



From the Institute of Mathematics
of the the University of Lübeck

Director:
Prof. Dr. Jürgen Prestin

Mathematical Models of Cell Population Dynamics

Dissertation for Fulfillment
of Requirements for the
Doctoral Degree
of the University of Lübeck

from the Department of Computer Sciences/Engineering

Submitted by
Benjamin Werner
Born 15.09.1984 in Leipzig

Research Group for Evolutionary Theory
Max Planck Institute for Evolutionary Biology

Supervisor:
Prof. Dr. Arne Traulsen

Printed: Plön, 24th of April 2013

First referee - Prof. Dr. Arne TRAULSEN
Second referee - Prof. Dr. Andreas DEUTSCH

Date of oral examination: 9. September 2013

Approved for publication: 9. September 2013

Signed:

Abstract

Cancers result from altered cell proliferation properties, caused by mutations in specific genes. An accumulation of multiple mutations within a cell increases the risk to develop cancer. However, mechanisms evolved to prevent such multiple mutations. One such mechanism is a hierarchically organized tissue structure. At the root of the hierarchy are a few, slow proliferating stem cells. After some cell differentiations all functional cells of a tissue are obtained. In the first two chapters of this thesis, we mathematically and computationally evaluate a multi compartment model that is an abstract representation of such hierarchical tissues. We find analytical expressions for stem cell and non stem cell driven cell populations without further mutations. We show that non stem cell mutations give rise to clonal waves, that travel through the hierarchy and are lost in the long run. We calculate the average extinction times of such clonal waves. In the third chapter we allow for arbitrary many mutations in hierarchically organized tissues and find exact expressions for the reproductive capacity of cells, highlighting that multiple mutations are strongly suppressed by the hierarchy. In the fourth chapter we turn to a related problem, the evolution of resistance against molecular targeted cancer drugs. We develop a minimalistic mathematical model and compare the predicted dynamics to experimental derived observations. Interestingly we find that resistance can be induced either by mutation or intercellular processes such as phenotypic switching. In the fifth chapter of this thesis, we investigate the shortening of telomeres in detail. The comparison of mathematical results to experimental data reveals interesting properties of stem cell dynamics. We find hints for an increasing stem cell pool size with age, caused by a small number of symmetric stem cell divisions. We also implement disease scenarios and find exact expressions how the patterns of telomere shortening differ for healthy and sick persons. Our model provides a simple explanation for the pronounced increase of telomere shortening in the first years of life, followed by an almost linear decrease for healthy adults. In the final chapter, we implement a method to introduce arbitrary many random mutations into the framework of frequency dependent selection. We show how disadvantageous mutations can reach fixation under a deterministic scenario and discuss possible applications to cancer modeling.

Kurzfassung

Krebs entsteht durch von Mutationen induzierte Veränderungen der Zellteilungseigenschaften. Das Anhäufen einer Vielzahl unterschiedlicher Mutationen in einer Zelle erhöht das Risiko der Krebsentstehung. Demgegenüber haben sich Mechanismen entwickelt, die eine solche Anhäufung verhindern. Einer dieser Mechanismen ist eine hierarchisch organisierte Gewebestruktur. Stammzellen können sowohl differenziertere Zellen produzieren, als auch die Stammzellpopulation aufrecht erhalten. Nach einer bestimmten Anzahl von Zelldifferenzierungen werden alle funktionalen Zellen eines Gewebes produziert. In den ersten zwei Kapiteln dieser Arbeit untersuchen wir mit mathematischen und computerorientierten Methoden ein Mehrkompartimentmodell, das eine solche hierarchische Gewebeorganisation repräsentiert. Wir entwickeln analytische Ausdrücke für Zellpopulationen mit und ohne Stammzellen und ohne zusätzliche Mutationen. Zellpopulationen ohne Stammzellunterstützung resultieren in klonalen Wellen, die durch die Hierarchie wandern und letztlich verloren gehen. Zusätzlich berechnen wir die Aussterbezeiten dieser Wellen. Im dritten Kapitel erlauben wir beliebig viele Mutationen und erreichen exakte Ausdrücke für die Reproduktionsfähigkeit von Zellen. Es stellt sich heraus, dass Zellen mit mehrfachen Mutationen innerhalb einer Hierarchie besonders stark unterdrückt werden. Im vierten Kapitel wenden wir uns dem Problem der Resistenzentstehung gegen bestimmte Krebsmedikamente zu. Wir entwickeln ein einfaches mathematisches Modell und vergleichen dessen Voraussagen zu experimentellen Ergebnissen. Interessanterweise können Resistenzen durch Mutationen und intrazelluläre Prozesse hervorgerufen werden. Im fünften Kapitel dieser Arbeit untersuchen wir die zeitliche Änderung von Telomerlängen. Der Vergleich von analytischen und experimentellen Ergebnissen liefert interessante Eigenschaften von Stammzelldynamiken. Wir finden Hinweise auf eine Zunahme der Stammzellen im Alter, hervorgerufen durch symmetrische Zellteilungen. Wir implementieren auch Krankheitsszenarien und erreichen exakte Ausdrücke für die erwartete Telomerlängenverteilung von gesunden und erkrankten Personen. Darüberhinaus erlaubt unser Modell eine einfache Erklärung der beschleunigten Abnahme der Telomerlänge in den ersten Lebensjahren. Im letzten Kapitel entwickeln wir eine Methode, um beliebig viele zufällige Mutationen in das Gebiet der frequenzabhängigen Selektion zu integrieren. Wir zeigen auf, wie nachteilige Mutationen die gesamte Population auch in einem deterministischen Szenario übernehmen können.

Contents

1	Introduction	1
1.1	Motivation	1
1.2	Biological basics	3
1.2.1	Hierarchically organized tissues	4
1.2.2	Cancer biology	6
1.2.3	Molecular targeted treatment strategies	9
1.2.4	Telomeres and telomerase	11
1.3	Stochastic dynamics	12
1.3.1	Gillespie algorithm	12
1.3.2	Moran process	15
1.3.3	Towards deterministic dynamics	22
1.4	Deterministic dynamics	23
1.4.1	Replicator equation	24
2	Single mutations in hierarchical tissues	27
2.1	Introduction	28
2.2	Mathematical model and results	30
2.2.1	Stem cell driven dynamics	31
2.2.2	Non stem cell driven dynamics	34
2.2.3	Mutant extinction times	37
2.2.4	Example: Dynamics of PIG-A mutants	39
2.3	Discussion	43
3	Multiple mutations in hierarchical tissues	45
3.1	Introduction	46
3.2	Mathematical model and results	47
3.2.1	Time continuous dynamics of multiple mutations	49
3.2.2	Cell reproductive capacity	54
3.2.3	Reproductive capacity of neutral mutants	55
3.2.4	Number of distinct neutral mutations	56
3.2.5	Example: clonal diversity in acute lymphoblastic leukemia	58
3.3	Discussion	59

4	Resistance evolution	63
4.1	Quasi species equation with time dependent fitness	64
4.2	Introduction	66
4.3	Experimental setup	67
4.4	Mathematical model and results	71
4.4.1	Analytical approximation for large population size	73
4.4.2	Development of Imatinib resistance in cell culture	74
4.4.3	Fitting the mathematical model to the experimental data	76
4.5	Discussion	78
5	A mathematical model of telomere shortening	81
5.1	Mathematical model and results	83
5.1.1	Asymmetric cell divisions	84
5.1.2	Symmetric cell divisions	89
5.1.3	T-cell mediated stem cell death	93
6	Impact of random mutations on population fitness	99
6.1	Random mutant games	100
6.2	Introduction	102
6.3	Mathematical model and results	103
6.3.1	Games with two types	103
6.3.2	Games with n types	109
6.3.3	Games with equal gains from switching	112
6.3.4	Diploid populations with two alleles	115
6.4	Discussion	115
6.5	Random mutations and cancer	118
7	Summary and Outlook	121
7.1	Summary	121
7.2	Outlook	123
	Bibliography	125

Introduction

1.1 Motivation

A scientific theory is a well substantiated explanation of some aspects of the natural world, that has been repeatedly confirmed through observation and experiment (Popper, 1959). In physics, theories are usually formulated in a mathematical framework and derived from a small set of postulations, for example classical electrodynamics described by Maxwell's equations (Jackson and Fox, 1999) or gravity described by Einsteins general relativity (Wald, 1984). The all embracing theory in biology is evolution, formulated by Darwin and Wallace by means of natural selection (Darwin, 1859). Based on competition within and between species, the principle of evolution is formulated orally. Ever since, scientists intended to derive quantitative theories of evolution, starting with classical population genetics by Fisher, Wright and Haldane (Fisher, 1930; Wright, 1932; Haldane, 1932) towards evolutionary game theory proposed by Maynard Smith (Maynard Smith, 1982). Although all these theories suffer from the enormous complexity of biological systems, they provide important contributions to the understanding of biological systems. This includes explanation of the widely observed sex ratio of 1:1 by Fisher (Fisher, 1930; Hamilton, 1967), the red queen hypothesis (van Valen, 1973) or eusociality by Hamilton's rule (Hamilton, 1964).

In recent years, the scientific community became increasingly interdisciplinary and methods of physics, chemistry and informatics are applied to biological systems and vice versa. These combined efforts yield substantial progress in our understanding of biological systems. One measure of this progress is the improvement of medical treatments and the accompanied increase of life expectancy for every new generation. Of course, the success of antibiotics in the treatment of infectious diseases is outstanding. However, the joint work of theory, experimental and clinical

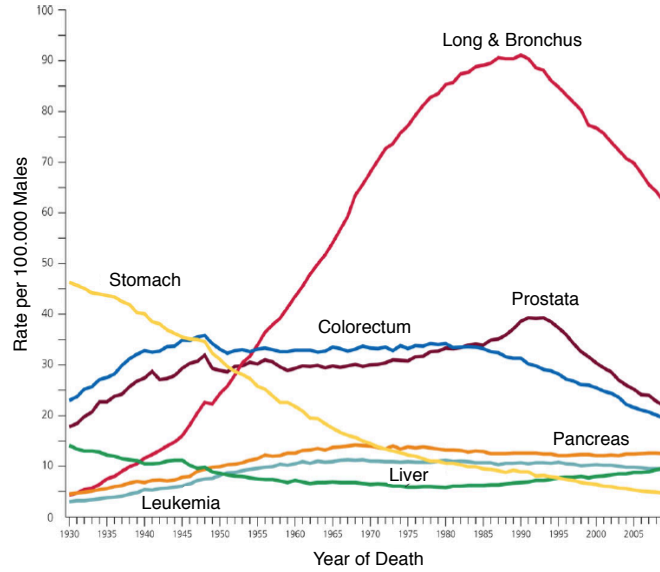


Figure 1.1: **Development of cancer specific death rates.** Shown is the development of the death rates per 100000 male individuals in the United States of America for different cancer types between 1930 and 2012. Mortality rates for cancers as lung or prostate are decreasing significantly, others as pancreatic cancers or leukemias remain unchanged. The figure was taken from (Siegel et al., 2013).

research show promising developments in a different area as well, that is cancer research. Understood as genetically caused, cancers fit into the framework of evolution as they can be seen as a competition of healthy and cancerous cells (Nowak, 2006a; Hanahan and Weinberg, 2000). Since the early 1990's there is a considerable decline in cancer related deaths, see figures 1.1 and 1.2. This decline is more pronounced for lung or prostate cancers and rather unchanged for leukemias or pancreatic cancers (Siegel et al., 2013). However, in the United States of America approximately 1.1 million cancer deaths were averted in the last 20 years, see figure 1.2. There are hopes to continue this promising trend. In this thesis, we investigate different aspects of cell populations dynamics. We mathematically and computationally tackle problems of cancer initiation in hierarchically organized tissues, resistance evolution and telomere dynamics. In the cases of resistance evolution and telomere dynamics we are also able to compare our deduced dynamics directly to experimental results. We hope, that this work provides a humble contribution to the broad area of biological modeling. In the rest of this chapter, we provide a short introduction to the basic biological terminology used throughout this thesis, followed by some the-

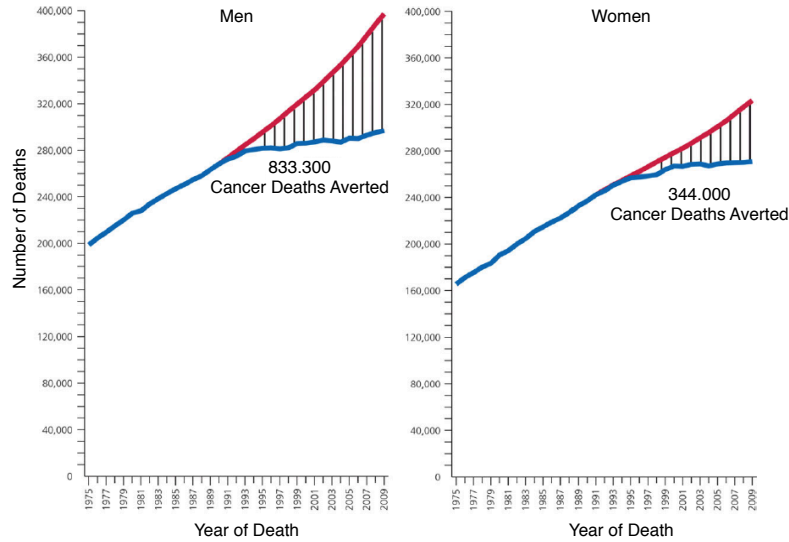


Figure 1.2: **Total number of cancer caused deaths for males and females in the United States of America between 1975 and 2012.** The blue line represents the actual number of cancer deaths recorded in each year, and the red line represents the number of cancer deaths that would have been expected, if cancer death rates had remained at their peak. The figure was taken from (Siegel et al., 2013).

oretical tools. In the main body of this thesis, we first discuss the dynamics of cell populations in hierarchically organized structures. We then shift our attention to aspects of the resistance evolution to molecular targeted treatment strategies. We further we derive telomere length distributions of (a)symmetric dividing stem cell populations, providing insights in the aging process of tissues. In the last chapter of the main body, we discuss a more general approach to model the impact of random frequency dependent mutations on a population. We conclude this thesis with some final remarks.

1.2 Biological basics

Throughout this thesis, we develop mathematical and computational models to describe various problems in biological systems. In this chapter we introduce the corresponding biological terminology. We explain the concept of hierarchically organized tissues and discuss the importance of mutations in cancer development. We give some insights in molecular targeted cancer treatments and its drawbacks. We also introduce telomeres and explain the importance of telomeres and telomerase on

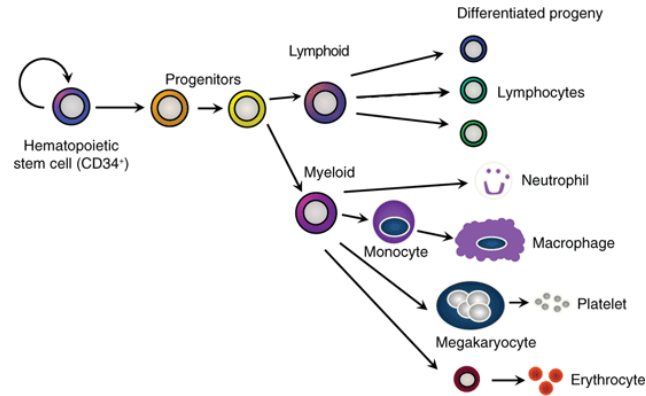


Figure 1.3: **Schematic simplified sketch of the hierarchical organization of hematopoiesis.** At the root of the hierarchy are a few tissue specific stem cells, able to self renewal and differentiate. Stem cells give raise to more specialized (differentiated) progenitor cells. During cell proliferation, cells commit further within the hierarchy. After a certain tissue dependent number of differentiation steps, all mature cells can be obtained. The figure was taken from (King and Goodell, 2011).

the regulation of cell cycle activity, damage control and cell senescence.

1.2.1 Hierarchically organized tissues

The seemingly unchanging physical appearance (at least over medium time scales) of organisms might give the impression of a very static tissue organization with very little cell proliferation. However, for most tissues the opposite is true. Cells continuously die and are replenished (Hayflick et al., 1961; Raff, 1992). For example in human hematopoiesis, on average 3.5×10^{11} mature blood cells are produced per day (Berardi et al., 1995; Abkowitz et al., 1996; Brümmendorf et al., 1998). Often such enormous cell turnover rates are realized by a hierarchical tissue organization, described for example for hematopoiesis (Dingli et al., 2007a; Nowak, 2006a), epidermal cell turnover in the skin (Tumbar et al., 2004; Fuchs, 2008), or in the colonic crypt (Potten et al., 2009). Mature functional tissue cells are produced throughout a series of cell differentiation steps (Pittenger et al., 1999; Michor et al., 2005), see Fig. 1.3 for a schematic sketch of human hematopoiesis. At the root of the hierarchy are a few tissue specific stem cells, combining two indispensable properties, self renewal and differentiation potential (McCulloch and Till, 2005; Grove et al., 2004; Dingli et al., 2007a; Werner et al., 2011). Self renewal is essential to avoid an exhaustion of the stem cell pool and the differentiation potential to enable mature cell production.

However, a clear distinction of mature tissue specific and embryonic stem cells is important. Where latter are pluripotent (can potentially differentiate into all three germ layers, endoderm, mesoderm and ectoderm) are former multipotent and the differentiation potential is limited to specific tissue lineages (Grove et al., 2004), although a reprogramming into pluripotent stem cells in vitro seems possible (Jiang et al., 2002; Yu et al., 2007). As we are concerned with tissue homeostasis, stem cell dynamics in our models relate to multipotent stem cells.

The exact phenotypic definition of stem cells appears to be difficult. The standard approach to characterize stemness is the identification and classification of expressed cell surface markers (Bunting, 2002). In recent years, an increasing number of stem like cells and their corresponding cell surface markers have been identified (amongst them CD34+ and CD38-)(Hao et al., 1995; Lavker and Sun, 2000). In addition gene expression profiles were analyzed (Berardi et al., 1995), however the interpretations and conclusions are still under discussion. Hayflick showed, that any cell line, cultured in vitro, has a limited repopulation capability of approximately 40-60 proliferation cycles per cell (Hayflick et al., 1961; Shay et al., 2000; Morrison et al., 1995). The underlying causes are incompletely understood, but one possible explanation is the progressive shortening of telomeres during cell proliferations, and an induced cell death for too short telomeres (Allsopp et al., 1992; Olovnikov, 1996). We discuss this in more detail below. This experimental result seems a contradiction to the observed unlimited reproduction capabilities of tissues in vivo. This leads to the postulation of stem cell niches. A stem cell niche is defined as spatial location, that induces stemness by providing special growth conditions to cells (Scadden, 2006; Lin, 2002). In certain tissues, as for example the colon, these locations were identified (Lavker and Sun, 2000; Tumbar et al., 2004), but in other tissues evidence for a stem cell niche is still lacking. For example hematopoietic stem cells seem to be distributed within the whole bone marrow (Zhang et al., 2003). Thus, especially in hematopoiesis stemness is ensured by a tight regulation of cell proliferation properties at early stages of the hierarchy. Very recent publications show, that hematopoietic stem and progenitor cells reside in distinct niches and the regulation of Cxcl12 seems key for self renewal maintenance (Ding and Morrison, 2013; Greenbaum et al., 2013). To some extent, the deviation of in vitro and in vivo results might be explained by experimental difficulties in providing such a highly regulated environment.

The lack of an exact phenotypic stem cell definition leads to a functional determination by transplanting cell lines in stem insufficient mice. The stemness of this cell line is then determined by its capability to repopulate the tissue. However, the incomplete knowledge leads to an uncertainty in other very basic parameters, as for example the exact number of actively dividing stem cells. In human hematopoiesis the estimates vary from approximately 400 to 10^5 actively dividing stem cells (Dingli and Pacheco, 2006; Buescher et al., 1985; Rufer et al., 1999; Roeder et al., 2006). Furthermore, experimental results in mice suggest temporal fluctuating stem cell numbers. The stem cell pool expands with age, whereas the proliferation rate of stem cells declines (Rossi et al., 2005, 2008). So far, this effect is not conclusively shown in humans, but there is no evidence to expect the opposite. Indeed, in chapter 5, we compare results from a mathematical model to experimental data of in vivo telomere shortening, and find hints of an increasing stem cell pool size caused by symmetric cell divisions.

The further differentiation steps within hierarchies are tissue dependent. In the hematopoietic system, stem cells proliferate into progenitor cells, which give rise to the two major subtrees of hematopoiesis, myeloid and lymphocyte progenitors. Within these subtrees a fixed amount of cell differentiations are necessary to obtain all mature blood cells, reaching from red and white blood cells, to macrophages, platelets and others, see Fig. 1.3.

1.2.2 Cancer biology

In the previous subsection we introduced the concept of hierarchically organized tissue structures. These tissues are highly controlled and regulated. However, for various reasons cells potentially escape these control mechanisms and start increased cell proliferation. Such uncontrolled cell behaviors, that give a fitness advantage to a cell, are considered cancerous (Hanahan and Weinberg, 2000). The reasons, why cells escape these control mechanisms are manifold. A few cases are known, where viruses (oncovirus) alter cell proliferation properties to be cancerous. The most prominent example is the discovery of the papillomavirus and its potential to cause cervical cancer (Zur Hausen, 2002). This discovery allowed the development of a vaccination, yielding the nobel price in medicine to zur Hausen, Barre-Sinoussi and Montagnier in 2008. Another example is the statistical correlation of Burkitt's

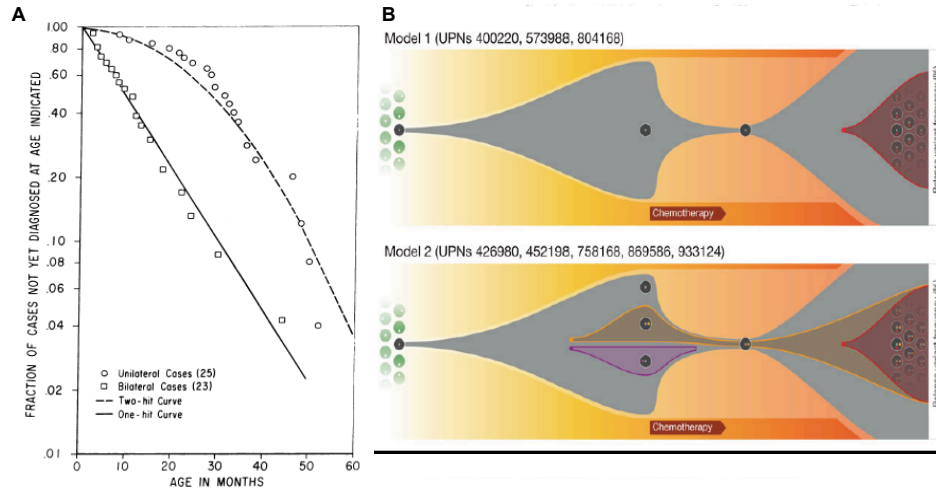


Figure 1.4: **A Discovery of tumor suppressor genes by Knudson in 1971.** It needs in general two hits, to deactivate both alleles of a tumor suppressor gene. Where the first activation is (almost) neutral, the second hit alters the phenotype of a cell. The time for two hits scales $\propto t^2$ (upper dashed line). However, rarely people carry one mutated allele due to a germline mutation. In this case one hit is sufficient and the time scales $\propto t$ (lower line). The figure was taken from (Knudson, 1971). **B Clonal evolution of single patients with acute myeloid leukemia.** Cells accumulate different mutations over time. Thus the clonal composition of patients vary. The diversity influences the treatment outcome. In cases of relapses, the secondary disease is driven by different clones and is difficult to predict. The figure was taken from (Ding et al., 2012).

lymphoma (an aggressive B-cell lymphoma) and the Epstein-Barr virus (Thorley-Lawson and Gross, 2004). However, the underlying cause is the trigger of mutations in specific oncogenes (here Myc-mutation) by these viruses (Allday, 2009). This lead to the modern view, that cancer is a genetic disease, caused by mutational hits in specific cancer initiating genes. Often these genes control cell proliferation or cell senescent properties and mutations can cause an escape of these control mechanisms (Vogelstein and Kinzler, 2004).

In general, multiple mutational hits within a cell are necessary for cancer progression. This lead to the multi-stage theory of carcinogenesis (Armitage and Doll, 1954). It assumes that mutations accumulate within cells, and the interplay of these mutations can provide a fitness advantage to those cells. In its basics the theory is still widely accepted today (Hornsby et al., 2007). Originally based on statistical arguments, it was possible to give a quantitative explanation of the appearance

of cancers. An early success of the multi-stage theory was the explanation of the statistical observations in retinoblastomas (blastoma developing in the retina) by Knudson in 1971 (Knudson, 1971). A mutation in a single gene is sufficient to cause a retinoblastoma. However, the mutation is recessive, thus mutations in both gene copies (alleles) are necessary for cancer initiation and effectively two mutation hits are needed. Such recessive genes, that need two hits to be cancerous, are called tumor suppressor genes (a prominent tumor suppressor gene is p53, mutated in approximately 50% of all human cancers) (Weinberg et al., 1991; Nowak, 2006a). By comparing data from patients with a germ line mutation to patients with somatic mutations, Knudson showed empirically, that the time for the first hit is proportional to t and proportional to t^2 for the second hit, see Fig. 1.4 A. Later, mathematical models revealed, that the scaling of these waiting times depend on population size. The second hit is proportional to t^2 for small population sizes to t for intermediate sizes and proportional to t^2 for large population sizes. In the last case these two hits are not rate limiting (Komarova et al., 2003; Nowak, 2006a).

However, tumor suppressor genes need to be deactivated to potentially lead to cancer. In contrast for many genes a single mutational hit is sufficient to alter the proliferation properties of a cell. Such genes are called oncogenes, a concept introduced by Bishop and Varmus (Varmus, 1984; Schwab et al., 1984). An important and well understood oncogene is the BCR-ABL translocation of the 9th and 22nd chromosome in hematopoietic stem cells. This fusion gene codes for a tyrosine kinase, altering the cell proliferation properties. This results in an accumulation of immature white blood cells (Lugo et al., 1990), and patients are diagnosed with chronic myeloid leukemia (Epstein et al., 1988; Bartram et al., 1983; Michor et al., 2005). However, in most cases a single mutated gene is insufficient to cause cancer, but mutations in several oncogenes are necessary. The exact number is tissue and cancer specific. Approximations reach from 1 (in some blood cancers) to 30 mutated genes in some colon cancers (Greenman et al., 2007; Bozic et al., 2010). The classification of oncogenes increased rapidly with the recent progress in sequencing techniques. The efficient sequencing of DNA of patients with different types of cancer revealed complicated mutational landscapes (Sjöblom et al., 2006; Bozic et al., 2010; Shah et al., 2012; Ding et al., 2012). Besides some known oncogenes (for example Myc, BCR-ABL, PML-RARA, PI3-A) scans show a high diversity. In some cancers one might not find a single common mutation, although over 100 mutated genes can be found in samples of single patients (Lee et al., 2010; Ding et al., 2012).

Presumably not all 100 mutations contribute to oncogenesis. Thus mutations are classified into driver and passenger mutations. Driver mutations are necessary for cancer initiation and are frequently found in patients. Passenger mutations are understood as (almost) neutral background mutations, unrelated to cancer progression (Stratton et al., 2009; Pleasance et al., 2009). However, the identification of driver mutations is difficult. In most cases a minor subset of a larger set of possible mutations seems sufficient for cancer progression. The mutational landscape strongly depends on the clonal history of patients (Ding et al., 2012), see Fig. 1.4 B and influences treatment strategies as well. An individualized cancer therapy is necessary and in some cases there are already promising developments towards this direction (Diaz Jr et al., 2012).

Besides the occurrence of passenger mutations, there are further complications in modeling cancer initiation and tumor progression, for example the hierarchical organization of tissues (Werner et al., 2011). The hierarchy is a natural protection mechanism against cancer, as the life time of most cells is limited, and potential cancerous mutations get washed out of the hierarchy. In chapter 2 we calculate solutions, describing such clonal waves within hierarchical tissues, and show the suppression of multiple mutations within hierarchies. However, such hierarchies increase the diversity of possible clonal dynamics. The stage of a mutation influences further dynamics. Some cancers clearly occur in stem cells, others evolve at later stages of the hierarchy. Especially latter case lead to the postulation of cancer stem cells, which emerge either by mutation within stem cells, or the gain of stemness within non stem cells by certain mutations (Jordan et al., 2006; Kelly et al., 2007). Much efforts have been done recently to identify stem like sub populations in cancerous tissues. An important question to modelers is, if a successful therapy requires the erasure of all cancer stem cells, or the extinction of a subset is sufficient (Lenaerts et al., 2010, 2011; Glauche et al., 2012). Given the complexity of the underlying processes, it seems difficult to obtain predictions by theory only. But the investigation of population dynamics within hierarchically organized tissues might provide one step towards a better understanding of cancer progression and cancer diversity.

1.2.3 Molecular targeted treatment strategies

We discussed the clonal theory of cancer initiation in the former subsection. A hallmark of most cancers is the high clonal diversity within patients, which complicates

cancer therapies significantly (Greaves and Maley, 2012). Treatment strategies are as diverse as cancer, and discussing all might be a dissertation by itself. Here we focus on the treatment of chronic myeloid leukemia, as this raises questions, we address in chapter 4. Chronic myeloid leukemia is caused by a translocation of the ends of the 9th and 22nd chromosome within a hematopoietic stem cell. This translocation results in a functional fusion gene (BCR-ABL) that codes for a tyrosine kinase. The activation of the resulting receptor is independent of cellular signaling molecules, causing an unnatural accumulation of immature white blood cells in the blood stream of patients (Epstein et al., 1988; Bartram et al., 1983; Michor et al., 2005). Chronic myeloid leukemia remains in a chronic phase for 3-7 years, followed by a transition into a blast crisis. This transition is lethal if untreated (Wetzler et al., 1993; Sawyers, 1999). Ten years ago chemotherapy followed by bone marrow transplantation was the standard therapy, imposing significant risks to patients. However, the treatment protocol changed with the invention of molecular targeted drugs, amongst them Imatinib (Druker et al., 2006), Nilotinib (Weisberg et al., 2005), Dasatinib (Shah et al., 2004) and Bosutinib (Cortes et al., 2011).. These drugs take advantage of the distinct spatial structure of the BCR-ABL encoded tyrosine kinase. They specifically occupy the binding sites of these receptors and inhibit the initiation of further signaling cascades within the cell, lowering the proliferation capabilities of BCR-ABL positive cells significantly (Komarova and Wodarz, 2005; Michor et al., 2005; Dingli et al., 2008a; Traulsen et al., 2010; Werner et al., 2011). This leads to a total remission in approximately 95 % of all treated patients, most of them staying in remission since at least 5 years (Kantarjian et al., 2007). Since these molecules target BCR-ABL positive cells very specifically and have lesser effects on healthy tissues, this was the first successful example of such targeted therapies, raising hopes to find similar drugs for other cases. However, there are drawbacks. It is unknown, if the medication can be stopped and if so, when (Lenaerts et al., 2010, 2011). In single cases the disease relapses, often due to an evolving drug resistant clone (Hochhaus et al., 2002; Werner et al., 2011). One tries to tackle this problem by combining different drugs with slightly different binding mechanisms (Komarova and Wodarz, 2005). Unfortunately mutants exist, resistant to all currently available drugs. Thus scientists are interested in a better understanding of mechanisms and dynamics of resistance evolution. In chapter 4 we discuss a resistance inducing experiment. By applying a minimalistic mathematical

model we are able to infer basic model parameters. Interestingly the results hint towards an interplay of genetic and epigenetic effects.

1.2.4 Telomeres and telomerase

Telomeres are noncoding DNA structures expressed at the ends of all eukaryotic chromosomes. In vertebrates they consist of numerous repeats of the amino acids TTAGGG (Blackburn et al., 1991). Telomeres prevent the loss of coding DNA regions during cell proliferation and are essential to ensure chromosomal integrity (Hande et al., 1999; d’Adda di Fagagna et al., 2001; Boukamp et al., 2005; Brümendorf and Balabanov, 2006). During cell proliferation telomere repeats shorten because of the inability of the DNA polymerase to fully replicate the 3’ end of chromosomes. This is known as the end replication problem and as a result a progressive shortening of telomeres can be observed in many aging tissues (Levy et al., 1992; Allsopp et al., 1995). Accelerated telomere shortening is associated with genetic instability and increased cell senescence. It accompanies diseases, for example cancers, HIV and increased stress (Epel et al., 2004; Brümendorf et al., 2006). Further telomere shortening may explain the limited proliferation capability of cell lines (Hayflick limit) and thus are proposed to play an important role in general aging processes of organisms (Blasco, 2007). The observed loss of telomere repeats during cell proliferation motivated the idea, that telomeres might be a mitotic clock, mirroring the replication history of cells (Vaziri et al., 1994; Morrison et al., 1996). But telomeres are susceptible to oxidative stress and stochasticity has an impact on telomere dynamics (von Zglinicki, 2002). In chapter 5 we show that stochasticity is important at the level of single telomeres, but deterministic modeling can describe telomere dynamics on the cell and the tissue level.

The shortening of telomeres can be counteracted. In vertebrates, the end replication problem is solved by the enzyme telomerase. For its discovery, Greidel, Szostak and Blackburn were awarded with the Nobel price for medicine in 2009. Telomerase is a reverse transcriptase, that can add additional TTAGGG repeats to the 3’ end of chromosomes and thus is able to counteract accelerated telomere shortening (Blackburn et al., 1991). Telomerase is highly expressed on the germline and on the stem cell level, but also many cancerous tissues show increased telomerase activity (Shay and Bacchetti, 1997; Sharpless et al., 2004). That provides one possible explanation

how cancer cells escape cell senescence. Therefore a combination of traditional cancer therapies with a telomerase blockage were proposed, unfortunately with limited success. However, telomerase is insufficiently expressed in most tissues to avoid a progressive shortening of telomeres with age (Harley et al., 1990; Vaziri et al., 1994; Rufer et al., 1999). The complete lack of telomerase can lead to severe diseases, such as anemias or premature aging (dyskeratosis congenita). Anemias result from a stem cell pool exhaustion as a consequence of increased cell death due to faster telomere shortening. This leads to an underproduction of blood cells, causing bone marrow failure and immunosenescence. Although diseases occurring from telomere dysfunctions are clinically diverse, they are closely related and a compact classification as telomere syndromes was proposed (Ball et al., 1998; Brümmendorf et al., 2001; Armanios and Blackburn, 2012). However, the underlying causes for the telomere dysfunction may differ, resulting in expected observable differences in the dynamics of telomere shortening. In chapter 5, we address telomere dynamics. We derive expected telomere distributions for different cell proliferation properties and diseases. A comparison to experimental data allows us to deduce some basic system parameter and to get insights in the *in vivo* stem cell dynamics.

1.3 Stochastic dynamics

Natural biological systems of proliferating cells are stochastic (Goel and Richter-Dyn, 1974; Abkowitz et al., 1996). The source of this stochasticity is manifold, reaching from mutations, resource competition, differentiation processes, clonal evolution, cell senescence and many others. Our aim is to describe the dynamics within such biological systems. However, in general it is difficult to catch all these effects in solvable closed mathematical models, which necessitates approximations of such system (McKane and Newman, 2004). These approximations can take the form of a deterministic description of the underlying stochastic dynamics, a reduction of the systems complexity by neglecting effects of seemingly lesser influence, or stochastic computer based simulations of biological processes.

1.3.1 Gillespie algorithm

In some cases, it is advantageous to implement a stochastic model with kinetics that is close to the underlying biological system. To understand the dynamics it is

helpful to run exact individual based stochastic simulations of the system. However, that immediately confronts us with the problem of how to implement an algorithm, that reproduces the exact stochastic trajectories.

For example in chapter 2 we describe the dynamics of cells in hierarchically organized tissues. Effectively, such systems contain a varying number of different cell types, each cell type with dissimilar cell proliferation properties. This difference could be differentiation probabilities or proliferation rates (Werner et al., 2011). The question becomes: Given a certain number of different types and given cell proliferations are discretized in a sense that there is exactly one proliferating cell at a single time, which cell is next to proliferate and which time does it take until the next cell proliferation? To implement such an algorithm, we can utilize results of stochastic chemical kinetics. The Gillespie algorithm and variants of it are used to describe the stochastic kinetics of chemical reactions (Gillespie, 1976, 2007). Here, we shortly describe the algorithm and derive a variant, that allows us to perform stochastic simulations of many type cell populations, used in chapters 2 and 5.

Assume $X_i(t)$ denotes the number of cells of type $i \in (1, \dots, n)$ at time t . A rigorous derivation of the underlying chemical master equation can be found in (Gillespie, 1992). Our goal is to simulate stochastic trajectories of the state space vector $\mathbf{X}(t) = (X_1(t), \dots, X_n(t))$, given some initial condition $\mathbf{X}(t_0) = \mathbf{x}_0$. Therefor we follow the argumentation in (Gillespie, 1976, 2007). We assume M possible reactions (R_1, \dots, R_M) in our system, for example different mutation and proliferation pathways. The key is the probability

$$P(\tau, j) d\tau = \text{probability that the next reaction will be within the} \\ \text{time interval } (t + \tau, t + \tau + d\tau) \text{ and will be a type} \\ j \in (1, \dots, M) \text{ reaction.}$$

To derive an expression for $P(\tau, j)$ we need to specify the possible state variables h_j of our system,

$$h_j = \text{number of distinct reaction combinations for reaction } R_j \\ \text{present at time } t.$$

For example if we assume reaction j to be $X_j \rightarrow X_k$, we have $h_j = X_j$. Assuming reaction j to have a rate parameter c_j , the probability that a reaction will occur in

the time interval $d\tau$ to first order is simply $h_j c_j d\tau$. Given, that $P_0(\tau)$ is the probability, that no reaction will occur within the time interval $(t, t + \tau)$, the probability of having a reaction j in this time interval is

$$P(\tau, j) d\tau = P_0(\tau) h_j c_j d\tau \quad (1.1)$$

To find an expression for P_0 , we assume the interval $(t, t + d\tau)$ to be divided into K subintervals of equal length, given by $\Delta = \tau/K$. Using the multiplication theorem for probabilities, the probability of having no reaction in the interval $(t, t + \Delta)$ is

$$P_0(\Delta) = \prod_{l=1}^M (1 - h_l c_l \Delta + o(\Delta)) = 1 - \sum_{l=1}^M h_l c_l \Delta + o(\Delta). \quad (1.2)$$

As the intervals are of equal length, the probability is the same for every single of the K intervals. This allows us to write

$$P_0(\tau) = \left[1 - \sum_{l=1}^M \frac{h_l c_l \tau}{K} + o\left(\frac{1}{K}\right) \right]^K. \quad (1.3)$$

Choosing the time intervals infinitesimally small, $K \rightarrow \infty$, we get

$$P_0(\tau) = \lim_{K \rightarrow \infty} \left[1 - \sum_{l=1}^M \frac{h_l c_l \tau}{K} + o\left(\frac{1}{K}\right) \right]^K = \exp \left[- \sum_{l=1}^M h_l c_l \tau \right], \quad (1.4)$$

where we used $\lim_{n \rightarrow \infty} (1 - x/n)^n = \exp(-x)$ in the last step. Putting this into (1.1) gives the exact expression for the reaction probability density function

$$P(\tau, j) = h_j c_j \exp \left[- \sum_{l=1}^M h_l c_l \tau \right]. \quad (1.5)$$

This allows us to construct exact stochastic simulations. There are different equivalent implementations. Maybe the simplest is the so called direct method. The method works as follows: Draw two random numbers ζ_1 and ζ_2 from a uniform distribution in the interval $[0, 1]$. The time until the next reaction is,

$$\tau = \frac{1}{\sum_{l=1}^M h_l c_l} \ln \left[\frac{1}{\zeta_1} \right], \quad (1.6)$$

and the next reaction to occur is the first j fulfilling the condition

$$\sum_{l=1}^j h_l c_l > \zeta_2 \sum_{s=1}^M h_s c_s. \quad (1.7)$$

With this, the time until the next reaction and the next reaction to take place are specified and the algorithm of the direct method can be implemented as follows:

- (i) Initialize the system according to the initial conditions.
- (ii) Calculate all h_l and $\sum_{l=1}^M h_l c_l$.
- (iii) Generate values for τ and j using the equation (1.6) and (1.7).
- (iv) Update the system by the last reaction and return to step (ii).

