MPHYS PROJECT

Evolutionary dynamics of cancer initiation

August 22, 2020

9013483
University of Manchester
School of Physics & Astronomy
Thomas J Leyshon

Contents

INTRODUCTION

Cancer is the consequence of an unwanted evolutionary process. Cells receive mutations that alter their phenotype and functionality [1]. A normal cell turns into a rogue cell at the start of all types of cancer. The cell begins to proliferate uncontrollably, spreading into surrounding tissues [2]. This uncontrolled dividing leads to the development of solid lumps (tumors). Cancer development is understood to be a multistage process. Present day oncology recognizes three main phases of cancer growth: *initiation, promotion and progression* [2].

A Mathematical approach to understanding Cancer

The initiation, promotion and progression of cancer is a consequence of somatic evolution [3]. Cells acquire successive mutations, mutated cells acquire an increased reproduction rate, and over time this leads to a population dominated by mutants [3]. Mathematical models have, and will continue to help build our understanding on the ways in which cancer evolves. They do so by providing insight into questions such as: how long does it take for a population of N cells to accumulate mutations? Can particular compositions of parameter values lead to stable points in mutation space that isn't cancer initiation? What is the average fixation time for a mutated species for different scenarios?

The application of mathematics to examine the evolution of cancer began in the 1950s [3]. Nordling, Armitage and Doll began using stochastic models for carinogenesis with the aim to understand the age dependence of human cancer [4]. Nordling concluded that around seven mutations fitted the age distribution for a range of human cancers [5]. Armitage and Doll further developed this early genetic model. Concluding that the epidemiological data was consistent with cancer developing in steps. This multi-stage theory of carcinogenesis [1] [6] still required an understanding of how many individual genetic events initiated cancer? Knudson's two-hit hypothesis supplied an answer, as it described the mechanics behind the loss of tumor suppressor genes, implying that both alleles[2] that code a particular protein must be mutated before a cancer-like effect is expressed [7]. Our current understanding of cancer growth is built on these findings. More recently, theoretical developments on the understanding of cancer initiation has come from attempts to produce mathematical machinery that capture the stochastic nature and stability of the mutation selection dynamics.

[1] the initiation of cancer is a sequence of successive individual genetic events

[2] each of two or more alternative forms of a gene that arise by mutation and are found at the same place on a chromosome.

Basic Biology behind Cancer initiation

In order to develop a mathematical framework that accurately describes the evolutionary dynamics of cancer; it is important to highlight and distinguish a few key biological properties of cancerous cells.

- Cancer cells possess a selfish behavior reminiscent of unicellular organisms, therefore mutated cell will have a superior fitness (growth rate) than the preceding species[3].
- Cancer susceptible genes are known as tumor suppressor genes. The normal function of these are to preserve genetic integrity, control cell interactions and cell growth rates. These functions are lost upon mutation [8].
- Cancerous cells undergo clonal expansion, that is, the production of many daughter cells all arising from a single cell [9].

CHAPTER 1: OVERVIEWS

Foundations for mathematical modeling

Cancer is a complex and highly dynamic process [10]. It is important to bear in mind the intrinsic trade-off between simplicity and accuracy when developing a model of such a complex system. However, the scope of the model and it's ability to describe cancer initiation accurately is essential most important. In this project, only the most crucial mechanics of cancer initiation were considered and formulated.

Deterministic models & numerical methods

A deterministic model assumes complete certainty. They are comprised of differential equations which describe the time dependent evolution of a process [11]. Deterministic models do not account for 'noise' contributions to the behavior. However, their predictability provide a strong null basis to build stochastic models upon. In this report a numerical method[3] was used to provide approximate solutions to deterministic equations.

[3] A Runge Kutta method to the 4^{th} order. The Runge Kutta method uses a trial step at the midpoint of an interval to cancel out lower-order error terms.

Markov chain processes

The birth death model is the foundation of the stochastic models considered in this project. Figure 1 shows an illustration of a birth death process. It is a Markov chain with successive transitions from one distinct state to another. If the chain is currently in state s_i, then it moves to state s_j at the next step with a probability denoted by p_{ij}. This probability does not depend upon which states the chain was in before the current state [12].

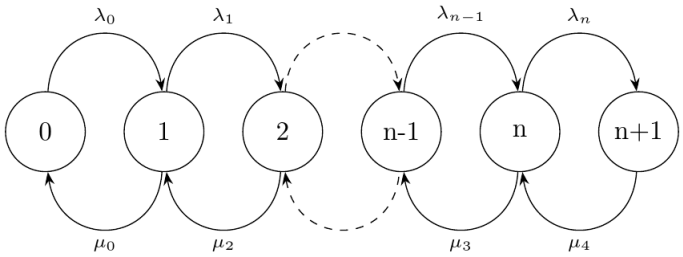

Figure 1: A birth death process, with transition rates λ and μ

The Wright-Fisher & Moran processes

The Wright-Fisher process is a discrete time Markov chain that describes genetic drift[4] of a population with constant size N. In a Wright-Fisher process, cell generations are assumed to be non-overlapping, so that many birth and death events occur during each time step but the overall population stays constant [13]. The Wright-Fisher model results in a decay of genetic variation. Exact calculations of the Wright - Fisher model with mutations are often impossible, although diffusion approximations can be used to find calculations. The Wright Fisher process was implemented in the replications of Haeno *et al* (2013) [14].

The Moran process is a variant of the Wright - Fisher model, and commonly used to study evolution in well mixed finite populations [1]. Unlike the Wright-Fisher model the Moran process is a continuous time model. Exact calculations of the Moran model are often possible. The Moran model was implemented in the replications of Ashcroft *et al* (2015). Both of these processes capture the intrinsic randomness in population genetics.

[4]Genetic drift is a random selection mechanism for reproduction and death

Simulation method & Gillespie algorithm

The Gillespie algorithm is a standard method used for simulating stochastic processes [11]. The method used was as follows. A process is discretized into an ensemble of states. All possible elemental transitions have a probability associated to them, which together sums to 1. These probabilities give the weighting of each transition, indicating which is most likely to occur next. One is chosen at random, this is the next transition to occur in the process. The time that elapses, τ until this next transition is exponentially distributed and determined via

$$\tau = -\frac{1}{\lambda} ln(r) \tag{1}$$

where $\lambda = \sum_{i=1}^{M} W_{n \Rightarrow m_i}$, the sum of all possible transition probabilities per unit time from the current state. r is a random number between 0 to 1 [12].
Appendix A shows the divergence of the numerical method and the stochastic simulation used for a general mutation model.

Two mutation models

Several researchers have studied the population dynamics of two mutations arising sequentially in a population of a fixed finite number of cells [15]. Such models give insight into Kundson's two hit rule and subsequently the dynamics of both tumorigenesis (long time scales) and the inactivation of a tumour suppressor gene (short time scales) [16].

Foundations for a two mutation model

Haeno *et al* used a birth death model with three variations of mutated cell [14], consisting of an original wild type (type-0), a single mutated species (type-1) and a double mutated species (type-2). Their populations are denoted as n_0, n_1, and n_2, respectively. Time was measured in cell cycles, and each cell has one opportunity to divide per day. During each time unit, a cell division and death event occurs, keeping a constant total population. A cell is chosen to divide at random proportional to their fitness, a cell is also randomly chosen to die out of the total population [14]. The probabilities of the system are shown below

$$Pr_{(n_0,n_1,n_2)\Rightarrow(n_0+1,n_1,n_2)} = \frac{r_0 n_0 (1-u_1)}{(r_0 n_0 + r_1 n_1 + r_2 n_2)} \qquad Pr_{(n_0,n_1,n_2)\Rightarrow(n_0-1,n_1,n_2)} = \frac{n_0}{N} \tag{2}$$

$$Pr_{(n_0,n_1,n_2)\Rightarrow(n_0,n_1+1,n_2)} = \frac{r_0 n_0 u_1 + r_1 n_1 (1-u_2)}{(r_0 n_0 + r_1 n_1 + r_2 n_2)} \qquad Pr_{(n_0,n_1-1,n_2)\Rightarrow(n_0,n_1-1,n_2)} = \frac{n_1}{N} \quad (3)$$

$$Pr_{(n_0,n_1,n_2)\Rightarrow Pr(n_0,n_1,n_2+1)} = \frac{r_1 n_1 u_2 + r_2 n_2}{(r_0 n_0 + r_1 n_1 + r_2 n_2)} \qquad Pr_{(n_0,n_1,n_2)\Rightarrow Pr(n_0,n_1,n_2-1)} = \frac{n_2}{N} \quad (4)$$

with initial conditions of $n_0 = N$ and $n_1 = n_2 = 0$. Where r_i is the fitness of type i and u_i is the mutation rate of type i to type $i+1$. The model presented in (Haeno *et al*, 2013) considers the dynamics of a fixed sized heterogeneous population in which two alterations may emerge [14]. Back mutations are assumed not possible[5]. Haeno *et al*, studied the effects of changing the relative fitnesses, r_i between each species. Replications of these effects were produced.

