i}{\partial t} + \frac{\partial}{\partial x}(vS_i) = \alpha S_i, \quad i = 1 \ldots 6, \tag{6.79}$$

$$S_1 + S_2 + S_3 + S_4 + S_5 + S_6 = 1, \tag{6.80}$$

where $\alpha = \alpha_3 - \alpha_1 - \alpha_2$, and (6.80) is the no-voids condition. We assume that the six stem cells are initially lined up along the crypt wall in numerical order, and so

$$S_i(x,0) = \begin{cases} 1, & (i-1)/6 < x < i/6, \\ 0, & \text{otherwise.} \end{cases} \tag{6.81}$$

This initial configuration of the probabilities of cell placements is shown in the first panel of Figure 6.9. There is one cell which is at the bottom of the crypt, and the boundary conditions, ensuring that this cell remains there, are given by

$$S_1(0,t) = 1, \quad S_2(0,t) = S_3(0,t) = S_4(0,t) = S_5(0,t) = S_6(0,t) = 0. \tag{6.82}$$

The solution to the model (6.79)–(6.82) is plotted in Figure 6.9, and the numerics give good agreement with the analytic solution, which is given by

$$S_i(x,t) = \begin{cases} 1, & (i-1)/6 < xe^{-\alpha t} < i/6, \\ 0, & \text{otherwise.} \end{cases} \tag{6.83}$$

This shows that the progeny of the one cell fixed at the bottom of the crypt will take over the stem cell niche within a small number of divisions, and the dominant cell line will always be the one occupying the bottom position on the crypt axis.

On a similar note to the results obtained in Section 6.4.1, this one-dimensional spatial model is inadequate for modelling niche succession and a stochastic or probabilistic model will be required for this purpose.

6.5 Discussion

In this chapter we have introduced a spatial model of the crypt in order to try to address questions regarding the stem cell niche hypothesis, tracking the spread of mutant cell populations and to determine the mechanisms underlying the spatial positioning of different cell populations.

This model is a spatial extension of the continuous model that was introduced in Chapter 3, and uses advection equations linked by a no-voids condition to impose the velocity in order to follow the cell population volume fractions as they vary with space and time. Given the restriction to one spatial dimension there is no need to consider constitutive relations such as a Darcy law, which would be difficult to validate biologically. In order to maintain the cell population spatial ordering, we allowed the rate parameters in the model to vary with spatial position. As a simple first approximation, considering only variation in one dimension along the crypt axis, the model was solved using the method of lines and the solutions were shown in Section 6.1. In Section 6.1.4 we considered a different form of feedback by using a chemotactic velocity of stem cells towards the bottom of the crypt to fix the stem cells at the base. However this was not sufficient to explain the spatial structure of the transit and fully-differentiated cell populations, and so we would require additional spatial cues to capture the expected crypt structure. In Section 6.2 we showed that integrating the spatial model over space returns the original continuous model, where the death rate is modified to incorporate loss due to sloughing at the top of the crypt in addition to actual death.

In Section 6.3 we considered three different crypt geometries: the cylinder, the test tube and the parabola. We found that the simplest model in one dimension, representing the cylinder, over-estimated the number of cells in the lower regions of the crypt, which is a significant failing given that this is the location of the stem cells which control crypt dynamics. We have shown that determining a realistic tapered base shape for the crypt is an important feature which cannot be neglected.

As a general theme throughout the chapter, we considered whether this simple modelling framework was sufficient for addressing key biological questions. Two such questions were:

following a mutation, how long does it take for the mutant clone to either be wiped out or take over the crypt; and what is the probability that the progeny of one cell will take over a stem cell niche? Clonal domination, if it exists, is a subject of speculation and we would like to determine the time scale at which it occurs. Our results have shown that a simple spatial model in one dimension is not sufficient to address these questions because the cell at the bottom of the crypt will always dictate the behaviour. In the mutant cell model, if the mutation is introduced to any cell other than the one at the base of the crypt, then it will be swept out within a few cell divisions. Alternatively, if the mutant cell appears at the bottom of the crypt, then within a few divisions the entire crypt will contain cells with that mutation. Similarly, in the niche model, the cell at the bottom of the crypt will always take over the stem cell compartment. A model in two or three dimensions is necessary to accurately capture the observed behaviour, and the model must also contain a stochastic element. The Integrative Biology Project[1] has developed a simulation model which can be used for this purpose, and this is explained in more detail in Chapter 7.

In conclusion, this spatial model has verified the continuous model and can be used to help explain simple observed behaviour such as the mechanisms determining the spatial ordering of the cell populations, but is insufficient for addressing issues regarding competition between different cell populations. Once again, in trying to address these questions we see how important stochastic models are, and we have not used this approach either here or in Chapter 4.

[1]See http://www.integrativebiology.ox.ac.uk/ for details of the Integrative Biology Project.

Chapter 7

Discussion and conclusions

In this final chapter we summarise the results of the thesis in Section 7.1, and discuss advantages and limitations to our approach in Section 7.1.1. In Section 7.2 we suggest possible further work for modelling and experiments, and we conclude in Section 7.3.

7.1 Thesis summary

Colorectal cancer is a well-characterised disease, and the prognosis for long-term survival is good if the cancer is diagnosed early. Consequently much experimental and modelling work is directed towards understanding the initial steps of tumourigenesis, which is what we have focused on in this thesis. Copious biological data describe the well-ordered homeostatic renewal cycle in the crypt, and the formation of polyps and adenomas, but it is unclear what mechanisms maintain homeostasis and how they are disrupted during tumourigenesis. In this thesis we have attempted to use mathematical modelling to explain these observations.

A first step towards this was taken by Tomlinson and Bodmer [1995], who proposed a cell population compartment model to predict that the failure of apoptosis or differentiation leads to tumour growth. This model only holds in very special cases of synchronous cell division, when the stem cell cycle time is an integer multiple of the transit cell cycle time. In addition, the model exhibits structural instability whereby small parameter changes lead to exponential growth or decay in the cell populations, which is biologically unrealistic. We have significantly extended the Tomlinson and Bodmer model to capture asynchronous cell division for any cell cycle times, and to incorporate feedback control into the system to stabilise the models to infinitesimal parameter changes.

In order to describe crypt dynamics in a healthy colon, we found that an age-structured model is suitable to follow specific variations at times on the order of the cell cycle, whereas a continuous model is simpler and more appropriate when a population changes over a time scale of many cell cycles. In order to capture both homeostasis and tumourigenesis in the same model, we presented two alternative types of feedback. A linear feedback model, which maintains a steady state regardless of any parameter changes, could be used to describe the stem cell population, which is believed to be tightly regulated in numbers. Alternatively, a saturating feedback model can describe transit cell regulation, which captures homeostasis but also allows these cells to develop into cancer-driving cells and exhibit unbounded population expansion at high growth rates. Different combinations of the feedbacks can lead to different types of system behaviour, with cancer-driving cells deriving from both stem and transit cells.

Our model predicts that a non-neutral stem cell mutation will either be driven to extinction or take over the stem cell compartment, meaning that there can be no coexistence of healthy and mutant cells. In addition, the saturating feedback model can be used to explain the long lag phases observed in carcinogenesis, which occur between periods of rapid tumour growth, before the cancer finally takes over. Significantly, we have shown that, contrary to general belief, the proportion of cancer-driving cells (thought to be transit cells) in the exponential growth phase of a tumour may vary depending on tumour type.

Komarova and Wang [2004] proposed a model that challenged the widely-held belief that only stem cells were the target of the two mutations in the *APC* gene, by predicting that both mutations could occur in transit cells if the mutation rate was sufficiently high. The authors made several simplifying assumptions in order to reach their conclusions and we tested their validity using stochastic simulation and analytic models. Our results showed that Komarova and Wang overestimated the likelihood of both mutations occurring in transit cells, verifying the original belief that the first mutation is most likely to occur in a stem cell.

These models have assumed a well-mixed system of cells in the crypt, and so we also developed a model of the cell population volume fractions at each point along the crypt axis in order to capture spatial variations of the populations. We demonstrated that variable proliferation rates or cell velocities were sufficient to capture the expected cell population configurations in the crypt, particularly that the stem cells are located at the base. We also used this simplified description of a one-dimensional spatial model to show that our ordinary differential equation model is valid. Our model revealed that the simplest choice of a one-dimensional model, representing the curved side of a cylinder, overestimates the number of stem cells in a crypt compared to models with tapered crypt bases. Finally, we concluded that our one-dimensional spatial model was not sufficient to determine the time scale of clonal dominance in a competition between healthy and mutant populations, or between different stem cells within a niche. To properly address this question would require a fully two- (or three-) dimensional model in which stochastic effects were incorporated.

As a result of our work, the following experimental predictions are made. We predict that changes in any of the death, differentiation or population renewal rates could lead to exponential cell population growth, and that the key model parameters are the net per-capita growth rates of stem and transit cells and the speed of feedback response.

7.1.1 Advantages and limitations of our approach

An advantage of the mathematical modelling approach is that it can be used to make predictions about system behaviour which would otherwise not be considered. For example, through verbal reasoning a linear model (such as Tomlinson and Bodmer [1995] and the models in Chapter 3) appears to be realistic, but mathematics is used to analyse the model in a rigorous context, which illuminates the structural instability in the model and the necessity for system feedback.

Given the uncertainty about specific parameter values in the literature, a strength of the deterministic approach that we have adopted is that we can capture experimentally observed behaviour for a range of parameter values, which is not easily possible with an individual cell-based model.

Our model allows for different interpretations of what it means to be a stem or a transit cell, and does not prove or disprove the existence of stem cells. An alternative view of "stemness" is that it is not an intrinsic property of a cell, but it is the feedback that causes

cells to behave in certain ways.

In Chapter 3 we showed that the long-term behaviour of the continuous model is comparable with that of an age-structured model. However, a criticism of our modelling approach is that the small number of cells in the crypt could invalidate the continuum assumption, and it is not appropriate when considering stochastic processes such as cell division or mutation. Individual cell-based models, such as the CHASTE model produced by the Integrative Biology Project, could be thought of as a natural way to include stochasticity. Our models are intended to qualitatively reproduce experimentally observed crypt dynamics, and complement individual cell-based models. While the latter are more appropriate for a quantitative analysis, the former can provide important insights into the effects of certain processes and parameters.

7.2 Further work

There are many possible directions in which this work can be extended. Firstly, we pose some additional modelling questions that could be addressed, and secondly we discuss experiments that would elucidate the results from our work and in turn initiate further modelling opportunities.

7.2.1 Modelling directions

In addition to the comparison discussed in Chapter 3, another method to verify the continuum model is to include stochastic variation of the cell cycle times in the age-structured model, which will lead to a distribution of cell ages and reduce to the continuous model [Gaffney, 2004].