A small variation of this is the method of the first moments, that however can be shown to be equivalent to the direct method (Gillespie, 2007). If there are M possible reactions in our system, we draw ζ_1, \dots, ζ_M in the interval $[0, 1]$ uniform distributed random numbers, to create in total M time intervals,

$$\tau_j = \frac{1}{h_j c_j} \ln \left[\frac{1}{\zeta_j} \right]. \quad (1.8)$$

The smallest τ_k , $k \in \{1, \dots, n\}$, determines the next reaction k to take place. In chapters 2 and 5 we will use the method of the first moment to simulate stochastic trajectories of proliferating cells in hierarchical organized tissues. There different compartments i contain a certain number of cells N_i , that proliferate with a proliferation rate r_i . The cell in compartment i to proliferate next is then determined by the smallest time τ_i

$$\tau_j = \frac{1}{N_j r_j} \ln \left[\frac{1}{\zeta_j} \right]. \quad (1.9)$$

1.3.2 Moran process

The Moran process is a classical model of population genetics (Moran, 1962) and belongs to the class of stochastic birth-death processes. As such, a discrete finite population of fixed size progresses in time. In general there are n types in a population of size N . During one time step one type reproduces and some other type dies, enforcing constant population size N . Assume the number of type A_j ($1 \leq j \leq n$ and $n < N$), individuals is i_j ($0 \leq i_j \leq N$). During one time step, the number i_j of type A_j can either increases by one, stays constant or decreases by one, which occurs with transition probabilities $T_{i_j}^+$, $T_{i_j}^0$ and $T_{i_j}^-$ respectively. The concrete implementation of these transition probabilities defines the properties of the Moran process at hand. Usually, the transition probabilities are chosen such, that they only depend on the population composition, $T = T(i_1, \dots, i_n)$, where $\sum_{l=1}^n i_l = N$, and thus are fully determined by the current state of the stochastic process (Traulsen and Hauert, 2009).

We assume, that the process $X = X_t$ has a finite state space $I = \{i_1, \dots, i_n\}$ and proceeds in some discrete time $t = 0, 1, 2, \dots$. The process is a Markov chain of k -th order, if the probability distribution P of the stochastic process X fulfills

$$P(X_{t+1} = I_{t+1} | X_t = I_t, \dots, X_{t-k+1} = I_{t-k+1}). \quad (1.10)$$

More informally in words: the probability distribution of the stochastic process at time $t + 1$ depends on the last k previous states of the process (Kemeny and Snell, 1960). The process fulfills the Markov property and is called Markovian if condition (1.10) reduces to

$$P(X_{t+1} = I_{t+1} | X_t = I_t). \quad (1.11)$$

Such processes are memoryless, as the probability distribution of the following state is fully determined by the present state of the process and is independent of the previous transitions. A Moran process with transition probabilities $T = T(i_1, \dots, i_n)$ fulfills the Markov property and thus falls into the class of Markov processes (Moran, 1958).

In the following we want to collect some important results that are obtained for the well understood two type Moran process (Nowak, 2006a; Traulsen and Hauert, 2009). Assume, a population of size N at time t is composed of i type A individuals and $N - i$ type B individuals respectively. Each time step, one individual is chosen to reproduce and one individual is chosen to die. Important properties of the process are then the fixation probability Φ_i^A , and the fixation time τ_i^A of type A (or Φ_{N-i}^B , and τ_{N-i}^B for type B). As the arguments can be exchanged, it is sufficient to focus on a single type. In the following we aim to calculate the fixation probability Φ_i , meaning the probability of i type A individuals to take over the whole population. Obviously, if the probability to decrease the number of A individuals from N to $N - 1$ is $T_N^- > 0$ one has $\Phi_i = 0$. Thus, for a non vanishing fixation probability it is necessary to assume $T_N^- = 0$, which, due to the constant population size, implies $T_{0,B}^+ = 0$ for the transition probability of 0 type B individuals. Usually, the process is assumed to be symmetric and thus we have the additional properties $T_0^+ = 0$ and $T_{N,B}^- = 0$. This gives at least two absorbing states $\Phi_0 = 0$ and $\Phi_N = 1$. For the probability of fixation of an intermediate state i we can write the recurrence relation (Traulsen and Hauert, 2009)

$$\Phi_i = T_i^- \Phi_{i-1} + (1 - T_i^- - T_i^+) \Phi_i + T_i^+ \Phi_{i+1}. \quad (1.12)$$

By rearranging terms in (1.12) we obtain

$$T_i^- \underbrace{(\Phi_i - \Phi_{i-1})}_{\xi_i} = T_i^+ \underbrace{(\Phi_{i+1} - \Phi_i)}_{\xi_{i+1}}, \quad (1.13)$$

yielding the recurrence relation $\xi_{i+1} = \gamma_i \xi_i$, with the abbreviation $\gamma_i = T_i^- / T_i^+$. Taking the absorbing state $\Phi_0 = 0$ into account, the solution to this equation is

$$\xi_i = \Phi_1 \prod_{j=1}^{i-1} \gamma_j. \quad (1.14)$$

Next we note that

$$\sum_{l=1}^N \xi_l = \Phi_1 - \Phi_0 + \Phi_2 - \Phi_1 + \dots + \Phi_N - \Phi_{N-1} = 1 \quad (1.15)$$

is a telescoping series, as all terms except the first and the last cancel each other out. Making use of the absorbing states $\Phi_0 = 0$ and $\Phi_N = 1$ allows us to obtain the result. Finally by combining (1.14) and (1.15) we can derive an expression for the fixation probability Φ_1

$$1 = \sum_{l=1}^N \xi_l = \Phi_1 \sum_{l=1}^N \prod_{j=1}^{l-1} \gamma_j = \Phi_1 \left(1 + \sum_{l=1}^{N-1} \prod_{j=1}^l \gamma_j \right), \quad (1.16)$$

where in the last step we wrote the empty product $\prod_{j=N}^{N-1} \gamma_j = 1$ in the sum explicitly. Thus the probability that a single type takes over the whole population of fixed size N is given by

$$\Phi_1 = \frac{1}{1 + \sum_{l=1}^{N-1} \prod_{j=1}^l \gamma_j}. \quad (1.17)$$

Expression (1.17) is an important property of biological systems. New types within a population are often considered to occur via mutation and thus are rare (a single type) initially, and their probability of fixation is often of interest. However, we can also derive an expression for the fixation probability Φ_i , given by

$$\Phi_i = \sum_{l=1}^i \xi_l = \Phi_1 \sum_{l=1}^i \prod_{j=1}^{l-1} \gamma_j = \frac{1 + \sum_{l=1}^{i-1} \prod_{j=1}^l \gamma_j}{1 + \sum_{l=1}^{N-1} \prod_{j=1}^l \gamma_j}. \quad (1.18)$$

Note in the special case $T_i^+ = T_i^-$, we have $\gamma_i = 1$ and the expression (1.18) simplifies to $\Phi_i = i/N$. If $i = 1$ the fixation probability becomes $\Phi_1 = 1/N$. This situation is called neutral selection and it corresponds to an unbiased random walk (Kimura, 1983; Nowak, 2006a). Neutral selection is an important concept in evolutionary

biology, as many processes are believed to take place at neutral or close to neutral conditions (Kimura, 1983, 1991). Further, neutral selection is used as a benchmark and often the deviation of fixation probabilities and fixation times are compared to the neutral case.

Another very important property in biological processes is the fixation time τ_i . Even if the fixation probability might be one, the process could proceed on very different time scales and thus appears different in biological systems (Altrock and Traulsen, 2009a,b; Altrock et al., 2010, 2012). A prominent example is oncogenesis. Although a mutation might be cancerous and can be proven to fixate with certainty in a tissue, it might take longer than the expected life time of a healthy person and thus is effectively harmless. The unconditional average fixation time τ_i can be formulated by a recurrence relation again (Taylor et al., 2006; Traulsen and Hauert, 2009)

$$\tau_i = 1 + T_i^- \tau_{i-1} + (1 - T_i^- - T_i^+) \tau_i + T_i^+ \tau_{i+1}, \quad (1.19)$$

where $\tau_0 = \tau_N = 0$. Again we can rearrange the terms to

$$\underbrace{\tau_{i+1} - \tau_i}_{\zeta_{i+1}} = \gamma_i \underbrace{(\tau_i - \tau_{i-1})}_{\zeta_i} - \frac{1}{T_i^+} \quad (1.20)$$

resulting in the recurrence relation $\zeta_{i+1} = \gamma_i \zeta_i + 1/T_i^+$. This can be solved iteratively

$$\zeta_i = \tau_1 \prod_{l=1}^{i-1} \gamma_l - \sum_{k=1}^{i-1} \frac{1}{T_k^+} \prod_{l=k+1}^{i-1} \gamma_l. \quad (1.21)$$

Further we observe $\sum_{l=2}^N \zeta_l = -\tau_1$ and thus arrive at the following expression

$$\tau_1 = - \sum_{l=2}^N \zeta_l = -\tau_1 \sum_{l=1}^{N-1} \prod_{k=1}^l \gamma_k + \sum_{l=1}^{N-1} \sum_{k=1}^l \frac{1}{T_k^+} \prod_{m=k+1}^l \gamma_m. \quad (1.22)$$

Rearranging terms finally gives the fixation time

$$\tau_1 = \frac{1}{\Phi_1} \sum_{l=1}^{N-1} \sum_{k=1}^l \frac{1}{T_k^+} \prod_{m=k+1}^l \gamma_m. \quad (1.23)$$

We derived expressions for fixation probability (1.18) and fixation time (1.23) of a two type Moran process without mutations. These expressions are very general and only depend on the transition probabilities T^- and T^+ . There are a few standard implementations of transition probabilities describing different situations in biological systems. The simplest case are constant transition probabilities, where

$T_i^+ = T^+ = r_1$ and $T_i^- = T^- = r_2$. This implies that $\gamma = T^-/T^+ = r_2/r_1 = r$, where $r \in \mathbb{R}_{\geq 0}$, is a constant number. This is called the constant selection case (Nowak, 2006a). Plugging these transition probabilities into (1.18) leaves us with two geometric series to be solved, and the fixation probability simplifies to

$$\Phi_i = \frac{1 - r^i}{1 - r^N}. \quad (1.24)$$

Note, for $r = 1$ by applying the rule of l'Hospital we obtain

$$\lim_{r \rightarrow 1} \frac{1 - r^i}{1 - r^N} = \lim_{r \rightarrow 1} \frac{ir^{i-1}}{Nr^{N-1}} = \frac{i}{N}, \quad (1.25)$$

which is the former result we derived above for neutral selection. We find $\Phi_i = 1$ only if $r = 0$, which is fulfilled for $T^- = 0$ and $0 < T^+ \leq 1$. Note that in all other cases $\Phi_i < 1$, even in the limit of large N . However, either one of both types will reach fixation in the long run.

The situation becomes slightly more complicated if the transition probabilities become state dependent. Assume there are i type A individuals and $N - i$ type B individuals. A standard way to introduce state dependency is

$$T_i^+ = \frac{if_i}{\underbrace{if_i + (N-i)g_i}_{\text{birth of A}}} \underbrace{\frac{N-i}{N}}_{\text{death of B}} \quad (1.26a)$$

$$T_i^- = \frac{(N-i)g_i}{\underbrace{if_i + (N-i)g_i}_{\text{birth of B}}} \underbrace{\frac{i}{N}}_{\text{death of A}}, \quad (1.26b)$$

where f_i and g_i are fitnesses of type A and B . We explain the exact concept of fitness and its implementation below. The logic of above terms is as follows. To increase the number of type A individuals by one, an A type needs to reproduce and a B type needs to die. Usually the birth is assumed to be proportional to the types fitness, where death is assumed to be random. This implementation implies selection acting on reproduction but not on death (Nowak, 2006a). However, this is an arbitrary choice and fitness dependence could be implemented in death terms as well (Altrock and Traulsen, 2009a).

Fitness in an traditional evolutionary framework is interpreted as an abstract reproduction capacity of a type (Ariew and Lewontin, 2004). Higher fitness corresponds to potentially higher number of offspring and thus a faster spread of this type within the population. Note, if fitness of type A and B are assumed to be constant numbers $f_i = \alpha$ and $g_i = \beta$, we find $\gamma_i = T_i^-/T_i^+ = \beta/\alpha = r$. This is the case of

constant selection we discussed above. So called frequency dependent selection was originally introduced in classical game theory and was later adapted to evolutionary game theory by John Maynard Smith and others (von Neumann and Morgenstern, 1944; Maynard Smith, 1974, 1982). First we have to implement a game between type A and B . Assume types A and B carry fixed strategies, that can be genetically in a biological sense or behavioral in an economic or social context. The strategies are determined by the payoff matrix of a game that takes the form

$$\begin{array}{cc} & \begin{array}{cc} A & B \end{array} \\ \begin{array}{c} A \\ B \end{array} & \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \end{array} \quad (1.27)$$

where a, b, c, d are real numbers called payoffs. The matrix has to be read as follows: if a type A individual interacts with another type A individual both will obtain payoff a , if type A interacts with type B , type A will obtain b and type B obtains c . Type B individuals interacting with type B individuals obtain d . Type interactions are specified in concrete models and depend on population structure (Nowak and May, 1992; Lieberman et al., 2005; Tarnita et al., 2009). However, a common approach is a well mixed population, where the probability to meet another type is given by its frequency (Nowak, 2006a). The payoff matrix above captures the special but important case of two players and two strategies. If we increase the number of strategies, the payoff matrix increases in size (3×3), an increase of players increases the dimension of the matrix (for example $2 \times 2 \times 2$ for the case of 2 strategies and 3 players) (Gokhale and Traulsen, 2010). In general a game between d players and n strategies can be captured by a (n, d) tensor. But so far we introduced payoffs without any connection to the actual dynamics of the stochastic process. This connection is achieved by a mapping of these payoffs into fitness (coefficients f_i and g_i in (1.26a) and (1.26b) respectively). Therefore we need to define a fitness mapping function, that is to some extend arbitrary again (Wu et al., 2010). We just may assume the function to be smooth, differentiable and monotonically decreasing for decreasing payoffs. However, traditionally a linear fitness mapping of the form

$$f_i = 1 - w + w \underbrace{\frac{a(i-1) + b(N-i)}{N-1}}_{F_i} \quad (1.28a)$$

$$g_i = 1 - w + w \underbrace{\frac{ci + d(N-i-1)}{N-1}}_{G_i}, \quad (1.28b)$$

is chosen, where 1 is the baseline fitness, $w \in [0, 1]$ is a control parameter that represents selection intensity and F_i, G_i are the expected payoffs of type A and type B individuals in a well mixed population (Nowak, 2006a). If $w = 0$, the fitnesses are independent of the game and we recover the case of neutral selection. For $w \rightarrow 0$ the contribution of the game becomes minimal, which generally is characterized as weak selection (Wild and Traulsen, 2007; Wu et al., 2010). The dynamical patterns for frequency dependent selection are more versatile compared to the constant selection case. A common question to ask is, given a certain set of strategies, which strategy should I play to maximize my personal fitness. In economics this would mean to maximize profit, in biology one could interpret it as evolutionary superior strategy that produces most offspring. We do not discuss this in too much detail, as this is not the aim of this thesis. We only want to collect a few basic facts.

The question of which strategy to chose is to some extend answered by the concept of the Nash equilibrium (Nash, 1950). It states, that a strategy is Nash, if any divergence from this strategy gives at most the same payoff. It is strict Nash, if a divergence always decreases the payoff. However, Nash equilibria do not necessarily maximize fitness on a population level. This is easily demonstrated by a minimalistic example. Consider the following payoff matrix

$$\begin{array}{cc} & \begin{array}{cc} C & D \end{array} \\ \begin{array}{c} C \\ D \end{array} & \begin{pmatrix} 2 & 0 \\ 3 & 1 \end{pmatrix} \end{array}. \quad (1.29)$$

A player that choses strategy C obtains a payoff of 2, if he meets another player with strategy C and a payoff of 0, if he meets a player with strategy D . A player with strategy D obtains a payoff of 3 by meeting a player with strategy C and 1 by meeting another D . In this situation it is always advantageous to play D and the only Nash equilibrium in this case is all D , thus everyone obtains payoff 1. However, if everyone played C the payoff per player is 2 and the populations payoff is higher. But this state is invaded by players with strategy D and the equilibrium shifts towards all D . This problem became famous as the Prisoners Dilemma and much research is done on additional mechanisms that are required to lead the dynamics towards the populations fitness maximum (Hamilton, 1964; Hauert, 1999; Macy and Flache, 2002; Hauert and Doebeli, 2004; Nowak, 2006b). Nash proved, that every game has at least 1 Nash equilibrium (if mixed strategies are allowed). We give a

full characterization of possible equilibria of the N player two strategy case in the chapter 1.4.1.

1.3.3 Towards deterministic dynamics

In the previous section we described properties of the Moran process, a stochastic birth-death process with finite population size. In simple situations, important properties such as fixation probabilities and fixation times can be derived. However, often the underlying dynamics of the process are also interesting. For example in oncogenesis the dynamics can be key. Although a mutation might extinct with certainty, intermediate states could be lethal. It is convenient to describe the dynamics of a system by time continuous differential equations, often enabling analytical solutions. In the following we want to derive the time continuous description of the Moran process by a system size expansion (limit of large N). Formally we have to perform a Kramers-Moyal expansion of the Master equation (Kampen, 1997; Traulsen et al., 2005; Traulsen and Hauert, 2009). The Master equation in general is a set of differential equations, describing the temporal change of the probability distribution of a stochastic process with finite state space (Gillespie, 1992). For the Moran process we can write

$$\begin{aligned} P^{\tau+1}(i) - P^{\tau}(i) &= P^{\tau}(i-1)T^{+}(i-1) - P^{\tau}(i)T^{-}(i) \\ &\quad + P^{\tau}(i+1)T^{-}(i+1) - P^{\tau}(i)T^{+}(i), \end{aligned} \quad (1.30)$$

where $P^{\tau}(i)$ denotes the probability that the system is in state i at time τ and T^{\pm} are the transition probabilities introduced above. Performing the transformations $x = i/N$, $t = \tau/N$ and $\rho(x, t) = NP^{\tau}(i)$ we obtain

$$\begin{aligned} \rho\left(x, t + \frac{1}{N}\right) - \rho(x, t) &= \rho\left(x - \frac{1}{N}, t\right)T^{+}\left(x - \frac{1}{N}\right) \\ &\quad + \rho\left(x + \frac{1}{N}, t\right)T^{-}\left(x + \frac{1}{N}\right) \\ &\quad - \rho(x, t)T^{-}(x) - \rho(x, t)T^{+}(x). \end{aligned} \quad (1.31)$$

The next step is a Taylor expansion of the transition probabilities and the probability densities around x and t , yielding

$$\begin{aligned}\rho\left(x, t + \frac{1}{N}\right) &= \rho(x, t) + \frac{1}{N} \frac{\partial}{\partial t} \rho(x, t) + O\left(\frac{1}{N^2}\right) \\ \rho\left(x \pm \frac{1}{N}, t\right) &= \rho(x, t) \pm \frac{1}{N} \frac{\partial}{\partial x} \rho(x, t) + \frac{1}{2N^2} \frac{\partial^2}{\partial x^2} \rho(x, t) + O\left(\frac{1}{N^3}\right) \\ T^\pm\left(x \pm \frac{1}{N}\right) &= T^\pm(x) \pm \frac{1}{N} \frac{\partial}{\partial x} T^\pm(x) + \frac{1}{2N^2} \frac{\partial^2}{\partial x^2} T^\pm(x) + O\left(\frac{1}{N^3}\right).\end{aligned}$$

Plugging these terms into (1.31) we obtain

$$\frac{\partial}{\partial t} \rho(x, t) = - \frac{\partial}{\partial x} \underbrace{[T^+(x) - T^-(x)]}_a \rho(x, t) + \frac{1}{2N} \frac{\partial^2}{\partial x^2} \underbrace{[T^+(x) - T^-(x)]}_{b^2} \rho(x, t),$$

which has the form of the well studied Fokker-Planck equation. Applying Ito's Lemma (Kampen, 1997) this can be transformed into a Langevin equation with multiplicative noise ζ

$$\dot{x}(t) = a(x) + b(x) \zeta. \quad (1.32)$$

In the limit $N \rightarrow \infty$ we have $b \rightarrow 0$ and thus we find

$$\dot{x}(t) = \lim_{N \rightarrow \infty} [T^+(x) - T^-(x)] = x(1-x) \left(\frac{f-g}{\phi} \right), \quad (1.33)$$

where $\phi = xf + (1-x)g$ is the average fitness of the population. Equation (1.33) is the well known replicator equation, frequently used in evolutionary game theory to describe the deterministic dynamics of a game (Taylor and Jonker, 1978; Zeeman, 1980; Page and Nowak, 2002a). Above calculations can be extended, allowing multiple types and non vanishing mutation rates, see (Traulsen et al., 2006). In chapter 4, we will use the Moran process and the replicator equation as deterministic approximation to describe the dynamics in a resistance inducing experiment.

1.4 Deterministic dynamics

So far, we have described stochastic approaches to capture biological systems. As these systems consist of a finite number of interacting agents, cells molecules etc. and due to numerous random influences, such as mutations or changing environments, it is natural to assume these dynamics to be stochastic. However, deterministic descriptions can offer interesting insights into these often very complicated systems, either as an approximation of large system size or by revealing general patterns

underlying stochastic fluctuations. A classical approach is the use of differential equations, to capture the deterministic dynamics of a system. Throughout this thesis we develop several model systems, describing different biological systems by applying different types of differential equations.

1.4.1 Replicator equation

In the former subsection 1.3.3 we derived the replicator equation for two possible strategies by a system size expansion of the Moran process. This equation can be generalized to n interacting types (Nowak, 2006a). We assume n strategies and denote with $\mathbf{x} = \{x_1, \dots, x_n\}$ the distribution of the population, where x_i is the frequency of strategy (or type) i . We call $f_i(\mathbf{x})$ the fitness of type i . The general replicator equation is given by

$$\dot{x}_i = x_i (f_i(\mathbf{x}) - \phi), \quad (1.34)$$

where ϕ is chosen such, that $\sum_{i=1}^n x_i = 1$. The parameter ϕ turns out to be the average fitness of the population, $\phi = \sum_i x_i f_i(\mathbf{x})$. The dynamics of this system can be studied on a simplex with n vertices, defined by all sets of frequencies that fulfill $\sum_{i=1}^n x_i = 1$. An important property of this system are its equilibria, which can be derived by solving the in general nonlinear system of equations $x_i (f_i(\mathbf{x}) - \phi) = 0$. The vertices of the simplex are either 0 or 1 and are trivial fixed points. However, of biological interest are often interior fixed points ($x_i > 0 \forall i$), as this provides a possible mechanism of maintaining biodiversity (Huang et al., 2012). Recently, it was shown, that for a d player game with n strategies, the maximum number of interior fixed points is $(d-1)^{n-1}$ (Gokhale and Traulsen, 2010; Han et al., 2012). Importantly, for any two player game, there is at most 1 interior fixed point, that either can be stable or unstable. In the special case of a two player two strategy game this leaves us with only 4 possible dynamical scenarios. These scenarios can easily be calculated by solving the system

$$\dot{x}_1 = x_1 (f_1(x_1, x_2) - \phi(x_1, x_2)) = 0 \quad (1.35a)$$

$$\dot{x}_2 = x_2 (f_2(x_1, x_2) - \phi(x_1, x_2)) = 0. \quad (1.35b)$$

Using $x_1 + x_2 = 1$ this system can be reduced to a single equation, given by

$$x_1 (1 - x_1) (f_1(x_1, x_2) - f_2(x_1, x_2)) = 0. \quad (1.36)$$

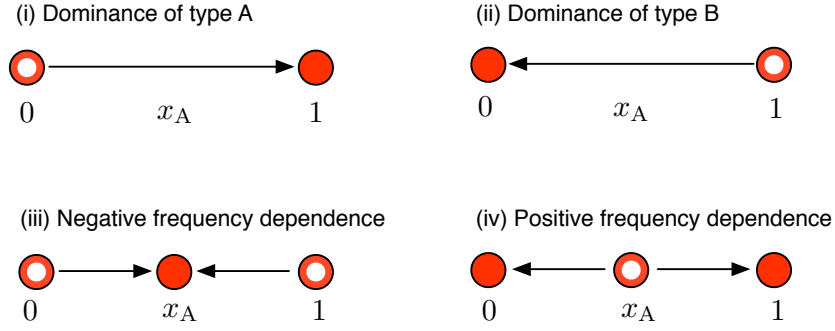


Figure 1.5: **All possible dynamical patterns of the two player two strategy scenario.** Here full circle represent stable equilibria and the direction of arrows give the direction of selection. See main text for a detailed discussion of all four cases.

Thus we find the trivial fixed points $x_1 = 0$ and $x_1 = 1$ at the vertices of the simplex (a line in this case). Further we assumed $f(x)$ to be strictly monotonic. Thus there is at most a single intersection point of f_1 and f_2 , leading to one potential interior fixed point if $f_1(x_1, x_2) = f_2(x_1, x_2)$. However, this fixed point can be asymptotically stable or unstable, resulting in two different dynamical patterns (Nowak, 2006a). Here, a fixed point is called stable if all eigenvalues of the Jacobian matrix are strict smaller than 1 (Browder, 1979). Referring to the implementation of the payoff matrix (1.27) and the fitness mapping functions (1.28a) and (1.28b), equation (1.36) can be written as

$$x_A (1 - x_A) [(a - b - c + d)x_A + b - d] = 0. \quad (1.37)$$

The interior fixed point becomes $x_A^* = \frac{a-b-c+d}{b-d}$. This equilibrium is stable if $d(f_1 + f_2)/dx_A < 0$, which is fulfilled for $a - b < c - d$. The 4 different dynamical scenarios are determined by the entries of the payoff matrix:

- (i) Dominance of type A, ($x_A \rightarrow 1$), if $a > c$ and $b > d$.
- (ii) Dominance of type B, ($x_A \rightarrow 0$), if $a < c$ and $b < d$.
- (iii) Negative frequency dependent selection (x_A^* is stable), if $a < c$ and $b > d$.
- (iv) Positive frequency dependent selection (x_A^* is unstable), if $a > c$ and $b < d$.

Here the terms negative or positive frequency dependent selection refer to a type, advantageous at either low or high frequencies respectively, see Fig. 1.5. The situation $a = b = c = d$ refers to the static case, where the system stays at its initial state.

We note, that equation (1.34) is closely related to some other important equations. If fitness is state independent $f_i(\mathbf{x}) = f_i$, the replicator equation corresponds to the quasi species equation without mutations. However, in chapter 4 mutations are necessary and we use a generalized quasi species equation that allows for mutations (Page and Nowak, 2002a). Further a transformation $x_i = y_i/(1 + y)$ converts the replicator equation into the Lotka-Volterra equation. An incorporation of mutations leads to the replicator mutator equation, given by

$$\dot{x}_i = \sum_{j=1}^n x_j f_j(\mathbf{x}) q_{ji} - x_i \phi. \quad (1.38)$$

Here q_{ji} gives the probability that a replication of type i gives rise to type j and $Q = [q_{ji}]_{n \times n}$ is called the mutation matrix or kernel.

Single mutations in hierarchical tissues

In chapter 1.2.1, we discussed the basic biological properties of hierarchical organized tissue structures. We emphasized, that cell production in most tissues follows such a hierarchical organization. This chapter is based on the publication (Werner et al., 2011), coauthored by David Dingli, Tom Lenaerts, Jorge M Pacheco and Arne Traulsen. A detailed summary of the authors contributions to this publication can be found at the end of this thesis. We evaluate a multi compartment model to describe cell dynamics in an abstract general hierarchy. We analyze the dynamics by stochastic computer simulations capitalizing the Gillespie algorithm, see chapter 1.3.1, and by rate equations, that capture the averages of the simulations. We obtain closed expressions for stem cell and non stem cell driven cell dynamics. Stem cell driven cell proliferation lead to a steady state of homeostasis and non stem cell driven cell proliferation result in clonal waves, that travel trough the hierarchy. Cells at higher stages of the hierarchy without stem cell support are washed out of the hierarchy and get lost in the long run. We further calculate the average extinction time of such clonal waves and are able to compare our theoretical results to the extinction of neutral PIG-A mutants, yielding a good agreement. However, in this chapter we neglect the possibility of additional mutations during cell proliferations and thus do not investigate the dynamics of cells with multiple mutational hits. We extend our model in chapter 3 to account for this aspect of homeostasis.

2.1 Introduction

Many tissues have a hierarchical multi compartment structure in which each compartment represents a cell type at a certain stage of differentiation. This architecture has been well described for hematopoiesis (Dingli et al., 2007a; Nowak, 2006a) and epidermal cell turnover in the skin (Fuchs, 2008; Tumber et al., 2004) or in the colonic crypt (Potten et al., 2009). At the root of this process are the tissue specific stem cells that have the capacity to differentiate into more specialized cells (McCulloch and Till, 2005). Each cell undergoes a series of cell divisions and differentiation steps until the whole diversity of the tissue is obtained (Michor et al., 2003; Komarova and Cheng, 2006; Dingli et al., 2007a; Nowak, 2006a; Johnston et al., 2007; Michor et al., 2005; Wodarz and Komarova, 2005). The model presented here closely follows this concept. We introduce in total $k + 1$ compartments, where each compartment i represents a certain stage of cell differentiation with $i = 0$ representing the stem cell pool. Each cell in a compartment i replicates at a rate r_i . If a cell in a non stem cell compartment $i > 0$ replicates, it can undergo three different processes: With probability ε_i , it divides into two more differentiated cells that migrate into the adjacent downstream compartment $i + 1$. With probability λ_i , the cell dies. With probability $1 - \varepsilon_i - \lambda_i$, it divides into two cells that retain the properties of their parent cell and therefore remain in the same compartment i (self-renewal), as shown in Fig. 2.1. Thus in compartment i , the number of cells N_i is increased by influx from the adjacent upstream compartment $i - 1$ and self-renewal within compartment i , and decreased by cell death in compartment i and cell differentiation into the adjacent downstream compartment $i + 1$. One could also allow asymmetric cell divisions in non stem cell compartments. But the average dynamics in this case can be captured by modifying the differentiation probabilities ε . Thus, this case is implicitly included in our model. In the following, we shall assume a constant number of stem cells N_0 , following (Dingli et al., 2007a; Abkowitz et al., 2002). This can be achieved via asymmetric cell division (Dingli et al., 2007; Lenaerts et al., 2010). However, one can also assume a process at the stem cell level in which cell differentiation, cell death and self renewal are balanced such that the average number of cells remains constant, i.e. $2\varepsilon_0 + 2\lambda_0 = 1$. For immortal stem cells, $\lambda_0 = 0$, this means $\varepsilon_0 = 0.5$. However, for our purpose details of the dynamics in the stem cell compartment are not relevant, as long as the number of stem cells is constant.

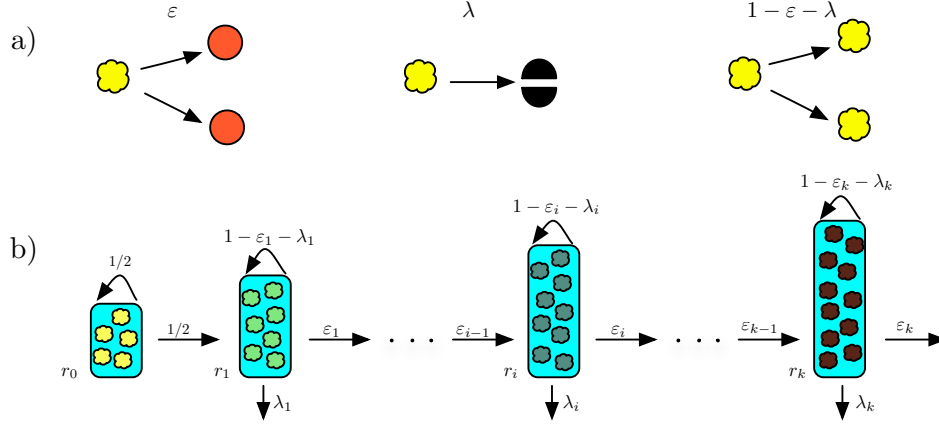


Figure 2.1: **Cell proliferation and compartment structure.** a) We consider three possible events during the cell division of a non-stem cell. Cells can differentiate, die, or reproduce. This happens with probabilities ε , λ , and $1 - \varepsilon - \lambda$, respectively. b) Compartment structure: The first compartment represents stem cells dividing asymmetric. These stem cells replicate with a rate r_0 and their number N_0 is constant. Cells in a non stem cell compartment i replicate with a rate r_i . They will differentiate into the next compartment with probability ε_i , die with probability λ_i or produce cell types of compartment i with probability $1 - \varepsilon_i - \lambda_i$.

This model does neglect several aspects that may have an impact on the dynamics of the system under consideration, such as biochemical feedback or spatial population structure (Jackson, 2003; Nowak et al., 2004; Johnston et al., 2007; Marciniak-Czochra et al., 2009). However, due to the generality, our model can be seen as a benchmark and thus allows to infer when such aspects are of importance and when they can be ignored by a comparison between the different model classes.

One special case of our framework is the model of hematopoiesis discussed in (Dingli et al., 2007a). There, cell death is neglected, $\lambda_i = 0$ for all i . Furthermore, an exponentially increasing proliferation rate $r_i = \gamma^i r_0$ and a constant differentiation probability $\varepsilon_i = \varepsilon$ are assumed for all non stem cell compartments. Here, we relax these conditions and therefore our analytical arguments hold for general values of r_i , ε_i and λ_i in each compartment and thus for a wide class of related models and different tissues.

2.2 Mathematical model and results

The individual cell model is based on a finite number of cells that divide and differentiate with certain probabilities. Thus, it is a stochastic process (Kampen, 1997) and fits the current view of the stochastic nature of such cell differentiation processes (Abkowitz et al., 1996). However, the average cell numbers can be captured by a system of coupled differential equations that is deterministic. These equations follow from a proper counting of incoming and outgoing cells within each compartment i .

Let us assume that the number of stem cells N_0 is constant, following (Dingli et al., 2007a; Abkowitz et al., 2002). The number of cells in the first non stem cell compartment $i = 1$ increases by influx from the stem cell pool at a rate $r_0 N_0$ and due to self renewal at a rate $(1 - \varepsilon_1 - \lambda_1) r_1 N_1$. In addition, the average number of cells in the compartment $i = 1$ is lowered by cell differentiation into the next compartment $i = 2$ at rate $\varepsilon_1 r_1 N_1$. Cell death in compartment $i = 1$ occurs at rate $\lambda_1 r_1 N_1$. The dynamics in all other compartments is the same, except that the number of cells N_i in the compartment i increases due to influx from the adjacent upstream compartment at rate $2\varepsilon_{i-1} r_{i-1} N_{i-1}$. Self renewal occurs at rate $(1 - \varepsilon_i - \lambda_i) r_i N_i$. N_i decreases due to cell death at rate $\lambda_i r_i N_i$ and cell differentiation at rate $\varepsilon_i r_i N_i$. Combining these terms and assuming in total $k + 1$ compartments, we obtain a system of coupled differential equations

$$\dot{N}_0(t) = 0 \quad (2.1a)$$

$$\dot{N}_1(t) = - \underbrace{(2\varepsilon_1 + 2\lambda_1 - 1)}_{\alpha_1} r_1 N_1 + r_0 N_0 \quad (2.1b)$$

$$\dot{N}_i(t) = - \underbrace{(2\varepsilon_i + 2\lambda_i - 1)}_{\alpha_i} r_i N_i + 2\varepsilon_{i-1} r_{i-1} N_{i-1} \quad (2.1c)$$

where $1 < i \leq k$ and the dots denote derivatives with respect to the time t . From now on, we use the abbreviation α_i to denote the difference between the loss from compartment i due to differentiation and cell death and the gain from self renewal. Thus, $\alpha_i = \varepsilon_i + \lambda_i - (1 - \varepsilon_i - \lambda_i) = 2\varepsilon_i + 2\lambda_i - 1$. Typically, we will have $\alpha_i > 0$, and this net loss of cells will be compensated from the influx of cells from the upstream compartment.

The simulations presented in this chapter are individual based stochastic simulations. We implement all elements of the first i compartments separately, thus we are able to record the dynamics of every single cell. Every cell division is called an

event. We use a standard Gillespie algorithm (Gillespie, 1976), described in detail in chapter 1.3.1, to determine in which compartment the next event takes place. After the compartment is determined, one cell in this compartment is chosen to divide proportional to the reproduction rate. The outcome of this event is determined by the cell death and differentiation probabilities λ_i and ε_i . The dynamics in the stem cell compartment is different: in our realization stem cells are allowed to divide asymmetrically only, thus we keep the number of stem cells constant. One could implement a Moran process on the stem cell level also and therefore allow dynamics on the stem cells (Lenaerts et al., 2010). However this would not change the aspects we look at here. The number of stem cell events determine the time scale. We define 1 time unit as n stem cell events. For example, in the hematopoietic system of a healthy adult human we assume that there are approximately 400 stem cell divisions a year.