The effects of varying relative fitness on the type 2 mutant population & evidence of stochastic tunneling

Figure 2 (see next page) shows plots of type 2 population concentration, $\frac{n_2}{N}$ against type 1 fitness, r_1 for a given time period. The figure highlights the existence of optimal values of r_1 which maximizes the type 2 population. The optimal r_1 values differ depending on the fitness landscape. **Part a)** presents a fitness landscape with an advantageous type 0 fitness, $r_0 > r_2$. It can be seen for $u_2 = 0.1$ trajectory that the optimal r_1 value for this scenario is ≈ 1, which is between the values of r_0 and r_2. **Part b)** presents a neutral fitness landscape, $r_0 = r_2$. It is clear from this plot that as type 2 mutation rate, u_2 increases, both type 2 concentration and the optimal r_1 value also increases. **Part c)** presents a fitness landscape with an advantageous type 2 fitness, $r_2 > r_0$. The optimal value of r_1 is again seen to exist between r_0 and r_2 values. The optimum r_1 in this scenario also increases with u_2. Fixation is reached for $u_2 = 0.1$ trajectory when r_1 is of a similar value to r_2. **Part d)** shows the dependence of type 2 population on type 1 mutation rate u_1. The trends show an initial increase of u_1 causes an increase in type 2 population, further u_1 increase has no influence and the concentration plateaus, indicating independence of u_1 after a certain value. An increased u_2 is seen to shift the trends towards type 2 fixation.

[5] human genome is very large, so the probability of mutating a specific base $\approx 10^{-11}$ [15]

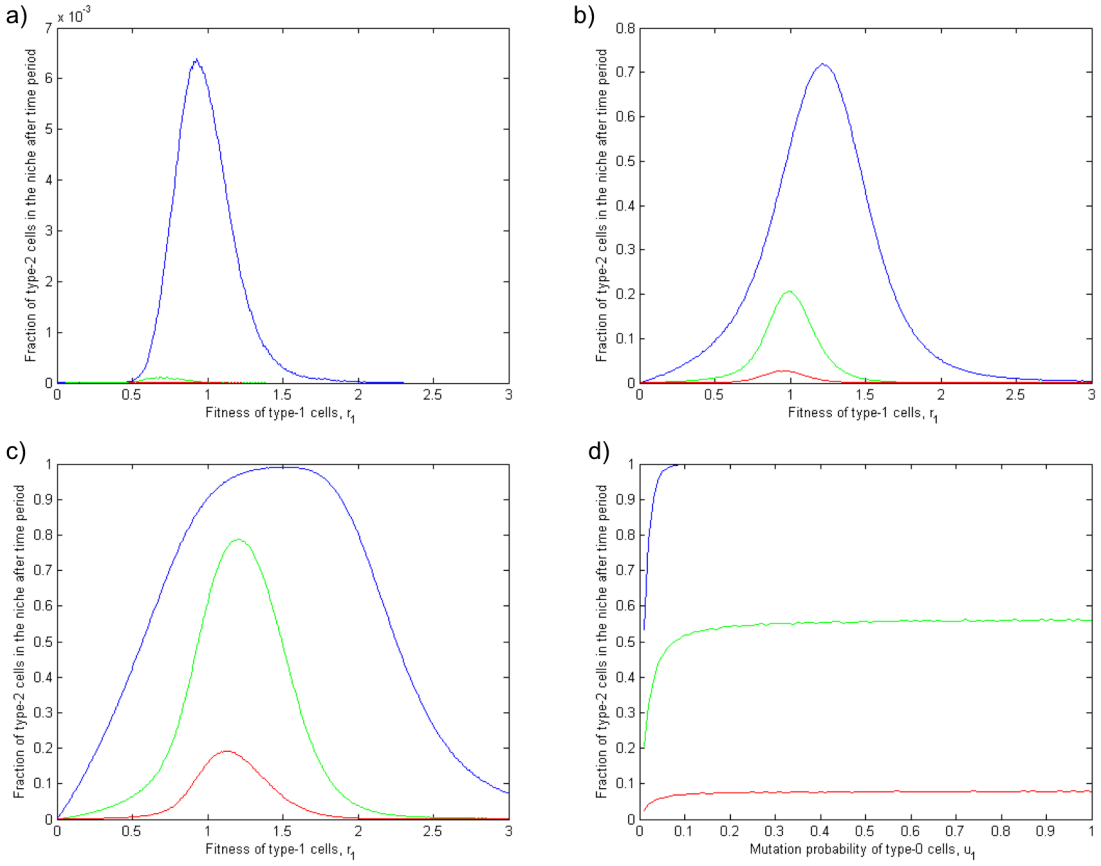

Figure 2: Parts a),b), c) and d) show the fraction of type 2 cells in the population after time period t, against type 1 fitness, r_1 for a),b),c) and against type 1 mutation rate, u_1 for d). Parameter values are: $N = 19$, $t = 100$ days, $r_0 = 1$, $u_1 = 0.01$. $u_2 = 0.001$ for red, 0.01 for green, 0.1 for blue. For a) $r_2 = 0.5$, b) $r_2 = 1$ c) $r_2 = 1.5$. d) $r_2 = 1$, $u_2 = 0.01$

Interpretations

Figure 2 indicates that a **mutation selection trade off** exists in maximizing the type 2 concentration. As r_1 increases from zero, the type 2 population rises from zero until the optimum r_1 is met, further r_1 increase then reduces type 2 concentration to zero. A physical explanation for this is that, r_1 needs to be of substantial size to increase the chance of initial type 2 mutations occurring. However, very large r_1 will result in the majority of selected division cells to be type 1. The optimal value of r_1 produces the most beneficial selection mutation composition for type 2 growth. In general, the **optimum r_1 value for non-neutral fitness landscapes** (see part a) and c)) **lies between that of r_0 and r_2**. For a relatively neutral mutation rate landscape, this r_1 range optimizes the stochastic tunneling rate to type 2 fixation. **Comparisons of part a) and c)** in figure 2 suggests that whilst between r_0 and r_2

values, the optimum r_1 is also **skewed towards the r_2 value** for a given u_2 trajectory. Part a), b) and c) all show r_2 causing progression of r_1 optimum. **Part b)** shows that **increased u_2 also skews the optimal r_1 towards higher values**. Larger proportion of type 1 divisions will now mutate to type 2 cells. The rapid increase seen in **Part d)** corresponds to the direct relationship of u_1 value with the abundance of type 1 mutants. A larger pool of type 1 mutants increases the probability of producing type 2 mutants. However, once wild type extinction has occurred, a larger u_1 does not increase type 2 population. This is represented by the plateau of the trends. The findings indicate the multi-dependence of the tunneling rate on the system's fitness and mutation parameter. The existence of an optimum r_1 value for a given scenario and a mutation selection trade off between these parameters has been shown.

Metastable states

Ashcroft *et al*, (2015) analyzed potential existence of quasi equilibria due to particular mutation selection balances were studied. Replications of the research presented in Ashcroft *et al* (2015) were produced. To do so, a Moran process with 2 successive mutant species was used. Transition equations were defined and the deterministic equations formulated. Fixed points associated with 5 distinct regions of phase space were identified, stability analysis was used to distinguish the type of fixed points that existed, and interpretations made.