The continuous model could be extended in several ways. One method would be to separate the transit and fully-differentiated cell populations into sub-compartments. This method would allow for the three different functional cells (colonocytes, Goblet and enteroendocrine cells) to be incorporated, but it would also produce a more complex system of equations with more parameters. For this it would be necessary to have a clearer understanding of each differentiated cell type.

In Chapter 4 we assumed that feedback acted through cell populations regulating their own size, but it is unclear to what extent our results would vary if the feedback was chosen to act between different populations: namely, fully-differentiated cells could influence the stem or transit cells, or the feedback could be dependent on the total crypt size. Would other functional forms of feedback produce the same conclusion that a mutation results in either the healthy or mutant populations being driven to extinction unless the mutation is neutral? De-differentiation from transit to stem cells could also be included, which would allow for the possibility of the stem cells being replenished if the entire stem cell population was removed. However, de-differentiation is thought to be unlikely, and a more likely outcome of stem cell removal would be that the crypt would die and be replaced by crypt fission (W. F. Bodmer, pers. comm.).

In Chapter 6 we showed that our one-dimensional deterministic spatial model was insufficient to describe the competition between different cell lines, and there are several extensions to this model which would allow it to capture more complex behaviour. A stochastic individual cell-based model is required to capture the variability in the niche, both in terms of symmetric and asymmetric divisions, and in the positioning of daughter cells after division. Including some type of transport process, or relaxing the assumption that a stem cell

is fixed at the bottom of the crypt, would not make the dominance of the lowest cell an inevitability. Also, if the model were extended to two or three dimensions, possibly using cylindrical or ellipsoidal polar coordinates, it would allow us to track spatially the spread of clones as they divide and determine the lateral dispersion of each clonal population, which would more accurately capture the competition between different cell lines.

Our argument, in Chapter 4, that mutations lead to an increased compartment size could be tested against experimental data, for example using data from the studies by Araki *et al.* [1995] and Batman *et al.* [2007]. They observed that mutated crypts increased in length, and so the percentage increase could be assumed to correlate to a percentage population increase after the mutation. This would make it possible to estimate the effect that mutations have on changing the rate parameters.

Modelling can be used to calculate the expected number of mutations that are likely to accumulate within a cell before the cancer reaches a critical size (say 10^{12} cells). Most of these mutations would occur before the exponential growth phase. Key questions to address are: how many mutations that are neutral could accumulate in the cell population, and to what extent does a slight selective advantage help mutations reach fixation in the crypt?

This thesis is part of a wider collaboration within the Integrative Biology Project[1] to describe the initiation of colorectal cancer using multi-scale modelling on different spatio-temporal scales. In particular, the CHASTE computer simulation environment can capture individual cell behaviour within the crypt. Our work is complementary to that of CHASTE, and can be used to predict parameter values which it is harder to achieve from an individual cell-based model. An important next step will be to integrate our work with that on the sub-cellular and tissue scales of modelling [van Leeuwen *et al.*, 2008]. In particular, our model approach can provide rate parameters that can be incorporated into the model of crypt budding and fission by Edwards and Chapman [2007]. The concentrations of APC and β-catenin vary along the crypt axis, and so their concentrations could be found from a sub-cellular model and incorporated into our spatial model in order to estimate the parameters. One of the problems with analysing one component of a network is that we do not know how the others are connected, and one of the future challenges is to determine and incorporate all of the important components into the system.

7.2.2 Possible experiments

Several features of our models could be illuminated or determined if certain experiments were undertaken. In order to verify our model predictions it is imperative to find the steady state parameters for the cell renewal system in the crypt from Chapter 3: specifically the death, differentiation and renewal rates, cell cycle times and numbers of cells for the stem and transit cell populations, and the fully-differentiated cell population removal rate. It is possible that the Clevers group [Barker *et al.*, 2007] will be able to determine these with the new marker, GPR49. We would also like to determine how these parameters are altered by mutations, and where these mutations occur in the crypt. If this was possible we could determine whether cancer-driving cells originate from stem or transit cells in order to inform our models from Chapter 4.

An experiment to test whether differentiation is dependent on the number of divisions a cell has undergone, or on spatial cues, is to move a transit cell from a position near the bottom of the crypt to the top. If that cell terminally differentiates immediately then differentiation can be concluded to be dependent on local chemical cues, whereas if it continues

[1] See http://www.integrativebiology.ox.ac.uk/ for details of the Integrative Biology Project.

to divide then it must require a certain number of divisions before terminally differentiating. This would be difficult to perform *in vivo* (A. Altaba, pers. comm.), but it is hoped that techniques will be developed within the next few years to enable such experiments to be carried out.

Homeostasis could be illuminated if it were possible to disturb a stem cell from the crypt base and then record the cell population sizes as they return to steady state. Experiments to perturb the steady state are not currently possible (A. Altaba, pers. comm.), but it is hoped that this will be achievable shortly. Similarly, if a stem cell was moved from the crypt base to the apex, would it move back down to the base or differentiate? This would illustrate the actions of EphB/EphrinB on cell positioning in the crypt.

Shih *et al.* [2001] observed that within a column of epithelial cells in mutant crypts there is an abrupt change between dysplastic cells in the upper regions and morphologically normal cells at the bottom of the crypt, which were genetically unrelated to the cells above them. Using these observations, experiments could be performed to test the allelic state of the cells beneath this junction in order to determine where the mutations occurred and to validate our results from Chapter 5. Under our proposed experiment, if both copies of the APC gene in the displaced cell are mutated, then both mutations would have occurred in stem cells; if one copy of the APC gene is mutated, then the first mutation would have occurred in a stem cell and the second in a transit cell; and if neither copy of the APC gene is mutated, then both mutations would have struck in transit cells. It is important to note that numerous properties of the crypts depend a substantial amount on the stage of the carcinogenesis process at which the experiments are being conducted, and the experiments suggested here are only suitable for early stage adenomas.

7.3 Final words

The most important results from this thesis are the similarity between the age-structured and continuous models and their relevance at different time scales; the use of linear and saturating feedback to model homeostasis and tumourigenesis in the crypt; the realisation of the importance of including death, appropriate cell cycle times and compartment sizes in a model for two mutational hits; and the insufficiency of a one-dimensional deterministic spatial model in capturing the competition between clonal populations. We hope that further modelling developments will inspire experiments to validate and then extend these models. As modelling and experiments become ever more intertwined, and increasing levels of system complexity are found, we believe that mathematical modelling will play a significant part in future research developments.

Appendices

Appendix A

Table of data

Parameter	Data	Species[1]	Reference
Stem cell numbers (N_0)	Max. 16	Mouse[2]	Loeffler *et al.* [1986]
	16	Mouse[2]	Potten and Loeffler [1987]
	4–16	Mouse	Potten and Loeffler [1990]
	16	Mouse[2]	Paulus *et al.* [1992]
	4–8	Mouse	Paulus *et al.* [1993]
	4	Mouse	Li *et al.* [1994]
	4–6	Mouse[2]	Loeffler *et al.* [1997]
	11–16	Mouse[2]	Boman *et al.* [2001]
	16	Mouse[2]	Meineke *et al.* [2001]
	1–40	Mouse	Yatabe *et al.* [2001]
	4–6	Mouse[2]	Marshman *et al.* [2002]
	4–6	Mouse[2]	Nowak *et al.* [2002]
	4–6	Mouse	Potten *et al.* [2003]
	4–6	Mouse	Moore and Lemischka [2006]
	4–6	Mouse	Barker *et al.* [2007]
	6 ± 2	Mouse	H. Clevers, pers. comm.
	1–40	Human	Yatabe *et al.* [2001]
	1–10	Human	Michor *et al.* [2004b]
	1–10	Human	W. F. Bodmer, pers. comm.
Transit cell numbers (N_1)	200–500	Human	W. F. Bodmer, pers. comm.
Differentiated cell numbers (N_2)	500–800	Human	W. F. Bodmer, pers. comm.

[1]Data refer to the murine small intestine or the human large intestine.
[2]Model parameter for simulations.

Parameter	Data	Species[1]	Reference
Total proliferating cells	150–160	Mouse	Potten and Loeffler [1990]
$(N_0 + N_1)$	150–160	Mouse	Li *et al.* [1994]
Total cells in a crypt	235–250	Mouse	Potten and Loeffler [1990]
$(N_0 + N_1 + N_2)$	250	Mouse	Li *et al.* [1994]
	300	Mouse	Drasdo and Loeffler [2001]
	235–250	Mouse	Meineke *et al.* [2001]
	2048	Human[2]	Yatabe *et al.* [2001]
	2000	Human	Kim and Shibata [2002]
	1000–4000	Human	Nowak *et al.* [2002]
	2000	Human[2]	Komarova and Wang [2004]
	1000–4000	Human	Michor *et al.* [2004b]
	1000	Human	W. F. Bodmer, pers. comm.
Stem cell cycle time	Av. 16	Mouse[2]	Loeffler *et al.* [1986]
(hours) (t_0)	16	Mouse[2]	Potten and Loeffler [1987]
	Av. 24	Mouse	Potten and Loeffler [1990]
	24	Mouse	Paulus *et al.* [1992]
	12–20	Mouse	Paulus *et al.* [1993]
	12–32	Mouse	Li *et al.* [1994]
	8–24	Mouse[2]	Loeffler *et al.* [1997]
	Av. 24	Mouse	Potten *et al.* [2003]
	24	Mouse	Barker *et al.* [2007]
	24	Mouse	H. Clevers, pers. comm.
	88	Human[2]	Tomlinson *et al.* [2002]
	168	Human	Potten *et al.* [2003]
	72	Human	W. F. Bodmer, pers. comm.
Transit cell cycle time	10–14	Mouse[2]	Loeffler *et al.* [1986]
(hours) (t_1)	11–12	Mouse[2]	Potten and Loeffler [1987]
	12	Mouse	Paulus *et al.* [1992]
	11–12	Mouse	Paulus *et al.* [1993]
	12–13	Mouse	Bach *et al.* [2000]
	12	Mouse	H. Clevers, pers. comm.
Velocity at crypt apex	300	Mouse	Potten and Loeffler [1990]
(cells/day)	1000/3	Human	W. F. Bodmer, pers. comm.

[1]Data refer to the murine small intestine or the human large intestine.
[2]Model parameter for simulations.