2.2.1 Stem cell driven dynamics

The equilibrium of the process is obtained from setting the left hand side of our system of differential equations to zero. Biologically, this corresponds to tissue homeostasis. In this case, we have

$$N_i = N_0 \frac{r_0}{r_i \alpha_i} \prod_{j=1}^{i-1} \frac{2\varepsilon_j}{\alpha_j}. \quad (2.2)$$

Next, we turn to the process of filling empty compartments by a continuous influx from the stem cell pool. Because we do not consider interactions between different cell clones in our differential equations, this corresponds also to the dynamics of a mutation arising in the stem cell pool. Thus, we choose the initial condition

$$N_i(0) = \begin{cases} N_0 & i = 0 \\ 0 & \text{otherwise.} \end{cases} \quad (2.3)$$

The differential equation (2.1b) for the compartment $i = 1$ is an inhomogeneous linear differential equation of first order and can be solved by methods such as the variation of parameters. Assuming initial condition (2.3), one obtains the solution for compartment $i = 1$

$$N_1(t) = N_0 \frac{r_0}{r_1 \alpha_1} [1 - e^{-\alpha_1 r_1 t}]. \quad (2.4)$$

Because the differential equation for compartment $i = 2$ depends on $N_1(t)$ and $N_2(t)$ only, one can insert (2.4) into (2.1c) for $i = 2$ and solve the resulting inhomogeneous equation through variation of parameters again,

$$N_2(t) = N_0 \frac{r_0}{r_2} \frac{2\varepsilon_1}{\alpha_1 \alpha_2} [1 - e^{-\alpha_2 r_2 t}] + N_0 \frac{2\varepsilon_1}{\alpha_1} \frac{r_0}{\alpha_1 r_1 - \alpha_2 r_2} [e^{-\alpha_1 r_1 t} - e^{-\alpha_2 r_2 t}]. \quad (2.5)$$

Continuing with this procedure one can find the general pattern, which leads to a solution for general i ,

$$N_i(t) = N_0 \frac{r_0}{r_i \alpha_i} \prod_{j=1}^{i-1} \left(\frac{2\varepsilon_j}{\alpha_j} \right) [1 - e^{-\alpha_i r_i t}] + N_0 r_0 \prod_{l=1}^{i-1} (2\varepsilon_l r_l) \sum_{j=1}^i \frac{(-1)^i}{r_j \alpha_j R_{ji}^{(1)}} [e^{-\alpha_j r_j t} - e^{-\alpha_i r_i t}], \quad (2.6)$$

where we have introduced $R_{ji}^{(1)} = \prod_{l=1, l \neq j}^i (\alpha_j r_j - \alpha_l r_l)$ to shorten our notation. Equation (2.6) allows any choice of ε_i , λ_i and r_i . Within the basic model assumptions depicted in Fig. 2.1, this represents the most general case. All thinkable stem cell driven effects can now be described and followed in detail, as for example any change in the equilibrium compartment sizes or any change of cell division properties during cell differentiation. Compartments are continuously filled with cells until they reach the equilibrium described above. This can easily be deduced from (2.6), because all terms involving decaying exponential functions in time will ultimately be irrelevant for the cell counts.

If we choose (i) an exponentially increasing proliferation rate $r_i = \gamma^i r_0$, (ii) constant differentiation probability $\varepsilon_i = \varepsilon$ and (iii) constant cell death $\lambda_i = \lambda$ for each non stem cell compartment i , solution (2.6) simplifies to

$$N_i(t) = N_0 \frac{1}{\gamma^i} \frac{(2\varepsilon)^{i-1}}{\alpha^i} \left[1 - e^{-\alpha \frac{\gamma^i}{N_0} t} \right] + N_0 \sqrt{\gamma^{i(i-1)}} \frac{(2\varepsilon)^{i-1}}{\alpha^i} \sum_{j=1}^i \frac{(-1)^i}{\gamma^j \Gamma_{ji}^{(1)}} \left[e^{-\alpha \frac{\gamma^j}{N_0} t} - e^{-\alpha \frac{\gamma^i}{N_0} t} \right] \quad (2.7)$$

with $\Gamma_{ji}^{(1)} = \prod_{l=1, l \neq j}^i (\gamma^j - \gamma^l)$ as a short cut.

In Fig. 2.2, equation (2.7) is compared to averages obtained from an individual based stochastic simulation. Note that $\alpha > 0$ is required to maintain an equilibrium. In this case we have $N_i(t \rightarrow \infty) = N_0 (2\varepsilon)^{i-1} / (\gamma \alpha)^i$, the cell count in a

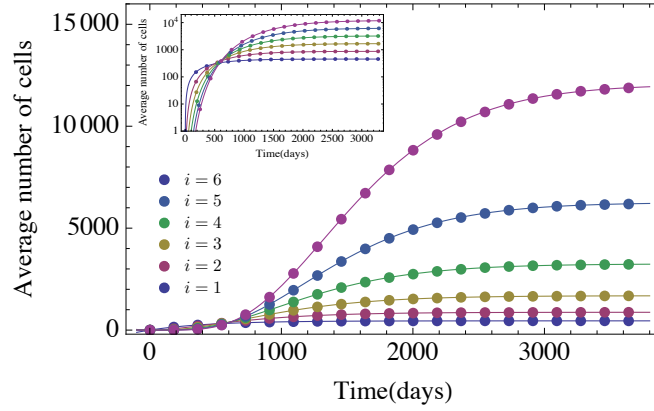


Figure 2.2: **Deterministic dynamics of hematopoiesis modeled as a hierarchical multi compartment process.** The colored symbols are averages of an individual based stochastic simulation with 10^4 realizations. The colored lines are our analytical solution (2.7) for the first six non stem-cell compartments i . Parameters are based on the model of hematopoiesis in (Dingli et al., 2007a) ($N_0 = 400$, $\varepsilon = 0.85$, $\lambda = 0$, and $\gamma = 1.26$, stem cells divide approximately once per year). The stem cell division rate fixes the time scale. Compartments are continuously filled with cells until they reach the equilibrium.

compartment i under equilibrium conditions. This is in agreement with former results (Dingli et al., 2008). While we focus on the biologically relevant case of $\alpha > 0$, we can also consider more general values of α . For $\alpha < 0$, the compartment produces more cells than it loses even in the absence of cell influx from upstream. Thus, the number of cells would grow exponentially according to equation (2.7) in each non stem cell compartment. For $\alpha = 0$ the gain of cells due to self renewal and the loss of cells due to differentiation and cell death in a compartment is equal. Thus the number of cells are not changed by processes in the compartment, despite a continuous output of cells into the next downstream compartment. The case $\alpha = 0$ can be solved directly from Eqs. (2.1a)-(2.1c), which gives $N_i(t) = N_0 \left(\prod_{j=0}^{i-1} r_j \right) \frac{t^i}{i!}$.

Solution (2.6) describes the deterministic process of filling empty compartments within hierarchical organized tissue structures, as can occur during wound healing, recovery from hematopoietic stem cell transplantation (Marciniak-Czochra et al., 2009) or of in vitro experiments with fetal liver cells (Brümmendorf et al., 1998). However, it can also be viewed as the dynamics of a mutant clone arising from a single cell in the stem cell pool, $N_0 = 1$. Thus, it is also possible to describe the

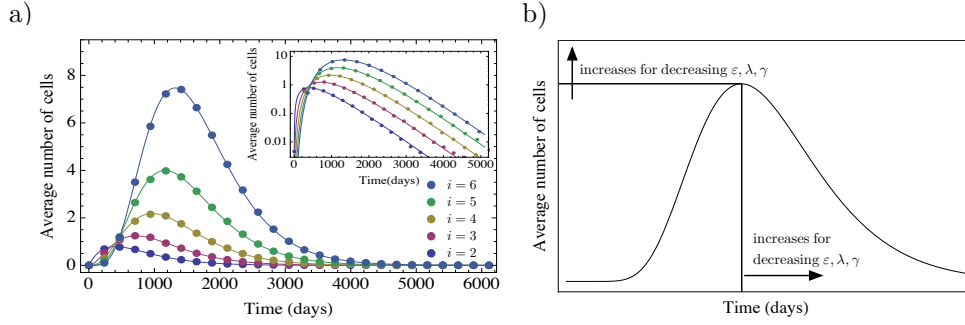


Figure 2.3: **Deterministic dynamics of non stem cell driven mutants in a hierarchical multi compartment process.** a) A single mutant occurred in the compartment $j = 1$. Shown are the number of mutants in the downstream compartments $i = \{2, \dots, 6\}$. Lines are due to equation (2.9), the colored dots are due to an individual based stochastic simulation. The corresponding parameters have been chosen from a model of hematopoiesis (Dingli et al., 2007b) ($N_0 = 400$, $\varepsilon = 0.85$, $\lambda = 0$ and $\gamma = 1.26$). Averages are over 10^4 realizations. In the long run, mutants will always be washed out, because of the missing input from the stem cell pool. However, this process can take a very long time, in our example it would be of the order of 10 years until the average number of mutant cells becomes smaller than one in compartment 6. But it will take significantly longer until all cells in downstream compartments are washed out. b) Parameter dependence of the maximum. For decreasing ε , λ and γ the maximum number of average cells increases, but the time to reach this maximum increases. The more differentiated the cell of the origin of the mutation, the lower the maximum number of cells and the quicker this maximum is reached.

average time development of diseases caused by mutations at the stem cell level such as the chronic myeloid neoplasms. Again, because we assume there is no interaction between normal cells and mutated cells, the dynamics of mutated cells proceeds independently, albeit with different differentiation parameters.

2.2.2 Non stem cell driven dynamics

Next, we turn to mutations occurring downstream of the stem cell compartment. The occurrence of a mutation in a non stem cell compartment is more likely than a mutation in the stem cell pool due to the higher numbers and proliferation rates of non stem cells. The dynamics of such a mutant is not driven by the stem cell pool and thus is not described by the solution form above, equation (2.6). However, the

compartment structure is unchanged and thus the dynamics of such mutants is also described by equations (2.1a)-(2.1c), but with altered initial conditions. Assuming there is a mutation in compartment j , the initial condition is

$$N_{ij}(0) = \begin{cases} 1 & i = j \\ 0 & \text{otherwise.} \end{cases}$$

Here $N_{ij}(t)$ represents the number of mutant cells in compartment i at time t , whereas the mutation occurred in compartment j at time $t = 0$. According to this initial condition, the system of coupled differential equations (2.1a)-(2.1c) turns into a homogenous system and the dynamics of mutant cells is independent from the first j equations. Using the same tools as above, we obtain the general solution

$$N_{ij}(t) = \begin{cases} 0 & i < j \\ \prod_{l=j}^{i-1} (2r_l \varepsilon_l) \sum_{h=j}^i \frac{(-1)^{i-j}}{R_{hi}^{(j)}} e^{-\alpha_h r_h t} & i \geq j \end{cases} \quad (2.8)$$

with $R_{hi}^{(j)} = \prod_{l=j, l \neq h}^i (\alpha_h r_h - \alpha_l r_l)$. Note that this equation describes the dynamics of chronic myeloid leukemia clones in (Michor et al., 2005) analytically and reduces to the solution in (Johnston et al., 2007) in a special case. If, as in (Dingli et al., 2007a), we assume (i) an exponentially increasing proliferation rate, (ii) a constant differentiation probability, (iii) constant cell death across all compartments, then the solution simplifies to

$$N_{ij}(t) = \begin{cases} 0 & i < j \\ \sqrt{\frac{\gamma^{i(i-1)}}{\gamma^{j(j-1)}}} \left(\frac{-2\varepsilon}{\alpha}\right)^{i-j} \sum_{h=j}^i \frac{1}{\Gamma_{hi}^{(j)}} e^{-\alpha \frac{\gamma_h}{N_0} t} & i \geq j. \end{cases} \quad (2.9)$$

Fig. 2.3 a) shows the dynamics of mutants in the first compartments, when the mutation arises in compartment $j = 1$. Note that for the most biologically plausible case $\alpha > 0$, for large t the exponential functions in (2.8) vanish. Thus the mutants will be washed out from the non stem cell compartments. Thus, the absence of mutants is a stable state of such hierarchical compartment structures. However, this equilibrium may not be of any biological or medical relevance, since the time to get rid of the last mutant cells of the clone may be longer than the normal expected lifetime of the healthy organism, cf. Fig. 2.4. Fig. 2.3 b) shows how the maximum of equation (2.9) and the time to reach it depends on ε , λ , and γ . As the proliferation rate γ of the mutant population decreases, the size of the mutant population in downstream compartments increases, although it will take 'longer' for

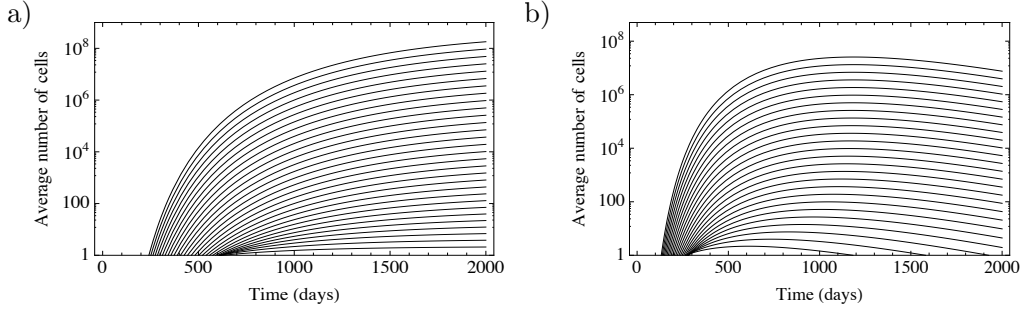


Figure 2.4: **Average cell number for first 31 compartments** a) Average number of mutants in the first 31 compartments of a hematopoiesis model ($N_0 = 400$, $\varepsilon = 0.85$, $\lambda = 0$ and $\gamma = 1.26$) driven by a single stem cell mutation, cf. equation (2.7). The number of mutants increases exponentially per compartment and approaches an equilibrium. b) The same dynamics driven by a single mutant in the third non stem cell compartment $j = 3$, due to equation (2.9). The average mutant cell count reaches a maximum, but vanishes in the long run, because of the missing influx from the stem cell compartment. However, it will be difficult to distinguish between the two cases in the initial phase of the disease (here ≈ 3 years).

the population to reach high levels. A mutation that increases the net loss of cells α (either by increasing cell differentiation ε or cell death λ) in a compartment lowers the number of mutants at maximum size of the clone in downstream compartments, which is also reached earlier. Note also that mutations occurring later in the cell differentiation process will lead to smaller maxima that vanish faster (Traulsen et al., 2010).

Based on equation (2.8), other mutant dynamics are also possible. If $\alpha_h < 0$ in a single compartment h , mutant counts diverge exponentially in all downstream compartments. If $\alpha_h = 0$ in a single compartment h and $\alpha > 0$ otherwise, in the long run mutants will reach an equilibrium in all downstream compartments, which is given by

$$N_{ij}(t \rightarrow \infty) = \begin{cases} 0 & i < h \\ 2^{i-j-1} \frac{r_h}{r_i} \frac{\alpha_h}{\varepsilon_i} \prod_{l=j}^i \frac{\varepsilon_l}{\alpha_l} & i \geq h. \end{cases} \quad (2.10)$$

This equilibrium is robust against variations of N_i and thus is a stable fixed point. However a small change in α_h would lead either to extinction or the divergence of the mutant cell count. Initially, the difference between the dynamics of a clone arising from the stem cell compartment and an early non stem cell compartment is

small, see Fig. 2.4.

2.2.3 Mutant extinction times

In the long run the average mutant cell count is given by the dynamics of the slowest decaying exponential function of equation (2.8). It is often natural to assume that this corresponds to the dynamics in the compartment of the mutant origin j . Thus, if we assume that $\alpha_j r_j < \alpha_i r_i$ for all $i > j$ (as in the hematopoiesis model in (Dingli et al., 2007a)), in the long run mutants will die out at a rate

$$N_{ij}(t) = \left(\prod_{l=j}^{i-1} 2r_l \varepsilon_l \right) \frac{(-1)^{i-j}}{R_{ji}^{(j)}} e^{-\alpha_j r_j t}. \quad (2.11)$$

This is shown in Fig. 2.5 a). For this special choice of parameters equation (2.11) becomes

$$N_{ij}(t) = \sqrt{\frac{\gamma^{i(i-1)}}{\gamma^{j(j-1)}}} \left(\frac{-2\varepsilon}{\alpha} \right)^{i-j} \frac{1}{\Gamma_{ji}^{(j)}} e^{-\alpha \frac{\gamma_j}{N_0} t}. \quad (2.12)$$

Thus, the mutant cell count in the long run is given by a decaying exponential function. This enables us to calculate the average extinction time t_{ij}^{ext} of mutants in the i -th compartment, if the mutation occurred in compartment j ,

$$t_{ij}^{\text{ext}} = \frac{1}{\alpha_j r_j} \ln \left[\frac{\prod_{l=j}^{i-1} (-2r_l \varepsilon_l)}{R_{ji}^{(j)}} \right]. \quad (2.13)$$

If we assume a constant differentiation probability, constant cell death and an exponential increasing proliferation rate again this simplifies to

$$t_{ij}^{\text{ext}} = -\frac{N_0}{\alpha \gamma^j} \ln \left[\sqrt{\frac{\gamma^{j(j-1)}}{\gamma^{i(i-1)}}} \left(\frac{-2\varepsilon}{\alpha} \right)^{j-i} \Gamma_{ji}^{(j)} \right]. \quad (2.14)$$

In Fig. 2.5 b) we compare the extinction time due to equation (2.14) to simulation results. This approximation does not allow to calculate the extinction time of the mutant of the compartment of origin, but a more detailed consideration of this case can be found in (Dingli et al., 2008).

A special case of interest is a mutation with $\varepsilon_j = 1/2$ and $\lambda_j = 0$. This results in a mutant cell that shows stem cell like properties in compartment j and non

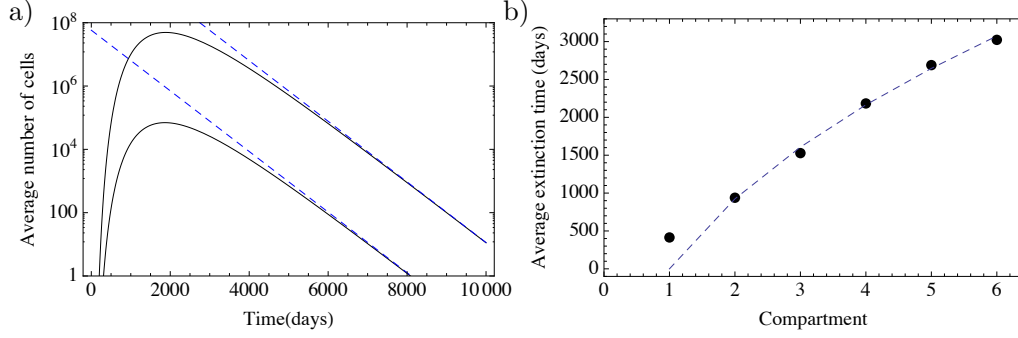


Figure 2.5: **Extinction times.** a) The black lines show the average mutant cell count based on equation (2.9) in the compartments $i = 20$ and $i = 30$, where the mutant occurred in the first non stem cell compartment $j = 1$. The dashed blue lines are given by equation (2.11), extinction occurs when the average cell count drops below 1. Again the parameters are chosen from a model of hematopoiesis as in Fig. 2.3 and 2.4. b) Average extinction time due to simulation (black dots) and due to equation (2.14) (dashed blue line). This approximation does not work for the mutant in the compartment of origin, where an alternative approach is necessary (Dingli et al., 2008).

stem cell like properties in higher compartments. In this case the set of differential equations (2.1a)-(2.1c) becomes

$$\dot{N}_i(t) = \begin{cases} 0 & i \leq j \\ -\alpha_i r_i N_i + r_{i-1} N_{i-1} & i = j + 1 \\ -\alpha_i r_i N_i + 2\varepsilon_{i-1} r_{i-1} N_{i-1} & i > j + 1. \end{cases} \quad (2.15)$$

Equation (2.15) transforms into equation (2.1a)-(2.1c) if one shifts the index i to $i \rightarrow i - j$. Thus we have to shift the index of the general solution (2.6) and find

$$N_{ij}(t) = \frac{r_j}{r_i \alpha_i} \prod_{l=j+1}^{i-1} \left(\frac{2\varepsilon_l}{\alpha_l} \right) [1 - e^{-\alpha_i r_i t}] + r_j \prod_{l=j+1}^{i-1} (2\varepsilon_l r_l) \sum_{h=j+1}^i \frac{(-1)^i}{r_h \alpha_h R_{hi}^{(j)}} [e^{-\alpha_h r_h t} - e^{-\alpha_i r_i t}]. \quad (2.16)$$

Thus, the average dynamics of such a mutation is exactly as described above for a stem cell mutation. The averages of the simulation are described by equation (2.6) and (2.8), but a single run is still stochastic. The relative standard deviation of lower compartments σ_i/N_i is of order 1, but decreases with increasing compartment number. Stochastic effects on the stem and the early progenitor cell level are important

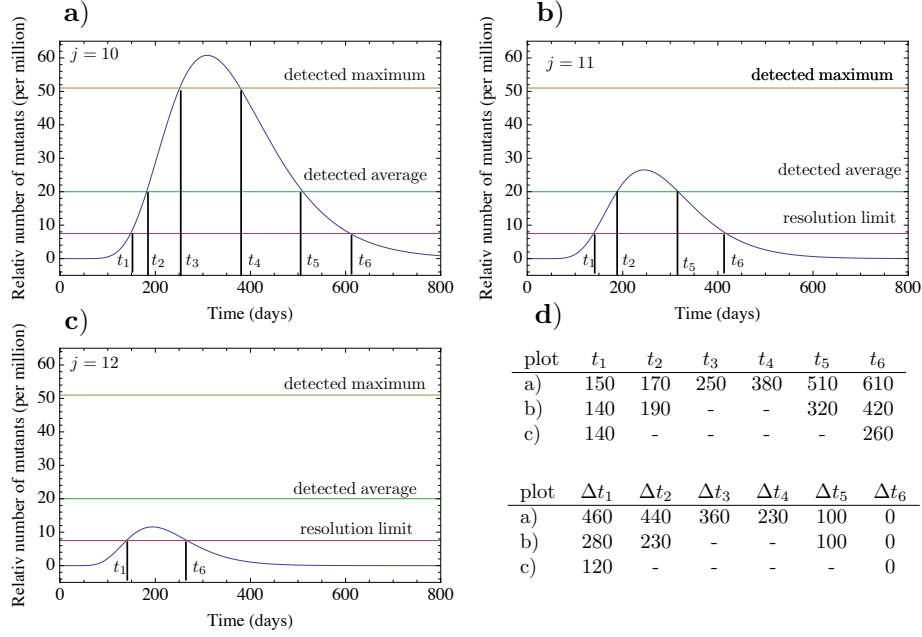


Figure 2.6: **Dynamics of PIG-A mutants.** Panels a) to c) show the number of PIG-A mutants per million healthy cells in compartment $k = 31$, based on equation (2.8). The mutation of origin occurred at time $t = 0$ in compartment $j = \{10, 11, 12\}$. The parameters of the mutant cells are $\gamma_P = 1.26$, $\varepsilon_P = 0.77$ and $\lambda_P = 0$ in all compartments. The horizontal lines correspond to the detected maximum, the detected average and the detection limit of PIG-A mutants observed in in vivo studies (Araten et al., 1999). The upper table in panel d) shows the time (in days) of these events after the original mutation occurred at time $t = 0$ based on equation (2.8). The bottom table shows the corresponding extinction times ($\Delta t_l = t_6 - t_l$). We predict an extinction time of 230 days for the measured maximum and 100 days for the average mutant cell count.

and can be crucial in a medical context (Dingli et al., 2007b; Lenaerts et al., 2010), but to understand some fundamental properties, the deterministic view seems to be sufficient.

2.2.4 Example: Dynamics of PIG-A mutants

Here, we will utilize the model to illustrate the dynamics of a mutation that is seen in virtually every healthy human being. Sensitive flow cytometric analysis of circulating blood cells will identify a small clone that lacks expression of CD55 and CD59 (amongst others)(Araten et al., 1999). CD55 and CD59 belong to a class

of proteins that inhibit complement activation and their absence renders red blood cells sensitive to intravascular destruction. These proteins are normally displayed on the surface of cells since they are anchored to the plasma membrane via a glycosylphosphatidyl inositol (GPI) moiety. Synthesis of GPI requires a series of steps. The PIG-A gene encodes a protein that is an essential component of the complex responsible for the first step of GPI biosynthesis. Mutations in this gene can lead to a partial or complete deficiency of GPI synthesis resulting in low level or complete absence of such proteins from the cell surface, as for example the complement inhibitors CD55 and CD59 (Hillmen et al., 1993; Longo et al., 2008). Red blood cells lacking CD55 and CD59 are destroyed by complement, leading to hemolytic anemia. As a result, mutations in PIG-A can explain the phenotype of paroxysmal nocturnal hemoglobinuria (PNH), an acquired hematopoietic stem cell disorder characterized by anemia, hemoglobinuria and other manifestations (Luzzatto et al., 1997). A recent mathematical model suggests that a PIG-A mutation in a HSC is sufficient to explain the incidence and natural history of PNH (Dingli et al., 2008). However circulating blood cells with the PNH phenotype (due to a mutation in PIG-A) can be found in virtually every healthy adult human (Araten et al., 1999). Such clones generally disappear with time. With this background, we will now apply the analytical solution (2.8), to assess extinction times of PIG-A mutants and compare these results to in vivo data derived from healthy adult humans.

The model parameters were fixed to represent hematopoiesis following (Dingli et al., 2007b). In this approach, cell death is neglected, $\lambda_i = 0$ for all i , and an exponentially increasing proliferation rate $r_i = \gamma^i r_0$ as well as a constant differentiation probability $\varepsilon_i = \varepsilon$ is assumed for all non stem cell compartments. Further, limited self-renewal is considered across many stages of differentiation, a prediction that is finding increasing support: For example, this was noted recently for cells at the proerythroblast stage, a cell type far removed from the stem cell or primitive progenitor cell pools (England et al., 2011). Finally, the model parameters for human hematopoiesis become $N_0 = 400$, $r_0 = 1/N_0$, $\gamma = 1.26$ and $\varepsilon = 0.85$. The number of cells per compartment increases exponentially and one needs $k \approx 31$ compartments (Dingli et al., 2007b) to ensure that in a healthy adult human, on average, the daily bone marrow output is of the order of $3.5 \cdot 10^{11}$ blood cells (Vaziri et al., 1994). The same model can also be fitted to other mammals (Dingli et al., 2008b). PIG-A mutants are considered to be neutral (Araten et al., 2002), supported by in vivo measurements (Maciejewski et al., 1997; Dunn et al., 1996). Thus we chose, as

for normal cells, $\lambda_P = 0$ and $\gamma_P = 1.26$ as mutant parameters. For fully neutral mutants, the clone would either be present for too much time or it would not reach the level observed in vivo. We explored various values of ε and found that a slightly lower differentiation probability $\varepsilon_P = 0.77$ gave the best results. Note that this slight difference compared to healthy cells is still consistent with the experimental evidence.

In Araten et al (Araten et al., 1999) the blood of 19 healthy adult humans was sampled and tested for clones with PIG-A mutations. Mutants were found in every person ranging from 8 to 51 mutants per million (with an average of 21) normal blood cells. Blood samples from the same patients were taken at later times to determine survival of these clones. The lower limit of detection in (Araten et al., 1999) was approximately 7.5 mutant cells per million. The detected maximum of 51 mutants per million healthy cells decreased after 164 days and was undetectable after 192 days. Individuals with the average cell count of 21 PIG-A mutants per million still had the clone present after 65 days, but it was not detectable after 174 days. We need to determine the compartment j , where a mutation in PIG-A occurred, such that the clone that arises would grow to reach the detection threshold and remain detectable for a time compatible with observations. Using equation (2.8), we record the dynamics of mutant cells in compartment 31 for different compartments of origin. In Fig. 2.6, the mutant cell count per healthy cells in compartment 31 is shown, where the mutation took place in a) compartment 10, b) compartment 11 and c) compartment 12. With these curves, one can predict extinction times for different origins of the mutation. Comparing the total size of the mutant population and the corresponding extinction times to values obtained in humans (Araten et al., 1999) allows to predict the compartments where the mutant clone originated.

In Fig. 2.6, we show the corresponding times from the mathematical model calculated from equation (2.8). The same figure also illustrates that the time of origin of the mutation can be much earlier than the detection time. For example, if the mutation occurred in compartment 10, we predict an extinction time of 230 days for the maximum of 51 mutants per million, when the initially sample was taken at time t_3 , see figure 2.6 a). The extinction time for the average cell count is 100 days. However, it should be clear that such mutant cells will survive for significantly longer than what is detectable by technology due to issues of sensitivity. With this in mind, there is good agreement between what the model predicts and the results described in (Araten et al., 1999), since it is unlikely that the clones in

all the individuals were either found as soon as they were detectable or when they were at their peak concentration. Thus what is relevant are (i) the distribution of times that these mutant cells remain in circulation and (ii) the size of the clones one observes. In this respect our model provides a very good approximation of the dynamics of such clones and is able to infer the cell of origin. If the mutation occurred in earlier compartments, the clone would be expected to expand to a higher cell count and will stay in the circulation for a longer time, but such clones are less likely to occur due to the lower number of progenitor cells and slower proliferation rate. Mutation events in higher compartments as $j = 12$ are more likely to happen but these mutants would not be detectable by most current clinical flow cytometry techniques due to the small size of such clones in compartment 31 although they may be detected perhaps with polymerase chain reaction technology. Thus, compartments $\{10, 11, 12\}$ are the most likely compartments of the mutant origin for the cases described in (Araten et al., 1999). Mutations arising in these compartments correspond to mutations in early progenitor cells. This agrees with the experimental results, since the PNH phenotype is present in several different cell lineages and thus has to occur early in the hematopoietic tree. Note that besides the compartment of origin, the only free parameter is the differentiation probability of mutated cells ε_p . Here we assumed that these circulating mutant cells all originate from one founder cell. Two or more independent, simultaneous or contemporaneous clones in early compartments would be unlikely (Traulsen et al., 2007; Dingli et al., 2008). If a second independent mutation occurs during the presence of an earlier mutant population, the total mutant cell number is the sum of both single populations, see the supporting information for more details.

Moreover the hierarchical structure of hematopoiesis provides an explanation why almost all humans carrying PIG-A mutations do not have symptoms of PNH. Only mutations in the most 'primitive' compartments have an impact and only mutations in a HSC will lead to disease. In general, one can predict the dynamics for mutants with very different properties using equation (2.8). The compartment of the mutant origin can be inferred if one follows the mutant count by taking blood samples at regular intervals.

2.3 Discussion

In this work, we presented closed analytical solutions for the deterministic dynamics of stem cell and non stem cell driven mutants in a multi compartment model of tissues such as hematopoiesis, the skin and the colon. This enables us to describe the dynamics of mutant cells in a general approach. We can predict the time development of a mutant depending on its origin and its specific proliferation properties. The process of cell differentiation is conceptually fairly well understood, but it is of course a challenge to estimate the various parameters in our model for real systems. Fortunately, very often, simplifying assumptions, e.g. exponentially increasing cell proliferation rates, can lead to insights (Dingli et al., 2007a). However, our analytical solution allows us to incorporate more involved parameter dependencies, which could immediately be analyzed.

Let us turn to hematopoiesis to address some of the implications of our model because recent technological developments allow the detection of well known mutations in many otherwise healthy people. Perhaps the best examples are derived from blood disorders, since repeated blood sampling is a minimal invasive procedure and molecular probes for many blood disorders are available. The case of PIG-A mutant cells present in healthy humans has been analyzed extensively in an earlier section. There are several other specific examples (Traulsen et al., 2010).

- (i) A mutation in the janus like kinase 2 where phenylalanine substitutes valine (JAK2V617F) is a common mutation in patients with chronic myeloid neoplasms. However, one can find this mutation in a substantial fraction of healthy adults (perhaps 0.2 - 0.4 percent) and with an even higher frequency (0.94 percent) in hospitalized patients who do not have an overt hematologic disorder (Nielsen et al., 2011; Xu et al., 2007). JAK2V617F is expected to give a survival and reproductive advantage to cells, and probably also enhances self renewal of progenitor cells. Knowing the dynamics of such clones, can lead to an understanding of the cell of origin in these patients as well as its impact on the fitness of mutant cells.
- (ii) Finally, the classic oncogene BCR-ABL (Daley et al., 1990) that is associated with chronic myeloid leukemia can be found in healthy adults (Bose et al., 1998). In some of these individuals the mutant clone resolves while in others it persists but to our knowledge, none of the individuals in the cohort described

have progressed to develop CML. There are various potential explanations for this observation including (i) non-stem cell origin of the mutant clone, (ii) stochastic extinction (Lenaerts et al., 2010), (iii) immune response to the clone, (iv) additional mutations may be needed to lead to CML. Independent of the multitude of possibilities, it is safe to conclude that the cell of origin of a mutant is of importance and the impact of a mutation is cell context dependent. Our model can provide plausible explanations for the frequency and cell of origin of these mutations and perhaps why they do not lead to disease.

- (iii) We can also think of other mutations altering cell division properties. For instance, one can consider a mutation occurring in compartment j with $\varepsilon_j = 1/2$, $\lambda_j = 0$ and normal properties in all the other upstream compartments. This would be the special case described by equation (2.16), and can be understood as a mutation that enables a cell to reacquire stem cell-like renewal capacities again. Such a behavior can explain the origin of various subtypes of acute leukemia as has been reported recently (Krivtsov et al., 2006; Huntly et al., 2004; Guibal et al., 2009).

Our model provides a mathematical abstraction of hierarchically structured tissues and neglects many factors that can have an important impact on the dynamics, as for example spatial population structure or temporal changes of cell division properties, e.g. due to aging or injury. Nonetheless, the most important aspects of such tissue structures are captured by our model. It takes the form of ordinary differential equations that allows analytical solutions in many cases. An alternative would be a numerical solution, but such a solution has to be implemented for specific sets of parameters. We are convinced that our model can readily be applied to various hierarchical tissues and expect that general features of mutant dynamics will be conserved across different tissues.

Multiple mutations in hierarchical tissues

In chapter 2 we evaluated the cell dynamics in hierarchically organized tissue structures, neglecting the possibility of multiple mutational hits within a single cell. However, as discussed in 1.2.2 such multiple mutational hits are essential for many cancers to occur. This chapter is based on a manuscript, coauthored by David Dingli and Arne Traulsen, currently under revision in the Journal Interface of the Royal Society. We now extend our multi compartment model to allow multiple mutations within the hierarchy. We investigate the scaling of multiple mutations in hierarchies. Most importantly we introduce the reproductive capacity of a non stem cell, defined as the total cell load produced from a single cell somewhere in the hierarchy. We obtain analytical expressions for most general cell proliferation parameters and find interesting scaling properties of hierarchical tissue structures. Cells carrying multiple mutations are strongly suppressed by the hierarchy. However, the scaling of those cells is sensitive to their proliferation parameters and a small divergence from the healthy state can already lead to a substantial increase of the number of cells with multiple mutations. Thus small differences in cell proliferation parameters result in significant differences of the expected number of distinct mutations, derived from a single ancestor cell. This finding might contribute to the explanation of very diverse mutational landscapes for different cancer types.

3.1 Introduction

The lifespan of most cells in biological organisms is limited and usually the life expectancy of the organism exceeds this time by orders of magnitudes ([Hayflick et al., 1961](#); [Raff, 1992](#)). As cells are continuously lost, mechanisms to replenish the cell pool evolved in organisms, enabling sustained cell production during life time ([Pardee, 1989](#)). Often this is realized by hierarchically organized tissue structures. At the root of the hierarchy are few tissue specific stem cells, combining two properties, self renewal and differentiation potential ([McCulloch and Till, 2005](#)). During cell proliferation, cells differentiate and become increasingly specialized to perform specific functions within the hierarchy. After some differentiation steps the complete spectrum of functional cells can be obtained ([Michor et al., 2003](#); [Dingli et al., 2007a](#); [Nowak, 2006a](#); [Michor et al., 2005](#); [Wodarz and Komarova, 2005](#); [Werner et al., 2011](#)). A prominent example is the hematopoietic system ([Loeffler and Wichmann, 1980](#); [Michor et al., 2005](#); [Dingli et al., 2007a](#); [Nowak, 2006a](#); [Wodarz and Komarova, 2005](#); [Werner et al., 2011](#)), but other tissues as for example skin ([Fuchs, 2008](#); [Tumbar et al., 2004](#)) or colon ([Potten et al., 2009](#)) are also hierarchically organized.

A large number of cell divisions are indispensable to life, however, they are unavoidably accompanied by mutations. Typically, these cells are washed out of the hierarchy and thus, especially if they arise in relatively differentiated cells, the associated mutations are lost in the long run ([Araten et al., 1999](#); [Werner et al., 2011](#)). But cells with multiple mutational hits might persist for a long time, causing the risk of the accumulation of additional mutations during cell proliferation, which can ultimately lead to cancer. Although some cancers seem to be caused by single mutation hits, for example the BCR-ABL translocation occurring in stem cells in chronic myeloid leukemia ([Daley et al., 1990](#)) or the PML-RARA translocation occurring in more differentiated cells in acute promyelocytic leukemia ([Guibal et al., 2009](#)), they are rare. The majority of cancers are triggered by at least a handful of mutations ([Stratton et al., 2009](#); [Beerenwinkel et al., 2007](#); [Gerstung and Beerenwinkel, 2010](#)). The recent progress in genome sequencing techniques allowed in some cases the classification of cancer initiating mutations, in other cases the underlying mutations remain unknown ([Welch et al., 2012](#)). However, many of these studies reveal a very diverse mutation landscape, indicating the existence of several cancer initiating driver mutations and additional alterations that have a small or even no impact on

cancer development, so called passenger mutations (Hanahan and Weinberg, 2000; Vogelstein and Kinzler, 2004; Sjöblom et al., 2006; Jones et al., 2008; Pleasance et al., 2010; Shah et al., 2012; Ding et al., 2012; Walter et al., 2012). The precise impact of passenger mutations on cancer progression is still under discussion, but the typical assumption is that they are neutral and do not affect the proliferation properties of cells (Haber and Settleman, 2007; Bozic et al., 2010). In particular, this holds for synonymous mutations that do not have any consequences on protein structure or function (Greenman et al., 2007).

In mathematical and computational approaches, compartment models are frequently used to describe cell dynamics in hierarchically organized tissue structures. Many of these studies investigate effects of stem cell mutations and related clinical implications, see for example (Michor et al., 2005; Rodriguez-Brenes et al., 2011; Roeder et al., 2006; Dingli et al., 2007a; Marciniak-Czochra et al., 2009; Glauche et al., 2011). Also stochastic effects of tissue homeostasis are analyzed (Lenaerts et al., 2010, 2011; Traulsen et al., 2013), highlighting, that cancer driving mutations can in principle disappear by chance (stochastic extinction). The interplay of stem cell and progenitor cell mutations and their impact on cancer initiation are discussed (Komarova and Wang, 2004) and game theoretical approaches allow to model evolutionary aspects of tissue homeostasis and inter cell competition (Basanta et al., 2011; Gerlee et al., 2011). Often, these studies investigate the effects of cells carrying one or very few specific mutations and assume either constant population size or only minimal hierarchies.

Here, we focus on the presence of cells carrying multiple mutations within a hierarchically organized tissue. We show mathematically, that the hierarchical organization strongly suppresses cells carrying multiple mutations and thus reduces the risk of cancer initiation. Closed solutions for the total cell population that arises from a single (mutant) cell are derived and from this the expected diversity of the mutation landscape and the clonal size can be described. This enables a better understanding of the expected diversity in mutation landscapes that are observed in both healthy and cancerous tissues.