Transition equations & formulating the master equation

For a Moran process with the potential for 2 mutation types, 6 elementary transition rates per a time step[6] collectively describe the population dynamics in full. u_i represents specie $i-1$'s tendency to mutate into specie i, r_i is specie i's fitness, the average fitness of the total population is denoted as $\bar{r} = \frac{(r_0 n_0 + r_1 n_1 + r_2 n_2)}{N}$. Transition rates consist of a particular species dying and another reproducing, $T^{i \Rightarrow j}$ is the death of type i and a type j daughter cell [15].

$$T^{0\Rightarrow 1} = \frac{((u_1) r_0 \frac{n_0}{N} + (1-u_2) r_1 \frac{n_1}{N}) n_0}{\bar{r}} \qquad T^{1\Rightarrow 0} = \frac{(1-u_1) r_0 \frac{n_0}{N} n_1}{\bar{r}}$$

$$T^{0\Rightarrow 2} = \frac{((u_2) r_1 \frac{n_1}{N} + r_2 \frac{n_2}{N}) n_0}{\bar{r}} \qquad T^{2\Rightarrow 0} = \frac{(1-u_1) r_0 \frac{n_0}{N} n_2}{\bar{r}}$$

$$T^{1\Rightarrow 2} = \frac{(u_2 r_1 \frac{n_1}{N} + r_2 \frac{n_2}{N}) n_1}{\bar{r}} \qquad T^{2\Rightarrow 1} = \frac{((u_1) r_0 \frac{n_0}{N} + (1-u_2) r_1 \frac{n_1}{N}) n_2}{\bar{r}}$$

[6] measured in cell generations

The master equation describes a continuous time process. It is the collective expression of the transition rates and determines the evolution of the probability that the system is in state n at time t. It is defined as

$$\dot{P}_n(t) = \sum_m [T^{m \Rightarrow n} P_m(t) - T^{n \Rightarrow m} P_n(t)] \tag{5}$$

The probability of being in state n at time t is the sum of the total influx of probability into state n minus the total outflux of probability from state n per unit time [12].

Derivation of deterministic equations

Using the van Kampen's system-size expansion on the master equation one can derive the deterministic equations of the system, seen in 6. A brief overview of how this was done is supplied in appendix B. x_i refers to the relative concentrations of cells of type i in the population, $x_0 + x_1 + x_2 = 1$

$$\dot{x_0} = \frac{[(1-u_1)r_0 - \bar{r}]x_0}{\bar{r}} \qquad \dot{x_1} = \frac{u_1 r_0 x_0 + [(1-u_2)r_1 - \bar{r}]x_1}{\bar{r}} \qquad \dot{x_2} = \frac{u_2 r_1 x_1 + (r_2 - \bar{r})x_2}{\bar{r}} \tag{6}$$

Fixed points, stability & phase diagrams

Within this model an obvious fixed point (FP) exists when $x_2 = 1$. This is the absorbing state. However, non trivial FPs can also exist away from the absorbing state [15]. There are two types of non trivial FPs that exist in this Moran process, these are interior ($x_i > 0$) and boundary FPs ($x_0 = 0$). As $x_0 + x_1 + x_2 = 1$, the system can be reduced to two variables $(1 - x_1 - x_2, x_1, x_2)$.

- Boundary FPs

  Boundary FPs are defined as having no wild type cells, $x_0 = 0$ and are of the form

$$\bar{r}\dot{x}(x*) = \begin{bmatrix} 0 \\ [(1-u_2)r_1 - \bar{r}]x* \\ u_2 r_1 x_1 + (r_2 - \bar{r})x_2* \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \tag{7}$$

  substituting in the condition for boundary FPs $x_2* = 1 - x_1*$ and rearranging for x_1 gives the boundary FP locations

$$x_1* = 1 - \frac{u_2 r_1}{r_1 - r_2} \qquad x_2* = 1 - x_1* \tag{8}$$

considering the requirement for a non-trivial FPs is $0 < x_1* < 1$, leads to the condition for non trivial boundary FPs to exist

$$0 < 1 - \frac{u_2 r_1}{r_1 - r_2} < 1 \qquad\qquad r_1 > \frac{r_2}{1 - u_2} > r_2 \tag{9}$$

The system's Jacobian matrix was evaluated in order to gain information about the stability of existing boundary FPs and the behaviour near them

$$J(x_1*, 1 - x_1*) = \frac{1}{\hat{r}} \begin{bmatrix} -u_1 r_0 - (r_1 - r_0)(1 - \frac{u_2 r_1}{r_1 - r_2} & -u_1 r_0 - (r_2 - r_0))(1 - \frac{u_2 r_1}{r_1 - r_2}) \\ u_2 r_1 - (r_1 - r_0)\frac{u_2 r_1}{r_1 - r_2} & r_2 - r_1(1 - u_2) - (r_2 - r_0)(\frac{u_2 r_1}{r_1 - r_2}) \end{bmatrix} \tag{10}$$

The determinant and trace of this Jacobian govern the stability of a boundary FP in the system and are given by

$$\Delta = \frac{[(1 - u_2) r_1 - r_2][(1 - u_2) r_1 - (1 - u_2) r_0]}{(1 - u_2)^2 r_1^2} \qquad \tau = \frac{(1 - u_1) r_0 + r_2 - 2(1 - u_2) r_1}{(1 - u_2) r_1} \tag{11}$$

$\tau < 0$ if $(1 - u_2) r_1 > (1 - u_1 r_0)$. Considering the condition for non trivial boundary FPs (equation 9) the stability of existing boundary FPs were determined. Boundary FPs exist at

$$x = \begin{bmatrix} 0 \\ 1 - \frac{u_2 r_1}{r_1 - r_2} \\ \frac{u_2 r_1}{r_1 - r_2} \end{bmatrix} \tag{12}$$

and are **stable** if $(1 - u_2) r_1 > (1 - u_1) r_0$ and are **saddle nodes** if $(1 - u_2) r_1 < (1 - u_1) r_0$.

- Interior FPs

  Interior FPs have the property that all $x_i > 0$. Simultaneously equating the deterministic equations and using the property $x_i > 0$ gave the locations of interior FPs to be

$$x* = \begin{bmatrix} 1 - \frac{[(1-u_1) r_0 + u_2 r_1 - r_2] u_1 r_0}{u_2 r_1 (r_0 - r_2) + (r_0 - r_1)[(1-u_2) r_0 - r_2]} \\ \frac{[(1-u_1) r_0 - r_2] u_1 r_0}{u_2 r_1 (r_0 - r_2) + (r_0 - r_1)[(1-u_1) r_0 - r_2]} \\ \frac{u_1 u_2 r_0 r_1}{u_2 r_1 (r_0 - r_2) + (r_0 - r_1)[(1-u_1) r_0 - r_2]} \end{bmatrix} \tag{13}$$

  and the determinant and trace were found to be

$$\Delta = \frac{[(1-u_1)r_0 - r_2][(1-u_1)r_0 - (1-u_2)r_1]}{(1-u_1)^2 r_0^2} > 0 \quad \tau = \frac{(1-u_2)r_1 - (1-u_1)r_0}{(1-u_1)r_0} < 0 \quad (14)$$

leading to the **existence conditions for interior FPs** to be $(1-u_1)r_0 > (1-u_2)r_1$ and $(1-u_1)r_0 > r_2$. These interior FP existence conditions means that **all interior FPs are stable**.

A numerical method was used to find approximations to the deterministic equations. The stability analysis reveals 3 existence conditions which identifies 5 distinct regions of parameter space. Phase diagrams were produced to show the deterministic flow in these 5 regions, 3 of which contain non-trivial FPs.