Parameter	Data	Species[1]	Reference[2]
Epithelium renewal time (Crypt residence time) (days)	2–3	Mouse	Tomlinson and Bodmer [1999]
	1–2	Mouse	Drasdo and Loeffler [2001]
	2–3	Mouse	Okamoto and Watanabe [2004]
	2	Mouse	H. Clevers, pers. comm.
	6	Human	Ro and Rannala [2001]
	5–7	Human	Marshman *et al.* [2002]
	3–6	Human	Nowak *et al.* [2002]
	4–6	Human	Ross *et al.* [2003]
	5	Human	Blanpain *et al.* [2007]
	3–5	Human	W. F. Bodmer, pers. comm.
Proportion of cells that are proliferative	2/3	Mouse	Potten and Loeffler [1990]
	1/2–4/5	Mouse	Li *et al.* [1994]
	2/3	Mouse	Drasdo and Loeffler [2001]
	2/3	Mouse	Meineke *et al.* [2001]
	1/3	Mouse	Ro and Rannala [2001]
	1/3	Mouse	Katz *et al.* [2002]
	1/5	Mouse	Rozen *et al.* [2002]
	1/3	Mouse	van de Wetering *et al.* [2002]
	1/3	Mouse	Giles *et al.* [2003]
	3/4	Mouse	Bjerknes and Cheng [2006]
	3/5	Human	Marshman *et al.* [2002]
Number of transit cell generations	4	Mouse[3]	Loeffler *et al.* [1986]
	4	Mouse[3]	Potten and Loeffler [1987]
	4–6	Mouse	Potten and Loeffler [1990]
	4–6	Mouse	Paulus *et al.* [1992]
	5–6	Mouse	Paulus *et al.* [1993]
	4–6	Mouse	Marshman *et al.* [2002]
	4	Mouse	H. Clevers, pers. comm.
Probability of stem cell apoptosis	0.05–0.1	Mouse	Bach *et al.* [2000]
	0–0.1	Human	W. F. Bodmer, pers. comm.
Mutation rate per cell division	10^{-8}–10^{-6}	Human	Goldie and Coldman [1998]

[1]Data refer to the murine small intestine or the human large intestine.

[2]In several cases where there are repeated references to one value, we have tried to quote the primary reference.

[3]Model parameter for simulations.

Appendix B

Model variables and parameters

N_0	Stem cell population size
N_1	Semi-differentiated cell population size
N_2	Fully-differentiated cell population size
n_i	Initial size of cell population i
N_i^*	Equilibrium size of cell population i
$\hat{N}_i$	Size of cell population i, averaged over all ages in the age-structured model
t_0	Stem cell cycle time (days)
t_1	Semi-differentiated cell cycle time (days)
t_2	Time interval over which fully-differentiated cell death/removal is measured (days)
α_1	Stem cell death rate (days^{-1})
α_2	Stem cell differentiation rate (days^{-1})
α_3	Stem cell renewal rate (days^{-1})
β_1	Semi-differentiated cell death rate (days^{-1})
β_2	Semi-differentiated cell differentiation rate (days^{-1})
β_3	Semi-differentiated cell renewal rate (days^{-1})
γ	Fully-differentiated cell death/removal rate (days^{-1})
α	$\alpha_3 - \alpha_1 - \alpha_2$, the net stem cell growth rate (days^{-1})
β	$\beta_3 - \beta_1 - \beta_2$, the net semi-differentiated cell growth rate due to semi-differentiated cells (days^{-1})
a_1	Proportion of stem cells dying after each cell cycle
a_2	Proportion of stem cells differentiating after each cell cycle
a_3	Proportion of stem cells renewing after each cell cycle
b_1	Proportion of semi-differentiated cells dying after each cell cycle
b_2	Proportion of semi-differentiated cells differentiating after each cell cycle
b_3	Proportion of semi-differentiated cells renewing after each cell cycle
c	Proportion of fully-differentiated cells dying/leaving the crypt after each t_2 time interval

Appendix C

Approximating the summations

In this section we seek approximate closed form expressions for the summations

$$\sigma_1 = \sum_{n=1}^{\lfloor t/t_0 \rfloor} (2a_3)^{n-1}(2b_3)^{\lfloor t/t_1 - nt_0/t_1 \rfloor}, \tag{C.1}$$

$$\sigma_2 = \sum_{m=1}^{\lfloor t/t_1 \rfloor} (2b_3)^{m-1}(1-c)^{\lfloor t/t_2 - mt_1/t_2 \rfloor}, \tag{C.2}$$

$$\sigma_3 = \sum_{n=1}^{\lfloor t/t_0 - t_1/t_0 \rfloor} \sum_{m=1}^{\lfloor t/t_1 - nt_0/t_1 \rfloor} (2a_3)^{n-1}(2b_3)^{m-1}(1-c)^{\lfloor t/t_2 - nt_0/t_2 - mt_1/t_2 \rfloor}, \tag{C.3}$$

which were given in (3.70)–(3.72) in Chapter 3. Firstly we evaluate (C.1)–(C.3) in the most general case where all the cell cycle times t_0, t_1 and t_2 are unknown, and secondly we evaluate (C.2) and (C.3) in the special case where $t_1 = t_2$ in Section C.4.

C.1 A first approximation to the summations

As a simple first attempt, since we assume that time t is large compared with the cell cycle times t_0, t_1 and t_2, we make the assumption that we can approximate the summations (C.1)–(C.3) by removing the floor signs and replacing them by the actual values of the arguments. In addition, we also assume that $\lfloor t/t_0 - t_1/t_0 \rfloor \approx t/t_0$ when $t \gg t_1$. The resultant summations can now be evaluated exactly as

$$\sigma_1 \approx \sum_{n=1}^{t/t_0} (2a_3)^{n-1}(2b_3)^{t/t_1 - nt_0/t_1} = \frac{(2a_3)^{t/t_0} - (2b_3)^{t/t_1}}{2a_3 - (2b_3)^{t_0/t_1}} = \sigma_{1a}, \tag{C.4}$$

$$\sigma_2 \approx \sum_{m=1}^{t/t_1} (2b_3)^{m-1}(1-c)^{t/t_2 - mt_1/t_2} = \frac{(2b_3)^{t/t_1} - (1-c)^{t/t_2}}{2b_3 - (1-c)^{t_1/t_2}} = \sigma_{2a}, \tag{C.5}$$

and

$$\sigma_3 \approx \sum_{n=1}^{t/t_0 - t_1/t_0} \sum_{m=1}^{t/t_1 - nt_0/t_1} (2a_3)^{n-1}(2b_3)^{m-1}(1-c)^{t/t_2 - nt_0/t_2 - mt_1/t_2} \tag{C.6}$$

$$= \frac{1}{2b_3 - (1-c)^{t_1/t_2}} \left[\frac{(2a_3)^{t/t_0} - (2b_3)^{t/t_1}}{2a_3 - (2b_3)^{t_0/t_1}} - \frac{(2a_3)^{t/t_0} - (1-c)^{t/t_2}}{2a_3 - (1-c)^{t_0/t_2}} \right] = \sigma_{3a}. \tag{C.7}$$

To test these approximations, we plot the ratios of the exact summations to their approximations σ_i/σ_{ia} in Figure C.1, for $i = 1, 2, 3$, and where the parameter values are $t_0 = 24$, $t_1 = 12$, $t_2 = 48$, $a_3 = 0.5$, $b_3 = 0.5121$ and $c = 0.6094$. The graph shows that the ratio for σ_1/σ_{1a} tends towards 1, but the other two ratios tend to approximately 1.35. The error appears to be constant in time, and is the same for both σ_2 and σ_3, which suggests that the approximations are missing out the same constant factor. Since σ_1 and σ_2 are of the same form, and the same approximation is made in each, the error is solely due to the particular ratios of t_0/t_1 and t_1/t_2 chosen. In the next section we discuss a better approximation to the summations.

C.2 A better approximation to σ_1 and σ_2

The first approximation to the summations misses out a constant factor, so it becomes necessary to make a better approximation. Let us consider a generalisation of σ_1 and σ_2 denoted by

$$I = \sum_{n=1}^{t/t_0} A^{n-1} B^{\lfloor t/t_1 - nt_0/t_1 \rfloor}. \tag{C.8}$$

Consider the summation when $t = mt_0t_1$, and assume that m, t_0, $t_1 \in \mathbb{N}$. In this case we can write

$$I = \sum_{n=1}^{mt_1} A^{n-1} B^{\lfloor mt_0 - nt_0/t_1 \rfloor} = B^{mt_0} \sum_{n=1}^{mt_1} A^{n-1} B^{\lfloor -nt_0/t_1 \rfloor}. \tag{C.9}$$

Splitting up I into m summations, each of t_1 terms, we get

$$\begin{aligned} IB^{-mt_0} &= \left(\sum_{n=1}^{t_1} + \sum_{n=t_1+1}^{2t_1} + \sum_{n=(m-1)t_1+1}^{mt_1} \right) A^{n-1} B^{\lfloor -nt_0/t_1 \rfloor} \\ &= \sum_{n=1}^{t_1} A^{n-1} B^{\lfloor -nt_0/t_1 \rfloor} + \sum_{k=1}^{t_1} A^{t_1+k-1} B^{\lfloor -(t_1+k)t_0/t_1 \rfloor} + \ldots \\ &\qquad + \sum_{k=1}^{t_1} A^{(m-1)t_1+k-1} B^{\lfloor -[(m-1)t_1+k]t_0/t_1 \rfloor}. \end{aligned} \tag{C.10}$$

Factorising (C.10), we get

$$\begin{aligned} IB^{-mt_0} &= \left(A^0B^0 + A^{t_1}B^{-t_0} + \ldots + A^{(m-1)t_1}B^{-t_0(m-1)} \right) \sum_{n=1}^{t_1} A^{n-1} B^{\lfloor -nt_0/t_1 \rfloor} \\ &= \left(\sum_{n=0}^{m-1} A^{nt_1} B^{-nt_0} \right) \left(\sum_{n=1}^{t_1} A^{n-1} B^{\lfloor -nt_0/t_1 \rfloor} \right). \end{aligned} \tag{C.11}$$

The first summation gives

$$\sum_{n=0}^{m-1} A^{nt_1} B^{-nt_0} = B^{-t_0(m-1)} \left(\frac{A^{mt_1} - B^{mt_0}}{A^{t_1} - B^{t_0}} \right), \tag{C.12}$$

and for the second summation we can change the index of summation using $n-1 = t_1 - k$, where k runs from 1 to t_1. Hence, substituting (C.12) back into (C.11) and rearranging, we see that

$$\sum_{n=1}^{mt_1} A^{n-1} B^{\lfloor mt_0 - nt_0/t_1 \rfloor} = \frac{A^{mt_1} - B^{mt_0}}{A^{t_1} - B^{t_0}} \sum_{k=1}^{t_1} A^{t_1-k} B^{\lfloor (k-1)t_0/t_1 \rfloor}. \tag{C.13}$$

