Evolutionary Dynamics of Cancer: Spatial and Heterogeneous Effects

by

Ali Mahdipour Shirayeh

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Applied Mathematics

Waterloo, Ontario, Canada, 2017

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner	Dr. DOMINIK WODARZ Professor of Theoretical/ Computational Biology, University of California, Irvine, USA
Supervisor(s)	Dr. MOHAMMAD KOHANDEL Associate Professor of Applied Mathematics Dr. SIVABAL SIVALOGANATHAN Professor of Applied Mathematics
Internal Member	Dr. ZORAN MISKOVIC Professor of Applied Mathematics Dr. EDWARD VRSCAY Professor of Applied Mathematics
Internal-external Member	Dr. LAURENT MARCOUX Professor of Pure Mathematics, Department of Pure Mathematics

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Despite significant advances in the study of cancer and associated combination therapeutic treatments, cancer still remains one of the most common and complex often-terminal diseases. Acquisition of high–throughput experimental data from diverse cellular perspectives has thrown light on some of the regulatory mechanisms underlying the development of cancer. However, in general there is a lack of a general pattern and coherent model which can explain the development and evolution of the disease. To this end, evolutionary dynamics has been used, as a mathematical tool, in numerous studies to model various aspects of cancer over time periods. Our main focus in this thesis is on the use of stochastic and statistic methods to study cellular interactions within cancer tissues in order to understand the role of spatial structure, heterogeneity, and the microenvironment in cancer development. By constructing multi-cellular structures and using both analytic calculations and stochastic simulations, we have investigated the phenotypic hierarchy of stem cells within a heterogeneous system and in the presence of environmentally induced plasticity. Moreover, the effect of a random environment on the development of cancer has been explored in a general framework. As an important application of the multi-stage hierarchical model, the structure of the colonic/intestinal crypt has been taken into account to show the crucial role of these stem cells in the initiation and progression of colorectal/intestinal cancer. From an alternative viewpoint, we have envisaged the hierarchy of mutations as an evolutionary mechanism in the context of acute myeloid leukemia and carried out a statistical analysis of genetic data. Our findings in this thesis are general and most likely have many implications across a wide array of fields including different blood and solid cancers, bacterial growth, drug resistance and social networks. Moreover, the introduced methods and analyses should have important applications in diverse branches of evolution, ecology, and population genetics.

Acknowledgements

First of all, I would like to express my sincere gratitude to my suppervisors Dr. Mohammad Kohandel and Dr. Sivaloganathan for their research guidance and scientific insight. My sincere thanks to Dr. Zoran Miskovic (Applied Mathematics Department, University of Waterloo) and Dr. Edward Vrscay (Applied Mathematics Department, University of Waterloo), Dr. Laurent Marcoux (Pure Mathematics Department, University of Waterloo), and Dr. Dominik Wodarz (University of California, Irvine) also on my thesis defence committee, for their advice and support. I really appreciate my collaborators Drs. Natalia Komarova, Kamran Kaveh, Leili Shahriyari, Amir Hossein Darooneh, Anthony Long, Liran Shlush and John Dick's Lab members for amazing discussions. I also would like to thank the Department of Applied Mathematics at University of waterloo and John Dick's Lab for their help and support. Thanks to the members of the Biomedical Research Group at the Department of Applied Mathematics of University of waterloo which made my graduate study as a great experience. Finally, thanks to my family for their extreme love and support.

Dedication

This is dedicated to my parents, Hassan and Khadijeh, and also my two brothers Amir Hossein and Akbar.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Cancer can be thought of as a dynamical disease that is initiated through microenvironmental effects and genetic variations in living organisms. This type of malignancy occurs through a multi-level process originating in mutations and epimutations during carcinogenesis (cancer generation) which alters the delicate homeostatic balance between cell reproduction and death. Among diverse diseases, cancer as a result of genetic/epigenetic alteration(s) on somatic cells has been widely studied because of its abundance, diversity, and high risk. Mutation (genetic alteration) is clearly one of the crucial factors leading to cancer and this may occur in different types of genes: caretaker, gatekeeper, and/or landscaper genes [110, 121] genes. In fact, genetic/epigenetic mutations have been thought of as one of the important halmarks of various cancers [51, 122]. More precisely, mutations are natural responses to microenvironmental challenges, often leading to the Darwinian evolution and adaptability of an organism. Mutants mostly have higher net reproduction rates [24]. Epigenetic changes are fundamentally associated to non–mutational events that affect the hierarchy of phenotypic-genotypic change in cells [81]. In most cases, cancer arises from multiple changes in cellular pathways of various phenotypes and thus can be subdivided into compartments of such phenotypes. In a healthy tissue, homeostasis maintains constant cell numbers in each of the compartments.

Cancer initiation has been thought of as a hierarchy of mutations, which are genetic changes that occur as a result of microenvironmental effects and heterogeneity (diversity of cell types) [4, 120, 154, 166, 185, 186]. Figure 1.1 represents how different mutational effects may lead to cancer initiation. In this figure, typical changes in the diversity of different subclones is given over a cancer initiation period. Diversity, in fact, is a notion of multiplicity for existing subclones within the population which may present phenotypically/genotypically different or may reveal diverse functionalities. Within tumorigenesis

(tumor initiation), the crucial and common effects of mutations on cells are related to uncontrolled proliferation, avoidance of cell death, avoidance of anti–growth signaling, sustaining angiogenesis, and metastasis to the targeted tissue in invasive types of tumors [23, 121]. There has been a long-standing discussion and debate on the three existing carcinogenesis models: clonal evolution [58], cancer stem cell hypothesis [81], and plastic cancer stem cell [51, 101]. Recent thinking suggests that the notion of a '*cancer initiating state*' also called a "*stemness state*", for a pool of stem cells (SCs) and certain non–SCs, might be a better interpretation of tumor development in the human body [81]. Stem cells, in fact, differentiate to produce progenitor cells, which give rise to fully differentiated cells. The reverse procedure, that is, the switch back from progenitor cells to stem cells, termed as plasticity or dedifferentiation, does not usually occur. However, there exists important recent evidence suggesting plasticity of cancer cells [101, 139, 147, 150, 164, 174]. Mathematical models have also been constructed to derive switching rates between these two states [48]. In the present thesis, mutation and plasticity are considered as possible underlying mechanisms in carcinogenesis.

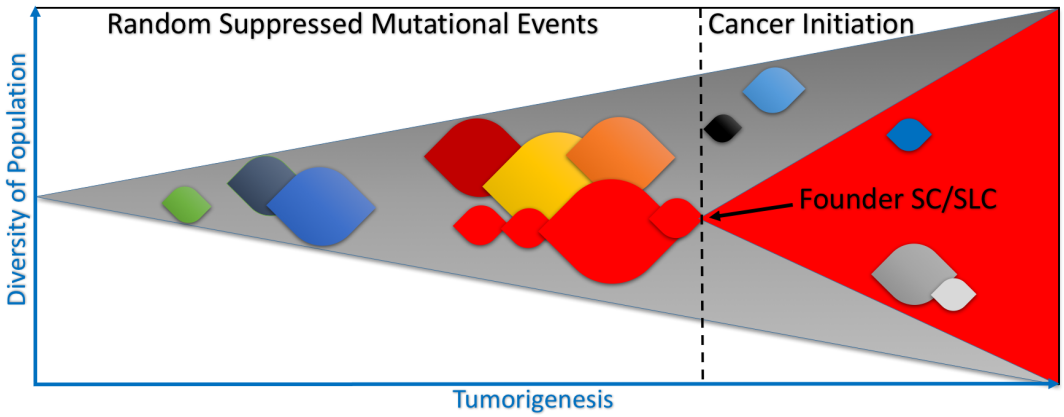


Figure 1.1: **Cancer initiation process.** Before the tumor initiation, many random mutations may occur within the population. Such mutational events lead to new clones and increase the heterogeneity (diversity) of the system. However, only a few of the clones can survive and initiate the cancer. The crucial mutation(s) which can trigger tumorigenesis are usually termed as *founder mutation(s)*: either as founder stem cells (SCs) or founder stem like cells (SLCs). SLCs are those non-stem cells which can phenotypically switch to a stemness state. Different suppressed or survived clonal expansions are shown with diverse colores.

To study the dynamics of cancer, there exists a variety of mathematical approaches.

Evolutionary dynamics is one of the approaches that has been played an important role to improve our understanding of such complex mechanisms. Considering cancer as a stochastic evolutionary process, Markov chains of finite or infinite populations have been used in evolutionary dynamics to analyze development of cancer. In this thesis, our main focus is on evolutionary dynamics of cancer. We assume different types of Moran processes [116, 121]: the process of competition between two types of individuals within a population of constant size. The birth–death (BD) and death–birth (DB) Moran models are two important mechanisms to study such dynamics. In a conventional (non-spatial) BD Moran model, a cell is chosen at random for proliferation first, while its offspring replaces another randomly chosen cell. In DB model, a cell is chosen at random for elimination first, then another random cell sends its offspring for substitution. These procedures occur proportional to the ability of each type of cell for reproduction, which is usually referred to as the fitness (proliferation rate). The fitness of each type of individual can be greater (or less) than that of the other types, then this type of cell would be called advantageous (disadvantageous). Applying evolutionary dynamics methods and considering finite Markov chains of finite dimension and finite variable numbers, two important types of cancer are studied in this thesis: colorectal/intestinal cancer and acute myeloid leukemia. We first summarize some fundamental biological facts about each of these cancers separately and explain how mathematical oncology can be used to probe and understand some aspects of tumorigenesis.

Firstly we study colon cancer which is the second and third most common cancer, respectively in women and men all over the world [32]. To start, we firstly review the structure of the human colonic/intestinal crypt and then focus on the cancer development process in such a tissue. There are millions ($\sim 10^7$) of crypts in the human colon [58]. In the epithelium, the colonic mucosa consists of an assay of a single-cell layer including stem–cells, progenitor cells, and fully differentiated cells. Precursor cells are gradually produced by intestinal stem–cells (ISC) as they evolve to turn into more mature levels; from early progenitor cells to fully differentiated cells. The rate of renewal in these organs is about 5 days [182] and thus the epithelium is one of the most rapidly renewing organs in the human body. Such a rapid renewal makes the crypt capable of hosting a wide range of malignancies. The high frequency of such malignancies makes colorectal cancer a fatal disease. The biology of cancerous colon/intestine is correlated with the natural homeostasis mechanisms at the bottom of the crypt or in the rest of the epithelium. Colon/intestinal cancer can also occur as a result of inheritable genetic and/or epigenetic factors. More explanation about the mechanism of a cancerous intestine can be found in [51, 58, 75, 110, 122, 123, 181, 182, 183].

In most colorectal cancers, mutation occurs over the patient's lifetime, and it is not

inherited. The most frequent mutation in colon cancer is the inactivation of the tumor suppressor gene adenomatous polyposis coli (APC) [78]. Moreover, the absence of APC immediately perturbs Wnt signalling pathways (the processes of receiving signals in cell surface receptors), and causes aberrant migration [149]. Another tumor suppressor gene that becomes inactivated in many cancers as well as in some colon cancers is p53 (often called '*the guardian of the genome*'). There is evidence that p53 inactivation leads to multiple mutations during a single cell cycle, like chromothripsis [59]. Chromothripsis is a massive genomic rearrangement occurring only on a single chromosome or a few chromosomes during a single cell cycle [203]. Moreover, in the majority of colorectal carcinomas, the loss of a large portion of chromosomes 17p and 18q has been observed [36]. These massive changes in the cell's genomes can evolve a cell to become immortal like the Hela cell line [113]. In addition, the over-expression of polycomb ring finger oncogene BMI1 transforms epithelial cells and influences telomerase functionality that can lead to cell immortalization [101, 201, 195]. Telomeres, repetitive nucleotide sequences at each end of all chromosomes, get slightly shorter with each normal cell division until they shorten to a critical length. This leads to cell aging and ultimately to apoptosis, or cell death [25, 40, 71, 84, 101, 171, 201]. Normal cells have a maximum number of divisions, i.e. '*Hayflick limit*', before these telomeres are depleted. Rare mutations or transformation events, highly associated with telomerase activity, allow cells to escape from the first mortality phase checkpoint (M1 or Hayflick limit) and the second M2 crisis checkpoint and become immortal [21].

Colon and intestinal cells are commonly divided into three categories: stem cells (SCs), transit amplifying (TA), and fully differentiated cells (FD). Colonic and intestinal stem cells, which are located at the bottom of the crypt, generate themselves and TA cells. TA cells, which are the most dividing cells in the crypt, give rise to the specialized differentiated cells that fulfill the physiological functions of the intestine [181]. Sometimes in response to an injury, TA cells that have some potential characteristics of stem cells can re-acquire some stem cell functionality and regenerate the crypt [181]. We do not model this scenario, and instead investigate the normal cell dynamics of crypts with no injuries.

Recently, it has been observed that the intestinal crypts contain two stem cell compartments: border stem cells (BSCs) and central stem cells (CeSCs) [144]. BSCs, which are located between CeSCs and transit amplifying (TA) cells, have high potential for differentiation; in contrast, the CeSCs, which are located at the very bottom of the crypt, are biased towards proliferation. The first computational cell dynamics model, which considered the existence of two stem cells groups, was developed by Shahriyari and Komarova [154] to obtain the probability of second-hit mutant production in the stem cell niche. Recently, by improving the model provided in [154], i.e. adding the possibility of cell mi-

grations from BSCs to CeSCs, we obtained the optimal structure for the stem cell niche, which minimizes the probability of the progeny of mutants taking over the entire stem cell niche [156]. Upgrading the existing models with mutational events and spatial structures [14, 51, 57, 86, 98, 101, 107, 137, 144, 184], the general and relevant mechanism in colorectal/intestinal cancer will be studied in this thesis.

Another important type of cancer which we study in this thesis is acute myeloid leukemia (AML). The dynamics of such a complex mechanism has still remained unclear even at steady state or even after the reconstitution of the blood [108]. High level of cell turnover at the reproduction rate of approximately 10^{12} cells per day occurs in the human body. For this purpose, progenitor cells which are raised from hematopoietic stem cells (HSCs) undergo the hierarchy of generating functionally matured hematopoietic cells. This hierarchy starts from HSCs and then to multi–potential SCs, then to lymphoid/myeloid progenitors and finally to white cells, red cells, and platelets. As HSCs cannot proliferate and differentiate *in vitro* and since still there is no accurate biomarker to distinguish SCs and non–SCs repopulated from the blood and bone marrow, there are many ambiguities in the study of HSCs.

Among various types of blood cancers, AML is the result of quick growth of affected myeloid cells tending to the accumulation of leukaemia blast cells in the blood and bone marrow. This type of leukemia is the most common hematological disease in adults with fewer occurrence of mutations [85]. The main source of leukaemia has still remained unclear but mostly arises via defection in transcription factor in the regulatory mechanism of homeostasis-related genes. According to the results of [85], on average 13 mutations observed in the genes of 200 clinical adult cases, in which 5 mutations occurred subsequently and in most of the cases, in which there was one potential driver mutation.

Overall, the pathogenesis of leukemia can be studied through different perspectives: the hierarchical method and clonal succession. There exists some evidence on both the hierarchical models of blood cancer formation, clonal succession and cancer stem cell hypothesis in the mouse blood [29, 68, 152]. Repopulation mechanism in mammals most probably occurs at different HSC self-renewal rates which makes hematopoiesis stochastic. This means that we may rely on numerical simulation and other mathematical approaches to understand human clonal kinetics [1, 108]. Such results then can be transferred to clinical trails and therapeutic treatments.

In spite of much research into therapeutic treatments and prognosis schedules, AML has endured as a case–dependent and high–death–rate malignancy [145]. The existing methods in prognosis, diagnosis, relapse, and therapy are related to genetic alterations which defines various subtypes of AML. The classical categories of AML, also called FAB

(French-American-British) classification, has been performed according to the morphology and maturation of hematopoietic cells [142]. Another classification of AML which has been defined by the World Health Organization (WHO) is based on the occurrence of dysplasia (the early stage in cancer development) and chromosome translocations [3]. However, we try to generalize these two classifications by considering phenotypic-genotypic features as new subtypes. Each of the subtypes is determined by a set of driver genes which have been thought to be genetically interconnected.

For the given list of the new AML subtypes, we perform gene expression analysis of AML data. Variations in genomes may occur to activate oncogenes or deactivate tumor suppressor genes, that can be observed over a range of information from raw data to classified data. Novel sequencing methods have intensely enhanced the biological and evolutionary understanding of hematopoiesis manipulating the analysis of genomes, exomes, transcriptomes, and proteins [145]. A comprehensive analysis of AML clinical/experimental samples would help us to map out the landscape of somatic mutations or the phylogeny of recurrent genetic alterations based on the signature of AML genes. Such a phylogenetics tree may provide worthwhile information about tumor initiation and progression, prognosis and chemotherapy as well as targeted–therapy, and also detecting other related genes to expand the tree. The cell of origin and those mutations that initiate and contribute to AML have been poorly identified. However, highly purified fractionation of HSCs, progenitor, and differentiated blood cells revealed that pre–leukemic cells which are resistant to chemotherapy and undergo a relapse, may be the origin of Leukemic cells in AML. More precisely, the appearance of DNMT3A mutation as an early event may play a prominent role in the evolution of pre–leukemic cells and in AML pathogenesis [159].

We investigate some crucial aspects of heterogeneity, spatial structure, and pathogenesis of various tumor types, which may have important effect on the fate of the system. The first aspect to study relates to different types of division which may occur at different rates depending on the inter/intra–cellular nature of the organ. Symmetric division produces two daughter cells of the same type where either both are the same as the ancestor or both different from the ancestor. An asymmetric division reproduces one cell of the same type as the ancestor and the other one of different type. Various types of cell division play important roles in the dynamics of cancer where a defect in one type of division can affect the whole mechanism.

The second aspect is associated with the dynamics of multistage cancer initiation, which has been studied in the literature for both well–mixed (independent of structure) and position–dependent conditions and by using stochastic analysis [75, 121, 122, 123]. In this thesis, we use mutation as an important factor that changes a specific genotype of individual to another genotype of cell. Considering the competition for life between two

genotypes, one may explore the hierarchy of mutations (in terms of tumor initiation and progression) as a Moran process [4, 75, 115, 121, 122, 120, 123, 124, 166].

Another important aspect in the evolution of cancers relates to spatial structure of the cancerous organism. In a spatially structured population, each individual within the population has a specific position and the mechanism is not necessarily the same as in the conventional Moran model [12, 39, 63, 75, 76]. Mathematically, the Moran model of a finite population can be assumed on a finite (di)graph where each of the individuals is located on a node of the graph whilst cellular interactions can only occur through the existing edges. There has been much effort in evolutionary dynamics to investigate various features of Moran models. The spatial Moran process for a structured population was firstly investigated by Maruyama [103]. Later, it was shown that the location of the new mutant is important in estimating the fixation probability, that is the probability of a newly arisen mutant to take over the entire population [90, 163].

Houchmandzade *et al* [54, 55] tried to give a general mechanism for the Moran model on graphs based on Marayuma's discoveries, but without repeating his simplifying assumptions. Their methods generalize spatial Moran models to consider islands over nodes of a finite graph and investigate the effect of such a pattern on the fixation probability. They emphasize the fact that for small selection pressures in DB models, the fixation probability is reduced for an advantageous mutant. Moreover, in BD models, the probability for an advantageous mutant to take over the whole system can exceed the neutral fixation probability for a non–structured population of the same size [54, 55]. A one-dimensional spatial generalization of the DB Moran process is discussed in [75].

Finally, the last aspect that we consider in this thesis is related to the application of the multi–compartment models. Our motivation for considering such structures is that the majority of human cancers have origins in epithelial tissues, that are tiny compartments of cells. In a multi–compartment model, each type of cell has its own particular compartment. The idea of using the multi–compartment models has mostly been used for a 1D spatial model. Combining all of the above aspects in a multi–compartment model, we attempt to understand those patterns leading to cancer. Particularly, for the case of colorectal/intestinal cancer, we generalize the existing models to more complex system with more structural details to conclude a better understanding of the cellular interactions among various compartments.

The thesis is organized as follows. Starting with some background materials about the evolutionary dynamics of structured and unstructured populations, in Chapter 2, we briefly review two important concepts: the fixation probability and fixation time, which will be used in subsequent chapters. In Chapter 3, the general form of a multi-compartmental

model is given where the impact of various types of division, in the form of asymmetric differentiation and self–renewal is compared with the effect of phenotypic plasticity in a general multi-stage framework. We consider the phenotypic and genotypic alterations to individuals and obtain the general formula for the survival probability of mutants when a malignant mutation occurs.

Then micro-environmental fluctuations on the fitness of mutants will be investigated in Chapter 4 to show how a neutral system can be affected under influence of randomness in the system and how such an impact may provide a selection advantage for mutants. We show that the minority enhances the fixation probability of mutants even though the normal cells are selectively advantageous. As the simplest case of the 1D spatial model, the line and circle have been studied under random effect. In the current research, we have generalized some ideas about 1D spatial models of the circle, the line, and complete graph in Chapter 4. Studying the effect of randomness on diverse 1D spatial models, we also show how such an environmental effect may influence the survivability of mutants on a 2D structure.

In Chapter 5, a general model of 4-compartmental mechanism is taken into account to study the probable structure of the colon/intestinal crypt, where two compartments of central and border stem cells are considered. We also examine the impact of immortal cells on the fate of mutants within the fully differentiated subpopulation. Much effort has been also expended in this chapter to characterize the role of different types of stem cells in colorectal/intestinal cancer and confirm and develop some very recent experimental results. The structure of the colonic crypt, different types of cells in epithelial layer, and their various types of division have been assumed in our models to quantify some important parameters for the first time and based on our analysis and existing experimental data.

Chapter 6 is devoted to the study of AML, in which statistical genetic analysis is used to investigate the lineage of hematopoietic stem cells within the initiation and progression of leukemia. Therefore in this chapter, the general mechanism of acute myeloid leukemia is studied along with some gene expression analysis of this type of blood cancer. Phylogenetics and the signature genes of AML are other important concepts which will be presented in this chapter.

Finally in Chapter 7, we summarize all the models described in the chapters 3-6 and suggest some new problems which may provide fruitful directions of research for future work and further endeavours.

Chapter 2

Mathematical Background Materials

2.1 Evolutionary Dynamics

Those mathematical principles of natural events in life that may repeatedly occur over time are the main topics of evolutionary dynamics. The origin of evolutionary dynamics formulation is genetic evolution while the genetic changes are considered over a long period of time. Mutation and proliferation are the main mechanisms underlying evolutionary dynamics. Mutation leads to generating various types through selection process whilst proliferation relates to the reproduction of copies of existing individuals.

The conventional Moran model is a dynamic condition between deterministic selection and stochastic model of generations. Finding the fixation probability as the probability that a beneficial mutant can take over the entire system (if applicable to spread), is an important quantity in evolutionary dynamics and population genetics. To attain the fixation probability, we usually assume that the number of individuals in a specific population is stochastic. Furthermore, this quantity is sensitive with respect to the time distribution. To investigate a finite Markov process, uniform distribution is usually assumed. In small populations, the probability of extinction is greater than that of larger populations due to the higher fluctuations in the system. Furthurmore, the fixation probability of a beneficial mutant in a large population is lower than that of in genetic drift. In large populations, stochastic variation represents a lower effect on extinction or survivability [56] changing the population size.

Among various methods in studying stochasticity of the evolutionary dynamics of a population, one of the simplest approaches is the Moran model in which two types of individuals compete within a population of constant size. The Moran model can be considered

for a well-mixed population in which any pair of individuals can randomly participate in birth/death events. However, the model can be assumed as a structured population (assuming to be located on vertices of a graph) where every individual has a specific position and birth/death occur under certain rules and in proportion to their fitnesses and the weight of the connecting edges. Weights are modeled with respect to the structure of the considered graph and presumably the relative fitness of individuals. Inhabitants in both of these models (structured and non-structured) can be exposed to selection pressure under random drift or neutral drift with no selection.

The most important parameters in population genetics, fixation probability and time to fixation, can be obtained by following the evolutionary dynamics of a newborn mutant until its absorption. Proliferation/death rates can be frequency dependent (game theory) or remain unchanged during the absorption process (in which the recently arisen mutant experiences extinction or fixation). However, in reality, these parameters may change due to some random effects and alter from location to location.

First, let us suppose a finite Markov chain for a random variable X which takes one (or other) of the values $0, 1, \cdots, N$ (N is the population size) respectively at time steps $0, 1, 2, \cdots$. We suppose the system is at state i when $X = i$. In a Markov process, the probabilities $P_{i\to j}$ of reaching from state i to state j, do not depend on time and the process at time $t + 1$ is dependent only on the procedure at time t. One question of interestt in population genetics is associated with the absorption of the system to either 0 or N.

2.1.1 Fixation Probability for a one dimensional Markov change

When a new mutant in a homogeneous population takes over the whole population, the probability that it successes is called the fixation probability. The fixation probability π_A of an individual A is, in fact, the probability that A takes over the whole population.

Let π_i be the probability of reaching state N starting from state i, when X takes its values at consecutive time steps, the π_i satisfy the following Kolmogorov system [41, **?**]

$$\begin{cases} \pi_i = \sum_{j=0}^{N} P_{i\to j}\, \pi_j, \\ \pi_0 = 0, \;\; \pi_N = 1. \end{cases} \tag{2.1}$$

The solution to this system is

$$\pi_i = \pi_1 \Big(1 + \sum_{j=1}^{i} \prod_{k=1}^{j} \gamma_k\Big), \qquad \pi_1 = \frac{1}{1 + \sum_{j=1}^{N-1} \prod_{k=1}^{j} \gamma_k},$$

for $\gamma_k = \dfrac{P_{i\to i-1}}{P_{i\to i+1}}$, where $P_{i\to i-1}$ and $P_{i\to i+1}$ are the transition probabilities of changing sate i to $i-1$ and $i+1$ respectively.

2.1.2 Fixation Probability for a higher dimesional Markov change

Now let us consider a higher dimensional Markov chain. For the sake of simplicity, one may consider a two dimensional random walk. Starting from an initial state (j,k), through the two dimensional random walk, where j and k are respectively the number of some specific subpopulations of two populations J and K with constant sizes $|J|$ and $|K|$. We follow the all possible changes in the number of j and k. We find the generalize form of the probability of fixation either in the J group, K compartment, and in the entire population of J and K. We assume that no more mutations would occur after placing an initial mutant within the population.

To envisage the fixation probability of starting from a state j in a 1-dimensional single variable Markov change, we define π_j as the fixation probability of staring from j mutants, i.e. π_j is the probability that the progeny of the j number of mutants taking over the entire system. Then, the fixation probability π_1 is the probability that the progeny of a single mutant will take over the entire population. Thus, we have the following system of equations:

$$\begin{cases} \pi_j = \sum_{m=0}^{N} P_{j\to m}\pi_m, \quad 1 < j < N-1, \\ \pi_1 = \sum_{m\geq 1}^{N} P_{1\to m}\pi_m, \\ \pi_{N-1} = P_{N-1\to N} + \sum_{m\leq N-1}^{N} P_{j\to m}\pi_m, \\ \pi_N = 1, \end{cases} \tag{2.2}$$

where $P_{j\to m}$ is the transition probability of moving from state j to state m and N is equal to either $|J|$ or $|K|$.

In higher dimensional state spaces, the procedure is similar while j is a vector, where each coordinate in this vector is associated to the number of mutants in its corresponding compartment. In the following, we illustrate a method to solve a bi–variable finite Markov chain to find the fixation probability of mutants starting from the state (j,k) of having j

mutant(s) in one of the compartments and k mutant(s) in the other compartment. This method can be generalized to higher dimensional state spaces with more than two variables.

We suppose that $\pi_{(j,k)}$ is the fixation probability of having j and k number of mutants in two different J and K compartments at the initial time. We may start from the state $(j=1,k=0)$ or $(j=0,k=1)$ which respectively associate to the case of having one initial mutant in compartments J or K, respectively. Moreover, we assume that $\pi^J_{(j,k)}$ is the fixation probability of mutants in the compartment J starting from the state (j,k). Thus $\pi^J_{(1,0)}$ and $\pi^J_{(0,1)}$ are respectively the fixation probability of an initial mutant in J and K in the J compartment. Similarly, the probability $\pi^K_{(1,0)}$ (or $\pi^K_{(0,1)}$) is the fixation probability in the K compartment starting from a mutant in the J (or K) group. In the following, we describe a model to calculate the fixation probability $\pi^J_{(1,0)}$ of starting from one mutation in the J compartment applying a two–variable Markov chain model moving through all possible states (j,k) where $(j,k)\neq(0,0)$ and $0\leq j\leq |J|-1, 0\leq k\leq |K|$. Overall, there exist $|J|(|K|+1)-1$ distinct states in this case. We define the $(|J|(|K|+1)-1)\times(|J|(|K|+1)-1)$ transition matrix A where the entries are the coefficients of $\pi^J_{j,k}$ in the following Kolmogorov equation:

$$\pi^J_{(j,k)}=\sum_{\tilde{j},\tilde{k}} P_{(j,k)\to(\tilde{j},\tilde{k})}\,\pi^J_{(\tilde{j},\tilde{k})}. \tag{2.3}$$

We also define $B=\Big[\,(j,k)\Big]_{0\leq k\leq|K|,0\leq j\leq|J|-1}$ as the matrix of all possible states (j,k). One may correspond the matrix B to a vector in $\mathbb{R}^{|J|(|K|+1)-1}$, considering the below isomorphism ignoring the entry $(0,0)$ in B is not taken into account:

$$\begin{cases} F: B\longrightarrow \mathcal{S}\subseteq\mathbb{R}^{|J|(|K|+1)-1} \\ (j,k)\longrightarrow (|J|-1)+(k-1)|J|+j+1, \quad 1\leq j\leq|J|-1, 0\leq k\leq|K| \end{cases}.$$

Now to calculate the fixation probability $\pi^J_{(1,0)}$, taking the advantage of using the vector representation of matrix B (ignoring the entry $(0,0)$), one can define the vector $X=[\pi^J_s]_{s\in\mathcal{S}}$ as a list of all fixation probabilities associated to various states s, then the Kolmogorov equation (2.3) can be viewed as a system: $A\cdot X^T=b^T$ for the vector $b\in\mathbb{R}^{|J|(|K|+1)-1}$ is the matrix of values on the right hand side of the system (2.3) constructed by taking the constants of each equation to the right hand side. The notation ^{T} accounts for the transpose operator of matrices. We solve this equation for X by assuming the initial conditions are $\pi^J_{(0,0)}=0$ and $\pi^J_{(|J|,k)}=1$ for any $0\leq k\leq|K|$. Then, the fixation probability $\pi^J_{(1,0)}$ is the first entry of the matrix X. Similarly, the $|J|^{\text{th}}$ entry of X is in fact the $\pi^J_{(0,1)}$, the fixation probability in the J compartment starting from one initial mutant in K. similarly, one can

find the other two fixation probabilities $\pi^K_{(1,0)}$ and $\pi^K_{(0,1)}$. We will consider this calculation in the subsequent sections.

Example. To describe the method we introduced in this section, for instance one can suppose that $|S_c| = 2$ and $|S_b| = 2$. Considering the transition probabilities in this case, we derive the following system of equations for $\pi_{(j,k)} = \pi^J_{(j,k)}$ where $0 \leq j < 2$ and $0 \leq k \leq 2$:

$$\begin{aligned}
&P^{+,0}_{(1,0)} + P^{0,+}_{(1,0)}\,\pi_{(1,1)} + P^{-,+}_{(1,0)}\,\pi_{(0,1)} + P^{-,+2}_{(1,0)}\,\pi_{(0,2)} \\
&\qquad\qquad - \left(P^{+,0}_{(1,0)} + P^{-,0}_{(1,0)} + P^{0,+}_{(1,0)} + P^{-,+}_{(1,0)} + P^{-,+2}_{(1,0)}\right)\,\pi_{(1,0)} = 0, \\
&P^{+,0}_{(1,0)}\,\pi_{(1,1)} + P^{0,+}_{(1,0)}\,\pi_{(0,2)} + P^{+,-}_{(1,0)}\,\pi_{(1,0)} - \left(P^{+,0}_{(1,0)} + P^{0,+}_{(1,0)} + P^{0,-}_{(1,0)} + P^{+,-}\right)\,\pi_{(0,1)} = 0, \\
&P^{+,0}_{(1,1)} + P^{+,-}_{(1,1)} + P^{-,0}_{(1,1)}\,\pi_{(0,1)} + P^{0,+}_{(1,1)}\,\pi_{(1,2)} + P^{0,-}_{(1,1)}\,\pi_{(1,0)} + P^{-,+}_{(1,1)}\,\pi_{(0,2)} \\
&\qquad\qquad - \left(P^{+,0}_{(1,1)} + P^{-,0}_{(1,1)} + P^{0,+}_{(1,1)} + P^{0,-}_{(1,1)} + P^{+,-}_{(1,1)} + P^{-,+}_{(1,1)}\right)\,\pi_{(1,1)} = 0, \qquad (2.4) \\
&P^{+,0}_{(0,2)}\,\pi_{(1,2)} + P^{0,+}_{(0,2)}\,\pi_{(0,3)} + P^{0,-}_{(0,2)}\,\pi_{(0,1)} + P^{+,-}_{(0,2)}\,\pi_{(1,1)} + P^{+,-2}_{(0,2)}\,\pi_{(1,0)} \\
&\qquad\qquad - \left(P^{+,0}_{(0,2)} + P^{0,-}_{(0,2)} + P^{+,-}_{(0,2)} + P^{+,-2}_{(0,2)}\right)\,\pi_{(0,2)} = 0, \\
&P^{+,0}_{(1,2)} + P^{+,-}_{(1,2)} + P^{+,-2}_{(1,2)} + P^{-,0}_{(1,2)}\,\pi_{(0,2)} + P^{0,-}_{(1,2)}\,\pi_{(1,1)} \\
&\qquad\qquad - \left(P^{+,0}_{(1,2)} + P^{-,0}_{(1,2)} + P^{0,-}_{(1,2)} + P^{+,-}_{(1,2)} + P^{+,-2}_{(1,2)}\right)\,\pi_{(1,2)} = 0.
\end{aligned}$$

Based on the range of change for variables j, k, the matrix B is

$$B = \begin{bmatrix} (0,0) & (0,1) & (0,2) \\ (1,0) & (1,1) & (1,2) \end{bmatrix} \qquad (2.5)$$

which defines the all possible states which should be taken into account in the calculation ignoring $(0,0)$ state. Using the isomorphism defined in above B then is assumed to be be equivalent to the vector $[(|J|-1)+(k-1)|J|+j+1]_{1\times 5} = (1,2,3,4,5)$ in $\mathbb{R}^5$. The latter system defines the 5×5 matrix A as follows

$$\begin{bmatrix}
-\mathcal{A}_1 & P^{-,+}_{(1,0)} & P^{0,+}_{(1,0)} & P^{-,+2}_{(1,0)} & 0 \\
P^{+,-}_{(0,1)} & -\mathcal{A}_2 & P^{+,0}_{(0,1)} & P^{0,2}_{(0,1)} & 0 \\
P^{0,-}_{(1,1)} & P^{-,0}_{(1,1)} & -\mathcal{A}_3 & P^{-,+}_{(1,1)} & P^{0,+}_{(1,1)} \\
P^{+,-2}_{(0,2)} & P^{0,-}_{(0,2)} & P^{+,-}_{(0,2)} & -\mathcal{A}_4 & P^{+,0}_{(0,2)} \\
0 & 0 & P^{0,-}_{(1,2)} & P^{-,0}_{(1,2)} & -\mathcal{A}_5
\end{bmatrix} \qquad (2.6)$$

where

$$\begin{aligned}
\mathcal{A}_1 &= P^{+,0}_{(1,0)} + P^{-,0}_{(1,0)} + P^{0,+}_{(1,0)} + P^{-,+}_{(1,0)} + P^{-,+2}_{(1,0)}, \\
\mathcal{A}_2 &= P^{+,0}_{(0,1)} + P^{0,+}_{(0,1)} + P^{0,-}_{(0,1)} + P^{+,-}_{(0,1)}, \\
\mathcal{A}_3 &= P^{+,0}_{(0,2)} + P^{-,0}_{(1,1)} + P^{0,+}_{(1,1)} + P^{0,-}_{(1,1)} + P^{+,-}_{(1,1)} + P^{-,+}_{(1,1)}, \\
\mathcal{A}_4 &= P^{+,0}_{(0,2)} + P^{0,-}_{(0,2)} + P^{+,-}_{(0,2)} + P^{+,-2}_{(0,2)}, \\
\mathcal{A}_5 &= P^{+,0}_{(1,2)} + P^{-,0}_{(1,2)} + P^{0,-}_{(1,2)} + P^{+,-}_{(1,2)} + P^{+,-2}_{(1,2)}.
\end{aligned} \tag{2.7}$$

Moreover, assuming the matrix X as

$$X = \left[\begin{array}{ccccc} \pi_{1,0)} & \pi_{0,1)} & \pi_{1,1)} & \pi_{0,2)} & \pi_{1,2)} \end{array} \right] \tag{2.8}$$

and the matrix b in this case as following

$$b = \left[\begin{array}{ccccc} -P^{+,0}_{(1,0)} & 0 & -P^{+,0}_{(1,1)} - P^{+,-}_{(1,1)} & 0 & -P^{+,0}_{(1,2)} - P^{+,-}_{(1,2)} - P^{+,-2}_{(1,2)} \end{array} \right] \tag{2.9}$$

then we can have the linear non-homogenous system $A \cdot X^T = b^T$.

2.1.3 Fixation Time

Defining $\bar{t}_{ij} = \sum_{n=0}^{\infty} P^{(n)}_{i,j}$ as the mean time between states i and j which is the total number of times which X has been at state j before absorption into 0 or N (i.e. the population size) starting from the state i, and $\bar{t}_i = \sum_{j=1}^{N-1} \bar{t}_{ij}$ as the mean time at state i until absorption into 0 or N. This is the fixation time as the average time of passing through different states until fixation which satisfies in the below system of equations

$$\begin{cases} \bar{t}_i = \sum_{j=0}^{N} P_{i\to j}\, \bar{t}_j + 1, \\ \bar{t}_0 = \bar{t}_N = 0, \end{cases} \tag{2.10}$$

where the term 1 in the above equations of $\bar{t}_i$ is due to the last system and the fact that $\sum_{j=1}^{N-1} \delta_{ij} = 1$.

The mean time to fixation in the unconditional Moran process at state i is

$$\bar{t}^B_i = -\bar{t}_1 \sum_{k=i}^{N-1} \prod_{m=1}^{k} \gamma_m + \sum_{k=i}^{N-1} \sum_{l=1}^{k} \frac{1}{P_{l\to l+1}} \prod_{m=l+1}^{k} \gamma_m, \tag{2.11}$$

where

$$\bar{t}_1 = -\bar{t}_1 \sum_{k=1}^{N-1} \prod_{m=1}^{k} \gamma_m + \sum_{k=1}^{N-1} \sum_{l=1}^{k} \frac{1}{P_{l\to l+1}} \prod_{m=l+1}^{k} \gamma_m, \tag{2.12}$$

Moreover, in conditional case we get

$$\bar{t}_1^B = \sum_{k=1}^{N-1} \sum_{l=1}^{k} \frac{\pi_l}{P_l^+} \prod_{m=l+1}^{k} \gamma_m, \tag{2.13}$$

$$\bar{t}_i^B = -\frac{\pi_1}{\pi_i} \left(\bar{t}_1^B \sum_{k=i}^{N-1} \prod_{m=1}^{k} \gamma_m - \sum_{k=i}^{N-1} \sum_{l=1}^{k} \frac{1}{P_{l\to l+1}} \prod_{m=l+1}^{k} \gamma_m \right). \tag{2.14}$$

Generally speaking, a very useful approach to derive the time to fixation is to use Kramers–Moyal expansion equation (see e.g [179, 180] for more details) when an approximation of this equation for a given model lead to a system of ODEs, such as replicator dynamics. In this case, solving such a system subject to initial conditions will provide an estimation of fixation time of the procedure. This might be very useful to obtain the time of tumor growth. Moreover, it looks crucial to have an estimation for the time of metastasis in an invasive mutant population when the epithelial markers switch to mesenchymal markers in a somatic cancer.

2.2 Moran Model

The Moran process is the simplest possible stochastic model to study evolutionary processes in a finite population when generations are allowed to overlap each other. Let us assume that we have a population of two type A and B individuals. During a birth–death (BD) Moran model, a random individual is chosen for proliferation while one of its daughter cells replaces another random individual at the same time step. Death–birth (DB) Moran model in turn, at first has a death event for a random individual and then a birth event simultaneously occurs for a random individual which send its offspring for substitution. Individuals in a Moran model either are in a mixed population in which every individual could move freely (often referred to as *conventional*, *well–mixed*, or *unstructured* model) or on a digraph (directed graph) which every individual has a specific position on a particular vertex (also called as *structured* or *spatial* Moran model). At each time step, there would be three different mechanisms happening for individuals through a Moran process where

$P_i^+(P_i^-)$ is the probability of increasing (decreasing) i by one while $P_i^\circ = 1 - P_i^- - P_i^+$ is the probability of that i remain unchanged through the finite Markov chain.

Furthermore, deriving the master equation from the Kolmogorov equation in terms of transition probabilities, the probability generating function (PGF) can be used to find survival/extinction probability. Let the finite summation $\phi(z,t) = \sum_{n=0}^{\infty} z^n p(n,t)$ be the corresponding PGF for continuous variable z. Since the Markov process has two absorbing states at $n = 0$ and $n = N$, this polynomial would be of degree N.

2.2.1 Well-Mixed Moran Model

Consider a well-mixed BD model of a finite population size N of two types of individuals: A and B. There are two different models.

(i) Neutral drift model. Let reproduction rates (fitness) be the same ($r_A = r_B$). When $\gamma = 1$, $\pi_i = \frac{i}{N}$ for $1 \leq i \leq N-1$ and the fixation probability is equal to $\frac{1}{N}$, that is, the neutral variant with respect to selection. In this case, the fixation time for large N will be

$$\begin{aligned} \bar{t}_i &= -N^2 \left(\pi_i \ln(\pi_i) + (1-\pi_i)\ln(1-\pi_i)\right) \\ &\approx -\frac{N(1-\pi_i)}{\pi_i} \ln(1-\pi_i). \end{aligned} \tag{2.15}$$

(ii) Random drift (constant selection) model. When $r_A = r > 1 = r_B$ or $\gamma = 1/r$ then

$$P_i^- = \frac{N-i}{r\,i+N-i}\frac{i}{N}, \quad P_i^+ = \frac{r\,i}{r\,i+N-i}\frac{N-i}{N}. \tag{2.16}$$

So $\pi_i = \frac{1-\left(\frac{1}{r}\right)^i}{1-\left(\frac{1}{r}\right)^N}$ and then $\rho_A = \frac{1-\frac{1}{r}}{1-\left(\frac{1}{r}\right)^N}$ and $\rho_B = \frac{1-r}{1-r^N}$. The approximate fixation time in this case is

$$\bar{t} = \frac{r+1}{r-1} \ln N. \tag{2.17}$$

The ratio of the fixation probabilities is given by $\frac{\rho_A}{\rho_B} = r^{1-N}$ which for $N \gg 1$ leads to the outcompeteness of A with $\rho_A = 1 - 1/r$. Now considering a DB Moran model, repeating the similar mechanism as we did for BD model, the same result will be determined, if we consider $r = 1$ and then exchange d with $1/r$ then we can come up with the same expression.

Let u be the probability of mutation from A to B in a BD model, the absorbing state in this system would be $i = N$. Let N be the population size. Thus probability of having a mutant from all A state is equal to Nu and $\mathcal{R}_{A\to B} = Nu\rho$ is the rate of fixation from A to B, where the time–scale for this event is N times as much as that of a proliferation. Then the probability for a new mutant to be absorbed at time t is approximately

$$1 - \mathrm{e}^{-N\,u\,\rho\,t} \approx N\,u\,\rho\,t \tag{2.18}$$

where time is scaled with N. This is, in fact, the solution of the following two-state master equation:

$$\begin{cases} \dot{x}_0 = -\mathcal{R}_{A\to B}\, x_0, \\ \dot{x}_1 = \mathcal{R}_{A\to B}\, x_0, \\ x_0(0) = 1, x_1(0) = 0, \end{cases} \tag{2.19}$$

where x_0 and x_1 denote the number of individuals at state $i = 0$ and $i = N$ respectively.

2.2.2 Spatial Structures and Moran Model

An important question in population genetics and evolutionary dynamics relates to the effect of spatial structures on the behavior of a given heterogeneous system [95, 97, 54, 55, 74]. It has been shown that by applying branching process and finite Markov chain, the compartmental structure of a well-mixed model has no effect on the fixation probability compared with that of a similar well-mixed system [74]. Moreover, Houchmandzade et al. [54, 55] have studied a general framework of spatial Moran model. They argue that under small selection pressure in a DB model (but not in the neutral drift), the fixation probability declines for an advantageous mutant. However, in BD model, the chance for a beneficial mutant to take over the whole system can exceed the fixation probability of a non-structured population of the same size with neutral drift.

Furthermore, there have been lots of efforts in the literature to understand the general formalism regarding the fixation probability of a certain individuals in a BD versus DB spatial structures (see e.g. [75, 66, 121, 122, 12]). Among diverse structures, tree-shaped structures are very conductive (suppressive) for advantageous mutants where the underlying mechanism is BD (DB). In a BD Moran model, the star or star–like trees are very beneficial for mutants; however, the line and circle graphs are not among the best conductive structures. Among various DB structures, the line and circle models are very conductive and the star (star–like) model decreases the survival property of mutants.

Another interesting feature of the spatial structure relates to the microenvironmental effect on tumor development. To capture such effects one can consider a heterogeneous population located on a one or two dimensional structure under the influence of epigenetic/environmental alteration. The 1D or 2D spatial structures can be seen as lattices (i.e. graphs in which cells are positioned on its nodes). One may also assume that within a 1D structure, there exist two neighbors for each individual and for 2D structure one may consider the Van Neuman's model with four adjacent cells for each internal node.

Most of the cancers are happening in a tissue when carcinoma is generated on a lattice of epithelial cells when this lattice usually is 1D or 2D graph. Moreover for other types of cancer, such as blood cancers, when a well–mixed population exist, a complete graph can be considered as the structure of the epithelium hosting wild–type or cancerous cells. While a population consists of two types of cells, A and B, located on a weighted digraph. The weight matrix $W = (w_{i,j})$ also referred to as edge–transition probability matrix defines the probability of choosing edges between two nodes of the graph as the weight of the connecting edge. Moreover, one can look at the set of vertices (located cells) as a binary vector $\mathbf{v} = (v_1, v_2, \cdots, v_N)$ in which for individuals of type A the corresponding vertex index v_i is equal to zero and is 1 for type B individuals. Let us assume that the relative fitness of B individuals is equal to r compared to the fitness of A which is supposed to be normalized to 1.

The major difference between spatial and well–mixed models relates to the fact that, when the birth event (death event in DB case) randomly occurs within the population, the death event (birth event in DB model) is a local mechanism associated to the neighbor cells surrounding the first individual apparently chosen for birth (death in DB regime). In the BD case we have

$$\begin{aligned} P^{+}_{BD}(m,i) &= \frac{r\, v_i}{rm + (N-m)} \; \frac{\sum_j \omega_{ij}\,(1-v_j)}{d\,\sum_j \omega_{ij}\, v_j + \sum_j \omega_{ij}\,(1-v_j)}, \\ P^{+}_{BD}(m,i) &= \frac{(1-v_i)}{rm + (N-m)} \; \frac{d\,\sum_j \omega_{ij}\, v_j}{d\,\sum_j \omega_{ij}\, v_j + \sum_j \omega_{ji}\,(1-v_j)}, \end{aligned} \tag{2.20}$$

where index i in $P^{\pm}_{BD}(m,i)$ is indicating the position of the event regarding the configuration index v_i and m is the number of mutant cells. Also $\sum_j \omega_{ij}\, v_j$ and $\sum_j \omega_{ij}\,(1-v_j)$ terms are counting respectively the total number of type B and A neighbor cells. Moreover, having the term v_i guarantees that we are dealing with a B cell whilst $1-v_j$ is related to a A cell.

Similarly in DB case we obtain

$$
\begin{aligned}
P^{+}_{DB}(m,i) &= \frac{(1-v_i)}{dm+(N-m)}\ \frac{r\ \sum_j \omega_{ji}\, v_j}{r\ \sum_j \omega_{ji}\, v_j + \sum_j \omega_{ji}\,(1-v_j)}, \\
P^{+}_{DB}(m,i) &= \frac{d\, v_i}{dm+(N-m)}\ \frac{\sum_j \omega_{ji}\,(1-v_j)}{r\ \sum_j \omega_{ji}\, v_j + \sum_j \omega_{ji}\,(1-v_j)},
\end{aligned}
\tag{2.21}
$$

The following important theorem investigates those conditions under which the spatial system is equivalent to a well–mixed Moran model, meaning that the fixation probability of mutant cells is free of position. The prerequsite for such an equivalence to occur is that the given structure be equipped with a periodic boundary. In the following theorem, a special case of the above construction is assumed when ω_{ij} is not only the weight of edges but also comprises the death probability (or birth probability in DB model) itself. To have a $\rho-$equivalence graph to the neutral drift model, the fixation probability of every node (which in fact defines a balance between selection and drift (see [121])) must be equal to that of any other point independent of the configuration of the system of A and B s. Thus a graph is isothermal if we have the equality $\sum_{j=1}^{N} \omega_{ij} = \sum_{i=1}^{N} \omega_{ij}$.

Isothermal Theorem [121]. *A graph is $\rho-$equivalent (has equivalent fixation probability at each vertex) to the Moran process iff it is isothermal, i.e., $\sum_{j=1}^{N} \omega_{ji} = \sum_{j=1}^{N} \omega_{ij}$.*

Isothermal theorem is not hold for some structured Moran models such as the cycle model (see e.g. [121]). In addition, we show how it differs for BD and DB Moran models, revealing the fact that BD and DB Moran models are not generally the same. For instance, see [67] for the the difference and the symmetry of these two types of preccesses in the circle model.

BD Case. Considering N individuals of types A and B in a finite population as vertices of a directed and weighted graph of weights ω_{ij} between vertices v_i and v_j.

If we define the state m to be as the total number of B individuals existing in the population, hence $m = \sum_i v_i$. The only possible case to have an increase in the number of B individuals is due to the case which in BD process one Bat site i is proliferating and an A at site j is selected to be killed (sites i and j are in correspondence with vertices v_i and v_j respectively). We know that in a BD model, the main process is the proliferation which forces the equality $\sum_j \omega_{ij} = 1$. Hence

Thus the probabilities of increasing and decreasing of population B are respectively

$$P_m^+ = \sum_{i,j} \left(\frac{r v_i}{rm + N - m} \right) \left(\omega_{ij} \left(1 - v_j \right) \right), \tag{2.22}$$

$$P_m^- = \sum_{i,j} \left(\frac{1 - v_i}{rm + N - m} \right) \left(\omega_{ij} v_j \right). \tag{2.23}$$

Thus

$$\frac{P_m^-}{P_m^+} = \frac{\sum_{i,j} \omega_{ji} v_i (1 - v_j)}{r \sum_{i,j} \omega_{ij} v_i (1 - v_j)}. \tag{2.24}$$

Thus the isothermal theorem holds in this case when this ratio is equal to $\frac{1}{r}$ which leads to the general form of the fixation probability in a perfect mixed Moran model: $\rho = \frac{1-1/r}{1-(1/r)^N}$. If the graph be isothermal then in BD model $\sum_{j=1}^{N} \omega_{ij} = \text{const.}$ taking summation of both sides of this equality leads to the result. For the reverse side, having (2.24) equal to $\frac{1}{r}$ for any configuration vector v which leads to $\rho-$equivalency of the system.

DB Case. In this case, the probability of increasing and decreasing the population by one for B individuals could be respectively

$$\begin{aligned} P_m^+ &= \sum_{i,j} \frac{r(1 - v_i)\, v_j}{N} \frac{\omega_{ji}}{1 + \sum_k \omega_{ki}\, v_k (r - 1)}, \\ P_m^- &= \sum_{i,j} \frac{v_i\, (1 - v_j)}{N} \frac{\omega_{ji}}{1 + \sum_k \omega_{ki}\, v_k (r - 1)}. \end{aligned} \tag{2.25}$$

and thus

$$\frac{P_m^-}{P_m^+} = \frac{\sum_{i,j} \frac{v_i\, (1 - v_j)\, \omega_{ij}}{1 + \sum_k \omega_{ik}\, v_k (r-1)}}{\sum_{i,j} \frac{(1 - v_i)\, v_j\, \omega_{ji}}{1 + \sum_k \omega_{ki}\, v_k (r-1)}}, \tag{2.26}$$

Isothermal theorem doesn't generally hold for DB Moran procedures. A generalization of this modeling for more than two types of individual has been considered in Chapter 5 as has been well–defined theoretically.

2.2.3 Spatial Moran Model for Cycle

As a very simple and regular structure, a cycle (circular digraph) could be thought of as an appropriate structure to study the evolutionary mechanism of a cell when a new mutant is located on such a framework. For instance, when a new mutant appears within the host cells in the niche of a crypt which usually assumed to be a circle of 4-8 wild–type stem–cells [97].

Consider a cycle of N nodes with weight (transition probability) matrix W with $w_{i,j} = 1/2$ for neighbor points and 0 for distant ones. Suppose that the vertices of the graph is full of individuals of types A and B with relative fitness 1 and r respectively

BD case. Considering a circle as the model graph with N individuals of types A and B the weight matrix is the same as we got for BD model

$$\begin{cases} P_1^+ = \frac{r}{r+N-1}, & P_1^- = \frac{2\,d}{(d+1)(r+N-1}, \\ P_m^+ = \frac{2\,r}{(d+1)(r(m-1)+N}, & P_m^- = \frac{2\,d}{(d+1)(r(m-1)+N}, \quad \text{for } 1 < m < N-1 \\ P_{N-1}^+ = \frac{2\,r}{(d+1)(r(N-1)+1)}, & P_{N-1}^- = \frac{1}{r(N-1)+1}. \end{cases} \tag{2.27}$$

Kolmogorov equation leads to the following form of fixation probability

$$\pi_1 = \frac{d(d+1)(r-d)}{d\Big(d(r-1)+(r+d)\Big) + \Big((r-d)-d(r+1)\Big)\Big(\frac{d}{r}\Big)^{N-1}}, \tag{2.28}$$

where for $d = 1$ leads to $\pi_1 = \frac{1-1/r}{1-(1/r)^N}$ which is exactly equal to that of for well–mixed Moran model with the same population. Converse to the BD model, the DB Moran model of the cycle provides the following transition probabilities

$$\begin{cases} P_1^+ = \frac{2\,r}{(r+1)(d+N-1)}, & P_1^- = \frac{d}{d+N-1}, \\ P_m^+ = \frac{2\,r}{(r+1)(d(m-1)+N)}, & P_m^- = \frac{2d}{(r+1)(d(m-1)+N)}, \quad \text{for } 1 < m < N-1 \\ P_{N-1}^+ = \frac{1}{d(N-1)+1}, & P_{N-1}^- = \frac{2d}{(r+1)(d(N-1)+1)}. \end{cases} \tag{2.29}$$

In this case we have

$$\pi_1 = \frac{2(r-d)}{2r + d(r-1) + d(r-1-2d)\Big(\frac{d}{r}\Big)^{N-2}}. \tag{2.30}$$

When $d = 1$ we obtain

$$\pi_1 = \frac{2(r-1)}{3r - 1 + (r-3)\Big(\frac{1}{r}\Big)^{N-2}}. \tag{2.31}$$

The same result as what is reported in [75] would be obtained when $\pi_1 \longrightarrow \frac{2(r-1)}{3r-1}$ as $N \longrightarrow \infty$. Comparing to the BD and DB Moran models on a cycle for $N \gg 1$ and assuming that $r = 1 + s$ for $s \ll 1$)

$$\begin{aligned} \pi_1^{\text{Moran}} &= 1 - \frac{1}{r} = \frac{s}{s+1} \approx s - s^2, \\ \pi_1^{DB} &= \frac{2(r-1)}{3r-1} = \frac{s}{1+3/2s} \approx s - \frac{3}{2}s^2. \end{aligned} \tag{2.32}$$

Chapter 3

Evolutionary dynamics of stem cell hierarchy with phenotypic plasticity

In this chapter, phenotypic hierarchy of a compartment of individuals and effect of phenotypic plasticity on differentiated individuals on the fate of mutants is exhaustively investigated. The majority of results presented in the current chapter are reported in [95]. This study reveals how unwelcome evolution of malignancy during cancer progression emerges through a selection process in a complex heterogeneous population structure. More precisely, we investigate evolutionary dynamics in a phenotypically heterogeneous population of stem cells (SCs) and their associated progenitors.