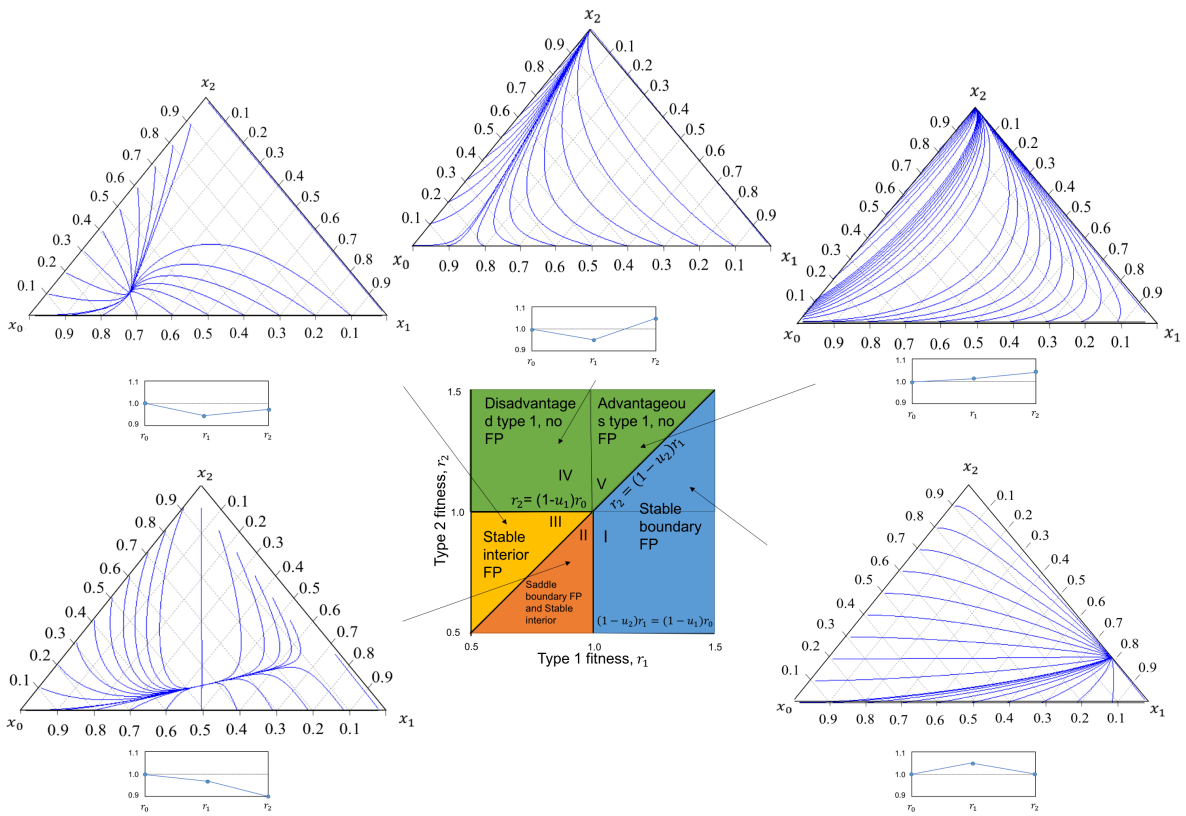

Figure 3: Replication of Phase diagram with satellite diagrams showing deterministic flow, in Ashcroft *et al* (2015). Regions 1 to 5 are labeled in the diagram, green = no FPs, yellow = stable interior FPs, blue = stable boundary FPs, orange = saddle boundary/stable interior FPs. The equations in the diagram govern the region boundaries. The relative fitnesses of each region are provide under each satellite diagram.

Region 1 presents an advantageous r_1 scenario. It can be seen that **a boundary FP** ($x_0 = 0$) exists, corresponding to **the mutation selection balance between type 1 and type 2 species**. Stability analysis determined that a FP on the boundary, $x_0 = 0$, which satisfies $(r_2 < (1-u_2)r_1)$ is stable. So once the wild type species is extinct the dynamics become trapped and the system maintains a mixed population of concentration $x_1 \approx 0.2$, $x_2 \approx 0.8$. **Region 2** presents a fitness landscape of $r_0 > r_1 > r_2$. The region contains an **interior FP** corresponding to **the mutation selection balance between all three species**, this is met when $x_0 \approx x_1 \approx 0.45$. There also exists a boundary FP. However, due to r_0 being the maximum fitness this is a **saddle FP** [15].

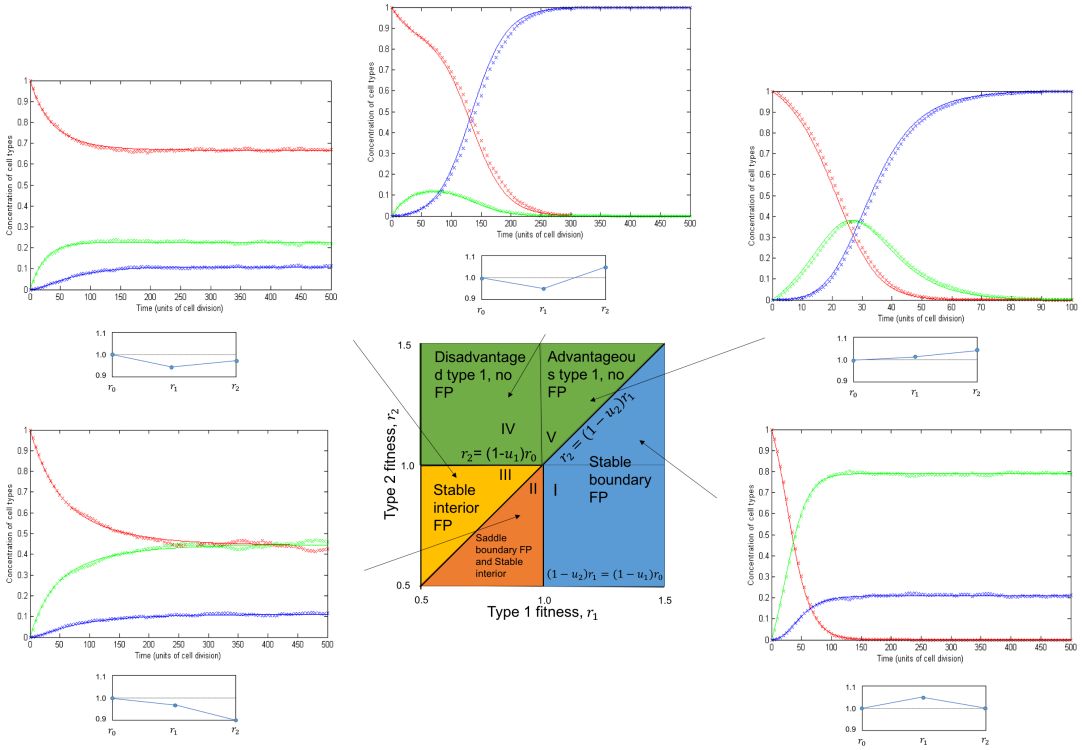

Figure 4: Phase diagram with satellite diagrams of stochastic simulation & deterministic approximation. Regions 1 to 5 are labeled in the diagram, green = no FPs, yellow = stable interior FPs, blue = stable boundary FPs, orange = saddle boundary/stable interior FPs. The equations in the diagram govern the region boundaries. The relative fitnesses of each region are provide below each satellite diagram. The satellite diagrams plot time vs population concentration, red = x_0, green = x_1 and blue = x_2.

Region 3 presents a fitness landscape of $r_0 > r_2 > r_1$, here a **single interior FP** exists when $x_0 = 0.7$. This FP corresponds to the bottleneck effect of a low type 1 fitness keeping the majority of the population as wild types. **Region 4 & 5** present scenarios of **advantageous**

type 2 fitness. These regions contain **no FPs**. As both selection and mutation processes make type 2 production preferential, subsequently no balance is met and the system tunnels to the absorbing state, type 2 fixation. **Region 4** shows a **disadvantaged type 1** and thus a more direct tunneling from to type 2 fixation state. **Region 5** shows an **advantaged type 1** and subsequently a more successive evolution.

Figure 4 shows the time evolution of each specie concentration for each distinct region of phase space. The time dependence of the specie population trajectories due to different fitness landscapes. **Region 1** shows a rapid transit to a large type 1 concentration until wild type extinction is reached, balance is met. **Region 2** is seen to have the slowest transition to a metastable state, this is due to the region having the overall lowest fitness landscape. **Region 3** r_1 is the minimum fitness and creating a bottleneck effect resulting in a relatively fast transition to a majority type 0 equilibrium. **Region 4 & 5** present fixation of type 2 mutants, a rise and fall of the type 1 population is observed in both regions as tunneling from type 0 to type 2 fixation takes place. **region 4** represents a **fast tunneling rate**, this is represented by the type 1 population slumping much lower than that of region 5, as well as region 4 reaching fixation quicker.

CHAPTER 2: COLORECTAL CANCER

Justification

There is a growing body of evidence supports the idea that human cancers are a stem cell disease [17]. In tumour development, mutations are said to accumulate over many years. However, tissue renewal is typically on a time scale of days [18]. Suggesting tumorigenic mutations must occur in long living and self replicating cells. Stem cells have two defining properties; self-renewal and the ability to generate differentiated cells, pluripotency [19]. These properties enable cancerous stem cells to clonally expand [18].