Therefore, since $t = mt_0t_1$, we propose that a good approximation for I when m is large is given by

$$I = \sum_{n=1}^{t/t_0} A^{n-1} B^{\lfloor t/t_1 - nt_0/t_1 \rfloor} \approx \frac{A^{t/t_0} - B^{t/t_1}}{A^{t_1} - B^{t_0}} \sum_{n=1}^{t_1} A^{t_1-n} B^{\lfloor (n-1)t_0/t_1 \rfloor}, \tag{C.14}$$

which is identically equal to I at every time point $t = mt_0t_1$. Using this better approximation, we obtain

$$\sigma_1 = \sum_{n=1}^{\lfloor t/t_0 \rfloor} (2a_3)^{n-1} (2b_3)^{\lfloor t/t_1 - nt_0/t_1 \rfloor} \approx f_1 \left(\frac{(2a_3)^{t/t_0} - (2b_3)^{t/t_1}}{(2a_3)^{t_1} - (2b_3)^{t_0}} \right) = \sigma_{1b}, \tag{C.15}$$

$$\sigma_2 = \sum_{m=1}^{\lfloor t/t_1 \rfloor} (2b_3)^{m-1} (1-c)^{\lfloor t/t_2 - mt_1/t_2 \rfloor} \approx f_2 \left(\frac{(2b_3)^{t/t_1} - (1-c)^{t/t_2}}{(2b_3)^{t_2} - (1-c)^{t_1}} \right) = \sigma_{2b}, \tag{C.16}$$

where

$$f_1 = \sum_{n=1}^{t_1} (2a_3)^{t_1-n} (2b_3)^{\lfloor (n-1)t_0/t_1 \rfloor} \qquad \text{and} \qquad f_2 = \sum_{n=1}^{t_2} (2b_3)^{t_2-n} (1-c)^{\lfloor (n-1)t_1/t_2 \rfloor}. \tag{C.17}$$

We see that if the floor functions in f_1 and f_2 are replaced by their arguments, or alternatively if $t_0 = pt_1$ and $t_1 = qt_2$, where p and q are integers, then σ_{1b} and σ_{2b} from (C.15) and (C.16) reduce to the original approximations σ_{1a} and σ_{2a} from (C.4) and (C.5).

To test these approximations we plot the ratios σ_i/σ_{ib} in Figure C.1 for $i = 1, 2$, where we again take the parameter values to be $t_0 = 24$, $t_1 = 12$, $t_2 = 48$, $a_3 = 0.5$, $b_3 = 0.5121$ and $c = 0.6094$. The graph shows that the ratios σ_1/σ_{1b} and σ_2/σ_{2b} oscillate towards 1, taking the value of 1 at regular time intervals, and so (C.15) and (C.16) are good approximations for σ_1 and σ_2. In the next section we discuss a better approximation for σ_3.

C.3 A better approximation to σ_3

In this section we wish to derive a better approximation to (C.3). Firstly, we calculate the summation over m, then we calculate the summation over n, and finally we consider a remainder term.

C.3.1 Summation over m

Consider the summation over m within σ_3 from (C.3) with general base numbers B and C, where $B = 2b_3$ and $C = 1 - c$, which is given by

$$J = \sum_{m=1}^{\lfloor t/t_1 - nt_0/t_1 \rfloor} B^{m-1} C^{\lfloor t/t_2 - nt_0/t_2 - mt_1/t_2 \rfloor}. \tag{C.18}$$

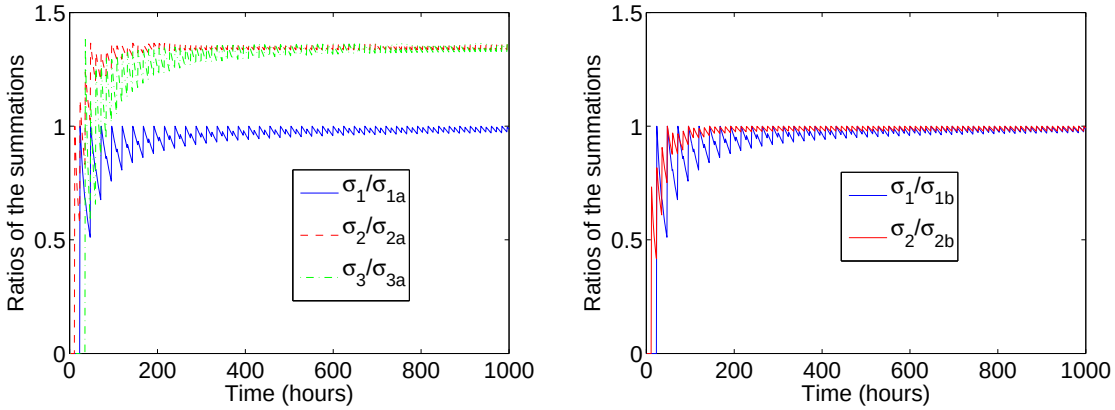

Figure C.1: Ratios of the exact summations to their approximations. Left: Plot of σ_1/σ_{1a}, σ_2/σ_{2a} and σ_3/σ_{3a}. Right: Plot of σ_1/σ_{1b} and σ_2/σ_{2b}. The summations σ_1, σ_2, σ_3, σ_{1a}, σ_{2a}, σ_{3a}, σ_{1b} and σ_{2b} are given by, respectively, (C.1), (C.2), (C.3), (C.4), (C.5), (C.7), (C.15) and (C.16). The parameter values for both plots are taken to be $t_0 = 24$, $t_1 = 12$, $t_2 = 48$, $a_3 = 0.5$, $b_3 = 0.5121$ and $c = 0.6094$.

At time points $t = pt_1t_2$, in the case where p, t_1, $t_2 \in \mathbb{N}$, the summation is given by

$$J = \sum_{m=1}^{\lfloor pt_2 - nt_0/t_1 \rfloor} B^{m-1} C^{\lfloor t_1(p-m/t_2) - nt_0/t_2 \rfloor}, \tag{C.19}$$

which can be split up into $\lfloor p - nt_0/(t_1t_2) \rfloor$ summations, each of t_2 terms, and a remainder term R_n at the end. This gives

$$\begin{aligned} J = &\left(\sum_{m=1}^{t_2} + \sum_{m=t_2+1}^{2t_2} + \ldots + \sum_{m=t_2\lfloor p-1-nt_0/(t_1t_2)\rfloor+1}^{t_2\lfloor p-nt_0/(t_1t_2)\rfloor} + \sum_{m=t_2\lfloor p-nt_0/(t_1t_2)\rfloor+1}^{\lfloor pt_2-nt_0/t_1\rfloor} \right) \\ &\qquad \times B^{m-1} C^{\lfloor t_1(p-m/t_2)-nt_0/t_2\rfloor} \\ = &\sum_{m=1}^{t_2} B^{m-1} C^{\lfloor t_1(p-m/t_2)-nt_0/t_2\rfloor} + \sum_{k=1}^{t_2} B^{t_2+k-1} C^{\lfloor t_1(p-1-k/t_2)-nt_0/t_2\rfloor} + \ldots \\ &+ \sum_{k=1}^{t_2} B^{t_2\lfloor p-1-nt_0/(t_1t_2)\rfloor+k-1} C^{\lfloor t_1(p-\lfloor p-1-nt_0/t_1t_2\rfloor-k/t_2)-nt_0/t_2\rfloor} + R_n, \end{aligned} \tag{C.20}$$

where R_n is given by

$$R_n = \sum_{m=t_2\lfloor p-nt_0/(t_1t_2)\rfloor+1}^{\lfloor pt_2-nt_0/t_1\rfloor} B^{m-1} C^{\lfloor t_1(p-m/t_2)-nt_0/t_2\rfloor}. \tag{C.21}$$

Factorising (C.20), we get

$$J = \left(B^0C^0 + B^{t_2}C^{-t_1} + \ldots + B^{t_2\lfloor p-1-nt_0/(t_1t_2)\rfloor}C^{-t_1\lfloor p-1-nt_0/(t_1t_2)\rfloor}\right) \times \sum_{m=1}^{t_2} B^{m-1}C^{\lfloor t_1(p-m/t_2)-nt_0/t_2\rfloor} + R_n$$
$$= \left(\sum_{k=0}^{\lfloor p-1-nt_0/(t_1t_2)\rfloor} B^{kt_2}C^{-kt_1}\right)\left(\sum_{m=1}^{t_2} B^{m-1}C^{\lfloor t_1(p-m/t_2)-nt_0/t_2\rfloor}\right) + R_n. \tag{C.22}$$

The first summation in (C.22) can be performed exactly, and the second summation can be rewritten using the index $k = t_2 - m$, so that J can be written as

$$J = S\left(\frac{B^{t/t_1+t_2\lfloor -nt_0/(t_1t_2)\rfloor} - C^{t/t_2+t_1\lfloor -nt_0/(t_1t_2)\rfloor}}{B^{t_2} - C^{t_1}}\right) + R_n, \tag{C.23}$$

where

$$S = \sum_{k=1}^{t_2} B^{t_2-k}C^{\lfloor (k-1)t_1/t_2-nt_0/t_2\rfloor - t_1\lfloor -nt_0/(t_1t_2)\rfloor}. \tag{C.24}$$

Rewriting the summation index $m = t_2\lfloor p - nt_0/(t_1t_2)\rfloor + k$ in (C.21), we get

$$R_n = B^{t/t_1+t_2\lfloor -nt_0/(t_1t_2)\rfloor}C^{-t_1\lfloor -nt_0/(t_1t_2)\rfloor}\sum_{k=1}^{\lfloor -nt_0/t_1\rfloor - t_2\lfloor -nt_0/(t_1t_2)\rfloor} B^{k-1}C^{\lfloor -kt_1/t_2-nt_0/t_2\rfloor}, \tag{C.25}$$

which is of the form $R_n = FB^{t/t_1}$, where F is independent of time.