3.2 Mathematical model and results

The hierarchical tissue organization is typically modeled by a multi compartment approach (Dingli et al., 2007a; Werner et al., 2011). Each compartment represents

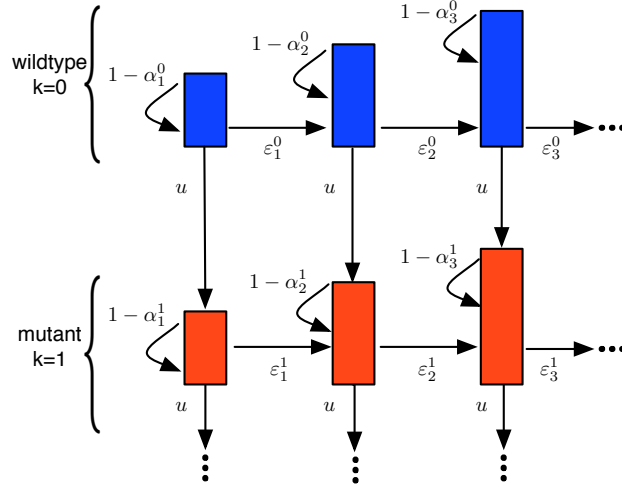


Figure 3.1: **Schematic representation of the compartment structure of multiple mutations and corresponding transition rates.** The top compartments (blue) contain cells carrying no mutation. The bottom compartments (red) contain cells carrying one mutation. Compartments to the right represent more specialized cell stages and arrows transition probabilities, where ε denotes the differentiation probability, u the mutation rate of cells and $\alpha_i^k = \varepsilon_i^k + u$. Initially no mutated cells are present in the hierarchy. We then ask, how many cells are acquired from the founder compartment (top left) and investigate how many cells with k mutations are on average expected at any stage of the hierarchy.

a certain differentiation stage of cells. At the root of the hierarchy are stem cells ensuring a continuous influx of cells. A proliferating cell in compartment i divides and the two daughter cells differentiate and migrate into the next downstream compartment ($i + 1$) with probability ε , increasing the downstream compartment by 2 cells, mutates with probability u or self renews within its own compartment with probability $1 - \varepsilon - u$. Mutated cells stay in the hierarchy. If a mutated cell proliferates, it differentiates with probability ε into the next downstream compartment, it self renews with probability $1 - \varepsilon - u$, or it mutates with probability u again, leading to a cell with two (or more) mutations. All possible outcomes of a cell proliferation are depicted in Fig. 3.1. The direction of the arrows point towards the accessible cell states and the labels give the transition probabilities. We allow arbitrary parameters and introduce ε_i^k as differentiation probability of cells in compartment i carrying k mutations. Asymmetric cell divisions are not explicitly implemented, as they can be absorbed in the differentiation probabilities on the population level. The fate of a cell's offspring is determined based on the probabilities ε_i^k . Cells proliferate

erate with a rate r_i in each compartment i . Usually cells in upstream compartments proliferate slow and cell proliferation speeds up in downstream compartments (i.e. $r_i < r_{i+1}$). This general framework is very flexible and different tissue structures can be represented.

We implement individual based stochastic simulations of the cell dynamics in hierarchically organized tissue structures. We utilize an implementation of the Gillespie algorithm (Gillespie, 1976, 2007). Originally introduced to simulate chemical reactions, it allows us to reproduce exact stochastic trajectories of the system. Each cell has an individual representation. Thus the complete clonal history of cells within the hierarchy can be recorded. If a cell is chosen for reproduction (determined by the Gillespie method), it either differentiates, self renews or mutates according to the probabilities ε_i^k and u . The parameters of the simulated system are described below and chosen to represent human hematopoiesis.

In the following we focus on the hematopoietic system. There about 400 stem cells replenish the hematopoietic cell pool (Dingli and Pacheco, 2006; Buescher et al., 1985). Each stem cell divides approximately once a year (Rufer et al., 1999; Dingli and Pacheco, 2006). Cell proliferation is assumed to increase exponentially $r_i = \gamma^i r_0$, with $\gamma = 1.26$ and r_0 corresponds to the proliferation rate of stem cells. The differentiation probability is assumed to be constant, $\varepsilon = 0.85$, for all non stem cell compartments and in total $i = 31$ compartments are needed, to ensure a daily bone marrow output of approximately 3.5×10^{11} cells (Dingli et al., 2007a; Werner et al., 2011).

3.2.1 Time continuous dynamics of multiple mutations

We describe the deterministic dynamics of a cell population within a hierarchically organized tissue structure, that initially carries no mutation. A cell may commit further into the hierarchy (differentiate), mutate, or self renew itself. This occurs with probability ε , u and $1 - \varepsilon - u$, respectively. In Fig. 3.1 a schematic representation of the resulting hierarchical structure is shown. Compartments to the right represent downstream compartments of more specialized (differentiated) cells, while compartments to the bottom represent states of cells, that accumulated an additional mutation. During one cell division, a cell either mutates and moves one compartment to the bottom, differentiates and produces two cells in the next downstream compartment to the right, or self renews and produces an additional

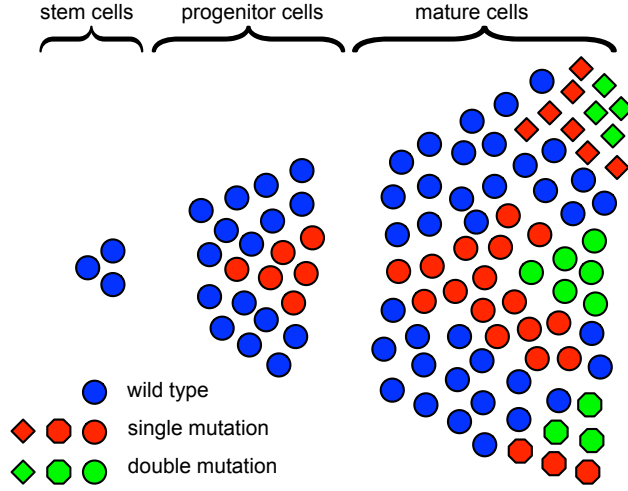


Figure 3.2: **Clonal expansion within a hierarchically organized tissue.** Cell proliferation is driven by a few slow dividing stem cells, giving rise to faster dividing progenitor cells. After some differentiation steps the mature tissue cells are obtained. Initially cells have no mutations, but mutants can arise and expand within the hierarchy. These cells either vanish or gain an additional mutation, which again potentially spreads within the hierarchy. Different colors code for a different number of mutations, whereas different shapes indicate different mutations.

cell within its original compartment. This leads to an expansion of clonal populations within the hierarchy, that potentially accumulates several (distinct) mutations during the differentiation process. This is schematically shown in Fig. 3.2.

The above transition probabilities can be used in an individual based stochastic simulation. In the following, we provide a deterministic description of the average dynamics of cells carrying multiple mutations in such hierarchical structures. Thus, we describe the dynamics by transition rates instead of transition probabilities, but only averages enter in our description. By doing so, we neglect certain effects, such as stochastic extinction of cells. However, the approach allows us to investigate the averages of the underlying stochastic simulations.

Assume that in compartment i there are $N_i^k(t)$ cells carrying k mutations at time t . The number of these cells increases due to influx from the upstream compartment at a rate $2r_{i-1}\varepsilon_{i-1}^k$, mutations with a rate $r_i u$ and self renewal at a rate $r_i(1 - \varepsilon_i^k - u)$. Cells are lost either by mutation at a rate $r_i u$, or differentiation at a rate $r_i \varepsilon_i^k$. The deterministic description of the hierarchical compartment model

becomes a system of coupled differential equations (Werner et al., 2011), given by

$$\dot{N}_i^k(t) = \begin{cases} r_i (1 - 2\alpha_i^k) N_i^k(t) + 2r_i \varepsilon_{i-1}^k N_{i-1}^k(t) & k = 0 \\ r_i (1 - 2\alpha_i^k) N_i^k(t) + 2r_i \varepsilon_{i-1}^k N_{i-1}^k(t) + r_i u N_i^{k-1}(t) & k > 0. \end{cases} \quad (3.1)$$

Here $\alpha_i^k = \varepsilon_i^k + u$ denotes the probability that a cell with k mutations leaves compartment i . Typically, α_i^k is very close to ε . A model for stochastic cell dynamics in the stem cell compartment for neutral and non neutral mutations can be found in (Dingli et al., 2007b; Traulsen et al., 2013). In that paper, the stochastic Moran process is used to investigate the extinction and fixation probabilities of stem cell mutations. The deterministic stem cell driven cell replenishment in hierarchical tissues is studied in detail in Werner et al. (2011). However, in that prior work, the effects arising from additional mutations were neglected. Here we focus on non stem cell driven clonal dynamics. We explicitly allow for an arbitrary number of mutational hits at any stage of the hierarchy, but we neglect a continuous influx of mutated cells from the stem cell level. This assumption gives the condition $N_0^k(t) = 0$. The initial condition

$$N_i^k(0) = \begin{cases} n_0 & i = 1 \quad k = 0 \\ 0 & \text{otherwise,} \end{cases} \quad (3.2)$$

corresponds to initially n_0 cells in compartment 1 carrying no mutation. One can think of a neutral marker approach, where one cell in the hierarchy is genetically marked, and one considers the clonal population arising from this marked cell (Gerrits et al., 2010). Although we neglect a continuous influx of mutated cells from the stem cell compartment, stem cell mutations can be implemented indirectly. Our approach allows for altered cell proliferation properties of the founder cell, potentially derived by a mutation at the stem cell level. For a constant differentiation probability, i.e. the case where all α_i^k and all ε_i^k are identical, Eq. (3.1) can be solved recursively. The number of cells in compartment i carrying no mutation changes in time as

$$N_i^0(t) = n_0 \frac{(2\varepsilon)^{i-1}}{(1-2\alpha)^{i-1}} \left(\prod_{j=1}^{i-1} r_j \right) \sum_{l=1}^i \frac{e^{r_l(1-2\alpha)t}}{\prod_{\substack{h=1 \\ h \neq l}}^i (r_h - r_l)}. \quad (3.3)$$

The solution can also be derived recursively for cells carrying k mutations in com-

partment i and becomes

$$N_i^k(t) = n_0 \sum_{h=0}^k \frac{u^{k-h} g_i^h}{k!} \frac{(2\varepsilon)^{i-1}}{(2\alpha-1)^{i+h-1}} \left(\prod_{j=1}^{i-1} r_j \right) \sum_{l=1}^i \frac{(r_l t)^{k-h} e^{r_l(1-2\alpha)t}}{\prod_{\substack{h=1 \\ h \neq l}}^i (r_h - r_l)}, \quad (3.4)$$

where g_i^h is a combinatoric parameter denoted in 3.5 for cells carrying up to three mutations,

	$h = 0$	$h = 1$	$h = 2$	$h = 3$
$k = 0$	1			
$k = 1$	1	$i - 1$		
$k = 2$	1	$(i - 1)k$	$i(i - 1)$	
$k = 3$	1	$(i - 1)k$	$i(i - 1)k$	$(i + 1)i(i - 1)$

(3.5)

If $\alpha < 0.5$, non stem cells will continuously accumulate in downstream compartments. The probability of self renewal in this case is larger than the probability of differentiation. This scenario seems to be realized in certain blood cancers. For example in acute promyelocytic leukemia an abnormal increase of immature granulocytes and promyelocytes is observed, resulting from a block of cell differentiation at a late progenitor cell stage (Raymond et al., 1993; Guibal et al., 2009). However these cases are rare.

For $\alpha > 0.5$, the solution becomes a clonal wave, traveling through the hierarchy in time. In this case the probability of differentiation is larger than the probability of self renewal and thus cells progressively travel downstream, see Fig. 3.3 as an example. The cell population founded by a single non stem cell expands within the hierarchy initially, but gets washed out and vanishes in the long run. This is believed to be true for healthy homeostasis. For example for the hematopoietic system the differentiation probability was estimated to be $\varepsilon = 0.85$ (Dingli et al., 2007a). As by far most cell proliferations occur at the progenitor and more committed differentiation stages, this provides a natural protection of the organism against the accumulation of multiple mutations, as the survival time of most (non stem cell like) mutations is finite.

The maximum mutant cell count of the clonal wave and the time to reach this maximum can be calculated for the compartment of the mutant origin, in our case the first compartment. The time is given by

$$t_{\max}^k = \frac{k}{(2\alpha - 1) r_1}. \quad (3.6)$$

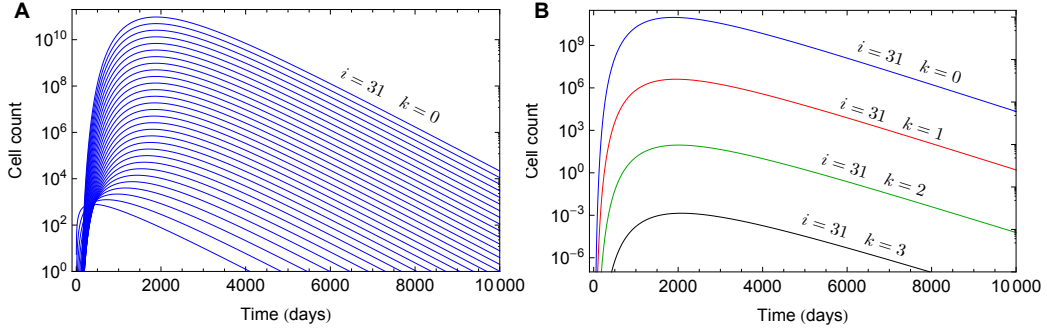


Figure 3.3: **A** Number of cells carrying no mutation in compartments 1 to 31 arising from compartment 1 containing 1000 cells. Lines show equation (3.3), with parameters $n_0 = 1000$, $\varepsilon = 0.85$, $\gamma = 1.26$, $u = 10^{-6}$ and $r_0 = 1/400$. Cells are more likely to differentiate than to self renew and thus progressively travel into more committed compartments. Initially, the cell count increases, but cells get washed out in the long run. The time scale is determined by the number of stem cell divisions. A stem cell is assumed to divide once a year, thus after 400 stem cell divisions a year passed. **B** Count of cells carrying zero to three mutations in compartment 31, given by equation (3.4). We used same parameters as in A. Cells carrying multiple mutations are exponentially suppressed.

The time to reach the maximum increases linearly with the number of additional mutations k . The cell count at the maximum becomes

$$N_1^k(t_{\max}) = \frac{(uk)^k}{k! (2\alpha - 1)^k} e^{-k} \approx \frac{u^k}{(2\alpha - 1)^k} \frac{1}{\sqrt{2\pi k}}, \quad (3.7)$$

where we used the Stirling formula to approximate $k!$. The maximum scales with u^k and thus decreases exponentially with k . In addition the factor $1/\sqrt{k}$ leads to a further suppression of the maximum for increasing k . However, the risk of additional mutations depends not on the maximal cell count, but on the reproductive capacity of a cell line. This reproductive capacity can be captured by the cumulative cell count. The number of cells within compartment i carrying k mutations produced until time t is given by

$$m_i^k(t) = r_i \alpha_i^k \int_0^t ds N_i^k(s), \quad (3.8)$$

and the reproductive capacity can be derived by taking the time limit to infinity. The general solution (3.4) allows to carry out the integral exactly by integration by parts. However the problem can be tackled from a different perspective, leading to a more transparent solution of (3.8) that is easier to handle.

3.2.2 Cell reproductive capacity

We call the cell subpopulation within a compartment i , that is derived by a single founder cell in an upstream compartment, the reproductive capacity of this founder cell. This idea directly corresponds to the method of neutral markers. We imagine a genetically marked cell somewhere in the hierarchy and count the offspring of this cell at any stage of the hierarchy. This corresponds to the total count of cells with same color in Fig 3.2.

Assume a single cell carrying no mutation in compartment 1. This cell differentiates with probability ε_1^0 into the next downstream compartment, mutates with probability u or produces an additional cell in compartment 1 with probability $1 - \varepsilon_1^0 - u$. We first discuss the probability, that a cell leaves compartment 1 after exactly l cell divisions. A cell can leave a compartment either by mutation or differentiation, before which the cell has to undergo $l - 1$ self renewals. Thus this probability becomes $(\varepsilon + u)(1 - \varepsilon - u)^{l-1}$. During this time, the cell population in compartment 1 derived from this single cell increased to 2^{l-1} cells, if all daughter cells share the same proliferation probabilities. With this, the reproductive capacity of a single cell in compartment 1 is on average

$$m_1^0 = \sum_{l=0}^{\infty} \alpha_1^0 2^l (1 - \alpha_1^0)^l = \begin{cases} \frac{\alpha_1^0}{2\alpha_1^0 - 1} & \alpha_1^0 > \frac{1}{2} \\ \infty & \alpha_1^0 \leq \frac{1}{2}. \end{cases} \quad (3.9)$$

The sum becomes infinite if $\alpha \leq 0.5$, as the probability to produce offspring in the founder compartment is higher than the probability to leave the compartment. Thus, the cell population continuously increases. Of course, under normal conditions, cells do not have an unlimited capacity to divide and serial telomere erosion amongst others will impose a physical limit on the number of divisions a cell can undergo (Hayflick et al., 1961; Armanios and Blackburn, 2012). The biologically more relevant case is $\alpha > 0.5$ and cells tend to differentiate into more committed compartments. In this case, the total number of offspring cells that arise from a single cell (i.e. a clone) is finite and given by (3.9). The number of cells m_i^0 in compartment i carrying no mutation, increases due to influx of cells via differentiation from compartment $i - 1$ and the expansion of these cells due to self renewal in compartment i . Thus we can write for the reproductive capacity of cells in compartment i without mutation

$$m_i^0 = 2 \frac{\varepsilon_{i-1}^0}{\alpha_{i-1}^0} m_{i-1}^0 \frac{\alpha_i^0}{2\alpha_i^0 - 1} = \frac{\alpha_i^0}{2\alpha_i^0 - 1} \prod_{l=1}^{i-1} \frac{2\varepsilon_l^0}{2\alpha_l^0 - 1}. \quad (3.10)$$

This can be generalized, and an expression for the reproductive capacity of cells in compartment i carrying k mutations can be derived. Cells in compartment i with k mutations are acquired either by differentiation of cells from compartment $i - 1$ that carry k mutations, by mutation of cells in compartment i carrying $k - 1$ mutations, or by self-renewal of cells already in compartment i and k mutations. With this, we can write

$$m_i^k = \left(2 \frac{\varepsilon_{i-1}^k}{\alpha_{i-1}^k} m_{i-1}^k + \frac{u}{\alpha_i^{k-1}} m_i^{k-1} \right) \frac{\alpha_i^k}{2\alpha_i^k - 1}, \quad (3.11)$$

where the two terms in the bracket represent cells either produced by differentiation or mutation, multiplied with the self renewal potential of these cells. This recurrence relation can be solved recursively

$$m_i^k = \frac{\alpha_i^k}{2\alpha_i^k - 1} \sum_{l=1}^i 2^{i-l} \frac{u}{\alpha_l^{k-1}} m_l^{k-1} \prod_{h=l}^{i-1} \frac{\varepsilon_h^k}{2\alpha_h^k - 1}. \quad (3.12)$$

Since, m_i^0 is given by (3.10) one can construct the explicit solution iteratively. Equation (3.12) allows for arbitrary parameters α_i^k and thus incorporates any mutational induced change in cell proliferation parameters. However, if $\alpha_i^k \leq 0.5$ the sum diverges. Cells with at least k mutations in and downstream compartment i will accumulate over time. In the following we discuss the general solution of equation (3.12) for mutations that are neutral relative to the founder cell.

3.2.3 Reproductive capacity of neutral mutants

We call a mutation neutral if the reproductive capacity of the mutant and the founder cell is equal. In the former subsection we have shown, that the reproductive capacity of a cell depends on its differentiation probability ε and its mutation rate u , but interestingly it is independent of the reproduction rate r . Thus, our definition of neutral mutations only requires constant differentiation probabilities and mutation rates relative to the founder cell. This assumption allows us to write $\varepsilon_i^k = \varepsilon_i$ and thus the number of parameters is reduced from $(k + 1)i + 1$ for the general case to $i + 1$ for the neutral case. This number can be reduced to two parameters u and ε , if a constant differentiation probability for all non stem cell stages is assumed, $\varepsilon_i = \varepsilon$. This simplifies the evaluation of the recurrence relation (3.12) significantly. The reproductive capacity m_i^k of neutral mutations in compartment i carrying k mutations becomes

$$m_i^k = \alpha \frac{u^k}{k!} \frac{(2\varepsilon)^{i-1}}{(2\alpha - 1)^{i+k}} \prod_{l=1}^k (i + l - 1) = \alpha u^k \frac{(2\varepsilon)^{i-1}}{(2\alpha - 1)^{i+k}} \binom{i + k - 1}{k}. \quad (3.13)$$

Mutants carrying k mutations are suppressed by a factor u^k and thus are rare in early differentiation stages. The number increases exponentially for downstream compartments, and a significant load of cells carrying few mutations can be observed in late differentiation stages, see Fig. 3.4.

Equation (3.13) reveals interesting properties of hierarchical tissue structures. The ratio of cells carrying k mutations to cells carrying $k - 1$ mutations in compartment i is

$$\frac{m_i^k}{m_i^{k-1}} = \frac{u}{2\alpha - 1} \left(1 + \frac{i-1}{k} \right). \quad (3.14)$$

The ratio increases with compartment number, but the increase becomes flatter for increasing k . The compartment structure leads to an additional suppression of cells carrying multiple mutations and thus is a protection mechanism against cancer initiation. The ratio is constant for $i = 1$, the compartment of the mutant origin. The protection mechanism affects downstream compartments only.

On the other hand, the scaling properties for more differentiated cells show interesting properties also. The ratio of cells with k mutations in compartment $i + 1$ to cells with k mutations in compartment i is given by

$$\frac{m_{i+1}^k}{m_i^k} = \frac{2\varepsilon}{2\alpha - 1} \left(1 + \frac{k}{i} \right). \quad (3.15)$$

The increase of cells is constant for cells carrying no mutations $k = 0$. It increases with k , but is suppressed within the hierarchy by the factor of $1/i$.

3.2.4 Number of distinct neutral mutations

So far, we have discussed the reproductive capacity of cells. However we did not distinguish between different mutations, but grouped together cells with an equal number of mutations. Often experimental studies focus on mutation landscapes, investigating the variation in clonal loads in healthy and sick individuals. Our approach allows us to estimate the expected number of distinct mutations that arise from a single founder cell, corresponding to the count of distinct symbols with same color in Fig. 3.2. The founder cell can have arbitrary cell proliferation parameters.

Let us assume that every mutation event is unique. Thus we neglect the possibility that the same mutation is derived twice independently. The diversity in compartment i increases either by additional mutations of cells in compartment i , or differentiation of clones from compartment $i - 1$ into compartment i . Assuming

a differentiation probability $\varepsilon > 0$ the expected diversity n_i^{k+1} of cells with $k + 1$ mutations in compartment i is

$$n_i^{k+1} = u \sum_{j=1}^i m_j^k = \alpha u^{k+1} \sum_{j=1}^i \frac{(2\varepsilon)^{j-1}}{(2\alpha - 1)^{j+k}} \binom{j+k-1}{k}. \quad (3.16)$$

As an example, if we use the parametrization of the hematopoietic system, $\varepsilon = 0.85$ and $u = 10^{-6}$, we find approximately $n_{31}^1 \approx 30$ distinct single mutations in compartment 31, that were derived from the clonal progeny of a single founder cell in compartment 1. But the founder cell by chance could carry a mutation, that changes their cell proliferation parameters. For example, the differentiation probability of the founder cell could change to $\varepsilon = 0.75$. In this case, Eq. 3.16 estimates 28000 distinct mutations derived from this single founder cell. We note, that the above change of the differentiation probability from 0.85 to 0.75 is sufficient to explain the manifestation of chronic myeloid leukemia in otherwise healthy adults (Dingli et al., 2008a; Lenaerts et al., 2011). However, Eq. (3.16) represents an average and in individual cases fluctuations, caused by stochasticity are expected. But also small changes in ε or u influence the expected diversity significantly. A linearized error analysis (Taylor, 1997) reveals the dependency of (3.16) on the uncertainties of u and ε , which is given by

$$\Delta n_i^{k+1} = \sum_{j=1}^i \frac{(2\varepsilon)^{j-1}}{(2\alpha - 1)^{j+k+1}} (g_j^\varepsilon + g_j^u) \binom{j+k-1}{k}, \quad (3.17)$$

with $g_j^\varepsilon = u^{k+1} |2\alpha(1-j-k) - 1| \Delta\varepsilon$ and $g_j^u = u^k |(k+1)\varepsilon(2\varepsilon-1) + u(2(k-j+3))\varepsilon - k - 2 + 2u^2(2-j)| \Delta u$. If we assume $\Delta u = 10^{-7}$ and $\Delta\varepsilon = 0.01$ the uncertainty given by (3.17) becomes $\Delta n_{31}^1 \approx 35$, where the individual error contributions in u and ε are 7 and 28 respectively. Especially note the strong dependency on $\Delta\varepsilon$. If we choose $\Delta\varepsilon = 0.05$, a deviation, that might be difficult to detect in vivo, one gets $\Delta n_{31}^1 \approx 150$. Thus small variations in ε lead to significant differences in the expected diversity of clonal populations, one aspect that might contribute to the explanation of the observation of very diverse mutation landscapes. Note also, that an increasing mutation rate $\Delta u = 10^{-6}$ gives $\Delta n_{31}^1 \approx 65$. Of course, higher mutation rates increase the expected diversity of clonal populations. However, a higher mutation rate (or genomic instability) is neither the exclusive nor necessarily the dominant underlying cause of the diversity in the mutation landscape that is observed.

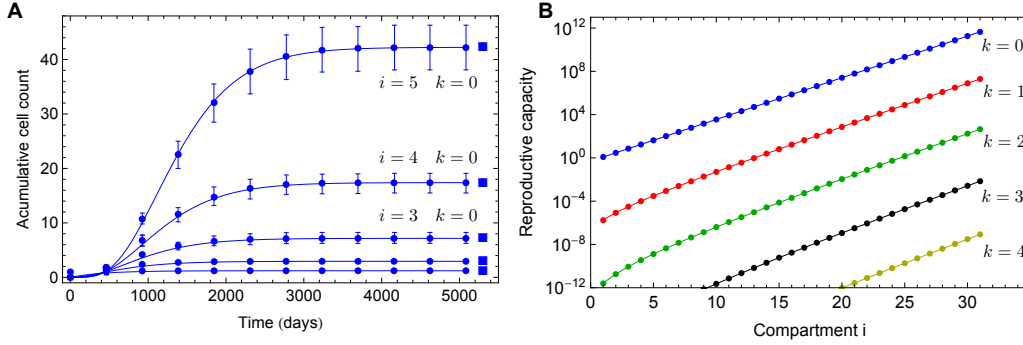


Figure 3.4: **A** Number of cells without mutations in compartments 1 to 5, arising from a single cell in compartment 1. Lines are equation (3.8), symbols are averages with corresponding standard deviations over 10^3 independent runs of stochastic individual based computer simulations and squares are equation (3.13). Parameters are $n_0 = 1$, $\varepsilon = 0.85$, $\gamma = 1.26$, $u = 10^{-6}$ and $r_0 = 1/400$. **B** Reproductive capacity of a single founder cell in compartment 1. Shown is the number of cells with 0 to 4 mutations in the first 31 compartments, acquired from a single cell in compartment 1. Symbols are numerical solutions of (3.8) in the limit of infinite time and lines are equation (3.13). The reproductive capacity increases exponentially for increasing compartment number. Cells carrying multiple mutations are strongly suppressed within a hierarchical tissue structure, see equation (3.14).

3.2.5 Example: clonal diversity in acute lymphoblastic leukemia

Let us now consider a specific example using data from childhood acute lymphoblastic leukemia (ALL). The most common chromosomal abnormality in this disease is the t(12;21) translocation that results in the fusion gene ETV6-RUNX1 (also known as TEL/AML1). There is evidence that this mutation often arises in utero and this has been confirmed to be the case in at least one pair of monozygotic twins (Hong et al., 2008). There is evidence that this mutation is a founder mutation and is considered critical for the disease. Cells that express this fusion gene appear to have a higher self-renewal and enhanced survival compared to normal cells (Fischer et al., 2005; Hong et al., 2008). In our model, enhanced self-renewal implies a reduced differentiation probability for the cells carrying the mutation ($\varepsilon < 0.85$). Recently, Ma et al. performed whole genome sequencing on leukemic cells isolated from two pairs of monozygotic twins. In one pair of twins, the initial event occurred in utero since the ETV6-RUNX1 fusion was shared by both siblings (Ma et al., 2013). They found that the incidence of non-synonymous single nucleotide mutations between

the samples ranged from 708 to 1237 (Ma et al., 2013). These mutations must have occurred after the founder mutation and independent of each other. The time from the putative appearance of the shared ETV6-RUNX1 mutation and disease was 48 to 55 months. The second pair of monozygotic twins shared a mutation in NF-1, and the time to diagnosis of ALL was 72 to 77 months after birth. The tumors in these two children had 949 to 975 unique non synonymous single nucleotide mutations. Using these constraints and an estimate of 10%-75% cancer cells at diagnosis as well as $N_0 = 100$ Dingli and Pacheco (2007), we use equation (3.4) to estimate the differentiation probability ε for the mutant cells, which is then in the range 0.78 to 0.81, ie. only slightly lower than that of normal cells. Based on equation (3.4), it takes approximately 50 to 80 months to reach this load. We further predict a range of 350 to 2600 distinct mutations by utilizing equation(3.16), assuming a mutation rate of $u = 10^{-6}$ and the above range of differentiation probabilities. Thus, we expect slightly more distinct mutations than found in patients. But some of those theoretically predicted clones are very small and may escape detection. For example, if we neglect mutations, that occur in the last differentiation step of cells in compartment 30 to 31 (expected to be very small cell populations), we predict only 150 to 950 distinct mutations. Clearly, the model as presented can explain the large number of passenger mutations that can be expected in a typical patient with ALL and likely other types of leukemia.

3.3 Discussion

The accumulation of multiple mutations in cells is considered to be critical for cancer initiation. However, mutations unavoidably accompany cell proliferation and thus cancer can potentially occur in any multicellular organism. Hierarchical tissue structures contribute to the protection against such mutations. So far, the suppression of single mutations in hierarchical tissue structures has been the focus. Here, we have shown that in addition, the risk of the accumulation of multiple mutations is dramatically reduced by a hierarchical tissue organization. The cell population is divided into few slow proliferating stem cells and many faster proliferating progenitor cells. Stem cells have an almost infinite cell reproductive capacity, but the manifestation of a critical mutational load often exceeds an organisms expected natural life time. Progenitor cells proliferate faster, but their reproductive capacity usually is limited and thus they give rise to clonal waves traveling through the hierarchy. Still

both cases can be observed. There are cancers that presumably originate from stem cell mutations for example chronic myeloid leukemia, and there are cancers that originate in later stages of hematopoiesis, for example acute promyelocytic leukemia and various other subtypes of acute myeloid leukemia (Guibal et al., 2009). Understanding the clonal dynamics for both cases is of importance. Here, we have focused on the accumulation of non stem cell driven mutations in hierarchically organized tissues. We arrived at closed solutions for the clonal waves traveling through the hierarchy. From this the reproductive capacity of cells can be deduced. This allows us to predict the expected risk to acquire any number of mutations from one single cell, given the proliferation properties of this cell. We derived equations that allow the quantitative classification of multiple mutations in hierarchically organized structures, highlighting the strong suppression of clones carrying multiple mutations by this architecture.

Another important question emerged more recently with the accessibility of whole genome sequencing technology. These techniques revealed very complex and diverse mutation landscapes for many different cancers, with the classification of driver and passenger mutations as the final goal. Knowing the exact driver mutations might help to understand the properties of specific cancer cells, allowing the development of effective treatment strategies. A promising example is the design of molecularly targeted agents such as the various tyrosine kinase inhibitors (Imatinib, Nilotinib etc) to treat patients with chronic myeloid leukemia. These molecules specifically bind to kinase domains, encoded by the BCR-ABL oncogene and strongly suppress the proliferative capacity of these cells (Druker et al., 2006; Saglio et al., 2010; Werner et al., 2011). Our work can contribute to this question by predicting the average number of distinct neutral (passenger) mutations, acquired from a single cell at any stage of the hierarchy. This approach directly corresponds to the method of a neutral marker. The genetically marked cell represents the founder cell of the clonal population and one follows the offspring of this founder cell throughout the hierarchy. This enables the prediction of the size and the variability of the clonal population. For example, in normal hematopoiesis, we expect cells to have a differentiation probability of $\varepsilon = 0.85$, leading to approximately 30 distinct single mutations (subclones) in adult cells, acquired from this single cell. If this founder cell acquired a mutation, that changed the differentiation probability to $\varepsilon = 0.75$ by chance, the expected number of distinct single mutations increases to approximately 28000.

We have shown how even a slight change in the self-renewal probability of progenitor cells can lead to substantial differences in the number of passenger mutations observed in ALL. This likely holds true for other malignancies. In acute promyelocytic leukemia, apart from the t(15;17) that is a critical event in the origin of this disease (akin to the ETV6-RUNX1 discussed above for ALL), approximately 440 non synonymous single nucleotide mutations were found that were unique to the tumor clone (Welch et al., 2012). Interestingly, it is highly likely that the cell of origin of APL is downstream (a progenitor cell) of the cell of origin of ETV6-RUNX1 driven ALL and this may, in part, explain the less diverse mutational landscape reported in APL compared to ALL and would fit well with our model. Of course, any genomic instability will further increase the repertoire of passenger mutations that is observed in any given tumor.

We also note that the effect of a specific mutation on a cell need not be large for the effect to spread throughout the tumor. Tissue architecture and dynamics as in hematopoiesis serve as a deterrent against the accumulation of mutations, in particular multiple mutations occurring in one cell. Once a driver mutation appears, if this changes either the self-renewal of the cell or the mutation rate, then the appearance of many passenger mutations becomes inevitable in such an architecture due to the amplification of cells that occurs. Thus, a minor change in the differentiation probabilities that might be difficult to detect in vivo drastically changes the expected number of passenger mutations that would be observed. Concomitantly, this will increase the risk of acquiring mutations in 'driver genes' and so lead to malignancy. However, if the initiation of disease requires several co-occurring mutations, a hierarchical tissue structure is a powerful mechanism of tumor suppression.

Resistance evolution

We argued in the former chapter, that most cancers are caused by multiple mutational hits in cells. However, cancers exist that evolve from a single mutational hit, for example chronic myeloid leukemia. We discussed in 1.2.3, that this mutation codes for an active kinase domain with a distinct spatial structure. This allowed the development of molecular targeted drugs, that strongly suppress the proliferation capacity of those cancer cells. Unfortunately, resistance to those drugs can occur, imposing an increasing problem in daily clinical care.

In this chapter we discuss the dynamics of a resistance inducing experiment in BCR-ABL positive cells. The experiments were performed by David Lutz under supervision of Stefan Balabanov and Tim Brümmendorf at the university hospital in Hamburg Eppendorf. They harvested in total 10 cultures of BCR-ABL positive cells. Each cell population contained approximately 2×10^5 cells. These cell cultures were exposed to increasing concentrations of the tyrosine kinase inhibitor Imatinib. After 12 days, the Imatinib concentration was increased and death cells were replaced by BCR-ABL positive wild types, to assure an approximately constant population size. The ratio of death and alive cells was recorded, allowing a rough estimate of cell fitness. The experiment ran for 120 days and all populations were scanned for resistance inducing mutations. Initially a pronounced decrease of cell fitness could be observed. However, in the long run cells adapted and high concentrations of tyrosine kinase inhibitors affected cell fitness less. The fitness of the cell population passes through a very characteristic minimum to reach the resistance state. We are especially interested in the underlying deterministic dynamics. By applying a minimalistic mathematical model we hope to infer some important parameters of the system.

In the following we shortly discuss the solution of the quasi species equation with time dependent fitness functions. The subsequent main body of this chapter is then based on the publication (Werner et al., 2011).

4.1 Quasi species equation with time dependent fitness

We apply the individual based stochastic Moran process, in detail described in section 1.3.2, to model the population dynamics observed in the experiment. The population consists of initially 2×10^5 wild type cells. Importantly, as the concentration of Imatinib increases, the fitness f_w of wild types decreases. We approximate this decrease by a linear time dependence of the wild types fitness,

$$f_w = \begin{cases} 1 - \frac{t}{t^*} & t < t^* \\ 0 & \text{otherwise.} \end{cases} \quad (4.1)$$

The baseline fitness of wild types is 1. The linear time decay is an arbitrary choice, however, other decaying functions such as a negative exponential does not change the dynamics significantly. The fitness of resistant types is chosen constant $f_r = b$, where the experimental results suggest $0 < b < 1$. In chapter 1.3.3 we derived a deterministic approximation of the Moran process in the limit of large population size N . The deterministic equations for the two type case became replicator equations of the form

$$\dot{x} = x(1-x)(f_1 - f_2). \quad (4.2)$$

Here we use a modified version of the replicator mutator equation (1.38). We chose $f_1(x_1, x_2) = f(t)$ and $f_2(x_1, x_2) = f_2(t)$, where t represents time. Additionally type 1 can mutate into type 2 with rate u , but backward mutations are neglected. With this restrictions, equation (1.38) becomes

$$\dot{x}_1(t) = x_1(t)[1 - x_1(t)][f_1(t) - f_2(t)] - u f_1(t) x_1(t), \quad (4.3)$$

and $x_2 = 1 - x_1$. By reordering terms, this can be written as

$$\dot{x}_1(t) + \underbrace{[f_2(t) - (1-u)f_1(t)]}_{a(t)} x_1(t) = \underbrace{[f_2(t) - f_1(t)]}_{b(t)} x_1^2(t), \quad (4.4)$$

which is a special case ($\alpha = 2$) of the nonlinear differential equation

$$\dot{x}(t) + a(t)x(t) = b(t)x^\alpha(t) \quad \alpha \neq 0, 1, \quad (4.5)$$

known as the Bernoulli differential equation (Bernoulli, 1695). Thus the time dependent quasi species equation with directed mutation (4.3) corresponds to the Bernoulli differential equation and the solution of (4.5) can be reduced to quadrature for general time dependence of $a(t)$ and $b(t)$. In our case $\alpha = 2$, the first

step to the general solution is the variable transformation $y(t) = x^{-1}(t)$ and the multiplication with an integrating factor $\lambda(t)$. This transforms (4.5) into

$$\lambda(t) \dot{y}(t) + \lambda(t) a(t) y(t) = \lambda(t) b(t). \quad (4.6)$$

Further we observe, that

$$\frac{d}{dt} [\lambda(t) y(t)] = \underbrace{\lambda \dot{y} + \dot{\lambda} y}_{\dot{\lambda} = a\lambda} = \lambda \dot{y} + \lambda a y, \quad (4.7)$$

yielding a simple differential equation in λ with the solution

$$\lambda(t) = e^{\int ds a(s)}. \quad (4.8)$$

However, equation (4.7) can also be written as

$$\frac{d}{dt} \underbrace{[\lambda(t) y(t)]}_{g(t)} = b(t) \lambda(t), \quad (4.9)$$

giving a differential equation $g(t)$, that can be solved by a separation of variables, finally leading to the general solution

$$y(t) = \frac{\int dt' b(t') e^{\int ds a(s)}}{e^{\int ds a(s)}}. \quad (4.10)$$

Thus, equation (4.3) can be reduced to the calculation of some integrals for general functions $f_1(t)$ and $f_2(t)$. With our choice $f_1 = f_w = 1 - \frac{t}{t^*}$ and $f_2 = f_w = \text{const}$, these integrals can be solved.