Colorectal cancer is the second biggest cause of death from any cancer in the United States [17]. A human colon crypt contains about 2,000 cells, these are replenished by a small number of stem cells (around 4 to 6) [20].

Foundations of the stem cell crypt model

The basic repeated building block of the colon lining is known as the crypt [19]. This is a tube like gland. The crypt's function in the colon is to make mucus and renew the lining of the intestine [21]. The crypt is an organized structure where both differentiated cells and stem

cells coexists, as illustrated in figure 5 a). These stem cells are confined to a small niche and tightly regulated [19].

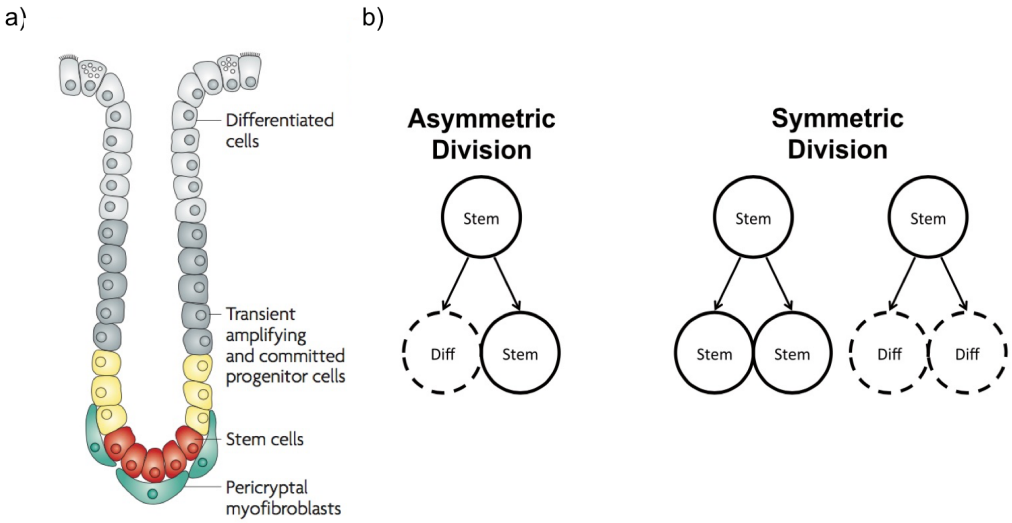

Figure 5: Illustrations of a) The organized structure of a colon crypt [19]. b) Possible stem cell divisions [22]

Differentiated cells, once produced, migrate to the higher region to perform their cell's function. The differentiated cell's populations are replenished via the population of pluripotent stem cells [19]. In addition to differentiated division, stem cells also self replicate. Stem cells can divide both symmetrically and asymmetrically with respect to daughter cell type, this is illustrated in figure 5 b) [22].

Stem cells regulate the relative cell population of the niche, making their division process dependent on the current environment. The population of stem cells per niche is still uncertain. However it has been claimed to contain around 19 stem cells [17]. The stem cell's division process aims to regulate around this average. Cancerous stem cells lose their environment dependence. They self reproduce regardless of the current stem cell population or niche capacity [23]. A model was produced describing the population growth of wild type stem cells with the potential for n consecutive mutated stem cell species.

Transition equations & formulating the master equation

10 elementary transition equations were used to describe the population dynamics

$$
\begin{aligned}
s_i \Rightarrow s_i s_i &= x_i r_i \sigma_i \gamma_i (1-u_{i+1}) & s_i \Rightarrow dd &= x_i r_i \sigma_i (1-\gamma_i)\\
s_i \Rightarrow s_i d &= x_i r_i (1-\sigma_i)(1-u_{i+1}) & s_i \Rightarrow s_{i+1} s_i &= x_i r_i \sigma_i \gamma_i u_{i+1}\\
s_i \Rightarrow s_{i+1} d &= \frac{1}{2} x_i r_i (1-\sigma_i) u_{i+1} & s_{i-1} \Rightarrow s_{i-1} s_i &= x_{i-1} r_{i-1} \sigma_i \gamma_i u_i\\
s_{i-1} \Rightarrow s_i d &= \frac{1}{2} x_{i-1} r_{i-1} (1-\sigma_{i-1})(1-u_{i-1}) & s_{i-1} \Rightarrow s_{i-1} d &= x_{i-1} r_{i-1} (1-\sigma_i)(1-u_{i-1})\\
s_{i-1} \Rightarrow dd &= x_{i-1} r_{i-1} \sigma_{i-1} (1-\gamma_{i-1}) & s_{i-1} \Rightarrow s_{i-1} s_{i-1} &= x_{i-1} r_{i-1} \sigma_{i-1} \gamma_{i-1} (1-u_i)
\end{aligned}
$$

where s_i is the stem cell population under consideration and s_{i+1} represents the next consecutive mutated population. d represents the differentiated daughter cells. The model describes the stem cell dynamics, therefore the transition $s_i \Rightarrow dd$ is interpreted as a stem cell death of species i. r_i and u_i represents the fitness and mutation rate respectively. σ_i represents the probability of symmetric over asymmetric division of species i. γ_i is a feedback mechanism that regulates the population size. γ_i is defined as the probability to produce two stem cell daughters over two differentiated cells. γ_i was chosen to be

$$\gamma_i = 1 - (\frac{N}{N+N_w})^{C_i} \tag{15}$$

where N is the total number of stem cells in the crypt, N_w is the crypt capacity, the 'desired' stem cell population, 19. C can vary between specie types. By choosing C_0 to be 1, the stem cells modulate their population around Nw.

A master equation can be formulated from all possible transition rates into and out of a system at a time t for a duration δt. A Taylor expansion to first order in δt and dividing through by δt gave the master equation a final form of

$$\partial_t x_i \bar{r} = x_i r_i a_i + x_{i-1} r_{i-1} b_{i-1} \tag{16}$$

where the coefficients a_i and b_i are given by

$$a_i = \sigma_i [2\gamma_i - u_{i+1}(\gamma_i - \frac{1}{2}) - 1] - \frac{u_{i+1}}{2} \qquad b_{i-1} = u_i [\sigma_{i-1}(\gamma_{i-1} - \frac{1}{2}) + \frac{1}{2}]. \tag{17}$$

as x_0 has no influx, the rate equation reduces to

$$\partial_t x_0 = x_0 \bar{r_0} a_0 \tag{18}$$

The effects of varying relative parameters

Figure 6 shows 4 plots of the average time before the first type 2 mutant appears (appearance time) against type 1 fitness r_1, for varying σ_0, σ_1, u_1 and u_2.

Part a) shows the effects of a varying σ_0 value. It indicates that as σ_0 increases the quicker a type 2 mutant will appear. The influence of σ_0 varying is seen to be more prominent at low r_1. At large type 1 fitness the σ_0 varying trajectories converge, suggesting appearance time becomes independent of σ_0. As r_1 increases the proportion of type 0 cell chosen for division decreases, reducing the influence of σ_0. An explanation for these observations is that when $\sigma_0 = 0$, this corresponds to all type 0 divisions being asymmetric, thus effectively halving the mutation rate per a division (as for every division a 50% chance of mutating the differentiated cell). From another view point, $\sigma_0 = 1$ corresponds to all symmetric divisions. Therefore, for any mutation to occur an additional σ_0 will be produced, increasing the type 0 population and reducing the probability of type 1 divisions/mutation selection.