C.3.2 Summation over n

Using (C.23), (C.3) can be written as

$$\sigma_3 = \sum_{n=1}^{\lfloor t/t_0-t_1/t_0\rfloor} S(n)(2a_3)^{n-1}\left(\frac{(2b_3)^{t/t_1+t_2\lfloor -nt_0/(t_1t_2)\rfloor} - (1-c)^{t/t_2+t_1\lfloor -nt_0/(t_1t_2)\rfloor}}{(2b_3)^{t_2} - (1-c)^{t_1}}\right) + R, \tag{C.26}$$

where

$$R = \sum_{n=1}^{\lfloor t/t_0-t_1/t_0\rfloor} (2a_3)^{n-1}R_n \tag{C.27}$$

and

$$S(n) = \sum_{k=1}^{t_2}(2b_3)^{t_2-k}(1-c)^{\lfloor (k-1)t_1/t_2-nt_0/t_2\rfloor - t_1\lfloor -nt_0/(t_1t_2)\rfloor}. \tag{C.28}$$

In order to obtain a closed form approximation to (C.3) we must first approximate (C.28). If we assume that

$$\left\lfloor (k-1)\frac{t_1}{t_2} - \frac{nt_0}{t_2} \right\rfloor \approx \left\lfloor (k-1)\frac{t_1}{t_2} \right\rfloor + \left\lfloor -\frac{nt_0}{t_2} \right\rfloor, \tag{C.29}$$

then

$$S(n) \approx (1-c)^{\lfloor -nt_0/t_2 \rfloor - t_1 \lfloor -nt_0/(t_1t_2) \rfloor} \sum_{k=1}^{t_2} (2b_3)^{t_2-k} (1-c)^{\lfloor (k-1)t_1/t_2 \rfloor} \tag{C.30}$$

$$= f_2 (1-c)^{\lfloor -nt_0/t_2 \rfloor - t_1 \lfloor -nt_0/(t_1t_2) \rfloor} = S_1(n). \tag{C.31}$$

$S_1(n)$ is a good approximation to $S(n)$, which we illustrate in Figure C.2. Substituting (C.31) into (C.26) gives

$$\sigma_3 \approx \frac{f_2}{(2b_3)^{t_2} - (1-c)^{t_1}} \left[(2b_3)^{t/t_1} \sum_{n=1}^{\lfloor t/t_0 - t_1/t_0 \rfloor} (2a_3)^{n-1} (2b_3)^{t_2 \lfloor -nt_0/(t_1t_2) \rfloor} \right.$$
$$\left. \times (1-c)^{\lfloor -nt_0/t_2 \rfloor - t_1 \lfloor -nt_0/(t_1t_2) \rfloor} - (1-c)^{t/t_2} \sum_{n=1}^{\lfloor t/t_0 - t_1/t_0 \rfloor} (2a_3)^{n-1} (1-c)^{\lfloor -nt_0/t_2 \rfloor} \right] + R. \tag{C.32}$$

When time t is much greater than the cell cycle times t_i, we can approximate $\lfloor t/t_0 - t_1/t_0 \rfloor \approx t/t_0$, and so both summations take the form

$$K = \sum_{n=1}^{t/t_0} A^{n-1} B^{\lfloor -nt_0/(t_1t_2) \rfloor} C^{\lfloor -nt_0/t_2 \rfloor}, \tag{C.33}$$

where $B = (2b_3)^{t_2}(1-c)^{-t_1}$ and $B = 1$ in the first and second summations of (C.32), respectively, and $A = 2a_3$, $C = 1-c$. Considering time points $t = qt_0t_1t_2$, where q, t_0, t_1, $t_2 \in \mathbb{N}$, K can be split into q summations, each of t_1t_2 terms, which can be evaluated by the method described in Section C.3.1. Hence

$$K \approx \frac{A^{t/t_0} B^{-t/(t_1t_2)} C^{-t/t_2} - 1}{A^{t_1t_2} B^{-t_0} C^{-t_0t_1} - 1} \sum_{n=1}^{t_1t_2} A^{n-1} B^{\lfloor -nt_0/(t_1t_2) \rfloor} C^{\lfloor -nt_0/t_2 \rfloor}. \tag{C.34}$$

Finally, applying the approximation (C.34) to (C.32), we find that

$$\sigma_3 \approx \frac{f_2}{(2b_3)^{t_2} - (1-c)^{t_1}} \left[f_3 \frac{(2a_3)^{t/t_0} - (2b_3)^{t/t_1}}{(2a_3)^{t_1t_2} - (2b_3)^{t_0t_2}} - f_4 \frac{(2a_3)^{t/t_0} - (1-c)^{t/t_2}}{(2a_3)^{t_1t_2} - (1-c)^{t_0t_1}} \right] + R, \tag{C.35}$$

where

$$f_3 = \sum_{n=1}^{t_1t_2} (2a_3)^{t_1t_2-n} (2b_3)^{t_2 \lfloor (n-1)t_0/(t_1t_2) \rfloor} (1-c)^{\lfloor (n-1)t_0/t_2 \rfloor - t_1 \lfloor (n-1)t_0/(t_1t_2) \rfloor}, \tag{C.36}$$

$$f_4 = \sum_{n=1}^{t_1t_2} (2a_3)^{t_1t_2-n} (1-c)^{\lfloor (n-1)t_0/t_2 \rfloor}. \tag{C.37}$$

C.3.3 The remainder term

Finally we must consider the remainder term. Substituting (C.25) into (C.27), where we recall that $B = 2b_3$ and $C = 1 - c$, we find that

$$R = (2b_3)^{t/t_1} \sum_{n=1}^{\lfloor t/t_0 - t_1/t_0 \rfloor} \Bigg[(2a_3)^{n-1} (2b_3)^{t_2 \lfloor -nt_0/(t_1 t_2) \rfloor} (1-c)^{-t_1 \lfloor -nt_0/(t_1 t_2) \rfloor} \times \sum_{k=1}^{\lfloor -nt_0/t_1 \rfloor - t_2 \lfloor -nt_0/(t_1 t_2) \rfloor} (2b_3)^{k-1} (1-c)^{\lfloor -kt_1/t_2 - nt_0/t_2 \rfloor} \Bigg]. \quad \text{(C.38)}$$

At time points $t = rt_0 t_1 t_2$, assuming that $t \gg t_i$ so that we can approximate $\lfloor t/t_0 - t_1/t_0 \rfloor \approx \lfloor t/t_0 \rfloor$, (C.38) can be written as $R \approx (2b_3)^{t/t_1} L$, where

$$L = \sum_{n=1}^{rt_1 t_2} (2a_3)^{n-1} \sum_{k=1}^{\lfloor -nt_0/t_1 \rfloor - t_2 \lfloor -nt_0/(t_1 t_2) \rfloor} (2b_3)^{t_2 \lfloor -nt_0/(t_1 t_2) \rfloor + k - 1} \times (1-c)^{-t_1 \lfloor -nt_0/(t_1 t_2) \rfloor + \lfloor -kt_1/t_2 - nt_0/t_2 \rfloor}. \quad \text{(C.39)}$$

L can be split up into r summations, each containing $t_1 t_2$ terms, and the terms that contain r can be factored out to give

$$L = M \sum_{k=1}^{r} \left[(2a_3)^{t_1 t_2} (2b_3)^{-t_0 t_2} \right]^{k-1} = M (2b_3)^{t_0 t_2 - r t_0 t_2} \frac{(2a_3)^{r t_1 t_2} - (2b_3)^{r t_0 t_2}}{(2a_3)^{t_1 t_2} - (2b_3)^{t_0 t_2}}, \quad \text{(C.40)}$$

where

$$M = \sum_{n=1}^{t_1 t_2} (2a_3)^{n-1} \sum_{k=1}^{\lfloor -nt_0/t_1 \rfloor - t_2 \lfloor -nt_0/(t_1 t_2) \rfloor} (2b_3)^{t_2 \lfloor -nt_0/(t_1 t_2) \rfloor + k - 1} \times (1-c)^{-t_1 \lfloor -nt_0/(t_1 t_2) \rfloor + \lfloor -kt_1/t_2 - nt_0/t_2 \rfloor} \quad \text{(C.41)}$$

is independent of time. Rewriting $n = t_1 t_1 - k + 1$ as the second summation variable in (C.41), we can substitute back in for $r = t/(t_0 t_1 t_2)$ and estimate (C.38) as

$$R \approx f_5 \left(\frac{(2a_3)^{t/t_0} - (2b_3)^{t/t_1}}{(2a_3)^{t_1 t_2} - (2b_3)^{t_0 t_2}} \right), \quad \text{(C.42)}$$

where

$$f_5 = \sum_{k=1}^{t_1 t_2} \sum_{m=1}^{\lfloor (k-1)t_0/t_1 \rfloor - t_2 \lfloor (k-1)t_0/(t_1 t_2) \rfloor} \Bigg[(2a_3)^{t_1 t_2 - k} (2b_3)^{t_2 \lfloor (k-1)t_0/(t_1 t_2) \rfloor + m - 1} \times (1-c)^{-t_1 \lfloor (k-1)t_0/(t_1 t_2) \rfloor + \lfloor (k-1)t_0/t_2 - mt_1/t_2 \rfloor} \Bigg]. \quad \text{(C.43)}$$

Hence the full approximation for σ_3 can be written in the same form as (C.35) using (C.42) to give

$$\sigma_3 \approx \frac{f_2}{(2b_3)^{t_2} - (1-c)^{t_1}} \left[\hat{f}_3 \frac{(2a_3)^{t/t_0} - (2b_3)^{t/t_1}}{(2a_3)^{t_1 t_2} - (2b_3)^{t_0 t_2}} - f_4 \frac{(2a_3)^{t/t_0} - (1-c)^{t/t_2}}{(2a_3)^{t_1 t_2} - (1-c)^{t_0 t_1}} \right] = \sigma_{3b}, \quad \text{(C.44)}$$

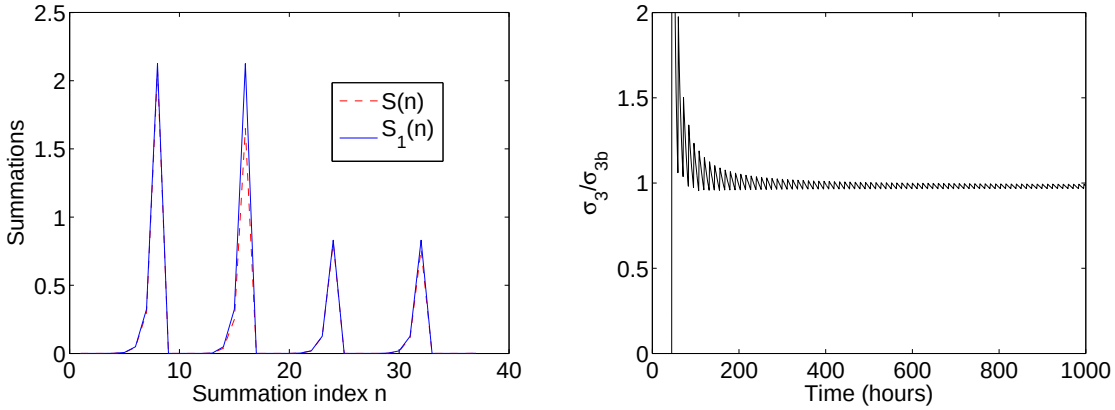

Figure C.2: Approximations for σ_3. Left: A comparison of the summations $S(n)$ with $S_1(n)$, given by (C.28) and (C.31), respectively. The parameters are taken to be $t_0 = 27$, $t_1 = 17$, $t_2 = 13$. Right: Plot of the ratio σ_3/σ_{3b}, where σ_3 and σ_{3b} are given by (C.3) and (C.44), respectively. The parameter values are taken to be $t_0 = 24$, $t_1 = 12$, $t_2 = 48$. In both cases $a_3 = 0.5$, $b_3 = 0.5121$ and $c = 0.6094$.

where

$$\hat{f}_3 = f_3 + \frac{f_5}{f_2}\left[(2b_3)^{t_2} - (1-c)^{t_1}\right]. \tag{C.45}$$

σ_{3b} produces a good estimate for σ_3, as we illustrate in Figure C.2.