The fate of a malignant mutation is determined not only by overall stem cell and differentiated cell growth rates but also differentiation and dedifferentiation rates. We investigate the effect of such a complex population structure on the evolution of malignant mutations. We derive exact analytic results for the fixation probability of a mutant arising in each of the subpopulations. The analytic results are in almost perfect agreement with the numerical simulations. Moreover, a condition for evolutionary advantage of a mutant cell versus the wild type population is given in the present study. We also show that microenvironment-induced plasticity in invading mutants leads to more aggressive mutants with higher fixation probability. Our model predicts that decreasing polarity between stem and differentiated cells' turnover would raise the survivability of non-plastic mutants; while it would suppress the development of malignancy for plastic mutants. We discuss our model in the context of colorectal/intestinal cancer (at the epithelium). This novel mathematical framework can be applied more generally to a variety of problems concerning selection in heterogeneous populations, in other contexts such as population genetics, and ecology.

3.1 Introduction

Cancer can be thought of as a complex ecosystem in which not only tumor cells but also other cell types (phenotypes) may influence the overall health of an organism. Experimental results have recently shown that cancer cells may mimic the functional features of normal cells [81]. The most important features are associated with a small subpopulation of cells, namely the stem cells. Stem cells (SCs) are defined to be cells with self-renewal capacity and pluripotency. For instance, they can replenish and regenerate the whole epithelial cell population in normal tissues. It has been proposed that cancer stem cells (CSCs) maintain invasive characteristics, such as (undesirable) multipotency and uncontrolled growth and tumor initiating capacity [9, 127, 141, 165, 101]. The differentiated progenies of SCs are the cells with specialized distinct functions, within the organism. They are produced via a hierarchical division scheme. As the differentiated cells (DCs) become more mature along the hierarchy, their replication potential decreases [16, 88, 94, 144].

CSCs reside in small niches and manifest characteristics similar to somatic SCs [9]. In solid cancers, CSCs are usually imputed as a result of the expression of similar biomarkers as those used to identify SCs [45, 119, 161, 7, 196]. In colon cancer, the over–expression of the polycomb ring finger oncogene BMI1 leads to the down–regulation of proteins p16INK4a and p14ARF. These proteins override cellular proliferation restriction and generate cancer SLCs [28, 88, 101]. For mammary stem cells, $CD44^+$ and $CD24^-$ are reported as markers for stemness. In acute–myeloid leukemia (AML) $CD34^+$ $CD38^-$ cells are a leukemia–initiating subpopulation[101, 147].

Despite the new established dogma that cancer cells originate from a small niche of cells [101, 144, 181], a range of experiments have now investigated and reported on the cancer initiating capacity of committed progenitor cells [127, 91, 101]. In other words, non-SCs can undergo a dedifferentiation process and regain stemness (these cells are also called stem like cells). In breast cancer, epithelial–mesenchymal transition (EMT) factors have been implicated in the production of stem like cells from non–stem cells [100, 16, 17]. Gupta *et al* [48] have also observed that the epithelial differentiated cells with basal markers can convert to cells with stem cell markers (see also [16]). There are several other experimental observations supporting dedifferentiation of committed progenitor cells [14, 57, 91, 101, 137]. In addition, this dedifferentiation has been observed, under certain microenvironmental conditions, in normal SCs [81, 172, 30]. In fact, it is becoming apparent that cellular trans–differentiation is activated in a number of organs to produce stem like cells in support of SCs in tissue regeneration [81, 91, 101, 181, 16, 35, 37, 88].

In *de novo* AML, expression of NUP98-HOXA9 and also loss of TSG PML (promyelocytic leukemia protein) has been observed to enhance symmetric divisions in WT hematopoi-

etic SCs [5, 60] and shift the balance between symmetric and asymmetric divisions. Our model reveals that in such a circumstance and where there is no DC plasticity in the system, decreasing the asymmetric division rate tends to an increase in the survivability of mutants. On the other hand, there exist some evidence relying on the dedifferentiation potential of non-leukemic cells [46, 205]. As a result of our investigation, the specific type of division in leukemic initiating cells may determine the fate of such an invasive malignancy. More precisely, when the tissue is endowed with a potential for DCs to transform back to stemness state, decrease in the rate of differentiation may decline the fixation probability of an originated cell in SC compartment. Our model may help to obtain a better understanding of the evolutionary dynamics of resistant leukemic cells at diagnosis, during treatment, or within relapse. In this viewpoint, it may improve the therapeutic strategies to prevent or delay the development of AML. For instance, controlling the dynein binding protein Lis1 (Pafah1b1) been shown to regulate the asymmetric division in HSCs [206]. Also, focusing on some translocation–associated oncogenes such as MLL-AF9 which relate to dedifferentiation and the HSC self–renewal properties leading to higher differentiation rates in LSCs, has thought of to be as another potential therapeutic target in AML [82]. Moreover, defect in the expression of $\beta-$catenin Ctnnb1 would tend to lower survival chance of mutants [188]. Furthermore, loss of the dynein-binding protein Lis1 increases differentiation of LSCs [206] and can be accounted as another method to pretend the asymmetric differentiation of LSCs and thus control the growth rate in AML [5].

A variety of quantitative approaches have been utilized to investigate the effect of the stem cell hierarchy and phenotypic heterogeneity on tumor growth [22, 26, 178, 202, 42]. In the context of cancer evolution, the deterministic population dynamics of the stem cell hierarchy in the absence of plasticity, is discussed by Werner *et al.* [193, 192]. Plasticity and dedifferentiation is explored under a diffusion approximation [62] and replicator equation [66]. In [155], the authors discuss the rate of evolution in a simple hierarchical stem and non–stem cell population. They argue that stem cell symmetric division is preferred under natural selection for two-hit mutations.

The evolutionary dynamics of malignant and normal genotypes in the presence of phenotypic transformations (differentiation and dedifferentiation) is not well understood. In this study, we consider a general framework to study natural selection in heterogeneous populations. We analyze competition between resident and mutant populations which are genotypically different. Each of these types divide into phenotypically different subpopulations (stem cell and differentiated subtypes). Due to homeostasis, the size of SC and DC subpopulations are assumed to remain constant. Stem cells can self–renew and replenish their own population or contribute to differentiated cell population via differentiation events. Differentiated cells can also divide into differentiated cells or dedifferentiate into

stem like cell states.

We investigate conditions for the successful selection of a malignant mutation in this complex population structure. Due to the plastic nature of the early malignant progenitors, there is a finite chance for an advantageous mutant to exit the differentiated group and become part of the SC niche. We derive analytic results that predict the fixation probability of a mutant (either in the SC or DC subpopulations) to establish a finite colony. We assume arbitrary population sizes and division rates and selection intensities as well as (de)differentiation rates. The analytic results are in excellent agreement with stochastic simulations in finite populations. We apply our findings to colorectal cancer and predict dedifferentiation rates that can confer a selection advantage for p53 mutants.

3.2 Evolutionary dynamics of the model

Consider two populations of resident or wild type (type 1) and mutant or invader (type 2). Mutants are the result of an oncogenic mutation in the resident population. Each genotype is divided into phenotypically different subpopulations of stem cells (SC) and differentiated cells (DC). Stem cells can self-renew symmetrically where the offspring are stem cells. They can differentiate (symmetrically or asymmetrically) to produce differentiated progenies of the same genotype. We denote the probability of asymmetric differentiation (per division) by $\hat{u}_1, \hat{u}_2$ and symmetric differentiation by $\hat{v}_1, \hat{v}_2$. The overall probability of differentiation is $u_{1,2} = \hat{u}_{1,2} + 2\hat{v}_{1,2}$. This is due to the fact that symmetric differentiation produces two differentiated cells. Similarly the self-renewal probability is denoted by $1-u_{1,2}$. The indexes 1 or 2 denote the corresponding probabilities for a wild type or mutant. The division rate of a normal (or mutant) stem cell is denoted by r_1 (r_2) respectively. Similarly, the division rates of progenitors/differentiated cells are denoted by $\hat{r}_1$ and $\hat{r}_2$. For evolutionary dynamics we consider a birth-death (BD) Moran process as follows: At each time step, an individual is chosen to reproduce proportional to its fitness within the SC or DC compartments. If a normal (mutant) cell in the SC compartment is chosen to reproduce, its offspring replaces a randomly chosen cell in the stem cell compartment with probability $1 - u_1$ ($1 - u_2$). Otherwise, with probability u_1 (u_2), the (differentiated) offspring replaces a randomly chosen cell in the DC compartment. Similarly, if a differentiated cell is chosen to reproduce, its offspring replaces another cell in the differentiated cell compartment with the probability $1 - \eta$. Alternatively, the offspring can dedifferentiate into a stem-like cell and replace a randomly chosen individual in the stem cell compartment with a rate η, where $\eta = \eta_1(\eta = \eta_2)$ denote the dedifferentiation probability for normal (mutant) DCs. For simplicity we assumed death rates of all types to be equal and set this to unity.

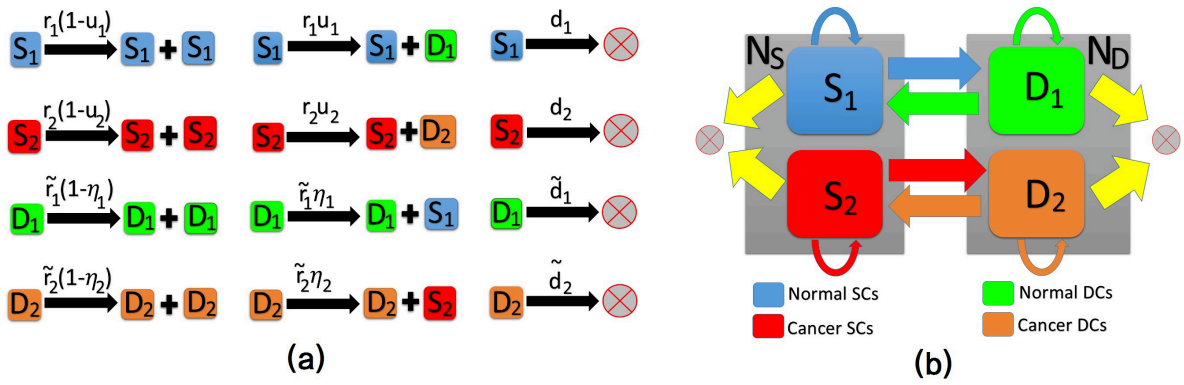


Figure 3.1: **Phenotypic–genotypic changes in individuals within a four– compartmental structure.** We consider constant population sizes N_{S} and N_{D} for SCs and DCs respectively. With respect to the finite Markov chain, we consider a generalized model to take into account the competition between normal and malignant individuals in each of the SC and DC subpopulations. Differentiation and dedifferentiation events connect the selection dynamics between the two niches. In (a), all possible differentiation, dedifferentiation, and death events with their corresponding rates are represented. The SC-DC compartmental structure is depicted in (b) with the associated self–renewal and differentiation/plasticity possibilities.

Table 3.1: Model parameters

Notation	Description
$N_{\mathrm{S}}, N_{\mathrm{D}}$	Total number of stem and differentiated cells
r_1, r_2	Net reproduction rate of wild type and mutant stem cells
$\tilde{r}_1, \tilde{r}_2$	Net reproduction rate of wild type and mutant differentiated cells
u_1, u_2	Asymmetric differentiation rate of normal and mutant stem cells
η_1, η_2	Dedifferentiation rate of normal and mutant differentiated cells

The above dynamics models the differentiation mechanism with an *effective* asymmetric division with the probabilities u_1, u_2. Thus in the following we use the terms differentiation (of stem cells) and asymmetric division interchangeably.

The above Moran process can be written as a continuous time process ($1/N$ is the

duration of each time step for $N = N_{\rm S} + N_{\rm D}$)

$$\begin{aligned}\frac{1}{N}\frac{\partial p(n_{\rm S},n_{\rm D};t)}{\partial t} &= W_{\rm S}^{+}(n_{\rm S}-1,n_{\rm D})\,p(n_{\rm S}-1,n_{\rm D};t) + W_{\rm S}^{-}(n_{\rm S}+1,n_{\rm D})\,p(n_{\rm S}+1,n_{\rm D};t)\\ &+ W_{\rm D}^{+}(n_{\rm S},n_{\rm D}-1)\,p(n_{\rm S},n_{\rm D}-1;t) + W_{\rm D}^{-}(n_{\rm S},n_{\rm D}+1)\,p(n_{\rm S},n_{\rm D}+1;t)\\ &- \big(W_{\rm S}^{+}(n_{\rm S},n_{\rm D}) + W_{\rm D}^{+}(n_{\rm S},n_{\rm D}) + W_{\rm S}^{-}(n_{\rm S},n_{\rm D}) + W_{\rm D}^{-}(n_{\rm S},n_{\rm D})\big)\,p(n_{\rm S},n_{\rm D};t). \quad (3.1)\end{aligned}$$

where $p(n_{\rm S},n_{\rm D};t)$ denotes the probability of having $n_{\rm S}$ mutant stem cells and $n_{\rm D}$ mutant differentiated cells, at time t (given $n_{\rm S}^0$ and $n_{\rm D}^0$ at $t=0$). The population of normal cells are given by $N_{\rm S}-n_{\rm S}$ and $N_{\rm D}-n_{\rm D}$ correspondingly. The probabilities $W_{\rm S}^{\pm}$ and $W_{\rm D}^{\pm}$ are the transition probabilities corresponding to an increase or decrease by one in the number of mutant SCs and DCs resp. They are given by

$$\begin{aligned}W_{\rm S}^{+}(n_{\rm S},n_{\rm D}) &= {\rm Prob}(n_{\rm S},n_{\rm D}\to n_{\rm S}+1,n_{\rm D})\\ &= \left(\frac{r_2\,(1-u_2)\,n_{\rm S}+\tilde{r}_2\,\eta_2\,n_{\rm D}}{{\rm N_r}}\right)\frac{N_{\rm S}-n_{\rm S}}{N_{\rm S}},\\ W_{\rm S}^{-}(n_{\rm S},n_{\rm D}) &= {\rm Prob}(n_{\rm S},n_{\rm D}\to n_{\rm S}-1,n_{\rm D})\\ &= \left(\frac{r_1\,(1-u_1)\,(N_{\rm S}-n_{\rm S})+\tilde{r}_1\,\eta_1\,(N_{\rm D}-n_{\rm D})}{{\rm N_r}}\right)\frac{n_{\rm S}}{N_{\rm S}},\\ W_{\rm D}^{+}(n_{\rm S},n_{\rm D}) &= {\rm Prob}(n_{\rm S},n_{\rm D}\to n_{\rm S},n_{\rm D}+1)\\ &= \left(\frac{\tilde{r}_2\,(1-\eta_2)\,n_{\rm D}+r_2\,u_2\,n_{\rm S}}{{\rm N_r}}\right)\frac{N_{\rm D}-n_{\rm D}}{N_{\rm D}},\\ W_{\rm D}^{-}(n_{\rm S},n_{\rm D}) &= {\rm Prob}(n_{\rm S},n_{\rm D}\to n_{\rm S},n_{\rm D}-1)\\ &= \left(\frac{\tilde{r}_1\,(1-\eta_1)\,(N_{\rm D}-n_{\rm D})+r_1\,u_1\,(N_{\rm S}-n_{\rm S})}{{\rm N_r}}\right)\frac{n_{\rm D}}{N_{\rm D}}.\end{aligned} \quad (3.2)$$

The denominator ${\rm N_r}$ denotes the total fitness of SC and DC individuals:

$${\rm N_r} = r_1\,(N_{\rm S}-n_{\rm S}) + r_2\,n_{\rm S} + \tilde{r}_1(N_{\rm D}-n_{\rm D}) + \tilde{r}_2\,n_{\rm D}. \quad (3.3)$$

The above Markov process has two absorbing states corresponding to fixation or extinction of the mutant or WT. The competition between the two genotypes in the stem cell compartment is tied to the competition inside the differentiated compartment via differentiation and dedifferentiation mechanisms. In the absence of plasticity we have a hierarchical population structure where only mutations in the stem cell compartments can give rise to fixation in the whole population.

3.2.1 Fixation probability in a heterogeneous Moran process

One of the most important questions to address within a heterogeneous population is the chance of success for a mutation in different subtypes.

The fixation probability of a mutant originating in the stem cell compartment, ρ_{S} or the differentiated cell compartment, ρ_{D} is a measure of the tumor initiating capacity of each subpopulation. For a completely hierarchical population, only mutants that arise in the stem cell niche have a chance of fixating in the whole population thus $\rho_{\mathrm{S}} = \rho$. If the progenitors can dedifferentiate into stem-like cells, the comparison between the two fixation probabilities, (ρ_{S} and ρ_{D}), is a good measure of how the tumor initiating capacity correlates with the notion of stemness.

The use of the probability generating function (PGF) method to study a constant population Moran process is discussed in [54, 55, 67]. It is used to present an alternative derivation of the (well-mixed) Moran fixation probability, by identifying a martingale for the process. A martingale is a sequence of random variables where the average of a given random variable is equal to the average of its previous random variable in the sequence. In this stochastic model, the fate of future events only depends on its previous event, not on the earlier events. Here we generalize this technique for a heterogeneous population under selective pressure in the presence of phenotypic plasticity. A martingale for the above four population model, Eqs. (3.1), (3.3), can be written as

$$\langle \left(z_{\mathrm{S}}^{\star}\right)^{n_{\mathrm{S}}} \left(z_{\mathrm{D}}^{\star}\right)^{n_{\mathrm{D}}} \rangle, \tag{3.4}$$

where $\langle \cdot \rangle$ denotes the stochastic average.

Taking advantage of the generating function, an exact analytic approach for the fixation probability can be derived even in the presence of plasticity (when mutation and mutation–back do not occur). The results are obtained for a BD Moran process; however, a similar calculation can be performed for a Voter (DB) Model with presumably different fixed points for the same initial conditions. The boundary and initial conditions for the corresponding generating function are resp. as follows

$$F(z_{\mathrm{S}} = 1, z_{\mathrm{D}} = 1, t) = 1, \quad \text{for any } t > 0, \tag{3.5}$$

$$F(z_{\mathrm{S}}, z_{\mathrm{D}}, t = 0) = z_{\mathrm{S}}^{i}\, z_{\mathrm{D}}^{j}, \tag{3.6}$$

where i and j are the initial number of cancer SCs and DCs resp. For the following special cases, the system has two absorbing states at equilibrium which signify fixation and extinction for mutant cells in either compartments. We denote the probability of

reaching extinction and fixation states by B_0 and B_1 resp. Thus we obtain the following result from the PGF at steady state

$$\begin{aligned} F(z_{\rm S}, z_{\rm D}, t \to \infty) &= p(n_{\rm S} = 0, n_{\rm D} = 0, t \to \infty) \\ &\quad + p(n_{\rm S} = N_{\rm S}, n_{\rm d} = N_{\rm D}, t \to \infty)\, z_{\rm S}^{N_{\rm S}}\, z_{\rm D}^{N_{\rm D}} \\ &= B_0 + B_1\, z_{\rm S}^{N_{\rm S}}\, z_{\rm d}^{N_{\rm D}}, \end{aligned} \tag{3.7}$$

based on the boundary condition (3.5), $B_0 + B_1 = 1$. Biologically, since the mutant DCs are produced by cancer SCs, extinction of cancer SCs will result in replenishment of mutant DCs. Moreover, the co-operation between SCs and DCs suggests that it is impossible to have mutant SCs fixate while DCs become completely extinct. We conclude that

$$F(z_{\rm S}, z_{\rm D}, t \to \infty) = 1 - B_1 \left(1 - z_{\rm S}^{N_{\rm S}}\, z_{\rm D}^{N_{\rm D}}\right). \tag{3.8}$$

Finally, applying the initial condition, $F(z_{\rm S}, z_{\rm D}, t = 0) = z_{\rm S}^i\, z_{\rm D}^j$ and the boundary condition, $F(z_{\rm S} = 1, z_{\rm D} = 1, t) = 1$, the fixation probability for an initial population of i malignant stem cells and j progenitors is derived as

$$\rho_{ij} = \frac{1 - (z_{\rm S}^\star)^i\, (z_{\rm D}^\star)^j}{1 - (z_{\rm S}^\star)^{N_{\rm S}}\, (z_{\rm D}^\star)^{N_{\rm D}}}, \tag{3.9}$$

where $(z_{\rm S}^\star, z_{\rm D}^\star)$ is the nontrivial fixed point of the generating Eq. (3.18). For $i = 1$ and $j = 0$ (starting with one initial SC mutant) the fixation probability is

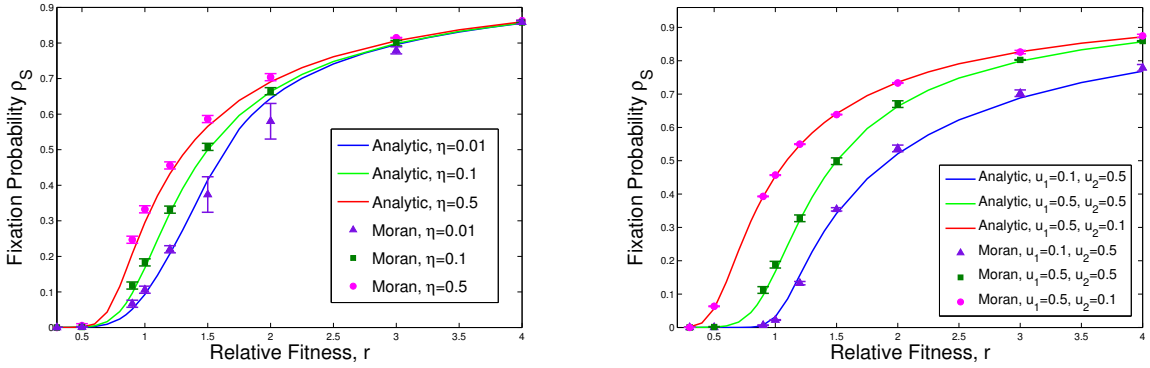
$$\rho_{\rm S} \equiv \rho_{10} = \frac{1 - z_{\rm S}^\star}{1 - (z_{\rm S}^\star)^{N_{\rm S}}\, (z_{\rm D}^\star)^{N_{\rm D}}}. \tag{3.10}$$

Similarly, the fixation probability of a newborn mutant in the DC compartment ($i = 0, j = 1$) is

$$\rho_{\rm D} \equiv \rho_{01} = \frac{1 - z_{\rm D}^\star}{1 - (z_{\rm S}^\star)^{N_{\rm S}}\, (z_{\rm D}^\star)^{N_{\rm D}}}. \tag{3.11}$$

Moreover, assuming random mutations, i.e. uniform mutation rates in both compartments, the average fixation probability is given by

$$\rho = \frac{1 - (N_{\rm S}/N_{\rm tot}) z_{\rm S}^\star - (N_{\rm D}/N_{\rm tot}) z_{\rm D}^\star}{1 - (z_{\rm S}^\star)^{N_{\rm S}}\, (z_{\rm D}^\star)^{N_{\rm D}}}. \tag{3.12}$$



(a) (b)

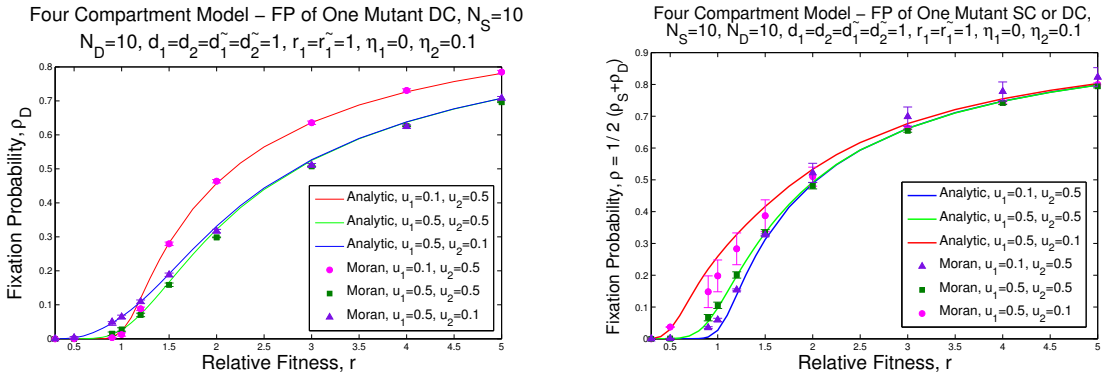
Figure 3.2: **Effect of change in asymmetric differentiation and plasticity rates on survivability of mutants.** We assume that $N_S = N_D = 10$, $r_1 = \tilde{r}_1 = 1$, and $r_2 = \tilde{r}_2 = r$. In subfigure (a), the fixation probability of SCs as a function of η is given, where $\eta = 0.01, 0.1, 0.5$ while $u_1 = u_2 = 0.5$, and $\eta_1 = 0$. In (b) $\eta_1 = 0$ and $\eta = 0.1$. Changing parameters u_1 and u_2, which are the asymmetric division rates of normal and tumor SCs resp., the fixation probability as a function of u_1, u_2 is shown. Solid lines represent the analytic calculation and points correspond to simulation results (error bars are based on the standard error of the mean).

with $N_{\text{tot}} = N_S + N_D$. The probability of a successful emergent mutant *before* time t (from a background of N_{tot} normal cells) is given by

$$P(t) = 1 - e^{-N_{\text{tot}} \cdot \mu \rho t}, \tag{3.13}$$

where μ denotes the mutation rate. In Figure 3.2 a comparison is represented to show how the fixation probability would change with respect to different regimes having asymmetric division and plasticity in the system. In this figure, the effect of variation in parameters u_1, u_2, η_1, and η_2 reveals the fact asymmetric division rate compare with plasticity rate would have more effect on the survivability of cancer (and thus w.t.) cells. Moreover, the probabilities u_1, u_2 of asymmetric division represent a symmetric relation with respect to the case with $u_1 = u_2$. Interestingly when mutants are highly advantageous or disadvantageous dedifferentiation shows a negligible effect. More precisely, when system is approximately close to neutral drift, the fate of the system can be brittle by change in the rate of plasticity.

Now, changing the initial location of the newly born mutant, the dynamics of the system may change (see Figure 3.3-(a)). Regarding the experimental evidence, as this location is



(a) (b)

Figure 3.3: **Dependency of the fixation probability on the initial mutant's phenotype.** Let us suppose that $N_{\rm S} = 10, N_{\rm D} = 10, r_1 = \tilde{r}_1 = 1, r_2 = \tilde{r}_2 = r, \eta_1 = 0$ and $\eta_2 = 0.1$. Having a recently born mutant in DC compartment, the fate of the system may differ than what concluded for the initiation process of a new mutant in SC compartment. In part (a) Moran simulation and analytic calculation , in a perfect agreement, have shown to represent the trend for the fixation probability ρ_D in terms of r. Now, when the location of the newly imposed mutant is either in SC or DC compartment which is the topic of subfigure (b). In this figure the total (average) fixation probability $\rho = \frac{1}{2}(\rho_S + \rho_D)$ is drawn as a function of the relative fitness r. The trends are also compared with the Moran simulation. Moran process was run 5 times with 20,000 iterations in (a) and 50,000 iterations in (b) and the error bars are prepared based on the standard error of the mean.

not determined and mutation may can appear in either compartments due to epigenetic random events as considered in Figure 3.3-(b). One can compare these two scenarios with the previous results of starting from one newborn mutant in SC compartment. Biological observation and mathematical exact approach have shown that SCs control the cellular automata in some organs such as the colon/intestine [144]. However, according to the cancer initiation state hypothesis this might occur in other groups of individuals which has been addressed in Figure 3.3.

The phase change diagram of the case in which only mutants have plastic properties has been sketched in Figure 3.6 for two different values of η_2. Interestingly, there exist a coexistence region (the blue domain) between extinction and fixation phases in which both normal and cancer cells remain in the system according to an imposed balance forced by the system. In such a system, as the asymmetric division rate stays small enough, this coexistence area will not appear and the process would have only two absorbing states as

mutant and w.t. individuals are dominant. For such small enough u, although plasticity exists, the procedure follows the conventional Moran procedure for disadvantageous, neutral, and advantageous mutants. All possible destinies of the system in terms of stability of dominant sub–populations, coexistence of subgroups, and the absorption speed of mutant SCs or DCs can be observed in Figure 3.7 for some given parameters of relative fitness r, and u_i, η_j.

3.2.2 Stochastic simulation

Using the model described above, we performed numerical simulations using such updates until each of the runs tends to saturation in the fraction of SCs and DCs, or until we reach the maximum updating time of T=15,000 for each realization. Then running the whole procedure for 20,000 realizations, we calculated the fraction of results for the fixation probability of SCs and DCs in those runs. Then repeating each calculation for a set of five iterations, we calculated the mean and error bars. Errors are calculated as the standard deviation of the mean.

3.3 Absorbing states: Exact stochastic analysis

Starting from the master equation (3.1) discussed in the previous section, the generating equation can be derived by assuming the probability generating function (PGF) in which coefficients define the probabilities of different possible states after a given time. In continuation, we conclude the fixed points of the given finite Markov chain which in turn leads to the probability of extinction or fixation for malignant individuals starting from a particular state. Moreover, we conclude the phase change diagram in terms of the fixation probability, differentiation, and dedifferentiation rates.

Now let us define the probability generating function as follows for the probability density function $p(n_{\mathrm{S}}, n_{\mathrm{D}}; t)$ of having n_{S} mutant SCs and n_{D} mutant DCs at time t. Then the PGF is

$$F(z_{\mathrm{S}}, z_{\mathrm{D}}; t) = \sum_{n_{\mathrm{S}}, n_{\mathrm{D}}} z_{\mathrm{S}}^{n_{\mathrm{S}}} z_{\mathrm{D}}^{n_{\mathrm{D}}}\, p(n_{\mathrm{S}}, n_{\mathrm{D}}; t) \tag{3.14}$$

It is straightforward to see that the PGF now satisfies in the following partial differential

equation

$$\begin{aligned}\frac{\partial F}{\partial t} &= (z_{\rm S}^{-1}-1)\left\langle (W_{\rm S}^{-}-z_{\rm S}\,W_{\rm S}^{+})\,z_{\rm S}^{n_{\rm S}}z_{\rm D}^{n_{\rm D}}\right\rangle + (z_{\rm D}^{-1}-1)\left\langle (W_{\rm D}^{-}-z_{\rm D}\,W_{\rm D}^{+})\,z_{\rm S}^{n_{\rm S}}z_{\rm D}^{n_{\rm D}}\right\rangle \\ &\simeq (z_{\rm S}-1)\,W_{\rm S}^{+}(\hat{n}_{\rm S},\hat{n}_{\rm D}) + (z_{\rm S}^{-1}-1)W_{\rm S}^{-}(\hat{n}_{\rm S},\hat{n}_{\rm D}) \\ &\quad +(z_{\rm P}-1)W_{\rm D}^{+}(\hat{n}_{\rm S},\hat{n}_{\rm D}) + (z_{\rm D}^{-1}-1)W_{\rm D}^{-}(\hat{n}_{\rm S},\hat{n}_{\rm D}),\end{aligned} \tag{3.15}$$

where the operators are $W_{\rm S}^{\pm}$ and $W_{\rm D}^{\pm}$ are the same as before and $\hat{n}_{\rm S}$ and $\hat{n}_{\rm D}$ are assumed to be

$$\hat{n}_{\rm S} = z_{\rm S}\frac{\partial}{\partial z_{\rm S}}, \qquad \hat{n}_{\rm D} = z_{\rm D}\frac{\partial}{\partial z_{\rm D}}. \tag{3.16}$$

In large-$N_{\rm S}$ ($N_{\rm D}$), we can simplify the large-t limit of the equation for the probability generating function by keeping the linear derivative terms as the leading orders in $N_{\rm S}$ ($N_{\rm D}$). This tends to

$$\begin{aligned}\frac{\partial F}{\partial t} &= (z_{\rm S}-1)\left\{\frac{r_2(1-u_2)\,z_{\rm S}\,\frac{\partial F}{\partial z_{\rm S}} + \tilde{r}_2\,\eta_2\,z_{\rm S}\,\frac{\partial F}{\partial z_{\rm D}}}{\hat{N}_{\rm r,S,D}} - \frac{r_1\,(1-u_1)\,N_{\rm S}\,\frac{\partial F}{\partial z_{\rm S}} + \tilde{r}_1\,\eta_1\,N_{\rm D}\,\frac{\partial F}{\partial z_{\rm S}}}{\hat{N}_{\rm r,S,D}\,N_{\rm S}}\right\} \\ &+(z_{\rm D}-1)\left\{\frac{\tilde{r}_2(1-\eta_2)\,z_{\rm D}\,\frac{\partial F}{\partial z_{\rm D}} + r_2\,u_2\,z_{\rm S}\,\frac{\partial F}{\partial z_{\rm S}}}{\hat{N}_{\rm r,S,D}} - \frac{\tilde{r}_1\,(1-\eta_1)\,N_{\rm D}\,\frac{\partial F}{\partial z_{\rm D}} + r_1\,u_1\,N_{\rm S}\,\frac{\partial F}{\partial z_{\rm D}}}{\hat{N}_{\rm r,S,D}\,N_{\rm D}}\right\},\end{aligned} \tag{3.17}$$

where the operator $\hat{N}_{\rm r,S,P} = r_1\,(N_{\rm S}-\hat{n}_{\rm S}) + r_2\,\hat{n}_{\rm S} + \tilde{r}_1(N_{\rm D}-\hat{n}_{\rm D}) + \tilde{r}_2\,\hat{n}_{\rm D}$ can be considered constant when $N_{\rm S}$ and $N_{\rm D}$ are set equal to $N_{\rm S}, N_{\rm D} \gg 1$.

Setting to zero the coefficients of the derivatives $\frac{\partial F}{\partial z_{\rm S}}$ and $\frac{\partial F}{\partial z_{\rm D}}$, we obtain approximate quasi-stationary points for the constant population model. This relates to the corresponding martingales for $N \to \infty$ branching process limit. Denoting the solutions with $z_{\rm S}^{\star}$ and $z_{\rm P}^{\star}$, one obtains

$$\begin{aligned}&(z_{\rm S}^{\star}-1)\,[r_2(1-u_2)\,z_{\rm S}^{\star} - r_1(1-u_1) - \tilde{r}_1\,\eta_1] + (z_{\rm D}^{\star}-1)\,z_{\rm S}\,r_2\,u_2 = 0 \\ &(z_{\rm D}^{\star}-1)\,[\tilde{r}_2\,(1-\eta_2)\,z_{\rm D}^{\star} - \tilde{r}_1\,(1-\eta_1) - r_1\,u_1] + (z_{\rm S}^{\star}-1)\,z_{\rm D}^{\star}\,\tilde{r}_2\,\eta_2 = 0.\end{aligned} \tag{3.18}$$

An interesting scenario occurs when the normal component is not plastic, i.e. $\eta_1 = 0$, in which the above equations can be simplified to the following expression for $z_{\rm S}^{\star}, z_{\rm D}^{\star} \neq 1$

$$z_{\rm D}^{\star} = \frac{\Big(r_1\,(1-u_2)\,(\tilde{r}_1 + r_1\,u_1)\Big)\,z_{\rm S}^{\star} + r_1\,(1-u_1)\,(\tilde{r}_1 + r_1\,u_1)}{\Big(r_1\,r_2\,(1-u_2)\,(1-\eta_1) - \tilde{r}_2\,r_2\,u_2\,\eta_2\Big)\,z_{\rm S}^{\star} - r_1\,r_2\,(1-u_1)\,(1-\eta_2)}. \tag{3.19}$$

From the equations (3.18), we conclude that

$$\left(A-\frac{B}{z_{\rm S}^{\star}}\right)\left(E-\frac{F}{z_{\rm D}^{\star}}\right)=C\,.\,G, \tag{3.20}$$

where

$$\begin{aligned} A&=r_2\,(1-u_2), \quad & B&=r_1\,(1-u_1)+\tilde{r}_1\,\eta_1, \quad & C&=r_2\,u_2,\\ E&=\tilde{r}_2\,(1-\eta_2), \quad & F&=\tilde{r}_1\,(1-\eta_1)+r_1\,u_1, \quad & G&=\tilde{r}_2\,\eta_2. \end{aligned}$$

Equations (3.20) and (3.18) suggest the below solution for $z_{\rm S}$ which depicts another fixed point of the problem:

$$\begin{aligned} &(A^2E-ACG)\,(z_{\rm S}^{\star})^3+(ACF+C^2G+BCG+ACG-2ABE-A^2E-ACE)\,(z_{\rm S}^{\star})^2\\ &+(B^2E+BCE+2ABE-BCG-BCF)\,z_{\rm S}^{\star}=B^2E. \end{aligned} \tag{3.21}$$

3.3.1 Standard BD Moran process as a particular case

Let us assume that there exist no communication between SC and DC compartments, that is, $u_i,\eta_j\ll 1$ for $i,j=1,2$. In this case, each compartment will follow the conventional BD Moran model as we derive the following fixed points for $z_{\rm S}^{\star}$ and $z_{\rm D}^{\star}$. Regarding the associated transition probabilities in this case, we find the below fixed points in coincidence with what can be derived from two disjoint Moran processes:

(1) $z_{\rm S}^{\star}=1,\ z_{\rm D}^{\star}=1,$

(2) $z_{\rm S}^{\star}=1,\ z_{\rm D}^{\star}=\frac{\tilde{r}_1}{\tilde{r}_2},$

(3) $z_{\rm S}^{\star}=\frac{r_1}{r_2},\ z_{\rm D}^{\star}=1,$

(4) $z_{\rm S}^{\star}=\frac{r_1}{r_2},\ z_{\rm D}^{\star}=\frac{\tilde{r}_1}{\tilde{r}_2}.$

3.3.2 Stability in differentiated compartment

Another interesting case relates to a stationary condition in which no reproduction occurs in DC compartment: $\tilde{r}_1,\tilde{r}_2\simeq 0$. This case will most likely happen after large number of generations which DCs reach to an steady state and do not pursue the proliferation procedure anymore. In such a situation, SCs are also committed to reach a stationary stage. The possible fixed points will be maintained as

(1) $z_S^\star = 1$, $z_D^\star = 1$,

(2) $z_S^\star = \frac{r_1}{r_2}$, $z_D^\star = 1$.

Subfigure 3.4-(a) reveals the phase change for the survival probability of mutant SCs as r_1 and r_2 vary in the given range from 0 to 5. As r_2 increases and r_1 declines, the fixation probability climbs. The lower slopes of lines passing though origin (when slopes are less than one), the higher survivability will be emerged. In the domains under associated lines of the given parameters, the fixed points of the characteristic equation (3.18) for various values of r and u are attractive and the evolutionary scenario is called harmony as cancer SCs dominate.

3.3.3 Invasion model in the absence of plasticity

As a restriction to the model where there is no plastic potential in the system, assuming $r_1 = \tilde{r}_1 = 1$, $r_2 = \tilde{r}_2 = r$, $u_1 = u_2 = u$, and $\eta_1 = \eta_2 = 0$, the system can encounter different scenarios as are listed in the following

(1) $z_S^\star = 1$, $z_D^\star = 1$,

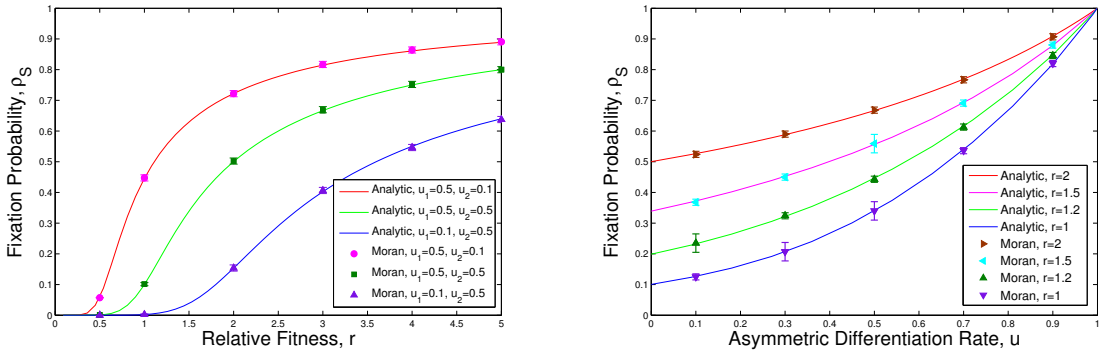
(2) $z_S^\star = \frac{1}{u_{\text{eff}}}$, $z_D^\star = 1$

(3) $z_S^\star = \frac{-r+u_1+u_2\,u_1-1+u_2+\Gamma}{2r(-1+u_2)}$, $z_D^\star = \frac{u_1+1}{r}$.

where $u_{\text{eff}} = \frac{r(1-u_2)}{1-u_1}$ is the effective asymmetric division rate and

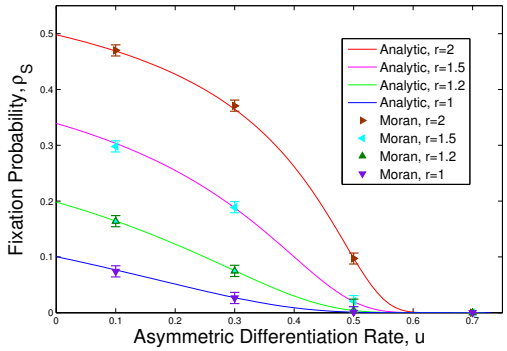
$\Gamma^2 = 1 - 6ru_2u_1 + u_1^2 + 2ru_2 - 2r - 2u_2 - 2u_1 + 2ru_1 + r^2 + 2u_2u_1^2 + 2u_2^2u_1 + u_2^2 + u_2^2u_1^2$.

Figure 3.4 represents the fixation probability variation in terms of the relative fitness of mutant SCs/DCs as well as the probability of asymmetric division in the SC compartment regarding solutions given of case (3). Figure 3.4-(a) shows how the trend for the fixation probability may differ via defecting the balance between asymmetric divisions of normal and malignant SCs. When $u_1 < u_2$ the trend for the survivability of mutants is similar to that for the balanced case $u_1 = u_2 = \frac{1}{2}$ when system remains almost neutral. As r increases, the survival chance of mutants for $u_2 > u_1$ behaves alike the balances case. Another feature to be addressed relates to the dependency of the fixation probability to asymmetric division rate $u_1 = u_2 = u$ where the role of change in $r_1 \leq r_2$ affect the system. Fixing r_2 and rising r_1 the chance for fixation of malignant cells increases for any u but the speed of growth in the fixation probability is linearly dependent to u.



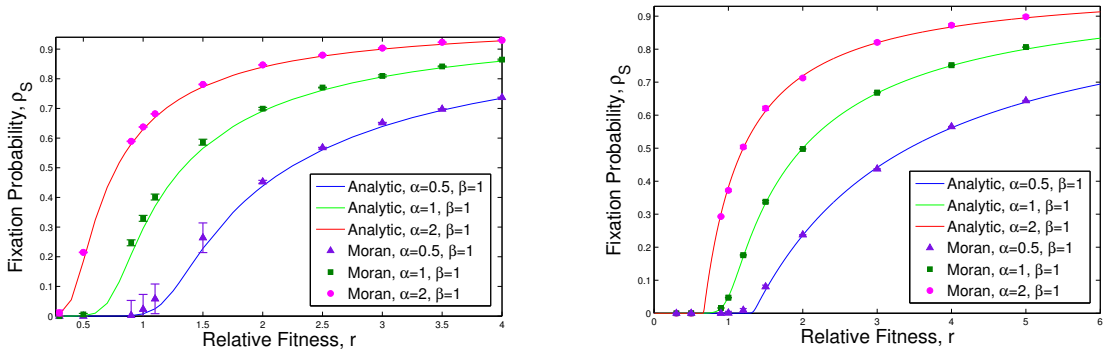
(a)

(b)



(c)

Figure 3.4: **(a) Fixation of mutants in the absence of plasticity.** We assume that $N_S = 10, N_D = 10, \eta_1 = \eta_2 = 0$. (a) changing parameters u_1, u_2, the differentiation rates of normal and tumor SCs respectively, the trends for the fixation probability has shown as a function of relative fitness of mutants: $r_1 = r_2 = r$. In (b) and (c) another observation can be concluded where the variation of fixation probability in terms of asymmetric differentiation rate $u_1 = u_2 = u$ is taken into account for various values of the relative fitness r and the ratio of the differentiation rates of normal SCs: $\alpha = u_2/u_1$ where in (b) $\alpha = 0.5$ and in (c) $\alpha = 1.5$. Numerical simulations are also done for all the case that are shown by colorful points (analytic results are shown as solid curves) with error bars as the standard error of the mean in a set of 5 iterations, each with 20,000 realization.



(a) (b)

Figure 3.5: **Comparison between analytic calculation and Moran simulation among SC and non-SC compartments.** We suppose that $N_{\rm S} = N_{\rm D} = 10, r_1 = \tilde{r}_1 = 1, r_2 = \alpha r, \tilde{r}_2 = \beta r, d_1 = d_2 = 1 = \tilde{d}_1 = \tilde{d}_2 = 1$. Changing α from 0.5 to 1 and then to 2 while β is remained fixed. The disconnected curves are related to the corresponding Moran simulations with a set of 5 iteration=50,000 and error is the standard error of the mean. The best match can be seen when there exist the symmetry in the system and between the two compartments (SCs and DCs) where they have the same population size. In subfigure (a) there exists no plastic potential situation for normal cells, that is $\eta_1 = 0$ while in (b) both w.t. and mutant cells can dedifferentiate to stem-like state.

3.3.4 Invasion model with plasticity

Suppose that dedifferentiation only occurs for the mutant DCs at a rate $\eta_2 = \eta$ for $\eta \gg \eta_1 \approx 0$. Also let assume that $r_1 = \tilde{r}_1 = 1$, $r_2 = \tilde{r}_2 = r$, $u_1 = u_2 = u$ and $\eta_2 = \eta$, then the solutions for $z_{\rm S}^\star$ and $z_{\rm D}^\star$ are

(1) $z_{\rm S}^\star = 1$, $z_{\rm D}^\star = 1$,

(2) $z_{\rm S}^\star$ satisfies in the equation

$$(A^2E - ACG)\,(z_{\rm S}^\star)^3 + (ACF + C^2G + BCG + ACG - 2ABE - A^2E - ACE)\,(z_{\rm S}^\star)^2 + (B^2E + BCE + 2ABE - BCG - BCF)\,z_{\rm S}^\star = B^2E,$$

where $A = r(1-u)$, $B = 1-u$, $C = ru$, $E = r(1-\eta)$, $F = u+1$, $G = r\eta$, and $z_{\rm D}^\star = \frac{(z_{\rm S}^\star+1)(1-u^2)}{r[(1-u-ru\eta)z_{\rm S}^\star-(1-u)(1-\eta)]}$.

The values for $z_{\rm S}^\star$ and $z_{\rm D}^\star$, in this case, satisfy the following relations (for more details refer to Supplementary Information):

$$\left(A-\frac{B}{z_{\rm S}^\star}\right)\left(E-\frac{F}{z_{\rm D}^\star}\right)=C\,.\,G,\quad z_{\rm D}^\star=\frac{(z_{\rm S}^\star+1)(1-u^2)}{r[(1-u-ru\eta)z_{\rm S}^\star-(1-u)(1-\eta)]},\quad (3.22)$$

where $A=r(1-u)$, $B=1-u$, $C=ru$, $E=r(1-\eta)$, $F=u+1$, $G=r\eta$.

The derived solutions introduce the possible fixed points of the characteristic equations. We investigate this case in more detail later and through analyzing the phase diagram of the generalized model for non-zero plasticity in the replicator dynamics.

3.4 Replicator dynamics of the invasion process

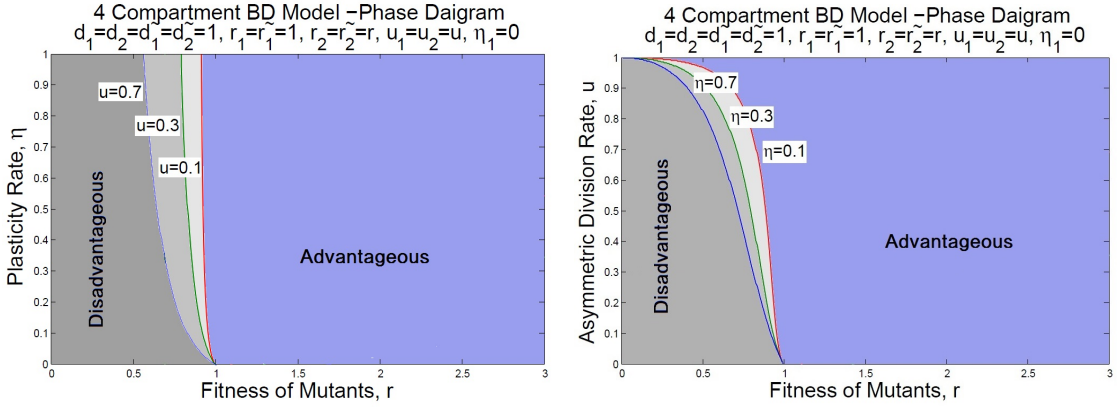
To capture more features of the BD Moran process of the considered multi–compartment model, studying the replicator dynamics of the system which envisages the average frequency of various phenotypes, can be insightful. First of all, the growth trend can be detected for each species, then the phase diagram for variations in parameters can be analyzed. Now, considering the replicator dynamics of the four compartment model which depict the alterations in the the average frequencies of mutant SCs and DCs, respectively

$$x_{\rm S}(t)=\left\langle\frac{n_{\rm S}(t)}{N_{\rm S}}\right\rangle,\quad \text{and}\quad x_{\rm D}(t)=\left\langle\frac{n_{\rm D}(t)}{N_{\rm D}}\right\rangle,\qquad (3.23)$$

one obtains the following system of equations:

$$\begin{aligned}
\frac{dx_{\rm S}}{dt} &\approx \left\langle W_{\rm S}^{+}-W_{\rm S}^{-}\right\rangle \\
&= \frac{[r_2\,(1-u_2)-r_1\,(1-u_1)]\,x_{\rm S}\,(1-x_{\rm S})+\tilde{r}_2\,\eta_2\,x_{\rm D}\,(1-x_{\rm S})-\tilde{r}_1\eta_1\,x_{\rm S}\,(1-x_{\rm D})}{r_1\,(1-x_{\rm S})+r_2\,x_{\rm S}+\tilde{r}_1\,(1-x_{\rm D})+\tilde{r}_2\,x_{\rm D}}, \\
\frac{dx_{\rm D}}{dt} &\approx \left\langle W_{\rm D}^{+}-W_{\rm D}^{-}\right\rangle \\
&= \frac{[\tilde{r}_2\,(1-\eta_2)-\tilde{r}_1\,(1-\eta_1)]\,x_{\rm D}\,(1-x_{\rm D})+r_2\,u_2\,x_{\rm S}\,(1-x_{\rm D})-r_1\,u_1\,x_{\rm D}\,(1-x_{\rm S})}{r_1\,(1-x_{\rm S})+r_2\,x_{\rm S}+\tilde{r}_1\,(1-x_{\rm D})+\tilde{r}_2\,x_{\rm D}}.
\end{aligned}\qquad (3.24)$$

At equilibrium, the fraction of mutant stem and non-stem cell groups would lead to specific states which are the pseudo–fixed points of the problem. These type of fixed points



(a) (b)

Figure 3.6: **Phase diagram of plastic mutant SCs.** The phase boundary for advantageous and disadvantageous mutant populations are given as differentiation and plasticity rates change. We assume that $r_1 = \tilde{r}_1 = 1$, $r_2 = \tilde{r}_2 = r$, $u_1 = u_2 = u$, and $\eta_1 = 0$. Different regions for advantageous and disadvantageous mutant SCs are given in (a) as u changes. A similar analysis has been carried out in (b) as η varies. In (a) $\eta = 0.1, 0.3, 0.7$, here the alteration in the plasticity rate of DCs results in a tendency to approach various regions of fixation for mutant SCs, while the extinction domain shrinks with increasing η. In (b) $u = 0.1, 0.3, 0.7$. Increasing the asymmetric division rate u, the region for advantageous mutants expands to provide a higher survival chance for mutant SCs. In both cases, advantageous criteria relate to either fixation of mutants or coexistence of mutants and WT individuals.

can be attractive or repulsive, depending on the initial conditions of the system. As we described in the paper, according to the cooperation between mutant SCs and DCs (and similarly between normal SCs and DCs), the malignant individuals may become extinct or survive together. Now defining the fixed point as $(x_{\mathrm{S}}^{\star}, x_{\mathrm{D}}^{\star})$, it may tend either to $(0,0)$ or to $(1,1)$. Such criteria suggest two distinct phases for the fate of the malignant cells which are separated by the phase boundary. At steady state, the following system can be derived for the fixed points $x_{\mathrm{S}}^{\star}$ and $x_{\mathrm{D}}^{\star}$:

$$\begin{aligned} &[r_2\,(1-u_2) - r_1\,(1-u_1)]\,x_{\mathrm{S}}^{\star}\,(1-x_{\mathrm{S}}^{\star}) + \tilde{r}_2\,\eta_2\,x_{\mathrm{D}}^{\star}\,(1-x_{\mathrm{S}}^{\star}) - \tilde{r}_1\eta_1\,x_{\mathrm{S}}^{\star}\,(1-x_{\mathrm{D}}^{\star}) = 0, \\ &[\tilde{r}_2\,(1-\eta_2) - \tilde{r}_1\,(1-\eta_1)]\,x_{\mathrm{D}}^{\star}\,(1-x_{\mathrm{D}}^{\star}) + r_2\,u_2\,x_{\mathrm{S}}^{\star}\,(1-x_{\mathrm{D}}^{\star}) - r_1\,u_1\,x_{\mathrm{D}}^{\star}\,(1-x_{\mathrm{S}}^{\star}) = 0. \end{aligned} \tag{3.25}$$

To analyze the latter system, we may consider two important cases:

Case 1. At first, we assume $u_1 = u_2 = \eta_1 = \eta_2 = 0$, then the solutions to the system (3.25) results in the extinction or fixation of mutant SCs and DCs. In this case, the phase diagram is simply divided to two advantageous ($r > 1$) and disadvantageous ($r < 1$) cases.