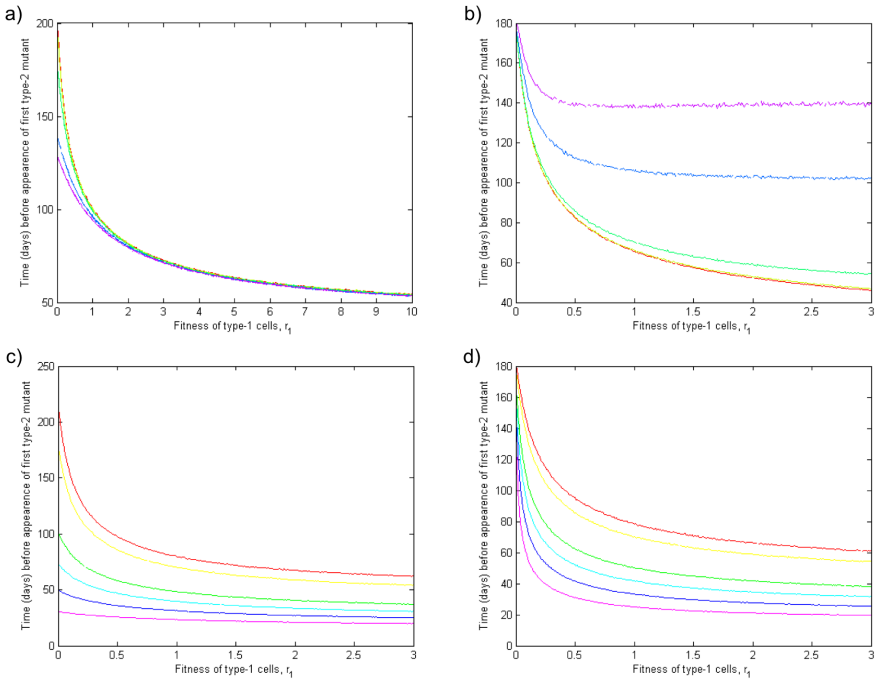

Figure 6: Time until first type 2 mutant appears, 'appearance time', against type 1 fitness, r_1 with $N = 19$, $r_0 = 1$, $r_2 = 1$ a) varying $\sigma_0 = 0$, 0.005,0.05,0.5 and 1, b) varying $\sigma_1 =$, 0, 0.005,0.05,0.5 and 1 c) $u_2 = 0.01$, varying $u_1 = 0.1, 0.05, 0.01, 0.03, 0.02, 0.008$ and d) $u_1 = 0.01$ varying $u_2 = 0.1$, 0.05, 0.01, 0.03, 0.02, 0.008

Part b) shows the effects of varying σ_1, the probability of symmetric division for type 1

mutants. The plot shows that as σ_1 increases the longer the appearance time is. These results agree with that of Shahriyari *et al* (2013), type 1 symmetric divisions delay cancer onset. An explanation of this is when $\sigma_1 = 0$, corresponding to all type 1 divisions being asymmetric, no type 1 mutants can die, over time producing a higher concentration of type 1 mutants and subsequently a type 2 mutant is more likely. The varying σ_1 trends diverge with increasing r_1. As more type 1 mutants will be chosen for division, increasing the influence of σ_1 variance. Higher values of σ_1 lead to a more even birth death process for stem cells and therefore plateau earlier. **Part c)**[7] presents the influence of u_1 on the appearance time. As u_1 increases the appearance time exponentially decreases from infinity then flattens off. The effects of u_1 variance decreases as r_1 increases, converging and flattening off. As large r_1 lowers the proportion of type 0 divisions and therefore limits the influence of u_1. **Part d)** presents the appearance time's dependence on u_2. The expected correlation between a higher mutation rate resulting in a quicker appearance time is observed. However, the plot also shows the plateau and convergence of u_2 varying trends with an increased r_1. This convergence is due to the required time to generate the type 1 mutants from wild type population, which does not dependent on u_1. This, therefore, produces a minimum appearance time for the scenario. As r_1 increase the difference between varying u_2 lessen as they all converge onto this finite minimum appearance time.

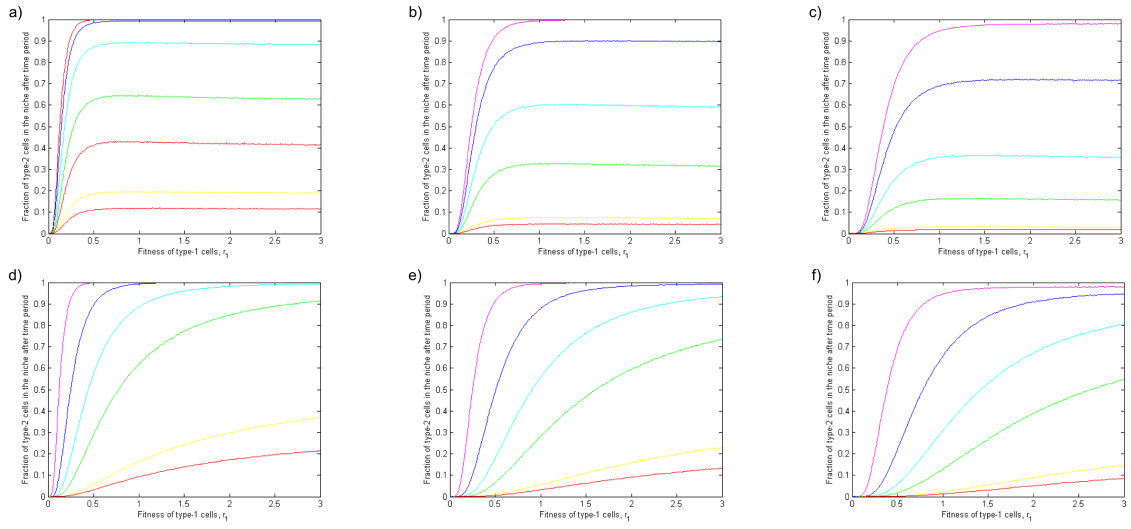

Figure 7: Type 2 population concentration against r_1 for $N = 19$, time = 200 days, $\sigma_0 = \sigma_1 = 0.05$ for a) $r_2 = 1.5$, u_1 , b) $r_2 = 1$, u_1 c) $r_2 = 1.5$,u_1 d) $r_2 = 1.5$,u_2 e) $r_2 = 1$, u_2 f) $r_2 = 0.5$, u_2 ... varying u_i values were 0.008,0.02, 0.03, 0.01, 0.05, 0.1 (red to purple) with $u_j = 0.01$

[7] trajectories don't go to infinity because the simulation didn't start at $r_1 = 0$

Stability analysis determined that no FPs other than the absorbing state exists in the stem cell model. Therefore, an increase in the parameters r_1, u_1, u_2 will never result in a decrease in type 2 concentration. Figure 7 shows the fraction of type 2 mutants against type 1 fitness with varying u_1 and u_2 for different r_2 values. Figure 7 clarifies increase in these parameters never leads to a decrease in type 2 concentration, ensuring no non trivial FPs exist.

One dimensional spatial model of the crypt

Within the crypt cells migrate outwards [1]. Nowak *et al* (2013) provide a model of this upwards stream of cells within a crypt, as illustrated in figure 8. A single stem cell at the bottom of the crypt maintains the homeostasis of the crypt by reproducing asymmetrically to a differentiated cell and a stem cell [1]. Cells are chosen to divide based on fitness. Upon division the two daughters replace both the parent and the next position in the array, figure 8 illustrates this process. A division will shift the last cell out of the array. This represents the displacement of a cell from the top of the crypt, undergoing apoptosis[8]. Nowak *et al* (2003) derived equations for the fixation probability due to a mutant stem cell. Expanding on Nowak et al's analysis, the time dependence of the probability for the mutant to leave the crypt was studied. Here the approximation that mutations are rare was made, $u << 1/N$. This approximation states no other type 0 cells will mutate within the time scale of the current mutant fixation time. To gain the time dependence of the probability the Poisson distribution for rates dependent on the current state must be derived.

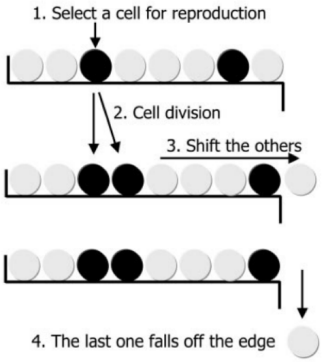

Figure 8: Linear model, cells are labeled i = 1,..., N. Dividing cells are replaced by two daughter cells, pushing all cells to the right by one position. The last cell undergoes apoptosis [1].

[8] cell death which occurs as a normal and controlled part of an organism's development.