C.4 Special case where $t_1 = t_2$

In Sections C.1–C.3 we presented approximations to the summations (C.1)–(C.3) for general values of t_0, t_1 and t_2, but in this section we consider the special case where $t_1 = t_2$. Now (C.2) can be evaluated exactly to give

$$\sigma_2 = \frac{(2b_3)^{\lfloor t/t_1 \rfloor} - (1-c)^{\lfloor t/t_1 \rfloor}}{2b_3 - (1-c)} = \sigma_{2c}. \tag{C.46}$$

We can also obtain a better approximation to σ_3, and in particular the summation over m can be performed exactly to give

$$\sigma_3 = \sum_{n=1}^{\lfloor t/t_0 - t_1/t_0 \rfloor} (2a_3)^{n-1}\left[\frac{(2b_3)^{\lfloor t/t_1 - nt_0/t_1 \rfloor} - (1-c)^{\lfloor t/t_1 - nt_0/t_1 \rfloor}}{2b_3 - (1-c)}\right]. \tag{C.47}$$

The approximation (C.44) is accurate at every interval $t = rt_0t_1t_2$, but now we can obtain an approximation to (C.47) that is accurate at every point $t = rt_0t_1$. Assuming that $t \gg t_i$ so that we can approximate $\lfloor t/t_0 - t_1/t_0 \rfloor \approx t/t_0$, (C.47) reduces into two summations of the form (C.8). Hence, we can approximate

$$\sigma_3 \approx \frac{1}{2b_3 - (1-c)}\left[f_1\frac{(2a_3)^{t/t_0} - (2b_3)^{t/t_1}}{(2a_3)^{t_1} - (2b_3)^{t_0}} - f_6\frac{(2a_3)^{t/t_0} - (1-c)^{t/t_1}}{(2a_3)^{t_1} - (1-c)^{t_0}}\right] = \sigma_{3c}, \tag{C.48}$$

where

$$f_6 = \sum_{n=1}^{t_1} (2a_3)^{t_1-n}(1-c)^{\lfloor (n-1)t_0/t_1 \rfloor}. \tag{C.49}$$

Bibliography

K. Araki, T. Ogata, M. Kobayashi, and R. Yatani. A morphological study on the histogenesis of human colorectal hyperplastic polyps. *Gastroenterology*, 109:1468–1474, 1995.

O. Arino and M. Kimmel. Asymptotic behavior of nonlinear semigroup describing a model of selective cell growth regulation. *Journal of Mathematical Biology*, 29:289–314, 1991.

O. Arino and M. Kimmel. Comparisons of approaches to modeling of cell population dynamics. *SIAM Journal on Applied Mathematics*, 53:1480–1504, 1993.

P. Armitage and R. Doll. The age distribution of cancer and a multi-stage theory of carcinogenesis. *British Journal of Cancer*, 8:1–12, 1954.

S. P. Bach, A. G. Renehan, and C. S. Potten. Stem cells: the intestinal stem cell as a paradigm. *Carcinogenesis*, 21:469–476, 2000.

N. Barker, J. H. van Es, J. Kuipers, P. Kujala, M. van den Born, M. Cozijnsen, A. Haegebarth, J. Korving, H. Begthel, P. J. Peters, and H. Clevers. Identification of stem cells in small intestine and colon by marker gene Lgr5. *Nature*, 449:1003–1007, 2007.

E. Batlle, J. Bacani, H. Begthel, S. Jonkeer, A. Gregorieff, M. van de Born, N. Malats, E. Sancho, E. Boon, T. Pawson, S. Gallinger, S. Pals, and H. Clevers. EphB receptor activity suppresses colorectal cancer progression. *Nature*, 435:1126–1130, 2005.

E. Batlle, J. T. Henderson, H. Beghtel, M. M. W. van den Born, E. Sancho, G. Huls, J. Meeldijk, J. Robertson, M. van der Wetering, T. Pawson, and H. Clevers. β-catenin and TCF mediate cell positioning in the intestinal epithelium by controlling the expression of EphB/EphrinB. *Cell*, 111:251–263, 2002.

P. A. Batman, D. P. Kotler, M. S. Kapembwa, D. Booth, C. S. Potten, J. M. Orenstein, A. J. Scally, and G. E. Griffin. HIV enteropathy: crypt stem and transit cell hyperproliferation induces villous atrophy in HIV/Microsporidia-infected jejunal mucosa. *Aids*, 21:433–439, 2007.

R. A. Beckman and L. A. Loeb. Efficiency of carcinogenesis with and without a mutator mutation. *Proceedings of the National Academy of Sciences U.S.A.*, 103:14140–14145, 2006.

M. Bjerknes. Expansion of mutant stem cell populations in the human colon. *Journal of Theoretical Biology*, 178:381–385, 1996.

M. Bjerknes and H. Cheng. Intestinal epithelial stem cells and progenitors. *Methods in Enzymology*, 419:337–383, 2006.

C. Blanpain, V. Horsley, and E. Fuchs. Epithelial stem cells: turning over new leaves. *Cell*, 128:445–458, 2007.

W. Bodmer. Genetic instability is not a requirement for tumor development. *Cancer Research*, 68:3558–3561, 2008.

W. F. Bodmer. The somatic evolution of cancer. *The Indian Journal of Statistics*, 61: 371–381, 1999.

W. F. Bodmer and I. Tomlinson. Population genetics of tumours. *Ciba Foundation Symposium*, 197:181–193, 1996.

B. M. Boman, J. Z. Fields, O. Bonham-Carter, and O. A. Runquist. Computer modeling implicates stem cell overproduction in colon cancer initiation. *Cancer Research*, 61:8408–8411, 2001.

M. Brittan and N. A. Wright. Gastrointestinal stem cells. *Journal of Pathology*, 197: 492–509, 2002.

M. Brittan and N. A. Wright. The gastrointestinal stem cell. *Cell Proliferation*, 37:35–53, 2004a.

M. Brittan and N. A. Wright. Stem cell in gastrointestinal structure and neoplastic development. *Gut*, 53:899–910, 2004b.

P. Calabresi, P. S. Schein, and S. A. Rosenberg, editors. *Medical oncology: basic principles and clinical management of cancer*, pages 41–60. MacMillan, New York, 1985.

Cancer Research UK, 2008. URL `www.cancerresearchuk.org/`.

S. J. Chapman, M. J. Plank, A. James, and B. Basse. A nonlinear model of age and size-structured populations with applications to cell cycles. *ANZIAM Journal*, 49:151–169, 2007.

C. Y. Chen, H. M. Byrne, and J. R. King. The influence of growth-induced stress from the surrounding medium on the development of multicell spheroids. *Journal of Mathematical Biology*, 43:191–220, 2001.

E. Clayton, D. P. Doupé, A. M. Klein, D. J. Winton, B. D. Simons, and P. H. Jones. A single type of progenitor cell maintains normal epidermis. *Nature*, 446:185–189, 2007.

H. Clevers. Wnt breakers in colon cancer. *Cancer Cell*, 5:5–6, 2004.

H. Clevers. Stem cells, asymmetric division and cancer. *Nature Genetics*, 37:1027–1028, 2005.

H. Clevers and E. Batlle. EphB/EphrinB receptors and Wnt signaling in colorectal cancer. *Cancer Research*, 66:2–5, 2006.

P. Dalerba, S. J. Dylla, I-K. Park, R. Liu, X. Wang, R. W. Cho, T. Hoey, A. Gurney, E. H. Huang, D. M. Simeone, A. A. Shelton, G. Parmiani, C. Castelli, and M. F. Clarke. Phenotypic characterization of human colorectal cancer stem cells. *Proceedings of the National Academy of Sciences U.S.A.*, 104:10158–10163, 2007.

A. d'Onofrio and I. P. M. Tomlinson. A nonlinear mathematical model of cell turnover, differentiation and tumorigenesis in the intestinal crypt. *Journal of Theoretical Biology*, 244:367–374, 2007.

D. Drasdo and M. Loeffler. Individual-based models to growth and folding in one-layered tissues: Intestinal crypts and early development. *Nonlinear Analysis*, 47:245–256, 2001.

C. M. Edwards and S. J. Chapman. Biomechanical modelling of colorectal crypt budding and fission. *Bulletin of Mathematical Biology*, 69:1927–1942, 2007.

R. Erban, S. J. Chapman, and P. K. Maini. A practical guide to stochastic simulations of reaction-diffusion processes. *arXiv:0704.1908v2*, 2007.

E. R. Fearon and B. Vogelstein. A genetic model for colorectal tumorigenesis. *Cell*, 61: 759–767, 1990.

R. Fodde. The multiple functions of tumour suppressors: it's all in *APC*. *Nature Cell Biology*, 5:190–192, 2003.

S. A. Frank and M. A. Nowak. Problems of somatic mutation and cancer. *BioEssays*, 26: 291–299, 2004.

E. A. Gaffney. The application of mathematical modelling to aspects of adjuvant chemotherapy scheduling. *Journal of Mathematical Biology*, 48:375–422, 2004.

J. Galle, M. Loeffler, and D. Drasdo. Modeling the effect of deregulated proliferation and apoptosis on the growth dynamics of epithelial cell populations in vitro. *Biophysical Journal*, 88:62–75, 2005.

R. A. Gatenby and P. K. Maini. Cancer summed up. *Nature*, 421:321, 2003.

T. G. Gerike, U. Paulus, C. S. Potten, and M. Loeffler. A dynamic model of proliferation and differentiation in the intestinal crypt based on a hypothetical intraepithelial growth factor. *Cell Proliferation*, 31:93–110, 1998.

R. H. Giles, J. H. van Es, and H. Clevers. Caught up in a Wnt storm: Wnt signaling in cancer. *Biochimica et Biophysica Acta*, 1653:1–24, 2003.

J. H. Goldie and A. J. Coldman. *Drug resistance in cancer: mechanisms and models*. Cambridge University Press, Cambridge, 1998.