In the next section, we use this deterministic description of the Moran process to derive analytical solutions of the underlying dynamics for the two type case. However, it turns out, that the two type case is insufficient to describe the experimental results completely. Thus we generalize our model to three possible types. Analytical results for the three type case are out of reach, but we can compare simulation results of the three type Moran process to the experimental dynamics.

4.2 Introduction

The variety of approaches in cancer research ranges from clinical studies, experimental studies *in vivo* and *in vitro* and genetic analysis to computational and mathematical models. Due to the insights from all these domains it was possible to develop specific targeted therapies for many cancer types and maybe cures for some are within reach (Mahon et al., 2010). A very promising example is the progress in treatment of patients with chronic myeloid leukemia (CML). CML is caused by a single reciprocal chromosomal translocation between the long arms of chromosome 9 and 22, the so called Philadelphia chromosome, in a hematopoietic stem cell (Rowley, 1973). The resulting BCR-ABL fusion gene encodes for a constitutive active tyrosine kinase, which is expressed and active in almost all haematopoietic cells and which leads to accelerated cell cycle activity (Deininger et al., 2000). As a result predominantly white blood cells at all stages of differentiation are significantly increased in peripheral blood and bone marrow of affected patients. Some years ago the standard treatment in the chronic phase was interferon-alpha and after disease progression chemotherapy with or without hematopoietic stem cell transplantation, a procedure that is associated with significant side effects and risks. The treatment algorithms changed due to the development of tyrosine kinase inhibitors (TKI) such as Imatinib (Druker et al., 2006), Nilotinib (Weisberg et al., 2005), Dasatinib (Shah et al., 2004) and Bosutinib (Cortes et al., 2011). These molecules bind specifically to the kinase domain of BCR-ABL and thereby strongly suppress the proliferation capability of CML cells. Since normal cells are less affected by this treatment, it allows normal hematopoiesis to be restored (Lenaerts et al., 2010). Long time clinical studies investigating the effect of TKIs show overwhelming success (Druker et al., 2006; Saglio et al., 2010; Kantarjian et al., 2010). Unfortunately, there are cases in which patients do not respond to the drug or develop resistance during the treatment. This raises the question of how BCR-ABL positive cells bypass inhibition and develop resistance. A large number of *in vivo* and *in vitro* studies were performed, which identified different mutations causing resistance to TKIs (O'Hare et al., 2007; Bixby and Talpaz, 2009). However, the spatial structure and thus the molecular mechanism of kinase domain binding varies between different tyrosine kinase inhibitors and thus clones resistant to one inhibitor may be sensitive to an alternative inhibitor. Unfortunately, there are clones resistant to several inhibitors at the same

time. For example the mutation T315I, the so called gatekeeper mutation, causes resistance to all approved TKIs. Agents as Ponatinib have been particularly designed to bind to such mutated kinase domains of BCR-ABL and suppress cell proliferation. These are in clinical trials and showing promising results (Hochhaus et al., 2011; Demidenko et al., 2005; O'Hare et al., 2009). Also many mathematical models concentrate on the probabilities and the dynamics of the development of such cross resistance and how to reduce the risk of resistance evolving to a minimum using combination therapies of different drugs (Katouli and Komarova, 2010; Komarova and Wodarz, 2009; Komarova et al., 2009). We will not focus on this aspect in the following, but address the dynamics of resistance development instead.

Here, we discuss an experimental setup that induces resistance to Imatinib in initially sensitive BCR-ABL positive Ba/F3-p210 cells. In the following, our aim is not to assess the probability of the occurrence of a specific resistance-conferring mutations nor to identify novel mutations, but to focus on the clonal dynamics that occurs during such an experimental setup. We show that not only mutations, but also intracellular processes that do not lead to any mutations can lead to the development of resistance. We investigate the average fitness of cancer cells exposed to a drug concentration that increases over time in this in vitro experiment. Basic features of this dynamic can be explained by a mathematical model, assuming a fitness of cancer cells that decreases with time as the drug concentration increases. By comparing the experimental data with simulation and analytical results we were able to assess individual parameters in this model and gain insights into the general cancer cell population dynamics under such environmental situations.

4.3 Experimental setup

We shortly describe some details of the resistance inducing experiment, that was performed by David Lutz and Stefan Balabanov at the university hospital in Hamburg Eppendorf. Below, we compare the population dynamics observed in the experiment to predictions of our mathematical model.

Materials

Imatinib was purchased from Toronto Research Chemicals Inc, Ontario, Canada. Stock solutions of IM (10mg/mL; in DMSO/H₂O (1:1)) was stored at 20°C. Murine

BCR-ABL transduced pro-B cell line Ba/F3-p210 was obtained from N.P. Shah and C.L. Sawyers (UCLA, USA). Cells were cultured under standard conditions as previously described (Balabanov et al., 2011) .

Cell culture approach for induction of Imatinib resistance

In this experiment, Ba/F3-p210 cells carrying the BCR-ABL kinase were used as a model for resistance development to IM treatment, following a well established protocol (Burchert et al., 2005). Flow cytometry based cell sorting using a FACS Aria (BD Biosciences) was employed to produce 10 clones of BaF/3-p210 cells. These clones were expanded to cultures of 2×10^5 cells each. These cultures were exposed to Imatinib in a concentration increasing over time, starting with $0.1 \mu\text{M}$. Cell viability was measured by the Trypan blue exclusion method (Patterson, 1979). The Imatinib concentration was periodically increased by $0.1 \mu\text{M}$ only if cell viability was above 70%, the concentration stays unchanged if the cell viability is between 30% and 70% and withdrawn in case of cell viability less than 30%. Thus the time the Imatinib concentration was increased differed for different clone lines, but the concentration was always increased within 12 days. Parallel, a control with untreated clones was performed. At the end of the experiment all Imatinib treated lineages were resistant to the drug, confirmed by a cytotoxicity MTT-assay.

MTT assay

The 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazoliumbromide (MTT) assay was performed as previously described (Balabanov et al., 2007). In brief, BAF/3-p210 cells were plated into 96-well flat-bottomed microtiter plates (Becton Dickinson, Heidelberg, Germany) at 1.5×10^4 cells/well in $150 \mu\text{M}$ of their respective media. Cells were preincubated for 24 hours before increasing concentrations of Imatinib (0 - $10 \mu\text{M}$) were added. All analyses were performed in triplicates. After 48 hours, the viable cells in each well were assayed. The dose-response effect for Imatinib at the point of inhibitory concentration (IC_{50}) was analyzed by the median-effect method with CalcuSyn Software (Biosoft, Cambridge, United Kingdom) (Chou and Talalay, 1981).

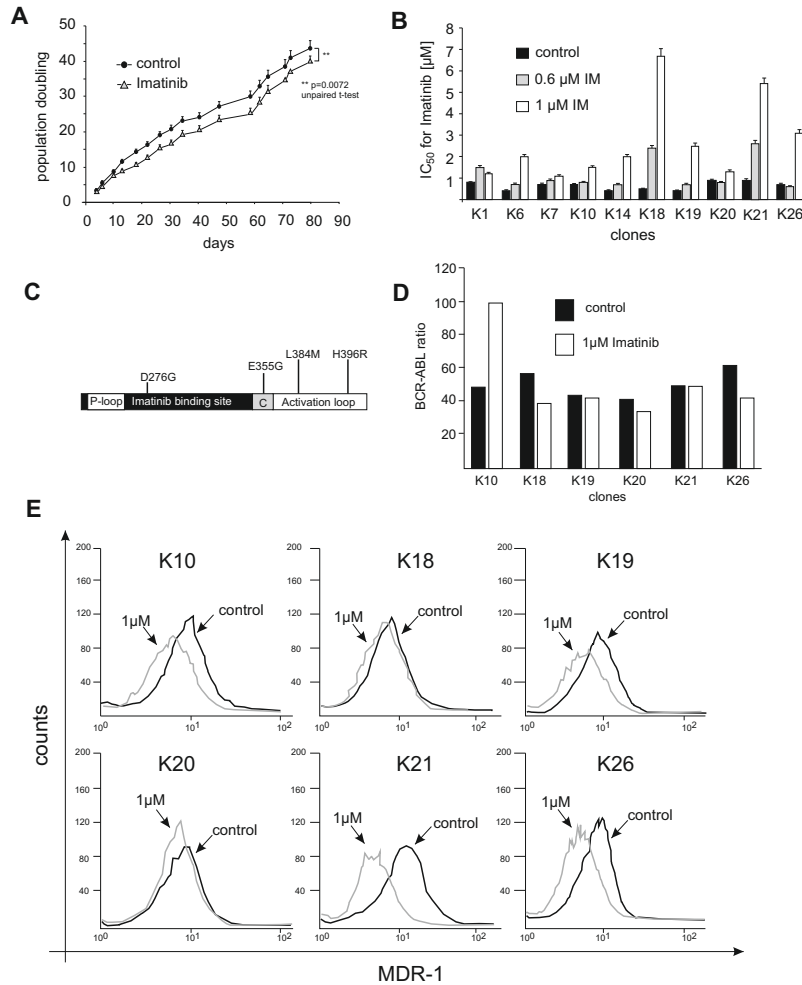


Figure 4.1: **Experimental results.** **A)** Average population doubling time observed for BAF/3-210 cells with (white triangles) and without (black dots) long-term Imatinib treatment. The number of population doubling for control cells was significantly higher compared to IM-treated cells. **B)** Measurement of IC_{50} for Imatinib in control and resistant clones. The long-term presence of Imatinib increase the IC_{50} in all resistant cell clones. In particular, cell clones which were growing at 1 μM Imatinib showed a marked increase of the IC_{50} value. The cell viability was determined using MTT assay and IC_{50} was calculated CalcuSyn software. **C)** Schematic representation of in vitro detected BCR-ABL mutations. **D)** Quantification of BCR-ABL transcript by real-time PCR in resistant clones without BCR-ABL mutations. Only one clone (K10) exhibited an increased expression of BCR-ABL. **E)** Quantification of MDR-1 expression by flow cytometry in resistant clones without BCR-ABL mutations. Imatinib resistance was not associated with an increased expression of MDR-1.

Mutagenesis screen and real time PCR for BCR-ABL transcript

To identify possible BCR-ABL kinase domain mutations as mechanism of Imatinib resistance, the coding cDNA of the kinase domain was sequenced. For all RT-PCR reactions, total RNA was isolated using TRIzol (Invitrogen). cDNA was prepared by reverse transcription of 250 ng RNA with oligo(dT)15 and Superscript II reverse transcriptase (Invitrogen) and amplified using REDTaq® ReadyMix® PCR Reaction Mix (Sigma-Aldrich). BCR-ABL allele was amplified by nested PCR using BCR forward primer B2A(5'-TTCAGAAGCTTCTCCCTGACAT-3') and ABL reverse primer 4065 (5'-CCTTCTCTAGCAGCTCATACACCTG-3'). Nested PCR reactions were performed using BCR F4 forward (5'-ACAGCATTCGCTGACCATCAATA-3') and reverse U396 (5'-GCCATAGGTAGCAATTTCCC). PCR products containing the kinase domain were sequenced with both forward primer 3306F (5'-TGGTTCATCATCATTCAACGG-3') and reverse primer 4000R (5'-GGACATGCCATAGGTAGCA-3'). Real time PCR for BCR-ABL was performed as previously described (Beillard et al., 2003; Gabert et al., 2003). Amplification of ABL was used as a internal housekeeping gene.

Quantification of MDR-1 on resistant clones

For the relative quantification of MDR-1 we used a flow cytometry-based assay. Cells were fixed in 2% formaldehyde for 10 minutes at 37°C, chilled on ice for 1 minute, and then permeabilized with ice-cold 90% methanol for 30 minutes on ice. From each sample, 2×10^5 cells were washed with 2 mL incubation buffer (PBS/0.5% bovine serum albumin), centrifuged at 50g for 5 minutes, and resuspended in 100 μ L of incubation buffer with 2.0 μ L of a MDR-1 specific antibody (abcam, Cambridge, UK). After 45 minutes of incubation at RT, cells were washed twice, resuspended in 100 μ L incubation buffer with 0.5 μ L of the secondary antibody (anti-rabbit IgG FITC-conjugate; Jackson ImmunoResearch Europe, Newmarket, United Kingdom), and incubated at RT for 30 minutes in the dark. After washing two times with washing buffer, cells were analyzed using flow cytometry.

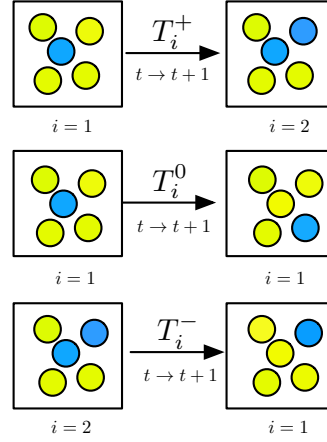


Figure 4.2: **Individual based stochastic simulation as two type Moran process.**

We assume two possible cell types, wild type leukemic cancer cells sensitive to Imatinib (yellow) and resistant cancer cells (blue). In total, $\approx 2 \times 10^5$ cells were included in the experiment and simulations. At time t there are i resistant cells and $N - i$ wild type cells. During the next time step $t + 1$ three cases are possible: (i) the number of resistant cells increases by one with probability T_i^+ , (ii) stays the same with probability $T_i^0 = 1 - T_i^+ - T_i^-$ (iii) or decreases by one with probability T_i^- , (see equation (4.12a) and (4.12b)).

4.4 Mathematical model and results

We compare the dynamics in the mathematical model to the experiment based on the fitness of cells. We assume that the average fitness of the experimental cell population is the ratio between the number of vital cells and the total number of cells. Mathematically we model the cell population dynamics observed in the experiment by an individual based stochastic approach, applying a standard two type Moran process (Moran, 1962; Nowak, 2006a). We assume a population of $N = 2 \times 10^5$ cells consisting of two subpopulations, wild type leukemic cancer cells sensitive to Imatinib and Imatinib resistant cancer cells. Initially, only wild type cancer cells are present, but transformations from wild types into the resistant type are possible. In each time step, one cell is chosen to reproduce and one cell is chosen to die, thus the total number of cells stays constant and the subpopulation sizes can change at most by 1 cell during one time step, see Fig. 4.2. If a wild type cancer cell is chosen to reproduce, an Imatinib resistant cancer cell is produced with a transformation probability u . To avoid complications, we assume that resistant types cannot switch back to nonresistant types. This would not change the basic

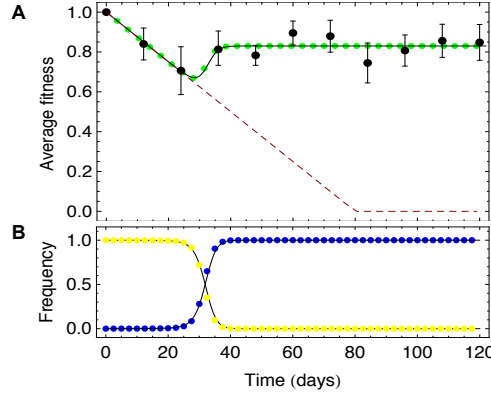


Figure 4.3: **Dynamics of resistance development in the experiment and the mathematical model.** **A)** Average fitness of the total population due to experiments (black dots), stochastic simulations (green dots) and analytical results (black line, due to equation (4.17)). The parameters of simulation and calculation were chosen to $N = 2 \times 10^5$, $u = 1 \times 10^{-4}$, $t^* = 80$ days and $b = 0.83$. The dashed red line shows the linear decrease of the wild type fitness, the slope is determined by the critical time t^* . **B)** The average frequency of wild type cells due to simulation (yellow dots) and calculation (black line), as well as the frequency of resistant cancer cells (blue dots) over time. We always start with wild types only, but since the system selects for resistant cells, they fixate in the long run.

results, however. The two types differ in their fitness f_w (for the wild type) and f_r (for the resistant type). We assume a time dependent wild type fitness function, due to the increasing concentration of Imatinib in the experiment. We set the fitness of untreated wild types f_w to one and assume that it decreases linearly with increasing Imatinib concentration,

$$f_w(t) = \begin{cases} 1 - \frac{t}{t^*} & 0 \leq t \leq t^* \\ 0 & \text{otherwise.} \end{cases} \quad (4.11)$$

The parameter t^* determines the impact of Imatinib on wild type cells. Choosing other decreasing functions, e.g. an exponentially decreasing function, does not change the results qualitatively. For the Imatinib resistant subpopulation we assume constant fitness, $f_r = \text{const.}$ For $f_r < f_w$, resistant cells proliferate slower than wild type cells, for $f_r > f_w$ they proliferate faster than wild type cells. The experimental data suggest $0 < f_r < 1$ and thus initially resistant types would initially be out-competed by wild type cells, see below. Due to the decreasing fitness of wild type cells, we can go from one domain to the other in time. We assume

that cell reproduction occurs proportional to fitness and simultaneously a random cell is chosen to die. Thus, the probability of increasing or decreasing the number of resistant cancer cells i at time t is notated by T_i^+ or T_i^- respectively,

$$T_i^+ = \left[\frac{if_r}{if_r + (N-i)f_w} + u \frac{(N-i)f_w}{if_r + (N-i)f_w} \right] \frac{N-i}{N} \quad (4.12a)$$

$$T_i^- = \left[(1-u) \frac{(N-i)f_w}{if_r + (N-i)f_w} \right] \frac{i}{N}. \quad (4.12b)$$

The number of resistant cells stays unchanged with probability $T_i^0 = 1 - T_i^+ - T_i^-$. One generation consists of N reproduction and death events, such that each cell reproduced on average once.

4.4.1 Analytical approximation for large population size

If the population size N is large, we can describe the averages of our individual based model by a deterministic approach. In this case, we consider the frequency (relative population size) of the two cell types, wild type leukemic cells with frequency $w(t) = (N-i)/N$ and Imatinib resistant cell types with frequency $r(t) = i/N$. We use the same fitness functions as above and also assume transformations from the wild type to the resistant type. The average fitness of the system, which has been measured in the experiment, is given by $\Phi(t) = w(t)f_w(t) + r(t)f_r$. The change in wild type and resistant type frequencies for large populations are given by a replicator mutator equation (Page and Nowak, 2002b; Nowak, 2006a)

$$\dot{w} = w(f_w - \Phi) - uf_w w \quad (4.13a)$$

$$\dot{r} = r(f_r - \Phi) + uf_w w, \quad (4.13b)$$

where the dot represents a derivative with respect to time t . With the functions from above, these differential equations take the form

$$\dot{w} = \begin{cases} w(1-w)(1 - f_r - \frac{t}{t^*}) - u(1 - \frac{t}{t^*})w & 0 \leq t \leq t^* \\ w(1-w)(-f_r) & t > t^* \end{cases} \quad (4.14a)$$

$$\dot{r} = \begin{cases} r(1-r)(f_r - 1 + \frac{t}{t^*}) + u(1 - \frac{t}{t^*})w & 0 \leq t \leq t^* \\ r(1-r)f_r & t > t^*. \end{cases} \quad (4.14b)$$

These are time dependent nonlinear differential equations of the Bernoulli type, for which a general solution scheme exist (Bernoulli, 1695). Thus the system of differential equations can be reduced to integrals, which in this situation are solvable.

Due to $w(t) + r(t) = 1$, we only have to solve one of the equations. The general solution for $w(t)$ with respect to the initial condition $w(0) = 1$ for $t \leq t^*$ reads

$$w(t) = \frac{1 - u}{1 - ue^{h_t^2 - h_0^2} + \sqrt{\frac{\pi t^*}{2}} \frac{uf_r}{\sqrt{1-u}} e^{h_t^2} [\text{Erf}(h_t) - \text{Erf}(h_0)]}, \quad (4.15)$$

where $h_t = \frac{(1-u)\frac{t}{\sqrt{t^*}} + \sqrt{t^*}(f_r - 1 + u)}{\sqrt{2(1-u)}}$ and $\text{Erf}(h_t) = \frac{2}{\sqrt{\pi}} \int_0^{h_t} ds e^{-s^2}$ is the error function. For $t \ll 1$, this can be approximated by

$$w(t) \approx 1 - ut + u \frac{1 + (1 + u - f_r)t^*}{2t^*} t^2 + o[t^3].$$

For $t \geq t^*$, the number of resistant cells grows logistically and consequently, the number of wild type cells decreases logistically

$$w(t) = \frac{1}{1 + \frac{1-w(t^*)}{w(t^*)} e^{-f_r(t-t^*)}}. \quad (4.16)$$

The average fitness of the total population becomes

$$\Phi(t) = f_r(1 - w(t)) + \left(1 - \frac{t}{t^*}\right) w(t). \quad (4.17)$$

For small t , when $w(t)$ is approximately constant, the average fitness decreases approximately linearly determined by t^* .

4.4.2 Development of Imatinib resistance in cell culture

Using an in vitro approach to induce Imatinib resistance under increasing selection pressure, we detected resistance in 10 out of 10 BAF/3-p210 clones. As depicted in Fig. 4.1A the development of Imatinib resistance was associated with a significant decrease in the population doubling compared to control cells. The observed development of Imatinib resistance resulted in an significant increase of the IC_{50} for Imatinib in all, but one clones (Fig. 4.1 B). For four resistant clones, it was possible to identify previously known resistance mutations in the BCR-ABL kinase domain. These were the D276G (prevents Imatinib from binding to BCR/ABL), E355G (blocks the TK catalytic centre) and L284M (inactivates the kinase) mutations (Fig. 4.1 C). No mutations of the kinase domain were found in the other clones, nevertheless they also developed resistance to Imatinib. In order to explore if the resistance in the non-mutated clones is based on an increased BCR-ABL expression we employed real time PCR for quantification of the BCR-ABL transcript. As depicted in Fig. 4.1 D we could only identify an up-regulation of BCR-ABL expression

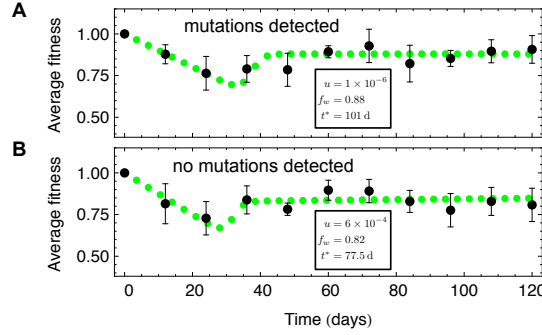


Figure 4.4: **Average fitness observed for samples A) with and B) without detected mutations.** The black dots are due to data and the green dots due to simulations. The evaluated parameters differ (see the inset of the plots), the fitness of the mutated cells is on average higher but their transformation rate is much smaller.

in one single clone. Furthermore, there was no increased expression of the ABC-transporter MDR-1 on resistance cell clones (Fig. 4.1 E). Based on these findings, we concluded that mechanisms other than kinase domain mutations, BCR-ABL over expression or MDR-1 up-regulation have lead to resistance in these cases.

We now focus on the dynamics of the cell population. Initially, the average fitness of the total population decreases as the Imatinib concentration increased. At day 24 of the experiment, a global fitness minimum of $\Phi = 0.706$ (experimental data) is obtained (Fig. 4.3 A). Subsequently, the average population fitness increased, although the environmental conditions become more challenging due to the ongoing increase of the Imatinib concentration. From day 48 until the end of the experiment the average fitness of the population was saturated and fluctuated around $\Phi = 0.83 \pm 0.03$. Our mathematical model offers a simple explanation for this result: In the beginning, only wild type cancer cells with decreasing fitness are present (Fig. 4.3 B). During the experiment, the frequency of resistant cell types increased and once they represent the bulk of the population, the average population fitness becomes independent of a further increase of Imatinib concentration. Consequently, the average fitness becomes constant. While it is challenging to analyze the underlying mechanisms of resistance evolution, we were able to address the parameter ranges in the experiment based on our model.

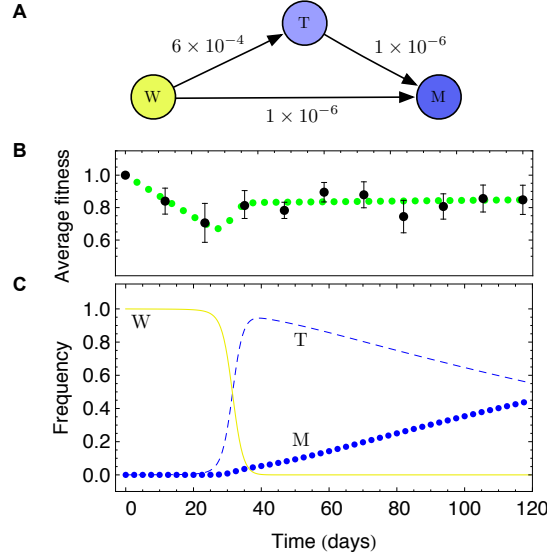


Figure 4.5: **Stochastic simulations with three subpopulations.** **A)** Transitions between the three types considered, wild types (W), resistant types without observed mutations (T) and resistant types with observed mutations (M). The switching probabilities are according to the arrows. **B)** Average fitness of the system described in a) due to simulations (green dots) and experimental data (black dots, as described in figure 4.3). The parameters are those observed in figure 4.4. **C)** Average frequency of the three types. Initially, only wild types (yellow line) are present, at day 37 of the experiment resistant types without observed mutations (T, dashed blue line) reach almost fixation, but in the long run mutated types (M, blue dots) take over due to their fitness advantage. At day 120 of the experiment we have 0.57 T types and 0.43 M types (6 T and 4 M types were observed in the experiment at day 120).

4.4.3 Fitting the mathematical model to the experimental data

First, we fixed time scales to make experimental and simulation results comparable. In the control experiment, untreated wild type cancer cells were grown in a regular growth medium and the population doubling times were recorded. We compare these doubling times to one generation in our simulation, which we define here as N reproduction events in a population of size N . We found on average a doubling rate of 0.48 ± 0.01 per day in the experiment, thus one generation in silico approximately relates to two days in vitro.

The initial fitness decrease was determined by the fitness decrease of wild type cells only and thus allowed us to fix the parameter t^* by linear regression of the

average population fitness on day 0, 12 and 24, leading to $t^* = 82 \pm 4$ days. Thus, assuming a linear decrease of fitness according to equation (4.11) the wild type fitness would reach zero on day 82 of the experiment (dashed black line in Fig. 4.3 A). Such an assumption of zero fitness is probably biologically not meaningful. However, our simulation results remain unchanged if we assume that the lowest possible fitness value of wild type cells is 0.5. This would be reached at day 40 of the experiment, where the resistant cell type already took over the population as shown in Fig. 4.3 B.

In our model, the average fitness of the total population becomes constant at high Imatinib concentrations, because the resistant cell types take over the cell cultures. This fixed the average fitness of the resistant cell types to $f_r = 0.83 \pm 0.02$ (averaged from day 48 on).

The only remaining free parameter is the switching probability u , which is more difficult to assess. To determine u , we fixed t^* and f_r as discussed above and performed individual based stochastic simulations, which suggests that a switching probability between $10^{-6} \leq u \leq 10^{-4}$ is compatible with the experiment. Smaller values of u would cause a later increase in frequency of resistant cell types, but as soon as a certain fraction of resistant cells are present they fixate faster. For higher values of u the minimum of the average fitness Φ would be reached earlier. We also performed a non linear fit of equation (4.17) with the full solution for $w(t)$, equation (4.15). The best fit yields a value of $u = 10^{-4}$. Note that $u = 10^{-4}$ is too high to call this parameter a mutation rate. However, in the experiment only in 4 out of 10 samples mutations known for causing resistance to Imatinib were found.

To take these two different experimental outcomes into account, we performed the procedure described above independently for samples with and without mutations. The result is shown in Fig. 4.4. In samples where mutations were detected, we observe $u = 1 \times 10^{-6}$, $f_r = 0.88$. In the case of no detected mutations, we found $u = 6 \times 10^{-4}$ and $f_r = 0.82$. Thus, the transformation probability differs by magnitudes and samples with mutations seemed finally fitter than samples without mutations. This leads to the possibility that two mechanism may cause resistance, mutation and a sort of phenotype switching. We extended our model to incorporate these two distinct effects, as shown in Fig. 4.5. With this extension, wild type cells switch into mutated cells (M) with probability $u_M = 1 \times 10^{-6}$ and fitness $f_M = 0.88$. Alternatively, they switch into transformed types (T) with probability

$u_T = 6 \times 10^{-4}$ and fitness $f_T = 0.82$. Additionally, transformed types (T) can mutate with probability u_M into mutated types (M). The dynamics for this scenario is shown in Fig. 4.5 C. Initially only wild type cells are present in the simulation. Due to the higher switching probability transformed types take over almost the whole population, but due to their fitness advantage and the directed mutations in the long run the mutant types fixate. However the experiment was stopped after 120 days. In the experiment we found mutant types in 4 out of 10 samples. At this time we observe frequencies of 0.43 for mutated types and 0.57 for transformed types in the simulation, compatible with the experiment. Also in the simulation, the average fitness of the population goes through a minimum, but the fitness still increases after 120 days, because the mutant type does not reach fixation yet.

4.5 Discussion

Resistance to TKI is a frequent clinical problem in treatment of CML patients. The appearance of resistant cell clones undermines our ability to control the malignant BCR-ABL positive cell clone and thereby prevent the successful treatment of CML patients. The mechanisms of TKI resistance are either BCR-ABL dependent or BCR-ABL independent (von Bubnoff et al., 2003). Mutations in the BCR-ABL kinase domain represent the most frequent BCR-ABL dependent mechanisms. On the other hand, overexpression of BCR-ABL and overexpression of drug transporters are known kinase independent mechanisms of TKI resistance. The rise of drug resistance is a evolutionary process characterized by a competition between sensitive and resistant cells but not an active process of resistance induction by the drug (Blagosklonny, 2002).

Using a combination of a clonal in vitro approach for induction of Imatinib resistance and mathematical modeling, we were able to explain the dynamics of the average fitness of Imatinib treated cancer cell populations in vitro in a simple model assuming that:

- (i) The population consists of only three subpopulation, the wild type cancer cells (the only cell type present at the beginning of the experiment) and resistant cell types (mutated and transformed cells).

- (ii) The fitness of wild type cells decreases with increasing drug concentration over time. The fitness of resistant types is lower than the wild type fitness, but both are constant.
- (iii) We introduced directed transformation. Wild types transform into resistant types, the reversal is neglected.

With these assumptions, we provided a method to fix the model parameters and thus get some insights into the mechanism that cause the dynamics. Comparing the analytical solution with experimental results enabled us to fix important parameter of the system. Accordingly, it is possible to derive a range for the transformation probability u , which is $10^{-6} \leq u \leq 10^{-4}$ for this experimental setup. This is high compared to values normally assumed for mutations, which seems to be about 10^{-8} for normal cells. But the mutation rate of malignant cells has been reported to be at least 10^{-1} to 10^{-2} higher (Drake et al., 1998; Hicks et al., 2010). If we perform our analysis separately for the samples with and without mutations, we find different parameter values. For the mutated cell lines, we found $u_M = 1 \times 10^{-6}$, which is in the expected range for mutation rates of such malignant cells. But for samples without mutations, we find $u_T = 6 \times 10^{-4}$. This supports the idea that not only mutations are responsible for drug resistance evolution in this case, but that also other mechanisms are likely to be involved (Lange et al., 2005). Noteworthy, we could not detect BCR-ABL- or MDR-1 overexpression as the reason for resistance. However, there is some evidence that an activation of additional pathways, like the phosphatidylinositol-3-kinase (PI3K)-AKT pathway (Burchert et al., 2005) or Src family kinase (Pene-Dumitrescu and Smithgall, 2010; Wu et al., 2008), can compensate for the BCR-ABL inhibition and can promote the proliferation of these cells in the presence of Imatinib. From our mathematical model and our experimental results, one can conclude that particularly the time frame between day 24 and day 48 is crucial for development of resistance. It can be speculated that starting with day 24, the cells launch out to adapt their metabolism to the increasing pressure of raising Imatinib concentrations. This kind of adaptations are normally accompanied by an activation of novel signaling pathways and can drive the cellular metabolism of resistant cells to an increased glycolytic activity and phospholipid turnover (Klawitter et al., 2009; Kominsky et al., 2009). In this context, our model represents a basis to determine the critical time point and Imatinib concentration for further studies to detect novel BCR-ABL independent resistance mechanisms. This should

be done by a comprehensive and integrative genome-, transcriptome-, proteome- and metabolome analysis and may lead to additional insights into activated pathways of resistant CML cells. Recent *in vitro* and *in silico* studies have revealed that a combination of drugs with different mode of actions can prevent treatment failures due to drug resistance (Katouli and Komarova, 2010; Komarova, 2011; Komarova et al., 2009; Komarova and Wodarz, 2009). However, the appropriate targets are not defined and must be elucidate in further studies.

From our study, it is apparent that the development of Imatinib resistance is associated with reduced fitness of resistant cells *in vitro*. Translating this into fitness *in vivo* can be challenging, due to additional aspects as the hierarchical structure of hematopoiesis (Dingli and Pacheco, 2006; Rodriguez-Brenes et al., 2011; Werner et al., 2011), failed targeting of cancer stem cells, for example due to quiescent cells, (Chomel and Turhan, 2011; Peng et al., 2010) and possible aging effects (Vicente-Dueñas et al., 2010). Furthermore very small differences *in vitro* can cause important measurable differences *in vivo*, but also in this case mathematical approaches can be helpful (Lenaerts et al., 2011). Notably, the fitness of clones harboring a BCR-ABL mutation was higher compared to the unmutated clones. This might be due to the proposed switch on alternative or additional programs in unmutated clones which may be costly in terms of proliferation, but enable the cell to function independent of the BCR-ABL kinase. Based on recent studies, it can be speculated that in mutated clones the mutations are preexisting and the cells do not need to adapt their metabolism to reach resistance (Jiang et al., 2007). In this case the reduced fitness reflect the differences between mutant and wild-type BCR-ABL. Recent studies have revealed that certain mutated BCR-ABL kinases are accompanied by reduced oncogenic activity (Griswold et al., 2006; Skaggs et al., 2006). These findings are supported by clinical studies. Hanfstein et al. have shown a deselection of mutant BCR-ABL positive clones after cessation of tyrosine kinase inhibitor (Hanfstein et al., 2011), as also suggested by mathematical models (Dingli et al., 2008; Lenaerts et al., 2010). Fitness costs are commonly associated with drug resistance. Particularly, resistance against antibiotics is frequently observed *in vitro* and *in vivo* and is associated with reduced fitness of resistant clones (Andersson, 2006; Andersson and Hughes, 2010). In comparison to that, the existence of a fitness cost in resistance to cancer drugs has not been analyzed in detail yet. Therefore, our data represent an appropriate model to further investigate in this cellular behavioral pattern.

A mathematical model of telomere shortening

This chapter contains work in progress about telomere dynamics on the cell and tissue level. We show, that telomere dynamics provide additional insights in the dynamics of stem cell populations. Although this chapter neglects hierarchical tissue organizations, the findings on stem cell dynamics promise interesting applications. The experimental data on the telomere distribution in lymphocyte and granulocyte cell populations of healthy adults that we use in this chapter were derived by Fabian Beier and Tim Brümmendorf, the modeling idea, the mathematical analysis and the performed simulations were done by Benjamin Werner and Arne Traulsen.

Telomeres are noncoding DNA structures expressed at the end of all eukaryotic chromosomes. In vertebrates they consist of hundreds to thousands of repeats of the amino acid blocks TTAGGG (Blackburn et al., 1991; Griffith et al., 1999). These structures are essential to ensure chromosomal integrity, as an accelerated telomere shortening is associated with genetic instability (Hande et al., 1999; d'Adda di Fagagna et al., 2001; Boukamp et al., 2005; Brümmendorf and Balabanov, 2006). However, due to the inability of the conventional DNA polymerase to fully replicate the 3' end of the DNA during replication, the so called "end-replication problem" (Levy et al., 1992) and due to the permanent suspension to oxidative stress, telomere repeats are sequentially lost in most somatic cells. The loss of telomere repeats is partially counteracted by the enzyme telomerase, as this specialized DNA polymerase adds additional amino acid blocks at the 3' end of chromosomes. Telomerase is most active on the stem cell level and the germ line (Chan and Blackburn, 2004; Wright et al., 1996), but increased activity is also found in various cancer types. In most tissues telomerase activity is insufficient to completely prevent the sequential loss of telomere repeats with age (Harley et al., 1990; Vaziri et al., 1994; Rufer et al., 1999), leading to a progressive shortening of telomeres. Shorten telomeres

were linked to the observation of a limited replication capacity of cell lines in vitro, and may be an important trigger of cell senescence in vivo as well.

Although the replication capacity of most cell lines is limited, a continuous cell reproduction is necessary to maintain a tissues functionality. This is often realized in a hierarchical organized tissue structure with multipotent stem cells at the root (Dingli et al., 2007a; Werner et al., 2011; Tumber et al., 2004; Potten et al., 2009). Stem cells differentiate into more committed cell lineages and in addition show self renewal potential. The rate of self-renewal is regulated by the ratio of symmetric and asymmetric cell divisions, either achieved in single stem cells or on the population level (Morrison et al., 1997; Morrison and Kimble, 2006). In the latter case symmetric stem cell divisions compensate the loss of stem cells, while in the former case each stem cell strictly undergoes asymmetric cell divisions. The ratio of symmetric versus asymmetric divisions may change with age or disease. For example stem cells undergo additional symmetric cell divisions after chemotherapy (Brümmendorf et al., 2001; Marciniak-Czochra et al., 2009). In contrast, stem cells potentially are lost because of replication errors, self immune-mediated damage, or external causes such as high radiation or chemotherapy. A progressive shrinking of the stem cell pool is found in patients diagnosed with aplastic anemia. These patients develop bone marrow failure syndromes and significantly shorter telomeres compared to healthy individuals (Ball et al., 1998; Brümmendorf et al., 2001). Recently accelerated telomere shortening was shown to correlate with a variety of conditions, reaching from cancers, infectious diseases to increased stress.

Besides experimental results, theoretical modeling can generate hypothesis on the dynamics of telomere shortening. Some results were achieved in the past, as the expected fraction of proliferating cells in a population of cells with decreasing telomeres (Levy et al., 1992; Arino et al., 1995; Olofsson and Kimmel, 1999) or changing properties of cell proliferation with age. Cells initially divide fast and enter a class of slower dividing cells during aging (Sidorov et al., 2004). Analytical results for the telomere length distribution of purely symmetrically dividing cells were derived and a phase transition between an ultimately senescent cell population and an infinitely proliferating cell population is determined (Antal et al., 2007).