Case 2. Suppose that $r_1 = \tilde{r}_1 = 1, r_2 = \tilde{r}_2 = r, u_1 = u_2 = u, \eta_1 = 0$, and $\eta_2 =: \eta$. This introduces an interesting scenario in which plasticity occurs only for the cancerous cells but not for the WT individuals. These restrictions simplify system (3.25) to the following possible cases for $x_{\mathrm{S}}^\star$ and $x_{\mathrm{D}}^\star$

(1) $x_{\mathrm{S}}^\star = 0, \quad x_{\mathrm{D}}^\star = 0,$

(2) $x_{\mathrm{S}}^\star = 1, \quad x_{\mathrm{D}}^\star = 1,$

(3) $x_{\mathrm{S}}^\star = 1, \quad x_{\mathrm{D}}^\star = \frac{r\,u}{1-r+r\,\eta}$, and

(4) $x_{\mathrm{S}}^\star = r\,\eta\,\mathcal{A}\,\mathcal{M}_1^{-1}, \quad x_{\mathrm{D}}^\star = \mathcal{A}\,\mathcal{M}_2^{-1},$

where

$$\begin{aligned}
\mathcal{M}_1 &= 3ru^2 + 3r + 2u - 3r^2u^2 + 6r^2u - 6ru - r\eta - 2r^3u - 2r^2\eta u + r\eta u + 2r^2\eta \\
&\quad + r^3\eta u - 3r^2 + r^3 + r^3u^2 - r^3\eta - 1 - u^2, \\
\mathcal{M}_2 &= r^2u + r^2\eta - r^2 - 2ru - r\eta + 2r - 1 + u, \\
\mathcal{A} &= 2r - 1 - r\eta + r^2u + r^2\eta + u^2 - ru - r^2 + r\eta u - ru^2.
\end{aligned} \tag{3.26}$$

Among the given solutions (1)-(4) of the present case, the only acceptable non–trivial solution is (4). The solution (3) does not satisfy the condition $0 \leq x_{\mathrm{D}} \leq 1$ for all possible values of r, u, and η. Moreover, as we described above, the cooperation among mutant cells results in the fixation of both mutant SC and DC groups to the state $(1, 1)$.

The last solution, then, implies that having $(x_{\mathrm{S}}, x_{\mathrm{D}}) \to (0, 0)$ leads to the following solution

$$\mathcal{A} = (u + \eta - 1)r^2 + \left[-u^2 + (\eta - 1)u + 2 - \eta\right] r - 1 + u^2 = 0, \tag{3.27}$$

which characterizes the advantageous and disadvantageous regions for mutants (see Figure 3.6 for more details). Another limit relates to the case where $(x_{\mathrm{S}}, x_{\mathrm{D}})$ approaches $(1, 1)$. One obtains

$$r\,\eta = (r - 1)(u - 1) \tag{3.28}$$

This condition, which is not defined for $r > 1$, does not change the phase diagram and would not have any effect on the selection pressure of the system on mutants.

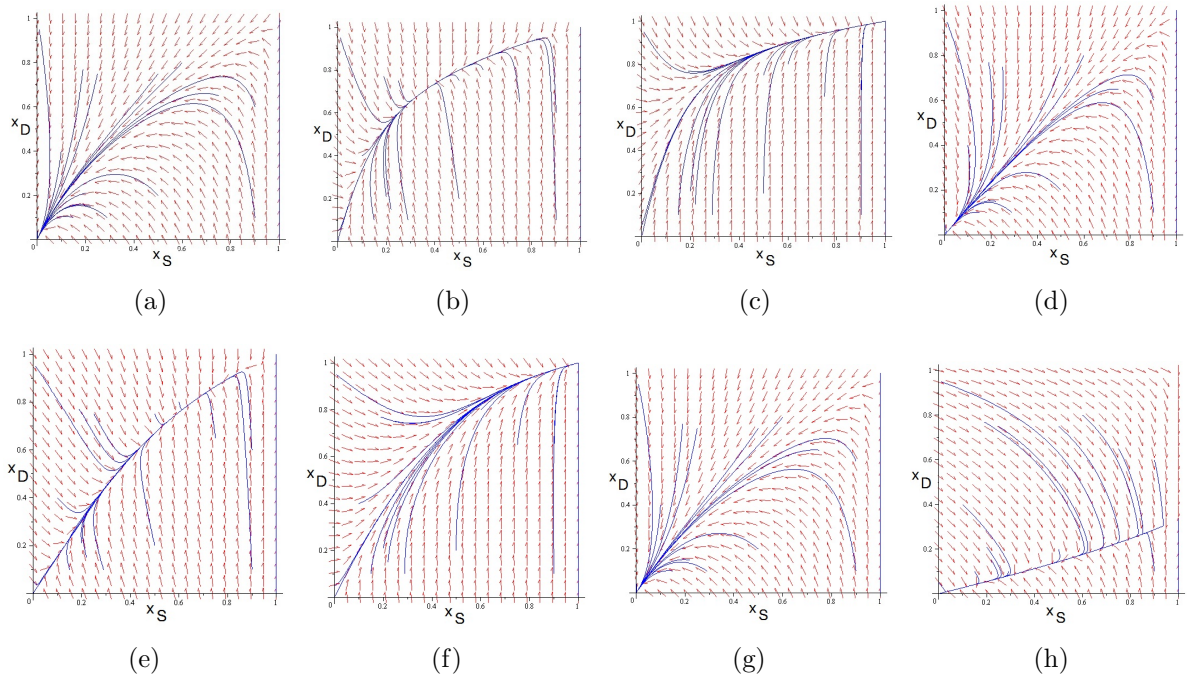


Figure 3.7: **Phase portrait of the four compartment BD model when $\eta_1 = 0$.** Suppose that $r_1 = \tilde{r}_1 = 1, r_1 = \tilde{r}_1 = r, u_1 = u_2 = u, \eta_1 = 0, N_S = 10$, and $N_D = 10$. Different initial conditions would tend to different fate of the system for cancer cells as dominant and coexistence at the stationary state. Process is dominant for Normal DCs in (a), (d), (g), and (h) where the mutants extinct. In (b) and (e) coexistence of both cancer SCs and DCs occurs at steady state. Dominance of cancer SCs/DCs and extinction of w.t. SCs/DCs occur in (c) and (f). We assumed that (a) $r = 1, u = 0.7, \eta_2 = 0.1$, (b) $r = 2, u = 0.7, \eta_2 = 0.1$, (c) $r = 3, u = 0.7, \eta_2 = 0.1$, (d) $r = 1, u = 0.8, \eta_2 = 0.3$, (e) $r = 1.75, u = 0.8, \eta_2 = 0.3$, (f) $r = 3, u = 0.8, \eta_2 = 0.3$, (g) $r = 0.95, u = 0.5, \eta_2 = 0.1$, and finally (h) $r = 1, u = 0.1, \eta_2 = 0.3$.

3.4.1 Application to colorectal cancer

Clonal expansion in colorectal cancer is known to be initiated as a result of mutations occurring at the bottom of the crypt [187]. More recently, Vermeulen *et al.* [181] considered the dynamics of cells at the bottom of a normal colonic/intestinal crypt. Such controlled cells in their circular model of size 5 undergo a selection process. The selection mechanism is investigated for several oncogenes and tumor suppressor genes imposed at the crypt base. Fixation probability of a single mutation and the relative fitness r of the mutant

cells, compared with the normal host, are reported. Their estimates reveal that the relative fitness of the original cell containing $APC^{-/-}$ is $r = 1.58$, while $r = 1.56$ for $Kras^{G12D}$, and $r = 0.96$ for $P53^{R172H}$ (compared with normal control cells mice). P53 mutation seems not to confer a fitness advantage and is weakly deleterious. However for $P53^{R172H}$, it has been observed that the fitness of a mutant elevates from 0.96 to 1.16 in comparison with the DSS-treated cells (colitis) which is in the presence of inflammatory injury. Thus, under inflammatory signaling effects, the $P53^{R172H}$ mutants appear to gain a selection advantage and thus a higher fitness [181, 183].

According to the recent *in vivo* study by Schwitalla *et al* [153], inflammatory signaling plays a role in elevating the rate of dedifferentiation. It has been also shown that inflammatory disease activates the transcription factor NF-κB. NF-κB can, in turn, elevate Wnt-signaling which leads to the phenotypic plasticity of non-SCs [65].

Thus we suggest that the higher survival chance of the P53 mutated SCs along with the DSS-treated cells, may be the result of dedifferentiation in the presence of inflammatory stroma [153]. The survival probability of mutants in colon/intestine, in the presence of inflammatory signaling, is presumably correlated with both the fitness and plastic nature of the epithelial cells. Thus the fixation probability reported in [181] of cells with P53 DSS colitis mutation, can be derived by the same fitness $r = 0.96$ for P53 when dedifferentiation occurs.

For instance, when $r = 0.96$ and $u_1 = 0.5, u_2 = 0.25$, having the plasticity rate at $\eta = 0.12$, one obtains the same fixation probability as [181] (see Table 3.2). This finding strongly suggests that there is an elevation in the fixation of a deleterious mutant into an advantageous trait due to the plastic properties of the mutant (see also the next section). In Figure 3.8, we considered a cylindrical model of the crypt-base, and immediate adjoining transit amplifier layers, cells in the colon/intestine. The bottom layer consists of central stem cells and border stem cells, while transit ampflifying and non-SC cells reside on top of this layer and at higher layers. A mouse crypt is comprised of 5–7 functional stem cells [97, 135, 181]. To compare the results of our model with experimental observations of [181] for the fixation probability of P53 DSS colitis, we consider the first two circular layers of SCs (see Figure 3.8) in which we roughly assume $N_S \approx 10$ [144] and note that the fitness reported in the experimental data [181] is for functional SCs and may be considered unchanged for the two layers ($N_S = 10$) of central (functional) and border SCs as well.

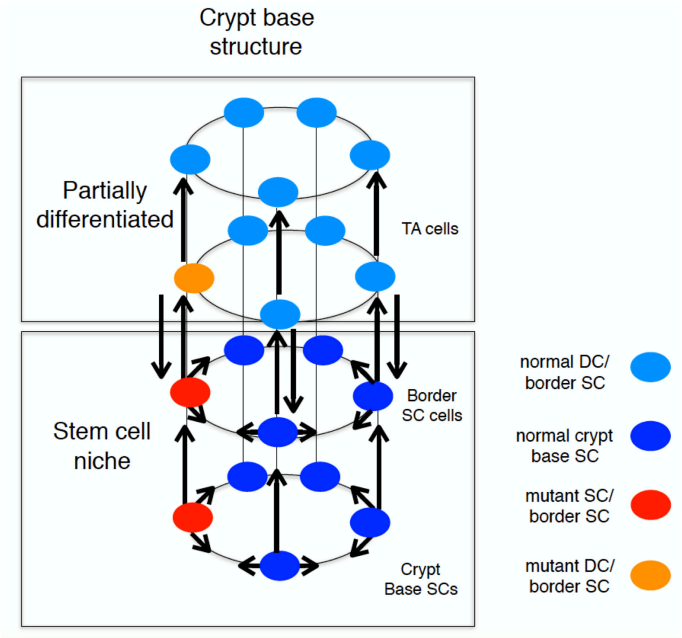


Figure 3.8: **Cellular interactions in the colonic crypt as a newborn mutant arises within the Stem or differentiated compartments.** Within this schematic cylindrical model, we represent how our model is structured through the four compartments of host and mutant stem and differentiated cells. In contrast to the circular model of five SCs considered in [183], we assume a cylindrical model of two circles, one on the top of the other. SCs are located at the bottom circle while the circle on the top is full of partially-differentiated cells.

Table 3.2: Comparison between our analytic result and experiment [181].

Case study	r	η_2	ρ_S	Reference
P53 Controlled	0.96	0	0.113	[181]
P53 DSS Colitis	1.16	0	0.343	[181]
P53 DSS Colitis+Plasticity	0.96	0.12	0.343	Section 2.2.1

3.5 Summary

The evolutionary implications of epigenetic heterogeneity are not very well understood in cancer biology. A known picture for phenotypic heterogeneity (when the genotypes are assumed to be identical) relates to the cancer stem cell hierarchy. In this picture, pluripotent cells with tumor initiating capacity can undergo mitotic events and either replenish their own population or produce a lineage of partially differentiated cells (including precursor and/or transit amplifying cells).

In the current study, we present a general model of four distinct subpopulations to investigate Darwinian evolution in such a hierarchical structure. We consider two genotypically different populations (mutant and wild-type). Mutations are results of unwanted oncogenic or TSG mutations. Each genotype has phenotypically different subtypes of stem cells (SCs) and differentiated cells (DCs). SCs ultimately generate their associated progenies, DCs, through proliferation and asymmetric differentiation. DCs have restricted proliferation capacity. Due to the tissue structure of the crypt, the population of different subtypes remains approximately unchanged. Genetic mutations can occur among SCs or DCs which we assume to be occurring through a uniform probability. Mutations can confer not only higher division rates but also different rates of differentiation and plasticity among mutant subtypes. These changes can be triggered, for example, by microenvironmental conditions.

Our model predicts the fixation probability of a newborn mutant as a function of the division rates of mutant and resident SCs/DCs, differentiation and dedifferentiation rates. Exact analytic calculation and numerical simulations – which are in almost perfect agreement – suggest that the asymmetric differentiation in the SC group has a major effect on the fate of mutants compared with dedifferentiation. More specifically, a greater impact on the fixation probability of SCs can be observed by the change in asymmetric differentiation of normal cells compared with that of malignant cells. Furthermore, we observe that the more plastic trait has an evolutionary advantage. This is most notable close to the neutral limit, i.e. when the proliferation rates of the mutants and residents are very close in value. Most interestingly, a sufficient increase in the rate of plasticity can turn a previously deleterious mutant into a beneficial one.

As an important application of this model, we consider the intestinal/colonic crypt with two groups of SCs and their neighboring, partially differentiated cells. The competition between malignant mutations and normal cells in the base of the crypt has attracted much research interest, and is one the most studied scenarios in cancer evolution. It has recently been shown that the Moran type process, initially suggested by [66, 67, 75, 76, 121, 122] are in almost perfect agreement with the experimental observations [183]. In this

in vivo experimental analysis, KRAS, $APC^{+/-}$, $APC^{-/-}$, and P53 DSS colitis mutations separately induced in the crypt base. Assessing clonal lineage tracing [183], the authors were able to observe the growth of mutants in the populations and thus measured the fixation probability of mutants. However, the experimental observations sometimes ignore the microenvironmental interaction between the neighboring transit amplifying (TA) cells and SCs in the crypt (See [97, 144, 183] for the models and estimated parameter values of population dynamics of the crypt, as well as numbers of central and border SCs at the base of the crypt).

In this investigation, we have suggested a plausible estimation method for the dedifferentiation rate which can give the same value of the fixation probabilities of P53 inactivation in the crypt base as observed by [181]. As the authors in [181] noticed, P53 mutations can maintain advantageous features to invade the crypt base in the presence of inflammatory signaling. We propose that this could potentially be a result of dedifferentiation caused by an inflammatory stroma [153]. This finding supports our idea about the elevation of a deleterious mutant into one with an advantageous trait due to the plastic properties of the mutant.

This research is one of many mathematical approaches recently used to investigate various aspects of tumorigenesis. In our approach, we attempt to dissect a part of the complex machinery in multi–compartmental models, and understand this in greater detail. This study provides various insights concerning some features of initiation and progression of cancer, and suggests possible experimental investigations to confirm some of the theoretical results. The approach taken here may lead to a better understanding of the natural pathological mechanisms in the colonic/intestinal epithelium or any similar four–compartmental structures in ecology, population genetics, and social networks. The particular application of this approach to carcinogenesis may lead to an intense treatment to stop or prevent the development of malignancy.

Chapter 4

The effect of randomness on the fixation probability of mutants: Minority rules

In the previous chapter, the main concentration was on a hierarchical compartmental structure of individuals within a population. More precisely, a mass-action (well-mixed) Moran model was considered within each compartment and compartments were located on the nodes of a graph. Moreover, the fitness of each phenotype assumed to be a constant and diverse types of division in the form of self-renewal and differentiation were taken into account. In this chapter, for which the majority of the presented results are reported in [96], our study is concentrated on a 1D spatial structure in which every individual is positioned on a node of a graph and fitness of each mutant is position and phenotype dependent. We, in fact, study the effect of spatial randomness on the chances of mutant fixation in a population of individuals of a constant size.

Such problems arise in models of cancer initiation and progression, bacterial dynamics, and drug resistance. It turns out that spatial heterogeneity redefines the notion of neutrality, allowing, e.g., a minority of cells (whose fitness values are drawn from the same distribution as that of the wild type) to behave as if they had a selective advantage. The effect can be very significant (increasing the probability of mutant invasion by orders of magnitude), it increases with the standard deviation of the underlying probability distribution and decreases with the skewness. It is the largest when the fitness values of the mutants and wild types are anti-correlated. We discuss the results for a spatial ring geometry of cells (such as that of a colonic crypt), a mass-action (complete graph) arrangement, and also for a line model with reflecting boundaries.

4.1 Introduction

Fixation is the replacement of the initially heterogeneous population with the offspring of just one individual, in a (finite population) birth-death process. The probability of fixation has been widely studied by physicists and mathematicians for almost a century, starting with the early works by Fisher [38], Haldane [49] and Wright [198], culminating in the work of Kimura [73]. These models have been extended in a number of ways, which can be roughly divided into three approaches: Markov chain methods, branching processes, and diffusion approximations (see [132] and the references therein).

The spatial structure and heterogeneity of the network are known to be of significant importance in evolutionary models. Maruyama [104] showed that regular spatial structures do not enhance or suppress selection. Liberman et al. [90] extended this result and showed that the fixation probability of mutants in a large group of graph structures (known as isothermal graphs) coincides with that for the mixed population. We have recently shown that this result is only valid for specific update rules [67]. Many authors have also discussed how the heterogeneity of the spatial structure affects fixation [2, 168, 105, 99, 53]. For instance, Antal et al. [2] showed that the introduction of randomness in a scale-free random graph suppresses the fixation probability. Considerably less effort has been devoted to understanding the effect of heterogeneity in spatial structures due to the spatial fitness distribution and environmental stress on the invasion probability. Here we use the Moran model and study the effect of spatial randomness on the chances of mutant fixation in a population of cells of a constant size.

Envisage the Moran process [116], where in a constant population of asexually reproducing agents (or cells), at each update, a cell is removed and replaced with the offspring of another cell according to some rules. Which cell gets to go and which one reproduces is decided probabilistically. For example, suppose that death (removal) happens with the same probability for all, and reproduction is performed for one of the '*neighboring*' cells, with probabilities proportional to the cells' fitness (see [67] for other rules). The notion of neighborhood is defined by spatial interactions and here we will specifically focus on the mass-action case where the whole population belongs to a cell's neighborhood, and on a circular arrangement where each cell has only two neighbors. Reproduction is assumed to be faithful, such that the offspring cell inherits the type of the parent cell. As the updates continue, the population size remains constant, and eventually with certainty the whole population will be replaced with the offspring of a single cell (and the corresponding type will "fixate" in the population).

Introducing a specific distribution, the main question to address is the effect of random fitness on the survival probability and time to fixation of mutants as one the moments

(of the distribution) varies and the rest of the moments are constant. To answer this question, we suppose that a new malignant mutation occurs among the resident cells with a constant population size. Moreover, all cells are located on the nodes of a graph. Due to the impact of randomness, the fitness of wild–type and mutant cells are then random variables of specific distributions. These random variables might change subject to the change of a moment (while the other moments are assumed to be fixed). More precisely in this chapter, we consider the effect of change in variance of the random fitnesses, while the mean of the random fitnesses are assumed to remain fixed. An important question is at what variance, mutants can take over the entire system. A similar model could be considered for a well–mixed model when randomness occurs within specific regions (we can consider them as some islands which are located on the nodes of a graph). We will not take the last scenario into account, but instead we tackle the more general case where each individual can take a random fitness.

One may also ask about the type of the distribution which makes the fixation probability of mutants increasing or decreasing as variance changes. In fact, the type of the distribution which might have an optimal effect on the survivability of mutants. Also effect of changes in other moments (e.g. skewness or other subsequent moments) of the considered distributions other than the means and variances, may have tremendous impact on the fixation probability or time to fixation. We also try to present an analytic approach which tends to a general mechanism of the fixation probability of circle, complete graph, and line models under moment variations.

4.2 Microenvironmental effects can make a minor subpopulation advantageous

Let us first assume that the fitness of all but one cell is equal to 1, and a single cell has a smaller fitness, $r < 1$ (we will refer to this cell as a mutant, and the rest of the cells as wild type cells). Not surprisingly, the mutant cell in this case will have a smaller fixation probability compared to any of the wild type cells. In fact, in the mass-action problem the probability of such disadvantageous mutant fixation, P_N, is exponentially small in terms of the population size N, and (in the case of a large population) is given by $(1/r - 1)r^N$. In the opposite case of an advantageous mutant ($r > 1$), the probability of mutant fixation is larger than the probability of fixation of wild type cells and is given by $1 - 1/r$ for large populations. Similar results also hold in the case of a circular geometry, where the probabilities of disadvantageous and advantageous mutant fixation are given respectively by $2r^{N-2}(1 - r)/(3 - r)$ and $2(r - 1)/(3r - 1)$ [75]. Not surprisingly, if the fitness of the

mutant is exactly the same as the fitness of the rest of the cells ($r = 1$, a "neutral" mutant), the probability that the whole population is eventually replaced by the offspring of such a mutant is exactly the same as the probability of fixation of any other cell, and is equal to $1/N$ (this result holds both for mass action and for circular geometry).

The simple situations described above are idealizations because, for example, they do not include the inhomogeneity of the environment. Let us suppose that the fitness of a cell depends not only on its type but also on the spatial location. We will assume that the fitness of a wild type cell at a given location is drawn from a fixed probability distribution. Also we suppose that $\langle P_N \rangle$ is the average fixation probability of an imposed mutant within a population of size N where the average is taken over the all possible configurations for the fitness of each individual located at N different locations. We further assume that the fitness of a mutant cell in a given location is also taken from exactly the same distribution. In other words, for a single realization of a Moran process we need to assign the finesses of wild type and mutant cells for each location, and these values remain fixed throughout the realization. What is the probability of mutant fixation (starting with one mutant cell at a randomly chosen location) averaged over realizations? Simple intuition would suggest that it is $1/N$, because the fitness values of the mutant are taken from the same distribution as those for the wild type, and the probability is averaged over all the possible realizations. We show that the answer $1/N$ however is only true for $N \leq 3$. Starting from $N = 4$, the probability of fixation, $\langle P_N \rangle$, of mutant in such a random landscape is greater than $1/N$. In fact, for relatively small values of N the function $\langle P_N \rangle N$ grows linearly with N and for larger population sizes it continues to grow but becomes logarithmic (to compare, $\langle P_N \rangle N = 1$ in the absence of randomness). One should note that when the mutants are advantageous then $\langle P_N \rangle N$ grows linearly with N and the slope increases with the standard deviation of the fitness distribution. If the fitness of the wild type is constant ($r = 1$), but the fitness of a mutant cell at a given location is selected from a fixed probability distribution, then the average fixation probability is negatively correlated with the variance of fitness distribution of mutants (for both advantageous and neutral mutants) [99]. This could be due to the fact that increments and decrements of fitness values compared to the background fitness do not have a symmetric effect on the fixation probability. For example, if $r = 1$ and $\tilde{r} = 1 \pm 0.5$, then the devastating effect of hitting the value $\tilde{r} = 0.5$ cannot be compensated by a relatively mild advantage of hitting $\tilde{r} = 1.5$.

Consider a circle with $N = 3$ individuals. We denote the fitness for the wild (mutant) type by a_i ($\tilde{a}_i$), $i = 1, 2, 3$. Further, P_s, where s is a binary number of length 3, stands for the probability of mutant fixation starting with configuration s, where "1" stands for

mutant and "0" for the wild type. Defining $f_\beta^\alpha := \alpha/(\alpha+\beta)$, we obtain,

$$\begin{aligned}
3P_{100} &= P_{110}f_{a_3}^{\tilde{a_1}} + P_{101}f_{a_2}^{\tilde{a_1}} + P_{100}(f_{\tilde{a_1}}^{a_2} + f_{\tilde{a_1}}^{a_3}),\\
3P_{010} &= P_{011}f_{a_1}^{\tilde{a_2}} + P_{110}f_{a_3}^{\tilde{a_2}} + P_{010}(f_{\tilde{a_2}}^{a_1} + f_{\tilde{a_2}}^{a_3}),\\
3P_{001} &= P_{101}f_{a_2}^{\tilde{a_3}} + P_{011}f_{a_1}^{\tilde{a_3}} + P_{001}(f_{\tilde{a_3}}^{a_1} + f_{\tilde{a_3}}^{a_2}),\\
3P_{110} &= 1 + P_{100}f_{\tilde{a_1}}^{a_3} + P_{010}f_{\tilde{a_2}}^{a_3} + P_{110}(f_{a_3}^{\tilde{a_1}} + f_{a_3}^{\tilde{a_2}}),\\
3P_{101} &= 1 + P_{100}f_{\tilde{a_1}}^{a_2} + P_{001}f_{\tilde{a_3}}^{a_2} + P_{101}(f_{a_2}^{\tilde{a_1}} + f_{a_2}^{\tilde{a_3}}),\\
3P_{011} &= 1 + P_{010}f_{\tilde{a_2}}^{a_1} + P_{001}f_{\tilde{a_3}}^{a_1} + P_{011}(f_{a_1}^{\tilde{a_2}} + f_{a_1}^{\tilde{a_3}}).
\end{aligned} \tag{4.1}$$

The solution to this system, averaged over all realizations of the fitness configurations for a given distribution is $P_{100} = P_{010} = P_{001} = 1/3$ and $P_{110} = P_{101} = P_{011} = 2/3$. This results in the fixation probability $P_3 = 1/3$, starting from one mutation. The above formulation can be generalized to larger N, but against expectations, the probability of fixation starting from one mutant for $N \geq 4$ is greater than $1/N$ (see Figure 4.1 for a circle with $N = 5$ and also Appendix A.3 for more details on analytic calculations).

Remark. As will be shown below, the trend for the fixation probability of mutant(s) for $N = 3$ is different than those for $N \geq 4$. The reason for such a difference may relate to the neutrality of the system with respect to increase and decrease in the number of mutants (and thus normal individuals) after averaging over all possible configurations. More precisely, in average the probability of increase by one from the state $i = 2$ is equal to that of decrease by one from the state $i = 1$. Moreover, in average probability of increase by one from the state $i = 1$ is equal to that of decrease by one from the state $i = 2$. These symmetric relations among the transition probabilities makes this case (with $N = 3$) neutral and does not occur for any larger populations.

The linear algebraic system can be solved exactly for relatively small values of N. For larger N, we perform numerical simulations of the Moran model (the results are verified by solving the corresponding Kolmogorov system of equations for 10,000 realizations). The results for a circle are presented in Figure 4.2, where the fixation probability starting from one mutant, $\langle P_N \rangle$, times N is plotted in terms of N. For relatively small values of N, this quantity increases linearly with N, suggesting that "neutral" (that is, obeying the same fitness probability distribution) mutants behave similar to the advantageous mutants in the non-random systems, where the probability of fixation approaches a constant as system size increases. For larger values of N, we observe that $\langle P_N \rangle N$ grows slower than linear but is still increasing. The large N behavior is obtained by fitting the data to a

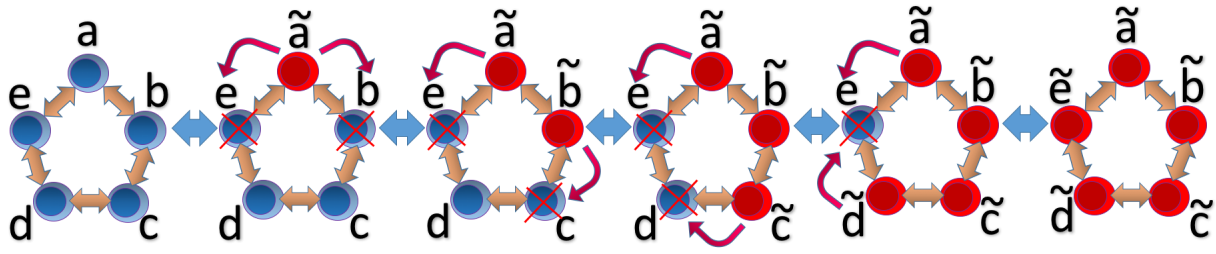


Figure 4.1: **Schematic evolution of a recently born mutant within the cycle model with $N = 5$.** The blue cells are assumed to be normal and red cells are cancerous. At each updating time, either a normal cell or a mutant is chosen for death at random and an adjacent neighbour cell which has a higher proportional fitness will replace its offspring. This DB Moran process may increase the number of mutants (moving to the right on the figure) or decrease their fraction (move to the left on the figure).

logarithmic function of the form of $a + b\ln(N + c)$, where the constants a, b and c depend on the standard deviation of the fitness distribution. In Figure 4.2 we used the fitness distribution given by $1 - \sigma$ and $1 + \sigma$ with equal probability; later in the chapter, we will refer to this distribution as the bimodal distribution.

4.2.1 Distribution of the average fixation probability

To understand the general behavior of the mechanism in the presence of randomness, we consider the whole hierarchical Markov chain. Then the average fixation probability will be calculated based on all the derived fixation probabilities. A question of interest is then, assuming that the distribution of the whole scenarios are the same, does the distribution of the average scenario have the same type? This can be investigated by applying the central limit theorem [41] . To detect such a similarity between the average distribution and the primitive distributions, the boundedness of the mean and variance of the average distribution and their relationships to those of the priori distributions should be investigated. In Appendix A, such an analysis has been performed for populations of size 10, 15, 20, and 50 and all verify that the distribution of the average trait represents a bounded distribution with skewness to the left (towards lower values than the mean of the fitness) and as we increase the variance σ^2 the distribution shifts more to the left. This can be also thought of as a planar wavelet moving towards the lower fitnesses as the variance climbs. This wavelet shows a more fat tail on the left (higher skewness to left) as the population size grows (see Figure (A.1) and the related section for more details.).

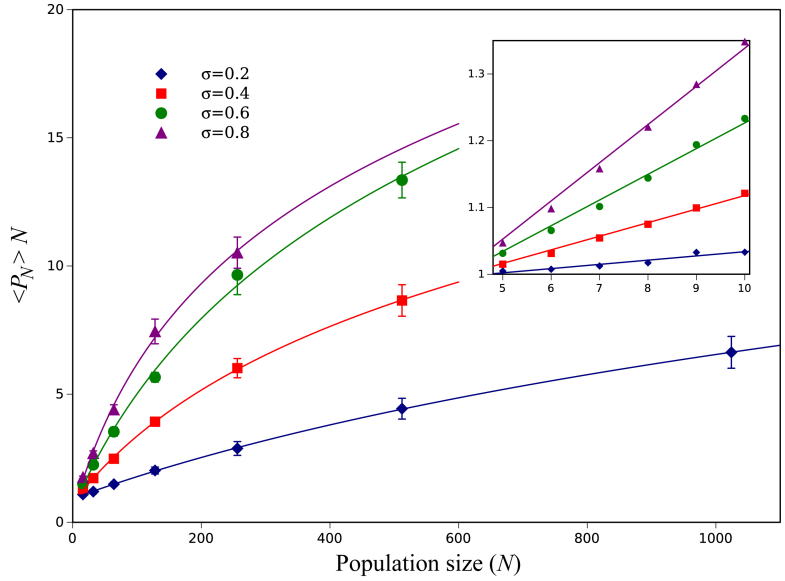


Figure 4.2: **The average fixation probability times N as a function of N for a circle.** The fitness of both mutants and wild types is given by $1-\sigma$ and $1+\sigma$ with equal probability (the bimodal distribution). The inset shows the behavior for small values of N. Points are the results of numerical simulations based on stochastic simulation and bars are the standard deviation of mean. Stochastic simulation are based on a set of 3 runs, each with 20,000 iterations . Solid curves reveal the analytic calculation results.

4.2.2 Randomness impact on the all individuals

Throughout this section, we will envisage the behavior of a population of normal and mutant individuals in the presence of randomness. Considering the same bimodal distribution for normal and mutant individuals with the same mean μ and variance σ^2 (for simplicity we assume that $\mu = 1$ or normalize the means to 1 otherwise), the system does not change the mean but variance can vary within a certain range. As can easily be detected, finding the fixation probability of a 1D structure under influence of random effects tends to a higher dimensional Markov chain which leads to a system of Kolmogorov equations based on the initial mutant's location, various types of individuals, and their fitness at each time step. In this situation, an interesting approach can be to observe the average treatment of the system by considering the whole possible scenarios from beginning to end separately, and then to take the average of their corresponding fixation probabilities with respect to a given probability distribution. Although the consequent results represent average features of the mechanisms, the mean field approximation captures the overall trend of the system.

As the fitness probability distributions of mutants and wild type cells are identical, the only factor that differentiates mutants from the wild types in this setting is the fact that initially, they are a minority. The results reported above (that is, the probability of mutant fixation greater than what is expected of a neutral mutant) hold as long as the mutants are initially a minority (i.e. $\langle P_N \rangle > i/N$ as long as $i < N/2$, where i is the initial number of mutants, see Figure A.7 in the Appendix).

The magnitude of the effect depends on the standard deviation and the higher moments of the underlying probability distribution, and can be quite significant. The difference between $\langle P_N \rangle N$ in the random and the non-random neutral cases is only about 1% for $N = 4$ and becomes orders of magnitude greater for larger N. The larger the standard deviation, the greater the deviation of $\langle P_N \rangle$ from the neutral result for small values of N, but at the same time the lower the size N at which the behavior of $\langle P_N \rangle N$ becomes sub-linear, see Figure 4.2. If we keep both the mean and the standard deviation of the fitness distributions constant, and vary the third moment (skewness), we observe that $\langle P_N \rangle N$ is the largest for the largest negative skewness, Figure 4.3.

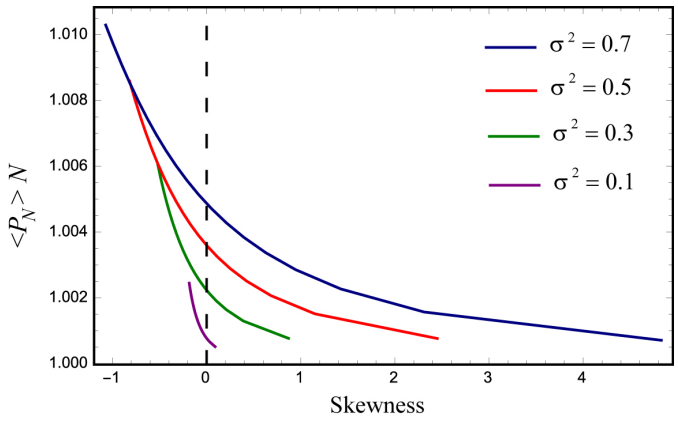


Figure 4.3: **The effect of skewness on the survivability of mutants.** The function $\langle P_N \rangle N$ in terms of the skewness ($N = 4$). We assume $\langle r \rangle = \langle \tilde{r} \rangle = 1$ and $\sigma = \tilde{\sigma}$. Curves are based on analytic calculations.

The beneficial effect of randomness on the chances of a minority fixation described here are observed in the presence of correlations between the wild type and mutant fitness values. We define $\rho = \langle (r - \langle r \rangle)(\tilde{r} - \langle \tilde{r} \rangle) \rangle / (\sigma \tilde{\sigma})$, where the Pearson correlation index ρ can vary within the interval $[-1, 1]$. Then r and $\tilde{r}$ are uncorrelated, fully-correlated, or anti-correlated if $\rho = 0, 1$ or -1, respectively. We plot the fixation probability as a function of the standard deviation for these three cases in Figure 4.4 (for $N = 4$). It is observed

that fully anti-correlated fitness values (with the Pearson correlation index of -1) yield the largest magnitude of the effect.

Remark. Interestingly, when r and $\tilde{r}$ are fully correlated, the fixation probability of mutants (and thus the wild-type individuals) remains fixed with respect to the change in standard deviation. Similar to the neutrality of the system for $N = 3$, one can show that for a population size $N \geq 4$ and for any $2 \leq i \leq N - 2$, after averaging over all possible configurations of the system, the average transition probabilities for an increase (a decrease) by one from the state i is equal to that for decrease from the state $i + 1$ $(i - 1)$. Thus the trend for fully correlated case in Figure 4.4 remains unchanged with respect to the standard deviation.

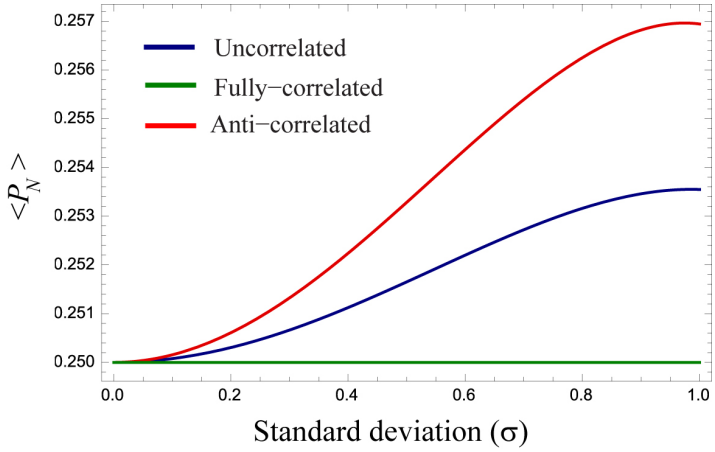


Figure 4.4: **Correlation among the fitnesses of mutants and normal individuals.** The average fixation probability as a function of the standard deviation of the bimodal fitness distribution for different degrees of correlation between mutant and non-mutant fitness values ($N = 4$). The solid curves are based on analytic calculations.

In a more general case one can assume that the distributions of wild type and mutant fitness values are not the same. Figure 4.5 demonstrates that mutants whose fitness values come from a distribution with a *lower* mean could behave as if they are advantageous (that is, fixate with a probability larger than $1/N$). This shows that being a minority can compensate even for having a lower mean fitness.

Next, let us suppose that while having equal means, the standard deviation of the two distributions can be different (where larger standard deviations signify larger randomness). It turns out that for the minority population, it is beneficial to be as deterministic as

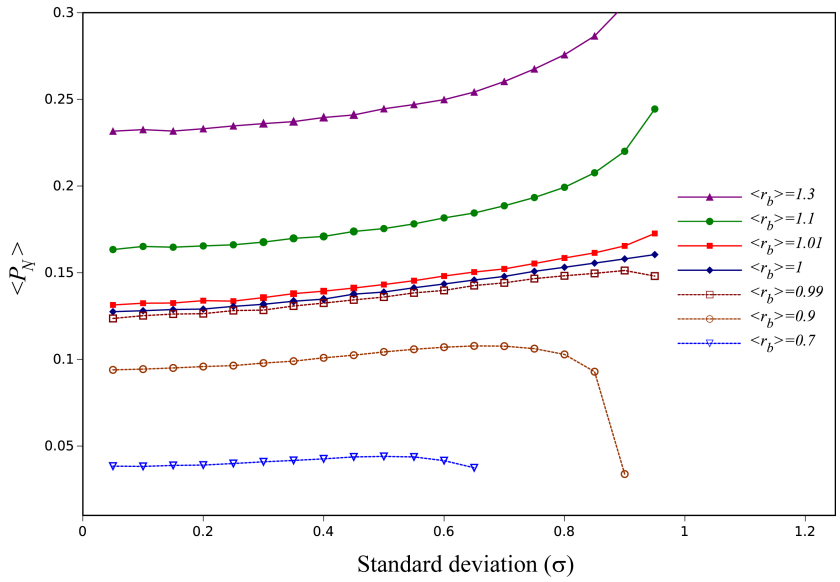


Figure 4.5: **Minority and the average fitness $\langle r\rangle > 1$, $\langle r\rangle = 1$, and $\langle r\rangle < 1$.** The average fixation probability as a function of the standard deviation of the bimodal fitness distribution ($N = 8$). Solid curves show the results of analytic approach while the points (with bars as the standard deviation of mean) represents the results of stochastic simulations. Stochastic simulation are based on a set of 3 runs, each with 20,000 iterations

possible. Larger standard deviations of the background distribution and smaller standard deviations of the mutant distribution lead to higher values of mutant fixation probabilities (Figure 4.6).

For a circle of fixed size N, the mutants grow as a cluster and we only need to consider the cells on the boundaries of mutants and wild type. Our results indicate that the fixation probability times N grows linearly with N for relatively small population sizes and becomes logarithmic for large N. For large N, the circular arrangement with random fitness distribution is similar to the problem of random walks in a random environment. Suppose a walker goes to the right (left) with a probability p_i ($q_i = 1 - p_i$), where p_i and q_i are independent, identically distributed random variables. We define $\rho_i = p_i/q_i$ and $\eta = \langle \ln \rho_0 \rangle$. The random walk is transient if $\eta \neq 0$ and it is recurrent when $\eta = 0$ [167, 160]. In the case that the random variables p_i take only two values σ and $1-\sigma$, with probabilities α and $1-\alpha$, respectively, then the walker is recurrent if $\alpha = 1/2$ or $\sigma = 1/2$.

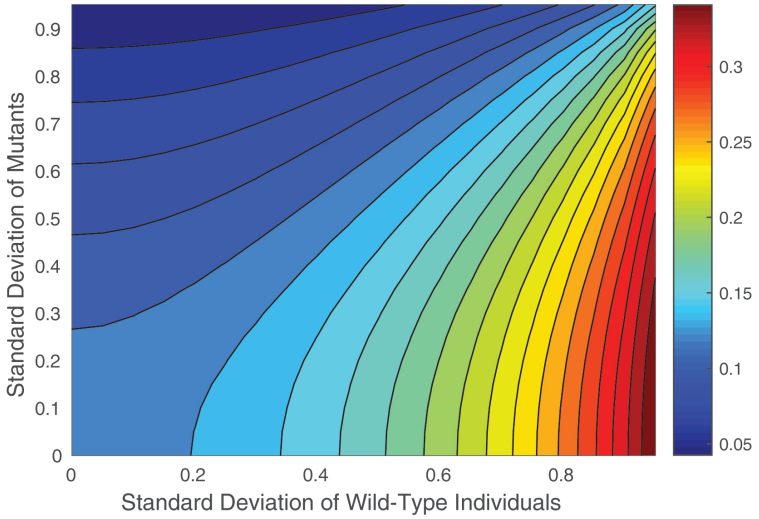


Figure 4.6: **Relationship between standard deviations of normal and cancer individuals.** The heat plot for $\langle P_N \rangle$ as a function of the two standard deviations ($N = 8$). This result is based on analytic calculation of the average fixation probability.

4.3 Complete graph (well-mixed) model equipped with random effect

An interesting observation would be to compare the behavior of the circle model with that for mixed population of the same size in the presence of randomness. To this end, we consider a complete graph with finite number of nodes where randomness will impose a random fitness to each particular node.

Similar to what we discussed about the analytic approach of the circle model, we assume a heterogeneous system of normal and cancerous cells located on a complete graph of size N. Again associating fitnesses $\tilde{a}_1, \cdots, \tilde{a}_N$ to mutant cells and $a_1, \cdots, a_N$ for normal cells based on the type of the cell locating on nodes labeled as 1 to N. We also consider P_σ as the probability of starting from state $\sigma = (\sigma_1, \cdots, \sigma_N)$ and absorbing to the state $(1, 1, \cdots, 1)$ for $\sigma_i \in \{0, 1\}$ $(1 \leq i \leq N)$ in which $\sigma_i = 0$ means that a normal individual is located at node i and otherwise we have a mutant cell in that location.

Writing the corresponding Kolmogorov equations, we have a system of $2^N - 2$ equations

of $2^N - 2$ variables as follows $(0 < M1 < N)$

$$\begin{cases} N\,P_\sigma = \sum_{i\in\mathcal{I}_1} P_{\sigma_i}\,[\mathcal{A}_i] + P_\sigma\,[-N + \mathcal{C}_2]\,, & \|\sigma\| = 1 \\ N\,P_\sigma = \sum_{i\in\mathcal{I}_1} P_{\sigma_i}\,[\mathcal{A}_i] + \sum_{j\in\mathcal{I}_2} P_{\sigma_j}\,[\mathcal{B}_j] + P_\sigma\,[-N + \mathcal{C}_1 + \mathcal{C}_2]\,, & 2 \le \|\sigma\| \le N-2 \\ N\,P_\sigma = 1 + \sum_{j\in\mathcal{I}_2} P_{\sigma_j}\,[\mathcal{B}_j] + P_\sigma\,[-N + \mathcal{C}_1]\,, & \|\sigma\| = N-1, \end{cases} \tag{4.2}$$

where $\sigma_i, \sigma_j \in \{0,1\}$ for $1 \le i,j \le N$ and the transition probabilities of increase and decrease by one from the state σ are respectively

$$\begin{aligned} \mathcal{A}_i(\sigma) &= \frac{\sum_{k\in\mathcal{I}_1} \tilde{a}_k}{\sum_{k\in\mathcal{I}_1} \tilde{a}_k + \sum_{l\in\mathcal{I}_2\setminus\{i\}} a_l}, \qquad \text{for} \quad 1 \le \|\sigma\| < N-1, \\ \mathcal{B}_j(\sigma) &= \frac{\sum_{k\in\mathcal{I}_2} a_k}{\sum_{l\in\mathcal{I}_1\setminus\{j\}} \tilde{a}_l + \sum_{k\in\mathcal{I}_2} a_k}, \qquad \text{for} \quad 1 < \|\sigma\| \le N-1. \end{aligned} \tag{4.3}$$

In the latter expressions, when $\|\sigma\| = N-1$ then $\mathcal{A}_1 = \mathcal{A}_2 = 1$ and for $\|\sigma\| = 1$ we can conclude that $\mathcal{B}_1 = \mathcal{B}_2 = 1$. Also the probabilities of having the state unchanged for mutant and normal cells are respectively

$$\begin{aligned} \mathcal{C}_1(\sigma) &= \sum_{i\in\mathcal{I}_1} \frac{\sum_{k\in\mathcal{I}_1\setminus\{i\}} \tilde{a}_k}{\sum_{k\in\mathcal{I}_1\setminus\{i\}} \tilde{a}_k + \sum_{l\in\mathcal{I}_2} a_l}, \\ \mathcal{C}_2(\sigma) &= \sum_{j\in\mathcal{I}_2} \frac{\sum_{k\in\mathcal{I}_2\setminus\{j\}} a_k}{\sum_{l\in\mathcal{I}_1} \tilde{a}_l + \sum_{k\in\mathcal{I}_2\setminus\{j\}} a_k}, \end{aligned} \tag{4.4}$$

where the initial conditions are $P_{00\cdots0} = 0, P_{11\cdots1} = 1$.

We calculate the fixation probability for a complete graph, where all cells are connected, using both analytic calculations and numerical simulations, Figure 4.7. The results indicate a similar behavior as the circle, i.e. $\langle P_N\rangle N$ grows linearly with N for small N ($N \lesssim 100$), and becomes sublinear for larger population sizes. Similar to the circular case, the lines in Figure 4.7 are obtained by fitting the results to a logarithmic function $a + b\ln(N + c)$.

To compare the circle and complete graph structures in terms of their phase change, in Figure (4.9) the phase change diagram of the circle is also given. There exist two different phases for the complete graph: (i) for the advantageous mean of the fitness, trends are similar to those for the circle model as the curve reveals a monotonic growth. (ii) for the neutral and advantageous mean of the fitness, there is a bump and downward motion for the curves as variance is large enough.

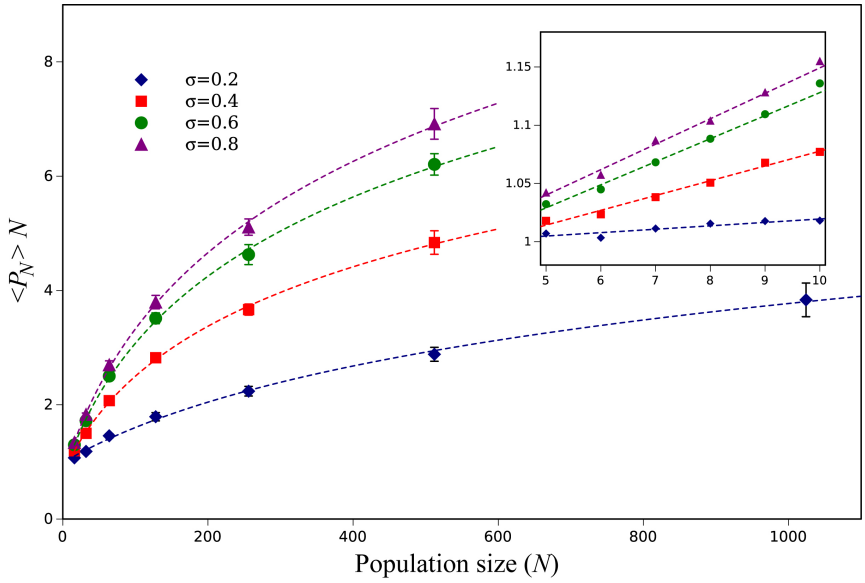
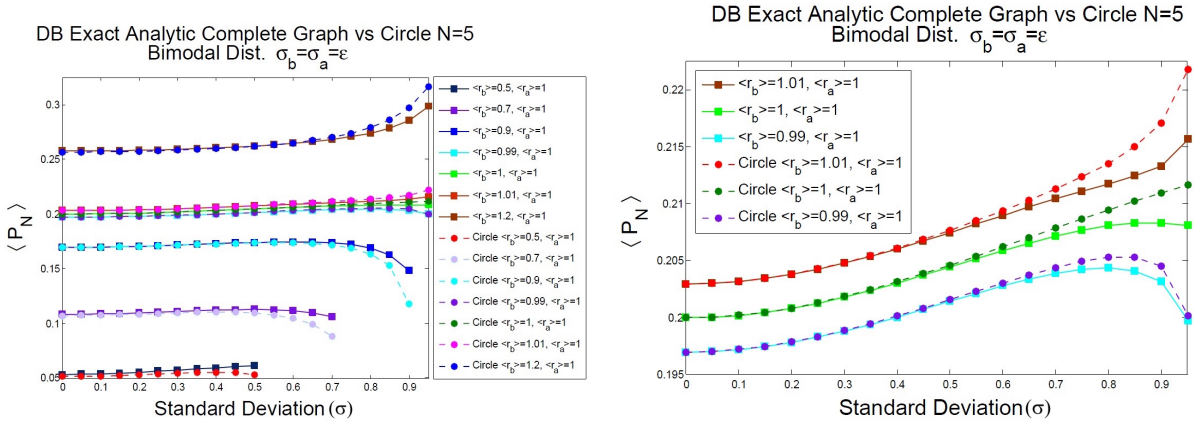


Figure 4.7: **Large population scale and randomness impact on the fixation of mutants.** The average fixation probability times N in terms of the standard deviation of the bimodal fitness distribution for a complete graph. The dash curves are based on analytic approach and points and error bars are based on stochastic simulation of a set of 3 runs, each with 20,000 iterations.

In contrast to what occurs for the circle model in the neutral mean of the fitness, the treatment of the average fixation probability is not monotonically increasing. It is in fact similar to the disadvantageous mean of fitness. In Figure (4.9)-(b) some curves around the neutral case are magnified and show this important difference between the circle and the complete graph structures.

When $N \geq 5$ and $\langle r_b \rangle = \langle r_a \rangle = 1$ as the tendency of the curve for the circle model climbs strictly, the graph for the complete graph provides a bump for large enough variances. More precisely, the trend moves downward as the standard deviation increases. Increasing the population size, the downward tendency occurs even for smaller values of the variance and the difference between the two graphs increases (in Figure (4.9) various case are given for various populations). For $N = 4$ the progression is different and the complete graph represents a stronger impact on the system to enhance the fixation probability. The case for a complete graph with $N = 3$ is similar to the graph of this case for the circle as a straight line due to their same structures.



(a) (b)

Figure 4.8: **Phase diagram of the complete graph vs the circle model.** Complete graph and the circle models are compared for different regimes for $N = 5$ and $\sigma_b = \sigma_a = \sigma$. Different regimes pertaining to different means of mutants have been considered in (a). In (b) a magnified version of the curves related to $\langle r_b \rangle = 0.99, 1, 1.01$ are given whilst $\langle r_a \rangle = 1$. Opposite to what we had for the circle model, a decline can be observed for larger values of σ in well-mixed model. We assumed that $N = 100$. data points are based on analytic calculation for the average fixation probability of mutants on complete graph and the circle model.

4.4 The dynamics of the line model: Significance of reflecting boundary and randomness

In some tissues in the body such as the brain or those epithelial layers surrounded by bones or vascular organs, the location of the newborn mutant within a homogeneous population can have a significant effect on its extinction or fixation.

Within the present section, we consider a new geographic structure in which the population is positioned on a line with finite nodes. Over this spatial structure, a higher dimensional Markov chain can be established to study the spatial effect of the structure on the fixation probability of the newborn mutant. We first start our investigation with the line model in the absence of randomness. Then taking advantage of this analysis, we continue our study to identify the effect of randomness on the survival probability in this 1D spatial model with a reflecting boundary. To maintain the crucial features of the structure, we start with the transition probabilities of a DB Moran process for the given

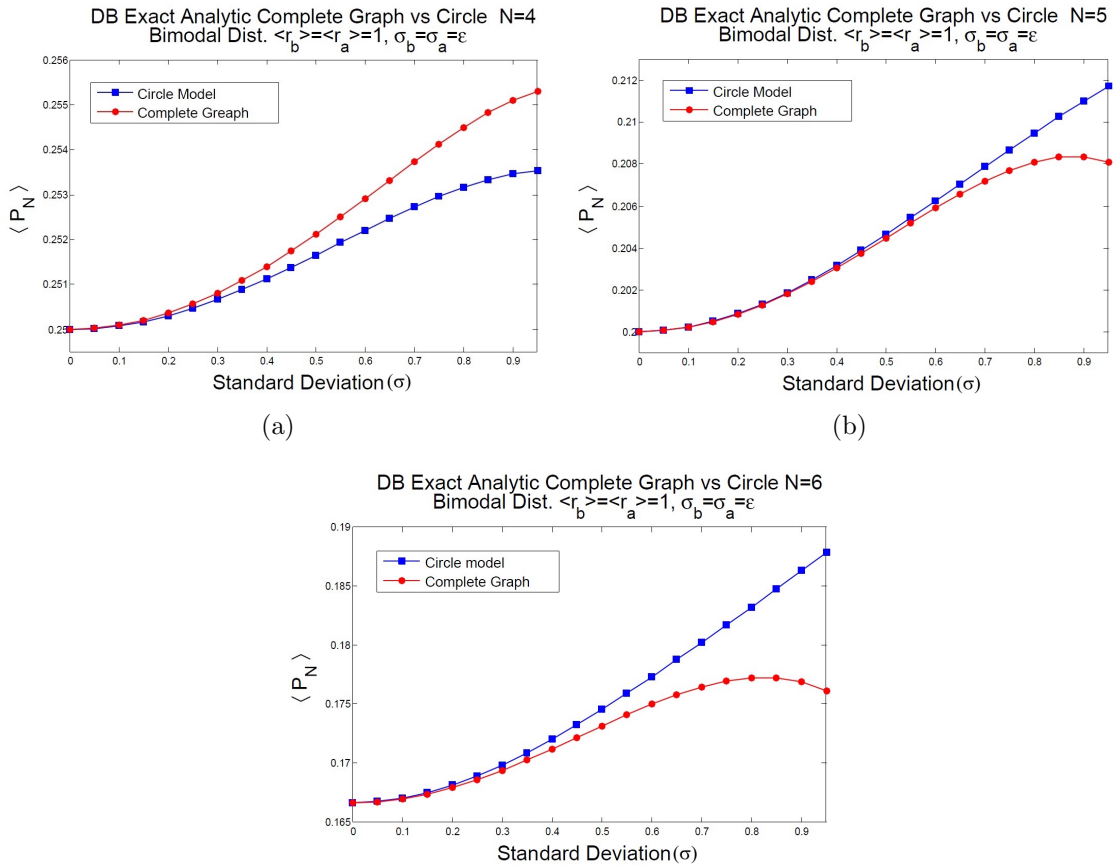


Figure 4.9: **A comparison between the well-mixed model and the circle model both with the same population size.** In figure (a) when $N = 4$ the well-mixed model shows a sharper enhancement for the fixation probability while for larger population size and when $\sigma_b = \sigma_a = \sigma$ is large enough the circle model represents more survival probabilities. For instance in (b) this behavior can be seen for $N = 5$. A similar figure is also given in (c) for $N = 6$. Points are based of exact calculations.

line Model.