L. C. Greaves, S. L. Preston, P. J. Tadrous, R. W. Taylor, M. J. Barron, D. Oukrif, S. J. Leedham, M. Deheragoda, P. Sasieni, M. R. Novelli, J. A. Z. Jankowski, D. M. Turnbull, N. A. Wright, and S. A. C. McDonald. Mitochondrial DNA mutations are established in human colonic stem cells, and mutated clones expand by crypt fission. *Proceedings of the National Academy of Sciences U.S.A.*, 103:714–719, 2006.

D. R. Halm and S. T. Halm. Secretagogue response of goblet cells and columnar cells in human colonic crypts. *American Journal of Physiology - Cell Physiology*, 278:C212–C233, 2000.

D. Hanahan and R. A. Weinberg. The hallmarks of cancer. *Cell*, 100:57–70, 2000.

K. Hardy and J. Stark. Mathematical models of the balance between apoptosis and proliferation. *Apoptosis*, 7:373–381, 2002.

T. F. R. Homfray, S. E. Cottrell, M. Ilyas, A. Rowan, I. C. Talbot, W. F. Bodmer, and I. P. M. Tomlinson. Defects in mismatch repair occur after *APC* mutations in the pathogenesis of sporadic colorectal tumours. *Human Mutation*, 11:114–120, 1998.

S. Hotton and M. E. Colvin. Analytic formulas for discrete stochastic models of cell populations with both differentiation and de-differentiation. *Journal of Theoretical Biology*, 245:610–626, 2007.

M. Ilyas. Wnt signalling and the mechanistic basis of tumour development. *Journal of Pathology*, 205:130–144, 2005.

M. D. Johnston, C. M. Edwards, W. F. Bodmer, P. K. Maini, and S. J. Chapman. Examples of mathematical modeling: tales from the crypt. *Cell Cycle*, 6:2106–2112, 2007a.

M. D. Johnston, C. M. Edwards, W. F. Bodmer, P. K. Maini, and S. J. Chapman. Mathematical modeling of cell population dynamics in the colonic crypt and in colorectal cancer. *Proceedings of the National Academy of Sciences U.S.A.*, 104:4008–4013, 2007b.

J. P. Katz, N. Perreault, B. G. Goldstein, C. S. Lee, P. A. Labosky, V. W. Yang, and K. H. Kaestner. The zinc-finger transcription factor Klf4 is required for terminal differentiation of goblet cells in the colon. *Development*, 129:2619–2628, 2002.

S. E. Kern and D. Shibata. The fuzzy math of solid tumor stem cells: a perspective. *Cancer Research*, 67:8985–8988, 2007.

K-M. Kim, P. Calabrese, S. Tavaré, and D. Shibata. Enhanced stem cell survival in familial adenomatous polyposis. *American Journal of Pathology*, 164:1369–1377, 2004.

K-M. Kim and D. Shibata. Methylation reveals a niche: stem cell succession in human colon crypts. *Oncogene*, 21:5441–5449, 2002.

K-M. Kim and D. Shibata. Tracing ancestry with methylation patterns: most crypts appear distantly related in normal adult human colon. *BMC Gastroenterology*, 4:8, 2004.

M. Kimmel and D. E. Axelrod. Mathematical models of gene amplification with applications to cellular drug resistance and tumorigenicity. *Genetics*, 125:633–644, 1990.

K. W. Kinzler and B. Vogelstein. Lessons from hereditary colorectal cancer. *Cell*, 87: 159–170, 1996.

S. C. Kirkland. Clonal origin of columnar, mucous, and endocrine cell lineages in human colorectal epithelium. *Cancer*, 61:1359–1363, 1988.

A. G. Knudson. Hereditary cancer: two hits revisited. *Journal of Cancer Research and Clinical Oncology*, 122:135–140, 1996.

A. G. Knudson. Two genetic hits (more or less) to cancer. *Nature Reviews Cancer*, 1: 157–162, 2001.

N. L. Komarova. Cancer, aging and the optimal tissue design. *Seminars in Cancer Biology*, 15:494–505, 2005.

N. L. Komarova. Spatial stochastic models for cancer initiation and progression. *Bulletin of Mathematical Biology*, 68:1573–1599, 2006.

N. L. Komarova and P. Cheng. Epithelial tissue architecture protects against cancer. *Mathematical Biosciences*, 200:90–117, 2006.

N. L. Komarova and L. Wang. Initiation of colorectal cancer: where do the two hits hit? *Cell Cycle*, 3:1558–1565, 2004.

Y. Q. Li, S. A. Roberts, U. Paulus, M. Loeffler, and C. S. Potten. The crypt cycle in mouse small intestinal epithelium. *Journal of Cell Science*, 107:3271–3279, 1994.

M. P. Little and G. Li. Stochastic modelling of colon cancer: is there a role for genomic instability? *Carcinogenesis*, 28:479–487, 2007.

P. N. Lobachevsky and I. R. Radford. Intestinal crypt properties fit a model that incorporates replicative ageing and deep and proximate stem cells. *Cell Proliferation*, 39:379–402, 2006.

L. A. Loeb. Mutator phenotype may be required for multistage carcinogenesis. *Cancer Research*, 51:3075–3079, 1991.

L. A. Loeb, J. H. Bielas, and R. A. Beckman. Cancers exhibit a mutator phenotype: clinical implications. *Cancer Research*, 68:3551–3557, 2008.

M. Loeffler, A. Birke, D. Winton, and C. S. Potten. Somatic mutation, monoclonality and stochastic models of stem cell organization in the intestinal crypt. *Journal of Theoretical Biology*, 160:471–491, 1993.

M. Loeffler, T. Bratke, U. Paulus, Y. Q. Li, and C. S. Potten. Clonality and life cycles of intestinal crypts explained by a state dependent stochastic model of epithelial stem cell organization. *Journal of Theoretical Biology*, 186:41–54, 1997.

M. Loeffler, C. S. Potten, U. Paulus, J. Glatzer, and S. Chwalinski. Intestinal crypt proliferation. II. Computer modelling of mitotic index data provides further evidence for lateral and vertical cell migration in the absence of mitotic activity. *Cell and Tissue Kinetics*, 21:247–258, 1988.

M. Loeffler, R. Stein, H-E. Wichmann, C. S. Potten, P. Kaur, and S. Chwalinski. Intestinal cell proliferation. I. A comprehensive model of steady-state proliferation in the crypt. *Cell and Tissue Kinetics*, 19:627–645, 1986.

E. G. Luebeck and S. H. Moolgavkar. Multistage carcinogenesis and the incidence of colorectal cancer. *Proceedings of the National Academy of Sciences U.S.A.*, 99:15095–15100, 2002.

M. C. Mackey. Unified hypothesis for the origin of aplastic anemia and periodic hematopoiesis. *Blood*, 51:941–956, 1978.

M. C. Mackey and L. Glass. Oscillation and chaos in physiological control systems. *Science*, 197:287–289, 1977.

C. C. Maley and S. Forrest. Exploring the relationship between neutral and selective mutations in cancer. *Artificial Life*, 6:325–345, 2000.

C. C. Maley, B. J. Reid, and S. Forrest. Cancer prevention strategies that address the evolutionary dynamics of neoplastic cells: simulating benign cell boosters and selection for chemosensitivity. *Cancer Epidemiology, Biomarkers and Prevention*, 13:1375–1384, 2004.

E. Marshman, C. Booth, and C. S. Potten. The intestinal epithelial stem cell. *BioEssays*, 24:91–98, 2002.

S. A. C. McDonald, S. L. Preston, M. J. Lovell, N. A. Wright, and J. A. Z. Jankowski. Mechanisms of disease: from stem cells to colorectal cancer. *Gastroenterology and Hepatology*, 3:267–274, 2006.

F. A. Meineke, C. S. Potten, and M. Loeffler. Cell migration and organization in the intestinal crypt using a lattice-free model. *Cell Proliferation*, 34:253–266, 2001.

L. M. F. Merlo, J. W. Pepper, B. J. Reid, and C. C. Maley. Cancer as an evolutionary and ecological process. *Nature Reviews Cancer*, 6:924–935, 2006.

F. Michor, Y. Iwasa, C. Lengauer, and M. A. Nowak. Dynamics of colorectal cancer. *Seminars in Cancer Biology*, 15:484–493, 2005.

F. Michor, Y. Iwasa, and M. A. Nowak. Dynamics of cancer progression. *Nature Reviews Cancer*, 4:197–205, 2004a.

F. Michor, Y. Iwasa, H. Rajagopalan, C. Lengauer, and M. A. Nowak. Linear model of colon cancer initiation. *Cell Cycle*, 3:358–362, 2004b.

I. M. Modlin, M. Kidd, K. D. Lye, and N. A. Wright. Gastric stem cells: an update. *Keio Journal of Medicine*, 52:134–137, 2003.

A. Mogilner, R. Wollman, and W. F. Marshall. Quantitative modeling in cell biology: what is it good for? *Developmental Cell*, 11:279–287, 2006.

K. A. Moore and I. R. Lemischka. Stem cells and their niches. *Science*, 311:1880–1885, 2006.

S. J. Morrison and J. Kimble. Asymmetric and symmetric stem-cell divisions in development and cancer. *Nature*, 441:1068–1074, 2006.

J. D. Murray. *Mathematical Biology. I: An Introduction*. Springer, London, 2002.

J. D. Nagy. Competition and natural selection in a mathematical model of cancer. *Bulletin of Mathematical Biology*, 66:663–687, 2004.

M. R. Novelli, A. Cossu, D. Oukrif, A. Quaglia, S. Lakhani, R. Poulsom, P. Sasieni, P. Carta, M. Contini, A. Pasca, G. Palmieri, W. Bodmer, F. Tanda, and N. Wright. X-inactivation patch size in human female tissue confounds the assessment of tumor clonality. *Proceedings of the National Academy of Sciences U.S.A.*, 100:3311–3314, 2003.

M. R. Novelli, J. A. Williamson, I. P. M. Tomlinson, G. Elia, S. V. Hodgson, I. C. Talbot, W. F. Bodmer, and N. A. Wright. Polyclonal origin of colonic adenomas in an XO/XY patient with *FAP*. *Science*, 272:1187–1190, 1996.

M. A. Nowak, N. L. Komarova, A. Sengupta, P. V. Jallepalli, I-M. Shih, B. Vogelstein, and C. Lengauer. The role of chromosomal instability in tumor initiation. *Proceedings of the National Academy of Sciences U.S.A.*, 99:16226–16231, 2002.

M. A. Nowak, F. Michor, N. L. Komarova, and Y. Iwasa. Evolutionary dynamics of tumor suppressor gene inactivation. *Proceedings of the National Academy of Sciences U.S.A.*, 101:10635–10638, 2004.