Below, we present a mathematical model of telomere shortening on the cell level. We do not consider the dynamics of single telomeres, but investigate the average shortening of telomeres in single cells. We analyze the case of healthy individuals, including the effects of stem cell pool expansion due to increased symmetric cell

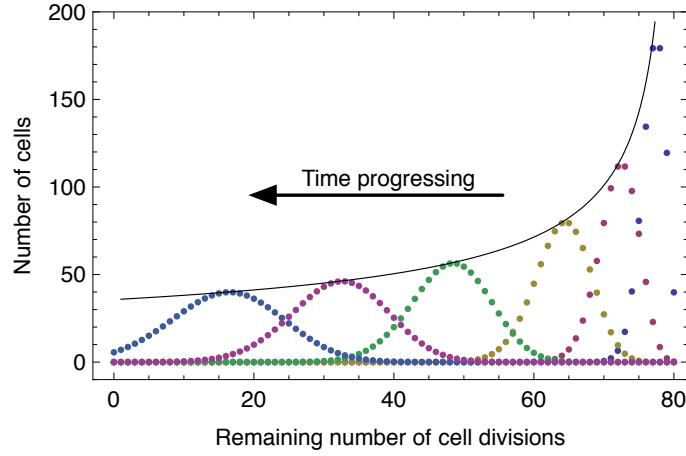


Figure 5.1: **Distribution of the remaining cell cycles of asymmetric dividing cells at six different points in time.** The distribution of the remaining cell cycles is traveling to the left with progressing time. The maximum of the distribution (black line) declines proportional to $\frac{1}{\sqrt{t}}$, see equation (5.4). Initially there are $N_0 = 800$ cells with $c = 80$ remaining cell cycles each.

divisions, which occurs in early childhood and after chemotherapy. We also address the consequences of stem cell pool exhaustion. This allows us to calculate the exact telomere length distribution at any time. We explicitly incorporate the effects of symmetric and asymmetric stem cell division and non linear effects caused by stem cell death due to external forces. We compare the theoretical predictions to clinical data, that measures the telomere length distribution of granulocytes and lymphocytes in healthy adults.

5.1 Mathematical model and results

We assume a population of N_0 cells, each cell initially with an average telomere length of c . Cells either divide symmetrically with probability p , or asymmetrically with probability $1-p$. We assume, that the average telomere length of a cell shortens by a fixed number (Δc) during one cell division. Thus a symmetric cell division leads to two stem cells, where the average telomere length of those cells shortened by Δc . An asymmetric cell division leads to one stem cell and one more committed cell, both cells with by Δc shorten telomeres. We note that the number of cells within

the stem cell pool is conserved by asymmetric cell divisions, whereas the number increases due to symmetric cell divisions. The implementation of asymmetric cell divisions reflects the hierarchical organization of most tissues. Stem cells are at the root of the hierarchy and a certain number of differentiation steps are necessary to produce all mature cell types of a tissue. The initial cell pool represents the stem cells and a committed cell reflects a progenitor cell. Each cell proliferates with a rate r and cells become senescent in state 0. This corresponds to the idea of a Hayflick limit, stating, that the proliferation capacity of any cell is limited (Hayflick et al., 1961). We will implement cell death other than reaching state 0 below and expand the model appropriately.

Below we compare our analytical results to individual based stochastic simulations. In these simulations every cell has an individual representation and the complete dynamics can be recorded. At each time step i one cell proliferates, where a Gillespie algorithm (Gillespie, 1976) is used to decide which cell reproduces next. If a cell reaches state 0 it drops out of the pool of proliferating cells. In the next section we formulate a mathematical description of the telomere dynamics, enabling us to find closed analytical solution of the telomere length distribution observed in exact individual based stochastic computer simulations. In addition we can compare those analytical results to measured telomere length distributions in healthy adults.

5.1.1 Asymmetric cell divisions

We discretize the available states of telomere length to derive analytical expressions for the telomere length distributions. We introduce in total $c + 1$ states. Each state contains cells of equal average telomere length. Initially all cells have telomeres of length c . In this section we focus on asymmetric cell divisions ($p = 0$). An asymmetric division of a cell in state i leads to one more committed cell and one cell functional identical to its mother cell. The committed (progenitor) cell leaves the pool of stem cells and does not further contribute to stem or progenitor cell production. The second cell keeps the stem cell properties and enters state $i - 1$. We call $N^{(i)}(t)$ the number of cells in state i at time t . We note, asymmetric cell divisions strictly conserve the size of the cell pool $\sum_{i=0}^c N^{(i)}(t) = N_0$. We further chose the initial condition $N^{(c)}(0) = N_0$. The next proliferating cell is randomly chosen amongst all cells not yet in state 0. We apply a deterministic, time continuous approximation and capture the dynamics of telomere shortening by

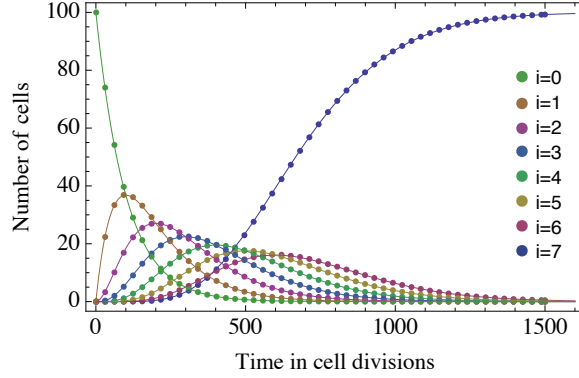


Figure 5.2: **Number of cells in state i given by equation (5.2) (lines) and exact individual based stochastic computer simulations (dots).** At $t = 0$ there are $N_0 = 100$ cells in state $c = 7$. Cells in state i are Gamma distributed, except for cells in state 0. Those cells accumulate over time to saturate at N_0 . The maximum number of cells in a state i is reached after $t_{\max}^{(i)} = (c - i) N_0$ cell divisions. The experimentally measured telomere length distribution (5.3) (see Fig. 5.1 for an example) can be obtained by drawing a vertical line at the associated time point.

a system of coupled differential equations,

$$\dot{N}^{(i)}(t) = \begin{cases} -\frac{rN^{(i)}}{N_0} & i = c \\ -\frac{rN^{(i)}}{N_0} + \frac{rN^{(i+1)}}{N_0} & 0 < i < c \\ \frac{rN^{(1)}}{N_0} & i = 0. \end{cases} \quad (5.1)$$

Here N_0 is the initial number of stem cells in state c , and r represents the proliferation rate of a cell. Cells progressively travel into lower states and accumulate in state 0 in the long run. In the following we shift the index to $j = c - i$ for convenience. The general solution of (5.1) can be derived recursively

$$N^{(c-j)}(t) = \begin{cases} \frac{N_0}{j!} \left(\frac{rt}{N_0}\right)^j e^{-\frac{rt}{N_0}} & 0 \leq j < c \\ N_0 \left(1 - e^{-\frac{rt}{N_0}} \sum_{l=0}^{c-1} \frac{1}{l!} \left(\frac{rt}{N_0}\right)^l\right) & j = c. \end{cases} \quad (5.2)$$

The number of cells in state j correspond to a Gamma distribution with rate parameter $\frac{r}{N_0}$ and shape parameter j , see Fig. 5.2 for a comparison of solution (5.2) to exact individual based stochastic computer simulations. The number of cells in state c decreases exponentially. Cells in downstream states are initially absent, undergo a maximum and vanish in the long run again. Only cells in state 0 accumulate over time. However, experimentally the distribution (5.2) can not be determined by one

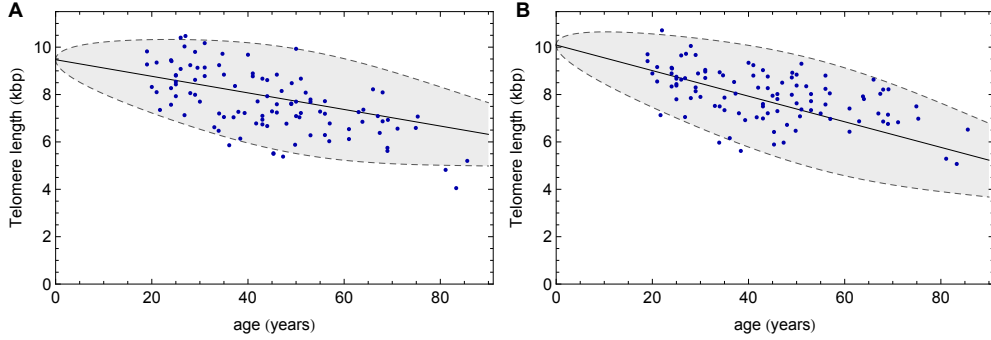


Figure 5.3: **Telomere length of 103 healthy individuals (blue dots) in the age of 18 to 85 for A Granulocytes and B Lymphocytes.** The lines correspond to fits of the analytical predicted linear approximation $E[c](t) \approx c - \frac{r}{N_0}t$. The gray area represents the theoretical predicted standard deviation (5.8) of the expected telomere distribution. We find for **A**) $c = 9.47 \pm 0.26$ kbp and $\frac{r}{N_0} = 0.035 \pm 0.006 \frac{\text{kbp}}{\text{year}}$ and for **B**) $c = 10.09 \pm 0.3$ kbp and $\frac{r}{N_0} = 0.054 \pm 0.006 \frac{\text{kbp}}{\text{year}}$. Using estimates for the telomere shortening of 50 to 100 bp per cell division and assume 0.5 to 2 stem cell divisions a day, one can estimate the number of actively dividing stem cells. For Granulocytes we find an approximate range of 250 to 2100 and for Lymphocytes a range of 150 – 1400.

sample at a single time point t' . A measurement of the telomere length distribution at time t' corresponds to the interception points of a vertical line, drawn at time t' , and the number of cells in every state given by equation (5.2). Thus the experimentally observed distribution at time t' is given by

$$f_{t'}(c) = \{N^{(c)}(t'), \dots, N^{(0)}(t')\}. \quad (5.3)$$

This distribution gives raise to a traveling wave shifting towards shorter average telomere length in time, see Fig. 5.1. The maximum of this wave reaches state i after $t_{\max}^{(i)} = \frac{(c-i)N_0}{r}$ cell divisions. Plugging this into equation (5.2), we find for the maximum of this wave

$$N^{(i)}\left(t_{\max}^{(i)}\right) = \frac{N_0}{(c-i)!} \left(\frac{c-i}{e}\right)^{c-i} \approx \frac{N_0}{\sqrt{2\pi(c-i)}}, \quad (5.4)$$

where in the last step the Stirling formula was used. The maximum of the telomere length distribution declines proportional to $\frac{1}{\sqrt{t}}$ in time. The telomere length distribution (5.2) allows us to calculate the expected average telomere length $E[c](t)$ at

any time t . This corresponds to the first moment of the distribution, given by

$$\begin{aligned} E[c](t) &= \frac{1}{N_0} \sum_{i=0}^{c-1} (c-i) N^{(i)}(t) = \sum_{i=0}^{c-1} \frac{c-i}{i!} \left(\frac{rt}{N_0}\right)^i e^{-\frac{rt}{N_0}} \\ &= \frac{1}{c!} \left(\frac{rt}{N_0}\right)^{c+1} e^{-\frac{rt}{N_0}} + \frac{cN_0 - rt}{N_0} \frac{\Gamma\left[1+c, \frac{rt}{N_0}\right]}{c!}. \end{aligned} \quad (5.5)$$

where $\Gamma\left[1+c, \frac{rt}{N_0}\right] = \int_{\frac{rt}{N_0}}^{\infty} dx x^c e^{-x}$ is the incomplete Gamma function. For a biological meaningful parameter range, we have $c \approx 10$ kbp and N_0 is in the order of a few hundred to thousands of cells. Thus the first term in Eq. (5.5) is negligible. The second term is dominated by the linear decaying term, as the incomplete gamma function is $\Gamma\left[1+c, \frac{rt}{N_0}\right] \approx c!$ for not too large t . In Fig. 5.4 we compare Eq. (5.5) to exact stochastic computer simulations, revealing, that the linear approximation is excellent, until most cells reach states of very short telomeres. Thus the complicated expression (5.5) is well approximated by

$$E[c](t) \approx \frac{cN_0 - rt}{N_0}. \quad (5.6)$$

Expression (5.6) allows the direct connection to experimental results. In Fig. 5.3 we fit this linear approximation to clinical data of telomere shortening in granulocytes and lymphocytes, derived from 103 healthy individuals in the age of 18 to 85. This linear fit yields for the model parameters: $c = 9.47 \pm 0.26$ kbp and $\frac{r}{N_0} = 0.035 \pm 0.006 \frac{\text{kbp}}{\text{year}}$ for granulocytes and $c = 10.09 \pm 0.3$ kbp and $\frac{r}{N_0} = 0.054 \pm 0.006 \frac{\text{kbp}}{\text{year}}$ for lymphocytes. Assuming a telomere shortening of 50 to 100 bp per cell division, we can estimate the ratio $\frac{r}{N_0}$. Assuming further 0.5 to 2 stem cell divisions per day, the granulocyte data predicts 250 to 2100 and the lymphocyte data predicts 150 to 1400 actively dividing stem cells. However, such linear fits were performed before, and the estimations derived here confirm former results (Rufer et al., 1999). But those studies lack an underlying kinetic model. Our approach allows us to calculate additional properties of the system. We can derive analytical expressions for the time dependence of the variance $\sigma^2(t)$ for example. Note, that the momentum generating function of the telomere length distribution $M_c(x) = E[e^{cx}](t)$ is given by

$$M_c(x) = 1 + \frac{e^{(e^{-x}-1)\frac{rt}{N_0}} \Gamma\left[c+1, \frac{e^{-x}rt}{N_0}\right]}{c!} - \frac{\Gamma\left[c+1, \frac{rt}{N_0}\right]}{c!}. \quad (5.7)$$

We can recover our result for the average (5.5) of the telomere length distribution via $E[c](t) = \frac{d}{dx}(M_c(0))$. In addition, the variance can be calculated via

$$\begin{aligned}\sigma^2(t) &= E[c^2] - E^2[c] = \frac{d^2}{dx^2}M_c(0) - \left(\frac{d}{dx}M_c(0)\right)^2 \\ &= \left(\frac{rt}{n}\right)^{c+1} \frac{nc - rt}{n c!} e^{-\frac{rt}{n}} + \left[\left(c - \frac{rt}{n}\right)^2 + \frac{rt}{n}\right] \frac{\Gamma\left[c + 1, \frac{rt}{n}\right]}{c!} - E^2[c].\end{aligned}\quad (5.8)$$

Again, the first term of equation (5.8) is negligible for a biological meaningful parameter range. The quadratic term $(c - rt/n)^2$ is compensated by an identical term in $E^2[c]$ (see Equ. (5.5)). Again, the gamma function is approximately equal to $c!$ for not too short telomeres. Thus, expression (5.8) is initially dominated by the linear term and consequently, the variance grows linear due to $\sigma^2 = \frac{rt}{n}$. The standard variance becomes

$$\sigma = \sqrt{\frac{rt}{n}}. \quad (5.9)$$

We compare the analytical results of Eq. (5.8) to results of exact individual based stochastic computer simulations in Fig. 5.5. The linear approximation of the variance is excellent. Only if cells start to accumulate in state 0 the variance decreases again.

In the following we fit our theoretical predicted telomere length distribution (5.3) to measured telomere length distributions of healthy individuals, see figure 5.7. We assume symmetric cell divisions with probability p and asymmetric cell divisions with probability $1 - p$, respectively. Further we denote the initial telomere length by c , the initial number of stem cells by N_0 and r is the proliferation rate of cells. We first discuss the special case of asymmetric cell divisions, $p = 0$. In this situation the number of actively dividing cells is strictly conserved. The expected number of cells with a telomere length of x at time t is given by (5.1). This yields for the expected telomere length distribution $\rho(t, x)$ at time t ,

$$\rho(t, x) = \frac{1}{(c - x)!} \left(\frac{rt}{N_0}\right)^{c-x} e^{-\frac{rt}{N_0}}. \quad (5.10)$$

We substitute $q = \frac{rt}{N_0}$ within the expected telomere length distribution and get

$$\rho(t, x) = \frac{q^{c-x}}{(c - x)!} e^{-q}. \quad (5.11)$$

We can fit Eq. (5.11) to measured telomere distributions in healthy individuals, leaving two free parameters q and c to be determined. Results of the nonlinear fits

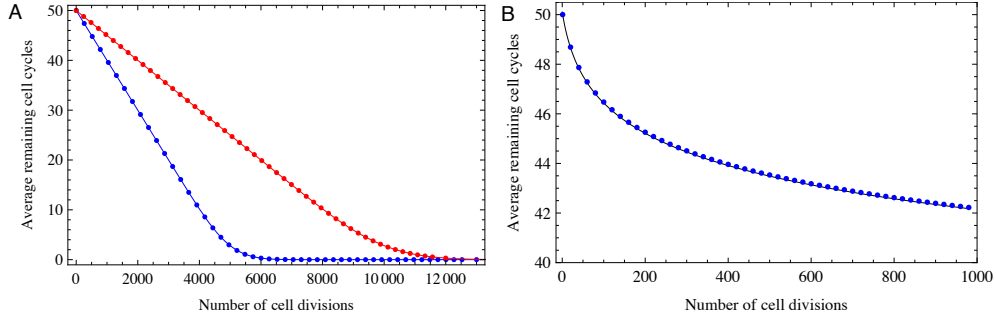


Figure 5.4: **A** Mean remaining cell cycles derived from equation (5.4) (lines) and simulations (dots) for $c = 50$, $r = 1$ and $N_0 = 200$ (red line) and for $c = 50$, $r = 1$ and $N_0 = 100$ (blue line) for strict asymmetric cell divisions. The average of the remaining cell cycles decreases linear before turning into a power law if it reaches 0. **B** Mean remaining cell cycles derived from equation (5.21) (line) and simulation results (dots) for symmetrically dividing stem cells with parameters $n_0 = 20$, $c = 50$, $r = 1$ and $p = 1$.

can be seen in Fig 5.7. The initial telomere length c varies between individuals but is within the expected range of 8 – 13 kbp. However, the parameter q shows an interesting property. The parameter is given by $q = \frac{rt}{N_0}$. We assumed $p = 0$ (asymmetric cell divisions only), thus the number of stem cells N_0 is constant. The parameter q is time dependent, and one would expect an increase of q with increasing age. However, the non linear fits suggest, that the parameter q is independent of age, see Fig. 5.7. This contradiction can be solved heuristically by assuming $p > 0$. Substituting $N_0 = n_0 + prt$ gives

$$q = \frac{rt}{N} = \frac{rt}{n_0 + prt} = \frac{1}{p + \frac{n}{rt}}. \quad (5.12)$$

We further observe $\lim_{t \rightarrow \infty} q = 1/p$. We expect q to converge towards a constant value for increasing t . However this is a heuristic argument. The calculated telomere length distribution (5.11) is only valid for $p = 0$. Incorporating $p > 0$ leads to a modified distribution.

5.1.2 Symmetric cell divisions

In the former subsection we derived the telomere length distribution under the restriction of asymmetric cell divisions ($p = 0$) and discussed its properties. However, a comparison of those analytical results to experimental data highlights the need to incorporate symmetric cell divisions into our mathematical framework. Symmetric

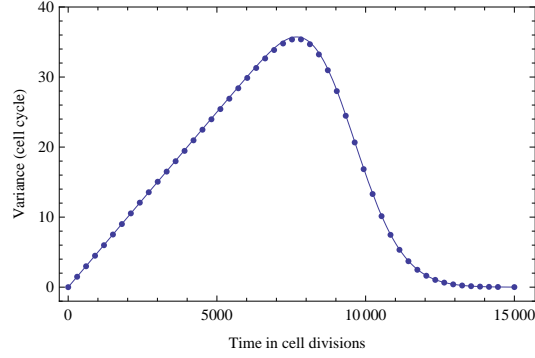


Figure 5.5: **Variance of the telomere length distribution for asymmetric cell divisions.** Shown are analytical (line, Equ. (5.8)) and simulation (dots) results for the variance of the telomere length distribution. The parameters are $N_0 = 200$, $r = 1$ and $c = 50$. Initially all cells are in state $c = 50$, thus the variance vanishes at $t = 0$. The variance increases linear with ongoing cell proliferations, but decreases again as cells accumulate in state 0.

cell divisions cause a stem cell pool expansion. To incorporate this into our model, we assume a cell division to be symmetric with probability p and asymmetric with probability $1 - p$ respectively. Initially there are n_0 cells with telomeres of length c . We assume the number of cell divisions within a fixed time interval to be constant and thus the cell pool increases on average linear, $N(t) = n_0 + rpt$. Including this, the system of differential equations changes to

$$\dot{N}_p^{(i)}(t) = \begin{cases} -\frac{rN_p^{(i)}}{rpt+n_0} & i = c \\ -\frac{rN_p^{(i)}}{rpt+n_0} + \frac{r(1+p)N_p^{(i+1)}}{rpt+n_0} & 0 < i < c \\ \frac{rN_p^{(1)}}{rpt+n_0} & i = 0. \end{cases} \quad (5.13)$$

The solution to (5.13) is

$$N_p^{(c-i)}(t) = \begin{cases} \frac{n_0}{i!} \left(\frac{1+p}{p}\right)^i \frac{\ln^i(t^*)}{t^*} & 0 \leq i < c \\ n_0 (1+p)^i \left(1 - \frac{\Gamma\left[i+1, \frac{1}{p}\ln(t^*)\right]}{i!}\right) & i = c, \end{cases} \quad (5.14)$$

where we used $t^* = \frac{rp}{n_0}t + 1$ as an abbreviation. Using l'Hospital and $e^x = \lim_{n \rightarrow \infty} \left(1 + \frac{x}{n}\right)^n$ we recover the result (5.2) for $p \rightarrow 0$ and the solution turns into a Gamma distribution again

$$\lim_{p \rightarrow 0} N_p^{(c-i)}(t) = \frac{n_0}{i!} \left(\frac{rt}{n_0}\right)^i e^{-\frac{rt}{n_0}} = N^{(c-i)}(t). \quad (5.15)$$

In the extreme case of $p = 1$ and $n_0 = 1$ Eq. (5.14) becomes

$$N_{p=1}^{(c-i)}(t) = \frac{2^i \ln^i(rt+1)}{i! (rt+1)}. \quad (5.16)$$

If symmetric cell divisions are possible $p > 0$, the number of stem cells increases. We can calculate the maximum possible number of additional produced stem cells N_a due to symmetric cell divisions

$$N_a = N_p^{(0)}(t \rightarrow \infty) - n_0 = n_0((1+p)^c - 1). \quad (5.17)$$

Thus for $p = 0$, corresponding to strict asymmetric cell divisions, the number of stem cells stays constant. We assumed a constant number of cell divisions within a fixed time interval. Due to the increasing stem cell pool size, this effectively causes a reduction in the proliferation rate of single cells with age.

Similar to the former subsection, the time dependence of the maximum of the distribution can be calculated. The time until the maximum of the telomere length distribution reaches length i , ($0 < i < c$) is given by

$$t_{\max,p}^{(i)} = n_0 \frac{e^{p(c-i)} - 1}{rp}. \quad (5.18)$$

The time to reach the maximum increases exponentially for symmetric cell divisions, in contrast to the linear increase for only asymmetric cell divisions. However, (5.18) reduces to the result we obtained in the former subsection in the limit $p \rightarrow 0$. The cell count at the maximum becomes

$$N^{(i)}(t_{\max,p}^{(i)}) \approx \frac{n_0 (1+p)^{c-i}}{\sqrt{2\pi(c-i)}}. \quad (5.19)$$

The maximum decreases considerably slower compared to the case of only asymmetric cell divisions (5.4). Similar to the former subsection we can calculate the average of the telomere length distribution. This time the average becomes

$$\begin{aligned} E_p[c](t) &= \frac{1}{N_p(t)} \sum_{i=0}^{c-1} (c-i) N_p^{(i)}(t) \\ &= \frac{\rho^{c+1} \ln^{c+1}(t^*)}{c! (t^*)^\rho} + \frac{\Gamma[1+c, g(t^*)]}{c!} (c - \rho \ln(t^*)) \end{aligned} \quad (5.20)$$

with $t^* = \frac{rp}{n_0}t + 1$, $\rho = \frac{p+1}{p}$ and $g(t^*) = \rho \ln(t^*)$. Similar to (5.5), this expression is dominated by the second term of the equation. The average decreases approximately logarithmically,

$$E_p[c](t) \approx c - \frac{p+1}{p} \ln\left(\frac{rp}{n_0}t + 1\right). \quad (5.21)$$

In Fig. 5.4 B we compare (5.21) to exact individual based stochastic simulations. The logarithmical approximation is in excellent agreement with the simulation results. The decrease of the average telomere length increases with decreasing p . In the limit $p \rightarrow 0$ we observe the former result (5.6) of a linear decreasing average. Similar to the former section we can derive the variance of the distribution, using the momentum generating function $M_p(x) = E_p[e^{cx}](t)$, via

$$\sigma_p^2(t) = \frac{d^2}{dx^2} M_p(0) - \left(\frac{d}{dx} M_p(0) \right)^2. \quad (5.22)$$

However, the result becomes lengthy and less accessible. Thus we restrict ourselves to a numerical solution of (5.22). The logarithmic decay of the average telomere length has consequences on the interpretation of experimental results of telomere length distributions. In infants an accelerated decrease of telomere length can be observed (Rufer et al., 1999). This accelerated decay is usually explained by increased cell turnover rates due to developmental growth. Parameters are estimated by two linear fits. Our model suggests a different dynamical pattern. The stem cell pool contains only few stem cells initially. These stem cells divide symmetrically with probability p and asymmetrically with probability $1 - p$ respectively. The symmetric cell divisions cause an increase of the stem cell pool size and an indirect decrease in cell proliferation rates. The logarithmic decay is pronounced initially, but becomes flatter after some time. Thus in adults the logarithmic decay is difficult to distinguish from a linear decay, see for example figure 5.4.

Again, we can compare the analytical solution (5.14) to experimental data of telomere length distributions in 11 healthy adult persons. The calculated distribution, if symmetric cell divisions are allowed, is

$$\rho_p(t, x) = \frac{1}{(c-x)!} \left(\frac{1+p}{p} \right)^{c-x} \frac{[\ln(q+1)]^{c-x}}{\sqrt[p]{q+1}}, \quad (5.23)$$

with $q = \frac{rpt}{n_0}$. Again, c gives the initial telomere length, p the probability of symmetric cell divisions, r the division rate and n_0 the initial number of active stem cells, leading to three free choosable parameters p , c and q in the non linear fit, see figure 5.8. The initial telomere length increases and is in the range of 9 – 15 kbp. The probability of symmetric cell divisions is in the range of 0.001 – 0.01 and rather low. However, assuming 400 stem cell divisions a year leads to approximately 200 symmetric cell divisions within a time span of 50 years.

5.1.3 T-cell mediated stem cell death

In the former section we investigated the interplay of symmetric and asymmetric cell divisions. If the dynamics is restricted to those, the total number of cells either stays constant or increases but never decreases (if cells in state 0 are counted). However, different effects may cause a shrinking of the cell pool. Cells may die with a certain probability q during cell division, or external effects cause an increase of cell mortality. One of these causes could be T-cell mediated damage to the stem cell pool, as thought to appear in certain diseases as for example in Aplastic anemia.

First we propose a minimalistic model to include the effect of T-cell mediated damage to the stem cell pool. Assume there are N stem cells in a pool of M non stem cells. T-cells selectively kill stem cells and thus the stem cell pool will shrink over time. If we assume a time continuous approximation this becomes

$$\dot{N}(t) = -\rho N = -\frac{N^2}{N+M}, \quad (5.24)$$

where $\rho = \frac{N}{N+M}$ is the probability to randomly kill a stem cell in the pool of non stem cells. It is natural to assume $M \gg N$ and thus the N in the denominator of equation (5.24) can be neglected. Thus, if we assume $N(t_0) = N_0$, the solution reads

$$N_q(t) = \frac{MN_0}{N_0(t-t_0) + M} = \frac{N_0}{qt+1}, \quad (5.25)$$

where in the last step $t_0 = 0$ was assumed and $q = \frac{N_0}{M}$ was chosen. This is a minimalistic description of the time dependence of the shrinking stem cell pool due T-cell mediated damage. As in the former subsection, we can use this time dependence and include it in our system of differential equations. This time we have to include the shrinking stem cell pool in our calculations, where we need to generalize the differential equations to

$$\dot{N}^{(c-i)} = \left(-1 + \frac{d}{dt} N(t) \right) \frac{rN^{(c-i)}}{N(t)} + \frac{rN^{(c-i+1)}}{N(t)}. \quad (5.26)$$

Incorporating this, the system of differential equations becomes

$$\dot{N}_q^{(i)}(t) = \begin{cases} -\left(1 + \frac{qN_0}{(qt+1)^2}\right) \frac{r(qt+1)N_q^{(i)}}{N_0} & i = c \\ -\left(1 + \frac{qN_0}{(qt+1)^2}\right) \frac{r(qt+1)N_q^{(i)}}{N_0} + \frac{r(qt+1)N_q^{(i+1)}}{N_0} & 0 < i < c \\ \sum_{l=1}^c \frac{rqN_q^{(l)}}{qt+1} + \frac{r(qt+1)N_q^{(1)}}{N_0} & i = 0. \end{cases} \quad (5.27)$$

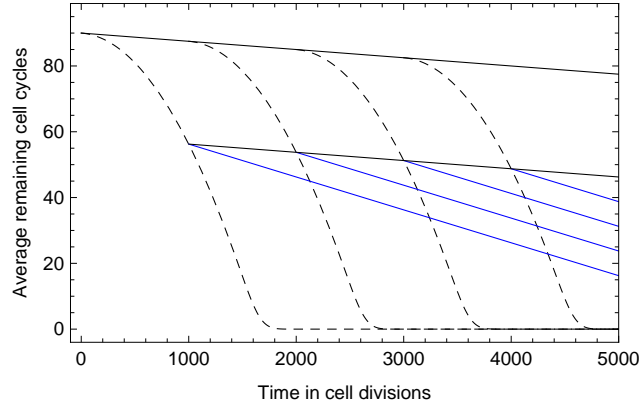


Figure 5.6: **Model of the telomere length reduction of patients with aplastic anemia.** The upper black line corresponds to the telomere decrease of healthy individuals. Aplastic anemia is caused by an immune-mediated damage to the stem cell pool, leading to an increased shortening of telomere repeats. The disease dynamics (dashed lines, disease initiation at 4 different time points) follows initially a quadratic decline due to equation (5.29). However after treatment the stem cell pool stabilizes and the telomeres decline linear again (blue lines), but possibly with a larger slope, as the stem cells were reduced in number.

The solution is

$$N_q^{(c-i)} = \frac{N_0}{i!} \left(\frac{rt}{N_0} \right)^i \frac{\left(\frac{q}{2}t + 1 \right)^i}{qt + 1} e^{-\frac{rt}{N_0} \left(\frac{q}{2}t + 1 \right)}, \quad (5.28)$$

if $j > c$. The solution for $j = 0$ is more difficult to obtain analytically and less intruding. We get

$$E[c](t) = \frac{1}{c!} \left(\frac{t^{**}}{n} \right)^{c+1} e^{-\frac{t^{**}}{n}} + (c - (t^{**})) \frac{\Gamma[1 + c, \frac{t^{**}}{n}]}{c!} \quad (5.29)$$

for the average, with $t^{**} = t \left(\frac{q}{2}t + 1 \right)$. Note the equality of (5.29) to (5.5), when one substitutes $t \rightarrow t \left(\frac{q}{2}t + 1 \right)$, the average telomere length decreases quadratic in time, see figure 5.6. Consequently, the average telomere length decreases faster than in the healthy adult case. Thus, lower average telomere length at disease diagnosis compared to healthy individuals are expected. However, in clinical studies a linear and not a quadratic decline in average telomere length is observed. This linear decrease is likely caused by the age distribution of the patients and corresponds to the linear decrease of telomere length with age, see Fig. 5.6 as example. The telomeres shrink linear for healthy individuals. As the disease initiates, the telomere

decrease accelerates and follows a quadratic decline. As the disease is diagnosed the telomeres are shorten considerable already, but treatment might stop further damage to the stem cell pool and telomeres shorten linear in time again. To measure the quadratic decline one would need to follow the telomere length of individual patients during disease progression without treatment. However, such data is not available as patients get treated immediately after they are diagnosed.

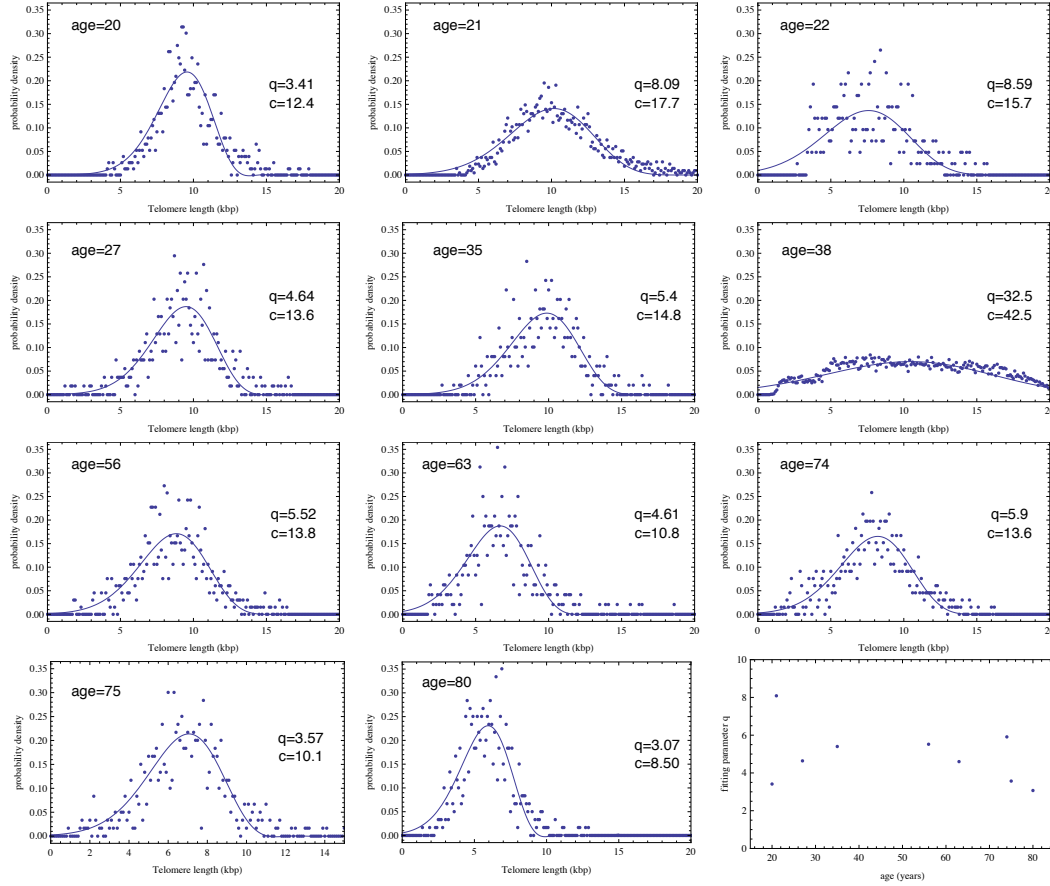


Figure 5.7: **Telomere length distribution of granulocytes in 11 healthy persons of different age.** Each panel represents a single person. Each dot in a panel represents the measured relative abundance of granulocyte cells with a certain average telomere length of this person. The line corresponds to the best fit of the calculated distribution (5.11), corresponding to asymmetric cell divisions. The initial telomere length c is in a expected range of 8 – 15 kbp. However, the second fitting parameter is time dependent $q = rt/N$ and thus should increase with increasing age. The lower right panel shows the parameters q for different persons, but no real pattern is revealed. This might be explained by symmetric cell divisions, see equation (5.12).

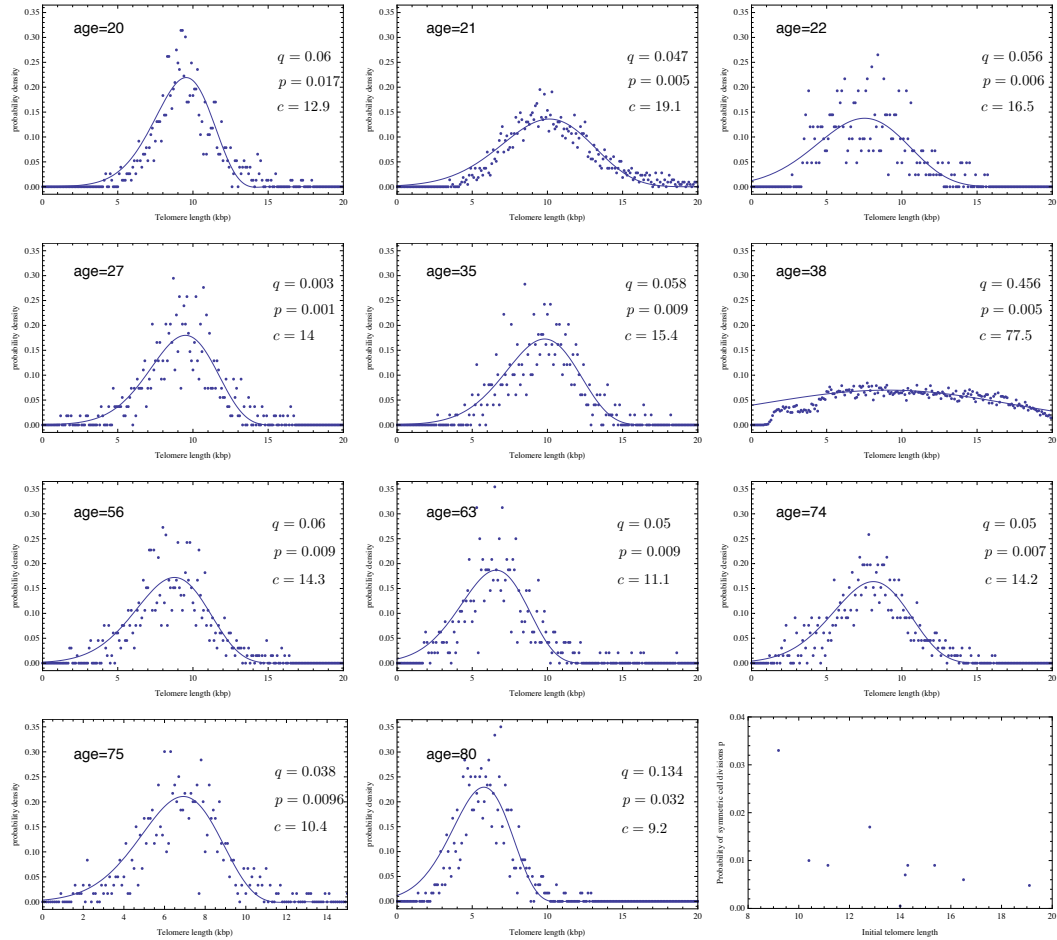


Figure 5.8: **Experimental telomere length distribution** as in figure 5.7. However, the line is the best nonlinear fit of the calculated distribution (5.11). This distribution accounts for symmetric cell divisions, giving the additional parameter p . The initial telomere length c increases to 9 – 17 kbp. The probability of symmetric cell divisions is in the range of 0.001 – 0.01 and thus rather low. However, if we assume 400 stem cell divisions a year, the number of actively dividing stem cells would on average increase by 200 within 50 years.