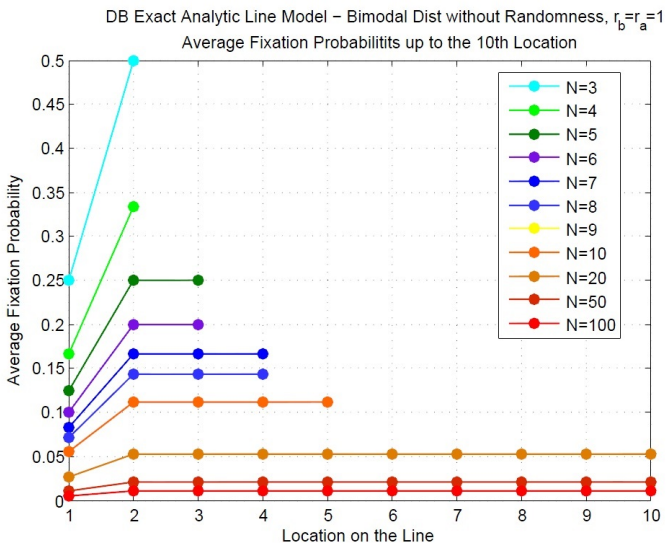


Figure 4.10: **Analytic results for the fixation probability of a newborn mutant located in different positions within the line model.** This figure relates to the neutral case and in the absence of randomness. Except for the fixation probability at boundary which shows a huge decline, rest of the location tend to the same value for the fixation probability in this model. Points are derived from analytic calculations.

4.4.1 Analytic approach without randomnes: Mutant at the boundary

At first, we suppose that the recently imposed mutant is located at one of the boundary points. Let the population size is N and there exist tw type of individulas: mutants, B and wild-type A. $r = \frac{r_b}{r_a}$ is the relative fitness of mutants vs. the normal individuals. Also $d = \frac{d_b}{d_a}$ is the relative death rate of mutants compered with that for normal individuals. The probability of increase and decrease by one in the number of mutants respectively are

$$P_m^+ = \begin{cases} \frac{r}{(r+1)(m(d-1)+N)} & 1 \leq m \leq N-2, \\ \frac{1}{d(N-1)+1} & m = N-1, \end{cases} \tag{4.5}$$

and also

$$P_m^- = \begin{cases} \frac{d}{d+N-1} & m = 1, \\ \frac{d}{(r+1)(m(d-1)+N)} & 2 \leq m \leq N-1. \end{cases} \tag{4.6}$$

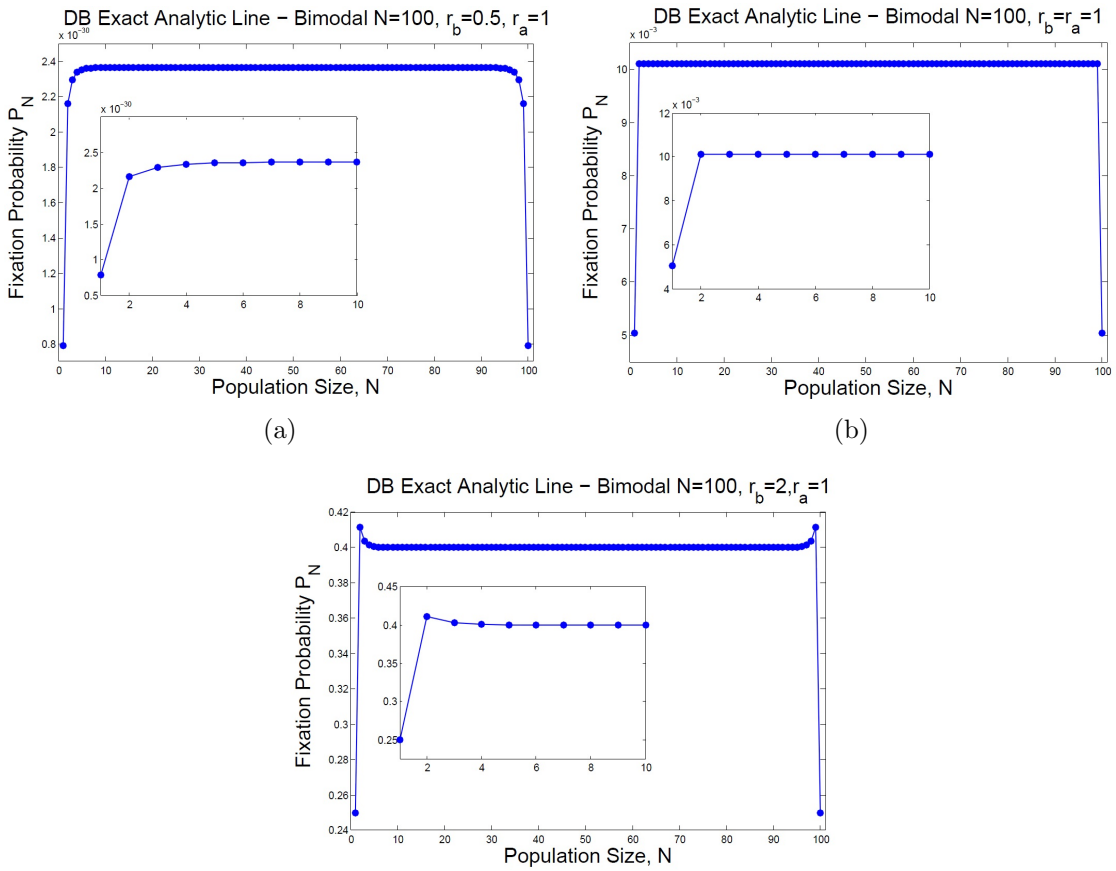


Figure 4.11: **Comparison among various analytic results for the fixation probability of disadvantageous, neutral, and advantageous mutants.** The new born mutant can arise in different locations within the population over a 1D assay of individuals with $N = 100$. In figure (a) $r_b = 0.5$ and $r_a = 1$ in which a smooth increasing trend can be seen for the fixation probability of an initial mutant at different nodes. In (b) $r_b = 1$ and $r_a = 1$ demonstrates the case of a neutral system. In (c) advantageous mutants with $r_b = 2$ and $r_a = 1$ depict a bump at adjacent-boundary point for the fixation probability while as the location changes towards the center point(s), the same fixation probabilities are achieved. Results are based on exact analytic approach.

Thus the fixation probability for $r \neq d$ is

$$\rho = \frac{r - d}{r(d+1) - d(d+1)(\frac{d}{r})^{N-2}}, \tag{4.7}$$

and when $d = 1$, in general we conclude

$$\rho = \begin{cases} \frac{r-1}{2r-2(1/r)^{N-2}} & r \neq 1, \\ \frac{1}{2(N-1)} & r = 1. \end{cases} \tag{4.8}$$

The first expression leads to the following simple form for large populations ($N \gg 1$):

$$\rho = \frac{r-1}{2r}. \tag{4.9}$$

4.4.2 Mutant at internal nodes

Having the new mutant at the internal locations, calculation can be done by finding the transition probability matrix. Finding the exact closed form for the probability of fixation is very complicated due to the complexity of higher dimensional Markov changes. Despite this, we study Kolmogrov equation of the system (the list of Kolmogorov equations are given in 4.11 setting the standard deviation equal to zero) similar to the result of [75].

The fixation probability of a newly–raised mutant at different locations on a line model except for the boundary points are the same when the relative fitness $r = 1$. In the neutral case ($r = 1$), as the population size increases the gap between the fixation probability at boundary and other points decreases. An interesting approach then would be about advantageous and disadvantageous mutants and the relationship between survival chance of different phenotypes at diverse locations on the line model.

The analytic calculation highlights the fact that in the DB Moran case converse to the BD case, the boundary points have the minimum fixation probability within the whole invasion probabilities compered with those for the internal nodes. According to Figure 4.11, the analytic results represent a different behavior for the advantageous and disadvantageous compared to the neutral case.

When $r > 1$, then mutants are beneficial in the system, although the fixation probability of the boundary points is minimum with a huge reduction. The fixation probability at adjacent–boundary points reaches a peak and then its value slightly decreases to be saturated at central points (see Figure 4.11-(c)).

Conversely, when $r < 1$ then there will be a monotonic increase in the probability of fixation moving towards the center/central points. In other words, The maximum survival probability occurs at center/central points. So one may conclude that for advantageous mutants, the adjacent–boundary location represent the best initial location for cancer cells to survive; however, the center (central points) are the best starting location for cancer initiation (refer to Figure (LineAnal)-(a)).

Remark 1. The reason for minimality of boundary points is due to their higher chance of decrease in the population size which increases the survival probability of their neighbor points. Such an enhancement is in favor of advantageous/disadvantageous mutants located besides the boundary. The closer location for advantageous/disadvantageous mutants to the boundary, the higher/lower survivability of mutants will be maintained compare to that of the neutral case.

Remark 2. Increasing the net reproduction rate will boost the maximum at adjacent–boundary point and make this peak sharper for $r > 1$. Similarly, the difference between the central and adjacent–boundary rises by decreasing the relative fitness of mutants.

Remark 3. Increasing the population size, the effect of the structure on the chance of mutants to get absorbed will be less and the chance of having higher survival probability will be distributed among central nodes.

A similar method can be performed for a BD Moran model to observe the symmetric correlation between results of the DB and BD 1D spatial structures.

4.4.3 Analytic approach in the presence of randomness

In the present section, we investigate the effect of randomness on the line model while reflecting boundaries are also taken into account. Detecting the location of the new raised mutant, biologically seems to be very difficult, if not impossible. To study such an event, we may want to put the mutant at a random position chosen through a specific distribution.

In the previous section, the exact analytic approach has provided a general picture for the landscape of the fixation probability on the line model. Now let us assume that the individuals are located on a line model of size N. Also, suppose that $\tilde{a}_1, \cdots, \tilde{a}_N$ are corresponding net growth rates of mutants while $a_1, \cdots, a_N$ are those for normal cells, each of which dedicated to the location 1 to N. Then one can define the probabilities P_σ where the Markov chain is considered at state $\sigma = (\alpha_1, \cdots, \alpha_N)$ within a certain time step. Suppose that $\mathcal{I} = \{i : \alpha_i = 1, 1 \leq i \leq N\}$, $\sigma_{max} = \max_{\mathcal{I}} i$, and $\sigma_{min} = \min_{\mathcal{I}} i$. Moreover, assume that $\mathcal{M}_1 = \max(\{\|\mathcal{I}\| - 2, 0\})$ and $\mathcal{M}_2 = \max(N - \{\|\mathcal{I}\| - 2, 0\})$. Let us also suppose that σ_1 (σ_2) is the element-wise maximum of σ with its permutation by one to the right (left) and $\tilde{\sigma}_1$ ($\tilde{\sigma}_2$) be the element–wise minimum of σ with its permutation by one to the left (right).

The associated Kolmogorov equations define a system of $N(N+1)/2-1$ equations of $N(N+1)/2-1$ variables as follows ($0<\|\sigma\|<N$). These equations are similar to those reported in [75] but when there exist randomness in the system. This generalized system assists us to investigate the behavior of a linear structure under microenvironmental as well as boundary effects which occur at epithelial layer of numerous tissues within the human body. For the transition probabilities by one from the state σ:

$$\begin{aligned}
&\mathcal{A}_1(\sigma)=\frac{\tilde{a}_{\sigma_{max}}}{\tilde{a}_{\sigma_{max}}+a_{\sigma_{max}+2}},\quad \mathcal{A}_2(\sigma)=\frac{\tilde{a}_{\sigma_{min}}}{\tilde{a}_{\sigma_{min}}+a_{\sigma_{min}-2}},\quad \text{for}\ \ 1\leq\|\sigma\|<N-1,\\
&\mathcal{B}_1(\sigma)=\frac{a_{\sigma_{max}+1}}{a_{\sigma_{max}+1}+\tilde{a}_{\sigma_{max}-1}},\quad \mathcal{B}_2(\sigma)=\frac{a_{\sigma_{min}-1}}{a_{\sigma_{min}-1}+\tilde{a}_{\sigma_{min}+1}},\quad \text{for}\ \ 1<\|\sigma\|\leq N-1.
\end{aligned}\tag{4.10}$$

One concludes

$$\begin{cases}
N\,P_\sigma=P_{\sigma_1}[\mathcal{A}_1]+P_{\sigma_2}[\mathcal{A}_2]+P_{\tilde{\sigma}_1}[\mathcal{B}_1]+P_{\tilde{\sigma}_2}[\mathcal{B}_2]\\
\qquad +P_\sigma\left[-N+\mathcal{M}_1+\mathcal{M}_2+\mathcal{C}_1+\mathcal{C}_2\right], & 2<\sigma_{min}<\sigma_{max}<N-1\\
N\,P_\sigma=P_{\sigma_1}[\mathcal{A}_1]+P_{\sigma_2}[\mathcal{A}_2]+P_\sigma\left[-N+\mathcal{M}_2+\mathcal{C}_2\right], & 2<\sigma_{min}=\sigma_{max}<N-1\\
N\,P_\sigma=P_{\sigma_1}[\mathcal{A}_1]+P_{\sigma_2}[1]+P_{\tilde{\sigma}_1}[\mathcal{B}_1]+P_{\tilde{\sigma}_2}[\mathcal{B}_2]\\
\qquad +P_\sigma\left[-N+\mathcal{M}_1+\mathcal{M}_2+\mathcal{C}_1+\mathcal{C}_4\right], & 2=\sigma_{min}<\sigma_{max}<N-1\\
N\,P_\sigma=P_{\sigma_1}[\mathcal{A}_1]+P_{\sigma_2}[1]+P_\sigma\left[-N+\mathcal{M}_2+\mathcal{C}_4\right], & 2=\sigma_{min}=\sigma_{max}<N-1\\
N\,P_\sigma=P_{\sigma_1}[\mathcal{A}_1]+P_{\tilde{\sigma}_1}[\mathcal{B}_1]+P_\sigma\left[-N+\mathcal{M}_1+\mathcal{M}_2+\mathcal{C}_3+\mathcal{C}_4\right], & 1=\sigma_{min}<\sigma_{max}<N-1\\
N\,P_\sigma=P_{\sigma_1}[\mathcal{A}_1]+P_\sigma\left[-N+\mathcal{M}_2+\mathcal{C}_4\right], & 1=\sigma_{min}=\sigma_{max}<N-1\\
N\,P_\sigma=P_{\sigma_1}[1]+P_{\sigma_2}[\mathcal{A}_2]+P_{\tilde{\sigma}_1}[\mathcal{B}_1]+P_{\tilde{\sigma}_2}[\mathcal{B}_2]\\
\qquad +P_\sigma\left[-N+\mathcal{M}_1+\mathcal{M}_2+\mathcal{C}_1+\mathcal{C}_6\right], & 2<\sigma_{min}<\sigma_{max}=N-1\\
N\,P_\sigma=P_{\sigma_1}[1]+P_{\sigma_2}[\mathcal{A}_2]+P_\sigma\left[-N+\mathcal{M}_2+\mathcal{C}_6\right], & 2<\sigma_{min}=\sigma_{max}=N-1\\
N\,P_\sigma=P_{\sigma_2}[\mathcal{A}_2]+P_{\tilde{\sigma}_2}[\mathcal{B}_2]+P_\sigma\left[-N+\mathcal{M}_1+\mathcal{M}_2+\mathcal{C}_5+\mathcal{C}_6\right], & 2<\sigma_{min}<\sigma_{max}=N\\
N\,P_\sigma=P_{\sigma_2}[\mathcal{A}_2]+P_\sigma\left[-N+\mathcal{M}_2+\mathcal{C}_6\right],P_\sigma\left[-N+\mathcal{M}_2+\mathcal{C}_6\right], & 2<\sigma_{min}=\sigma_{max}=N\\
N\,P_\sigma=P_{\sigma_1}[1]+P_{\sigma_2}[1]+P_{\tilde{\sigma}_1}[\mathcal{B}_1]+P_{\tilde{\sigma}_2}[\mathcal{B}_2]+P_\sigma\left[-N+\mathcal{M}_1+\mathcal{C}_1\right], & 2=\sigma_{min}<\sigma_{max}=N-1\\
N\,P_\sigma=P_{\sigma_1}[1]+P_{\sigma_2}[1]+P_\sigma\left[-N\right], & 2=\sigma_{min}<\sigma_{max}=N-1\\
N\,P_\sigma=P_{\tilde{\sigma}_1}[\mathcal{B}_1]+P_\sigma\left[-2+\mathcal{C}_3\right], & 1=\sigma_{min}\ \&\ \sigma_{max}=N-1\\
N\,P_\sigma=P_{\tilde{\sigma}_2}[\mathcal{B}_2]+P_\sigma\left[-2+\mathcal{C}_5\right], & 2=\sigma_{min}\ \&\ \sigma_{max}=N
\end{cases}\tag{4.11}$$

In the latter expressions, when $\|\sigma\|=N-1$ then $\mathcal{A}_1=\mathcal{A}_2=1$ and for $\|\sigma\|=1$ we can conclude that $\mathcal{B}_1=\mathcal{B}_2=1$. Also we have

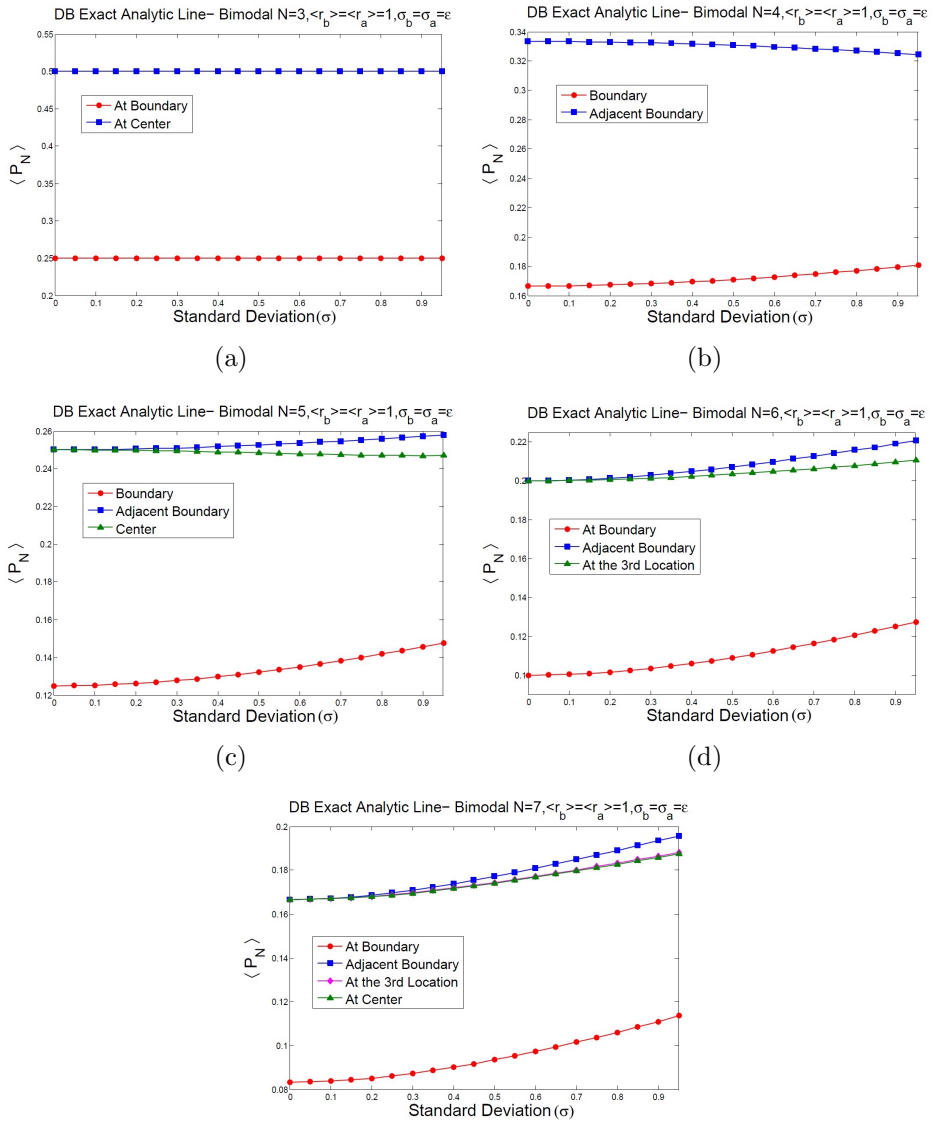


Figure 4.12: **The trend for the fixation probability starting at different locations for various variances and population sizes.** Analytic calculations show the trends for the average fixation probability of mutants for various population sizes: (a) $N = 3$, (b) $N = 4$, (c) $N = 5$, (d) $N = 6$, and (e) $N = 7$. In all cases we assumed that $\langle r_b \rangle = \langle r_a \rangle = 1$ and $\sigma_b = \sigma_a = \sigma$.

$$
\begin{aligned}
\mathcal{C}_1(\sigma) &= \mathrm{H}(\|\sigma\| - 3/2)\left(\frac{\tilde{a}_{\sigma_{max}-1}}{\tilde{a}_{\sigma_{max}-1} + a_{\sigma_{max}+1}} + \frac{\tilde{a}_{\sigma_{min}+1}}{\tilde{a}_{\sigma_{min}+1} + a_{\sigma_{min}-1}}\right), \\
\mathcal{C}_2(\sigma) &= \mathrm{H}(N - \|\sigma\| - 3/2)\left(\frac{a_{\sigma_{min}-2}}{a_{\sigma_{min}-2} + \tilde{a}_{\sigma_{min}}} + \frac{a_{\sigma_{max}+2-1}}{a_{\sigma_{max}+2} + \tilde{a}_{\sigma_{max}}}\right), \\
\mathcal{C}_3(\sigma) &= \mathrm{H}(\|\sigma\| - 3/2)\left(\frac{\tilde{a}_{\sigma_{max}-1}}{\tilde{a}_{\sigma_{max}-1} + a_{\sigma_{max}+1}}\right), \\
\mathcal{C}_4(\sigma) &= \mathrm{H}(N - \|\sigma\| - 3/2)\left(\frac{a_{\sigma_{max}+2}}{a_{\sigma_{max}+2} + \tilde{a}_{\sigma_{max}}}\right), \\
\mathcal{C}_5(\sigma) &= \mathrm{H}(\|\sigma\| - 3/2)\left(\frac{\tilde{a}_{\sigma_{min}+1}}{\tilde{a}_{\sigma_{min}+1} + a_{\sigma_{min}-1}}\right), \\
\mathcal{C}_6(\sigma) &= \mathrm{H}(N - \|\sigma\| - 3/2)\left(\frac{a_{\sigma_{min}-2}}{a_{\sigma_{min}-2} + \tilde{a}_{\sigma_{min}}}\right),
\end{aligned}
\tag{4.12}
$$

where $H(\,.\,)$ is the Heavyside step function. Additionally, the initial conditions of the system are $P_{00\cdots0} = 0, P_{11\cdots1} = 1$.

According to the analytic calculation for $N > 5$ when $\langle r_b \rangle = \langle r_a \rangle = 1$ and $\sigma_b = \sigma_a = \sigma$, the fixation probability of a newborn mutant increases starting from any initial location on the line (Figure 4.13). For smaller populations, there exist some exceptional cases in which the fixation probability decays for higher standard deviations (see Figure 4.13-(a)-(c)).

Remark 1. More pronounced effects can be seen at the boundary points and changing the location of the first mutant towards central location(s) the effect of randomness slightly declines. By increasing population size, a similar behavior to the circle model occurs in the line model as $\rho \times N$ monotonically elevates via its dependency on the standard deviation σ. Another interesting observation relates to the influence of randomness on the system which enhance the location–dependency of the initial mutant in this particular linear structure even for the case in which the average fitness of both normal and cancer individuals are neutral. Having the relative average fitness greater/less than 1 would lead to higher effect of randomness on the system considering the analytic approaches of the previous section for advantageous/disadvantageous mutants respectively.

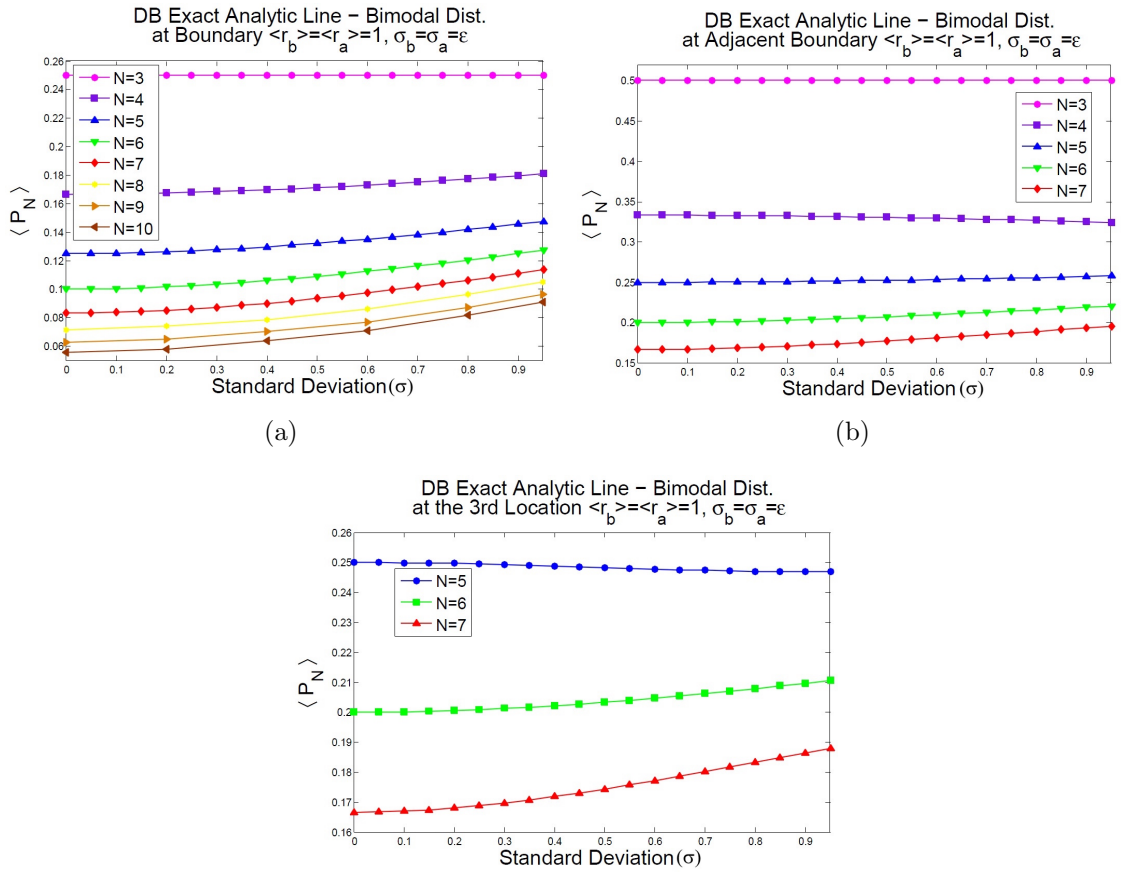
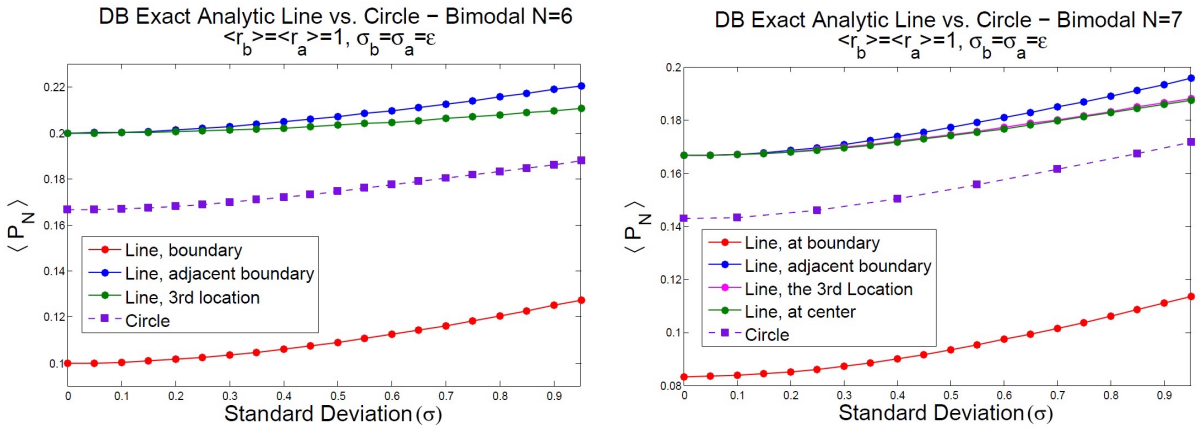


Figure 4.13: **The treatment of the fixation probability for different population sizes are compered separately at boundary, adjacent boundary, and the third location on the line model.** (a) the treatment at boundary, (b) at adjacent boundary, and (c) at the third location on the line, all show a similar increasing behavior as the circle model for larger population size. Points are drawn based on analytic calculations.

Remark 2. A comparison between line and circle model reveals that although their behavior (when $N > 5$) are similar, they do not coincide. More precisely, the fixation probability of circle will remain between the fixation probabilities of starting with initial mutant at boundary in one hand (as the lower bound) and at the internal nodes (each trend as an upper bound) in the line model (Figure 4.14). The relationship between these trends is given in Figure 4.14 for $N = 6, 7$ where the average net reproduction rate is 1.



(a) (b)

Figure 4.14: **Comparing the line and circle models for the same population sizes and when $\langle r_b \rangle = \langle r_a \rangle = 1$ and $\sigma_b = \sigma_a = \sigma$.** In (a) $N = 6$ and in (b) $N = 7$ in which the analytic result for the fixation probabilities in the circle model remains between those of teh line models ifor various variances.

4.5 Higher dimensional spatial structures

In the previous sections, we observed that the influence of microenvironmental effects on the fitness of mutants and/or host cells, changes the survival probability of mutants. Particularly, such an effect on a neutral system where the average fitness of both phenotypes remains approximately the same, may tend to a decrease (when randomness only affects on mutants) or an increase (when randomness affects both resident and malignant individuals). Such evidence which has been verified by our research and through analytic calculation and stochastic simulation [96], has been discussed in the context of one-dimensional spatial structures such as the circle and line models, as well as for complete graph (mass-action) models.

Regarding the tissue structures and epithelial layer of many organs in mammals, it turns out that the alterations to fitness in a heterogenous system is likely to occur in higher dimensions as well. For instance the epithelial layer of the colonic/intestinal crypt is a 2D essay of diverse types of cells. Our analysis shows that even for a two-dimensional lattice (without reflecting boundary) the same behaviour as that of 1D case can be observed. Figure 4.15 represents such a scenario when the average fitness of mutants and normal individuals is 1 and standard deviation of their fitness increases. Then again, minority of mutants makes this type of cell advantageous over the resident cells and in turn, tends

to an increase in the the average survival probability of mutants. In this figure, we have considered $L \times L$ lattices with $L = 3, \cdots, 10$.

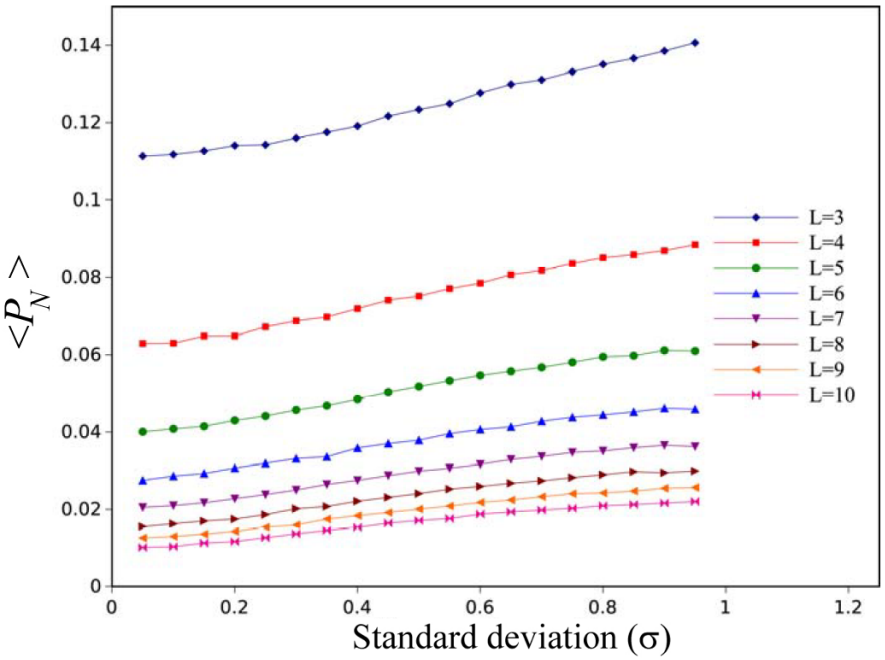


Figure 4.15: **The randomness effect in higher dimensions.** The average fixation probability in terms of the standard deviation of the bimodal fitness distribution for a 2D lattice $L \times L$ without reflecting boundary ($3 \leq L \leq 10$). The fluctuations in the (standard deviation of the) fitness of both normal and mutant cells with average fitness equal to 1, reveal the same trend for the average fixation probability as that of the 1D spatial/non-spatial case. and we run stochastic simulation for a set of 3 runs, each with 20,000 realization. Error bars as the standard deviation of mean are very small and have not shown in this figure.

4.6 Summary

The effects of spatial structure and heterogeneity are important in biological models, including evolutionary biological and biomedical models and social networks. It is known that structure of the network can suppress or amplify the fixation probability [90, 128]. In addition, the heterogeneity of the spatial structure has impacts on the fixation probability,

see for example [2, 168, 105, 99, 53]. Although several papers have focused on the study of heterogeneous networks, less effort has been devoted to understanding the effect of heterogeneity, due to the spatial fitness distribution and environmental stress, and its effect on the invasion probability. In real systems, the fitness of individuals strongly depends on the microenvironmental conditions. For example, in models of bacterial growth or cancer progression, fitness can be a function of the spatial distribution of nutrients and microenvironment. As demonstrated here, random fitness distributions can significantly influence biological and social systems dynamics, leading to an advantage of a minority.

Due to its connections to some of the physical systems, the voter (death-birth) model has been widely studied in the physics literature (see [105] and the references therein). An important biological application of the ring geometry studied here is the model of a human colonic crypt, where stem cells are situated along circular bands (in this context, fixation is referred to as monoclonal conversion [58]). These cells divide leading to proliferation or differentiation (equivalent to removal in our models), and the origins of colon cancer can be studied by examining selection dynamics of mutants in such a system. Of particular importance is the $\mathrm{APC}^{+/-}$ mutant, which in many models is considered neutral or slightly disadvantageous. Such a mutant taking over in one of about 10^7 crypts in a colon is often a first step in the pathway to colon cancer. If the randomness of the environment is taken into account the theoretical likelihood of such an event can be significantly higher than predicted by the standard theories. In fact, given the relatively small population size of the stem cell pool, such mutants will behave as if they are advantageous, leading to a very different dynamics, see e.g. [121, 197].

On a more general note, a large part of population genetics theory relies on the fact that neutral alleles arise with a probability inversely proportional to the population size, N. This fundamental notion plays a role in famous concepts such as Kimura's molecular clock, genetic drift, evolutionary divergence, and coalescence [72, 114], In this paper we show that neutral mutants can behave as if they are selected for, and fixate with a probability independent of N (or decaying much slower than $1/N$), given that they are in a minority, and given that a random environment plays a role. This result can impact prediction levels of DNA divergence between species, levels of standing variation within species, patterns of population subdivisions within species, tests for neutrality vs selection, and inferences about demography [44].

Chapter 5

The Significance of Central Stem Cells in Initiation of Colon and Intestinal Cancer

In the last two chapters, the effect of different spatial structures as well as microenvironmental impact on the development of malignancy within a heterogenous system were investigated. The fixation probability of mutants under certain circumstances was investigated. In the present chapter, we propose a generalized structure to the compartmental model given in Chapter 4, to capture an optimal structure for the cellular turnover within the colonic/intestinal crypt. The majority of the results included in this chapter are reported in [97]. This chapter accounts for another real-world implication of evolutionary dynamics in the study of the complex structure of the crypt as well as tumorigenesis and oncogenesis, and a potential therapeutic treatment to cure such a malignancy.

Rapidly dividing tissues, like colonic/intestinal crypts, are frequently chosen to investigate the process of tumor initiation, because of their high rate of mutations. To study the interplay between normal and mutant as well as immortal cells in the human colon or intestinal crypt, we developed a 4-compartmental stochastic model for cell dynamics based on current discoveries. Recent studies of the intestinal crypt have revealed the existence of two stem cell groups. Therefore, our model incorporates two stem cell groups (central stem cells (CeSCs) and border stem cells (BSCs)), plus one compartment for transit amplifying (TA) cells and one compartment of fully differentiated (FD) cells. However, it can be easily modified to have only one stem cell group.

We find that the main deficiency occurs when CeSCs are mutated, or an immortal cell

arises in the TA or FD compartments. The probability of a single advantageous mutant CeSC being able to transform all cells into mutants is more than 0.2, and one immortal cell always causes all FD cells to become immortals. Moreover, when CeSCs are either mutants or wild-type (w.t.) individuals, their progeny will take over the entire crypt in less than 100 days if there is no immortal cell. Unexpectedly, if the CeSCs are wild-type, then non-immortal mutants with a higher fitness are washed out faster than those with a lower fitness (net reproduction rate). Therefore, we suggest one potential treatment for colon cancer might be replacing or altering the CeSCs with the normal stem cells.

5.1 Introduction

In the present Chapter, we model the cell dynamics of the intestinal crypt based on the available experimental data sets. We investigate the spread of one initial mutant at different locations within the crypt. Furthermore, the proposed model accommodates the possibility that a mutant cell becomes immortal in a single cell division. In this model, we vary the fitness of mutants, which defines the probability that a mutant cell divides and replaces its neighbor cells. We assume that the fitness of all w.t. cells is 1, and the relative fitness of mutants is r. In other words, when there are j mutants and i w.t. cells competing to divide and fill out the available empty space, the probability that a mutant cell divides is $\frac{rj}{rj+i}$.

Recently, the probability P_R that a mutant stem cell replaces its neighbor for various mutants was empirically obtained [181]. Based on this mouse experiment, the fitness of mutant $APC^{+/-}$ is 1.6, while the fitness of mutant $APC^{-/-}$ is 3.8. Moreover, the fitness of the dominant-negative hotspot $P53^{R172H}$ mutant, which corresponds to the human hotspot $P53^{R175H}$, has been obtained. The $P53^{R172H}$ protein has been shown to inhibit the w.t. p53 function, and tumors expressing $p53^{R172H}$ are more metastatic than tumors deleted for p53 [129]. Surprisingly, the fitness of $P53^{R172H}$ mutants is 0.9 in the normal colon, while the fitness of the $P53^{R172H}$ mutant is 1.4 in the inflammatory environment. Thus, we consider a range of fitness values for r such that the mutants can be disadvantageous ($r < 1$), neutral ($r = 1$), or advantageous ($r > 1$) compared to w.t. cells. We consider the expansion of mutant or marked cells presented at the initial time of simulations. For simplicity, we assume all mutants in the tissue have the same fitness r. We also assume that when a mutant cell divides, one of its children is immortal with a very small probability.

Several computational models have been developed to study the dynamics of multistage carcinogenesis [146, 92, 155, 77, 197, 122, 123, 31, 156]. Moreover, there are many computational and mathematical models investigating crypt cell dynamics [63, 13, 39, 111, 138,

11, 64]. However, there are several recent experimental studies on animal colon/intestinal crypts that reveal new information about the cell dynamics. Therefore, new mathematical models are required to accommodate the latest experimental discoveries. Note, because of the inability to use the cell fate mapping experimental techniques in humans, the available data for the human colon/intestinal crypt cell dynamics is very limited. Therefore, in order to obtain more realistic computational models for humans, we need to incorporate the available observed data from animal studies.

Many cell dynamics models have been designed for the intestinal and colon crypts, because of their fairly simple cell dynamics compared to the other tissues, i.e one directional movement from bottom to the top of the crypt [162]. The cell dynamics of the intestinal crypt are very similar to the colon crypt. One of the first models is a homogenous model developed in 1992 to obtain cell division rates at each location of human crypts as well as cell cycles [136]. The time and probability of the progeny of neutral stem cells taking over the stem cell niche and the entire normal intestinal crypts have been modeled in [162]. This stochastic model, which treats all stem cells as one stem cell type, is in perfect agreement with their experimental data, and it predicts stem cells mostly divide symmetrically. Bravo et al. [11] also provided an agent-based computer simulations for cell dynamics in normal human colon crypts.

Although there are many mathematical models for cell dynamics of normal crypts, several models have been designed to investigate the process of mutants' production and their dynamics. Zhao and Michor [204] developed a one-dimensional homogeneous model, which includes only one column of cells, to track mutants in the crypt. They found that most divisions should occur at the bottom of the crypt in order to maximize the time to cancer. In contrast, a recent two-dimensional model, which only contains two columns of cells, shows that most divisions should occur at the top of the crypt to delay cancer [156]. In [204], at each time step, a cell at position i is selected to divide, two daughter cells are then placed into positions i and $i+1$, causing cells that previously resided in positions $i+1$ to shift by one position toward the top of the crypt. The difference, which accounts for the discrepancy between the results of [204] and [156], is that the model developed by Shahriyari et al. [156] considers the probability of two-hit mutant production from wild type cells. Instead, [204] starts with an $\mathrm{APC}^{+/-}$ mutation (a 1-hit mutant) at a given position and calculate the time to produce a $\mathrm{APC}^{-/-}$ mutation (a 2-hit mutant). The predictions of [204] agree with [156] when they consider the probability of second mutation production conditioned on the existence of 1-hit mutants. Additionally, Mirams et al. [111] developed a computer model for cell dynamics of colonic crypts to obtain the probability of mutants taking over or washing out from the crypt. In this model, cells are defined by the location of their centers, and cell movement was determined by assuming that each cell

exerts a linear spring force on its neighbors. They used an equation of motion to determine the center of each cell at each time.

5.2 Multi-stage phenotypic hierarchy of cells in the colon/intestinal crypt

We consider a 4-compartmental stochastic model for the tissue architecture. This model consists of one compartment for TA cells D_t and one compartment for FD cells D_f, and two stem cell (SC) groups S_b and S_c, corresponding to BSCs and CeSCs, respectively. We also denote the population sizes of S_c, S_b, D_t, and D_f with $|S_c|, |S_b|, |D_t|$, and $|D_f|$ respectively.

We assume at each updating time step, two FD cells die and two cells divide based on their fitness. If stem cells divide with probability σ the division is symmetric and with probability $1-\sigma$ is asymmetric. Asymmetric stem cells' divisions can only occur at the S_b compartment. There are two types of symmetric division; differentiation and proliferation. In this model, differentiation only happens in the S_b compartment, because it is close to the TA cells. However, with probability γ, proliferation happens in the S_c compartment and with probability $1-\gamma$ it happens at the S_b compartment. If it occurs at the S_c group then one random stem cell from the S_c compartment migrates to the S_b. We denote the total number of non-stem cells by $D = |D_t| + |D_f|$, the total number of stem cells by $S = S_b + S_c$, and the total number of cells by $N = D + S$. In summary, at each updating time step two FD cells die and two divisions occur based on the following algorithm:

- With probability λ_f, two FD cells divide. Or,
- with probability 1-λ_f, one TA cell differentiates to two FDs, and one of the following scenarios happens.
  - With probability $1-\lambda_s$, one D_t cell proliferates to replace the differentiated D_t cell. Or
  - with probability λ_s, one stem cell divides according to one of the following steps:
    - With probability $1-\sigma$, the division occurs in the S_b asymmetrically, i.e. a TA cell is generated. Or
    - with probability σ the division is symmetric. With probability $\delta = \frac{S^{10}}{{S_0}^{10}+S^{10}}$ this symmetric division is differentiation and happens in the S_b group, and with probability $1-\delta$ is proliferation (S_0 is the initial total number of SCs).

If it is proliferation then with probability γ, it happens at the S_c and with probability $1-\gamma$ it occurs at the S_b. If it occurs at the S_c group, then one random cell from the S_c migrates to the S_b.

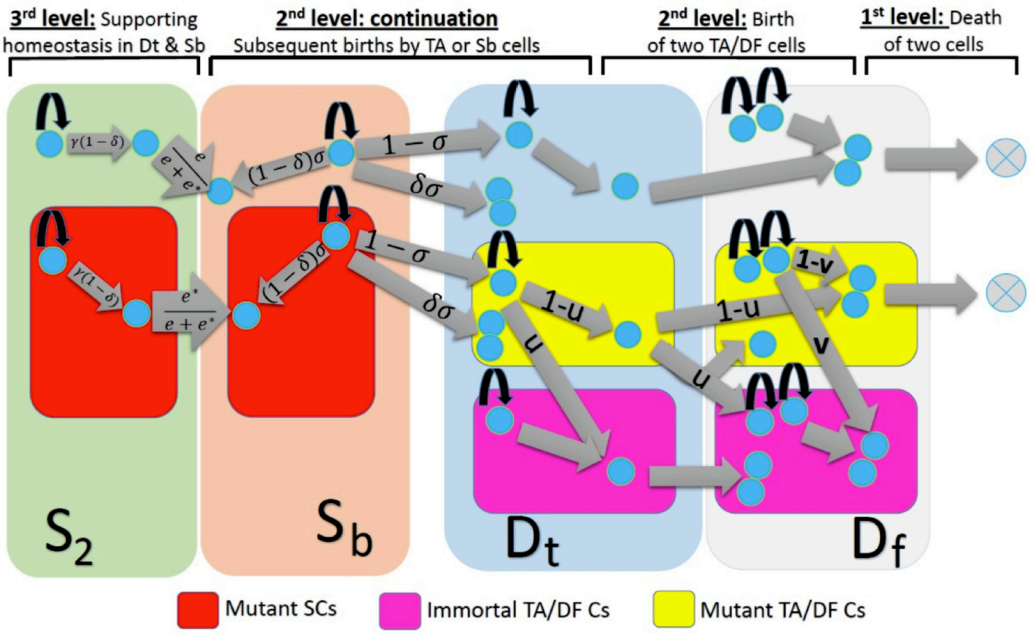


Figure 5.1: **A Schematic representation of the model with possible pathways.** This model includes four compartments: (i) central stem cells, S_c, (ii) border stem cells, S_b, (iii) transient amplifying cells, D_t, and (iv) fully differentiated cells, D_f. Different types of proliferation and differentiation of stem and non-stem cells occur in the system in order to preserve the constant population size. The model includes the possibility of dedifferentiation; mutant D_t or D_f cells are able to generate immortal D_t or D_f cells, respectively.

In this model, the total number of cells stays constant. However, because of the definition of the function δ, the number of stem cells varies, but its variation is very small. We assume, if a mutant divides, then its newborn children are mutants. To accommodate the possibility of dedifferentiation in the division of one mutant non-stem cell, we assume with probability u, that one of its children become immortal. In general, CeSCs can only proliferate, and BSCs are able to proliferate, differentiate, and divide asymmetrically. TA

cells are able to both proliferate and differentiate, and FD cells only able to proliferate. Also for simplicity, death only happens in the FD compartment. A summary of the model is given in Figure re provided in the supplementary ma

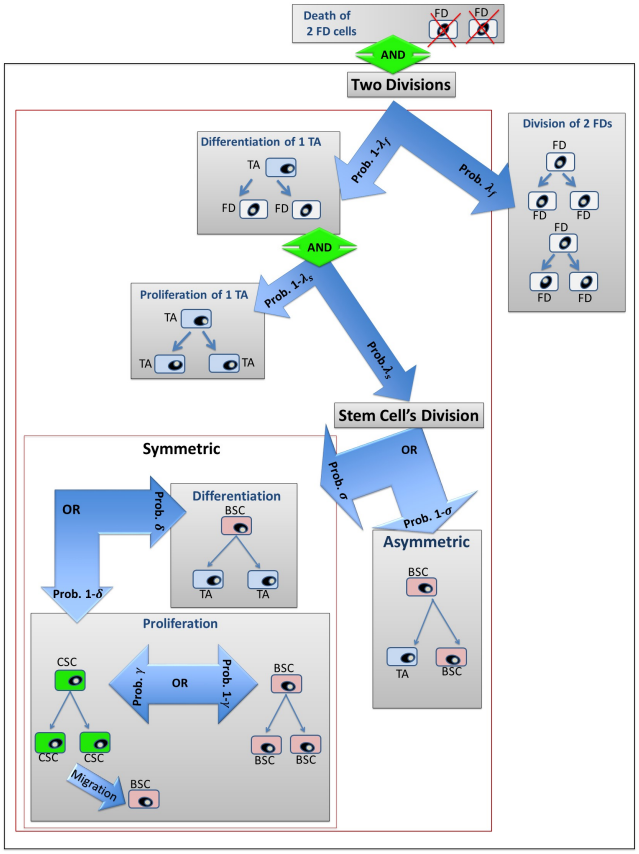


Figure 5.2: **The general algorithm.** The figure reveals the algorithm we used in this chapter for the natural mechanism of the colon.intestinal crypt: at each updating step, two FD cells die and two cells divide to replace the dead cells.

5.2.1 Parameter estimation

The number of studies on human colon crypt's cell dynamics is very limited. However, there are some works on inferring the parameters for human colon crypts [200, 6], as well as measurement of in vivo proliferation using bromodeoxyuridine (BrdU labeling) [136].

These experiments show that there are approximately 2000 cells in each human crypt, and the height of the crypt is around 80 cells [136]. Additionally, most divisions happen in the lower part of the crypt; cells at positions 10-50 (Figure 5.2), where 0 is the bottom of the crypt [135]. In order to calculate the division probabilities at each compartments (FD, TA, S_b, and S_c), we normalize the division rates experimentally obtained in [135].

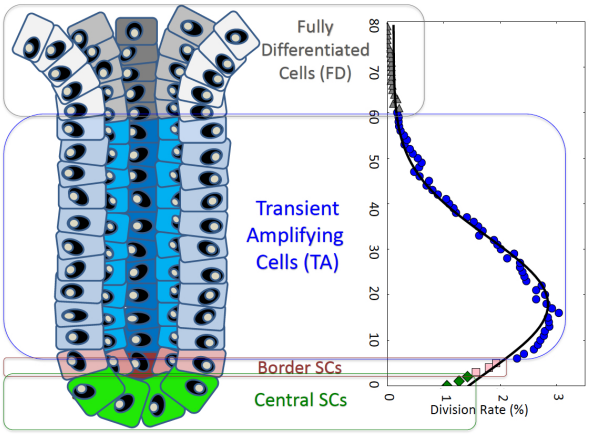


Figure 5.3: **A schematic view of the model.** The figure represents the normalized rate of cell's division at each location of the crypt obtained from [135]. The black solid line is the graph of the function g which shows the normalized division rate of cells, in Parameter estimation section, and the discontinuous curve represents the result of experiment.

We find a function using curve fitting application in MATLAB, which fits the normalized division rates of cells at each location x in the human crypt,

$$g(x) = [0.974 + 24.1 \exp(-\frac{(x-17.9)^2}{444})]/881.4. \tag{5.1}$$

Then we count cells at positions 70-80 as FD cells, therefore the probability that a division occurs in the FD compartment is given by $\lambda_f = \sum_{x=60}^{80} g(x) = 0.026$. Furthermore, each crypt contains around 6 actively dividing stem cells [6], they divide once every 2-3 days [135]. Hence, we assume cells at the locations 0-5 are stem cells, and TA cells are cells at the locations 6-69. Thus, the probability that a division occurs at the stem cell niche when FD cells do not replace two dead FD cells is $\lambda_s = \frac{1}{1-\lambda_f} \sum_{x=0}^{5} g(x) = 0.105$. Moreover, in this model the division probability of CeSCs is given by $\lambda_s \gamma \sigma (1-\delta)$. According to Ritsma et al., cells at the locations 0-2 belong to the CeSC group. Since one CeSC is able to colonize [144], then $\gamma\sigma > 0$. In our model, δ is approximately 0.5, hence $\sigma\gamma = \frac{2}{\lambda_s} \sum_{x=0}^{2} g(x) = 0.884$.

Since $0 < \sigma < 1$ and $0 < \gamma < 1$, we conclude σ and γ are greater than 0.884. The parameters are summarized in Table 5.1.

Table 5.1: Model Parameters estimated from [136, 144].