Derivation of Poisson distribution with state dependent rates

State i in this linear model can only move to one other state with a rate that is dependent on the current state, λ_i. A recurring theme was found in the solutions of the master equation for the probability distribution, $P_{i\to n}$ to see the system in any state $i+n \geq i$ at time t. This general theme was found by inspection than proved to hold in all cases. Solutions to the master equation are of the form

$$P_{i\to i+n} = \frac{\prod_{j\prime=0}^{n-1}\lambda_{i+j\prime}}{\lambda_{i+1}-\lambda_i}\sum_{j=0}^{n} A_{j,n}e^{-\lambda_{i+j}t} \tag{19}$$

where the coefficients $A_{j,n}$ are given by

$$A_{0,n} = \prod_{l=2}^{n}\frac{A_{k,1}}{\lambda_{i+l}-\lambda_i} \qquad A_{k,n} = \prod_{l=k+1}^{n}\frac{A_{k,k}}{\lambda_{i+l}-\lambda_{i+k}}, k = 1,2,...,n-1 \qquad A_{n,n} = -\sum_{l=0}^{n-1}A_{l,n} \tag{20}$$

the systems initial condition at the initial time $t = 0$ set $A_{0,1} = 1$, $A_{1,1} = 0$. Equation 20 states the probability to be in any state other than i at $t = 0$ is zero.

Proof

To prove that this equation is valid, it must satisfy the master equation

$$\frac{dP_{i+n}}{dt} = \lambda_{i+n-1}P_{i+n-1} - \lambda_{i+n}P_{i+n}. \tag{21}$$

Inserting equation 19 into equation 21 produces

$$-\prod_{j\prime=0}^{n-1}\lambda_{i+j\prime}\sum_{j=0}^{n}A_{j,n}\lambda_{i+j}e^{-\lambda_{i+j}t} = \prod_{j\prime=0}^{n-2}\lambda_{i+j\prime}\lambda_{i+n-1}\sum_{j=0}^{n-1}A_{j,n-1}e^{-\lambda_{i+j}t} - \lambda_{i+n}\prod_{j\prime=0}^{n-1}\lambda_{i+j\prime}\sum_{j=0}^{n}A_{j,n}e^{-\lambda_{i+j}t}. \tag{22}$$

where λ_{i+n-1} on the RHS is incorporated in the product and divided out. The $j = n$ terms for the two summations up to and including n can then be removed from their respective summations These will cancel, leaving

$$\sum_{j=0}^{n-1}A_{j,n}\lambda_{i+j}e^{-\lambda_{i+j}t} = \sum_{j=0}^{n-1}(\lambda_{i+n}A_{j,n} - A_{j,n-1})e^{-\lambda_{i+j}t} \tag{23}$$

now by equating the $j = 0$ terms from the summation as well as the remaining terms yields two equations. For the proof to hold these must be satisfied

$$A_{0,n}\lambda_i = \lambda_{i+n}A_{0,n} - A_{0,n-1} \qquad A_{j,n}\lambda_{i+j} = \lambda_{i+n}A_{j,n} - A_{j,n-1}, j = 1,2,...,n-1 \tag{24}$$

divided through by the exponential in both cases, and in equation 44 the contents of the remaining summation has been equated. Equations 43 and 44 can be proven to be true by using equations 38 and 39 respectively. This is shown below for equation 44. Substituting in equation 39 creates the proof condition

$$\prod_{l=j+1}^{n} \frac{A_{j,j}}{\lambda_{i+l} - \lambda_{i+j}} \lambda_{i+j} = \prod_{l=j+1}^{n} \frac{A_{j,j}}{\lambda_{i+l} - \lambda_{i+j}} \lambda_{i+n} - \prod_{l=j+1}^{n-1} \frac{A_{j,j}}{\lambda_{i+l} - \lambda_{i+j}} \tag{25}$$

cancelling terms $\prod_{l=j+1}^{n-1} \frac{A_{j,j}}{\lambda_{i+l}-\lambda_{i+j}}$ gives

$$\frac{\lambda_{i+j}}{\lambda_{i+n} - \lambda_{i+j}} = \frac{\lambda_{i+n}}{\lambda_{i+n} - \lambda_{i+j}} - 1 \tag{26}$$

which is true, and therefore the solution holds.

Probability for a mutation to leave the crypt

A cell in the position i can only move to position $i+1$. $i = 1$ represents the stem cell position, and $i = N$, where N is the total number of cells in the crypt. The probability distribution derived only require the rates functional forms. These functional forms depend on the current position i in the crypt. Recording when the position of the furthest mutant into the crypt is eliminated will give the probability that the entire mutant species is eliminated from the crypt.

For the deepest mutant at position i at time $t = 0$ to moved to position $i+1$ any cell before position i must divide. Only these divisions prior to the mutant will effect the mutants position. If a cell has divided somewhere in the crypt, there is a probability

$$P_{i \to i+1} = (i-1)\frac{r_0}{r_s} \tag{27}$$

where $r_s = \sum_{i=0}^{1} N_i r_i$, with r_0 and r_i being the relative fitness of the wild-type and type 1 mutant cells, and N_i is the i^{th} species population. $\lambda_{i \to i+1} = r_s P_{i \to i+1}$ defines the rate at which a cell positioned before the deepest mutant divides. Therefore after time t, the position

of the mutant is j places further towards the villus tip than i. The rate to move to the next position is $\lambda_{i+j} = (i-1+j)r_0$. Shifting the product index to $k = i-1+j$ and substituting these rates into equation 36 gives

$$P_{i\to i+n} = r_0^{n-1} \prod_{k=i-1}^{i+n-2} k \sum_{j=0}^{n} A_{j,n} e^{-(i-1)r_0 t} \tag{28}$$

The probability to leave the crypt is now the probability to not be in any state $j = i, i+1, ..., N$. Equation 49 shows that $P_{i\to i+n}$ is not affected by mutant fitness as a mutant division doesn't increase the position of the deepest mutant, as illustrated in figure 8, step 1 to 2.

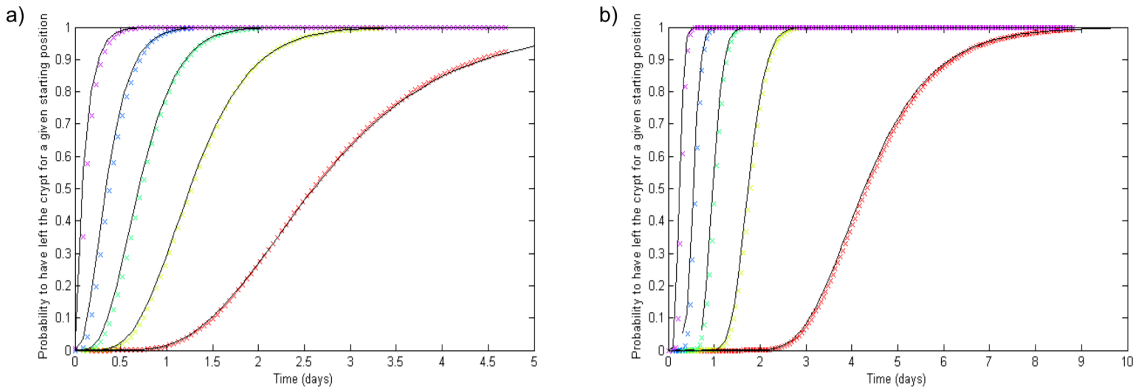

Figure 9: a) Total population of $N = 10$ and parameters of $r_0 = 1, r_1 = 1$ and starting at positions $i = 2,4,6,8,10$, b) total population of $N = 50$ and parameters of $r_0 = 1$, $r_1 = 1.5$ and starting at positions $i = 2,4,6,8,10$

Probability distributions for a variety of crypt set-ups were compared. Figure 9 shows the probability for the mutant to become extinct at time t, a) for parameters $N = 10$ and $r_0 = r_1 = 1$ b) a crypt size $N = 50$ with $r_0 = 1$ and $r_1 = 1.5$. Crosses on the plot represent the stochastic simulation data while the solid line is the plot of the probability distribution as seen in equation 28. The plots prove that solutions for rate dependent Poisson master equation holds for all cases with the approximation of rare mutations.

DISCUSSION & CONCLUSION

The first section of this project attempted to give a concise overview of how to begin developing a useful model of cancer initiation. Replications of the analysis of Haeno *et al* (2013) and Ashcroft *et al* (2015) provided evidence of the existence of optimal r_1 values for stochastic tunneling, and a mutation selection trade off when maximizing type 2 population. Stability analysis on deterministic equations of the Moran process revealed 5 distinct regions of parameter space, 3 of which contained metastable states relating to mutation selection balances. The findings produced agreed with that of the original studies. Two physical interpretations of two mutation models can be made; when the time until type 2 fixation is relatively long, this scenario models the inactivation of tumour suppressor genes. When fixation time is small this scenerio models tumorgenesis. To mimic the conditions of chromosomal instability, large mutation rates ($10^{-4} < u < 10^{-1}$) were used [24].