Impact of random mutations on population fitness

The theory of evolution formulated in means of natural selection by Darwin and Wallace is the unifying concept of biology ([Darwin, 1859](#)). Competition between different agents selects for an inheritable set of currently advantageous properties. Thus evolution is a dynamical process and its rate can range from a few days, for example in resistance evolution, to million of years in speciation processes ([Maynard Smith, 1995](#); [Ridley, 1996](#)). However, natural selection reduces its required diversity, as a fire burns its source. Darwin was aware of this objection, but unable to disprove it. Indeed, only the discovery of genetics solves this problem naturally ([May, 2004](#)). Random changes (mutations) in the DNA of individuals provide a source of diversity in a population, which natural selection can continuously act on. However, this immediately confronts us with new questions, what is the exact interplay of selection and mutation, which dynamical patterns can we expect and how can this be captured in a mathematical framework? In previous work [Huang et al.](#) proposed a mathematical framework, which allows the incorporation of arbitrary many randomly distributed mutations into the framework of evolutionary game theory ([Huang and Traulsen, 2010](#); [Huang et al., 2012](#)). Interestingly, such modeling accounts for an intermediate standing diversity in certain parameter regimes ([Huang et al., 2012](#)), but fitness dynamics can appear counter intuitive. For example dominant mutations can take over a population, but reduce the average fitness of the population ([Huang et al., 2012](#)). For a further quantitative investigation of latter effect, we apply the mathematical framework to infinite large populations described by replicator equations. In the following, we shortly introduce the concept of random mutant games and link this framework to possible applications for the modeling of cancer initiation and cancer progression.

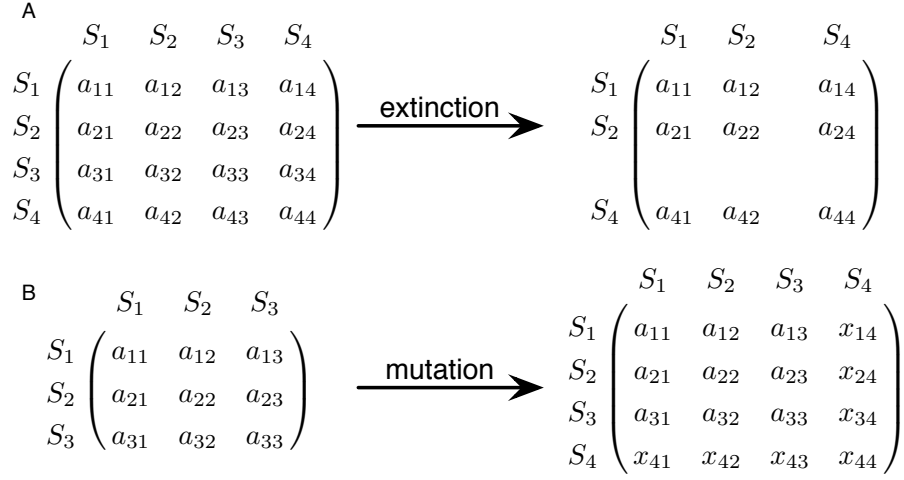


Figure 6.1: **Examples of a payoff matrix of variable size in a random mutant game.**

A Assume, there are initially four possible strategies, but strategy S_3 becomes extinct. Thus all corresponding payoff entries a_{x3} and a_{3x} are deleted and the payoff matrix reduces to a size of 3×3 . **B** In contrast, if a new mutation occurs one row and column are added to the payoff matrix and in general $2n + 1$ payoff entries are drawn from a chosen probability distribution.

6.1 Random mutant games

In the section 1.3.2, where we describe properties of the Moran process, we also introduced basics of game theoretical modeling. Key is the payoff matrix, that defines interactions between agents carrying strategies. Assume we have a set of n possible strategies, denoted by $\{S_1, \dots, S_n\}$. We need n^2 numbers to define all possible interactions and the payoff matrix becomes

$$\begin{array}{c}
 S_1 \quad S_2 \quad \dots \quad S_n \\
 \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix}
 \end{array} \tag{6.1}$$

Usually, the number of strategies n and all possible interactions are predefined. However, here we assume a process with arbitrary numbers of possible strategies. Therefore we implement a payoff matrix of variable size. If one strategy becomes extinct, the corresponding row and column of the payoff matrix are deleted. If a new mutation occurs one row and column are added to the payoff matrix and $2n + 1$

new payoff entries need to be specified, see figure 6.1 for an example. We assume these payoff entries to be random variables independently drawn from a probability distribution. The mean of the distribution corresponds to the payoff entries of the mutants descent. In addition to the payoff matrix, we need to define the dynamical rules of the process. We use standard stochastic approaches as the Moran or the Wright-Fischer process to model finite populations, and replicator equations for populations of infinite size. The probability of fixation of a random mutant given its payoff distribution or the resulting standing diversity for such random games in finite populations can be investigated. However, in the following publication we investigate the impact of such a random mutation on the average fitness of the system. It turns out that random frequency dependent mutations can decrease average fitness. Especially for low beneficial mutation rates (considered a relevant biological scenario) the probability of decreasing average fitness after the fixation of a mutant exceeds the probability to increase it. Thus, (assuming a non arbitrary shiftable probability distribution) we neither expect an infinite increase nor decrease of evolution driven populations fitness, but a balanced state of natural selection and beneficial and deleterious mutations.

6.2 Introduction

Mutations provide a continuous source of variation in natural populations, on which natural selection can act. When fitness is assumed to be constant, only those mutations with higher fitness values will be fixed in a haploid population under strong selection and negligible random drift. Thus, the average fitness of the population would monotonically increase in evolutionary time. There have been numerous hypotheses why this is not what is observed in nature: for instance, environmental changes require new adaptations (Remold and Lenski, 2001; Lalic et al., 2011) or co-evolution can imply continuous adaptation without increasing the average fitness (Brockhurst et al., 2003; Thompson and Cunningham, 2002). However, these are not aspects that we intend to include here. Instead, we focus on a haploid population in a constant environment, and explore frequency dependent fitness, which can be described by evolutionary game theory (Maynard Smith and Price, 1973; Maynard Smith, 1982; Hofbauer and Sigmund, 1998; Cressman, 2003; Nowak and Sigmund, 2004; Nowak, 2006a). In this framework, the fitness of a type depends on the frequencies of other types of individuals in the population. We address the very general question of how the average fitness changes when it is driven by random mutations under frequency dependent selection.

The fitness effects of new mutations have gained significant attention both in experimental research and theoretical work (Sanjuan et al., 2004; Orr, 2003). In experiments, the distribution of fitness effects depends on several aspects of the experimental setup, e.g. how well adapted the organism is to the environment and whether only single mutants or also double mutants (mutants differing from the wild type by two mutations) are considered. Different shaped distributions were proposed to capture the fitness distributions of random mutants under constant selection (Gillespie, 1983; Zeyl and DeVisser, 2001; Cowperthwaite et al., 2005; Orr, 2006). The concrete shape of fitness distributions of spontaneous mutations varies between species and even within the same species on different parts of DNA (Eyre-Walker, 2007). Although no common conclusion on this has been obtained yet – and a universal fitness distribution may as well not exist – it is often possible to estimate some general properties, such as the proportion of advantageous mutations and the mean value of the fitness of the mutations (Lynch et al., 1999; Eyre-Walker et al., 2006).

The concept of random distributed and frequency dependent fitness of mutations can be addressed by evolutionary game theory (Huang and Traulsen, 2010), which considers evolutionary processes under frequency dependent selection (Levin et al., 1997). In this framework, a population of interacting individuals is considered. In the simplest case of linear frequency dependence, the interactions of different types of individuals are captured by a payoff matrix for a game. Those types which are more successful in the game will have a higher reproduction rate. We introduce a payoff matrix with variable size to capture mutations and extinctions. The new payoff entries introduced by mutations are independently drawn from a probability distribution, which corresponds to the concept of randomly distributed fitness. By tracking the dynamics of the payoff matrix and the compositions of the population, we are able to investigate several aspects of an evolving system, such as the average fitness changes of the population, the impact of the fitness distribution on these changes and the expected level of diversity.

6.3 Mathematical model and results

6.3.1 Games with two types

Let us start with a population of a resident wild type (R) and a mutant type (M). Suppose the fitness of a wild type in a homogenous population is d . For constant selection, the fitness distribution is simply a one dimensional distribution around d . For frequency dependent selection, the fitness of a mutant must be defined based on more than a single number. We can write it as an evolutionary game based on a 2×2 payoff matrix with three new payoff entries, a , b and c

$$\begin{array}{cc} & \begin{array}{cc} M & R \end{array} \\ \begin{array}{c} M \\ R \end{array} & \left(\begin{array}{cc} a & b \\ c & d \end{array} \right). \end{array}$$

When a mutant and a wild type interact, the mutant obtains fitness a , and the wild type obtains c . When a mutant meets another mutant, it obtains b . Following the concept of randomly distributed fitness of mutations, the entries a , b and c are defined as random variables. We assume that a , b and c follow the same probability distribution given by a probability density function $f(x)$. While this is the simplest possibility, it may be more realistic to assume correlations between the payoff entries

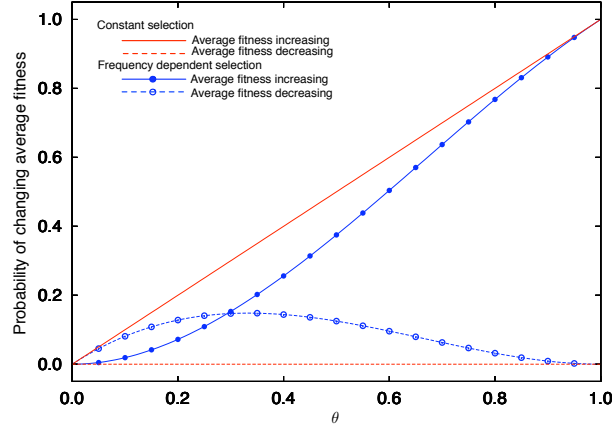


Figure 6.2: **Probability of increasing or decreasing the average fitness in the new equilibrium after one mutation event in an initial homogenous population.** θ is the probability that a random payoff entry of the mutant, a , b or c is larger than wild type initial fitness d . Blue symbols and lines are simulation and analytical results under frequency dependent selection (average over 10^6 runs). Red lines are analytical results under constant selection. For constant selection, the average fitness either increases or is unchanged by a new mutation, where the fraction of mutants that increases fitness is determined by θ . However, under frequency dependent selection, the average fitness of the population in the new equilibrium after a mutation can also decrease. The probability to increase, decrease the average fitness or maintain the same average fitness, depends on θ , for $\theta > \frac{\sqrt{2}-1}{\sqrt{2}}$ the probability to increase the average fitness in the new equilibrium is larger than the probability to decrease it.

characterizing each type, i.e. between a and b as well as between c and d . However, in the extreme case of $a = b$ and $c = d$, this would recover the case of constant selection, so we expect that such correlations would lead to results intermediate between constant and frequency dependent selection. We discuss how this distribution affects the changes in the average fitness during the evolutionary process. It turns out, the probability $\theta = \int_d^\infty dx f(x)$ that a payoff entry is larger than the fitness of the wild type (the parent type in the case of n types) d , is of particular interest and determines the change in the average fitness when initially only a single type is present. Remarkably, all other aspects of the fitness distribution turn out to be irrelevant for this observable.

The dynamics of evolving populations of interacting individuals shows stochastic fluctuations when selection is weak and when populations are small. In addition,

stochasticity can arise based on environmental changes or stochastic effects due to mutations. As we are interested in the effects of frequency dependent selection, we only consider stochasticity arising from random frequency dependent mutations and use the replicator equations to model evolutionary dynamics. The frequency of a certain type changes deterministically according to the difference of its own fitness to the average fitness in the population.

Suppose x is the frequency of the mutant type and $1 - x$ the frequency of the wild type, respectively. We can define the fitness of the mutant type, W_1 , and the fitness of the wild type, W_2 , as

$$\begin{aligned} W_1 &= ax + b(1 - x), \\ W_2 &= cx + d(1 - x), \end{aligned} \tag{6.2}$$

where a , b , c , and d are the entries in the payoff matrix. The average fitness of the population \bar{W} is given by

$$\bar{W} = xW_1 + (1 - x)W_2. \tag{6.3}$$

If the fitness of the mutant type is larger than the average fitness, its frequency will increase. If the fitness of the mutant type is below the average fitness, its frequency will decrease. We follow the usual assumption that the change of the frequency of the mutant type is given by the replicator equation (Taylor and Jonker, 1978; Zeeman, 1980; Hofbauer et al., 1982)

$$\dot{x} = x(W_1 - \bar{W}) = x(1 - x)(W_1 - W_2). \tag{6.4}$$

The change of the wild type frequency follows immediately as $-\dot{x}$. This dynamics is fully determined by the entries of the payoff matrix. Different constellations of the payoff entries cause different dynamical patterns. In the following, we discuss all generic cases of two-type interactions and how the average fitness of the population changes under the different situations.

First, we analyze the case where the mutant has higher fitness than the wild type for all frequencies x . This is the case for $a > c$ and $b > d$. The wild type goes extinct and the mutant type will be fixed in the population. Thus, the average fitness \bar{W} in the new equilibrium $x = 1$ is given by the payoff entry of the mutant type interacting with itself, a . We are interested in the probability, that the fitness of the population is increased after the fixation of the mutant. This becomes a

conditional probability of $a > d$ given that $a > c$ and $b > d$. Applying Bayes Rule, this can be expressed as

$$\begin{aligned} p(\overline{W}(1) > d \mid a > c, b > d) &= p(a > d \mid a > c, b > d) \\ &= \frac{p(a > d, a > c, b > d)}{p(a > c, b > d)} \\ &= \frac{p(a > d, a > c)}{p(a > c)}. \end{aligned} \quad (6.5)$$

We assume that the random variables a , b and c are independently derived from the same probability distribution. Hence, b does not depend on a or on c . Thus, the probability of $b > d$ is independent from the probability that $a > d$, which is used in Eq. (6.5). Since a and c are sampled from the same distribution, we have $p(a > c) = 1/2$ in the denominator. For the numerator, we have

$$\begin{aligned} p(a > d, a > c) &= \int_d^\infty da \int_{-\infty}^a dc f(c) f(a) \\ &= \int_d^\infty da F(a) F'(a) \\ &= \frac{1}{2} - \frac{F(d)^2}{2}, \end{aligned} \quad (6.6)$$

where $F(x)$ is the cumulative distribution function of a random variable with probability density function $f(x)$. As the probability that one of the new payoff entries a, b, c is greater than the wild type fitness d is $\theta = \int_d^\infty dx f(x) = 1 - F(d)$. Using this expression in Eq. (6.6), we arrive at

$$p(\overline{W}(1) > d \mid a > c, b > d) = 2\theta - \theta^2. \quad (6.7)$$

Strikingly, this only depends on θ , and is independent of the concrete choice of the probability density function $f(x)$. In population genetics, beneficial mutation rates are measured based on the concept of constant fitness, where the fitness of the mutant and the fitness of the wild type are both constant numbers. However, if we consider frequency dependent fitness, a new parameter is needed to represent the proportion of beneficial mutations. One option arising from our approach is to compare the payoff value of the mutant with the payoff value of the wild type when they are confronted by the same opponent. Since θ is the probability that the new payoff value of the mutant is larger than the wild type's payoff d , it corresponds to the probability that a mutation is beneficial under the constant selection scenario. If θ can be measured, the probability that the average fitness is increased by a random

mutant is independent of the payoff distribution according to Eq. (6.6). But different choices of probability density functions $f(x)$ will result in different values of θ , thus leading to different probabilities to increase the average fitness.

Next, we assume that a mutant type occurs with lower fitness than the wild type. With frequency dependence, there are two situations for such a mutant type. The mutant type can either have lower fitness than the wild type for all frequencies, or it can have a lower fitness only for small frequencies. In both cases, the mutant will go extinct and the average fitness will remain unchanged, since a mutant type is supposed to arise with a small amount.

Finally, a mutant type could be initially advantageous compared to wild types, but turn to be disadvantageous when it has reached a certain frequency. This occurs for $a < c$ and $b > d$. In this case neither the wild type nor the mutant type can take over the population, but there exists a mixed equilibrium consisting of mutant types at a frequency $x^* = \frac{b-d}{b-d-a+c}$ and wild types at a frequency $1 - x^*$. In this coexistence equilibrium, the fitness of the wild type subpopulation is equal to the fitness of the mutant type subpopulation. The average fitness of the system in the equilibrium is given by

$$\overline{W}(x^*) = ax^* + b(1 - x^*) = \frac{bc - ad}{b - d - a + c}. \quad (6.8)$$

Again, we ask for the probability of having a coexistence game that increases the average fitness. This is the conditional probability that $\overline{W}(x^*) > d$ given that $a < c$ and $b > d$, which can be written as

$$\begin{aligned} & p(\overline{W}(x^*) > d \mid a < c, b > d) \\ &= p((b - d)(c - d) > 0 \mid a < c, b - d > 0) \\ &= p(c > d \mid a < c) \\ &= \frac{p(c > a, c > d)}{p(c > a)} \end{aligned} \quad (6.9)$$

This is identical to Eq. (6.5) if one exchanges $a \leftrightarrow c$. Since a and c have the same distribution, we recover the result from Eq. (6.7),

$$p(\overline{W}(x^*) > d \mid a < c, b > d) = 2\theta - \theta^2. \quad (6.10)$$

In other words, the probability to increase fitness is the same in a coexistence game as in a game where the mutant dominates the wild type.

Let us now combine the results and consider the changes of the average fitness over all types of interactions. The probability to increase the fitness due to a new mutation is given by

$$\begin{aligned}
 p(\overline{W} > d) &= \underbrace{p(\overline{W} > d \mid a > c, b > d)}_{2\theta - \theta^2} \underbrace{p(a > c, b > d)}_{\frac{\theta}{2}} \\
 &+ \underbrace{p(\overline{W} > d \mid a < c, b > d)}_{2\theta - \theta^2} \underbrace{p(a < c, b > d)}_{\frac{\theta}{2}} \\
 &+ \underbrace{p(\overline{W} > d \mid b < d)}_0 \underbrace{p(b < d)}_{1 - \theta} \\
 &= 2\theta^2 - \theta^3
 \end{aligned} \tag{6.11}$$

In a similar manner, we can calculate the probability to decrease the average fitness due to a new mutation. When the mutant dominates the wild type, the average fitness may still decrease. This is exactly what happens in the Prisoner's Dilemma (Rapoport and Chammah, 1965; Nowak, 2006b). Equivalently to the calculation above, we have

$$\begin{aligned}
 p(\overline{W}(1) < d \mid a > c, b > d) &= \frac{p(a < d, a > c, b > d)}{p(a > c, b > d)} \\
 &= \frac{p(a < d, a > c)}{p(a > c)} \\
 &= (1 - \theta)^2.
 \end{aligned} \tag{6.12}$$

For the probability to decrease the average fitness in a coexistence game, we find

$$p(\overline{W}(x^*) < d \mid a < c, b > d) = (1 - \theta)^2. \tag{6.13}$$

Thus, using a calculation similar to Eq. (6.11), the overall probability to decrease the average fitness is given by

$$p(\overline{W} < d) = \theta - 2\theta^2 + \theta^3. \tag{6.14}$$

Also the probability to maintain a constant average fitness can be calculated in this way. For continuous fitness distributions, there are no strictly neutral mutations. As the fitness of the wild type is a specific value of the continuous random variable, the probability of having a strict neutral mutation, the fitness of which is equal to the fitness of the wild type, is 0. Thus, the average fitness is only maintained when the mutant goes extinct, which occurs with probability

$$p(\overline{W}(0) = d) = p(b < d) = 1 - \theta \tag{6.15}$$

We discussed the changes of the average fitness in a two-type population under frequency dependent selection above. Under constant selection, the average fitness will increase with probability θ and decrease with probability 0 . As for frequency dependent selection, it will remain constant with probability $1-\theta$. Fig. 6.2 illustrates these results and compares frequency dependent selection to constant selection for all values of θ . For frequency dependent selection, there is an intersection point θ_* , where the probability to increase the average fitness and to decrease the average fitness are equal. Using Eq. (6.11) and Eq. (6.14), this becomes $2\theta_*^2 - \theta_*^3 = \theta_* - 2\theta_*^2 + \theta_*^3$, and we have $\theta_* = \frac{\sqrt{2}-1}{\sqrt{2}}$. Small values of θ are typically considered to be of biological relevance. In this case, frequency dependent selection tends to decrease the average fitness: for $\theta < \frac{\sqrt{2}-1}{\sqrt{2}}$, it is more likely that the average fitness of the population is decreased by a single random frequency dependent mutation; for $\theta > \frac{\sqrt{2}-1}{\sqrt{2}}$, it is more likely that it is increased.

Frequency-dependent selection can arise from different mechanisms. In a haploid population, frequency-dependent selection are caused by the interactions of different types. In this case, the fitness of a particular type depends on the frequency of its own and other types in the population. However, in a diploid population, frequency dependent selection on alleles can arise also from the interactions of two alleles at one locus (Cressman, 1992; Hofbauer and Sigmund, 1998; Traulsen and Reed, 2012). Thus, our model can be easily extended to a diploid population in such a case, which leads to different results for the average change in fitness, see 6.3.4.

6.3.2 Games with n types

So far, we have discussed the change of the average fitness of a population consisting of at most two types. However, when two types coexist in a stable polymorphism, an additional type can enter the population and persist. To describe the interaction of individuals in a population with more than two types, we extend the 2×2 payoff matrix to a $n \times n$ payoff matrix A , where n is the number of types in the population. The entry in the i -th row and the j -th column, A_{ij} represents the fitness of an i -type individual interacting with a j -type individual. The fitness of type i on average can be written as $W_i(x) = \sum A_{ij}x_j$, where $j = 1, 2, 3, \dots, n$, and x_j is the frequency of type j , such that $\sum_{j=1}^n x_j = 1$.

In our model, n is not a fixed number. When a type goes extinct, the corresponding row and column are deleted in the payoff matrix. Thus, the value of n

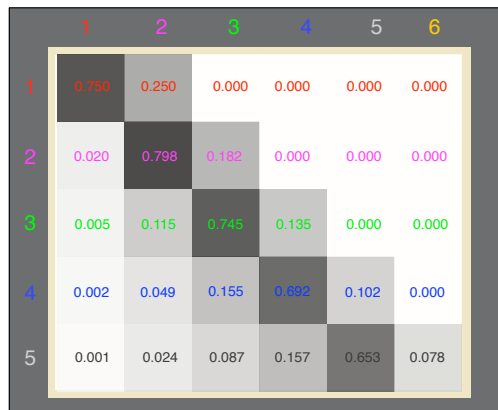


Figure 6.3: **Transition probabilities between different levels of diversity.** The entry in row i and column j is the transition probability from a stable coexistence of i types to a stable coexistence of j types, numbers are also color coded. The mutation rate is so low that the transitions between different states are caused by the appearance of a single mutation. The higher the number of coexisting types is, the more difficult the state is to be reached. Here we show the transition for up to six co-existing types ($\theta = 0.5$, averages obtained over 500 independent realizations and 20000 mutations per realization).

decreases by one. When a mutation occurs, one row and one column are added to describe the interactions of the mutant type and resident types, which increases the size of the payoff matrix by one. The new entries introduced by a mutation are generated based on the assumption that the interactions between the mutant type m and any resident type i are similar to those between the parent type p and the resident type i . In our case, we assume a_{mj} is a random variable which is drawn from a probability density function $f(x)$ and is larger than a_{pj} with probability θ .

Since the complexity of the population dynamics increases considerably with the number of types, it would be difficult to obtain the changes of the average fitness in a polymorphic population of $n > 2$ types analytically. Therefore, we use the replicator equations to simulate the dynamics of the system with several types. We start the simulation from a homogenous population. However, since we are interested in the average fitness changes and other stationary quantities averaged over a long time period, the initial number of types has no effects on the results. The time intervals are sufficiently small that at most one mutant type can appear during one time interval. The probability that a resident type i produces a mutant type is $\mu x_i W_i(x) / \bar{W}(x)$, where $i = 1, 2, 3, \dots, n$. Thus the probability that a mutant

arises from a resident type i increases with the fitness of this type. However, for the whole population, the probability that a mutant type appears is just the mutation rate, $\sum_{i=1}^n \mu x_i W_i(\mathbf{x}) / \bar{W}(\mathbf{x}) = \mu$.

We can choose arbitrary mutation rates in our simulations. However, when the mutation rate is very high, a population might experience a new mutation when it is still in a non-equilibrium state triggered by the previous mutation. In this case, the fate of a mutant is not only driven by selection, but also by the interplay of mutations. Since we are interested in the fitness consequences of frequency dependent selection, we choose the mutation rate small enough such that a population disturbed by a mutation reaches the new equilibrium before the next mutation arises.

We first look at the transition probability between different levels of diversity under mutation and selection. Once a mutation occurs it can coexist with all resident types, replace one resident type, outcompete some resident types, or go extinct. The transition matrix T describes this dynamics. Suppose the number of types in the current population is n . The element T_{ni} denotes the transition probability from n to i coexisting types, where $i = 1, 2, 3, \dots, n+1$, see Fig. 6.3. We obtain the values in the transition matrix from numerical simulations. Every transition event triggered by a mutation is recorded and the probability to go from a certain number of types to another number of types is averaged over many realizations. These transition probabilities show some interesting properties. The probability to keep the current diversity (the element in the main diagonal in a row) is always higher than the probabilities to decrease or increase the diversity (all the other elements in the same row), see Fig. 6.3 and Ref. (Huang et al., 2012). Interestingly, for a population consisting of less than 4 types, the probability to increase the diversity T_{ii+1} is higher than the probability to decrease the diversity $\sum_{j=1}^{i-1} T_{ij}$ in the parameter regime of Fig. 6.3. Once the population reaches the threshold of 4 types, this pattern reverses. Thus in the long run the population tends towards an intermediate level of diversity. Furthermore, we observe the ranking, $T_{12} > T_{23} > T_{34} > T_{45}$. This suggests that the probability to reach higher levels of diversity decreases with increasing diversity even for larger number of initial types. The transition probability from one type to a two-type coexistence can be calculated analytically based on the comparison of payoff entries, see above. Thus, $T_{12} = p(a < c)p(b > d) = \theta/2$, which is confirmed by our simulation results of T_{12} under different θ for the n -type model.

For a population with n types, the changes of the average fitness are more complicated, as the interactions between different types are much more diverse than in a two-type population. Even a classification of different types of interactions in such a population is difficult and of limited value to understand the change in average fitness. Instead, we evaluate the changes of the average fitness between these states numerically.

A mutation can increase, maintain, or decrease the diversity level of the population. We present the changes of the average fitness in these three scenarios, see Fig. 6.4, for those transitions which happen most frequently (see Fig. 6.3). For small θ , mutants are more likely to obtain lower fitness than their parents type does, in the interactions with the same resident type. This can cause the decrease of the average fitness in all three situations. If θ is sufficiently small, the average fitness will decrease all the time. When θ becomes larger, the average fitness can increase. The larger θ is, the larger the increase is. Thus, our results under the replicator dynamics provide not only the change of the average fitness under a constant θ , but also the direction and magnitude of the average fitness changes. In real systems, one may expect that θ decreases during the adaption of the population. However, e.g. environmental changes could also increase it.

6.3.3 Games with equal gains from switching

So far, we have assumed that the payoff of the mutant interacting with another resident type is derived from the payoff of its parent interacting with the same resident type. In a population with only two types, this leads to the case where the three random payoff entries, a , b and c , are all related to d . As a null model, we have assumed that a , b and c are uncorrelated. While this is the simplest possibility, it may not be the case for concrete biological systems. Therefore, we analyze an different case here which focuses on particular cases of frequency dependence, but includes such correlations.

We focus on an evolutionary game with the payoff matrix

$$\begin{array}{cc} & \begin{array}{cc} \text{M} & \text{R} \end{array} \\ \begin{array}{c} \text{M} \\ \text{R} \end{array} & \begin{pmatrix} d + \varepsilon + \delta & d + \delta \\ d + \varepsilon & d \end{pmatrix}, \end{array}$$

where ε and δ are independent random variables with probability distributions $f_\varepsilon(x)$

and $f_\delta(x)$ respectively. ε represents the effect of a mutation on the mutant type, and δ represents the effect of a mutation on those who interact with the mutant type. This game has the property of “equal gains from switching”, where the sum of the payoff values in the main diagonal is equal to the sum of the payoff values in the other diagonal (Nowak and Sigmund, 1990). It can arise from the assumption that the two types are close to each other in a continuous phenotype space (Wild and Traulsen, 2007). The case of $\delta = 0$ corresponds to constant selection. Note that there are no coexistence games when we assume such payoff matrices. If $\varepsilon > 0$, the mutant will take over the population ($d + \varepsilon + \delta > d + \delta$ and $d + \varepsilon > d$), and the new average fitness becomes $\bar{W} = d + \varepsilon + \delta$. Compared with the former average fitness d , the average fitness increases if $\varepsilon + \delta > 0$, and decreases if $\varepsilon + \delta < 0$. If $\varepsilon < 0$, the mutant will be outcompeted by the wild type ($d + \varepsilon + \delta < d + \delta$ and $d + \varepsilon < d$), and the average fitness of the population remains the same.

The probability to increase the average fitness becomes $p(\bar{W} > d) = (1 - \theta_\varepsilon) \cdot 0 + \theta_\varepsilon \cdot p(\varepsilon + \delta > 0 \mid \varepsilon > 0)$, where θ_ε is the probability that ε is larger than 0, and $p(\varepsilon + \delta > 0 \mid \varepsilon > 0)$ is the conditional probability that the sum of ε and δ is larger than 0 given ε is larger than 0. This conditional probability can be written as

$$\begin{aligned} p(\varepsilon + \delta > 0 \mid \varepsilon > 0) &= \frac{p(\delta > -\varepsilon, \varepsilon > 0)}{p(\varepsilon > 0)} \\ &= \frac{\int_0^\infty dx \int_{-x}^\infty dy f_\delta(y) f_\varepsilon(x)}{\theta_\varepsilon} \end{aligned} \quad (6.16)$$

The values of θ_ε and $p(\varepsilon + \delta > 0 \mid \varepsilon > 0)$, which determine the probability that the average fitness increases, depend on the concrete choice of $f_\varepsilon(x)$ and $f_\delta(x)$. The integrals can only be carried out in special cases.

It is worth to mention there is a difference between games with equal gains from switching and games with independent random payoff entries on the population dynamics. In an infinite population, where genetic drift has no effect on the population dynamics, the resulting dynamics under positive frequency dependent selection and under constant selection are similar, as there are no stable coexistences. Successful mutants will invade and take over the population sequentially. The diversity will only increase if the mutation rate is high enough. On the contrary, when different kinds of interactions, especially negative frequency dependent selection, are allowed (for example, the case with independent random payoff entries), diversity can increase even for lower mutation rates (see above).

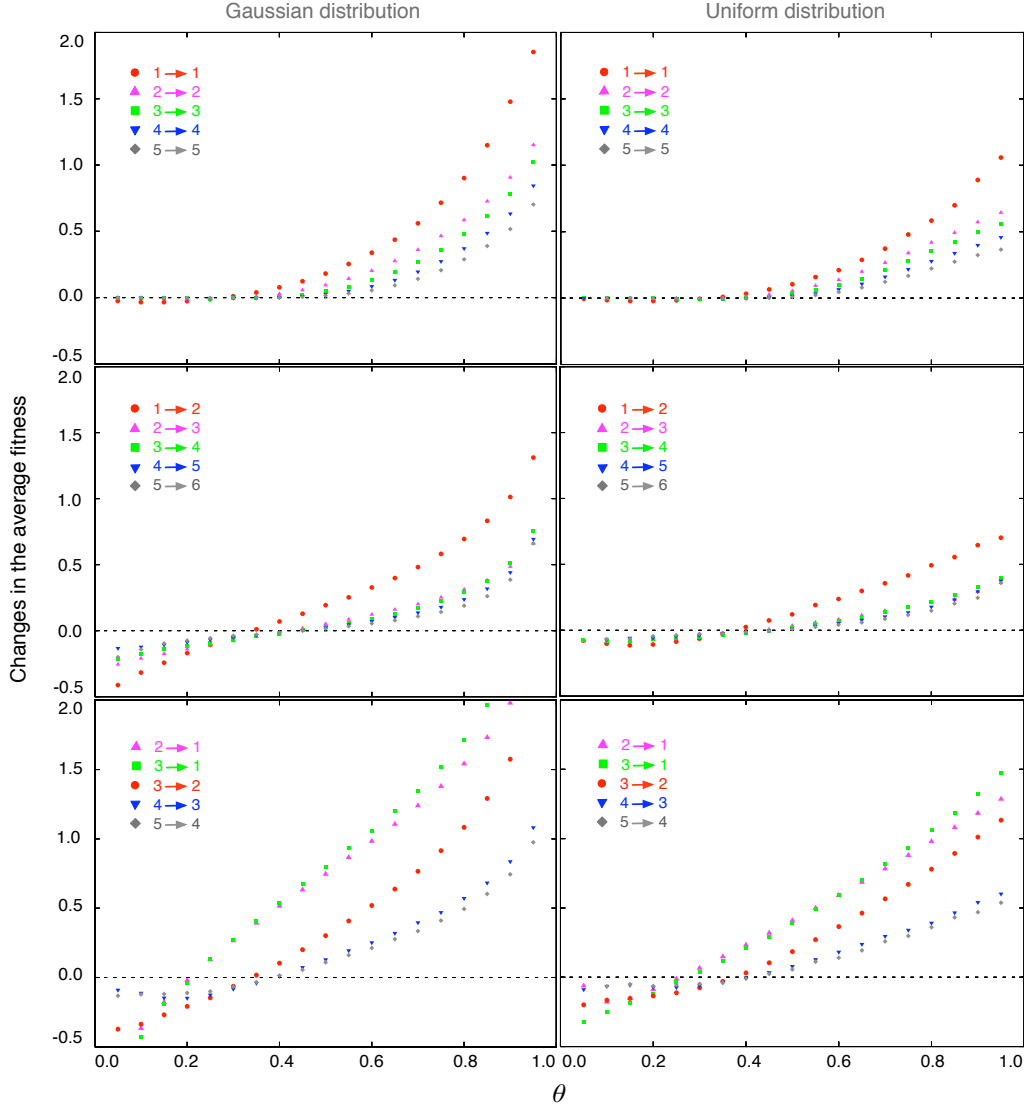


Figure 6.4: **Changes in the average fitness when a population evolves between different levels of diversity under various probabilities that a mutant payoff values is larger than the parent's θ .** The symbols are simulation results based on replicator dynamics. The number of different types can either stay the same, increase by one or decrease by any number, because at most a single mutation enters the population. Note that the average fitness of the population in the new equilibrium decreases for small θ in all three cases after a transition. Thus even if a mutant takes over a population, the average fitness can decrease. With increasing θ , the average fitness will increase over time, but the fitness gain reduces with increasing diversity. The difference among results under Gaussian distribution and uniform distribution with the same variance, shows that the absolute changes of the average fitness also depends the concrete shapes of the probability distribution (every symbol is averaged over 500 independent realizations and 20000 mutations per realization. The probability distribution $f(x)$ is Gaussian (left) or uniform (right) with variance 1).

6.3.4 Diploid populations with two alleles

The impact of Mendelian inheritance on the population dynamics has been discussed in the framework of evolutionary game theory before (Hines, 1980; Hofbauer et al., 1982; Eshel, 1982; Brown, 1983). In a diploid population, the combinations of two alleles at a given locus on a pair of homologous chromosomes, can be interpreted by a special two player game. Suppose there are allele A and allele B . The fitness of different genotypes, W_{AA} , W_{AB} and W_{BB} can be described by a 2×2 matrix

$$\begin{array}{cc} & \begin{array}{cc} A & B \end{array} \\ \begin{array}{c} A \\ B \end{array} & \begin{pmatrix} W_{AA} & W_{AB} \\ W_{AB} & W_{BB} \end{pmatrix}. \end{array}$$

This is mathematically identical to the game with two types discussed above. Here, W_{AA} corresponds to a , W_{AB} to $c = b$, and W_{BB} to d . For a population initially only with homozygotes BB , the probability of increasing the average fitness \bar{W} caused by a random new allele A , can be calculated by setting $c = b$ in Eq. (6.11). This becomes

$$\begin{aligned} p(\bar{W} > W_{BB}) &= \underbrace{p(a > d \mid a > b, b > d)}_1 \underbrace{p(a > b, b > d)}_{\theta - \frac{\theta^2}{2}} \\ &\quad + \underbrace{p(\frac{b^2 - ad}{2b - d - a} > d \mid a < b, b > d)}_1 \underbrace{p(a < b, b > d)}_{\theta - \frac{\theta^2}{2}} \\ &\quad + \underbrace{p(d > d \mid b < d)}_0 \underbrace{p(b < d)}_{1 - \theta} \\ &= 2\theta - \theta^2 \end{aligned} \tag{6.17}$$

The probability that the average fitness decreases in such a population is 0, because the diploid AB and the diploid BA is indistinguishable, $c = b$. In asymmetric diploids, where the maternal alleles and paternal alleles are not equally expressed, the average fitness changes are exactly the same as shown in a general case of haploid populations.

6.4 Discussion

Mutants with high individual fitness do not necessarily increase the average fitness of the population under frequency dependent selection. Similarly, the mutants which

maximize the average fitness of a population are not necessarily those leading to a stable equilibrium in this scenario. An example for a two-type population is that a mutant interacts with the wild type in a game like Prisoners' Dilemma (Rapoport and Chammah, 1965; Maynard Smith, 1982). This is a special case of a dominance game, where the defector (the mutant) outcompetes the cooperator (the wild type) and causes a reduction in the average fitness. For example, in the RNA phage $\phi 6$, the competitive interactions among the high multiplicities-of-infection phage (the defector) and the low multiplicities-of-infection phage (the cooperator) in the same host cell are studied, which conforms to the Prisoners' Dilemma (Turner and Chao, 1999). In this experiment, when the defector invades with a low frequency, it has higher fitness than the residents ($c > a$), but the average fitness decreases when the defector becomes fixed ($d > a$).

Since natural selection works on an individual level rather than a population level, it does not always lead to an increase of the average fitness. Our random mutant games model accommodates mutations under frequency-dependent selection, which can result in an increase or decrease in the average fitness, not only for the simplest case of two types but also for an arbitrary number of mutant types. An interesting aspect of our model is that even though it allows for an infinite number of mutant types, it does not result in a continuous growth of diversity in a population, but leads to an intermediate level of diversity (Huang et al., 2012). We assume that the payoffs are constant in time and identical for individuals of the same type. If individuals vary in their payoffs despite being of the same type, the results are altered by this additional source of randomness (Fudenberg and Harris, 1992; Johnson et al., 2002). In a population with two types, we calculate a particular value θ_* , where the probability that the average fitness increases is equal to the probability it decreases. The exact value of θ_* depends on the concrete implementation of the payoff matrix. An interesting result of our model is that the probability to decrease or increase fitness depends only on a particularly simple property of the fitness distribution. While this may not be of direct relevance to a concrete biological system, it illustrates conceptually that a decreasing fitness may not be counterintuitive even under the simplest possible assumptions of frequency dependence.

We have discussed the changes in the average fitness for an infinite asexual population under mutation and selection. Additional effects occur when the population size becomes finite and genetic drift is not negligible (Huang et al., 2012). However, our main observation is that the average fitness at equilibrium can only increase

or remain constant by random mutations under constant selection, but also decrease under frequency-dependent selection. This can shed new light on problems in evolutionary biology and leads to the exciting question on the dynamics of the average population fitness in real biological populations. In an asexual finite population, random genetic drift leads to the accumulation of deleterious mutations and an continuous decrease in the average fitness, which is well known as Muller's ratchet (Muller, 1964). Without any forms of recombination and epistasis, beneficial mutations are the only source to compensate the average fitness decline. Since the probability of increasing the average fitness by random mutations is lower under frequency-dependent selection (see Fig. 6.2), we must conclude that asexual populations face an even bigger challenge to maintain their average fitnesses under frequency dependent selection than under constant selection in a finite population. This is particularly striking when θ is small, a case that is typically thought of as the biologically most relevant case.