Symbol	Definition	Value
N	total number of cells	2000
σ	probability of symmetric division	0.884 - 1
γ	division prob. of CeSCs when SCs proliferate	0.884 - 1
r_1	fitness of mutants	0.9 - 3.8
λ_f	division probability of FD cells	0.026
λ_s	division probability of stem cells	0.105
$\lvert S_c\rvert$	number of stem cells in S_c (CeSC)	4 - 6
$\lvert S_b\rvert$	number of stem cells in S_b (BSC)	4 - 7
$\lvert D_t\rvert$	total number of transit amplifying cells	1500
$\lvert D_f\rvert$	total number of fully differentiated cells	500

5.3 The probability of fixation for central stem cells

The general system introduces a multi–variable Markov chain of dependent random processes. Let us consider a particular case of one-dimensional multi–variable Moran process in which there exist only one mutant stem cell in S_c compartment where $(e^*, b^*, d^*, d^{**}, f^*, f^{**}) = (1, 0, 0, 0, 0, 0)$. In other words, in this part we only investigate the cell dynamics in the S_c compartment. Therefore, we obtain the following non–zero transition probabilities for the cell dynamics in the S_C compartment:

$$
\begin{aligned}
P_{e^* \to e^*+1} &= \left(\frac{f}{f+f^*}\right)^2 (1-\lambda_f)\,\frac{d}{\mathcal{D}}\left[\lambda_s\,\sigma\,(1-\delta)\,\gamma\,\frac{r_1 e^*}{\mathcal{R}_2}\,\frac{e}{e+e^*}\right],\\
P_{e^*,d^* \to e^*+1,d^*-1} &= \left(\frac{f^*}{f+f^*}\right)^2 (1-\lambda_f)\,\frac{r_1 d^*}{\mathcal{D}}\,(1-u)\left[\lambda_s\,\sigma\,(1-\delta)\,\gamma\,\frac{r_1 e^* e}{\mathcal{R}_2(e+e^*)}\right],\\
P_{e^*,d^*,f^{**} \to e^*+1,d^*-1,f^{**}+1} &= \frac{2ff^*}{(f+f^*)^2}\,(1-\lambda_f)\,\frac{r_1 d^*}{\mathcal{D}}\,u\left[\lambda_s\,\sigma\,(1-\delta)\,\gamma\,\frac{r_1 e^*}{\mathcal{R}_2}\,\frac{e}{e+e^*}\right],
\end{aligned}
\tag{5.2}
$$

$$P_{e^*,b^*\to e^*-1,b^*+1} = \left(\frac{f}{f+f^*}\right)^2 (1-\lambda_f)\,\frac{d}{\mathcal{D}}\left[\lambda_s\,\sigma\,(1-\delta)\,\gamma\,\frac{e}{\mathcal{R}_2}\,\frac{e^*}{e+e^*}\right],$$
$$P_{e^*,b^*,d^*\to e^*-1,b^*+1,d^*-1} = \left(\frac{f^*}{f+f^*}\right)^2 (1-\lambda_f)\,\frac{r_1 d^*}{\mathcal{D}}\,(1-u)\,\left[\lambda_s\,\sigma\,(1-\delta)\,\gamma\,\frac{ee^*}{\mathcal{R}_2(e+e^*)}\right],$$
$$P_{e^*,b^*,d^*,f^{**}\to e^*-1,b^*+1,d^*-1,f^{**}+1} = \frac{2ff^*}{(f+f^*)^2}\,(1-\lambda_f)\,\frac{r_1 d^*}{\mathcal{D}}\,u\left[\lambda_s\,\sigma\,(1-\delta)\,\gamma\,\frac{e}{\mathcal{R}_2}\,\frac{e^*}{e+e^*}\right].$$

Let us denote the probability of fixation starting from e^* mutants located at the central stem cell compartment by π_{e^*}. Based on the above transition probabilities associated to increase and decrease in the number of S_c mutants at each time step, when stem cells divide only asymmetrically (i.e. when $\sigma = 0$), then π_{e^*} is zero. Note, when stem cells divide asymmetrically, no division occurs in the CeSC compartment, thus the number of CeSC mutants does not change. However, when $\sigma > 0$, the following system of equations can be derived

$$\begin{aligned}
&r_1\,\pi_{e^*+1} + \pi_{e^*-1} - (1+r_1)\,\pi_e^* = 0, \quad 1 < e^* < |S_c| - 1,\\
&r_1\,\pi_2 - (1+r_1)\,\pi_1 = 0,\\
&r_1 + \pi_{|S_c|-2} - (1+r_1)\,\pi_{|S_c|-1} = 0.
\end{aligned} \tag{5.3}$$

The solution to this system signifies that for $1 \le e^* \le |S_c| - 1$:

$$\pi_{e^*} = \frac{1 - \left(\frac{1}{r_1}\right)^{e^*}}{1 - \left(\frac{1}{r_1}\right)^{|S_c|}}. \tag{5.4}$$

Therefore, the fixation probability of one mutant central stem cell in the S_c, i.e. the probability of the progeny of one CeSC mutant taking over the entire S_c compartment, is

$$\pi_1 = \frac{1 - \left(\frac{1}{r_1}\right)}{1 - \left(\frac{1}{r_1}\right)^{|S_c|}}. \tag{5.5}$$

5.4 The fixation probability for mutant S_b stem cells

To obtain the fixation probability in the S_b group, i.e. the probability of progeny of a mutant border stem cell taking over the entire S_b compartment, we only consider the cell

dynamics in the BSC compartment. We assume the system has one mutant stem cell at the initial time in the S_b compartment, while no more mutants exist elsewhere, i.e. $(e^*, b^*, d^*, d^{**}, f^*, f^{**}) = (0, 1, 0, 0, 0, 0, 0))$. Then, we obtain the transition probabilities $p^+(b^*)$ and $p^-(b^*)$, which are the probabilities of transforming from b^* number of mutant BSCs to b^*+1 and b^*-1 number of mutant BSCs, respectively. We denote the probability of b^* number of mutants taking over the S_b by π_{b^*}. Therefore the the fixation probability in the S_b is obtained using the following system of equations.

$$\begin{aligned}
&p^+(b^*)\pi_{b^*+1} + p^-(b^*)\pi_{b^*-1} - (p^+(b^*) + p^-(b^*))\pi_{b^*} = 0, \qquad 1 < b^* < |S_b| - 1,\\
&p^+(1)\,\pi_2 - (p^+(1) + p^-(1))\,\pi_1 = 0,\\
&p^+(|S_b|-1) + p^-(|S_b|-1))\pi_{|S_b|-2} - (p^+(|S_b|-1) + p^-(|S_b|-1))\pi_{|S_b|-1} = 0,
\end{aligned} \tag{5.6}$$

where the total sum of transition probabilities of increase and decrease by one in the number of border stem cells located in S_b are respectively defined by the following formulas:

$$\begin{aligned}
p^+(b^*) &= \lambda_s\,\sigma\,(1-\delta)\,(1-\gamma)\,\left[\frac{d}{\mathcal{D}} + \frac{r_1 d^*}{\mathcal{D}}\,(1-u)\right],\\
p^-(b^*) &= \lambda_s\,\sigma\,\delta\,\left[\frac{d}{\mathcal{D}} + \frac{r_1 d^*}{\mathcal{D}}\,(1-u)\right].
\end{aligned} \tag{5.7}$$

Similar to what we had in the previous section, if the probability of stem cell division λ_s is zero or the probability of symmetric division σ is zero, then the number of S_b mutants does not alter and leads to $\pi_{b^*} = 0$. On the other hand, when $\lambda_s \neq 0$ and $\sigma \neq 0$ then both symmetric and asymmetric division can occur and the system will be reduced to the following recurrence system in the absence of mutation ($u = 0$):

$$\begin{aligned}
&(1-\delta)(1-\gamma)\,\pi_{b^*+1} + \delta\,\pi_{b^*-1} - ((1-\delta)(1-\gamma) + \delta)\pi_{b^*} = 0, \;\; 1 < b^* < |S_c| - 1,\\
&(1-\delta)(1-\gamma), \pi_2 - ((1-\delta)(1-\gamma) + \delta)\pi_1 = 0,\\
&(1-\delta)(1-\gamma) + \delta\,\pi_{|S_c|-2} - ((1-\delta)(1-\gamma) + \delta)\pi_{|S_c|-1} = 0.
\end{aligned} \tag{5.8}$$

This system of equations reveals that

$$\pi_{b^*} = \frac{1 - \left(\frac{\delta}{(1-\delta)(1-\gamma)}\right)^{b^*}}{1 - \left(\frac{\delta}{(1-\delta)(1-\gamma)}\right)^{|S_b|}}. \tag{5.9}$$

Therefore, the fixation probability of a single mutant border stem cell is given by

$$\pi_1 = \frac{1 - \left(\frac{\delta}{(1-\delta)(1-\gamma)}\right)}{1 - \left(\frac{\delta}{(1-\delta)(1-\gamma)}\right)^{|S_b|}}. \tag{5.10}$$

The dependency of the fixation probability on the initial number of S_b mutants has been shown in Figure 5.4.

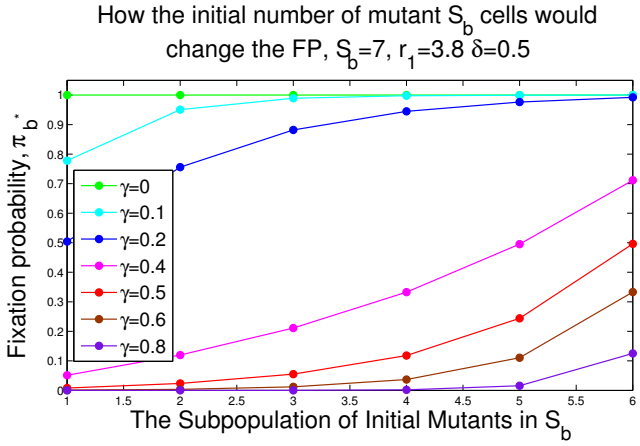


Figure 5.4: **Homeostasis in the number of border stem cells manages the compartmental growth via crucial factors δ and γ.** One mutant border stem cell arises in the S_b compartment and no more mutations are allowed in the system. We assume that $\lambda_s \neq 0$ and $\sigma \neq 0$ which means that both symmetric and asymmetric division can occur and $|S_b| = 7$. This figure shows how the fixation probability π_{b^*} varies w.r.t. the changes in the population size of mutants in the S_b compartment as γ takes various values and $\delta = 0.5, r_1 = 3.8$.

5.5 The probability of fixation for mutant progenitor D_t cells

In this case, we assume there exist only one mutant transit amplifying cell in the D_t compartment, i.e. $(e^*, b^*, d^*, d^{**}, f^*, f^{**}) = (0, 0, 1, 0, 0, 0)$. When at the initial time, there is only one mutant in the compartment of progenitor cells and no naive mutation appears

in the system, e^* and b^* will have no chance to arise, however, the number of d^{**}, f^*, and f^{**} might change. If we assume no new mutations occur ($u = v = 0$), then d^{**} and f^{**} will stay zero over time. Denoting the fixation probability of d^* number of mutant progenitor cells by π_{d^*}, we get

$$\begin{aligned}
&\left(p^+(d^*) + p^-(d^*)\right)\, \pi_{d^*} = p^+(d^*)\, \pi_{d^*+1} + p^-(d^*)\, \pi_{d^*-1} \quad 1 < d^* < |D_t| - 1,\\
&p^+(1)\, \pi_2 = \left(p^+(1) + p^-(1)\right)\, \pi_1,\\
&p^+(|D_t| - 1) + p^-(|D_t - |1)\, \pi_{|D_t|-2} = \left(p^+(|D_t| - 1) + p^-(|D_t| - 1)\right)\, \pi_{|D_t|-1}.
\end{aligned} \tag{5.11}$$

Here, the coefficient of π_{d^*+1} ($p^+(d^*)$) is, in fact, the sum of all transition probabilities which tend to an increase by one in the number of mutant TA cells, and $p^-(d^*)$ as the coefficient of π_{d^*-1} is the sum over all possible transition probabilities leading to a decrease by one in the number of mutant progenitor cells. Under the assumptions of this case and when no more mutation is occurring in the system, there would only exist normal and mutant cells in the D_t compartment and no immortal progenitor cells can arise or be produced by mutant cells in this compartment (i.e. d^{**} remains zero). Hence the coefficients of the system can be reduced to the following form when $\lambda_f \neq 1$, which means that the division is also allowed to occur in D_t compartment.

$$\begin{aligned}
p^+(d^*) &= (1 - \lambda_s)\, \frac{|D_t| - d^*}{|D_t| + (r_1 - 1)d^*},\\
p^-(d^*) &= \lambda_s + (1 - \lambda_s)\, \frac{|D_t| - d^*}{|D_t| + (r_1 - 1)d^*},
\end{aligned} \tag{5.12}$$

and the solution to this new system is in the following form

$$\pi_{d^*} = \frac{\sum_{k=0}^{d^*-1} \mathcal{H}(k)}{\sum_{k=0}^{|D_t|-1} \mathcal{H}(k)}, \tag{5.13}$$

$$\mathcal{H}(k) = \left(\frac{\lambda_s\, r_1 - 1}{\lambda_s - 1}\right)^k \frac{\Gamma\left(\frac{|D_t| + (\lambda_s\, r_1 - 1)(k+2)}{\lambda_s\, r_1 - 1}\right) \Gamma(2 - |D_t|)(1 - \lambda_s)}{\Gamma\left(\frac{|D_t| + 2(\lambda_s\, r_1 - 1)}{\lambda_s\, r_1 - 1}\right) \Gamma(k + 1 - |D_t|)(1 - \lambda_s r_1)}, \tag{5.14}$$

where $\Gamma(t) = \int_0^\infty x^{t-1}\, \mathrm{e}^x\, dx$ is the gamma function. Then the fixation probability π_1 can be derived from the following relation

$$\pi_1 = \mathcal{H}(0) \left[\sum_{k=0}^{|D_t|-1} \mathcal{H}(k)\right]^{-1}. \tag{5.15}$$

Comparing the coefficients $p^+(d^*)$ and $p^-(d^*)$, when $\lambda_f, \lambda_s \neq 0$, the chance of decreasing by one in the number of mutant progenitor cells is higher than that of increasing by one ($p^+(d^*) < p^-(d^*)$ for $1 \leq d^* \leq |D_t| - 1$), which implies that as reproduction rate takes larger values, the probability of fixation declines (See Figure 5.7-(c) in the paper for more details). This means progenitor cells are more capable of producing differentiated cells than being fixated. Given the assumptions of this section, if no division happen in the stem cell niche, i.e. $\lambda_s = 0$, then $\pi_{d^*} = \dfrac{1}{|D_t|}$. Figure 5.7-(b) in the results section, represents the role of initial mutant cell population size in the absorption mechanism for this case.

5.6 The fixation probability of immortal D_t cells in the TA compartment

In this part, we calculate the fixation probability of a single immortal TA cell in the TA compartment. Therefore, we assume at the initial time, the is only one mutant immortal cell, i.e. $(e^*, b^*, d^*, d^{**}, f^*, f^{**}) = (0, 0, 0, 1, 0, 0)$. Based on the model, in this case the number of e^*, b^*, and d^* will stay zero, while the rest of the variables might change. Assuming again $u = v = 0$, the system of equations for the fixation probability $\pi_{d^{**}}$, which is the probability of the progeny of b^{**} number of immortal cells taking over the entire TA compartment, is

$$\begin{aligned}
&p^+(d^{**})\,\pi_{d^{**}+1} + p^-(d^{**})\,\pi_{d^{**}-1} - \big(p^+(d^{**}) + p^-(d^{**})\big)\ \pi_{d^{**}} = 0, \quad 1 < d^{**} < |D_t| - 1,\\
&p^+(1)\,\pi_2 - \big(p^+(1) + p^-(1)\big)\ \pi_1 = 0,\\
&p^+(|D_t| - 1) + p^-(|D_t| - 1)\,\pi_{|D_t|-2} - \big(p^+(|D_t| - 1) + p^-(|D_t| - 1)\big)\ \pi_{|D_t|-1} = 0.
\end{aligned} \tag{5.16}$$

where the coefficients $p^+(d^{**})$ and $p^-(d^{**})$ are respectively the probabilities that the number of immortal TA cells increases by one and decreases by one in one time step. The above system of equations can be simplified to the following system, when $\lambda_f \neq 1$:

$$\begin{aligned}
p^+(d^{**}) &= (1 - \lambda_s)\,\frac{|D_t| - d^{**}}{|D_t| + (r_2 - 1)d^{**}},\\
p^-(d^{**}) &= \lambda_s + (1 - \lambda_s)\,\frac{|D_t| - d^{**}}{|D_t| + (r_2 - 1)d^{**}},
\end{aligned} \tag{5.17}$$

Thus, the probability of fixation of d^{**} immortal TA cells in the TA compartment is given by

$$\pi_{d^{**}} = \frac{\sum_{k=0}^{d^{**}-1} \mathcal{H}(k)}{\sum_{k=0}^{|D_t|-1} \mathcal{H}(k)}, \tag{5.18}$$

$$\mathcal{H}(k) = \left(\frac{\lambda_s\, r_1 - 1}{\lambda_s - 1}\right)^k \frac{\Gamma\left(\frac{|D_t|+(\lambda_s\, r_1-1)(k+2)}{\lambda_s\, r_1-1}\right) \Gamma(2-|D_t|)(1-\lambda_s)}{\Gamma\left(\frac{|D_t|+2(\lambda_s\, r_1-1)}{\lambda_s\, r_1-1}\right) \Gamma(k+1-|D_t|)(1-\lambda_s r_1)}, \tag{5.19}$$

where $\Gamma(t) = \int_0^\infty x^{t-1}\, \mathrm{e}^x\, dx$ is the gamma function. Then the fixation probability π_1 can be obtained as

$$\pi_1 = \mathcal{H}(0) \left[\sum_{k=0}^{|D_t|-1} \mathcal{H}(k)\right]^{-1}. \tag{5.20}$$

This result is similar to the one obtained for the mutant progenitor cells in section C. The behavior of the system is also the same as those given in Figure 5.7 parts (b) and (c). Therefore, the crucial role of immortal progenitor cells can be explained mainly by producing immortal differentiated cells. It might be worthy to remark that a small increase in the number of initial immortal D_t cells would not significantly affect the fixation probability (See Figure 5.7-(c)).

5.7 The fixation probability of mutant FD cells in the FD compartment

Here, we investigate the survival probability of a mutant D_f cell while environment imposes no further mutations, i.e. $u = v = 0$. Assuming the initial state is $(e^*, b^*, d^*, d^{**}, f^*, f^{**}) = (0, 0, 0, 0, 1, 0)$, only the number of f^* can vary. Therefore, the fixation probability of f^* number of mutant D_f cells, π_{f^*}, can be obtained from the following system of equations.

$$
\begin{aligned}
&p^+(f^*)\,\pi_{f^*+2} + q^+(f^*)\,\pi_{f^*+1} + p^-(f^*)\,\pi_{f^*-1} + q^-(f^*)\,\pi_{f^*-2} \\
&\qquad - \big(p^+(f^*) + q^+(f^*) + p^-(f^*) + q^-(f^*)\big)\ \pi_{f^*} = 0, \\
&\qquad\qquad 2 < f^* < |D_f| - 2, \\
&p^+(1)\,\pi_3 + q^+(1)\,\pi_2 - \big(p^+(1) + q^+(1) + p^-(1)\big)\ \pi_1 = 0, \\
&p^+(2)\,\pi_4 + q^+(2)\,\pi_3 + p^-(2)\,\pi_1 - \big(p^+(2) + q^+(2) + p^-(2) + q^-(2)\big)\ \pi_2 = 0, \\
&p^+(|D_f|-2) + q^+(|D_f|-2)\,\pi_{|D_f|-1} + p^-(|D_f|-2)\,\pi_{|D_f|-3} + q^-(|D_f|-2)\,\pi_{f^*-4} \\
&\qquad - \big(q^+(|D_f|-1) + p^-(|D_f|-1) + q^-(|D_f|-1)\big)\ \pi_{|D_f|-2} = 0, \\
&q^+(|D_f|-1) + p^-(|D_f|!-1)\,\pi_{|D_f|-2} + q^-(|D_f-1)\,\pi_{f^*-3} \\
&\qquad - \big(q^+(|D_f|-1) + p^-(|D_f|-1) + q^-(|D_f|-1)\big)\ \pi_{|D_f-1} = 0.
\end{aligned}
\tag{5.21}
$$

Where,

$$
\begin{aligned}
p^+(f^*) &= \lambda_f \left(\frac{r_1\, f^*}{|D_f| + (r_1-1)f^*}\right)^2, \\
q^+(f^*) &= 2\,\lambda_f\, \frac{r_1\, f^*}{|D_f| + (r_1-1)f^*}, \\
p^-(f^*) &= 2\,\lambda_f\, \frac{(|D_f| - f^*)}{|D_f| + (r_1-1)f^*} + 2(1-\lambda_f), \\
q^-(f^*) &= \lambda_f \left(\frac{|D_f| - f^*}{|D_f| + (r_1-1)f^*}\right)^2 + 2(1-\lambda_f),
\end{aligned}
\tag{5.22}
$$

If divisions never occur in the FD compartment, i.e. $\lambda_f = 0$, then the number of FD mutants (d^*) remains constant during the process, implying

$$
\pi_{f^*} = 0, \qquad \text{for} \quad 1 \le f^* \le |D_f| - 1. \tag{5.23}
$$

On the other hand, when $0 < \lambda_f \le 1$, the coefficients $p^{\pm}, q^{\pm}$ lead to a more complicated system. In Figure 5.7-(e),(f) the solutions to this system are given for some particular values of λ_f where the other parameters have chosen from Table 5.1: $|D_f| = 500, r_1 = 3.8$. In this figure, as λ_f tends to zero, a dramatic change will occur in the graph. In part (f) of this figure, the graphs reveal the fact that, even for a large value of the relative fitness of mutants, the survival chance of mutants remains very small.

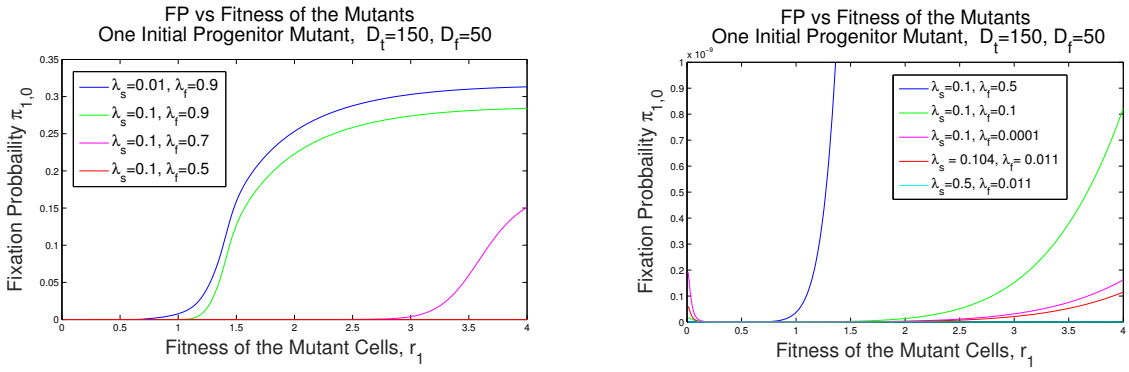


Figure 5.5: **Multi–variable Markov chain of mutants in non-stem cell compartments.** In the absence of mutation and plasticity, when a mutant cell appears in either D_t or D_f compartments, we calculate the probability of fixation for mutant differentiated cells. Assuming $|D_t| = 150$ and $|D_f| = 50$ we investigate three different approaches as λ_s and λ_f alter. Firstly, (a) represents the probability $\pi_{1,0}$ of starting from one initial mutant D_t cell for lower values of $0.01 \leq \lambda_s \leq 0.1$ and higher values of $0.5 \leq \lambda_f \leq 0.9$. We conclude that lower values for λ_s and higher values for λ_f tends to higher fixation probabilities. In contrast, changing the values of λ_s to lower values as well, leads to a huge drop in the survival probability. (c) depicts a landscape for the fixation probability for possible initial states (d^*, f^*) (for $0 \leq d^* \leq 150, 0 \leq f^* \leq 50$). A dramatic increase in the probability of fixation can be obtained by starting from larger initial mutant population of TA cells where $\lambda_s = 0.105, \lambda_f = 0.026$, and $r_1 = 3.8$.

5.8 The fixation probability of mutant D_t cells in the FD compartment

In this part, we investigate the probability of the progeny of one mutant TA cell taking over the entire FD compartment. Again, we assume no more mutation is expected to occur through the whole procedure, i.e. $u = v = 0$. Therefore, we assume that initially there is only one mutant TA cell, i.e. $(e^*, b^*, d^*, d^{**}, f^*, f^{**}) = (0, 0, 1, 0, 0, 0)$. This case is more complicated compared to the previous cases because of dependency of the system on both mutant progenitor and differentiated cells. Let π_{d^*,f^*} be the fixation probability of starting from d^* TA cells and f^* mutant differentiated cells (initial state is (d^*, f^*)). Here, we consider a bi–variable Markov chain to explore the cell dynamics in the FD and TA compartments. There are 10 corresponding transition probabilities of possible changes to the state (d^*, f^*). The initial state is $(1, 0)$ while the initial conditions are $\pi_{0,0} = 0$ and $\pi_{d^*,|D_f|} = 1$ for any d^*.

Since there are $D_f(D_t + 1) - 1$ different states (where $(d^*, f^*) \neq (0, 0)$ and $0 \leq f^* < |D_f|$), the transition matrix is a $|D_f|(|D_t|+1)-1$ by $|D_f|(|D_t+1)-1$ dimensional matrix A, where each entry of A corresponds to one of the states (d^*, f^*) and includes the coefficients of representing π_{d^*,f^*} in terms of all possible fixation probabilities $\pi_{\tilde{d^*},\tilde{f^*}}$:

$$\pi_{d^*,f^*} = \sum_{\tilde{d^*},\tilde{f^*}} P_{(d^*,f^*)\to(\tilde{d^*},\tilde{f^*})}\, \pi_{(\tilde{d^*},\tilde{f^*})}. \tag{5.24}$$

Let $B = \Big[(d^*, f^*)\Big]_{0\leq d^*\leq |D_t|, 0\leq f^*\leq |D_f|-1}$ be the matrix of all possible states (d^*, f^*) (for $(d^*, f^*) \neq (0, 0)$ and $0 \leq f^* < |D_f|$). The matrix B is isomorphic to a vector in $\mathbb{R}^{D_f(|D_t|+1)-1}$ by considering the subsequent rows in an ordered array as coordinates of this vector when the first array $(0, 0)$ is ignored. More precisely, we have the following isomorphism

$$(d^*, f^*) \longrightarrow |D_f| - 1 + \text{Heavyside}\left(d^* - \frac{3}{2}\right)(d^* - 2)|D_f| + f^*,$$

for $0 \leq d^* \leq |D_t|, 0 \leq f^* \leq |D_f| - 1$. Thus, using the vector representation of matrix B (after dropping the entry $(0, 0)$), we label rows and columns of the matrix A with entries of the isomorphic vector to matrix B. Figure 5.5-(c) reveals how the fixation probabilities depend on all possible states having the landscape of changes for possible states (d^*, f^*).

If all divisions occur in the FD compartment, i.e. $\lambda_f = 1$, then there is no chance that mutant TA cells divide, thus $\pi_{d^*,0} = 0$ for all $1 \leq d^* \leq |D_t|$. Moreover, when

λ_s decreases and λ_f increases, the fixation probability dramatically climbs as these two conditions reinforce the chance for proliferation in progenitor and differentiated cells rather than divisions within other types of cells. For lower rates of λ_f, a much higher mutants' fitness is required to slightly increase the fixation probability. However, when mutants are disadvantageous, the optimal fixation probability can occur for restrained birth rates (see Figure 5.5).

5.9 The fixation probability of immortal differentiated cells in the D_f compartment

Here, we calculate the probability of the progeny of $f^{*}*$ number of mutant FD cells taking over the entire FD compartment, $\pi_{f^{**}}$. For this reason, we assume the initial state of the system is $(e^*, b^*, d^*, d^{**}, f^*, f^{**}) = (0, 0, 0, 0, 0, 1)$, we also assume no new mutations or immortal cells arise ($u = v = 0$). The fixation probability $\pi_{f^{**}}$ satisfies the following system of equations, when $\lambda_f \neq 0$.

$$\begin{aligned} &2(|D_f| - f^{**})\,\pi_{f^{**}+1} + (r_2\, f^{**})\,\pi_{f^{**}+2} - (2|D_f| + (r_2 - 2)f^{**})\;\pi_{f^{**}} = 0, \quad 1 \leq f^{**} < |D_f| - 1, \\ &4\,\pi_{|D_f|-1} + r_2\,(|D_f| - 2) - (4 + r_2\,(|D_f| - 2))\;\pi_{|D_f|-2} = 0, \\ &2 - 2\,\pi_{|D_f|-1} = 0. \end{aligned} \tag{5.25}$$

The above recurrence system implies that

$$\pi_{f^{**}} = 1, \quad 1 \leq f^{**} \leq |D_f| - 1. \tag{5.26}$$

When $\lambda_f \neq 0$, a similar approach will be achieved as the case for $\lambda_f = 1$ since there exist no supporting divisions from immortal D_t cells to increase immortal D_f population. Therefore, immortal differentiated cells, in the absence of apoptosis, can exponentially grow. Appearance of a minor population of this type of cells will take over the whole population of D_f compartment. Another possible scheme is when $\lambda_f = 0$ which results in no chance for immortal cells to fixate even starting from $|D_f| - 1$ number of initial cells.

5.10 The fixation probability of immortal D_t cells in the D_f compartment

Now, we obtain the probability of the progeny of d^{**} number of immortal D_t cells taking over the FD, $\pi_{d^{**}}$, while no more mutations is expected in the whole system. Assuming

the initial state $(e^*, b^*, d^*, d^{**}, f^*, f^{**}) = (0, 0, 0, 0, 1, 0)$, we investigate the probability of fixation $\pi_{d^{**},f^{**}}$ of d^{**} immortal progenitor cell(s) and f^{**} immortal differentiated cell(s) in the FD compartment. When $\lambda_f = 1$, then there is no chance for any migration of immortal cells from D_t compartment into D_f compartment and thus

$$\pi_{d^{**},0} = 0, \qquad 0 \leq d^{**} \leq |D_t|. \tag{5.27}$$

Now if we consider the other extreme in which $\lambda_f = 0$, the only resource for population growth of immortal D_f cells is the immortal D_t compartment. In this case, when $\lambda_s = 1$, there is a chance for the only immortal cell in D_t to divide symmetrically to two immortal differentiated daughter cells but these cells will not growth or decay in D_f group and the the divided immortal cell in D_t will not be substituted. Thus $\pi_{d^{**},0} = 0$ for small values d^{**} compared with the population size of D_f. In this situation, if $\lambda_s < 1$ then proliferation can occur in D_t and each divided immortal cell in D_t will be replaced with a certain chance when $d^{**} \geq 2$ at the beginning. But we conclude again $\pi_{1,0} = 0$.

Now, let us assume that $0 < \lambda_f < 1$. As we explained in the above, the onset of any immortal cell in D_f compartment given the assumptions of this case will tend to fixation (although the time to fixation may vary). So the fixation probability equals to the probability of having first immortal cell in D_f. Therefore, if $\lambda_s = 1$, considering the probability of division for the initial immortal cell in D_t, we obtain

$$\pi_{1,0} = \frac{r_2}{|D_t| + r_2 - 1}. \tag{5.28}$$

This probability will increase linearly as the initial number of immortal D_t cells increases. Finally if $\lambda_s < 1$, we conclude that

$$\pi_{d^{**},f^{**}} = 1, \qquad 0 \leq d^{**} \leq |D_t|,\ 0 \leq f^{**} \leq |D_f| - 1. \tag{5.29}$$

In summary, these calculations reveal how different mechanisms would influence the system, the appearance of an immortal cell either in D_t or D_f compartments can trigger a cancer, and how fast this may develop to take over the whole normal population via maintaining the structured stability in the crypt.

5.11 Significant results for the general regulatory mechanism within the crypt

5.11.1 The probability of the progeny of mutant/marked CeSCs taking over the S_c and the entire crypt is high

The analytical methods reveal that the probability of the progeny of e^* number of mutated/marked CeSC taking over the S_c compartment is given by $\pi_{e^*} = \frac{1-(1/r_1)^{e^*}}{1-(1/r_1)^{|S_c|}}$, where r_1 is the fitness of mutants and $|S_c|$ is the total number of CeSCs. Although cells with high fitness have high proliferation rate, their differentiation rate is also high. As a result, the fixation probability of advantageous CeSC mutants is higher than disadvantageous mutants, however their fixation time is also higher than disadvantageous ones. Additionally, if the progeny of mutant CeSCs take over the FD compartment, its average occurrence time is less than 100 days (Figure 5.6).

If stem cells divide only asymmetrically (i.e. $\sigma = 0$), then no division happens at the CeSC compartment, thus CeSC mutants never divide. In other words, if stem cells divide fully asymmetrically then the probability that a mutant CeSC spreads to any other compartment is zero. The same scenario would occur, when the proliferation probability in the S_c is zero, i.e. $\gamma = 0$.

5.11.2 The probability of a mutant BSC's progeny taking over the S_b or D_f is approximately zero

The analytical calculations show that the probability of the progeny of b^* number of BSCs taking over the entire BSC is given by $\pi_{b^*} = \frac{1-(\delta/[(1-\delta)(1-\gamma)])^{b^*}}{1-(\delta/[(1-\delta)(1-\gamma)])^{S_b}}$, where δ and γ are respectively the differentiation probability of BSCs and the proliferation probability of CeSCs when stem cells divide symmetrically. This formula and the simulations imply that the probability that the progeny of one BSC will spread over the entire BSC is almost zero, because according to the parameter estimations δ is approximately half, and γ is between 0.884 and one. Additionally, the progeny of BSCs are always washed out from the crypt regardless of their fitness when there is no immortal cell. If at least half of BSCs are mutants, then they might have a small chance to colonize and take over the entire BSC compartment. If this rare event happens, it occurs in one day (Figure 5.6-

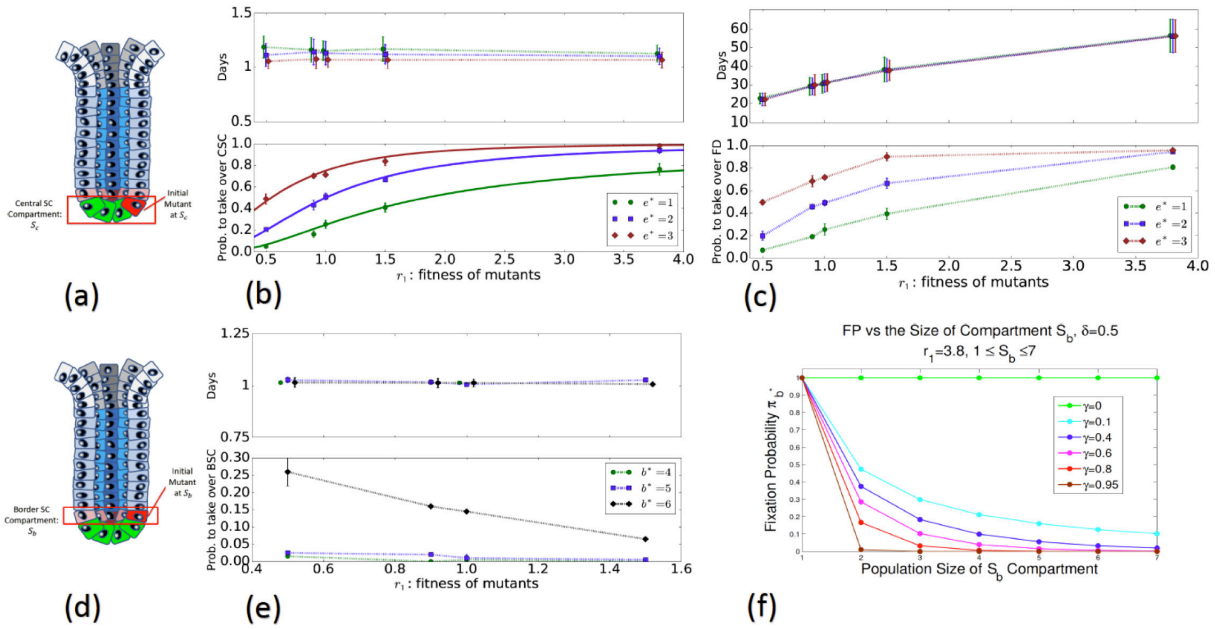


Figure 5.6: **(a)-(c) The probability and time that mutant CeSCs will take over the S_c and the FD.** The sub-figure (a) presents a schematic view of the model at the initial time. The simulations start with e^* mutants in the S_c, while the other cells are wild-type. The sub-figure (b) indicates the average time and the probability of the progeny of mutant CeSCs taking over the CeSCs. The plot (c) shows the probability and time that the progeny of CeSC mutants will take over the FD. In this figure $|S_b| = 7, |S_c| = 4$, and $u = v = 0$, other parameters are given in Table 5.1. The points are the average and the bars indicate the standard deviations of 5 batches of 100 runs, and the solid lines present the results of the formula. **(d)-(f) Time and probability of the progeny of mutant BSCs taking over the S_b.** The figure (d) shows that there are b^* number of mutants in the S_b at the initial time of simulations, and (e) presents the result of simulations. The bottom sub-figure of (e) indicates the probability that the progeny of b^* number of mutant BSCs will take over the entire S_b, and the top sub-figure shows the time of its occurrence. Plot (f) presents the analytic results, and it shows the effect of the number of BSCs, S_b, and the proliferation probability of CeSCs, γ, on the fixation probability, which is the probability of the progeny of mutant BSCs taking over the entire S_b. Here, $|S_b| = 7$, $|S_c| = 4$, and the rest of parameters are given in Table 5.1.

(d),(e),(f)). However, several days later the progeny of CeSCs will replace all mutant BSCs, and eventually the progeny of mutant BSCs are washed out from the crypt.

Although we expect the higher proliferation probability in the BSC compartment (i.e. smaller γ) to lead to the higher fixation probability in the BSC group, it does not increase the fixation probability π_{b^*} very much. The reason is each proliferation is coupled with one

differentiation. The differentiation of SCs increases the chance of transporting mutants from the stem cell niche to the TA group. Additionally, advantageous mutants disappear faster from the S_b than disadvantageous ones. This result is in agreement with the result of the experiments done by Ritsma et al. [144], where they observed that the probability of a BSC colonization is very small.

5.11.3 The progeny of a small number of FD or TA mutants never take over the entire TA or FD

Experimental data shows that FD cells do not divide much compared to TA cells. If FD cells do not divide at all ($\lambda_f = 0$), then non-immortal FD mutants are always removed from the crypt. Moreover, if $\lambda_f > 0$ and no more mutation or plasticity is allowed in the system, the probability $\pi_{f^*=1}$ that the progeny of one FD mutant will take over the whole FD is very small (Figure 5.7-(b),(c)); however, it is much higher than the probability of one TA cell's progeny taking over the TA compartment, $\pi_{d^*=1}$. When the rate of divisions in D_t is approximately zero (it rarely occurs in reality), then $\pi_{d^*=1} = 0$. If a small but non-zero number of divisions happen in stem cells and a high number of divisions occur in D_t, which corresponds to the experimental observations, then the probability of one TA mutant's progeny taking over the TA compartment decreases when the mutants' fitness r increases (Figure 5.7-(e),(f)). Meaning that TA cells are more capable of producing FD cells than being fixated. In the extreme scenario, when no divisions happen in SCs ($\lambda_s = 0$), then $\pi_{d^*=1}$ is $\frac{1}{|D_t|}$. Moreover, when no proliferation in D_t is permitted and TA cells only differentiate, i.e. $\lambda_s = 1$, then π_{d^*} is zero. Note, when TA cells are only able to differentiate but not proliferate, mutants differentiate to two FD mutants (and thus extinct), therefore mutants will be removed from the TA compartment (Figure 5.7-(e),(f)).

We also study the probability of a TA mutant's progeny taking over the FD compartment. In the human colon crypt, where $\lambda_f = 0.011$ and $\lambda_s = 0.104$, the probability of one TA or FD cell's progeny taking over the entire FD or TA compartment is approximately zero. This emphasizes the fact that although TA mutants generate FD mutants, the generated FD mutants are washed out from the crypt, before they get a chance to colonize. Interestingly, mutants with a high fitness will be removed from the crypt quickly, because most divisions occur in the TA compartment, and advantageous TA mutants quickly differentiate to two FD cells.

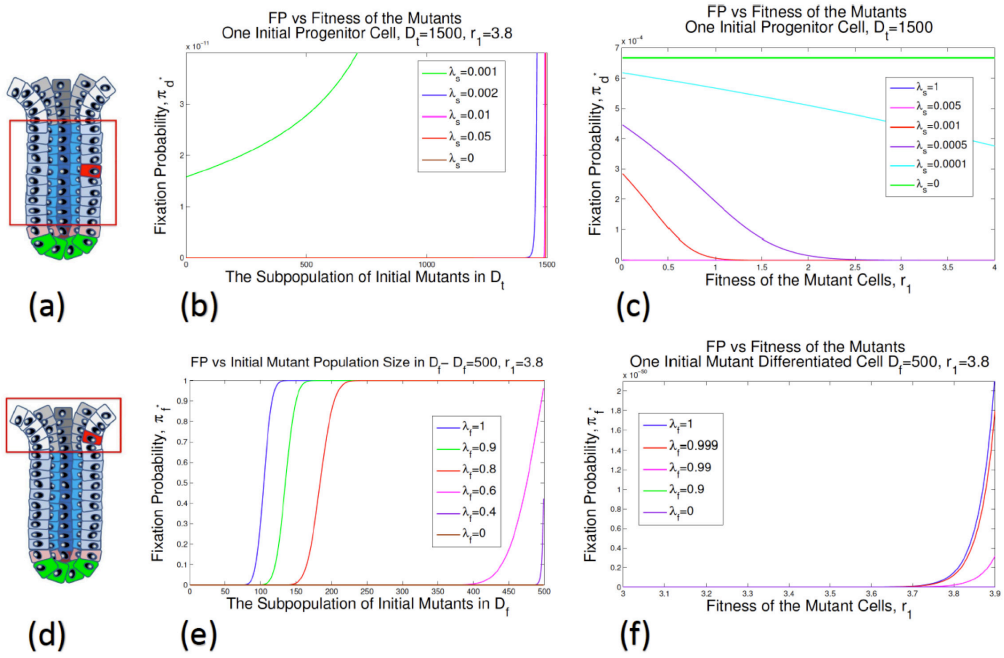


Figure 5.7: **(a)-(c) Role of TA mutants in generating FD mutants.** Figure (a) is a schematic view of the crypt at the initial time of the process. Plot (b) indicates the effect of the initial number of TA mutants and the probability of stem cells' division, λ_s, on the probability that mutants take over the entire TA. Plot (c) reveals the effect of the mutants' fitness on the fixation probability of mutants in the TA. **(d)-(f) Fixation probability in the FD compartment.** Sub-figure (d) is a schematic representation of the crypt at the initial time. (e) shows the behavior of the system for a range of the rate of divisions in the D_f group, λ_f, as the initial number of mutants varies. The curves in (f) illustrate the impact of the mutants' fitness r_1 and the division probability of FD cells, λ_f, on the fixation probability of mutants in the FD group. This figure shows the results of the analytical formulas, when the total number of TA cells is 1500, and the number of FD cells is 500, and in (b) and (c), the mutant's fitness is $r_1 = 3.8$.

5.11.4 Central stem cells control the entire crypt

The results of simulations reveal that with probability one, the progeny of CeSCs will take over the entire human colon crypt in less than three months. In other words, if all stem cells are wild-type, while the rest of the cells in the crypt are mutants, then all crypt cells become wild-type in less than 100 days. Surprisingly, the time that the progeny of the

CeSCs need to take over the crypt decreases when the mutants' fitness increases: meaning the advantageous mutants are washed out from the crypt faster than disadvantageous ones. Moreover, the probability that the progeny of a single normal stem cell will take over the entire crypt is more than zero. This probability is more than 0.25, if mutants are disadvantageous, and it is close to zero if mutants are advantageous. This implies that if only one of the CeSCs is wild-type, and the rest of the crypt's cells are $P53^{R172H}$ mutants, then with a probability of 0.25 all cells will become wild-type in 1-2 months in the non-inflammatory condition. Although the probability of w.t. cells taking over the entire crypt depends on the number of normal cells in the S_c, its concurrence time does not depend on the number of wild-type CeSCs (Figure 5.8).

If stem cells divide fully asymmetrically ($\sigma = 0$), then no division occurs in the CeSC compartment. Therefore, in this case CeSCs will not take over the crypt. Moreover, the time that the progeny of CeSCs need to take over the entire crypt is a decreasing function of σ. In other words, if stem cell divisions are mostly symmetric, then CeSCs progeny will rapidly spread over the crypt. In addition, the probability γ, which is the probability that a CeSC divides in the case of symmetric division, is not as important as the fitness of mutants r_1.

5.11.5 The progeny of a single immortal TA or FD cell always take over the entire FD in less than 70 days

Some environmental conditions or genetic/epigenetic changes lead to creation of an immortal cell in the TA compartment. Therefore, here we investigate the dynamics of potential immortal cells in the crypt. We observe that the immortal TA cells have higher desire to differentiate and generate more immortal FD cells than spreading over the TA compartment. Moreover, the progeny of even a single immortal TA or FD cell will spread over the entire FD in less than 70 days. Expectedly, the advantageous immortals spread faster than disadvantageous ones.

5.11.6 Existence of the bi-compartmental stem cell niche has some advantages and disadvantages.

Our 4-compartmental model can be easily modified as a 3-compartmental model with only one stem cell group by assuming there is no CeSCs ($|S_c| = 0$) and the probability of division in CeSCs γ is zero. Figure 5.10 shows that the probability that the progeny of one

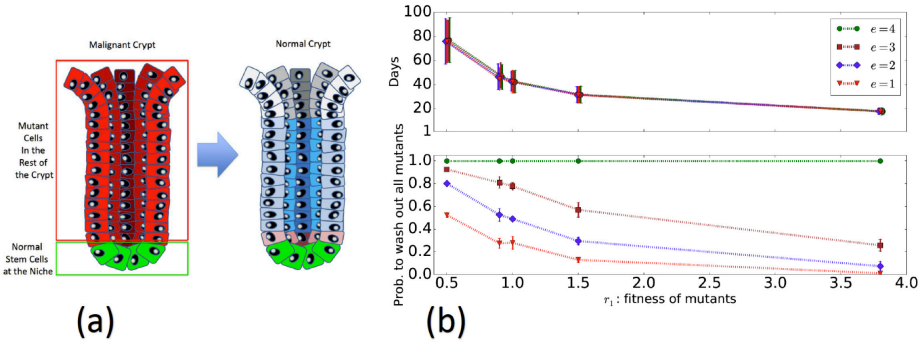


Figure 5.8: **Probability and time that mutants are washed out from the crypt.** Cartoon picture in (a) shows how a S_c compartment with all normal stem cell is able to wash out mutants in the rest of the crypt. The sub-figure (b) shows the result of simulations indicating the average time and the probability that the all crypt's cells become wild-type, i.e. all mutants are washed out from the crypt. In this figure $|S_b| = 7, |S_c| = 4$, and $u = v = 0$, other parameters are given in Table 5.1. The points are the average, and the bars indicate the standard deviation.

mutant stem cell will take over the FD compartment is small (between 0.01 and 0.14) in the one stem cell compartment model. However, this probability is zero for the progeny of a mutant BSC in the 4-compartmental model.

In the 4-compartmental model, if a mutant appears in the CeSCs, with a high probability it stays in the crypt, and its progeny will take over the entire crypt especially for advantageous mutants. However, in 3-compartmental model, the possibility that an advantageous mutant stem cell differentiates to two TA cells and be removed from the crypt is high. This result may suggest the existence of only one stem cell compartment is an advantage, however if half of SCs become mutants, then mutants take over the entire crypt with a high probability in the 3-compartmental model. In the 4-compartmental model, if all BSCs are mutants (more than half of SCs) and all CeSCs are wild-type, then mutants will be washed out from the crypt in less than 3 months. Furthermore, in the work done by Shahriyari et al. [154], it has been shown that the bi-compartmental stem cell niche delays the mutants generation.

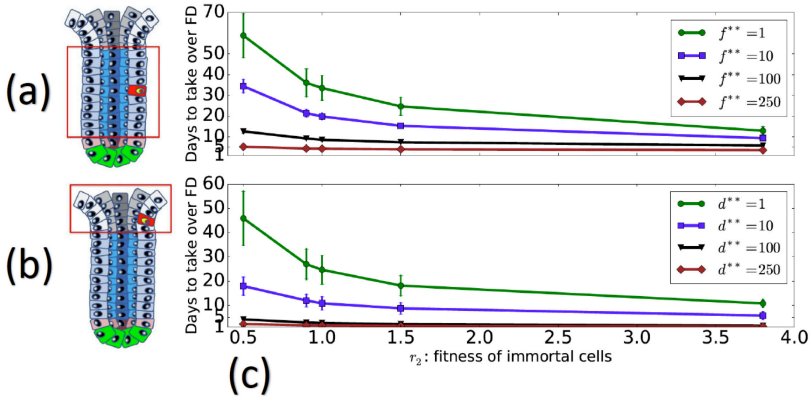


Figure 5.9: **Fixation of immortal cells in the FD.** The sub-figures (a) and (b) represent a schematics view of the system at the initial time of simulations generating the bottom and top sub-figures of (c), respectively. In the top sub-figure (c), the process starts with f^{**} immortal cells, while the rest are wild-type. In the bottom sub-figure (c), at the initial time there are d^{**} immortal cells in D_t and other cells are wild-type. In both sub-figures, we obtain the time that immortal cells take over the entire FD.

5.11.7 Time to Fixation and potential therapeutic treatments

An important concept corresponding to the fixation probability of a given finite Markov chain is the time to fixation. This quantity measures the time that the progeny of a single mutant cell require to take over an entire compartment. The time to fixation can be very important when it approximates the tumor growth period, the time needed for tumor initiation, or clonal conversion. Moreover, it seems crucial to have an estimation for the time of metastasis for an invasive mutant population when the epithelial markers divert to mesenchymal markers in a somatic cancer.

In the current study, assuming the Moran process for a four–compartmental model as described in the analytic tools section, we performed a wide variety of time estimations to maintain some critical features of tumor development within the crypt. We focus our attention on the fixation time of some initial mutant(s) in the central stem cell compartment. In Figures 5.11-(a) and (b) the average fixation time is given for different percentages of mutants in the central stem cell compartment opposed to the probability of symmetric division (σ) for parameters obtained based on the experimental data summarized in Table 1.

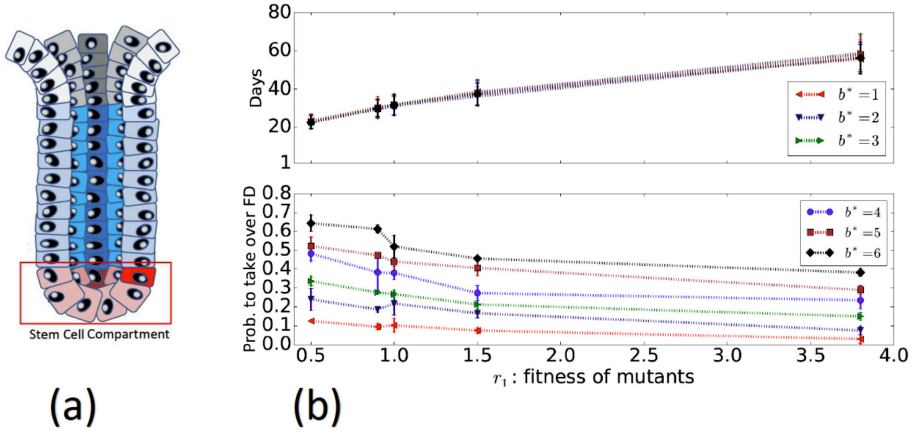


Figure 5.10: **One stem cell group instead of two compartments.** The figure (a) shows that there are b^* number of mutants in the one stem cell group at the initial time of the simulation, and (b) presents the result of simulations. The bottom sub-figure of (b) indicates the probability that the progeny of b^* number of mutant SCs will take over the entire FD, and the top sub-figure shows the time of its occurrence. Here, the total stem cell population is $|S| = 11$, and the rest of parameters are given in Table 5.1.

Another interesting result is represented in Figures 5.11-(c) and (d) in which the fixation time (washed-out time) of central stem cells depicts the number of days it takes for a central stem cell compartment full of wild-type individuals to completely sweep out the rest of the crypt covered by mutants. In these figures different regimes have been considered for neutral and advantageous mutants (various values for r_1) and different probabilities of proliferation in the central stem cell group (γ). Other important observations can be found in the results section of the paper.

In the present study, the high level of dependency on central stem cells has been observed for colonic/intestinal crypt in light of analytic analysis and numerical simulation which confirm the existing experimental data and predict some new consequences. According to the results, we suggest some significant and non–trivial methods which may improve diverse therapies. Firstly and most importantly, we found that in the absence of immortal cells a normal central stem cell compartment will sweep out all malignancies of the rest of the crypt, thus substituting the mutant stem cells in the niche with the wild-type ones will cure the crypt. This can be a potential method for chemotherapy and radiotherapy to

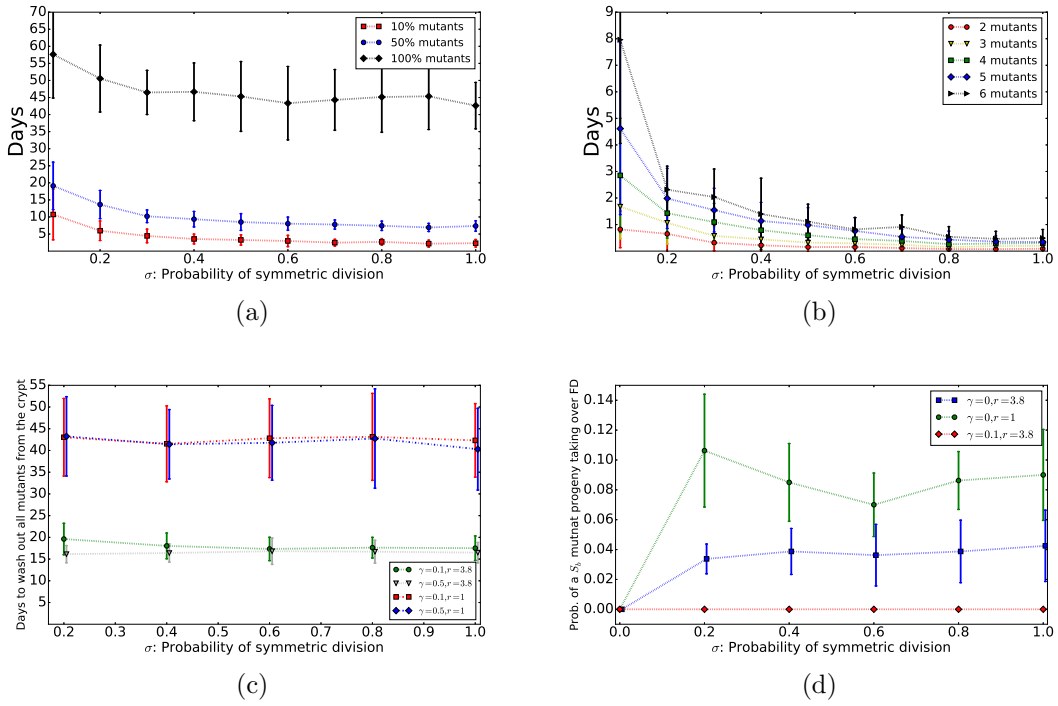


Figure 5.11: **(a)-(b) The average spreading time of one mutant central stem cell.** The sub-figure (a) shows the average time that the progeny of one mutant central stem cell will need to take over 10%, half, and the entire crypt. The sub-figure (b) shows the average time that one mutant central stem cell needs to generate 2-6 mutant central stem cells. In this figure $|S_b| = |S_c| = 6$, and $u = v = 0$, other parameters are given in Table 1. The points are the average time, and the bars indicate the standard deviations. **(c) The average time that the progeny of central stem cells need to take the entire crypt.** At the initial time of this simulation all cells are mutants except central stem cells. We calculate the average time that the crypt evolves, and all cells become wild-type. In this figure $|S_b| = |S_c| = 6$, and $u = v = 0$, other parameters are given in Table 1. The points are the average time, and the bars indicate the standard deviations. **(d) The probability that the progeny of one mutant stem cell takes over the FD in One and two stem cell compartment models.** In this plots circles and squares indicate the results of simulation for the one-stem cell compartment model, and diamonds are the results of two stem cell compartment model. In these simulations we start the system with one border stem cell mutant, and we obtain the probability that the progeny of the mutant cell takes over the FD group.

control malignancies of the crypt.

Another approach of this attempt relates to two important quantities in evolutionary dynamics and population genetics, the fixation probability and time to fixation. These two concepts have been investigated in the current work to presumably describe the complex intra/extra–cellular mechanisms in the colonic/intestinal crypt and then to predict the chance and required time of different critical events within the onset of tumorigenesis. Taking the derived results into account may tend to an improvement in arranging more effective schedules for clinical trials, drug delivery strategies and radiation exposures in anti–cancer treatments.

Numerical simulation

In order to obtain the fixation probability and time to fixation through simulation, we set the maximum updating time T equal to 10,000,000. Then, we run the algorithm for 100 times, and we calculate the ratio of the fixation occurrence number out of 100. We repeat this process for 5 times to obtain the mean and the standard deviation. Moreover, to achieve the time of occurrence, the occurrence time collected for each single run whenever the fixation appeared. Then, we obtained the average and standard deviation of these times. To convert the simulation time to be in terms of day, we assumed that the average cell cycle time of the crypt be equal to one day [20]. This means that having the total number of cells equal to N, then the time step t is equivalent to $t/N + 1$ days.

Summary

The absence of APC, which causes aberrant migration [149], is the most frequent mutation in colon cancer[78]. If all cells in one compartment lose their APC, then a tumor initiates because cells located lower than this compartment cannot migrate to the top of the crypt. For this reason, we calculate the probability that mutants are fixated at each of the compartments. The simulations and analytical calculations, which are in perfect agreement, show that the progeny of one non-stem non-immortal cell regardless of its fitness are not able to take over any compartment. However, the progeny of one immortal neutral FD cell will take over the entire FD in less than one month.

Only CeSCs are able to take over the entire crypt, and they are able to renew the CeSC compartment in fewer than 2 days, and the whole crypt in fewer than 3 months. Although the fitness of CeSCs does not make much difference in their fixation time in the CeSC compartment, it affects the time that they need to take over the entire crypt. The disadvantageous mutants take over the crypt quickly, but the probability of this occurrence

is small. For example, the probability of one disadvantageous CeSC mutant's progeny like $P53^{R172H}$, which has a fitness of 0.9, taking over the FD compartment is 0.2 in the non-inflammatory environment. However, if it happens, they will spread over the entire FD in less than one month. Moreover, the probability of the progeny of one CeSC $APC^{-/-}$ mutant, which has a fitness of 3.8, taking over the FD is 0.8. Importantly, they need at least 40 days to spread over the entire FD compartment (Figure 5.6). We conclude the probability of the progeny of the mutant CeSCs with lower fitness like $P53^{R172H}$ mutants, taking over over the crypt is small, but if it happens it is a fast process. Additionally, the probability of the progeny of the mutant CeSCs with a high fitness, like $APC^{-/-}$ mutants, taking over the crypt is high, but it is a slow process.