G. Odell, G. Oster, B. Burnside, and P. Alberch. A mechanical model for epithelial morphogenesis. *Journal of Mathematical Biology*, 9:291–295, 1980.

G. M. Odell, G. Oster, P. Alberch, and B. Burnside. The mechanical basis of morphogenesis. I. Epithelial folding and invagination. *Developmental Biology*, 85:446–462, 1981.

R. Okamoto and M. Watanabe. Molecular and clinical basis for the regeneration of human gastrointestinal epithelia. *Journal of Gastroenterology*, 39:1–6, 2004.

U. Paulus, M. Loeffler, J. Zeidler, G. Owen, and C. S. Potten. The differentiation and lineage development of goblet cells in the murine small intestinal crypt: experimental and modelling studies. *Journal of Cell Science*, 106:473–484, 1993.

U. Paulus, C. S. Potten, and M. Loeffler. A model of the control of cellular regeneration in the intestinal crypt after perturbation based solely on local stem cell regulation. *Cell Proliferation*, 25:559–578, 1992.

K. Polyak and W. C. Hahn. Roots and stems: stem cells in cancer. *Nature Medicine*, 11: 296–300, 2005.

C. S. Potten, C. Booth, and D. Hargreaves. The small intestine as a model for evaluating adult tissue stem cell drug targets. *Cell Proliferation*, 36:115–129, 2003.

C. S. Potten and M. Loeffler. A comprehensive model of the crypts of the small intestine of the mouse provides insight into the mechanisms of cell migration and the proliferation hierarchy. *Journal of Theoretical Biology*, 127:381–391, 1987.

C. S. Potten and M. Loeffler. Stem cells: attributes, cycles, spirals, pitfalls and uncertainties. Lessons for and from the crypt. *Development*, 110:1001–1020, 1990.

S. L. Preston, W-M. Wong, A. O-O. Chan, R. Poulsom, R. Jeffery, R. A. Goodlad, N. Mandir, G. Elia, M. Novelli, W. F. Bodmer, I. P. Tomlinson, and N. A. Wright. Bottom-up histogenesis of colorectal adenomas: origin in the monocryptal adenoma and initial expansion by crypt fission. *Cancer Research*, 63:3819–3825, 2003.

H. Rajagopalan, M. A. Nowak, B. Vogelstein, and C. Lengauer. The significance of unstable chromosomes in colorectal cancer. *Cancer*, 3:695–701, 2003.

A. G. Renehan, S. T. O'Dwyer, N. J. Haboubi, and C. S. Potten. Early cellular events in colorectal carcinogenesis. *Colorectal Disease*, 4:76–89, 2002.

L. Ricci-Vitiani, D. G. Lombardi, E. Pilozzi, M. Biffoni, M. Todaro, C. Peschle, and R. De Maria. Identification and expansion of human colon-cancer-initiating cells. *Nature*, 445:111–115, 2007.

S. Ro and B. Rannala. Methylation patterns and mathematical models reveal dynamics of stem cell turnover in the human colon. *Proceedings of the National Academy of Sciences U.S.A.*, 98:10519–10521, 2001.

I. Roeder and M. Loeffler. A novel dynamic model of hematopoietic stem cell organization based on the concept of within-tissue plasticity. *Experimental Hematology*, 30:853–861, 2002.

T. Roose, S. J. Chapman, and P. K. Maini. Mathematical models of avascular tumor growth. *SIAM Review*, 49:179–208, 2007.

M. H. Ross, G. I. Kaye, and W. Pawlina. *Histology, a text and atlas*. Lippincott, Williams & Wilkins, Philadelphia, 2003.

P. Rozen, G. P. Young, B. Levin, and S.J. Spann. *Colorectal cancer in clinical practice*. Martin Dunitz, London, 2002.

E. Sancho, E. Batlle, and H. Clevers. Live and let die in the intestinal epithelium. *Current Opinion in Cell Biology*, 15:763–770, 2003.

O. J. Sansom, K. R. Reed, A. J. Hayes, H. Ireland, H. Brinkmann, I. P. Newton, E. Batlle, P. Simon-Assmann, H. Clevers, I. S. Nathke, A. R. Clarke, and D. J. Winton. Loss of *APC in vivo* immediately perturbs Wnt signaling, differentiation, and migration. *Genes and Development*, 18:1385–1390, 2004.

I-M. Shih, T-L. Wang, G. Traverso, K. Romans, S. R. Hamilton, S. Ben-Sasson, K. W. Kinzler, and B. Vogelstein. Top-down morphogenesis of colorectal tumors. *Proceedings of the National Academy of Sciences U.S.A.*, 98:2640–2645, 2001.

O. M. Sieber, K. Heinimann, P. Gorman, H. Lamlum, M. Crabtree, C. A. Simpson, D. Davies, K. Neale, S. V. Hodgson, R. R. Roylance, R. K. S. Phillips, W. F. Bodmer, and I. P. M. Tomlinson. Analysis of chromosomal instability in human colorectal adenomas with two mutational hits at *APC*. *Proceedings of the National Academy of Sciences U.S.A.*, 99:16910–16915, 2002.

O. M. Sieber, K. Heinimann, and I. P. M. Tomlinson. Genomic instability - the engine of tumorigenesis? *Nature Reviews Cancer*, 3:701–708, 2003.

A. Simon and J. Frisén. From stem cell to progenitor and back again. *Cell*, 128:825–826, 2007.

S. L. Spencer, M. J. Berryman, J. A. Garcia, and D. Abbott. An ordinary differential equation model for the multistep transformation to cancer. *Journal of Theoretical Biology*, 231:515–524, 2004.

S. L. Spencer, R. A. Gerety, K. J. Pienta, and S. Forrest. Modeling somatic evolution in tumorigenesis. *PLoS Computational Biology*, 2:0939–0947, 2006.

R. M. Summers, J. Yao, P. J. Pickhardt, M. Franaszek, I. Bitter, D. Brickman, V. Krishna, and J. R. Choi. Computed tomographic virtual colonoscopy computer-aided polyp detection in a screening population. *Gastroenterology*, 129:1832–1844, 2005.

R. W. Taylor, M. J. Barron, G. M. Borthwick, A. Gospel, P. F. Chinnery, D. C. Samuels, G. A. Taylor, S. M. Plusa, S. J. Needham, L. C. Greaves, T. B. L. Kirkwood, and D. M. Turnbull. Mitochondrial DNA mutations in human colonic crypt stem cells. *The Journal of Clinical Investigation*, 112:1351–1360, 2003.

I. Tomlinson and W. Bodmer. Selection, the mutation rate and cancer: ensuring that the tail does not wag the dog. *Nature Medicine*, 5:11–12, 1999.

I. Tomlinson, P. Sasieni, and W. Bodmer. How many mutations in a cancer? *American Journal of Pathology*, 160:755–758, 2002.

I. P. M. Tomlinson and W. F. Bodmer. Failure of programmed cell death and differentiation as causes of tumors: some simple mathematical models. *Proceedings of the National Academy of Sciences U.S.A.*, 92:11130–11134, 1995.

I. P. M. Tomlinson and W. F. Bodmer. Modelling the consequences of interactions between tumour cells. *British Journal of Cancer*, 75:157–160, 1997.

I. P. M. Tomlinson, M. R. Novelli, and W. F. Bodmer. The mutation rate and cancer. *Proceedings of the National Academy of Sciences U.S.A.*, 93:14800–14803, 1996.

E. Trucco. Mathematical models for cellular systems. The von Foerster equation. Part I. *Bulletin of Mathematical Biophysics*, 27:285–304, 1965a.

E. Trucco. Mathematical models for cellular systems. The von Foerster equation. Part II. *Bulletin of Mathematical Biophysics*, 27:449–471, 1965b.

J-L. Tsao, Y. Yatabe, R. Salovaara, H. J. Järvinen, J-P. Mecklin, L. A. Aaltonen, S. Tavaré, and D. Shibata. Genetic reconstruction of individual colorectal tumor histories. *Proceedings of the National Academy of Sciences U.S.A.*, 97:1236–1241, 2000.

M. van de Wetering, E. Sancho, C. Verweij, W. de Lau, I. Oving, A. Hurlstone, K. van der Horn, E. Batlle, D. Coudreuse, A-P. Haramis, M. Tjon-Pon-Fong, P. Moerer, M. van den Born, G. Soete, S. Pals, M. Eilers, R. Medema, and H. Clevers. The β-catenin/TCF-4 complex imposes a crypt progenitor phenotype on colorectal cancer cells. *Cell*, 111: 241–250, 2002.

J. H. van Es, M. E. van Gijn, O. Riccio, M. van den Born, M. Vooijs, H. Begthel, M. Cozijnsen, S. Robine, D. J. Winton, F. Radtke, and H. Clevers. Notch/γ-secretase inhibition turns proliferative cells in intestinal crypts and adenomas into goblet cells. *Nature*, 435: 959–963, 2005.

I. M. M. van Leeuwen, H. M. Byrne, O. E. Jensen, and J. R. King. Crypt dynamics and colorectal cancer: advances in mathematical modelling. *Cell Proliferation*, 39:157–181, 2006.

I. M. M. van Leeuwen, H. M. Byrne, M. D. Johnston, C. M. Edwards, S. J. Chapman, W. F. Bodmer, and P. K. Maini. Modelling multiscale aspects of colorectal cancer. In Kamel Ariffin Mohd Atan and Isthrinayagy S. Krishnarajah, editors, *International Conference on Mathematical Biology 2007*, volume 971, pages 3–7. American Institute of Physics, 2008.

J. von Neumann. *The theory of self-reproducing automata.* University of Illnois Press, 1966.

J. P. Ward and J. R. King. Mathematical modelling of avascular-tumour growth. *IMA Journal of Mathematics Applied in Medicine and Biology*, 14:39–69, 1997.

J. P. Ward and J. R. King. Mathematical modelling of avascular-tumour growth II: Modelling growth saturation. *IMA Journal of Mathematics Applied in Medicine and Biology*, 16:171–211, 1999a.

J. P. Ward and J. R. King. Mathematical modelling of the effects of mitotic inhibitors on avascular tumour growth. *Journal of Theoretical Medicine*, 1:287–311, 1999b.

R. A. Weinberg. Using maths to tackle cancer. *Nature*, 449:978–981, 2007.

D. Wodarz. Effect of stem cell turnover rates on protection against cancer and aging. *Journal of Theoretical Biology*, 245:449–458, 2007.

Y. Yatabe, S. Tavaré, and D. Shibata. Investigating stem cells in human colon by using methylation patterns. *Proceedings of the National Academy of Sciences U.S.A.*, 98:10839–10844, 2001.