The second section covered the development of a model to describe the evolutionary dynamics of cancerous stem cells within a colonic crypt. A master equation was formulated, the effects of varying parameters on type 2 appearance time and concentration were distinguished. It was observed that symmetrically dividing stem cells resulted in longer times until type 2 appearances. This agrees with Shahriyari *et al* (2013), indicating that symmetric divisions can have a cancer delaying effect [22]. Both stability analysis and simulation data showed the model to have no other FPs apart from type 2 fixation. A spatial model of the crypt as presented in Nowak *et al* (2003) was built upon, the probability for a mutant to leave the crypt with respect to time was determined.

Further development could be by undertaking analysis of the effects of varying the feedback mechanism, γ in the stem cell model on type 2 appearance time and fixation.

References

[1] Nowak MA, Michor F, Iwasa Y. The linear process of somatic evolution. Proceedings of the national academy of sciences. 2003;100(25):14966–14969.

[2] Devi PU. Basics of carcinogenesis. Health Adm. 2004;17(1):16–24.

[3] Michor F. Evolutionary dynamics of cancer. The Department of Organismic and Evolutionary Biology, Harvard University; 2005.

[4] Moolgavkar SH, Venzon DJ. Two-event models for carcinogenesis: incidence curves for childhood and adult tumors. Mathematical Biosciences. 1979;47(1):55–77.

[5] Nordling C. A new theory on the cancer-inducing mechanism. British journal of cancer. 1953;7(1):68.

[6] Armitage P, Doll R. A two-stage theory of carcinogenesis in relation to the age distribution of human cancer. British journal of cancer. 1957;11(2):161.

[7] Knudson AG. Two genetic hits (more or less) to cancer. Nature Reviews Cancer. 2001;1(2):157–162.

[8] Weinberg RA, et al. Tumor suppressor genes. Science. 1991;254(5035):1138–1146.

[9] Greaves M, Maley CC. Clonal evolution in cancer. Nature. 2012;481(7381):306–313.

[10] Vicente-Dueñas C, Cobaleda C, Pérez-Losada J, Sánchez-García I. The evolution of cancer modeling: the shadow of stem cells. Disease Models and Mechanisms. 2010;3(3-4):149–155.

[11] Erban R, Chapman J, Maini P. A practical guide to stochastic simulations of reaction-diffusion processes. arXiv preprint arXiv:07041908. 2007;.

[12] Galla T. PHYS40571-Advanced Statistical Physic notes, University of Manchester, 2016: p.5-189;.

[13] Ewens WJ. Mathematical Population Genetics 1: Theoretical Introduction. vol. 27. Springer Science & Business Media; 2012.

[14] Haeno H, Maruvka YE, Iwasa Y, Michor F. Stochastic tunneling of two mutations in a population of cancer cells. PloS one. 2013;8(6):e65724.

[15] Ashcroft P, Michor F, Galla T. Stochastic tunneling and metastable states during the somatic evolution of cancer. Genetics. 2015;199(4):1213–1228.

[16] Mader C. The Biology of Cancer. The Yale journal of biology and medicine. 2007;80(2):91.

[17] Khalek FJA, Gallicano GI, Mishra L. Colon cancer stem cells. Gastrointestinal cancer research: GCR. 2010;(Suppl 1):S16.

[18] Boman BM, Wicha MS. Cancer stem cells: a step toward the cure. Journal of clinical oncology. 2008;26(17):2795–2799.

[19] Gatenby RA, Gillies RJ, Brown JS. The evolutionary dynamics of cancer prevention. Nature Reviews Cancer. 2010;10(8):526–527.

[20] Nicolas P, Kim KM, Shibata D, Tavaré S. The stem cell population of the human colon crypt: analysis via methylation patterns. PLoS Comput Biol. 2007;3(3):e28.

[21] Van der Flier LG, Clevers H. Stem cells, self-renewal, and differentiation in the intestinal epithelium. Annual review of physiology. 2009;71:241–260.

[22] Shahriyari L, Komarova NL. Symmetric vs. asymmetric stem cell divisions: an adaptation against cancer? PLoS One. 2013;8(10):e76195.

[23] Humphries A, Wright NA. Colonic crypt organization and tumorigenesis. Nature Reviews Cancer. 2008;8(6):415–424.

[24] Lengauer C, Kinzler KW, Vogelstein B. Genetic instability in colorectal cancers. Nature. 1997;386(6625):623–627.

APPENDIX A

Numerical method & stochastic simulation divergence

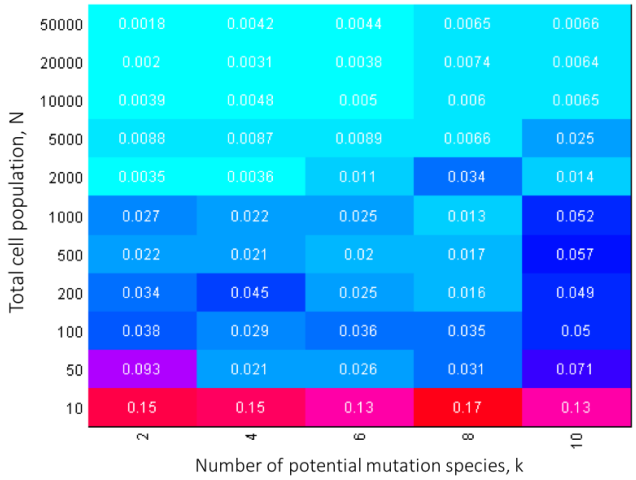

Figure 10: The divergence of stochastic simulation and numerical method trajectories for a cell population, N with k potential mutation species. For $k = 2,4,6,8,10$ and $N = 10,50,100,200,500,1000,2000,5000,10000,20000,50000$. Dark Red represents the largest divergence, while light blue represents the smallest.

Figure 10 presents the divergence of the numerical method and stochastic simulation trajectories for a population with various values of potential mutation species, k and total population, N. It can be seen that as the total population, N is increased the divergence decreases for all k values. As the number of potential species, k increases the overall divergence increases. For smaller populations the numerical method and deterministic equations will be less representative of the corresponding stochastic process.

APPENDIX B

System size expansion

Firstly, assume the variance of a species population, n probability distribution scales with the total population, N. Therefore n is governed by the sum of a deterministic term and a random term, $n = Nx(t) + \sqrt{N}\xi$, where ξ is a random variable. The probability distribution of n can therefore be expressed in terms of ξ

$$P(n,t) = P(Nx + \sqrt{N}\xi) = \Pi(\xi, t) \tag{29}$$

and therefore

$$\partial_t P(n,t) = \partial\Pi(\xi, t)\, t \tag{30}$$

taking the time derivative of n yields

$$\partial_t \xi = -\sqrt{N}\dot{x} \tag{31}$$

combining equations 30 and 31 gives

$$\partial_t P(n,t) = \frac{\partial\Pi(\xi,t)}{\partial t} - \frac{\partial\pi(\xi,t)}{\partial\xi}\sqrt{N}\dot{x} \tag{32}$$

The master equation can also be written in the form of raising and lowering operators $\hat{E}$, where $\hat{E}f(n) = f(n+1)$ and $\hat{E}^{-1}f(n) = f(n+1)$ [12].

$$\dot{P}_n(t) = (\hat{E}^{-1} - 1)\sum_m T^{m\Rightarrow n} P(n,t) + (\hat{E} - 1)\sum_m T^{n\Rightarrow m} P(n,t) \tag{33}$$

Due to the previous assumption the operator's action can now be defined in terms of N and $\frac{d}{d\xi}$ [12]. The operation $\hat{E}f(n) = f(n+1)$ is the equivalent to increasing ξ by $\sqrt{N}$. and therefore $\hat{E}f(\xi) = f(\xi + \sqrt{N}) = \sum_l \frac{1}{l!}(\sqrt{N})^l \frac{\partial^l}{\partial\xi^l} f(\xi)$. By substituting this expand expression for $\hat{E}$ and equating the two master equations 13 and 14 a comparison of $\sqrt{N}$ coefficients yields the deterministic equations of the species n, seen in 6.