Chapter 6

The Pathogenesis of the Acute Myeloid Leukemia

In the last three chapters, our main concentration was on the stochastic analysis of some important spatial structures, which mostly applied to the understanding of some aspects of colorectal/intestinal cancer. The results of these chapters are broadly applicable and can be used in diverse disciplines with similar structures. One important application of such results is in the study of blood cancers, where there exist sequential compartments of stem and differentiated cells through the lineage of stem cells. Moreover, symmetric and asymmetric divisions occur within different compartments and dedifferentiation may occur through different stem cell generations. Also blood cells, presumably more than any other tissue cells, are under the effect of environmental fluctuations. All these aspects can be taken into account to illustrate the stochastic behavior of blood cancers. Another interesting tool to investigate the general mechanism of blood cancer relates to the statistical analysis of genetic data where genetic/epigenetic changes are encoded.

In this chapter, which has been done in collaboration with John Dick's laboratory, we focus on the statistical analysis of genetic data for a common type of leukemia, referred to as acute myeloid leukemia (AML). The determination of the phylogenetic tree and the signature genes of AML are two important open problems in AML that are explored in this chapter. The phylogeny of AML and the order of signature gene mutations in this tree would describe the spatial structure of the different compartments of normal and leukemic cells. Being able to construct the phylogenetic tree of AML, one can employ the results of the preceding three chapters to presumably arrive at a better understanding of this heterogeneous malignancy.

According to experimental results, the interaction between genetic inheritable functions termed inter-cellular features on the one hand, and micro–environmental effects as intra–cellular processes on the other hand, would allow determination of the genetic pathway of leukemia cells. This means that in the hematopoetic system, not only oncogenes play a crucial role in cancer development, but also the alterations to the surrounding environment can influence the heterogeneity of the blood system. There is no well-known empirical and standard method to measure micro–environmental effects; however, genetic/epigenetic changes reveal alterations in the expression level of functional and non–functional genes compared with the normal gene expression. This would delineate important pathogenetic information about the lineage of AML. Investigation of the change in hypoxia levels and acidity of surrounding tissue(s), as well as the role of entropy (negative free energy) and signaling in cancer development are some of the very recent attempts to understand the environmental impacts [143, 93, 173]. Nevertheless, the most reliable evidence of phenotypic/geneotypic alterations can be found in the form of genetic data. Genetic/epigenetic data may illustrate the tumorigenic behavior of a specific cancer in terms of the hierarchy of mutations, epigenetic effects and dedifferentiation, selective advantage, and clonal conversion. In this chapter, we envisage some features of AML based on gene expression analysis and constructing phylogenetic trees. The main goal is to understand the hierarchical pathway of mutations leading to a lesion of leukemia cells.

6.1 Introduction

Cancer is a disease which can be seen as a result of genetic alterations and intratumoral influences of tumor heterogeneity. Intratumoral factors are related to metabolic fluctuations. For instance, when the tumor environment is hypoxic or when the functionality of tumor cells varies due to nutrition distribution, the survivability of tumor cells can be affected. Genetic diversity, tumor microenvironment, and epigenetic changes are interconnected in the lineage of hematopoietic cells within tumorigenesis [81]. Gene–expression signatures of cancer and normal stem cells have been used to find various driver mutations which define the early and late events in the hierarchy of oncogenes. The tumor microenvironment, in turn, plays an important role in the fate of tumors and has been thought of as a reason of chemotherapy failure [50].

AML has a complex mechanism which makes it, as a clonal abnormality, a dangerous type of leukemia. Such a malignancy affects myeloid cells and is characterized by the accumulation of stem/progenitor cells in the blood and bone marrow. AML has been thought of as the result of change in the transcription factor regulation, which causes

defects in the normal mechanism of differentiation in the blood system.

Similar to the general mechanism of hematopoiesis, the SC hierarchy in AML can also be designed with a four–compartmental model where the proliferation and differentiation of leukemic SCs maintains the diverse phenotypes within the hematopoietic system [134, 150]. AML has been shown to be produced from a rare set of $CD34^{+}CD38^{-}$ leukemic initiating cells at the apex of tumorigenesis [27].

Based on morphology and differentiation status (maturation), traditionally there exist eight different types of AMLs, referred to as French-American-British (FAB) classification. Figure 6.1 represents the seven different categories of malignant types; the 8th category if for normal samples with no symptoms for AML. There exist some other classification such as WHO (the world health organization) classification which characterizes various types of AMLs based on their disease development (dysplasia) and diverse chromosome translocations [3]. Nevertheless, we will try to define a new classification of AMLs based on some novel subtypes and regarding the most recent research for leukemia–initiating mutations.

Table 6.1: FAB classification of AML cells

FAB Class	Subtype Feature
M_0	dedifferentiated
M_1	Myeloblastic
M_2	Myeloblastic with maturation
M_3	Promyelocytic
M_4	Myelomonocytic
M_5	Monocytic
M_6	Erythroleukemic

As we described in Chapter 1, the accumulation of mutations which tends to generate a chain of sequential clonal expansions, is known to be the main reason for cancer development. Among various clonal populations, the clone of the fittest cells dominates the host tissue. After acquiring more advantageous mutations, the progeny of the colony which contains the mutations may have a higher growth rate [81].

As we described in above, because of the lack of a general mechanism to measure the microenvironmental and genetic/epigenetic alterations, initiation and progression of cancer can be traced from the dynamics of genomes via throughput genome sequencing of tumors. Recent technologies allow for whole genome and exome sequence analysis which

may provide important information for the hierarchy of tumor cells, compared with the genetic data of similar normal cells. Such an analysis on the clonal expansion of different advantageous mutations reveals a pathogenesis of a founder HSC as the root of the tree. When the founder HSC emerges then the subsequent mutations create new branches of the tree until the fittest subclone(s) take over the whole or majority of the organ. This procedures introduce the phylogenetic tree of mutational events leading to AML.

Phylogenetics, in fact, is a testable hypothesis about relationships among phenotypic or genotypic diversities and seeks to understand the corresponding phylogeny of some species based on their shared traits and common functionalities. Finding the best way of combining the information contained in numerous different gene-trees for the same set of species remains an open problem in contemporary biology. Biologists use different methods to construct trees from DNA sequence. The goal of phylogenetics is to obtain the true phylogeny for a given group (taxa) of species. Phylogenetics has applications to molecular biology, genetic evolution, epidemiology, ecology, conservation biology, forensics, and oncology.

Compared to solid cancers in adults, leukemias require the lowest number of genetic mutations in their genomes within tumorigenesis. On average, there are 13 mutations reported in the pathogenesis of AML [85] with only 5 of them being recurrent mutations [18]. This result has been observed in a comprehensive study of 200 AML samples [81, 85]. Moreover, 23 of samples showed distinguishable mutations while another set of 237 mutations rarely occurred within the samples. Despite huge progression in the study of AML, complexity of the inter/intra-tumor heterogeneity has not thrown light on how to divide mutational events into driver and passenger mutations. However, the origin of passenger mutations in AML were mostly observed in the founder HSC of the dominant subclone [191]. Another interesting approach is to construct the phylogenetic tree of single cells, in the lineage of cell of origin until maturations.

6.2 Exploring the expression data and sequencing manipulation

Genetic diversity (epigenetic differences) and Darwinian selection have been considered as the building blocks of modern biology. Reproductions in cells may lead to sequential mutations which occurs within the human genome and characterize the selection pressure on a certain type of individuals in a heterogeneous population. The alterations in functional genes describe how branching evolution leads to others species or various functionalities. To

study such a hierarchy of mutations, a wide variety of methods have been applied to derive applicable results from microarray, RNA, transcriptome, exome, chromatin, and protein sequencing methods. The human genome project has recently discovered human DNA to be 99.9% identical in different individuals with some differences in particular locations on the genome (by identifying more than 30,000 genes in the human body). Comparing these differences between individuals and especially between normal and malignant cells, RNA sequencing (RNA–seq) has generally been used as an important tool to study genetic diseases.

However, leukemia is a heterogeneous disease in which the heterogeneity in contribution with genetic diversity leads to tumor development and therapy failure [81]. To capture some features of the evolution of cancer, the only well–defined method is to pursue genetic alterations of tumor cells over time. Such an analysis may tend to a phylogenetic tree in leukemia. In [142], a phylogeny of acute myeloid leukemia (AML) has been represented based on a gene database of 362 AML patients and also 7 unclassified Myelodysplastic Syndrome samples [142]. More precisely, the authors construct a tree by considering a set of different FAB (French-American-British) classification of AML patients. They also considered some samples from peripheral blood (CD34 PB), bone marrow (CD34 BM), human mesenchymal precursor cells (hESCMPC), and fully differentiated mononuclear cells from peripheral blood (PB) and bone marrow (BM) to construct the phylogeny of AML (see Table 6.1).

After normalizing the microarray sequences, the authors used false discovery rate (FDR) of 0.05 (and less) to filter significant genes. Then they found the matrix of average gene–expression differences in each subgroup of samples to construct the most consensus tree. The leaves of the tree are the subgroups. Their results are general and reveal the hierarchy of diverse types of samples between stem and fully differentiated cells. In this section, applying a similar method, we analyzed the RNA–seq data of 179 AML samples from the Cancer Genome Atlas (TCGA). Then we constructed the phylogenetic tree of AML compared with normal skin samples from the Gene Expression Omnibus (GEO). In our investigation, we additionally use single cell data of hematopoietic cell with DNMT3A mutations, generated by John Dick's laboratory.

The idea is to construct a new classification of diverse subtypes of AML as opposed to the conventional FAB classification. Each subtype in this new classification includes all samples which experienced a specific set of mutation(s) which determines the label of that subtype. Then our analysis allows us to understand the relationships among various subtypes (various sets of genes). Using different statistical methods, we construct the phylogeny of the subtypes. Such a phylogeny may help us to understand the hierarchy of mutations and/or parallel hierarchies that occur(s) for this set of AML samples. Either

unsupervised or supervised clustering can provide useful information about the evolution of AML for the given database. Unsupervised clustering reveals how different mutations in different samples can be clustered and how diverse mutations are distributed all over the samples in comparison to the normal samples. However, by defining the new subtypes of AML, supervised clustering maintains the connections between mutual subtypes.

Table 6.2: The suggested new classification of AML based on [85].

Specific group	Genetics	Specific group	Genetics
G_1	PML-RARA	G_8	Other myeloid TFs
G_2	MYH11-CBFB	G_9	MLL-X fusions,MLL-PTD
G_3	RUNX1-RUNX1T1, RUNX1	G_{10}	NUP98-NSD1
G_4	PICALM-MLLT10	G_{11}	ASXL1
G_5	DNMT3A,DNMT3B, DNMT1	G_{12}	EZH2
G_6	TET1,TET2	G_{13}	Cohesin
G_7	IDH2	G_{14}	NPM1,TP53,WT1,PHF6, FLT3,KIT,Other Tyr kinases, Ser-Tyr kinases,KRAS/NRAS, PTPs,KDM6A,Other modifiers

According to the results of [85] for 200 AML samples from TCGA, a complete set of possible mutations has been reported for each of the patients. This analysis was performed on the whole genome sequencing (WGS) of 50 patients, the AML exome sequences of 150 samples, and also DNA Methylation results for these 50 samples. DNA Methylation is usually used to capture epigenetic events based on regulatory mechanisms, was also performed. Then based on these types of data, the frequencies of different mutations have been reported in 9 categories regarding to the transcription–factor fusions, gene encoding nucleophosmin, signaling genes, chromatin–modifying genes, tumor suppressor genes, myeloid–transcription factor, splicesome, DNA methylation–related, and cohesion. Then the most common mutations with the highest frequencies were found for the considered AML dataset. Among a wide variety of different mutations, NPM1, DNMT3A, and FLT3 have the top three frequencies. Moreover, new discoveries via targeted sequencing on AML genes suggest IDH1, IDH2, CEBPA, KIT, and TET2 may play crucial roles in the development of leukemia at prognosis and intermediate/high risk levels [118, 85, 131, 157]. Therefore, these types of mutations can be used to define some of the subtypes. Branching

Table 6.3: The Suggested classification of AML: Early vs. late events

No.	Early events	Late events
1	PML-RARA	NPM1
2	MYH11-CBFB	TP53
3	RUNX1-RUNX1T1/RUNX1	WT1
4	PICALM-MLLT10	PHF6
5	DNMAT3A/ DNMT3B/ DNMT1	FLT3
6	TET1/ TET2	KIT
7	IDH2	Other Tyr kinases: ABL1/ DYRK4/ EPHA2/ EPHA3/ JAK3/ MST1R/ OBSCN/ PDGFRB/ WEE1
8	NUP98-NSD1	SerThr kinases: ACVR2B/ ADRBK1/ AKAP13/ BUB1/ CPNE3/ DCLK1/ MAPK1/ YLK2/ MYO3A/ NRK/ PRKCG/ RPS6KA6/ SMG1/ STK32A/ STK33/ STK36/ TRIO/ TTBK1/ WNK3/ WNK4
9	ASXL1	KRAS/NRAS
10	EZH2	PTPs: PTPN11/ PTPRT/ PTPN14
11	Cohesin: SMC1A/ SMC3/ SMC5/ STAG2/ RAD21	KDM6A
12	Other myeloid TFs: GATA2/ CBFB/ ETV6/ ETV3/ GLI1 IKZF1/ MYB/ MYC/ MLLT10-CEP164	Other modifiers: ARID4B/ ASXL2/ ASXL3/ BRPF1/ CBX5/ CBX7/ / EED/ HDAC2/ HDAC3/ JMJD1C/ KAT6B/ KDM2B/ KDM3B/ MLL2/ MLL3/ MTA2/ PRDM9/ PRDM16/ RBBP4/ SAP130/ SCML2/ SUDS3/ SUZ12/ ZBTB33/ ZBTB7B/ CREBBP-KAT6A/ RPN1-MECOM/ RUNX1-MECOM
13	MLL-PTD	Spliceosome: CSTF2T/ DDX1/ DDX23/ DHX32/ HNRNPK/ METTL3/ PLRG1/ PRPF3/ PRPF8/ RBMX/ SF3B1/ SNRNP200/ SRRM2/ SRSF6/ SUPT5H/ TRA2B/ U2AF1/ U2AF1L4/ U2AF2
14	MLL-X fusions: MLL-ELL/ MLL-MLLT4/ MLL-MLLT3/ MLLT10-MLL	

processes in AML may occur through short–term evolutions such as those for acute promuelocytic leukemia (APL) which defines the class G1 mutations (see Table 6.2). Such a

branching process may also occur through long–term mechanisms involving the accumulation of many passenger mutations and a wide variety of subclones such as the evolution of myelodysplastic syndrome (MDS). However, the main question is about the order of occurrence of mutations. To this end, we first explored preleukemic and driver genes and specifically investigate early, intermediate, and late events in AML samples. Although the initiation and progression of this heterogeneous and invasive malignancy has been poorly understood.

Recent studies have shown some evidence which characterizes the role of some specific genes such as DNMT3A, NPM1, PML, RUNX1, TET2 and so on in the development of AML [159, 158, 85, 145, 170]. Acquiring the novel discoveries, our new classification of AML can be established in order to derive some features of such a complex disease. Our suggesting list comprises 14 subtypes which are given in Table 6.2. The potential early and intermediate events provide the majority of subtypes while the potential late events are basically categorized in one subtype. Recent research which is mostly focused on the hierarchy of mutational events in AML [83, 81, 89, 85, 130, 151] depicts a list of early and late events, given in Table 6.3.

6.3 A phylogenetic tree of AML

Constructing the phylogenetic tree of AML is not only important to find the hierarchy of mutational events, but also results in a better understanding of the whole mechanism tending to a set of driver mutations. It also results in finding a set of signature genes, which is a collection of possible mutations that cause initiation and progression of AML. Moreover, the existence of such a phylogeny provides information about the survival chance of patients, the type of drug which can be used to prevent the progression of the cancer, and also about the relapse mechanism.

6.3.1 General algorithm to construct the phylogenetic tree

There are a wide variety of different methods to construct a phylogenetic tree; however, the following algorithm sketches an overview of some crucial steps to construct such a tree.

Step 1. Firstly, a specific type of data from reliable resources is required. Particularly, this type of data can be found from a public nucleotide database. For instance, public data of different types, from gene expression and microarrays to SNV and raw data can be

found from DNA Data Bank of Japan (DDBJ), European Molecular Biology Laboratory (EMBL), or GenBank (in USA). Genomic data are accessible through the websites of the Institute for Genomic Research (TIGR), the Join Genome Institute (JGI), the Sanger Institute, and the National Center for Biotechnology Information (NCBI). Moreover, the evolutionary data of genomes-in-progress can be found in GenBank and NCBI which has more than 44 billion base pairs of DNA. The data uploaded from these resources are usually normalized but they might contain batch effects as they are produced in diverse situations. In such a case, all data needs to be normalized after releasing the batch effects.

Step 2. Acquiring the data, to specify the location of each gene throughout the genome, *annotation* process is used. Finding a general and standard annotation method of genomes has still remained unsolved. Annotation is mostly used through the following general tools:

- **Keywords:** which annotate sequences by looking through their written descriptions. This method is easy, intuitive, and mostly used for genomic data for which high throughput analysis seems to be essential. Some of the softwares associated with this method are Entrez (NCBI) and SRS.
- **Similarity:** which looks over the sequences themselves to find the similar parts of the sequences and annotate based on those similarities. For instance, BLAST is a software which works based on similarity. In this software, the hidden Markov model (HMM) is applied to predict coding–base regions by a high level of probability.

Step 3. After annotating the genetic data, the next important step is associated to the *alignment* process. Alignment stands for the method of sorting sequences through a columnar list in which similar annotated parts are aligned in the same column. Multiple sequence alignment plays a crucial role for this purpose and is placed in the heart of constructing phylogeny. There exist different approaches to perform the alignment:

- Using pairwise (or multiple) sequence alignment models, the evolution of sequences regarding insertion, deletion (insdel), and mutation could be characterized.
- Progressive sequence alignment: this method starts with the most similar sequences and show the guide tree which is the most probable case for alignment. Then progressively the more dissimilar (divergent) sequences will be added to the guide tree.
- Pair HMM for two (or more) sets of observations and hidden variables is also used as an other approach.

Some important programs for the process of alignment on annotated data are ClustalX, BioEdit, BCM, BLAST, and GCG.

Step 4. The next step is dedicated to the construction of a tree. The methods for calculating phylogenetic trees fall into two general categories:

- **Distance-matrix methods (clustering or algorithmic methods):** This method is relatively simple and straightforward and can be performed by applying simple calculations. The distance which is, in fact, based on differences between the distribution of expressions in each sequence. Each distribution is defines as the set of all expression levels over the landscape of different genes. Then the distances are assembled into a tree. Some of the methods to calculate the distance metric are UPGMA, Euclidean, Pearson, neighbour-joining, Fitch, and Margoliash.
- **Discrete data (tree-searching) methods:** This method examines each column of the alignment separately and search for the tree that best accommodates all the information. This methods has more fruitful information by tracing the evolution of specific sites such as catalytic sites or regulatory regions. Some of the main tools regarding the tree searching method are parsimony, maximum likelihood, and Bayesian methods. Moreover, Mega2, PHYLIP, and Treeview are some softwares to construct the phylogenetic tree based on this method.

Step 5. Finally and after constructing the tree we look for the *consensus tree* which depicts the most probable tree among all possible choices. There are two main methods to test the temporarily constructed phylogeny:

- **Bootstrapping:** This method tests whether the whole dataset supports the constructed tree, or if the tree has been chosen among nearly equal alternatives. Starting from some random subsamples of the dataset, this method builds the corresponding tree of each subsample and calculates the frequency of the random subsample compared with the whole database.
- **Long branches:** This method accounts for highly divergent sequences, i.e. those sequences that have long terminal branches. Then it groups them together in a tree regardless of their true relationships. This can be applied when sequences do not have close relatives but may have numerous unique mutations.

6.3.2 Pipelines and the reduction of data

In many genetic expression analysis, such as that for RNA-seq analysis which includes a big list of essential and redundant genes, various pipelines may be used to filter the redundant information which may add noise to the data. Such pipelines help to clean the data to be more efficient in terms of data analysis and tree construction. One of the most important pipelines, generally applied to reduce a range of useless information, is the 't-test'. This method can be used to compare the mean of a sample to the means of the other known samples for a randomly drawn sample from the data. The distribution of the mean of the chosen sample is assumed to be normal. For an unpaired (independent group) t-test, the following formula is used:

$$t(x,y) = \frac{\langle x\rangle - \langle y\rangle}{\sqrt{\frac{\sigma_x^2}{n(x)} + \frac{\sigma_y^2}{n(y)}}}, \tag{6.1}$$

where $\langle x\rangle, \langle y\rangle$ and σ_x, σ_y are respectively the means and standard deviations of x, y. Another pipeline, referred to as 'p-value' is based on the false discovery rate (FDR) method, which reveals the probability of error in rejecting the hypothesis of no difference between the two given groups. A low p-value for a test (usually less than 0.01 or 0.05) means that there is an evidence to reject the null hypothesis (that is, the means of the two groups are equal) in favor of the alternative hypothesis. Two tailed form of p-value distribution can be calculated in the following way:

$$p(t,v) = \frac{1}{v^{1/2}\,\mathrm{Bessl}(1/2, v/2)} \int_{-t}^{t} \left(1 + \frac{x^2}{v}\right)^{-(v+1)/2} dx, \tag{6.2}$$

where $\mathrm{Bessl} = \int_0^1 t^{b-1}(1-t)^{a-1}dt$ and $v = n(x) + n(y) - 2$ is the degree of freedom in unpaired case. Applying this pipeline, one can find those genes which have significant variations. Beyond these two fundamental methods which have been widely used to find a more accurate number of significant data, there are also many alternative pipelines associated to the specific purpose of the analysis.

6.3.3 AML signature genes

The main question to address in this section is about the most common ancestor of the new subtypes of AML (Table 6.2). Finding such a hierarchy of mutations which is basically constructed compared with some normal samples and the hematopoietic stem cells (HSCs),

depicts the signature genes of AML. However, since the number of mutations in AML is very low, there might exist so many rare mutations occurring within the lineage of HSCs based on the existing data [85]. Among the most important concepts to address, the features of the founder stem cell(s) and the pathogenesis of HSCs seems complicated to investigate.

Knowing about the signature genes derived from the data, one can describe a hierarchy or diverse parallel hierarchies of AML. Then the pathway(s) can be validated with experimental data. For instance, the derived signature genes can be compared with the data provided by the single cell analysis. Then using the biomarkers known for the maturation process of hematopoietic cells, we can examine the type of mutation in terms of being an early, an intermediate, or a late event. Similarly, one can follow the method to derive the signatures of important AML genes such as DNMT3A, FLT3, TET2, WT1 known as the most common mutations based on the data [85]. By obtaining signature genes of the entire data, or signature of certain genes, the main purpose will be to obtain the accurate relationships among mutations and genotypic-phenotypic alterations in the hierarchy of AML mutations.

As we explained before, based on the recently discovered AML genes, a new list of 14 subtypes is proposed in this chapter. Each sample, then, belongs at least to one of the subtypes. For each of the subtypes we find the distribution of gene expression over the landscape of all the genes. Then the derived distributions will be compared with that of the normal subtype and other additional subtypes (if exist). Such a comparison can be done more efficiently by applying t-test or p-value analysis. Using t-test or p-value analysis, one can find significant gene expression differences. Then ignoring the genes with lower differences with respect to those of the normal subtype, we can conclude the list of important genes. Such a pipeline will define a more accurate relationship between the subtypes.

Based on the above illustration, starting from the 179 AML samples from TCGA in the form of RNA-seqs, we use a complete table of mutational events for each patients reported in [85]. Then manipulating the data for HSC CD34+ CD38-/ HSC CD133 CD34dim samples from GEO, we firstly find the unsupervised clustering of the these samples. Thus at first, we find the list of common genes between these two databases. Applying the bioconductor packages 'Phylogenetics', 'dvtools', 'biocGenetics', 'bioDist', and 'affy' through R modules, we find the unsupervised clustering of all samples with roots as HSC CD34+ CD38- and HSC CD133 CD34dim. The result is given in Figure 6.1 in which the immediate cluster from the roots is the PML cluster. This consequence is coincidence with the short evolutional of PML [191]. The the next cluster also seems to mostly contain mutations in FLT3 (see Figure 6.1) which is thought of as an late event.

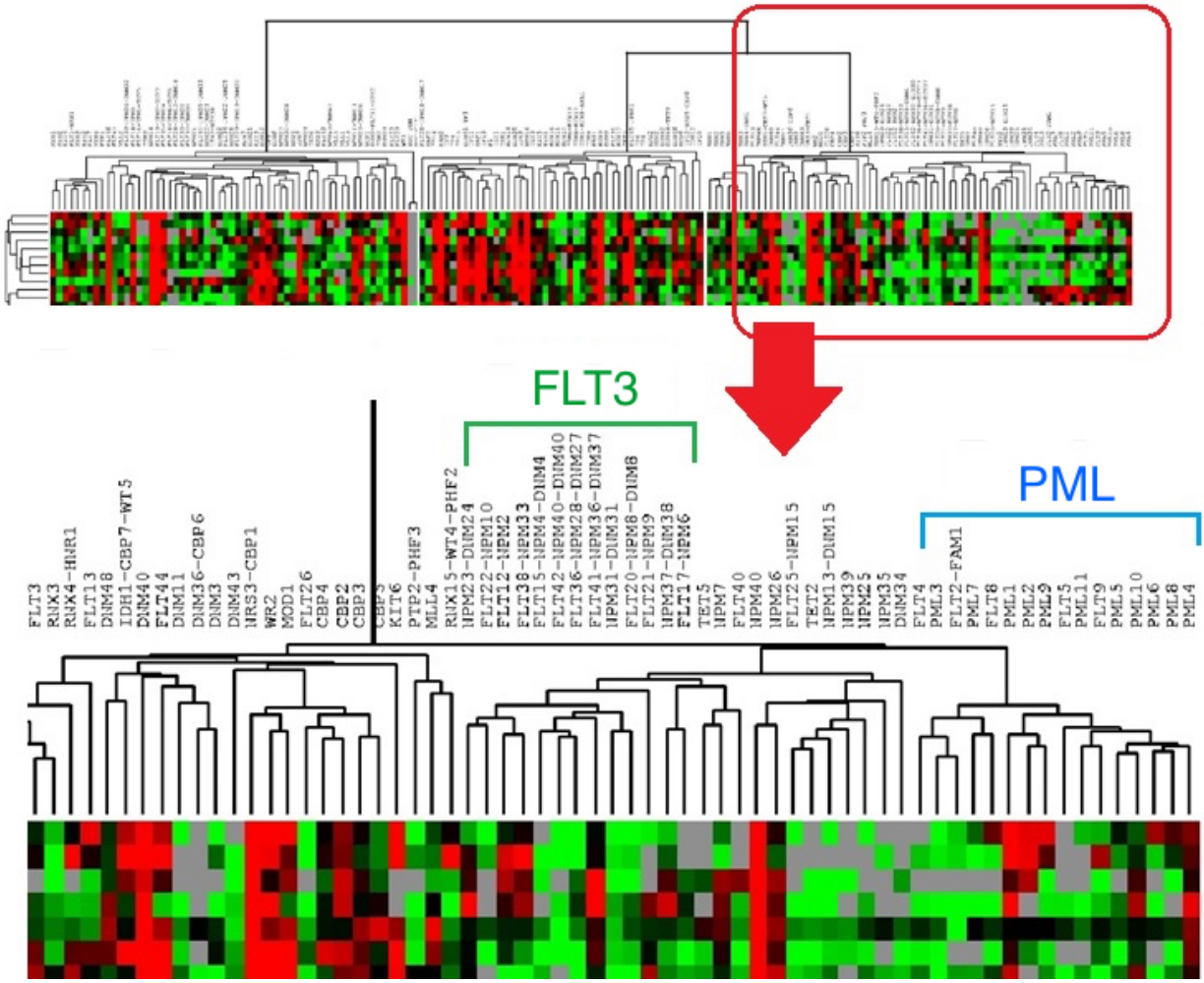


Figure 6.1: **Unsupervised clustering.** We perform the unsupervised clustering for 179 samples from TCGA and two more samples, HSC CD34+ CD38-/ HSC CD133 CD34dim from GEO.

Then we perform the following algorithm to find the hierarchy of mutations based on supervised clustering (classification) and considering those genes which exist in one subtype but not in at least one of the other subtypes:

- Applying (unpaired case) t-test between each subtype and mutant HSC samples, and normal single HSCs.

- Finding the p-value and significant genes for each subtype using FDR less than 0.01 (or 0.05).
- Through differentiation-based method [142], those genes are selected for the analysis which exist in one subtype but not in at least one other subtype.
- Finding the average expression matrix of the new list of differentiation-based genes.
- Finding the distance matrix of all subtypes and then constructing the tree by considering the large number of bootstrap (10,000 times).

6.3.4 Supervised and unsupervised classification

Firstly, we use the 179 normalized data from TCGA and HSC CD34+ CD38- and HSC CD133 CD34dim from GEO to understand the order of mutations on the lineage of HSCs. Figure 6.2 represents $G_1, \cdots, G_{14}$ as the novel subtypes described in the previous sections. Then we apply FDR p-value test at various levels, ranging from 0.01 to 0.05, to filter significant genes. Our analysis shows that 17,433 genes are in common out of 19,065 genes given in the database. Then using the selected set of genes we perform a hierarchical clustering with stability assessment based on bootstrap method by using different metrics. In continuation, we employ the consensus method by using Phylip software. The consensus method provides the most consensus set of trees. Then using TreeView software, we unify the produced trees to the most probable one.

Analysis shows that the potential signature genes for AML may contain PML mutation in the lineage or at least in one of the parallel lineages of pre–leukemic/leukemic HSCs. Also having FLT3 in the subsequent cluster shows that important mutations such as DNMT3A, TET2, WT1, FLT3, PHF6, and PTPs also need to be taken into account. These mutation are all located at the same subtype of FLT3 and can be another candidate mutations in the AML signature gene. Moreover, NUP98, MLLs, RUNX1, and NPM1 mutations are other potential candidates for the signature genes of AML regarding their close locations to the root of the phylogenetic tree (see Figure 6.2). This analysis is required to be experimentally validated and statistically confirmed through diverse investigations not only gene expression data, but also at the transcriptional and chromatin levels.

Secondly, we perform a classification for a set of patients which have a certain type of mutation, e.g. for DNMT3A, and restricting each subtype to this set of patients and relabeling the non-empty subtypes as $R_1, \cdots, R_{14}$ which are respectively the restricted sets of $G_1, \cdots, G_{14}$ to only include DNMT3A samples. To complete the analysis we have also

Figure 6.2: **Supervised clustering (classification) of AML samples.** Considering the 179 TCGA samples, and HSC CD34+ CD38- and HSC CD133 CD34dim samples from GEO, we derive a classification of our new subtype in order to understand early vs. late events and the hierarchy of mutations to detect the signature genes of AML.

considered 15 normal single–cell samples generated by John Dick's laboratory mmebers. Then we run two different scenarios to presumably capture the main features of such mutations and their relationship:

- Filtering out the PML-RARA, MYH-CBFB, and RUNX1 samples, which are known as early events, the unsupervised clustering of all samples performed in order to understand the distribution of mutations in DNMT3A, FLT3, TET2, and WT1 within the samples. See Figures 6.3 and 6.4 which represents the distributions of these mutations all over the tree instead of having them clustered in some particular subgroups. Then a comparison among these distributions results in table 6.5, which reveals the average distance of each mutation within the tree from the root and within each group of patients with such mutations in common. Again, we use the 179 TCGA data as well as 15 controlled samples, and 14 DNMT3A diagnosed/relapse data from John Dick's laboratory. The derived data suggests that TET2 and then DNMT3A respectively have the first and second largest distances from the root using Euclidean bootstrap

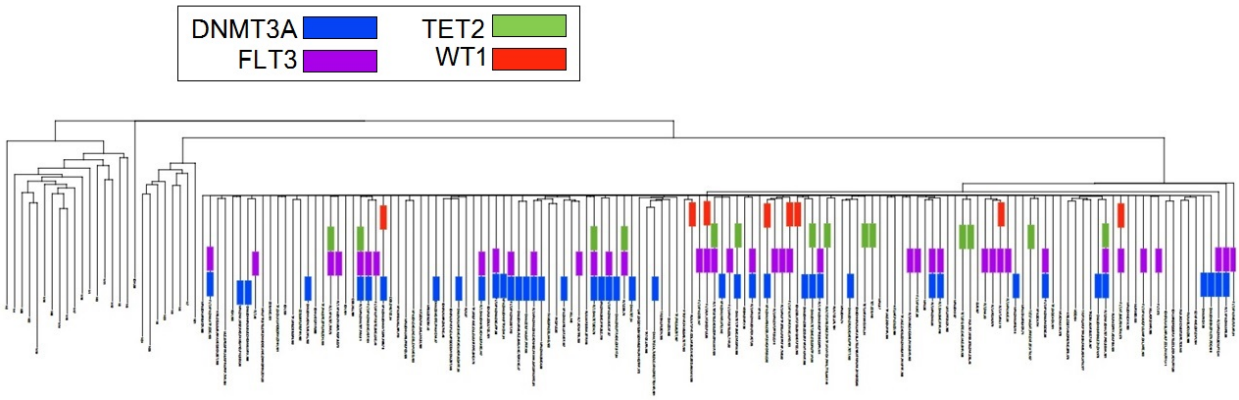


Figure 6.3: **Distribution of DNMT3A, FLT3, TET2, and WT1 mutation.** Using the 179 TCGA samples as well as 15 normal and 14 DNMT3A samples from John Dick's Laboratory, the unsupervised clustering is given using Euclidean metric for bootstrapping and neighbor–joining method to find the most consensus trees after filtering out the samples comprising PML-RARA, MYH-CBFB, and RUNX1 mutations.

and FastME consensus methods. This result is compatible with the experimental results which introduce these mutations as early events [83, 81, 89, 85, 130, 151]. Another evidence which show that TET2 and DNMT3A mutations as early events have the largest average distances compared with those of WT1 and FLT3 (late events). This result for DNMT3A mutation confirms the result of [83] arguing the high stability of DNMT3A and preceding of this mutation to NPM1 which is another mutation mostly referred to as a late event in the clonal evolution of AML.

- We, then, perform supervised clustering for the subtypes R_i $(i = 1, \cdots, 11)$ in comparison to the 15 normal samples (the controlled subtype), and 14 HSC DNMT3A single–cell data (the DNMT3A subtype) at diagnosis and relapse (see Figure 6.6). Different groups are denoted in different colors in the trees (blue, purple, green, and red are respectively related to DNMT3A, FLT3, TET2, and WT1 mutations). Our investigation about DNMT3A data results in a hierarchy of mutational events and assuming DNMT3A as the apex of mutations, reveals important facts about the potential signature genes of DNMT3A. It turns out that WT1, FLT3, NPM1, and TP53 which are located in the R_{14} cluster at the very first cluster within the phylogenetic tree. This analysis has done for Euclidean bootstrap and FastME consensus method for R_is.

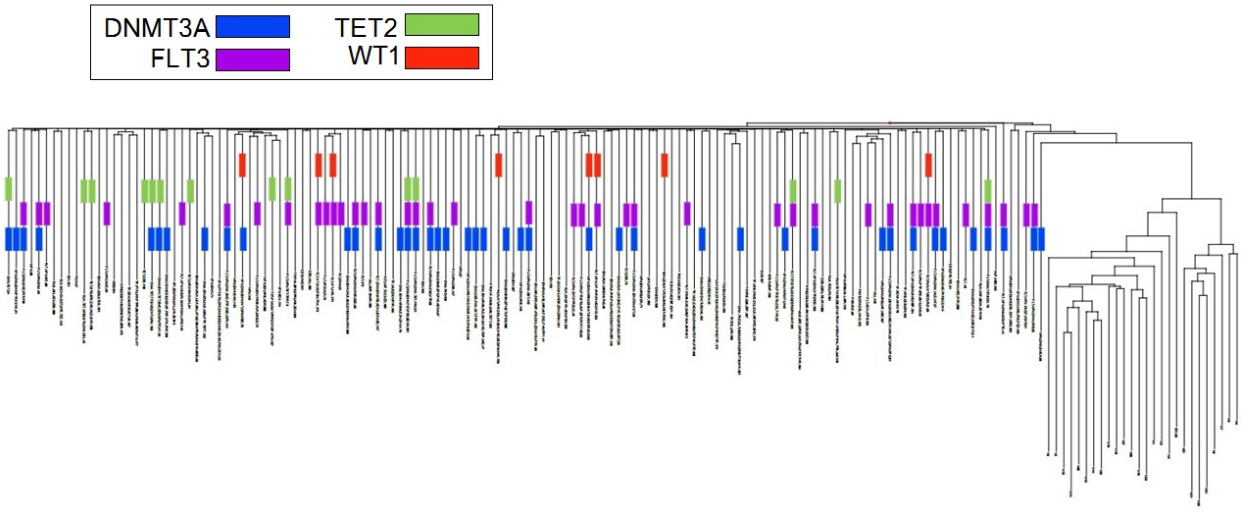


Figure 6.4: **Another distribution of DNMT3A, FLT3, TET2, and WT1 mutation.** Again we use the 179 TCGA samples as well as 15 normal and 14 DNMT3A samples from John Dick's Laboratory to find the distribution of DNMT3A, FLT3, TET2, and WT1 mutation within the unsupervised clustering filtering out the samples that include any of PML-RARA, MYH-CBFB, and RUNX1 mutations. We apply Euclidean bootstrapping method and neighbor–joining/fastme consensus facility. The blue, purple, green, and red strands are correspondence to DNMT3A, FLT3, TET2, and WT1 mutations respectively.

Summary

Despite years of research in diverse types of leukemia, different combination chemotherapy, and HSC transplantation over several decades, leukemia still remains among complex, dangerous, and poorly understood cancers. The low number of mutations leading to a founder (leukemia–initiating) stem cells, makes this type of disease complicated. Moreover, the high rate of relapse after chemotherapy among remitted cells also makes leukemia even more lethal. In this chapter, we tried to study some aspects of AML. Our main concentration was on the hierarchy of mutations. We used TCGA RNA-seq data and defined a potential list of new subtypes. Some recent techniques in statistical analysis of the genetic data, the construction of a phylogenetic tree of AML based on supervised and unsupervised clustering has been taken into account.

It turns out that having HSCs at the apex of such a phylogeny, PML mutations are occurring at the very first stage. This result was confirmed through the supervised cluster-

	Type of Mutation	Number of Samples	Euc-Nj Ave Distance from the Root	Euc-Fast Ave Distance from the Root	Euc-Nj Ave Distance Within the Group	Euc-Fast Ave Distance Within the Group
Late	WT1	8	4.725	4.59	7.1071	10.5429
Early	TET2	14	6	5.7278	7.7736	13.5527
Late	FLT3	40	6.315	5.1593	9.9610	13.1773
Early	DNMT3a	42	7.1571	5.2671	8.7359	14.1070

Figure 6.5: **Supervised clustering (classification) of AML samples.** Considering the 179 TCGA samples and HSC CD34+ CD38-/ HSC CD133 CD34dim from GEO samples, we derive a classification of our new subtype in order to to understand early vs. late events and the hierarchy of mutations to detect the signature genes of AML. The blue, purple, green, and red strands are correspondence to DNMT3A, FLT3, TET2, and WT1 mutations respectively.

ing (classification) of subtypes $G_1, \cdots, G_{14}$. This result may have substantial effect on the development of chemotherapy and may improve our understanding of drug-resistant traits of leukemic disorders. Another observation delineates that even FLT3 mutation which has though of as an early event, occurred in the majority of samples in the subsequent cluster to the PML cluster.

We also find the unsupervised and supervised clustering of specific types of mutations such as DNMT3A. This helps us to understand that early events such as DNMT3A and TET2 are more distributed within the samples and farther from the root. This observation has performed in comparison to the late events WT1 and FLT3, ignoring those samples which include early events PML-RARA, MYH-CBFB, and RUNX1. Our analysis shows that the potential set of AML signature genes should contain PML mutation in the lineage or at least in one of the parallel pathways of pre–leukemic HSCs. Also having FLT3 in the subsequent cluster shows that important mutations such as or at least in one of the parallel pathways of DNMT3A, TET2, WT1, FLT3, PHF6, PTPs are other candidates located in the same subtype as that of FLT3. Furthermore, MLLs, RUNX1, NUP98,

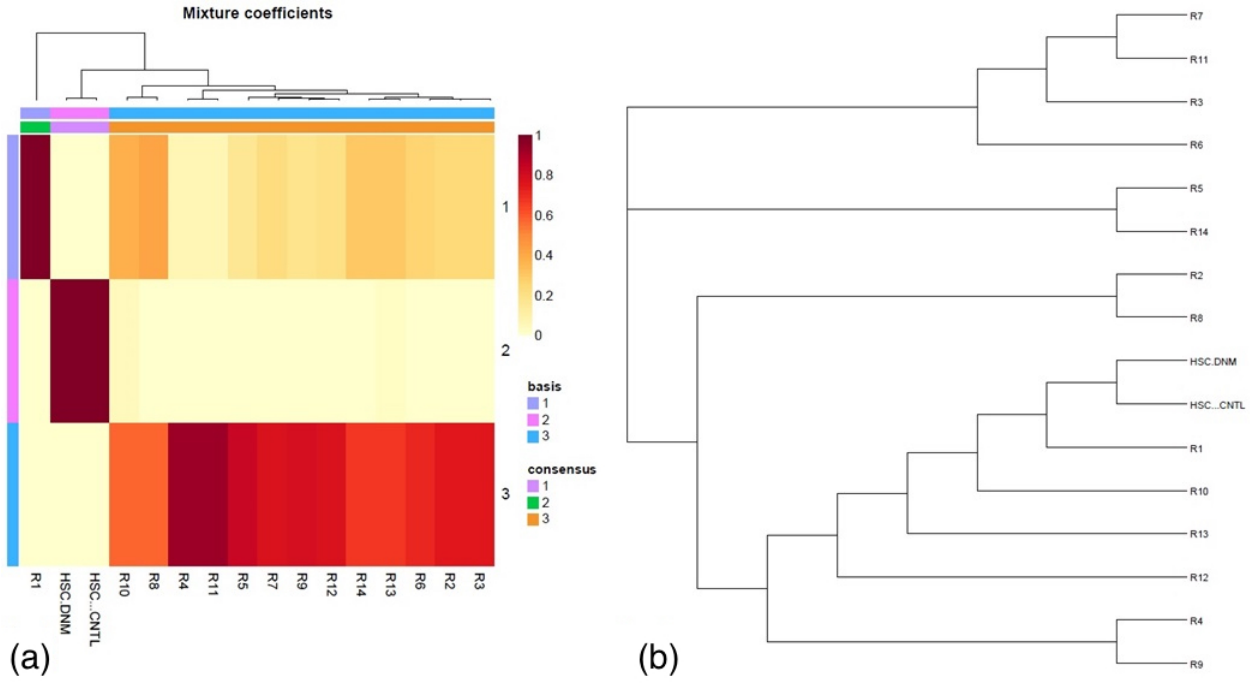


Figure 6.6: **Supervised clustering (classification) of DNMT3A samples.** Restricting the TCGA samples to a set of 59 samples with DNMT3A mutation classified in $R_1, \cdots, R_{11}$ subtypes along with normal and DNMT3A samples John Dick's laboratory using the same pipelines and methods described in the last two figures. (a) represents the heatmap of mixture coefficient method used to classify diverse subtypes based on the general mixture distribution model of 14 observation and 3 mixture component. Subfigure (b) depicts the supervised classification of subtypes R_i in compere with controlled and DNT3a subtypes.

and NPM1 mutations are some other potential candidates for the set of AML signature genes as they are located in the closer subtypes on the phylogenetic tree (see Figure 6.2). This analysis is required to be experimentally validated and statistically repeated through diverse investigation of not only a bigger gene expression database, but also on other type of data such as miRNAs, transcriptomes, and chromatins.

Similar analysis has performed for a restricted number of samples having a certain mutations such as DNMT3A. Our analysis shows that a hierarchy of mutational events assuming DNMT3A as the apex of mutations, would include WT1, FLT3, NPM1, and TP53 mutations as well. These mutations which are located in the R_{14} cluster at the

very close location on the phylogenetic tree (based on Euclidean bootstrap and FastME consensus method for R_is). This attempt is one of much research have been done to understand the regulatory mechanism of blood cells and the hierarchical mutational events leading to AML. Despite many developments in this regard, yet there is a big lack of a standard hierarchy of genetic and epigenetic changes tending to AML and also to an intense therapeutic treatment. Such a hierarchy or set of parallel hierarchies will, in addition, reveal the accurate relationships between mutations and genotypic–phenotypic events, and integrating leukemia initiation and progression. The next crucial step is about estimating the contribution of pre-leukemia vs. leukemia to relapse. Another crucial aspect relates to understand the correlation between clonal conversion and microenvironmental effect on the chance of pre–leukemic/leukemic cells to develop leukemia. This would assist us to illustrate the origin of AML and the mechanism(s) in corrolation with the transformation of pre–leukemic/leukemic cells to relapse.

Chapter 7

Conclusion and future work

In this Chapter, we briefly summarize the results of previous chapters and then suggest possible future work related to each of the aforementioned topics in Chapter 3 to Chapter 6. These, I believe, might be not only of great general interest, but also may provide novel insights for further endeavours to extend and/or apply these results in evolutionary dynamics, population genetics, and oncology.

In Chapter 3, starting with a 4-compartmental model of stem and non-stem cells, we asked about the most influential factors in the progression of tumor through comparing asymmetric differentiation of stem cells and phenotypic plasticity of plastic differentiated cells. After deriving a general analytic framework, which was in a perfect agreement with stochastic simulation, we found that the asymmetric differentiation has a more significant effect than plasticity on the fate of mutants under high selection pressure. The change in the rate of different types of division would increase the survivability of non-plastic mutants, but it would suppress the development of malignancy for plastic mutants. As the introduced mechanism and analytic approach are completely general, both can be used for similar mechanisms in other disciplines (e.g. in evolution, population genetics, and ecology). One can also add the symmetric differentiation to the system and use the Wright-Fisher model to study a similar mechanism to compare not only the asymmetric differentiation and plasticity, but also symmetric and asymmetric divisions in the presence of dedifferentiation. Another suggestion for future work relates to a generalization of the model to higher compartments of diverse phenotypes to have a more accurate estimation of natural mechanisms.

Then in Chapter 4, we defined and analyzed some aspects of microenvironmental effect on a structured population. The effect was basically investigated for the change in the

variance and higher moments of fitness of individuals proportional to their locations and genotypes. We found that even in a neutral system, when an environmental fluctuation impacts on both normal and mutant individuals, such an effect is in favor of mutants as the fluctuation increases. This interesting and non-trivial observation is fundamentally related to the minority power of mutants when a small portion of malignant mutations are imposed on the system. For the sake of simplicity, we considered 1D spatial structures, such as the circle, the line and complete graph structures, and allowed that the fitness of both genotypes be drawn from the same distribution. Our result suggests that minority power will also turn the disadvantageous mutants into advantageous mutants. We also found that the largest average fixation probability is associated to the anti-correlated fitnesses of mutants and resident individuals. For future study, a similar analytic investigation and/or stochastic simulation for 2D structures with reflecting boundaries would be of interest, where not only the mean, but also other moments of the random fitness are taken into account. A comparison among 1D and 2D structures would indicate how randomness is dependent on the dimension of the population and how the randomness effect may vary from lower dimensions to higher dimensions.

Chapter 5 presented a 4-compartmental model of the crypt where two various compartments of stem cells, central and border stem cells, are taken into account. In such a complex structure, we considered the migration of central stem cells towards border stem cells and since the reverse migration is reported to be negligible [144], we ignored the converse migration. We have also provided a space for the production of immortal cells. Then carrying out the required analytic calculations and stochastic simulation, we studied the role of each compartment in the initiation and progression of cancer within the crypt. We found that central stem cells may have a crucial role in the regulatory mechanism of the whole crypt. We also showed that a newborn mutant in central stem cell population has a high chance to take over the entire crypt. Conversely, using stem cell transplantation to replace all the central stem cells with normal stem cells will eventually cure the entire crypt, even if it is thoroughly occupied by mutants. Based on our research, another interesting question is about adding plasticity into the system, where plastic progenitor cells can dedifferentiate to either central or border stem cells. Then comparison between these two migration models would reveal the role of central and border stem cells in the progression of cancer in the presence of phenotypic plasticicty.

In Chapter 6, as opposed to the previous models that were basically concentrated on the stochastic study of the evolutionary dynamics of mutants, we performed a statistical analysis on genetic data. We analyzed the gene expression data of AML in the form of RNA-seq and used diverse pipelines to ignore redundant genes and construct a hierarchy of mutations. To derive the signature genes of AML, we took advantage of defining a new list

of AML subtypes and applied different unsupervised and supervised clustering methods to understand the correlation of the new subtypes. Our statistical analysis suggests DNMT3A and TET2 mutations are early events and exist in the signature genes of AML, which is in agreement with other existing experimental data. From unsupervised clustering we found an immediate cluster of PML samples, which also confirms the short evolution of this type of mutation. This can also be seen from the supervised clustering where the immediate subsequent cluster is associated to the NUP98 mutation. Our analysis revealed that after PML cluster, FLT3 mutation was found among the majority of samples in the subsequent cluster, showing that other than other rare mutations, those samples which are close to the root eventually experienced this type of mutation. Thus this type of mutation is almost certainly among the signature genes within the lineage or parallel lineages of sequential mutants in the pathogenesis of leukemia. Also having FLT3 in the subsequent cluster shows that important mutations such as DNMT3A, TET2, WT1, FLT3, PHF6, PTPs, MLLs, RUNX1, TP53, and NPM1 mutations are potential candidates as the signature genes of AML.

To further add to the results of Chapter 6, having a bigger cohort of samples may assist in deriving more clusters through unsupervised clustering and may provide a more accurate relationship among various subtypes within the supervised clustering (classification), repeating the general algorithm described in this chapter. Moreover, data analysis of exomes, transcriptomes, proteins, and chromatines as well as flow cytometry data analysis as complimentary sources for the RNA-seq data would describe other features of mutational events in AML. This would lead to a better understanding of founder stem cells or cancer initiating cells and demonstrate the cell of origin in leukemia.

References

[1] J.L. Abkowitz, S.N. Catlin and P. Guttorp, *Evience thathematopoiesis may be a stochastic process in vivo*, Nat. Med. **2** (1996), pp. 190–197.

[2] T. Antal, S. Redner and V. Sood, *Evolutionary Dynamics on Degree-Heterogeneous Graphs*, Phys. Rev. Lett. **96** (2006): 188104 .

[3] D. Arber et al., *The 2016 revision to the World Health Organization (WHO) classification of myeloid neoplasms and acute leukemia*, Blood (2016): blood-2016.

[4] P. Armitage, R. Doll, *The age distribution of cancer and a multi-stage theory of carcinogenesis*, British. J. Cancer **8**(1) (1954), pp. 1–12.

[5] J. Bajaj, Z. Bryan and R. Tannishtha, *Fearful symmetry: subversion of asymmetric division in cancer development and progression*, Cancer Res. *75*(5) (2015), pp. 792–797.

[6] A.-M. Baker et al., *Quantification of Crypt and Stem Cell Evolution in the Normal and Neoplastic Human Colon*, Cell Reports **8**(4) (2014), pp. 940–947.

[7] S. Bao et al., *Glioma stem cells promote radioresistance by preferential activation of the DNA damage response*, Nature. **444**(7120) (2006), pp. 756–760.

[8] B.E. Bernstein, A. Meissner, E.S. Lander, S. Eric, *The Mammalian Epigenome*, Cell **128**(4) (2007), pp. 669–681.

[9] T. Borovski, E.M. Felipe De Sousa EM, L. Vermeulen and J.P. Medema, *Cancer stem cell niche: the place to be*, Cancer Res. **71**(3) (2011), pp. 634–639.

[10] I. Bozic and M.A. Nowak, *Unwanted Evolution*, Science: Perspectives-Cancer **342**(6161) (2013), pp. 938–939.

[11] R. Bravo and D.E. Axelrod, *A calibrated agent-based computer model of stochastic cell dynamics in normal human colon crypts useful for in silico experiments*, Theo. Biol. Med. Model. **10**(66) (2013), pp. 1–24.

[12] M. Broom, J. Rycht and B. Stadler, *Evolutionary dynamics on graphs - the effect of graph structure and initial placement on mutant spread*, JMB **5** (2011), pp.369–381.

[13] P. Buske et al., *A Comprehensive Model of the Spatio-Temporal Stem Cell and Tissue Organisation in the Intestinal Crypt*, PLoS Comp. Biol. **7**(1): e1001045 (2011).

[14] M.C. Cabrera, *Cancer stem cell plasticity and tumor hierarchy*, World J. Stem Cells **7**(1) (2015), pp. 27–36.

[15] H.S. Callahan, M. Pigliucci and C.D. Schlichting, *Developmental phenotypic plasticity: where ecology and evolution meet molecular biology*, Bioessays **19**(6) (1997), pp. 519–525.

[16] C.L. Chaffer et al., *Normal and neoplastic nonstem cells can spontaneously convert to a stem-like state*, PNAS **108**(19) (2011), pp. 7950–7955.

[17] C.L. Chaffer et al., *Poised chromatin at the ZEB1 promoter enables breast cancer cell plasticity and enhances tumorigenicity*, Cell **154**(1) (2013), pp. 61–74.

[18] Cancer Genome Atlas Research Network, *Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia*, New Engl. J. Med. 2013.368 (2013). pp. 2059–2074.

[19] Y. Cao, D.T. Gillespie and L.R. Petzold, *Efficient step size selection for the tau-leaping simulation method*, J. Chem. Phys. **124**: 044109 (2006).

[20] L. Cassimeris, G. Plopper and V. Lingappa, *Lewin's CELLS*, Jones & Bartlett Publ., 2011.

[21] C.P. Chiu and C.B. Harley, *Replicative Senescence and Cell Immortality: The Role of Telomeres and Telomerase*, Expe. Biol. Med. **214**(2), pp. 99–106.

[22] E. Clayton et al., *A single type of progenitor cell maintains normal epidermis*, Nature **446**(7132) (2007), pp. 185–189.

[23] R.S. Datta, A. Gutteridge, C. Swanton, C.C. Maley and T.A. Graham, *modeling the evolution of genetic*, Evol. Appl. **6** (2013), pp. 20–33.

[24] J.A. de Visser, *The fate of microbial mutators*, Microbiology, **148**(5) (2002), pp. 1247-52.

[25] H.L. Del Puerto et al., *Canine distemper virus induces apoptosis in cervical tumor derived cell lines*, Virology J. **8**(334) (2011).

[26] A. Dhawan et al., *Tumour control probability in cancer stem cells hypothesis*, PloS one **9**(5) (2014): e96093.

[27] J.E. Dick, *Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell*, Nature Med **3** (1997), pp. 730–737.

[28] G.P. Dimri et al., *The Bmi-1 oncogene induces telomerase activity and immortalizes human mammary epithelial cells*, Cancer Res. **62**(16) (2002), pp. 4736–4745.

[29] N.J. Drize, J.R. Keller and J.L.Chertkov, *Local clonal analysis of the hematopoietic system shows that multiple small short-living clones maintain life-long hematopoiesis in reconstituted mice*, Blood **88**(8) (1996), pp. 2927–38..

[30] E. Dorantes-Acosta and P. Rosana, *Lineage switching in acute leukemias: a consequence of stem cell plasticity?*, Bone marrow research 2012 (2012).

[31] R. Durrett, J. Foo., and K. Leder, *Spatial Moran Models II. Tumor growth and progression*, preprint: http:// math.umn.edu/j̃yfoo/PAPERS/tau2DFL_final.pdf.

[32] M. El Zoghbi and L.C. Cummings, *New era of colorectal cancer screening*, World J. Gastrointestinal Endoscopy **8**(5) (2016), pp. 252–25.

[33] S.N. Ethier and M.F. Norman, *Error estimate for the diffusion approximation of the Wright–Fisher model*, PNAS **74**: 5096 (1977).

[34] W.J. Ewens, *Mathematical Population Genetics*, Springer-Verlag, Berlin, Heidelberg, 1979.

[35] N. Fausto, *Liver regeneration*, J. Hepatology. **32** (2000), pp. 19–31.

[36] E.F. Fearon and B. Vogelstein, *A Genetic Model for Colorectal Tumorigenesis*, Cell *61*(5) (1990), pp. 759–767.

[37] C.M. Fillmore and C. Kuperwasser, *Human breast cancer cell lines contain stem-like cells that self-renew, give rise to phenotypically diverse progeny and survive chemotherapy*, Breast Cancer Res. **10**(2) (2008), pp. 1–13.

[38] R. A. Fisher, *On the dominance ratio*, Proc. R. Soc. Edin. **50** (1922), pp. 204–219.

[39] A.G. Fletcher, Ch.J.W. Berward, S.J. Chapman, *Mathematical modeling of monoconal conversion in the colonic crypt*, J. Theor. Bio. **300** (2012), pp. 118-133.

[40] W. Fu, J.G. Begley, M.W. Killen and M.P. Mattson, *Anti-apoptotic Role of Telomerase in Pheochromocytoma Cells*, J. Biol. Chem. **274**(11), pp. 7264–7271.

[41] C. Gardiner, *Stochastic Methods: A Handbook for the Natural and Social Sciences*, Springer, 4th Ed., 2009.

[42] X. Gao et al., *Acute and fractionated irradiation differentially modulate glioma stem cell division kinetics*, Cancer research **73**(5) (2013), pp. 1481–1490.

[43] C.K. Ghalambor, J.K. McKay, S.P. Carroll and D.N. Reznick , *Adaptive versus non-adaptive phenotypic plasticity and the potential for contemporary adaptation in new environments*, Func. Ecol. **21**(3) (2007), pp. 394–407.

[44] J.H. Gillespie. *Population Genetics: a concise guide*, JHU Press (2010).

[45] C. Ginestier et al., *ALDH1 is a marker of normal and malignant human mammary stem cells and a predictor of poor clinical outcome*, Cell Stem Cell **1**(5) (2007), pp. 555–567.

[46] N. Goardon et al., *Coexistence of LMPP-like and GMP-like leukemia stem cells in acute myeloid leukemia*, Cancer cell **19**(1) (2011), pp. 138–152.

[47] J. Grime, J. Crick and J. Rincon, *The ecological significance of plasticity*, Symposia Soc. Exper. Biol. **40** (1985), pp. 5–29.

[48] P.B. Gupta, C.M. Fillmore, G. Jiang, S.D. Shapira, K. Tao, C. Kuperwasser, E.S. Lander, *Stochastic state transitions give rise to phenotypic equilibrium in populations of cancer cells*, Cell **146** (2011), pp. 633–644.

[49] J. B. S. Haldane, *The mathematical theory of natural and artificial selection*, Proc. Camb. Philos. Soc. **23** (1927), pp. 838–844 .

[50] D. Hanahan and L.M. Coussens, *Accessories to the crime: fucntions of cells recruited to the tumor microenvironment*, Cancer Cell **21** (2012), pp. 309-322.

[51] D. Hanahan and R.A. Weinberg, *Hallmarks of cancer: the next generation*, Cell **144** (2011), pp. 646–674.

[52] Ch. Hauert, *Evolutonary dynamics*, Springer-verlag, 2007.

[53] L. Hindersin and A. Traulsen, *Counterintuitive properties of the fixation time in network-structured populations*, J. R. Soc. Interface **11** (2014): 20140606.

[54] B. Houchmandzadeh and M. Vallade, *Alternative to the diffusion equation in population genetics*, Phys. Rev. E **82**: 051913 (2010).

[55] B. Houchmandzadeh and M. Vallade, *The fixation probability of a beneficial mutation in a geographically structured population*, New J. Phys. **13**: 3020 (2011).

[56] B. Houchmandzadeh and M. Vallade, *The fixation probability of a beneficial mutation in a geographically structured population*, New J. Phys. **13**: 3020 (2011).

[57] D.J. Huels and O.J. Sansom, *Stem vs non-stem cell origin of colorectal cancer*, British J. Cancer **113** (2015), pp. 1–5.

[58] A. Humphries and N.A. Wright, *Colonic crypt organization and tumorigenesis*, Nature Review 8 (2008), pp. 415-424.

[59] R. Ivkov and F. Bunz, *Pathways to chromothripsis*, Cell Cycle **14**(18) (2015), pp. 2886–2890 .

[60] K. Ito et al., *A PML-PPAR-[delta] pathway for fatty acid oxidation regulates hematopoietic stem cell maintenance*, Nature Med. **18**(9) (2012), pp. 1350–1358.

[61] S.M. Janes, S. Lowell and C. Hutter, *Epidermal stem–cell*, J. Pathol. **197** (2002), pp. 479–491.

[62] A. Jilkine and R.N. Gutenkunst, *Effect of dedifferentiation on time to mutation acquisition in stem cell-driven cancers*, PLoS Comp. Biol. **10**(3) (2014): e1003481.

[63] M. D. Johnston, C. M. Edwards, W.F. Bodmer, Ph. K. Maini and S.J. Chapman, *Mathematical modeling of cell population dynamics in the colonic crypt and in colorectal cancer*, PNAS **104**(10) (2007), pp. 4008–4013.

[64] Y. Kagawa, N. Horita, H. Taniguchi and S. Tsuneda, *Modeling of stem cell dynamics in human colonic crypts in silico*, J. Gastroenterology **49**(2) (2014), pp. 263–269.

[65] M. Karin and F.R. Greten, *NF-κB: linking inflammation and immunity to cancer development and progression*, Nat. Rev. Immun. **5**(10) (2005), pp. 749–759.

[66] K. Kaveh, M. Kohandel and S. Sivaloganathan, *Replicator dynamics of cancer stem cell: Selection in the presence of differentiation and plasticity*, Math. Biosci. **272** (2016), pp. 64–75.

[67] K. Kaveh, N.L. Komarova and M. Kohandel, *The duality of spatial death-birth and birth-death processes and limitations of the isothermal theorem*, R. Soc. Open Sci. **2**(4): 140465 (2015).

[68] H.E. Kay, *How many cell-generations? (Hypothesis)*, Lancet **1** (1965), pp. 418–419.

[69] J.P. Keener, *The Perron-Frobenius Theorem and the Ranking of Football Teams*, SIAM Rev. **35**(1) (1993), pp. 80–93.

[70] F.J.A. Khalek , G.I. Gallicano and L. Mishra, *Colon cancer stem cells*, Gastrointestinal Cancer Res. (Suppl 1):S16 (2010).

[71] A. Kilian et al., *Isolation of a Candidate Human Telomerase Catalytic Subunit Gene, Which Reveals Complex Splicing Patterns in Different Cell Types*, Human Molecular Gen. **6**(12), pp. 2011–2019.

[72] M. Kimura, *Evolutionary rate at the molecular level*, Nature **217** (1968), pp. 624–626.

[73] M. Kimura, *On the probability of fixation of mutant genes in a population*, Genetics **47** (1962), 713–719.

[74] M. Kimura, *Solution of a Process of Random Genetic Drift with a Continuous Model*, PNAS **41**(3) (1955), pp. 144–150.

[75] N.L. Komarova, *Spatial Stochastic models for cancer Initiation and Progression*, Bull. Math. Biol. **68** (2006), pp. 1573–1599.

[76] L. Komarova, *Loss- and Gain-of-Function Mutations in Cancer: Mass-action, Spatial and Hierarchical Models*, J. Stat. Phys. **128** (2007), pp. 413–446.

[77] N.L. Komarova, L. Shahriyari and D. Wodarz, *Complex role of space in the crossing of fitness valleys by asexual populations*, J. Royal Soc. Interface **11**(5): 20140014 (2014).

[78] K. W. Kinzler and B. Vogelstein, *Lessons from hereditary colorectal cancer*, Cell **87** (1996), pp. 159–170.

[79] J.A. Knoblich, *Mechanisms of asymmetric stem cell division*, Cell **132**(4) (2008), pp. 583–597.

[80] A. Kreso et al., *Self-renewal as a therapeutic target in human colorectal cancer* Nature Med. **20**(1) (2014), pp. 29–36.

[81] A. Kreso and J.E. Dick, *Evolution of the cancer stem cell model*, Cell Stem Cell **14**(3) (2014), pp. 275–291.

[82] A.V. Krivtsov et al., *Transformation from committed progenitor to leukaemia stem cell initiated by MLL?AF9*, Nature *442*(7104) (2006), pp. 818–822.

[83] J. Krnke et al., *Clonal evolution in relapsed NPM1-mutated acute myeloid leukemia*, Blood **122**(1) (2013), pp. 100–108.

[84] N. Kuranaga, N. Shinomiya and H. Mochizuki, *Long-term cultivation of colorectal carcinoma cells with anti-cancer drugs induces drug resistance and telomere elongation: an in vitro study*,BMC Cancer **1**(10) (2001).

[85] T.J, Ley et al, *Genomic and Epigenomic Landscape of Adult De Novo Acute Myeloid Leukemia*, NEJM **368**(22) (2013), pp. 2059–2074.

[86] S.A. Lamprecht and M. Lipkin, *Migrating colonic crypt epithelial cells: primary targets for transformation*, Carcinogenesis **23**(11) (2002), pp. 1777–80 .

[87] R.C. Langan et al., *Colorectal cancer biomarkers and the potential role of cancer stem cells*, J. Cancer **4**(3) (2013), pp. 241–250.

[88] C. Legraverend and P. Jay, *Hierarchy and plasticity in the crypt: back to the drawing board*, Cell Res. **21**(12) (2011), pp. 652–1654.

[89] F. Lemonnier, Satoshi Inoue and T.W. Mak, *Genomic Classification in Acute Myeloid Leukemia*, N. Engl. J. Med. **375** (2016), pp. 900–900.

[90] E. Lieberman, Ch. Hauert, and M.A. Nowak, *Evolutionary dynamics on graphs*, Nature **433**(7023), (2005), pp. 312–316.

[91] Y. Li and J. Laterra, *Cancer stem cells: distinct entities or dynamically regulated phenotypes?* Cancer Res. **72**(3) (2012), pp. 576–580.

[92] X. Liu et al., *Nonlinear Growth Kinetics of Breast Cancer Stem Cells: Implications for Cancer Stem Cell Targeted Therapy*, Sci. Reports **3**(2473) (2013), pp. 1–9.

[93] B. D. MacArthur and I.R. Lemischka, *Statistical mechanics of pluripotency*, Cell **154**(3) (2013), pp. 484–489.

[94] J.A. Magee, E. Piskounova and S.J. Morrison, *Cancer stem cells: impact, heterogeneity, and uncertainty*, Cancer Cell **21**(3) (2012), pp. 283–296.

[95] A. Mahdipour Shirayeh, K. Kaveh, M. Kohandel and S. Sivaloganathan, *Phenotypic heterogeneity in modeling cancer evolution*, preprint: arXiv:1610.08163 (2016)

[96] A. Mahdipour Shirayeh, A.H. Darooneh, N.L. Komarova and M. Kohandel, *Genotype by random environment interactions gives an advantage to neutral minor alleles*, submitted (2016).

[97] A. Mahdipour Shirayeh and L. Shahriyari, *New Insights into Initiation of Colon and Intestinal Cancer: The Significance of Central Stem Cells in the Crypt*, preprint: arXiv:1610.04089 (2016).

[98] T. Makino et al., *Primary signet-ring cell carcinoma of the colon and rectum: report of eight cases and review of 154 Japanese cases*, Hepato-gastroenterology **53**(72) (2006), pp. 845–849.

[99] V.S.K. Manem, K. Kaveh, M. Kohandel and S. Sivaloganathan, *Modeling Invasion Dynamics with Spatial Random-Fitness Due to Micro-Environment*, PloS One **10**(10) (2015), pp.1–20.

[100] S.A. Mani et al, *The epithelial-mesenchymal transition generates cells with properties of stem cells*, Cell **133**(4) (2008), pp. 704–715.

[101] N.D. Marjanovic, R.A. Weinberg and Ch. L. Chaffer, *Cell plasticity and heterogeneity in cancer*, Clin. Chem. **59**(1) (2013), pp. 168-179.

[102] E.A. Martens, R. Kostadinov, C.C. Maely, and O. Halltschek, *Spatial structure increases the waiting time for cancer*. New J. Phys. **13**: 115014 (2011).

[103] T. Maruyama, *A Markov process of genes frequency change in a geographically structured population*, Genetics **76**(2) (1974), pp. 367–377.

[104] T. Maruyama, *A simple proof that certain quantities are independent of the geographical structure of population*, Theor. Popul. Biol. **5** (1974), pp. 148–154.

[105] N. Masuda, N. Gibert and S. Render, *Heterogeneous voter models*, Phys Rev. E **82**: 010103(R) (2010).

[106] J. Matis and T. Wehrly, *17 Compartmental models of ecological and environmental systems*, Handbook of Statistics **12** (1994), pp. 583–613.

[107] S.A. McDonald et al., *Mechanisms of field cancerization in the human stomach: the expansion and spread of mutated gastric stem cells*, Gastroenterology **134**(2) (2008), pp. 500–510 .

[108] J.L. McKenzie, O.I. Gan, M. Doedens, J.C.Y. Wang and J.E. Dick, *Individual stem cells with highly variable proliferation and self-renewal properties comprise the human hematopoietic stem cell compartment*, Nat. Immun. **7** (2006) , pp. 1225–1233.

[109] F. Michor, *Patterns of Proliferative Activity in the Colonic Crypt Determine Crypt Stability and Rates of Somatic Evolution*, PLoS Comp. Biol. **9** (6) (2013), e1003082.

[110] F. Michor, Y. Iwasa, and M. A. Nowak, *Dynamics of Cancer Progression*, Nat. Rev. Cancer **4** (2004), pp. 197–205.

[111] G.R. Mirams, A.G. Fletcher, P.K. Maini, H.M. Byrne, *A theoretical investigation of the effect of proliferation and adhesion on monoclonal conversion in the colonic crypt*, J. Theo. Biol. **312** (2012), pp. 143–156.

[112] A. Mitrofanova, *Absorbing states in Markov chains. Mean time to absorption. Wright-Fisher Model. Moran Model*, New York Uni., Department of Computer Science, http://cs.nyu.edu/mishra/COURSES/09.HPGP/scribe2 (2007).

[113] W. Mittelman and J.H. Wilson, *The fractured genome of HeLa cells*, Genome Biol. **14**(111) (2013), pp. 1–4.

[114] M. Mohle, *Ancestral processes in population genetics-the coalescent*, J. Theor. Biol. **204** (2000), pp. 629-38.

[115] S.H. Moolgavkar, *The multistage theory of carcinogenesis and the age distribution of cancer in man*, J. Natl. Cancer Inst. **61**(1) (1978), pp. 49–52.

[116] P. Moran, *The Statistical Processes of Evolutionary Theory*, Oxford: Clarendon (1962).

[117] S.J. Morrison and J. Kimble, *Asymmetric and symmetric stem-cell divisions in development and cancer.*, Nature. **441**(7097) (2006), pp. 1068–1074.

[118] K. Mrózek et al., *Prognostic significance of the European LeukemiaNet standardized system for reporting cytogenetic and molecular alterations in adults with acute myeloid leukemia*, Journal of clinical oncology **30**(36) (2012), pp. 4515–4523.

[119] T.G. Natarajan and K.T. FitzGerald, *Markers in normal and cancer stem cells*, Cancer Biomarkers, **3**(4, 5) (2007), pp. 211–231.

[120] C.O. Nordling, *A new theory on cancer-inducing mechanism*, British J. Cancer 7(1) (1983), pp. 68–72.

[121] M. A. Nowak, *Evolutionary dynamics*, Harvard University Press (2006).

[122] M.A. Nowak, N.L. Komarova, A. Sengupta, P.V. Jallepalli, I.-M. Shih, B. Vogelstein, and Ch. Lengauer, *The role of Chromosomal instability in tumor initiation*, PNAS **99**(25), (2002), pp. 16226–16231.

[123] M.A. Nowak, F. Michor and Y. Iwasa, *The linear process of somatic evolution*, PNAS **100**(25) (2003), pp. 14966–14969.

[124] M. Nowak, F. Michor, N.L. Komarova, and Y. Iwasa, *Evolutionary dynamics of tumor suppressor gene inactivation*, Proc. National Acad. Sci. **101**(29) (2004), pp.10635–10638.

[125] R.A. Neumüller and J.A. Knoblich, *Dividing cellular asymmetry: asymmetric cell division and its implications for stem cells and cancer* Genes & development **23**(23) (2009), pp. 2675–2699.

[126] L.V. Nguyen, R. Vanner, P. Dirks and C.J. Eaves, *Cancer stem cells: an evolving concept*, Nature Rev. Cancer **12**(2) (2012), pp. 133–143.

[127] C.A. O'Brien, A. Kreso and J.E. Dick, *Cancer stem cells in solid tumors: an overview*, Seminars Rad. Onco. **19**. Elsevier; 2009, pp. 71–77.

[128] H. Ohtsuki, C. Hauert, E. Lieberman and M. A. Nowak, *A simple rule for the evolution of cooperation on graphs and social networks*, Nature **441** (2006), pp. 502-505.

[129] K.P. Olive et al., *Mutant p53 gain of function in two mouse models of Li-Fraumeni syndrome*, Cell **119** (2004), pp.847–860.

[130] E. Papaemmanuil et al., *Genomic Classification and Prognosis in Acute Myeloid Leukemia*, New Engl. J. Med. **374**(23) (2016), pp. 2209–2221.

[131] J.P. Patel et al., *Prognostic relevance of integrated genetic profiling in acute myeloid leukemia*, NEJM **366**(12) (2012), pp. 1079-1089.

[132] Z. Patwa and L. M. Wahl, *The fixation probability of beneficial mutations*, J. R. Soc. Interface **5** (2008), 1279-89.

[133] J.W. Pepper, K. Sprouffske and C.C. Maley, *Animal cell differentiation patterns suppress somatic evolution*, PLoS Comp. Biol. **3**(12) (2007), pp.2532-2545.

[134] D.A. Pollyea et al., *Targeting acute myeloid leukemia stem cells: a review and principles for the development of clinical trials*, haematologica **99**(8) (2014), pp. 1277–1284.

[135] C.S. Potten, M. Kellett, D.A. Rew and S.A. Roberts, *Proliferation in human gastrointestinal epithelium using bromodeoxyuridine in vivo: data for different sites, proximity to a tumour, and polyposis coli*, Gut **33**(4) (1992), pp. 524–529.

[136] C.S. Potten, M. Kellett, S.A. Roberts, D.A. Rew and G.D. Wilson, *Measurement of in vivo proliferation in human colorectal mucosa using bromodeoxyuridine*, Gut **33**(1) (1992), pp. 71–78.

[137] A. Philpott and D.J. Winton, *Lineage selection and plasticity in the intestinal crypt*, Current Opinion Cell Biol. **31** (2014), pp. 39–45.

[138] C. Pin, A.J.M. Watson and S.R. Carding, *Modelling the Spatio-Temporal Cell Dynamics Reveals Novel Insights on Cell Differentiation and Proliferation in the Small Intestinal Crypt*, PLoS ONE **7**(5): e37115 (2012).

[139] F. Prall, *Tumour budding in colorectal carcinoma*. Histopathology **50** (2007), pp.151–162.

[140] T. Reya and H. Clevers, *Wnt signalling in stem cells and cancer*, Nature **434** (2005), pp. 843-850

[141] T. Reya, S.J. Morrison, M.F. Clarke and I.L. Weissman, *Stem cells, cancer, and cancer stem cells*, Nature. **414**(6859) (2001), pp. 105–111.

[142] M. Riester , C.S.-O. Attolini , R.J. Downey, S. Singer and F. Michor, *A Differentiation-Based Phylogeny of Cancer Subtypes*, PLoS Comp. Biol. **6**5 (2010): e1000777.

[143] E.A. Rietman et al., *Personalized anticancer therapy selection using molecular landscape topology and thermodynamics*, Oncotarget (2016).

[144] L. Ritsma et al., *Intestinal crypt homeostasis revealed at single-stem-cell level by in vivo live imaging*, Nature **507** (2014), pp. 362–365.

[145] L. Riva, L. Luzi and P.G. Pelicci, *Genomics of acute myeloid leukemia: the next generation*, Frontiers Oncol. 2(40) (2012), pp. 1–12.

[146] I.A. Rodriguez-Brenes, N.L. Komarova and D. Wodarz, *Evolutionary dynamics of feedback escape and the development of stem-cell-driven cancers*, Proceedings of the National Academy of Sciences **108**(47) (2011), pp. 18983–18988.

[147] A. Roesch et al, *A temporarily distinct subpopulation of slow-cycling melanoma cells is required for continuous tumor growth*, Cell 141 (2010), pp.583–94.

[148] A.F. Salem et al. , *Two-compartment tumor metabolism: autophagy in the tumor microenvironment and oxidative mitochondrial metabolism (OXPHOS) in cancer cells*, Cell cycle **11**(13) (2012), pp. 2545–2559.

[149] O.J. Sansom et al., *Loss of Apc in vivo immediately perturbs Wnt signaling, differentiation, and migration*, Genes Devel. **18**(12) (2004), pp. 1385–1390.

[150] J.E. Sarry et al. *Human acute myelogenous leukemia stem cells are rare and heterogeneous when assayed in NOD/SCID/IL2Rgammac-deficient mice*, J. Clin. Invest. **121** (2011), pp. 384–395.

[151] T. Schoofs et al., *DNA methylation changes are a late event in acute promyelocytic leukemia and coincide with loss of transcription factor binding*, Blood **121**(1) (2013), pp. 178–187.

[152] M. Schmidt et al. *Polyclonal long-term repopulating stem cell clones in a primate model*, Blood **100**(8) (2002), pp. 2737–2743.

[153] S. Schwitalla et al., *Intestinal tumorigenesis initiated by dedifferentiation and acquisition of stem-cell-like properties*, Cell **152**(1) (2013), pp. 25–38.

[154] L. Shahriyari and N.L. Komarova, *Role of the bi-compartmental stem cell niche in delaying cancer*, Phys. Biol. **12**(5): 055001 (2015).

[155] L. Shahriyari and N.L. Komarova, *Symmetric vs. Asymmetric Stem Cell Divisions: An adaptation against Cancer?*, PLoS ONE **8**(10) (21013), pp. 1–16.

[156] L. Shahriyari L, A. Mahdipour-Shirayeh, *Optimal structure of heterogeneous stem cell niche: The importance of cell migration in delaying tumorigenesis*, preprint: bioRxiv 082982 (2016).

[157] Y. Shen et al, *Gene mutation patterns and their prognostic impact in a cohort of 1185 patients with acute myeloid leukemia*, Blood **118**(20) (2011), pp. 5593-5603.

[158] L. Shlush and A. Mitchell, *AML evolution from preleukemia to leukemia and relapse*, Best Practice & Research Clinical Haematology **28**(2) (2015), pp. 81–89.

[159] L. Shlush et al, *Identification of pre-leukaemic haematopoietic stem cells in acute leukaemia*, Nature **506**(7488) (2014), pp.328-333.

[160] Ya. G. Sinai, *The Limiting Behavior of a One-Dimensional Random Walk in a Random Medium*, Theo. Prob. Appl. **27** (1982): 256 .

[161] S.K. Singh et al., *Identification of a cancer stem cell in human brain tumors*, Cancer Res. **63**(18) (2003), pp. 5821–5828.

[162] H.J. Snippert, L.G. Flier and T. Sato, *Intestinal crypt homeostasis results from neutral competition between symmetrically dividing Lgr5 stem cells*, Cell **143**(1) (2010), pp. 134–144.

[163] V. Sood, T. Antal and S. Render, *Voter models on heterogeneous networks*, Phys. Rev. E Stat. Nonl. Soft Matter Phys. **77**: 041121 (2008).

[164] S. Spaderna et al, *A transient, EMT-linked loss of basement membranes indicates metastasis and poor survival in colorectal cancer*, Gastroenterology **131** (2006), pp. 830–40.

[165] K. Sprouffske et al., *An evolutionary explanation for the presence of cancer nonstem cells in neoplasms*, Evol. Appl. **6**(1) (2013), pp. 92–101.

[166] T. Strachan, A.P. Read, *Human Molecular Genetics 2*. Wiley, New York, 1999.

[167] F. Solomon, *Random walks in a random environment*, Annal. Prob. **3** (1975), pp. 1–31 .

[168] V. Sood, T. Antal and S. Redner, *Voter models on heterogeneous networks*, Phys. Rev. E **77** (2008): 041121.

[169] S.E. Sultan, *Phenotypic plasticity in plants: a case study in ecological development*, Evol. Devel. **5**(1) (2003), pp. 25–33.

[170] Q. Y. Sunet al., *Ordering of mutations in acute myeloid leukemia with partial tandem duplication of MLL (MLL-PTD)*, Leukemia Adv. **8** (2016), doi: 10.1038/leu.2016.160.

[171] D.G. Tang, *Understanding cancer stem cell heterogeneity and plasticity*, Cell Res. **22**(3) (2012), pp. 457–472 .

[172] P.R. Tata et al., *Dedifferentiation of committed epithelial cells into stem cells in vivo*, Nature **503**(7475) (2013), pp. 218–223.

[173] A.E. Teschendorff, P. Sollich and R. Kuehn, *Signalling entropy: A novel network-theoretical framework for systems analysis and interpretation of functional omic data*, Methods **67**(3) (2014), pp. 282–293.

[174] J.P. Thiery, H. Acloque, R.Y. Huang, M.A. Nieto, *Epithelial-mesenchymal transitions in development and disease*, Cell **139** (2009), pp. 871–890.

[175] C.J. Thalhauser, J.S. Lowengrub, D. Stupack and N.L. Komarova, *Selection in spatial stochastic models of cancer: Migration as a key modulator of fitness*, Biol. Direct **5**(21) (2010), pp. 1–17.

[176] C. Tomasetti and D. Levy, *Role of symmetric and asymmetric division of stem cells in developing drug resistance*, PNAS **107**(39) (2010), pp. 16766–16771.

[177] A. Traulsen and Ch. Hauert, *Stichastic evolutionary game dynamics, Reviews of Nonlinear Dynamics and Complexity*, Vol. II, Wiley-VCH, 2009.

[178] C. Turner et al., *Characterization of brain cancer stem cells: a mathematical approach* Cell proliferation **42**(4) (2009), pp. 529–540.

[179] N.G. Van Kampen, *The diffusion approximation for Markov processes*, De Gruyter (1982).

[180] N.G. Van Kampen, *The validity of nonlinear Langevin equations*, J. Statistical Physics **25**(3) (1981), pp. 431–442.

[181] L. Vermeulen et al., *Defining stem cell dynamics in models of intestinal tumor initiation*, Science **342**(6161) (2013), pp. 995-8.

[182] L. Vermeulen and J. Snippert, *Stem cell dynamics in homeostasis and cancer of the intestine*, Nature Reviews Cancer **14**(7) (2014), pp.468–480..

[183] L. Vermeulen et al., *Wnt activity defines colon cancer stem cells and is regulated by the microenvironment*, Nature cell biol. **12**(5) (2010), pp. 468–476.

[184] H. Vogelsang and J.R. Siewert, *Endocrine tumours of the hindgut. Best practice & research*, Clin. Gastroenterology **19**(5) (2005), 739–751.

[185] b. Vogelstein, K.W. Kinzler, *The Genetic Basis of Human Cancer*. McGraw-Hill, New York, 1997.

[186] M. Vogelstein, K.W. Kinzler, *Cancer genes and the pathways they control*, Nature Med. **10**(8) (2004), pp. 789–99.

[187] B. Vogelstein et al., *Cancer genome landscapes*,"science **339**(6127) (2013), pp. 1546–1558.

[188] Y. Wang et al., *The Wnt/?-catenin pathway is required for the development of leukemia stem cells in AML*, Science **327**(5973) (2010), pp. 1650–1653.

[189] S.L. Weekes et al., *multicompartment mathematical model of cancer stem cell-driven tumor growth dynamics*, Bull. Math. Biol. **76**(7) (2014), pp. 1762–1782.

[190] R.A. Weinberg, The biology of cancer, Garland Science, Taylor & Francis Publ., 2007.

[191] J.S. Welch et al, *The origin and evolution of mutations in acute myeloid leukemia*, Cell **150** (2012), pp. 264–278.

[192] B. Werner, D. David and A. Traulsen, *A deterministic model for the occurrence and dynamics of multiple mutations in hierarchically organized tissues*, J. R. Soc. Interface **10**(85) (2013): 20130349.

[193] B. Werner, D. Dingli, T. Lenaerts, J.M. Pacheco and A. Traulsen, *Dynamics of mutant cells in hierarchical organized tissues*, PLoS Comp. Biol. **7**(12):e1002290 (2011).

[194] J. Westman, B. Fabijonas, D. Kern, F. Hanson, *Compartmental model for cancer evolution: chemotherapy and drug resistance*, Math. Biosci. (submitted) (2001).

[195] R.H. Whitehead, P.E. VanEeden, M.D. Noble, P. Ataliotis, and P.S. Jat, *Establishment of conditionally immortalized epithelial cell lines from both colon and small intestine of adult H-2Kb-tsA58 transgenic mice*, Proc. National Acad. Sci. **90**(2), pp. 587–591.

[196] W.A. Woodward and E.P. Sulman, *Cancer stem cells: markers or biomarkers?* Cancer Metastasis Res. **27**(3) (2008), pp. 459–470.

[197] D. Wodarz and N. L. Komarova, *Dynamics of cancer: mathematical foundations of oncology*, World Scientific (2014).

[198] S. Wright, *Evolution in Mendelian populations*, Genetics **16** (1931), pp. 97–159.

[199] S. Wright, *The Differential Equation of the Distribution of Gene Frequencies*, PNAS **31**(12) (1945), pp. 382–389.

[200] Y. Yatabe, S. Tavare and D. Shibata, *Investigating stem cells in human colon by using methylation patterns*, PNAS **98**(19) (2001), pp. 10839–10844.

[201] C.C. Yu et al., *Bmi-1 regulates snail expression and promotes metastasis ability in head and neck squamous cancer-derived ALDH1 positive cells*, J. Oncology (2011), pii: 609259.

[202] S. Zapperi, C.A and La Porta, *Do cancer cells undergo phenotypic switching? The case for imperfect cancer stem cell markers*, Sci. Reports **2**(441) (2012), pp. 1–7.

[203] C.-Z. Zhang et al., *Chromothripsis from DNA damage in micronuclei*, Nature **522**(7555) (2015), pp. 179–184.

[204] R. Zhao and F. Michor, *Patterns of Proliferative Activity in the Colonic Crypt Determine Crypt Stability and Rates of Somatic Evolution*, PLoS CB **9**(6) (2013):e1003082.

[205] Zh. Zhao et al., *p53 loss promotes acute myeloid leukemia by enabling aberrant self-renewal*, Genes & Devel. **24**(13) (2010), pp. 1389–1402.

[206] B. Zimdahl et al., *Lis1 regulates asymmetric division in hematopoietic stem cells and in leukemia*, Nature Gen. **46**(3) (2014), pp. 245–252.

APPENDICES

Appendix A

Randomness Effects

A.1 Fixation probability distribution

We are investigating about the average treatment of the system as we have applied in the preceding sections of the paper for the circle model and will perform later for the line and complete graph structures. Prior to these investigations about the fixation probability of the structured populations, we need to understand the distribution of the fixation probability when an average behavior is taken into account. The main implement to capture the average behavior of the system can be the central limit theorem [41]. For instance to analyze the variation of the fixation probability on average when the mean and variance are not changing through a trial in which the proliferation rates are random variables of a normal distribution, the central limit theorem predicts another normal distribution with the same mean and a proportional variance. This will also predict that the derived distribution of average behavior is not a composition of two or more distributions and as the mean of the distribution remains fixed, the variance is also bounded with certain value.

In Figure A.1, a unique probability distribution for the average fixation probability can be detected compared to the individual events in the fixation process of a newborn mutant through all possible scenarios. Different histograms represent the existence of a skewness towards left (or towards lower values than the average). For a considered population size as σ grows, the fixation probability values shift to the smaller values. Moreover, when one increases N, the probability of distribution slides to the lower values. An interesting result is that the standard deviation of the fixation probability remains bounded as N raises to larger values.

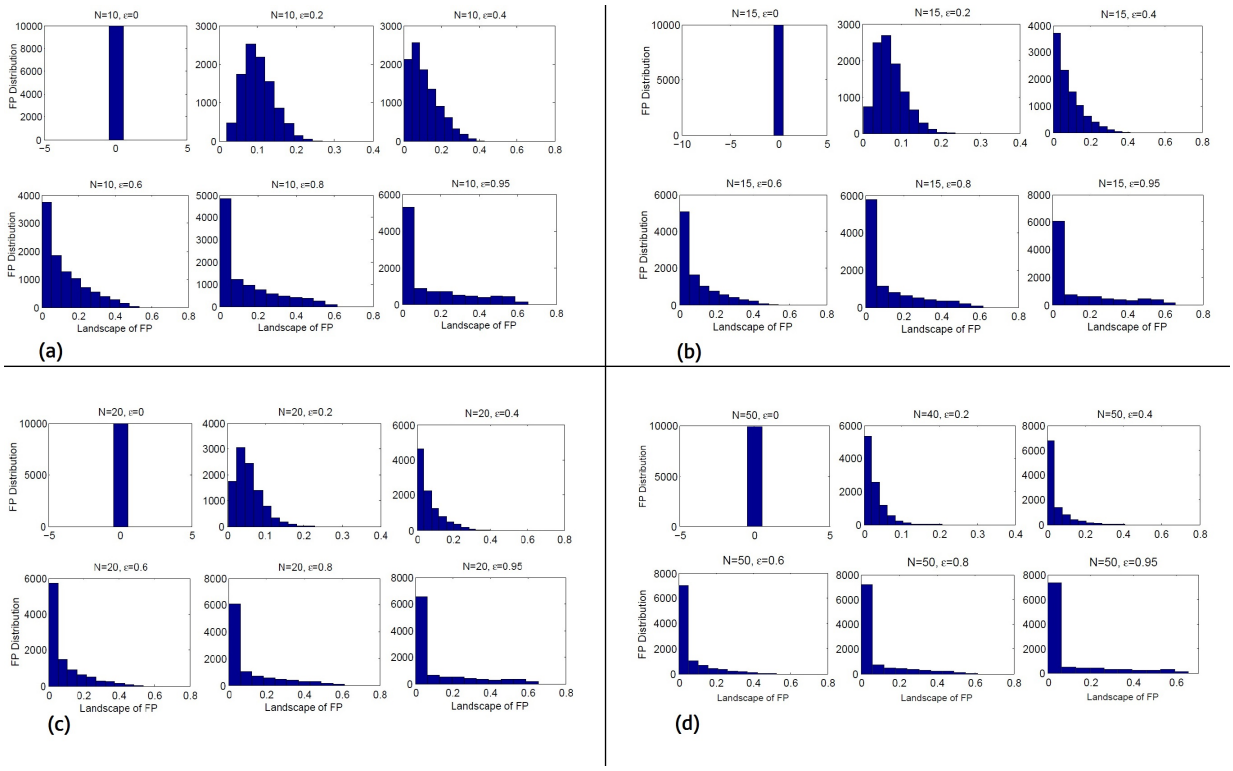


Figure A.1: **Distribution of the fixation probability shows skewness to the left hand side which as population size increases shifts towards left.** Distribution of the fixation probability for different populations are given in (a) for $N = 10$, in (b) for $N = 15$, in (c) for $N = 20$, and finally in (d) for $N = 50$. In all of the considered cases we assumed that $\langle r_b \rangle = \langle r_a \rangle = 1, \sigma_b = \sigma_a = \sigma$.

Overall this analysis confirm the fact that the central limit theorem can be applied to analytically envisage the treatment of the mean filed approximation of the considered simple structures in the present study.

Another interesting approach as is given in Figure A.2 relates to trends for the mean and variance of the average fixation probability of the circle. Based on these graphs, both mean and variance of the average fixation probability increases as the variance σ of the bimodal distribution of fitness increases. However, the both values can rise up to a limit and as a function of σ and population size N they are bounded for any given σ and N. In this figure, as the population size grows, the rate of change for the variance of the average fixation probability is larger than that for the mean and its corresponding graph passes through the graph of mean sooner.

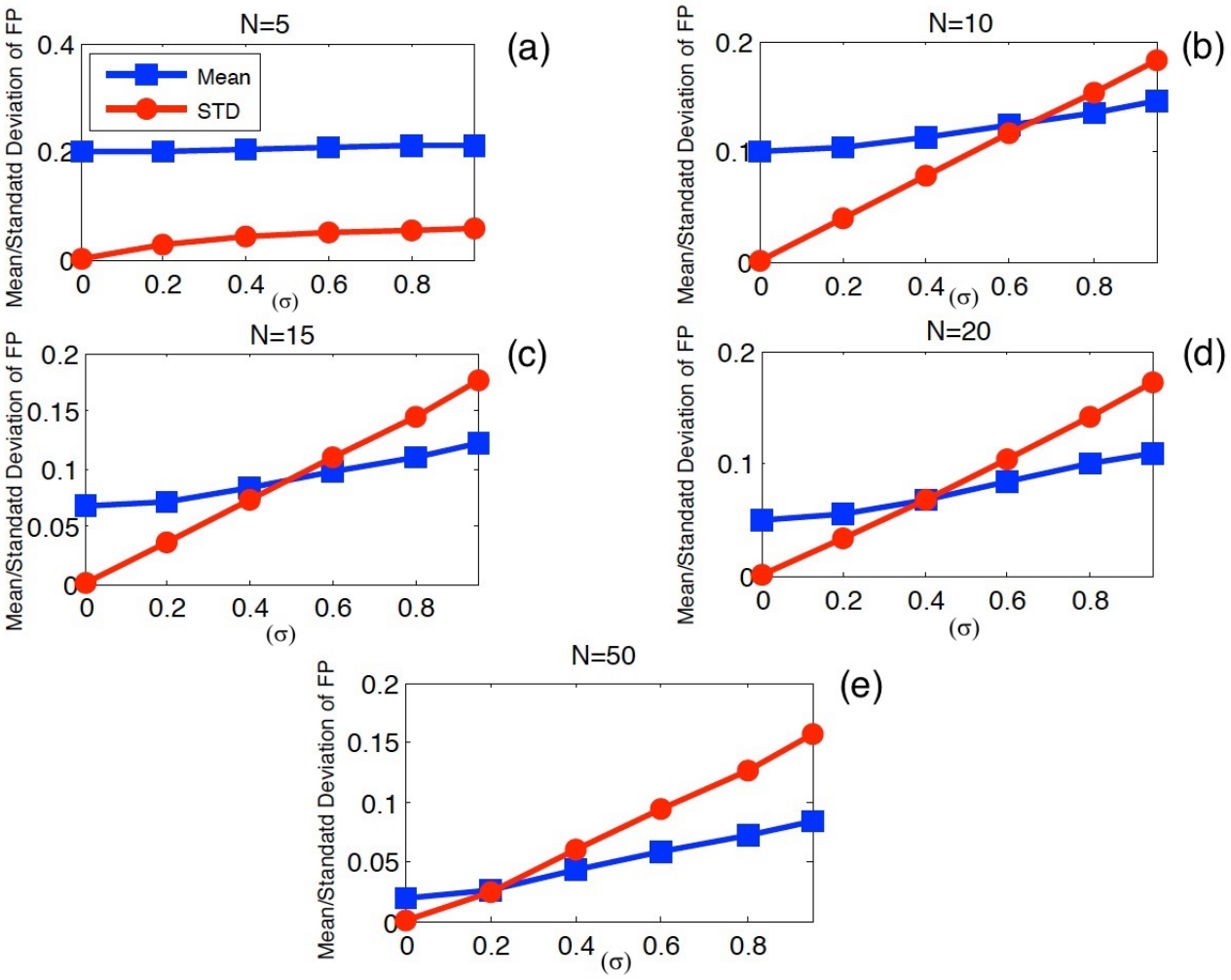
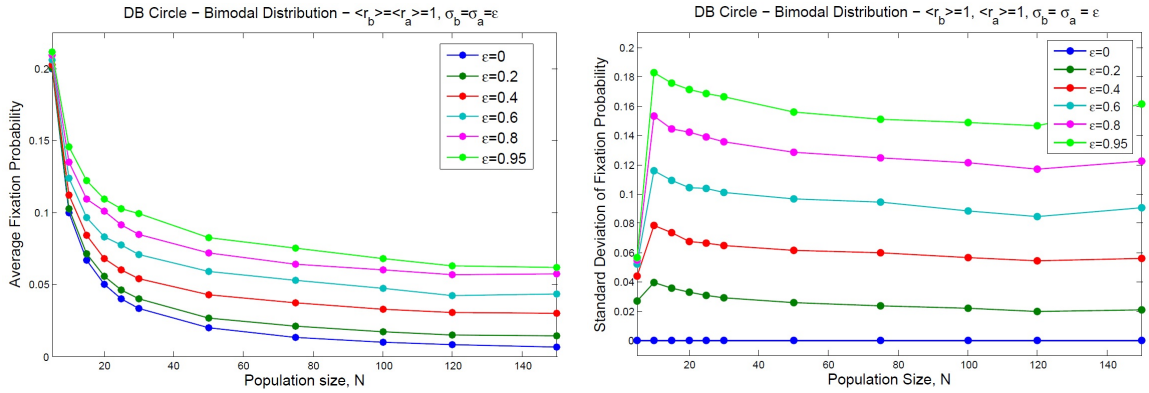


Figure A.2: **The relationship between differences in mean and variance of the fixation probability with respect to σ.** Assuming $\langle r_b \rangle = \langle r_a \rangle = 1, \sigma_b = \sigma_a = \sigma$, considering the probability distribution of the fixation probability in the neutral case, the mean and variance are approximately linearly changing in terms of σ. As can be seen in different figures, as population size increases, as the line for standard deviation remains less than 0.2 with some slight changes, the curve for the mean declines to intersect the line of means in lower values for σ. In (a), (b), (c), (d), and (e) respectively $N = 5, 10, 15, 20,$ and 50.

A.2 Trends for the mean and variance of the average fixation probability: Randomness for both types.

In this section, returning to the Section 4.5, we represent the main features of the change in the mean and variance of the average fixation probability for small, intermediate, and large populations. The main goal is to capture the long-term behavior of the average fixation probability for each given variance σ of the fitness which defines the randomness for both normal and mutant individuals. Complementary to what we explained in Section 4.5, we discuss about the mean and variance of the average fixation probability in more details as follows.

At first, when $\langle r_b \rangle > \langle r_a \rangle$ there is a monotonic decrease in the value of average fixation probability as population increase. The trend is sharper when σ is larger (see Figure A.1-(a) for more details). Moreover, the tendency for the standard deviation of the average probability of fixation declares an upper bounded in each graph. This confirms the result of the previous section about the existence of an upper bound for the variance. Figure A.1-(b) represents the trend for a variety of σs.



(a) (b)

Figure A.3: **The trends for the mean and standard deviation of the fixation probability when $\langle r_b \rangle = \langle r_a \rangle = 1, \sigma_b = \sigma_a = \sigma$.** In part (a), mean of the fixation probability is drawn in terms of population size. The trend for various values of σ show the same decreasing behaviors. In part (b), a similar result can be detected for intermediate and large population sizes. These figures confirms the fact that mean is decreasing and the variance is bounded for all values of σ.

The second scenario relates to the case in which $\langle r_b \rangle < \langle r_a \rangle$. The trends for various σ

signify a dramatic decrease for a range of different population sizes. More importantly, by increasing the σ there exist an upward shift for intermediate population sizes which starts to shift downward for large enough σs (this value for σ depends on the average fitness of the mutants. In Figure A.4-(a) the green curve that is associated to $\sigma = 0.9$ represents the lowest average fixation probabilities compared with the corresponding values for smaller σs. So the overall mechanism tends to the deterioration of the average fixation probability for large populations.

Similarly, the standard deviation of the mean of the survival chance of a recently born mutant declines and approaches to zero. As σ climbs to a certain level, then the trend sharply drops to converge to zero even for average populations.

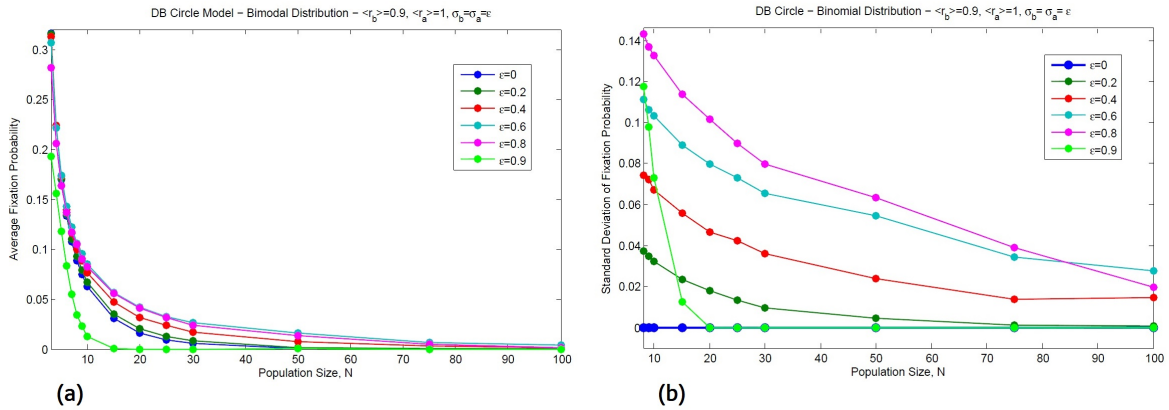


Figure A.4: **The trends for the mean and standard deviation of the fixation probability when** $\langle r_b \rangle = 0.9, \langle r_a \rangle = 1, \sigma_b = \sigma_a = \sigma$**.** In part (a), mean of the fixation probability dramatically decreases as population size raises. In (b), the trend for smaller values σ increases compared to smaller σs but for $\sigma = 0.9$ represents a huge decline as we conclude in the analytic results.

Finally when the average fitness of both mutant and host cells the same $\langle r_b \rangle = \langle r_a \rangle$ which can be simply normalized to 1, an intermediate behavior between the advantageous and disadvantageous average fitness of cancer cells can be detected. Although the trend for the standard deviation in Figure A.5-(b) shows almost constant trends for diverse values of σ as population size increases, the progression for the average survival chance reveals an slight but monotonic decrease over a range from small population size to intermediate and then to large population size (see A.5-(a) for a more details).

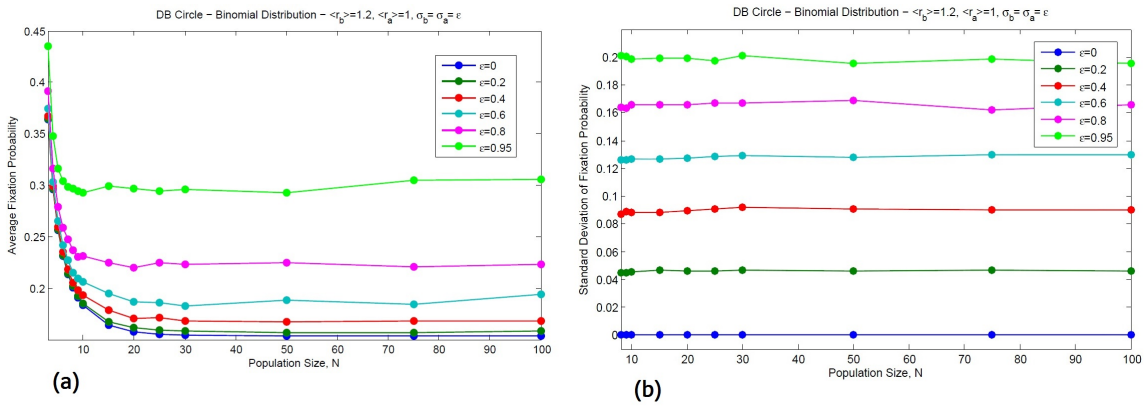


Figure A.5: **The trends for the mean and standard deviation of the fixation probability when** $\langle r_b \rangle = 1.2, \langle r_a \rangle = 1, \sigma_b = \sigma_a = \sigma$**.** In (a), mean of the fixation probability for larger populations will saturate to a certain amount which depends on σ. In (b), standard deviation of intermediate populations and larger populations stays unchanged on a certain value which is again only dependent to σ.

A.3 Analytic results for randomness effect on large-scale populations

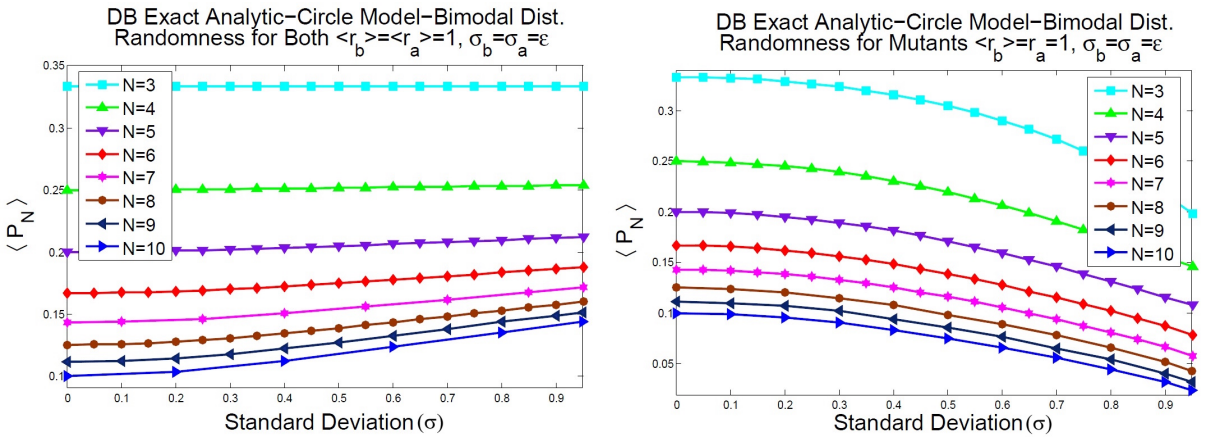
Starting from the smallest population size $N = 4$ on the circle model, in this section, we let both populations of normal and mutant individuals take random fitness under the influence of micro–environment when the associated stochastic procedure is based on a bimodal distribution with peaks r_1, r_2. Considering the nodes of this 1D lattice, at each vertex of the graph, both mutants and host cells can take either the value r_1 or r_2 without any correlation. In this case, each of the three nodes may take a value from the set $\{r_1, r_2\}$. Now having a new mutant at a specific point, there exist three possible location for the new mutant. Denoting the possible cases by a set of binary sequences $l_1 l_2 l_3 l_4$ for $l_i \in \{0, 1\}$ where $l_i = 1$ is associated to have mutant at location i on the circle and $l_i = 0$ for a normal cell. Considering different scenarios, based on the location and population size of mutants, the transition probabilities of increase and decrease by one define the dynamic conditions of progression of mutant cells. In this situation, one obtains the following system along with its corresponding initial conditions and when the reproduction rates for cancer cells

are assumed to be $\tilde{a}, \tilde{b}, \tilde{c}, tilded$ while those of normal cells are denoted by a, b, c, d:

$$
\begin{aligned}
4P_{1000} &= P_{1100}\left(\frac{\tilde{a}}{\tilde{a}+c}\right) + P_{1001}\left(\frac{\tilde{a}}{\tilde{a}+c}\right) + P_{1000}\left(1+\frac{2c}{\tilde{a}+c}\right),\\
4P_{0100} &= P_{0110}\left(\frac{\tilde{b}}{\tilde{b}+d}\right) + P_{1100}\left(\frac{\tilde{b}}{\tilde{b}+d}\right) + P_{0100}\left(1+\frac{2d}{\tilde{b}+d}\right),\\
4P_{0010} &= P_{0011}\left(\frac{\tilde{c}}{\tilde{c}+a}\right) + P_{0110}\left(\frac{\tilde{c}}{\tilde{c}+a}\right) + P_{0010}\left(1+\frac{2a}{\tilde{c}+a}\right),\\
4P_{0001} &= P_{1001}\left(\frac{\tilde{d}}{\tilde{d}+b}\right) + P_{0011}\left(\frac{\tilde{d}}{\tilde{d}+b}\right) + P_{0001}\left(1+\frac{2b}{\tilde{d}+b}\right),\\
4P_{1100} &= P_{1110}\left(\frac{\tilde{b}}{\tilde{b}+d}\right) + P_{1101}\left(\frac{\tilde{a}}{\tilde{a}+c}\right) + P_{1000}\left(\frac{c}{\tilde{a}+c}\right) + P_{0100}\left(\frac{d}{\tilde{b}+d}\right) + 2\,P_{1100},\\
4P_{1001} &= P_{1101}\left(\frac{\tilde{a}}{\tilde{a}+c}\right) + P_{1011}\left(\frac{\tilde{d}}{\tilde{d}+b}\right) + P_{1000}\left(\frac{c}{\tilde{a}+c}\right) + P_{0001}\left(\frac{b}{\tilde{d}+b}\right) + 2\,P_{1001},\\
4P_{0110} &= P_{0111}\left(\frac{\tilde{c}}{\tilde{c}+a}\right) + P_{1110}\left(\frac{\tilde{b}}{\tilde{b}+d}\right) + P_{0100}\left(\frac{d}{\tilde{b}+d}\right) + P_{0010}\left(\frac{a}{\tilde{c}+a}\right) + 2\,P_{0110},\\
4P_{0011} &= P_{1011}\left(\frac{\tilde{d}}{\tilde{d}+b}\right) + P_{0111}\left(\frac{\tilde{c}}{\tilde{c}+a}\right) + P_{0010}\left(\frac{a}{\tilde{c}+a}\right) + P_{0001}\left(\frac{b}{\tilde{d}+b}\right) + 2\,P_{0011},\\
4P_{1110} &= 1 + P_{1100}\left(\frac{d}{\tilde{b}+d}\right) + P_{0110}\left(\frac{d}{\tilde{b}+d}\right) + P_{1110}\left(1+\frac{2\tilde{b}}{\tilde{b}+d}\right),\\
4P_{1101} &= 1 + P_{1100}\left(\frac{c}{\tilde{a}+c}\right) + P_{1001}\left(\frac{c}{\tilde{a}+c}\right) + P_{1101}\left(1+\frac{2\tilde{a}}{\tilde{a}+c}\right),\\
4P_{1011} &= 1 + P_{1001}\left(\frac{b}{\tilde{d}+b}\right) + P_{0011}\left(\frac{b}{\tilde{d}+b}\right) + P_{1011}\left(1+\frac{2\tilde{d}}{\tilde{d}+b}\right),\\
4P_{0111} &= 1 + P_{0110}\left(\frac{a}{\tilde{c}+a}\right) + P_{0011}\left(\frac{a}{\tilde{c}+a}\right) + P_{0111}\left(1+\frac{2\tilde{c}}{\tilde{c}+a}\right).
\end{aligned}
\tag{A.1}
$$

Let us now consider a circle with $N(\geq 3)$ individuals of either cancerous or wild-type. Labeling the graph with numbers 1 to N, one may dedicate the fitnesses $\tilde{a}_1, \cdots, \tilde{a}_N$ to mutant cells and $a_1, \cdots, a_N$ for normal cells depending on the type of the individual located

on a specific node. Denoting the probability of fixation by P_σ which is the probability of starting from state $\sigma = (\alpha_1, \cdots, \alpha_N)$ and ending up at absorption state $(1, 1, \cdots, 1)$ where $\alpha_i \in \{0, 1\}$ with $1 \leq i \leq N$. When $\alpha_i = 0$, it means that a normal individual is located at node i, otherwise $(\alpha_i = 1)$ that is a mutant.



(a) (b)

Figure A.6: **Exact analytic calculations show increasing trends for the fixation probability for universal randomness and decreasing when only mutants are under effect of randomness.** Let the population size vary and is equal to $N = 3, 4, 5, 6, 7, 8, 10$ and a bimodal distribution imposes the randomness on the system. In (a) cancer and wild-type cells have random fitnesses r_b, r_a with average $\langle r_b \rangle = \langle r_a \rangle = 1$ and variance $\sigma_b = \sigma_a = \sigma$ respectively. In (b) the exact analytic calculations reveals a decreasing behavior of the fixation probability on a circle when only mutants are under the effect of randomness. More precisely, cancer cells have relative random fitness r_b with average $\langle r_b \rangle = 1$ and variance $\sigma_b = \sigma$ whilst normal cells have a fixed fitness $r_a = 1$ This result analytically confirms the result of [99].

Moreover, suppose that σ_1 (σ_2) be the element-wise maximum of σ with its permutation by one to the right (left) and $\tilde{\sigma}_1$ $(\tilde{\sigma}_2)$ be the element–wise minimum of σ with its permutation by one to the left (right). Furthermore, assume that $\sigma_{\max} = \max\{\, i \,|\, \sigma_i = 1\}$ and $\sigma_{\min} = \min\{\, i \,|\, \sigma_i = 1\}$, and $\mathcal{M}_1 = \max\{\|\sigma\| - 2, 0\}, \mathcal{M}_2 = \max\{N - \|\sigma\| - 2, 0\}$ with respect to Euclidean norm $\|\,.\,\|$. Then the forward Kolmogorov equations includes $N(N-1)$ equations of $N(N-1)$ variables and can be written in the following form (when

$0 < \|\sigma\| < N$ and all the indices are considered modulo N)

$$\begin{cases} N\,P_\sigma = P_{\sigma_1}\,[\mathcal{A}_1] + P_{\sigma_2}\,[\mathcal{A}_2] + P_\sigma\,\,[N-3+\mathcal{C}_2]\,, & \|\sigma\| = 1 \\ N\,P_\sigma = P_{\sigma_1}[\mathcal{A}_1] + P_{\sigma_2}[\mathcal{A}_2] + P_{\tilde{\sigma}_1}[\mathcal{B}_1] + P_{\tilde{\sigma}_2}[\mathcal{B}_2] \\ \qquad\qquad +P_\sigma\,[\mathcal{M}_1 + \mathcal{M}_2 + \mathcal{C}_1 + \mathcal{C}_2]\,, & 2 \le \|\sigma\| \le N-2 \\ N\,P_\sigma = 1 + P_{\tilde{\sigma}_1}\,[\mathcal{B}_1] + P_{\tilde{\sigma}_2}\,[\mathcal{B}_2] + P_\sigma\,\,[N-3+\mathcal{C}_1]\,, & \|\sigma\| = N-1, \end{cases} \tag{A.2}$$

where $\sigma_i \in \{0,1\}$ for $1 \le i \le N$ and transition probabilities by one from the state σ are

$$\begin{aligned} &\mathcal{A}_1(\sigma) = \frac{\tilde{a}_{\sigma_{max}}}{\tilde{a}_{\sigma_{max}} + a_{\sigma_{max}+2}}, \quad \mathcal{A}_2(\sigma) = \frac{\tilde{a}_{\sigma_{min}}}{\tilde{a}_{\sigma_{min}} + a_{\sigma_{min}-2}}, \quad \text{for} \;\; 1 \le \|\sigma\| < N-1, \\ &\mathcal{B}_1(\sigma) = \frac{a_{\sigma_{max}+1}}{a_{\sigma_{max}+1} + \tilde{a}_{\sigma_{max}-1}}, \quad \mathcal{B}_2(\sigma) = \frac{a_{\sigma_{min}-1}}{a_{\sigma_{min}-1} + \tilde{a}_{\sigma_{min}+1}} \quad \text{for} \;\; 1 < \|\sigma\| \le N-1, \end{aligned} \tag{A.3}$$

In the latter expressions, when $\|\sigma\| = N-1$ then $\mathcal{A}_1 = \mathcal{A}_2 = 1$ and for $\|\sigma\| = 1$ we can conclude that $\mathcal{B}_1 = \mathcal{B}_2 = 1$. Also we have

$$\begin{aligned} \mathcal{C}_1(\sigma) &= \mathrm{H}(\|\sigma\| - 3/2)\left(\frac{\tilde{a}_{\sigma_{max}-1}}{\tilde{a}_{\sigma_{max}-1} + a_{\sigma_{max}+1}} + \frac{\tilde{a}_{\sigma_{min}+1}}{\tilde{a}_{\sigma_{min}+1} + a_{\sigma_{min}-1}}\right), \\ \mathcal{C}_2(\sigma) &= \mathrm{H}(N - \|\sigma\| - 3/2)\left(\frac{a_{\sigma_{max}+1}}{a_{\sigma_{max}+1} + \tilde{a}_{\sigma_{max}-1}} + \frac{a_{\sigma_{min}-1}}{a_{\sigma_{min}-1} + \tilde{a}_{\sigma_{min}+1}}\right), \end{aligned} \tag{A.4}$$

where $H(\,.\,)$ is the Heavyside step function. Note that the initial conditions of the system are $P_{00\cdots0} = 0, P_{11\cdots1} = 1$.

Remark. A comparison among different populations is given in Figure A.6. These graphs emphasizes on the fact that the trend for the average fixation probability increases as the variance grows. For larger populations, the trend is even sharper and represents higher upward shifts compare to smaller population sizes. The reason of this increase in the average probability of fixation relates to two important factors: minority of mutant cells and majority of normal cells while both have random fitnesses. A simple analysis reveals that having the peak values of the binomial distribution, $\mu - \sigma$ has a stronger effect in reducing the fixation probability than $\mu + \sigma$ which makes an increase in the fixation probability. In this situation, where randomness effects both host and mutant cells, minority of mutants enforces a reduction in the fixation probability of mutants as the upper peak of the binomial distribution have a lower increasing effect to the lower peak of the binomial distribution for a given variance to decrease the fixation probability. This means that the reduction in the fixation probability of normal cells will happen more

often than the increase in the fixation probability of mutants and thus the average fixation probability shows an increasing treatment. The represents sharper effects for larger values of standard deviation. So overall, the effect of all different possible configurations on the average fixation probability, only in the viewpoint of mutants is decreasing. the same effect can be see in the viewpoint of normal cells when the same procedure will have a much higher impact on normal cells and in the favor of mutants and thus in the circumstance of the whole system while mutants are fighting against normal cells the average fixation probability is increasing. Moreover, the structure of the model would also fortify the conduction of mutants towards absorption.

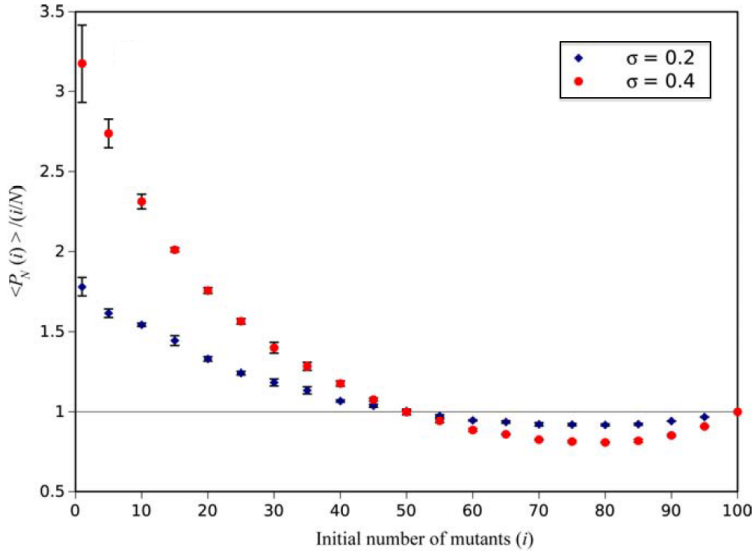
This achievement for a 1D structure is opposite to the result of [99] in a 2D structure where the trend for average fixation probability is thought to be decreasing as the variance increases.

A.4 Initial frequency dependency of the fixation probability

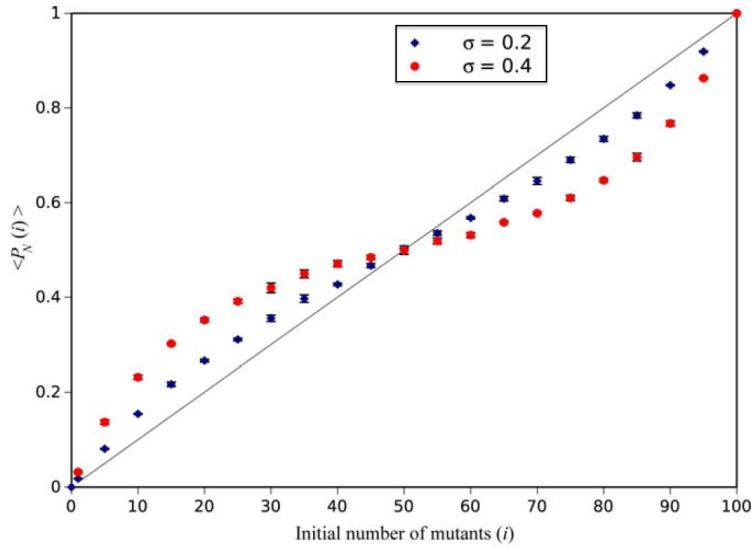
The small number of initial mutants which is, in fact, associated with the minority of malignancy has been observed in Chpater 4 to play a crucial role in the fate of this type of individual, where both host and cancer individuals are under random effect of the surrounding microenvironment.

Having an equal number of host and malignant individuals, tends to a neutral system in which the configuration of the system in terms of random fitness in different locations, tends to neutral drift. This means that although these types of cells are phenotypically different, the regeneration mechanism for both phenotypes is completely similar and the chance for any change in the size of either sub-populations remains the same. Therefore, the average behavior of the system follows Darwinian evolutionary mechanism.

However, increasing the number of initial mutants to have larger population of mutants compare with that of normal individuals, fluctuation will be in favor of normal individuals which comprises the minor population. Figure 3.6 depicts how randomness (for two different values of the standard deviation $\sigma = 0.2, 0.4$ would change the trend of the average fixation probability as the number of initial mutants alters. According to this figure, there exist a symmetric behavior for the fixation probability of the corresponding initial size of mutants i and $N - i$ compared with that of having $i = \frac{N}{2}$.



(a)



(b)

Figure A.7: **Dependency of the average fixation probability of mutants on the initial number of malignant mutations.** (a) The average fixation probability over the expected probability under neutrality ($\frac{i}{N}$), and (b) the average fixation probability, in terms of initial number of mutants (for the bimodal fitness distribution). In both subfigures, the total population size is assumed to be $N = 100$ and each each point represents the result of stochastic simulation for a given $1 \leq i \leq 100$ and error bars are the standard deviation of mean for a set of 3 iterations, each with 20,000 realization.

Appendix B

Evolutionary Mechanism of Colorectal/Intestinal Cancer

B.1 Analytic tools

We denote the number of wild-type (w.t.) and mutant stem cells in the S_b group respectively by b and b^*, in the S_c group by e and e^*, in the D_t group by d and d^*, and finally in the D_f compartment by f and f^*. In addition d^{**} and f^{**} are respectively the number of TA and FD immortal cells. We also assume that the fitness of mutant cells and immortal cells are respectively r_1 and r_2.

In this model at each updating time step, two D_f cells die. For each of these deaths, with probability $f^*/(f+f^*)$, one mutant D_f cell dies and with probability $f/(f+f^*)$ one wild-type (w.t.) D_f cell dies. Then, two randomly chosen cells divide according to the following algorithm:

- With a probability of λ_f, two D_f cells divide. For each of these divisions, with probability $\frac{r_1 f^*}{r_2 f^{**}+r_1 f^*+f}$, one mutant D_f cell divide and with probability v one of its children becomes immortal cell. However, with probabilities $\frac{r_2 f^{**}}{r_2 f^{**}+r_1 f^*+f}$ and $\frac{f}{r_2 f^{**}+r_1 f^*+f}$, respectively one immortal and one wild-type D_f cell divides. Or,
- With a probability of $(1-\lambda_f)\frac{r_2 d^{**}}{r_2 d^{**}+r_1 d^*+d}$ or $(1-\lambda_f)\frac{d}{r_2 d^{**}+r_1 d^*+d}$, respectively one immortal TA cell differentiates to two immortal FD cells or one normal TA cell differentiates to produce two w.t. D_f cells. However, with probability $(1-\lambda_f)\frac{r_1 d^*}{r_2 d^{**}+r_1 d^*+d}$,

one mutant TA cell divides and with probability u, one of its newborn daughter cells becomes an immortal FD cell, i.e. dedifferentiation happens, while the other offspring is a mutant TA cell. However, with probability $1-u$, both newborn individuals are mutant fully differentiated cells. Then one of below scenarios occurs

- With probability $(1-\lambda_s)$, one TA cell proliferates. This is the proliferation of a wild-type TA cell with a probability of $\frac{d}{r_2 d^{**}+r_1 d^{*}+d}$ to produce one w.t. TA cell, or the proliferation of immortal TA cell with a probability of $\frac{r_2 d^{**}}{r_2 d^{**}+r_1 d^{*}+d}$ to produce one immortal TA cell. Or with probability $\frac{r_1 d^{*}}{r_2 d^{**}+r_1 d^{*}+d}$, the proliferation of a mutant TA cell occurs, then with probability u one of the newborn members is immortal TA cell and the other one is mutant TA cell and with probability $1-u$ both are mutant TA cells. Or,
- with probability λ_s, one stem cell divides in the following way:
  - One mutant S_b stem cell divides asymmetrically and makes one mutant TA cell with probability $(1-\sigma)\frac{r_1 b^{*}}{r_1 b^{*}+b}$. With probability $(1-\sigma)\frac{b}{r_1 b^{*}+b}$, one w.t. S_b stem cell divides to generate one w.t. cell in the D_t compartment. Or,
  - with probability $\delta\frac{b}{r_1 b^{*}+b}$, one wild-type S_b stem cell differentiates to generate two wild-type TA cells. However, with probability $\delta\frac{r_1 b^{*}}{r_1 b^{*}+b}$ one mutant border stem cell differentiates to make two mutant TA cells. Or,
  - with probability $1-\delta$ proliferation happens in the stem cell niche.
    - With probability $\gamma\frac{e}{r_1 e^{*}+e}$, one normal S_c cell proliferates to produce one normal stem cell in the S_c group. Moreover, with probability $\gamma\frac{re^{*}}{r_1 e^{*}+e}$, one mutant stem S_c cell proliferates to generate one mutant S_c cell. Then with probability $\frac{e}{e+e^{*}}$ one random w.t. stem cell from S_c migrates to the S_b. In addition, with probability $\frac{e^{*}}{e+e^{*}}$ one mutant S_c cell migrates to the S_b compartment. Or,
    - with probability $(1-\gamma)\frac{b}{r_1 b^{*}+b}$, one wild-type S_b cell proliferates to make one wild-type stem cell in the S_b compartment. However, with probability $(1-\gamma)\frac{r_1 b^{*}}{r_1 b^{*}+b}$ one mutant S_b cell proliferates to generate another mutant S_b cell.

More precisely, σ is the probability of symmetric division in stem cell niche, while δ is the probability of differentiation in the S_b compartment in the case of symmetric stem cell division. Moreover, γ is the probability of proliferation in the S_c compartment, when a stem cell proliferates. Moreover, λ_f is the probability of choosing fully differentiated cells for birth event (see Figure 5.1), while λ_s is the probability of division in the stem cell niche. Figure B.1 and B.2 reveal different steps of the procedure in details.

B.2 Evolutionary dynamics of colorectal/Intestinal cancer

In this model, the total number of stem cells in the S_c and S_b compartments (separately) remain approximately fixed, which means that homeostasis controls each stem cell compartment's size. Moreover, the other two compartments of progenitor and fully differentiated cells are subject to the same assumption and their sizes remain approximately unchanged through the evolutionary dynamics of the system. Therefore, we have a 6-dimensional multi–variable Markov model as the system of random movements over possible states $(e^*, b^*, d^*, d^{**}, f^*, f^{**})$.

We denote the probability of moving from the state a to the state b in one time step by $P_{a\to b}$, where $a, b \in \{(e^*, b^*, d^*, d^{**}, f^*, f^{**}\}$. For simplicity, indexes a and b only includes the parameter(s), which are changing. For example, the probability $P_{e^*\to e^*+1}$ is the probability of moving from the state, which has e^* number of S_c mutants, to the state that has e^*+1 number of S_c mutants in one time step, while the number of the other mutants $(b^*, d^*, d^{**}, f^*, f^{**})$ has not changed. All possible non-zero transition probabilities are listed as follows.

B.2.1 Transition Probabilities

(1) $P_{f^*\to f^*+1} = \left(\frac{f}{f+f^*}\right)^2 \left\{2\,\lambda_f\,\left(\frac{f}{\mathcal{F}}\right)\,\left(\frac{r_1 f^*}{\mathcal{F}}\right)\,(1-v)\right\} + \frac{2ff^*}{(f+f^*)^2}\,\left\{\lambda_f\,\left(\frac{r_1 f^*}{\mathcal{F}}\,(1-v)\right)^2 + (1-\lambda_f)\,\frac{r_1 d^*}{\mathcal{D}}\,(1-u)\,\left[(1-\lambda_s)\,\frac{r_1 d^*}{\mathcal{D}}\,(1-u) + \lambda_s\,(1-\sigma)\,\frac{r_1 b^*}{\mathcal{R}_b}\right]\right\},$

(2) $P_{f^*\to f^*-1} = \frac{2ff^*}{(f+f^*)^2}\,\left\{\lambda_f\,\left(\frac{f}{\mathcal{F}}\right)^2 + (1-\lambda_f)\,\frac{d}{\mathcal{D}}\left[(1-\lambda_s)\,\frac{d}{\mathcal{D}} + \lambda_s\,(1-\sigma)\,\frac{b}{\mathcal{R}_b} + \lambda_s\,\sigma\left(\delta\frac{b}{\mathcal{R}_b} + (1-\delta)(1-\gamma)\frac{b}{\mathcal{R}_b} + (1-\delta)\gamma\,\frac{e}{\mathcal{R}_2}\frac{e}{e+e^*}\right)\right]\right\} + \left(\frac{f^*}{f+f^*}\right)^2\,\left\{2\,\lambda_f\,\frac{f}{\mathcal{F}}\,\frac{r_1 f^*}{\mathcal{F}}\,(1-v)\right\},$

(3) $P_{f^*\to f^*+2} = \left(\frac{f}{f+f^*}\right)^2\,\left\{\lambda_f\,\left[\frac{r_1 f^*}{\mathcal{F}}\,(1-v)\right]^2 + (1-\lambda_f)\,\frac{r_1 d^*}{\mathcal{D}}\,(1-u)\,\left[(1-\lambda_s)\,\frac{r_1 d^*}{\mathcal{D}}\,(1-u) + \lambda_s\,(1-\sigma)\,\frac{r_1 b^*}{\mathcal{R}_b}\right]\right\},$

(4) $P_{f^*\to f^*-2} = \left(\frac{f^*}{f+f^*}\right)^2\,\left\{\lambda_f\,\left(\frac{f}{\mathcal{F}}\right)^2 + (1-\lambda_f)\,\frac{d}{\mathcal{D}}\left[(1-\lambda_s)\,\frac{d}{\mathcal{D}} + \lambda_s\,(1-\sigma)\,\frac{b}{\mathcal{R}_b} + \lambda_s\,\sigma\left(\delta\frac{b}{\mathcal{R}_b} + (1-\delta)(1-\gamma)\frac{b}{\mathcal{R}_b} + (1-\delta)\gamma\,\frac{e}{\mathcal{R}_2}\frac{e}{e+e^*}\right)\right]\right\},$

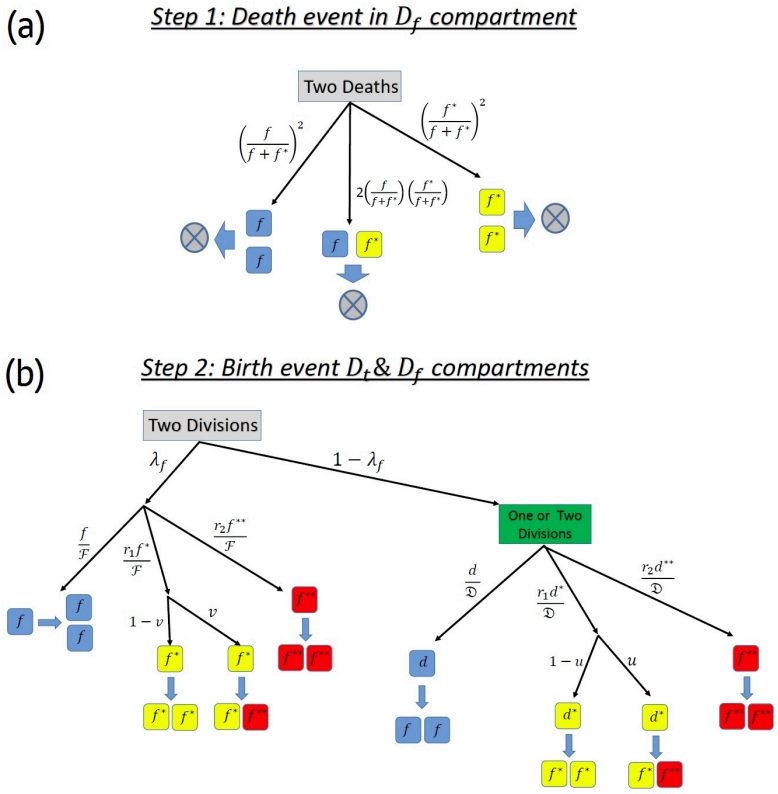


Figure B.1: **The cartoon figure of the possible death and birth.** The sub-figure (a) represents the three possible death in the D_f compartment. The sub-figure (b) shows the probable divisions occurring in either D_f or D_t compartments to replace the dead cells. With a probability of λ_f divisions occur in the D_f. Otherwise the replacements can be the result of divisions in D_t population with a probability of $1-\lambda_f$. We have $\mathcal{D} = r_2\, d^{**} + r_1\, d^* + d$ and $\mathcal{F} = r_2\, f^{**} + r_1\, f^* + f$.

$$(5)\quad P_{f^*,f^{**}\to f^*-1,f^{**}+1} = \frac{2ff^*}{(f+f^*)^2}\left\{2\,\lambda_f\,\left(\frac{f}{\mathcal{F}}\right)\left(\frac{r_1\,f^*}{\mathcal{F}}\,v + \frac{r_2\,f^{**}}{\mathcal{F}}\right)\right\} + \left(\frac{f^*}{f+f^*}\right)^2\left\{2\,\lambda_f\,\left(\frac{r_1\,f^*}{\mathcal{F}}\,(1-v)\right)\right.$$

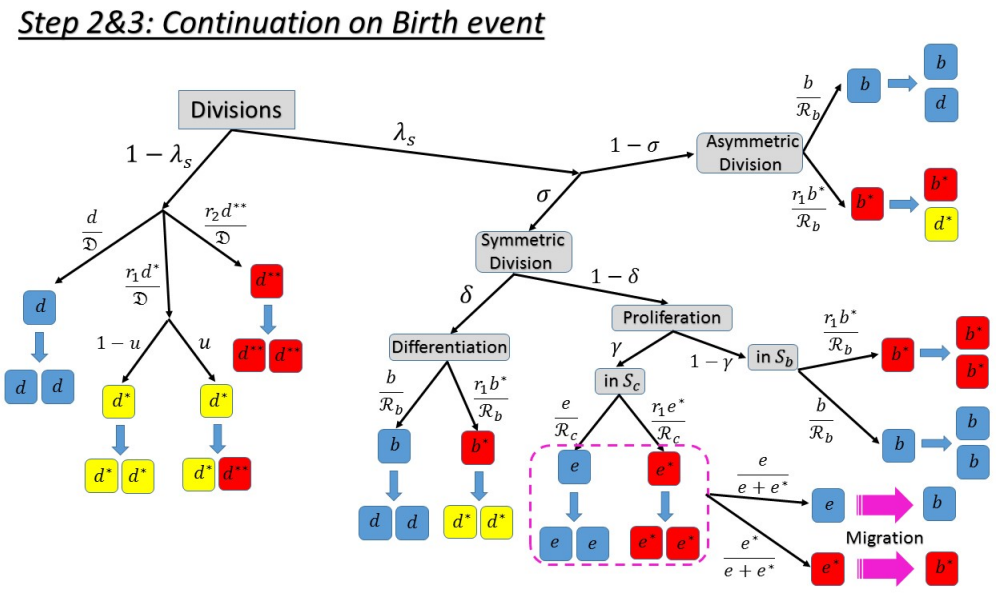


Figure B.2: **A representative cartoon picture representing the hierarchy of divisions occurring in the D_t, S_b, and S_c compartments as a continuation to the birth events.** Figure (a) reveals a continuation to the second step where all the possible cases happen in the D_t, S_b, S_c compartments. A cell divides in the D_t population with a probability of $1-\lambda_s$, and with a probability of λ_s in the S_b or S_c compartments. The sub-figure (b) indicates the possibilities of migration from the S_c compartment to the S_b. Considering $\mathcal{D}$ and $\mathcal{F}$ as those defined in the preceding figure, in this figure, we assume $\mathcal{R}_c = r_1\, e^* + e$ and $\mathcal{R}_b = r_1\, b^* + b$.

$$\times \left[\tfrac{r_1 f^*}{\mathcal{F}}\, v + \tfrac{r_2 f^{**}}{\mathcal{F}}\right] + (1-\lambda_f)\, \tfrac{r_1 d^*}{\mathcal{D}}\, u \left[(1-\lambda_s)\, \tfrac{r_1 d^*}{\mathcal{D}}\, (1-u) + \lambda_s\, (1-\sigma)\, \tfrac{r_1 b^*}{\mathcal{R}_b}\right]\Big\},$$

(6) $$P_{f^*, f^{**} \to f^*+1, f^{**}+1} = \left(\tfrac{f}{f+f^*}\right)^2 \Big\{2\,\lambda_f\, \left(\tfrac{r_1 f^*}{\mathcal{F}}\,(1-v)\right) \left[\tfrac{r_1 f^*}{\mathcal{F}}\, v + \tfrac{r_2 f^{**}}{\mathcal{F}}\right]$$
$$+(1-\lambda_f)\, \tfrac{r_1 d^*}{\mathcal{D}}\, u \left[(1-\lambda_s)\, \tfrac{r_1 d^*}{\mathcal{D}}\, (1-u) + \lambda_s\, (1-\sigma)\, \tfrac{r_1 b^*}{\mathcal{R}_b}\right]\Big\},$$

(7) $$P_{f^*, f^{**} \to f^*-2, f^{**}+1} = \left(\tfrac{f^*}{f+f^*}\right)^2 \Big\{2\,\lambda_f\, \left(\tfrac{f}{\mathcal{F}}\right) \left[\tfrac{r_1 f^*}{\mathcal{F}}\, v + \tfrac{r_2 f^{**}}{\mathcal{F}}\right]\Big\},$$

(8) $$P_{f^*, d^* \to f^*+2, d^*-1} = \left(\tfrac{f}{f+f^*}\right)^2 \Big\{(1-\lambda_f)\, \tfrac{r_1 d^*}{\mathcal{D}}\, (1-u) \left[(1-\lambda_s)\, \tfrac{d}{\mathcal{D}} + \lambda_s\, (1-\sigma)\, \tfrac{b}{\mathcal{R}_b}\right.$$

$$+\lambda_s\,\sigma\left(\delta\tfrac{b}{\mathcal{R}_b}+(1-\delta)(1-\gamma)\tfrac{b}{\mathcal{R}_b}+(1-\delta)\gamma\,\tfrac{e}{\mathcal{R}_2}\tfrac{e}{e+e^*}\right)\Big]\Big\},$$

(9) $$P_{f^*,d^*\to f^*+1,d^*-1}=\tfrac{2ff^*}{(f+f^*)^2}\,\Big\{(1-\lambda_f)\,\tfrac{r_1\,d^*}{\mathcal{D}}\,(1-u)\,\Big[(1-\lambda_s)\,\tfrac{d}{\mathcal{D}}+\lambda_s\,(1-\sigma)\,\tfrac{b}{\mathcal{R}_b}$$
$$+\lambda_s\,\sigma\left(\delta\tfrac{b}{\mathcal{R}_b}+(1-\delta)(1-\gamma)\tfrac{b}{\mathcal{R}_b}+(1-\delta)\gamma\,\tfrac{e}{\mathcal{R}_2}\tfrac{e}{e+e^*}\right)\Big]\Big\},$$

(10) $$P_{f^*,f^{**},d^*\to f^*+1,f^{**}+1,d^*-1}=\left(\tfrac{f}{f+f^*}\right)^2\,\Big\{(1-\lambda_f)\,\tfrac{r_1\,d^*}{\mathcal{D}}\,u\,\Big[(1-\lambda_s)\,\tfrac{d}{\mathcal{D}}+\lambda_s\,(1-\sigma)\,\tfrac{b}{\mathcal{R}_b}$$
$$+\lambda_s\,\sigma\left(\delta\tfrac{b}{\mathcal{R}_b}+(1-\delta)(1-\gamma)\tfrac{b}{\mathcal{R}_b}+(1-\delta)\gamma\,\tfrac{e}{\mathcal{R}_2}\tfrac{e}{e+e^*}\right)\Big]\Big\},$$

(11) $$P_{f^*,f^{**},d^*\to f^*-1,f^{**}+1,d^*-1}=\left(\tfrac{f^*}{f+f^*}\right)^2\,\Big\{(1-\lambda_f)\,\tfrac{r_1\,d^*}{\mathcal{D}}\,u\,\Big[(1-\lambda_s)\,\tfrac{d}{\mathcal{D}}+\lambda_s\,(1-\sigma)\,\tfrac{b}{\mathcal{R}_b}$$
$$+\lambda_s\,\sigma\left(\delta\tfrac{b}{\mathcal{R}_b}+(1-\delta)(1-\gamma)\tfrac{b}{\mathcal{R}_b}+(1-\delta)\gamma\,\tfrac{e}{\mathcal{R}_2}\tfrac{e}{e+e^*}\right)\Big]\Big\},$$

(12) $$P_{f^{**},d^*\to f^{**}+1,d^*-1}=\tfrac{2ff^*}{(f+f^*)^2}\,\Big\{(1-\lambda_f)\,\tfrac{r_1\,d^*}{\mathcal{D}}\,u\,\Big[(1-\lambda_s)\,\tfrac{d}{\mathcal{D}}+\lambda_s\,(1-\sigma)\,\tfrac{b}{\mathcal{R}_b}$$
$$+\lambda_s\,\sigma\left(\delta\tfrac{b}{\mathcal{R}_b}+(1-\delta)(1-\gamma)\tfrac{b}{\mathcal{R}_b}+(1-\delta)\gamma\,\tfrac{e}{\mathcal{R}_2}\tfrac{e}{e+e^*}\right)\Big]\Big\},$$

(13) $$P_{f^{**}\to f^{**}+1}=\left(\tfrac{f}{f+f^*}\right)^2\,\Big\{2\,\lambda_f\,\tfrac{f}{\mathcal{F}}\,\Big[\tfrac{r_1f^*}{\mathcal{F}}\,v+\tfrac{r_2f^{**}}{\mathcal{F}}\Big]\Big\}+\tfrac{2ff^*}{(f+f^*)^2}\,\Big\{2\,\lambda_f\,\tfrac{r_1\,f^*}{\mathcal{F}}\,(1-v)\,\Big[\tfrac{r_1\,f^*}{\mathcal{F}}\,v+\tfrac{r_2\,f^{**}}{\mathcal{F}}\Big]$$
$$+(1-\lambda_f)\,\tfrac{r_1\,d^*}{\mathcal{D}}\,u\,\Big[(1-\lambda_s)\,\tfrac{r_1d^*}{\mathcal{D}}\,(1-u)+\lambda_s\,(1-\sigma)\,\tfrac{r_1b^*}{\mathcal{R}_b}\Big]\Big\},$$

(14) $$P_{f^{**}\to f^{**}+2}=\left(\tfrac{f}{f+f^*}\right)^2\,\Big\{\lambda_f\,\Big[\tfrac{r_1f^*}{\mathcal{F}}\,v+\tfrac{r_2f^{**}}{\mathcal{F}}\Big]^2+(1-\lambda_f)\,\tfrac{r_2d^{**}}{\mathcal{D}}\,\Big[(1-\lambda_s)\,\Big(\tfrac{r_1d^*}{\mathcal{D}}\,u+\tfrac{r_2d^{**}}{\mathcal{D}}\Big)\Big]\Big\},$$

(15) $$P_{f^*,f^{**}\to f^*-2,f^{**}+2}=\left(\tfrac{f^*}{f+f^*}\right)^2\,\Big\{\lambda_f\,\Big[\tfrac{r_1f^*}{\mathcal{F}}\,v+\tfrac{r_2f^{**}}{\mathcal{F}}\Big]^2$$
$$+(1-\lambda_f)\,\tfrac{r_2d^{**}}{\mathcal{D}}\,\Big[(1-\lambda_s)\,\Big(\tfrac{r_1d^*}{\mathcal{D}}\,u+\tfrac{r_2d^{**}}{\mathcal{D}}\Big)\Big]\Big\},$$

(16) $$P_{f^*,f^{**}\to f^*-1,f^{**}+2}=\tfrac{2ff^*}{(f+f^*)^2}\,\Big\{\lambda_f\,\Big[\tfrac{r_1f^*}{\mathcal{F}}\,v+\tfrac{r_2f^{**}}{\mathcal{F}}\Big]^2$$
$$+(1-\lambda_f)\,\tfrac{r_2d^{**}}{\mathcal{D}}\,\Big[(1-\lambda_s)\,\Big(\tfrac{r_1d^*}{\mathcal{D}}\,u+\tfrac{r_2d^{**}}{\mathcal{D}}\Big)\Big]\Big\},$$

(17) $$P_{d^*\to d^*+1}=\left(\tfrac{f}{f+f^*}\right)^2\,\Big\{(1-\lambda_f)\,\tfrac{d}{\mathcal{D}}\,\Big[(1-\lambda_s)\,\tfrac{r_1d^*}{\mathcal{D}}\,(1-u)+\lambda_s\,(1-\sigma)\,\tfrac{r_1b^*}{\mathcal{R}_b}\Big]\Big\},$$

(18) $$P_{d^*,f^*\to d^*+1,f^*-1}=\tfrac{2ff^*}{(f+f^*)^2}\,\Big\{(1-\lambda_f)\,\tfrac{d}{\mathcal{D}}\,\Big[(1-\lambda_s)\,\tfrac{r_1d^*}{\mathcal{D}}\,(1-u)+\lambda_s\,(1-\sigma)\,\tfrac{r_1b^*}{\mathcal{R}_b}\Big]\Big\},$$

(19) $$P_{d^*,f^*\to d^*+1,f^*-2}=\left(\tfrac{f^*}{f+f^*}\right)^2\,\Big\{(1-\lambda_f)\,\tfrac{d}{\mathcal{D}}\,\Big[(1-\lambda_s)\,\tfrac{r_1d^*}{\mathcal{D}}\,(1-u)+\lambda_s\,(1-\sigma)\,\tfrac{r_1b^*}{\mathcal{R}_b}\Big]\Big\},$$

(20) $P_{b^*,d^*,f^*\to b^*-1,d^*+2,f^*-2} = \left(\frac{f^*}{f+f^*}\right)^2 \left\{(1-\lambda_f)\,\frac{d}{\mathcal{D}}\,\left[\lambda_s\,\sigma\,\delta\,\frac{r_1 b^*}{\mathcal{R}_b}\right]\right\},$

(21) $P_{b^*,d^*,f^*\to b^*-1,d^*+2,f^*-1} = \frac{2ff^*}{(f+f^*)^2}\,\left\{(1-\lambda_f)\,\frac{d}{\mathcal{D}}\,\left[\lambda_s\,\sigma\,\delta\,\frac{r_1 b^*}{\mathcal{R}_b}\right]\right\},$

(22) $P_{d^*\to d^*-1} = \left(\frac{f^*}{f+f^*}\right)^2 \left\{(1-\lambda_f)\,\frac{r_1 d^*}{\mathcal{D}}\,(1-u)\left[(1-\lambda_s)\,\frac{d}{\mathcal{D}} + \lambda_s\,(1-\sigma)\,\frac{b}{\mathcal{R}_b}\right.\right.$
$\left.\left. +\lambda_s\,\sigma\left(\delta\frac{b}{\mathcal{R}_b} + (1-\delta)(1-\gamma)\frac{b}{\mathcal{R}_b} + (1-\delta)\gamma\,\frac{e}{\mathcal{R}_2}\frac{e}{e+e^*}\right)\right]\right\},$

(23) $P_{f^*,f^{**},d^{**}\to f^*-2,f^{**}+2,d^{**}-1} = \left(\frac{f^*}{f+f^*}\right)^2 \left\{(1-\lambda_f)\,\frac{r_2 d^{**}}{\mathcal{D}}\,\left[(1-\lambda_s)\,\frac{d}{\mathcal{D}} + \lambda_s\,(1-\sigma)\,\frac{b}{\mathcal{R}_b}\right.\right.$
$\left.\left. +\lambda_s\,\sigma\left(\delta\frac{b}{\mathcal{R}_b} + (1-\delta)(1-\gamma)\frac{b}{\mathcal{R}_b} + (1-\delta)\gamma\,\frac{e}{\mathcal{R}_2}\frac{e}{e+e^*}\right)\right]\right\},$

(24) $P_{d^{**}\to d^{**}+1} = \left(\frac{f}{f+f^*}\right)^2 \left\{(1-\lambda_f)\,\frac{d}{\mathcal{D}}\,\left[(1-\lambda_s)\,\left(\frac{r_1 d^*}{\mathcal{D}}\,u + \frac{r_2 d^{**}}{\mathcal{D}}\right)\right]\right\},$

(25) $P_{d^{**},f^*\to d^{**}+1,f^*-1} = \frac{2ff^*}{(f+f^*)^2}\,\left\{(1-\lambda_f)\,\frac{d}{\mathcal{D}}\,\left[(1-\lambda_s)\,\left(\frac{r_1 d^*}{\mathcal{D}}\,u + \frac{r_2 d^{**}}{\mathcal{D}}\right)\right]\right\},$

(26) $P_{d^{**},f^*\to d^{**}+1,f^*-2} = \left(\frac{f^*}{f+f^*}\right)^2 \left\{(1-\lambda_f)\,\frac{d}{\mathcal{D}}\,\left[(1-\lambda_s)\,\left(\frac{r_1 d^*}{\mathcal{D}}\,u + \frac{r_2 d^{**}}{\mathcal{D}}\right)\right]\right\},$

(27) $P_{d^*,b^*\to d^*+2,b^*-1} = \left(\frac{f}{f+f^*}\right)^2 \left\{(1-\lambda_f)\,\frac{d}{\mathcal{D}}\,\left[\lambda_s\,\sigma\delta\,\frac{r_1 b^*}{\mathcal{R}_b}\right]\right\},$

(28) $P_{d^*,b^*\to d^*-1,b^*+1} = \left(\frac{f^*}{f+f^*}\right)^2 \left\{(1-\lambda_f)\,\frac{r_1 d^*}{\mathcal{D}}\,(1-u)\left[\lambda_s\,\sigma\,(1-\delta)\,\left((1-\gamma)\,\frac{r_1 b^*}{\mathcal{R}_b}\right.\right.\right.$
$\left.\left.\left. +\gamma\,\frac{r_1 e^*}{\mathcal{R}_2}\,\frac{e^*}{e+e^*}\right)\right]\right\},$

(29) $P_{d^*,b^*\to d^*+1,b^*-1} = \left(\frac{f^*}{f+f^*}\right)^2 \left\{(1-\lambda_f)\,\frac{r_1 d^*}{\mathcal{D}}\,(1-u)\left[\lambda_s\,\sigma\,\delta\,\frac{r_1 b^*}{\mathcal{R}_b}\right]\right\},$

(30) $P_{b^*\to b^*+1} = \left(\frac{f}{f+f^*}\right)^2 \left\{(1-\lambda_f)\,\frac{d}{\mathcal{D}}\,\left[\lambda_s\,\sigma\,(1-\delta)\,(1-\gamma)\,\frac{r_1 b^*}{\mathcal{R}_b} + \lambda_s\,\sigma\,(1-\delta)\,\gamma\,\frac{r_1 e^*}{\mathcal{R}_2}\,\frac{e^*}{e+e^*}\right]\right\},$

(31) $P_{b^*,f^*\to b^*+1,f^*-1} = \frac{2ff^*}{(f+f^*)^2}\,\left\{(1-\lambda_f)\,\frac{d}{\mathcal{D}}\,\left[\lambda_s\,\sigma\,(1-\delta)\,(1-\gamma)\,\frac{r_1 b^*}{\mathcal{R}_b}\right.\right.$
$\left.\left. +\lambda_s\,\sigma\,(1-\delta)\,\gamma\,\frac{r_1 e^*}{\mathcal{R}_2}\,\frac{e^*}{e+e^*}\right]\right\},$

(32) $P_{b^*,f^*\to b^*+1,f^*-2} = \left(\frac{f^*}{f+f^*}\right)^2 \left\{(1-\lambda_f)\,\frac{d}{\mathcal{D}}\,\left[\lambda_s\,\sigma\,(1-\delta)\,(1-\gamma)\,\frac{r_1 b^*}{\mathcal{R}_b}\right.\right.$
$\left.\left. +\lambda_s\,\sigma\,(1-\delta)\,\gamma\,\frac{r_1 e^*}{\mathcal{R}_2}\,\frac{e^*}{e+e^*}\right]\right\},$

(33) $P_{e^* \to e^*+1} = \left(\frac{f}{f+f^*}\right)^2 \left\{(1-\lambda_f)\,\frac{d}{\mathcal{D}}\,\left[\lambda_s\,\sigma\,(1-\delta)\,\gamma\,\frac{r_1 e^*}{\mathcal{R}_2}\,\frac{e}{e+e^*}\right]\right\},$

(34) $P_{f^{**},d^{**} \to f^{**}+2,d^{**}-1} = \left(\frac{f}{f+f^*}\right)^2 \left\{(1-\lambda_f)\,\frac{r_2 d^{**}}{\mathcal{D}}\,\left[(1-\lambda_s)\,\frac{d}{\mathcal{D}} + \lambda_s\,(1-\sigma)\,\frac{b}{\mathcal{R}_b}\right.\right.$
$\left.\left. +\lambda_s\,\sigma\left(\delta\,\frac{b}{\mathcal{R}_b} + (1-\delta)(1-\gamma)\,\frac{b}{\mathcal{R}_b} + (1-\delta)\gamma\,\frac{e}{\mathcal{R}_2}\,\frac{e}{e+e^*}\right)\right]\right\},$

(35) $P_{e^*,d^* \to e^*+1,d^*-1} = \left(\frac{f^*}{f+f^*}\right)^2 \left\{(1-\lambda_f)\,\frac{r_1 d^*}{\mathcal{D}}\,(1-u)\,\left[\lambda_s\,\sigma\,(1-\delta)\,\gamma\,\frac{r_1 e^*}{\mathcal{R}_2}\,\frac{e}{e+e^*}\right]\right\},$

(36) $P_{e^*,b^* \to e^*-1,b^*+1} = \left(\frac{f}{f+f^*}\right)^2 \left\{(1-\lambda_f)\,\frac{d}{\mathcal{D}}\,\left[\lambda_s\,\sigma\,(1-\delta)\,\gamma\,\frac{e}{\mathcal{R}_2}\,\frac{e^*}{e+e^*}\right]\right\},$

(37) $P_{e^*,b^*,d^* \to e^*-1,b^*+1,d^*-1} = \left(\frac{f^*}{f+f^*}\right)^2 \left\{(1-\lambda_f)\,\frac{r_1 d^*}{\mathcal{D}}\,(1-u)\,\left[\lambda_s\,\sigma\,(1-\delta)\,\gamma\,\frac{e}{\mathcal{R}_2}\,\frac{e^*}{e+e^*}\right]\right\},$

(38) $P_{e^*,d^*,f^{**} \to e^*+1,d^*-1,f^{**}+1} = \frac{2ff^*}{(f+f^*)^2}\,\left\{(1-\lambda_f)\,\frac{r_1 d^*}{\mathcal{D}}\,u\,\left[\lambda_s\,\sigma\,(1-\delta)\,\gamma\,\frac{r_1 e^*}{\mathcal{R}_2}\,\frac{e}{e+e^*}\right]\right\},$

(39) $P_{d^*,d^{**},f^* \to d^*-1,d^{**}+1,f^*+2} = \left(\frac{f}{f+f^*}\right)^2 \left\{(1-\lambda_f)\,\frac{r_1 d^*}{\mathcal{D}}\,(1-u)\,\left[(1-\lambda_s)\,\left(\frac{r_1 d^*}{\mathcal{D}}\,u + \frac{r_2 d^{**}}{\mathcal{D}}\right)\right]\right\},$

(40) $P_{d^*,d^{**},f^* \to d^*-1,d^{**}+1,f^*+1} = \frac{2ff^*}{(f+f^*)^2}\,\left\{(1-\lambda_f)\,\frac{r_1 d^*}{\mathcal{D}}\,(1-u)\,\left[(1-\lambda_s)\,\left(\frac{r_1 d^*}{\mathcal{D}}\,u + \frac{r_2 d^{**}}{\mathcal{D}}\right)\right]\right\},$

(41) $P_{d^*,d^{**} \to d^*-1,d^{**}+1} = \left(\frac{f^*}{f+f^*}\right)^2 \left\{(1-\lambda_f)\,\frac{r_1 d^*}{\mathcal{D}}\,(1-u)\,\left[(1-\lambda_s)\,\left(\frac{r_1 d^*}{\mathcal{D}}\,u + \frac{r_2 d^{**}}{\mathcal{D}}\right)\right]\right\},$

(42) $P_{d^*,b^*,f^* \to d^*+1,b^*-1,f^*+2} = \left(\frac{f}{f+f^*}\right)^2 \left\{(1-\lambda_f)\,\frac{r_1 d^*}{\mathcal{D}}\,(1-u)\,\left[\lambda_s\,\sigma\,\delta\,\frac{r_1 b^*}{\mathcal{R}_b}\right]\right\},$

(43) $P_{d^*,b^*,f^* \to d^*+1,b^*-1,f^*+1} = \frac{2ff^*}{(f+f^*)^2}\,\left\{(1-\lambda_f)\,\frac{r_1 d^*}{\mathcal{D}}\,(1-u)\,\left[\lambda_s\,\sigma\,\delta\,\frac{r_1 b^*}{\mathcal{R}_b}\right]\right\},$

(44) $P_{d^*,b^*,f^*,f^{**} \to d^*-1,b^*+1,f^*+1,f^{**}+1} = \left(\frac{f}{f+f^*}\right)^2 \left\{(1-\lambda_f)\,\frac{r_1 d^*}{\mathcal{D}}\,u\,\left[\lambda_s\,\sigma\,(1-\delta)\,(1-\gamma)\,\frac{r_1 b^*}{\mathcal{R}_b}\right.\right.$
$\left.\left. +\lambda_s\,\sigma\,(1-\delta)\,\gamma\,\frac{r_1 e^*}{\mathcal{R}_2}\,\frac{e^*}{e+e^*}\right]\right\},$

(45) $P_{d^*,b^*,f^*,f^{**} \to d^*-1,b^*+1,f^*-1,f^{**}+1} = \left(\frac{f^*}{f+f^*}\right)^2 \left\{(1-\lambda_f)\,\frac{r_1 d^*}{\mathcal{D}}\,u\,\left[\lambda_s\,\sigma\,(1-\delta)\,(1-\gamma)\,\frac{r_1 b^*}{\mathcal{R}_b}\right.\right.$
$\left.\left. +\lambda_s\,\sigma\,(1-\delta)\,\gamma\,\frac{r_1 e^*}{\mathcal{R}_2}\,\frac{e^*}{e+e^*}\right]\right\},$

(46) $P_{d^*,b^*,f^{**} \to d^*-1,b^*+1,f^{**}+1} = \frac{2ff^*}{(f+f^*)^2}\,\left\{(1-\lambda_f)\,\frac{r_1 d^*}{\mathcal{D}}\,u\,\left[\lambda_s\,\sigma\,(1-\delta)\,(1-\gamma)\,\frac{r_1 b^*}{\mathcal{R}_b}\right.\right.$
$\left.\left. +\lambda_s\,\sigma\,(1-\delta)\,\gamma\,\frac{r_1 e^*}{\mathcal{R}_2}\,\frac{e^*}{e+e^*}\right]\right\},$

$$(47)\ P_{d^*,b^*,f^*,f^{**}\to d^*+1,b^*-1,f^*+1,f^{**}+1} = \left(\frac{f}{f+f^*}\right)^2 \left\{(1-\lambda_f)\,\frac{r_1 d^*}{\mathcal{D}}\,u\,\left[\lambda_s\,\sigma\,\delta\,\frac{r_1 b^*}{\mathcal{R}_b}\right]\right\},$$

$$(48)\ P_{d^*,b^*,f^*,f^{**}\to d^*+1,b^*-1,f^*-1,f^{**}+1} = \left(\frac{f^*}{f+f^*}\right)^2 \left\{(1-\lambda_f)\,\frac{r_1 d^*}{\mathcal{D}}\,u\,\left[\lambda_s\,\sigma\,\delta\,\frac{r_1 b^*}{\mathcal{R}_b}\right]\right\},$$

$$(49)\ P_{d^*,b^*,f^{**}\to d^*+1,b^*-1,f^{**}+1} = \frac{2ff^*}{(f+f^*)^2}\left\{(1-\lambda_f)\,\frac{r_1 d^*}{\mathcal{D}}\,u\,\left[\lambda_s\,\sigma\,\delta\,\frac{r_1 b^*}{\mathcal{R}_b}\right]\right\},$$

$$(50)\ P_{f^*,f^{**},d^{**}\to f^*-1,f^{**}+2,d^{**}-1} = \frac{2ff^*}{(f+f^*)^2}\left\{(1-\lambda_f)\,\frac{r_2 d^{**}}{\mathcal{D}}\,\left[(1-\lambda_s)\,\frac{d}{\mathcal{D}} + \lambda_s\,(1-\sigma)\,\frac{b}{\mathcal{R}_b} + \lambda_s\,\sigma\left(\delta\frac{b}{\mathcal{R}_b} + (1-\delta)(1-\gamma)\frac{b}{\mathcal{R}_b} + (1-\delta)\gamma\,\frac{e}{\mathcal{R}_2}\frac{e}{e+e^*}\right)\right]\right\},$$

$$(51)\ P_{e^*,b^*,d^*,f^{**}\to e^*-1,b^*+1,d^*-1,f^{**}+1} = \frac{2ff^*}{(f+f^*)^2}\left\{(1-\lambda_f)\,\frac{r_1 d^*}{\mathcal{D}}\,u\left[\lambda_s\,\sigma\,(1-\delta)\,\gamma\,\frac{e}{\mathcal{R}_2}\,\frac{e^*}{e+e^*}\right]\right\},$$

$$(52)\ P_{d^*,b^*,f^*\to d^*-1,b^*+1,f^*+1} = \frac{2ff^*}{(f+f^*)^2}\left\{(1-\lambda_f)\,\frac{r_1 d^*}{\mathcal{D}}\,(1-u)\left[\lambda_s\,\sigma\,(1-\delta)\,\left((1-\gamma)\,\frac{r_1 b^*}{\mathcal{R}_b} + \gamma\,\frac{r_1 e^*}{\mathcal{R}_2}\,\frac{e^*}{e+e^*}\right)\right]\right\},$$

$$(53)\ P_{d^*,b^*,f^*\to d^*-1,b^*+1,f^*+2} = \left(\frac{f}{f+f^*}\right)^2\left\{(1-\lambda_f)\,\frac{r_1 d^*}{\mathcal{D}}\,(1-u)\left[\lambda_s\,\sigma\,(1-\delta)\,\left((1-\gamma)\,\frac{r_1 b^*}{\mathcal{R}_b} + \gamma\,\frac{r_1 e^*}{\mathcal{R}_2}\,\frac{e^*}{e+e^*}\right)\right]\right\},$$

$$(54)\ P_{e^*,b^*,d^*,f^*\to e^*-1,b^*+1,d^*-1,f^*+1} = \frac{2ff^*}{(f+f^*)^2}\left\{(1-\lambda_f)\,\frac{r_1 d^*}{\mathcal{D}}\,(1-u)\left[\lambda_s\,\sigma\,(1-\delta)\,\gamma\,\frac{e}{\mathcal{R}_2}\,\frac{e^*}{e+e^*}\right]\right\},$$

$$(55)\ P_{e^*,b^*,d^*,f^*\to e^*-1,b^*+1,d^*-1,f^*+2} = \left(\frac{f}{f+f^*}\right)^2\left\{(1-\lambda_f)\,\frac{r_1 d^*}{\mathcal{D}}\,(1-u)\left[\lambda_s\,\sigma\,(1-\delta)\,\gamma\,\frac{e}{\mathcal{R}_2}\,\frac{e^*}{e+e^*}\right]\right\},$$

where $\mathcal{F}, \mathcal{D}, \mathcal{R}_b, \mathcal{R}_2$ are defined in the following

$$\begin{aligned}
\mathcal{R}_c &= r_1\,e^* + e,\\
\mathcal{R}_b &= r_1\,b^* + b,\\
\mathcal{D} &= r_2\,d^{**} + r_1\,d^* + d,\\
\mathcal{F} &= r_2\,f^{**} + r_1\,f^* + f.
\end{aligned} \tag{B.1}$$

B.3 Fixation Probability

The structure of the model comprises a wide variety of different scenarios which may occur in the system. Therefore, to conclude a specific probability of fixation, different mechanisms

may need to be taken into account. Signifying the long-term behavior of the system after initiation of a mutant within compartments, would assist us to understand the epithelial cell dynamics in colorectal and intestinal cancer. Even though there might be a chance for new mutations, according to the long time-scaling of mutations in compare to the the period required for the dynamics of mutants as late of absorption. Moreover, since the washed out mechanisms within the crypt eliminates new mutants, one may investigate dynamics of newborn mutants in a specific compartment without any new mutations (forward or backward mutations), i.e. $u = v = 0$. Such a scenario, in turn, will help us to understand the survival chance of one mutant in each compartment as well as the fixation probability of mutants, which may subsequently appear in the other compartments as a result of mutants' division. Therefore, in this section we study the fixation probability of a mutant in a given compartment. To investigate the survival probability of a mutant in a particular population, we calculate the probability that the progeny of mutants will take over the whole compartment. We denote the probability of absorption of j number of mutants in a population of size $N > j$ by π_j, then the probability of one mutant's progeny taking over the entire population (π_1) can be obtained using the following system of equations.

$$\begin{cases} \pi_j = \sum P_{j\to m}\pi_m, \quad 1 < j < N-1, \\ \pi_1 = \sum_{m\geq 1}^{m} P_{1\to m}\pi_m, \\ \pi_{N-1} = P_{N-1\to N} + \sum_{m\leq N-1} P_{j\to m}\pi_m, \\ \pi_N = 1. \end{cases} \tag{B.2}$$

Here, $P_{j\to m}$ is the probability of transition from state j to state m and the initial condition is to have only one mutant of a certain type while there exist no other type of mutants in the system.

The analytic results, which are based on the transition probabilities derived in the previous section, are in a perfect agreement with the simulation results. Therefore, we can rely on simulation results to investigate more complicated scenarios for our generalized multi–compartmental model. A general mechanism for higher dimensional Markov chains is given in Chapter 2, Section 2.1.2.