In population genetics, the change of the average fitness has also been studied in diploid systems (Fisher, 1930, 1941). However, our approach starting from a different point of view, not only allows the interplay of mutation and selection, but also a wider interpretation of the fitness of heterozygotes. Suppose A and B are two alleles at the same locus. In population genetics, the fitness of genotype AB and BA is usually considered to be identical, which is a special case in our model called symmetric diploids. However, this does not hold in asymmetric diploids where the maternal allele and paternal allele are not equally expressed. Our model and our analysis allow both cases. In the framework of a well-mixed symmetric diploid population (corresponding to random mating), our result that the average fitness never decreases is consistent with the former statement in population genetics.

Frequency dependent interactions can lead to a decrease of the average fitness of a population during the process of evolution despite natural selection. This is because natural selection works on individual fitness instead of the average fitness of a population.

6.5 Random mutations and cancer

As discussed in chapter 3, the majority of cancers are caused by multiple mutational hits within single cells. The interplay of these mutational hits might give a fitness advantage to those cells, and they escape control mechanisms, ultimately leading to cancer. However, most mutations are either neutral or disadvantageous (Eyre-Walker, 2007). Thus, those mutations do not contribute or even compensate for potentially cancerous mutations. Game theoretical approaches in cancer modeling exist already. These approaches allow an alternative perspective on cancer progression and may also provide alternative treatment strategies (Basanta et al., 2008, 2011; Basanta and Deutsch, 2008). Random mutant games are a natural framework to combine game theoretical approaches and the impact of random mutations.

Besides the effect of random mutations, there is a constant cell flow due to the hierarchical tissue organization. To connect the concept of random games with a hierarchical tissue organization, we first have to implement a fitness concept. Usually in biology, fitness corresponds to the number of produced offspring. The more offspring one individual has compared to others the higher its fitness. Thus in hierarchical tissues, the fitness can be understood as the capability of a cell to produce daughter cells. Following (Traulsen et al., 2010) and the arguments in chapter 3 the number of cell divisions of a cell before leaving a compartment is

$$n = \sum_{i=1}^{\infty} i (1 - \varepsilon)^{i-1} \varepsilon = \frac{1}{\varepsilon}. \quad (6.18)$$

Cells leave the compartment with probability ε . Before this cell undergoes i cell divisions with probability $(1 - \varepsilon)^i$ within the compartment. The number of produced cells ρ is $n - 1$ as in the last division, the cell leaves the compartment. However, we also need to take the offspring in higher compartments into account. Assuming the direct number of offspring in compartment $i - 1$ is ρ_{i-1} , and all differentiation probabilities are constant, the offspring in compartment i is

$$\rho_i = \rho_{i-1} \frac{1}{\varepsilon} - \rho_{i-1} = \left(\frac{1 - \varepsilon}{\varepsilon} \right)^i. \quad (6.19)$$

However, if we assume differentiation probabilities ε_i for each compartment i this is

$$\rho_i = \prod_{j=1}^i \frac{1 - \varepsilon_j}{\varepsilon_j}. \quad (6.20)$$

However, this measures the direct offspring of a cell in all compartments. It seems more appropriate to measure all cells derived from a cell (the daughter cells of the daughter cells and so on). Including mutations, this leads to a recurrence relation for the number of produced cells in compartment i

$$m_i^k = \frac{\alpha_i^k}{2\alpha_i^k - 1} \sum_{l=1}^i 2^{i-l} \frac{u}{\alpha_l^{k-1}} m_l^{k-1} \prod_{h=l}^{i-1} \frac{\varepsilon_h^k}{2\alpha_h^k - 1}. \quad (6.21)$$

We derived and discussed this recurrence relation in chapter 3. Here i is the compartment number, k the number of mutational hits in a cell, ε the differentiation probability, u the mutation probability and $\alpha = \varepsilon + u$. Importantly, the total number of produced cells depends on u and ε only. This gives us an ansatz to introduce random mutations into the framework of hierarchical tissues. We assume the differentiation probability of a mutant follows a probability distribution. Thus if a cell mutates during proliferation, the differentiation probability of this single mutation ε^1 is drawn from this probability distribution. Note, the differentiation probability is restricted to the interval $[0, 1]$. Thus the distribution of the differentiation probability needs to be bounded within this interval. This can be achieved for any continuous probability distribution $f(x)$. This leads to a conditional (truncated) distribution of the original distribution, given by

$$f^c(x|a \leq x \leq b) = \frac{f(x)}{F(b) - F(a)}. \quad (6.22)$$

The probability distribution is restricted to the interval $[a, b]$ and $F(x)$ represents the cumulative distribution function. In figure 6.5 we show a realization of a truncated gaussian distribution. This distribution has its mean at the wild types differentiation probability and relatively narrow range (small variance). The actual distribution of fitness effects of mutations in human tissues is unknown. However, one would expect a pronounced peak around the wild types fitness, as most mutations are assumed neutral. Anyhow, due to the hierarchical tissue organization most mutants go extinct. Only mutations, that cause a low differentiation probability may give raise to additional mutations, which may lead to cancer. But such mutations are rare. A further possibility provide stepwise mutations towards a low differentiation probability, see figure 6.5. In chapter 3 we showed, that such multiple mutations are strongly suppressed by a hierarchical organized tissue organization. Thus we have at least two competing effects: (i) the rareness of mutations causing low differentiation probabilities and (ii) the inability of the mutant population to derive a state of

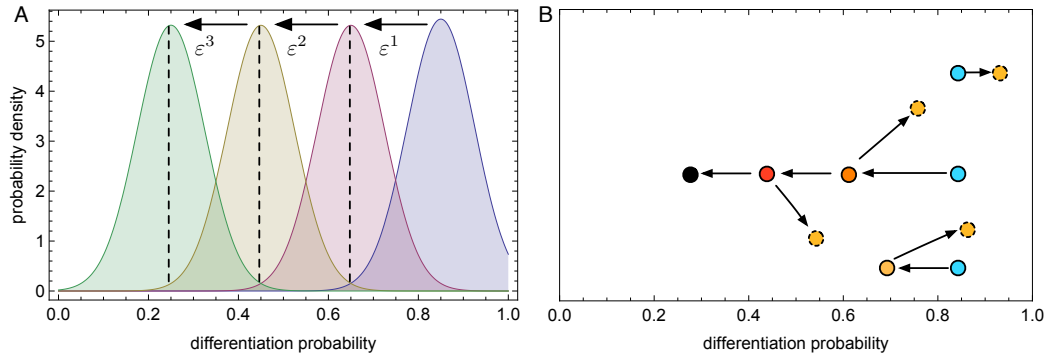


Figure 6.5: **Schematic representation of random mutations in a hierarchical tissue structure.** **A** If a mutation occurs, the differentiation probability of the mutant is drawn from a probability distribution. This distribution peaks around the differentiation probability of the ancestor cell (most mutations are neutral). However, the differentiation probability of the mutant might be lower compared to the ancestor cell, thus the distribution shifts to the left. In a few steps a state of low (potentially cancerous) differentiation probability can be derived. **B** Multiple mutations can lead to complex patterns. Jumps in the differentiation probabilities may occur, most of the mutants extinct (dashed circles), but there might be a path leading to a low differentiation probability (black circle) and potentially to cancer.

low differentiation probabilities by stepwise mutations because of the hierarchical tissue organization. We think, this might result in interesting dynamical patterns and investing them is worthwhile. This might be an interesting problem for future research.

Summary and Outlook

7.1 Summary

Throughout this thesis we developed models about different aspects of cell proliferation properties. In chapter 2, we focused on the dynamics of cells in hierarchically organized tissue structures. Such hierarchical organizations increase the complexity of modeling approaches. Cell flux causes varying compartment sizes. Further, population dynamics in such hierarchies usually span many orders of magnitude and very diverse dynamical patterns can be observed. We tackle these problems by (i) implementing exact individual based stochastic computer simulations, utilizing a Gillespie algorithm, and (ii) by analytically solving coupled systems of differential equations. The first part of chapter 2 contains the publication ([Werner et al., 2011](#)). Here we focus on the deterministic dynamics of a cell with a single mutational hit within such hierarchies. The dynamics differs for stem cell and non stem cell mutations. Stem cell mutations accumulate over time. For non stem cell mutations, the dynamics depends on the properties of the mutant. If the differentiation probability of the mutant is considerably lower than wild type cells, mutant cells accumulate. Otherwise, mutant cells will travel through the hierarchy and get washed out in the long run. We can calculate the extinction time of such mutants and find a good agreement to experimental results for neutral mutants in the literature.

The second part of chapter 2 contains a submitted manuscript and turns to the problem of multiple mutations in such hierarchically organized tissues. We investigate the accumulation of multiple mutations in a hierarchy. Although crucial for cancer initiation, such multiple mutations are strongly suppressed. Stem cells are rare and proliferate slowly, thus an accumulation of multiple mutations is unlikely. However, cells in more committed stages of the hierarchy undergo a limited number of cell division before becoming senescent. Thus, most mutations get washed out

of the hierarchies. We introduce the concept of the reproductive capacity of a cell. This property is defined as the total number of cells derived from a single cell in a compartment and directly corresponds to the concept of clonal markers. We derive at closed expressions for the reproductive capacity of cells. These solutions allow us to calculate the number of multiple mutations during the differentiation process of those cells. We find explicit scaling properties of such hierarchically organized tissues, highlighting the strong suppression of such multiple mutations.

In chapter 4, we turn to a related problem. The development of molecular targeted drugs to fight chronic myeloid leukemia was a breakthrough in the treatment of this particular cancer. These drugs turn an ultimately lethal into a chronic disease. Unfortunately, the occurrence of treatment resistant subclones and a disease relapse are an increasing problem in clinical daily routine. The main body of chapter 4 contains the publication (Werner et al., 2011). Here, we analyze a resistance inducing experiment to those molecular targeted drugs. We develop a minimalistic mathematical model based on the two and three type Moran process with time dependent fitness functions and compare the dynamics to experimental results. The fitness of the cancer cell population passes through a minimum before it reaches a relatively stable state. This state shows considerable lower fitness than the wild type, but is resistant to high concentrations of those molecular targeted drugs. It turns out that a minimalistic description of the experimental results requires at least three types, the wild type and two resistant types with different switching rates. We find hints for a fast switching type with slightly lower fitness compared to the slower switching type. However, due to the higher fitness, the fitter type takes over the population in the long run.

Chapter 5 contains work in progress on a model of telomere shortening on the cell and tissue level. We derive analytical expressions for the expected telomere length distribution and are able to compare those to clinical measurements of telomere length distributions in healthy adults. This reveals interesting properties of stem cell dynamics. We derive a microscopic model that explains basic observations of telomere shortening at different ages. We find an increased shortening of telomeres at young ages, that turns into an almost linear decrease for adults. Importantly, the model hints towards an expanding stem cell pool size with increasing age. In addition we can deduce dynamical patterns of telomere shortening for certain diseases, such as T-cell mediated stem cell death. In this case, we expect a quadratic decline of the average telomere length compared to the approximately linear decline in healthy

adults. However, understanding basic properties of stem cell dynamics is crucial for an improved description of hierarchically organized tissue structures.

The last chapter 6 of the main body of this thesis contains the publication (Huang et al., 2012), where we implement a method for an arbitrary number of random mutations into the framework of game theory. We investigate the effects of such random mutations on the average fitness of an infinite large population by applying replicator equations. We find fundamental differences for constant and frequency dependent selection. The average population fitness never decreases for fixed mutants, if selection is constant. In contrast, the average fitness of the population can decrease after the fixation of a mutant for frequency dependent selection. We calculate the probability for such events explicitly. This gives two conclusions: (i) The probability of such events only depends on the available fraction of advantageous mutants, but is independent of the actual payoff distribution (ii) If the proportion of available advantageous mutations is not too large, the probability to decrease the average fitness is higher than the probability to increase the average fitness. Our publication (Huang et al., 2012) explores basic properties of the random mutant model. We do not link this directly to cell population dynamics in hierarchies. However, in the last part of chapter 6 we discuss one mechanism, that allows to include the idea of random mutations into the framework of hierarchical tissue organization, which may enable us to tackle additional effects of non neutral multiple mutation hits.

7.2 Outlook

In this thesis we modeled different aspects of cell population dynamics. However, there are interesting questions for future research, that directly follow from results in this thesis. So far we focused on deterministic effects of cell dynamics in hierarchically organized tissue structures. The investigation of stochastic effects and their interference with deterministic predictions in such hierarchies is a promising task. We discussed one possibility in chapter 6, where we propose an implementation of random mutations. The differentiation probability of such mutants would follow a probability distribution. This approach would allow us to study the stochastic clonal evolution in such hierarchies. One could investigate the relative importance of different scenarios. Cancerous clones either evolve by a few mutational hits of big impact or many mutational hits with minor changes. Most cancers might fall into

the first scenario, as many mutational hits are unlikely. However, a mathematical description of such clonal histories is of interest.

A second interesting question concerns our model of telomere shortening. We modeled the shortening of telomeres on the cell level. This allowed us to neglect stochastic effects. However, stochastic effects are important on the single telomere level. Building a model on the single telomere level, that derives our model as an average of the cell telomere length would be of interest. We expect pronounced differences between the telomere length distributions on single telomere and cell level. Deriving exact expressions could reveal additional insights into the stem cell dynamics. A comparison of both solutions may allow an estimation of the importance of stochastic effects on the stem cell dynamics.

However, we also expect effects on the clonal dynamics due to the expending stem cell pool size. There are some studies, that investigate the dynamics of stem cell mutations. Usually, these studies assume a constant instead of an expanding stem cell pool. In addition the proliferation rate of stem cells slows down. Thus we expect an (maybe small) impact on the clonal dynamics.

I believe this thesis shows at least partly, how the cooperation of theoretical, experimental and clinical research can improve our understanding of fundamental questions in cell population dynamics and subsequent questions as cancer, treatment or aging. Usually the underlying systems are complex and so are the observed dynamical patterns. However, even more important are mathematical models, that can test for assumptions and may even result in predictions. Clarifying the driving processes are indispensable in developing potential cures for cancer.

Bibliography

- Abkowitz, J. L., S. N. Catlin, and P. Gutterp (1996). Evidence that hematopoiesis may be a stochastic process in vivo. *Nature Medicine* 2, 190–197. (Cited on pages 4, 12 and 30.)
- Abkowitz, J. L., S. N. Catlin, M. T. McCallie, and P. Gutterp (2002). Evidence that the number of hematopoietic stem cells per animal is conserved in mammals. *Blood* 100, 2665–2667. (Cited on pages 28 and 30.)
- Allday, M. J. (2009). How does epstein–barr virus (ebv) complement the activation of myc in the pathogenesis of burkitt’s lymphoma? *Seminars in Cancer Biology* 19, 366–376. (Cited on page 7.)
- Allsopp, R., H. Vaziri, C. Patterson, S. Goldstein, E. Younglai, A. Futcher, C. Greider, and C. Harley (1992). Telomere length predicts replicative capacity of human fibroblasts. *Proceedings of the National Academy of Sciences* 89, 10114–10118. (Cited on page 5.)
- Allsopp, R. C., E. Chang, M. Kashefi-Aazam, E. I. Rogaev, M. A. Piatyszek, J. W. Shay, and C. B. Harley (1995). Telomere shortening is associated with cell division in vitro and in vivo. *Experimental Cell Research* 220, 194–200. (Cited on page 11.)
- Altrock, P. M., C. S. Gokhale, and A. Traulsen (2010). Stochastic slowdown in evolutionary processes. *Physical Review E* 82, 011925. (Cited on page 18.)
- Altrock, P. M. and A. Traulsen (2009a). Deterministic evolutionary game dynamics in finite populations. *Physical Review E* 80, 011909. (Cited on pages 18 and 19.)
- Altrock, P. M. and A. Traulsen (2009b). Fixation times in evolutionary games under weak selection. *New Journal of Physics* 11, 013012. (Cited on page 18.)
- Altrock, P. M., A. Traulsen, and T. Galla (2012). The mechanics of stochastic slowdown in evolutionary games. *Journal of Theoretical Biology* 311, 94–106. (Cited on page 18.)
- Andersson, D. I. (2006). The biological cost of mutational antibiotic resistance: any practical conclusions? *Current Opinion in Microbiology* 9, 461 – 465. (Cited on page 80.)

- Andersson, D. I. and H. Hughes (2010). Antibiotic resistance and its cost: is it possible to reverse resistance? *Nature Reviews Microbiology* 8, 260–271. (Cited on page 80.)
- Antal, T., K. B. Blagoev, S. A. Trugman, and S. Redner (2007). Aging and immortality in a cell proliferation model. *Journal of Theoretical Biology* 248, 411–417. (Cited on page 82.)
- Araten, D., M. Bessler, S. McKenzie, H. Castro-Malaspina, B. Childs, F. Boulad, A. Karadimitris, R. Notaro, L. Luzzatto, et al. (2002). Dynamics of hematopoiesis in paroxysmal nocturnal hemoglobinuria (pnh): no evidence for intrinsic growth advantage of pnh clones. *Leukemia* 16, 2243. (Cited on page 40.)
- Araten, D., K. Nafa, K. Pakdeesuwan, and L. Luzzatto (1999). Clonal populations of hematopoietic cells with paroxysmal nocturnal hemoglobinuria genotype and phenotype are present in normal individuals. *Proceedings of the National Academy of Sciences* 96, 5209–5214. (Cited on pages 39, 40, 41, 42 and 46.)
- Ariew, A. and R. C. Lewontin (2004). The confusions of fitness. *British Journal for the Philosophy of Science* 55, 347–363. (Cited on page 19.)
- Arino, O., M. Kimmel, and G. F. Webb (1995). Mathematical modeling of the loss of telomere sequences. *Journal of Theoretical Biology* 177, 45–57. (Cited on page 82.)
- Armanios, M. and E. H. Blackburn (2012). The telomere syndromes. *Nature Reviews Genetics* 13, 693–704. (Cited on pages 12 and 54.)
- Armitage, P. and R. Doll (1954). The age distribution of cancer and a multi-stage theory of carcinogenesis. *British Journal of Cancer* 8, 1–12. (Cited on page 7.)
- Balabanov, S., A. Gontarewicz, G. Keller, L. Raddrizzani, M. Braig, R. Bosotti, J. Moll, E. Jost, C. Barrett, I. Rohe, C. Bokemeyer, T. L. Holyoake, and T. H. Brummendorf (2011). Abcg2 overexpression represents a novel mechanism for acquired resistance to the multi-kinase inhibitor danusertib in bcr-abl-positive cells in vitro. *PLoS ONE* 6, e19164. (Cited on page 68.)
- Balabanov, S., A. Gontarewicz, P. Ziegler, U. Hartmann, W. Kammer, M. Copland, U. Brassat, M. Priemer, I. Hauber, T. Wilhelm, G. Schwarz, L. Kanz, C. Bokemeyer, J. Hauber, T. L. Holyoake, A. Nordheim, and T. H. Brummendorf (2007).

- Hypusination of eukaryotic initiation factor 5a (eif5a): a novel therapeutic target in bcr-abl-positive leukemias identified by a proteomics approach. *Blood* 109, 1701–1711. (Cited on page 68.)
- Ball, S. E., F. M. Gibson, S. Rizzo, J. A. Tooze, J. C. Marsh, and E. C. Gordon-Smith (1998). Progressive telomere shortening in aplastic anemia. *Blood* 15, 3582–3592. (Cited on pages 12 and 82.)
- Bartram, C. R., A. De Klein, A. Hagemeijer, T. Van Agthoven, A. G. van Kessel, D. Bootsma, G. Grosveld, M. A. Ferguson-Smith, T. Davies, M. Stone, et al. (1983). Translocation of c-abl oncogene correlates with the presence of a philadelphia chromosome in chronic myelocytic leukaemia. *Nature* 306, 277–280. (Cited on pages 8 and 10.)
- Basanta, D. and A. Deutsch (2008). *A game theoretical perspective on the somatic evolution of cancer*. Springer. (Cited on page 118.)
- Basanta, D., H. Hatzikirou, and A. Deutsch (2008). Studying the emergence of invasiveness in tumours using game theory. *The European Physical Journal B-Condensed Matter and Complex Systems* 63, 393–397. (Cited on page 118.)
- Basanta, D., J. G. Scott, R. Rockne, K. R. Swanson, and A. R. Anderson (2011). The role of idh1 mutated tumour cells in secondary glioblastomas: an evolutionary game theoretical view. *Physical Biology* 8, 015016. (Cited on pages 47 and 118.)
- Beerenwinkel, N., T. Antal, D. Dingli, A. Traulsen, K. W. Kinzler, V. E. Velculescu, B. Vogelstein, and M. A. Nowak (2007). Genetic progression and the waiting time to cancer. *PLoS Computational Biology* 3, e225. (Cited on page 46.)
- Beillard, E., N. Pallisgaard, V. H. J. van der Velden, W. Bi, R. Dee, E. van der Schoot, E. Delabesse, E. Macintyre, E. Gottardi, G. Saglio, F. Watzinger, T. Lion, J. J. M. van Dongen, P. Hokland, and J. Gabert (2003). Evaluation of candidate control genes for diagnosis and residual disease detection in leukemic patients using 'real-time' quantitative reverse-transcriptase polymerase chain reaction (rq-pcr) - a europe against cancer program. *Leukemia* 17(12), 2474–2486. (Cited on page 70.)

- Berardi, A., A. Wang, J. Levine, P. Lopez, D. Scadden, et al. (1995). Functional isolation and characterization of human hematopoietic stem cells. *Science* 267, 104–108. (Cited on pages 4 and 5.)
- Bernoulli, J. (1695). Explicationes, annotationes, additiones ad ea, quae in actis sup. anni de curva elastica, isochrona paracentrica, velaria, hinc inde memorata, paratim controversa legundur; ubi de linea mediarum directionum, alliisque novis. *Acta Eruditorum* 1, 65–80. (Cited on pages 64 and 73.)
- Bixby, D. and M. Talpaz (2009). Mechanisms of resistance to tyrosine kinase inhibitors in chronic myeloid leukemia and recent therapeutic strategies to overcome resistance. *Hematology American Society of Hematology Education Program Book 1*, 461–476. (Cited on page 66.)
- Blackburn, E. H. et al. (1991). Structure and function of telomeres. *Nature* 350, 569–573. (Cited on pages 11 and 81.)
- Blagosklonny, M. V. (2002). Sti-571 must select for drug-resistant cells but 'no cell breathes fire out of its nostrils like a dragon'. *Leukemia* 16, 570–572. (Cited on page 78.)
- Blasco, M. A. (2007). Telomere length, stem cells and aging. *Nature Chemical Biology* 3, 640–649. (Cited on page 11.)
- Bose, S., M. Deininger, J. Gora-Tybor, J. M. Goldman, and J. V. Melo (1998). The presence of typical and atypical bcr-abl fusion genes in leukocytes of normal individuals: biologic significance and implications for the assessment of minimal residual disease. *Blood* 92, 3362–3367. (Cited on page 43.)
- Boukamp, P., S. Popp, and D. Krunic (2005). Telomere-dependent chromosomal instability. *Journal of Investigative Dermatology Symposium Proceedings* 10, 89–94. (Cited on pages 11 and 81.)
- Bozic, I., T. Antal, H. Ohtsuki, H. Carter, D. Kim, S. Chen, R. Karchin, K. W. Kinzler, B. Vogelstein, and M. A. Nowak (2010). Accumulation of driver and passenger mutations during tumor progression. *Proceedings of the National Academy of Sciences* 107, 18545–18550. (Cited on pages 8 and 47.)
- Brockhurst, M. A., A. D. Morgan, P. B. Rainy, and A. Buckling (2003). Populating mixing accelerates coevolution. *Ecology Letters* 6, 975–979. (Cited on page 102.)

- Browder, F. E. (1979). Fixed point theory and nonlinear problems. *Nonlinear and Global Analysis 1*, 31. (Cited on page 25.)
- Brown, R. L. W. (1983). Evolutionary game dynamics in diploid populations. *Theoretical Population Biology 24*, 313–322. (Cited on page 115.)
- Brümmendorf, T. H. and S. Balabanov (2006). Telomere length dynamics in normal hematopoiesis and in disease states characterized by increased stem cell turnover. *Leukemia 20*, 1706–1716. (Cited on pages 11 and 81.)
- Brümmendorf, T. H., W. Dragowska, J. Zijlmans, G. Thornbury, and P. M. Lansdorp (1998). Asymmetric cell divisions sustain long-term hematopoiesis from single-sorted human fetal liver cells. *Journal of Experimental Medicine 188*, 1117–1124. (Cited on pages 4 and 33.)
- Brümmendorf, T. H., J. P. Maciejewski, J. Mak, N. S. Young, and P. M. Lansdorp (2001). Telomere length in leukocyte subpopulations of patients with aplastic anemia. *Blood 97*, 895–900. (Cited on pages 12 and 82.)
- Brümmendorf, T. H., N. Rufer, G. M. Baerlocher, E. Roosnek, and P. M. Lansdorp (2001). Limited telomere shortening in hematopoietic stem cells after transplantation. *Annals of the New York Academy of Sciences 938*, 1–7. (Cited on page 82.)
- Brümmendorf, T. H., N. Rufer, T. L. Holyoake, J. Maciejewski, M. J. Barnett, C. J. Eaves, A. C. Eaves, N. Young, and P. M. Lansdorp (2006). Telomere length dynamics in normal individuals and in patients with hematopoietic stem cell-associated disorders. *Annals of the New York Academy of Sciences 938*, 293–304. (Cited on page 11.)
- Buescher, E. S., D. W. Alling, and J. I. Gallin (1985). Use of an x-linked human neutrophil marker to estimate timing of lyonization and size of the dividing stem cell pool. *The Journal of Clinical Investigation 76*, 1581–1584. (Cited on pages 6 and 49.)
- Bunting, K. (2002). Abc transporters as phenotypic markers and functional regulators of stem cells. *Stem Cells 20*, 11–20. (Cited on page 5.)

- Burchert, A., Y. Wang, D. CAI, N. Bubnoff, P. Paschka, S. Müller-Brüsselbach, O. G. Ottmann, J. Dyster, A. Hochhaus, and A. Neubauer (2005). Compensatory pi3-kinase/akt/mtor activation regulates imatinib resistance development. *Leukemia* 19, 1774–1782. (Cited on pages 68 and 79.)
- Chan, S. R. and E. H. Blackburn (2004). Telomeres and telomerase. *Philosophical Transactions of the Royal Society B* 359, 109–121. (Cited on page 81.)
- Chomel, J. C. and A. G. Turhan (2011). Chronic myeloid leukemia stem cells in the era of targeted therapies: resistance, persistence and long-term dormancy. *Oncotarget* 9, 713–727. (Cited on page 80.)
- Chou, T. C. and P. Talalay (1981). Generalized equations for the analysis of inhibitions of michaelis-menten and higher-order kinetic systems with two or more mutually exclusive and nonexclusive inhibitors. *European Journal of Biochemistry* 115, 207–216. (Cited on page 68.)
- Cortes, J., H. Kantarjian, T. Brümmendorf, D. Kim, A. Turkina, Z. Shen, R. Pasquini, K. H.J., S. Arkin, A. Volkert, N. Besson, R. Abbas, J. Wang, E. Leip, and C. Gambacorti-Passerini (2011). Safety and efficacy of bosutinib (ski-606) in chronic phase philadelphia chromosome-positive cml patients with resistance or intolerance to imatinib. *Blood* 118, 4567–4576. (Cited on pages 10 and 66.)
- Cowperthwaite, M. C., J. J. Bull, and L. A. Meyers (2005). Disributions of beneficial fitness effects in RNA. *Genetics* 170, 1449–1457. (Cited on page 102.)
- Cressman, R. (1992). *The stability concept of evolutionary game theory*, Volume 94. Springer. (Cited on page 109.)
- Cressman, R. (2003). *Evolutionary Dynamics and Extensive Form Games*. MIT Press, Cambridge. (Cited on page 102.)
- d’Adda di Fagagna, F., M. P. Hande, W. M. Tong, D. Roth, P. M. Lansdorp, Z. Q. Wang, and S. P. Jackson (2001). Effects of dna nonhomologous end-joining factors on telomere length and chromosomal stability in mammalian cells. *Current Biology* 11, 1192–1196. (Cited on pages 11 and 81.)
- Daley, G. Q., R. A. Van Etten, and D. Baltimore (1990). Induction of chronic myelogenous leukemia in mice by the p210bcr/abl gene of the philadelphia chromosome. *Science* 247, 824–830. (Cited on pages 43 and 46.)

- Darwin, C. (1859). *On the origins of species by means of natural selection*, Volume 1. John Murray. (Cited on pages 1 and 99.)
- Deininger, M., J. M. Goldman, and J. V. Melo (2000). The molecular biology of chronic myeloid leukemia. *Blood* 96, 3343–3356. (Cited on page 66.)
- Demidenko, Z. N., W. G. An, J. T. Lee, L. Y. Romanova, J. A. McCubrey, and M. V. Blagosklonny (2005). Kinase-addiction and bi-phasic sensitivity-resistance of bcr-abl- and raf-1-expressing cells to imatinib and geldanamycin. *Cancer Biology & Therapy* 4(4), 484–490. (Cited on page 67.)
- Diaz Jr, L. A., R. T. Williams, J. Wu, I. Kinde, J. R. Hecht, J. Berlin, B. Allen, I. Bozic, J. G. Reiter, M. A. Nowak, et al. (2012). The molecular evolution of acquired resistance to targeted egfr blockade in colorectal cancers. *Nature* 486, 537–540. (Cited on page 9.)
- Ding, L., T. J. Ley, D. E. Larson, C. A. Miller, D. C. Koboldt, J. S. Welch, J. K. Ritchey, M. A. Young, T. Lamprecht, M. D. McLellan, et al. (2012). Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. *Nature* 481, 506–510. (Cited on pages 7, 8, 9 and 47.)
- Ding, L. and S. J. Morrison (2013). Haematopoietic stem cells and early lymphoid progenitors occupy distinct bone marrow niches. *Nature* 495, 231–236. (Cited on page 5.)
- Dingli, D., L. Luzzatto, and J. M. Pacheco (2008). Neutral evolution in paroxysmal nocturnal hemoglobinuria. *Proceedings of the National Academy of Sciences USA* 105, 18496–18500. (Cited on pages 40 and 42.)
- Dingli, D. and J. M. Pacheco (2006). Allometric scaling of the active hematopoietic stem cell pool across mammals. *PLoS One* 1, e2. (Cited on pages 6, 49 and 80.)
- Dingli, D. and J. M. Pacheco (2007). Ontogenic growth of the haematopoietic stem cell pool in humans. *Proceedings of the Royal Society B* 274, 2497. (Cited on page 59.)
- Dingli, D., J. M. Pacheco, and A. Traulsen (2008). Multiple mutant clones in blood rarely coexist. *Physical Review E* 77, 021915. (Cited on pages 33, 37, 38 and 80.)

- Dingli, D., A. Traulsen, and F. Michor (2007). (A)symmetric stem cell replication and cancer. *PLoS Computational Biology* 3, e53. (Cited on page 28.)
- Dingli, D., A. Traulsen, and J. M. Pacheco (2007a). Compartmental architecture and dynamics of hematopoiesis. *PLoS One* 4, e345. (Cited on pages 4, 28, 29, 30, 33, 35, 37, 43, 46, 47, 49, 52 and 82.)
- Dingli, D., A. Traulsen, and J. M. Pacheco (2007b). Stochastic dynamics of hematopoietic tumor stem cells. *Cell Cycle* 6, e2–e6. (Cited on pages 34, 39, 40 and 51.)
- Dingli, D., A. Traulsen, and J. M. Pacheco (2008a). Chronic myeloid leukemia: origin, development, response to therapy, and relapse. *Clinical Leukemia* 2, 133–139. (Cited on pages 10 and 57.)
- Dingli, D., A. Traulsen, and J. M. Pacheco (2008b). Dynamics of haemopoiesis across mammals. *Proceedings of the Royal Society B* 275, 2389–2392. (Cited on page 40.)
- Drake, J. W., B. Charlesworth, D. Charlesworth, and J. F. Crow (1998). Rates of spontaneous mutation. *Genetics* 148, 1667–1686. (Cited on page 79.)
- Druker, B. J., F. Guilhot, S. G. O’Brien, I. Gathmann, H. Kantarjian, N. Gattermann, M. W. Deininger, R. T. Silver, J. M. Goldman, R. M. Stone, F. Cervantes, A. Hochhaus, B. L. Powell, J. L. Gabilove, P. Rousselot, J. Reiffers, J. J. Cornelissen, T. Hughes, H. Agis, T. Fischer, G. Verhoef, J. Shepherd, G. Saglio, A. Gratwohl, J. L. Nielsen, J. P. Radich, B. Simonsson, K. Taylor, M. Baccarani, C. So, L. Letvak, R. A. Larson, and IRIS Investigators (2006). Five-year follow-up of patients receiving imatinib for chronic myeloid leukemia. *New England Journal of Medicine* 355, 2408–2417. (Cited on pages 10, 60 and 66.)
- Dunn, D., J. Yu, S. Nagarajan, M. Devetten, F. Weichold, M. Medof, N. Young, and J. Liu (1996). A knock-out model of paroxysmal nocturnal hemoglobinuria: Pig-a (-) hematopoiesis is reconstituted following intercellular transfer of gpi-anchored proteins. *Proceedings of the National Academy of Sciences* 93, 7938–7943. (Cited on page 40.)

- England, S. J., K. E. McGrath, J. M. Frame, and J. Palis (2011). Immature erythroblasts with extensive ex vivo self-renewal capacity emerge from the early mammalian fetus. *Blood* 117, 2708–2717. (Cited on page 40.)
- Epel, E. S., E. H. Blackburn, J. Lin, F. S. Dhabhar, N. E. Adler, J. D. Morrow, and R. M. Cawthon (2004). Accelerated telomere shortening in response to life stress. *Proceedings of the National Academy of Sciences of the United States of America* 101, 17312–17315. (Cited on page 11.)
- Epstein, F. H., R. Kurzrock, J. U. Gutterman, and M. Talpaz (1988). The molecular genetics of philadelphia chromosome–positive leukemias. *New England Journal of Medicine* 319, 990–998. (Cited on pages 8 and 10.)
- Eshel, I. (1982). Evolutionarily stable strategies and viability selection in Mendelian populations. *Theoretical Population Biology* 22, 204–217. (Cited on page 115.)
- Eyre-Walker, A. (2007). The distribution of fitness effects of new mutations. *Genetics* 8, 610–618. (Cited on pages 102 and 118.)
- Eyre-Walker, A., A. Woolfit, and T. Phelps (2006). The distribution of fitness effects of new deleterious amino acid mutations in humans. *Genetics* 173, 891–900. (Cited on page 102.)
- Fischer, M., M. Schwieger, S. Horn, B. Niebuhr, A. Ford, S. Roscher, U. Bergholz, M. Greaves, J. Löhler, and C. Stocking (2005). Defining the oncogenic function of the tel/amll1 (etv6/runx1) fusion protein in a mouse model. *Oncogene* 24, 7579–7591. (Cited on page 58.)
- Fisher, R. A. (1930). *The Genetical Theory of Natural Selection*. Clarendon Press, Oxford. (Cited on pages 1 and 117.)
- Fisher, R. A. (1941). Average excess and average effect of a gene substitution. *Annals of Eugenics* 11, 53–63. (Cited on page 117.)
- Fuchs, E. (2008). Skin stem cells: rising to the surface. *The Journal of Cell Biology* 180, 273–284. (Cited on pages 4, 28 and 46.)
- Fudenberg, D. and C. Harris (1992). Evolutionary dynamics with aggregate shocks. *Journal of Economic Theory* 57, 420–441. (Cited on page 116.)

- Gabert, J., E. Beillard, V. H. van der Velden, W. Bi, D. Grimwade, N. Pallisgaard, G. Barbany, G. Cazzaniga, J. M. Cayuela, H. Cavé, F. Pane, J. L. Aerts, D. De Micheli, X. Thirion, V. Pradel, M. González, S. Viehmann, M. Malec, G. Saglio, J. van Dongen, J J Gabert, E. Beillard, V. H. van der Velden, W. Bi, D. Grimwade, N. Pallisgaard, G. Barbany, G. Cazzaniga, J. M. Cayuela, H. Cavé, F. Pane, J. L. Aerts, D. De Micheli, X. Thirion, V. Pradel, M. González, S. Viehmann, M. Malec, G. Saglio, and J. J. van Dongen (2003). Standardization and quality control studies of 'real-time' quantitative reverse transcriptase polymerase chain reaction of fusion gene transcripts for residual disease detection in leukemia - a europe against cancer program. *Leukemia* 17, 2318–2357. (Cited on page 70.)
- Gerlee, P., D. Basanta, and A. R. Anderson (2011). Evolving homeostatic tissue using genetic algorithms. *Prog Biophys Mol Biol* 106, 414–425. (Cited on page 47.)
- Gerrits, A., B. Dykstra, O. J. Kalmykowa, K. Klauke, E. Verovskaya, M. J. Broekhuis, G. de Haan, and L. V. Bystrykh (2010). Cellular barcoding tool for clonal analysis in the hematopoietic system. *Blood* 115, 2610–2618. (Cited on page 51.)
- Gerstung, M. and N. Beerenwinkel (2010). Waiting time models of cancer progression. *Mathematical Population Studies* 17, 115–135. (Cited on page 46.)
- Gillespie, D. (1976). A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *Journal of Computational Physics* 22, 403–434. (Cited on pages 13, 31, 49 and 84.)
- Gillespie, D. (1992). A rigorous derivation of the chemical master equation. *Physica A: Statistical Mechanics and its Applications* 188, 404–425. (Cited on pages 13 and 22.)
- Gillespie, D. (2007). Stochastic simulation of chemical kinetics. *Annual Reviews of Physical Chemistry* 58, 35–55. (Cited on pages 13, 15 and 49.)
- Gillespie, J. H. (1983). Some properties of finite populations experiencing strong selection and weak mutation. *The American Naturalist* 121, 691–708. (Cited on page 102.)

- Glauche, I., K. Horn, M. Horn, L. Thielecke, M. A. Essers, A. Trumpp, and I. Roeder (2012). Therapy of chronic myeloid leukaemia can benefit from the activation of stem cells: simulation studies of different treatment combinations. *British Journal of Cancer* 106, 1742–1752. (Cited on page 9.)
- Glauche, I., L. Thielecke, and I. Roeder (2011). Cellular aging leads to functional heterogeneity of hematopoietic stem cells: a modeling perspective. *Aging Cell* 10, 457–465. (Cited on page 47.)
- Goel, N. and N. Richter-Dyn (1974). *Stochastic Models in Biology*. Academic Press, New York. (Cited on page 12.)
- Gokhale, C. S. and A. Traulsen (2010). Evolutionary games in the multiverse. *Proceedings of the National Academy of Sciences USA* 107, 5500–5504. (Cited on pages 20 and 24.)
- Greaves, M. and C. C. Maley (2012). Clonal evolution in cancer. *Nature* 481, 306–313. (Cited on page 10.)
- Greenbaum, A., Y.-M. S. Hsu, R. B. Day, L. G. Schuettelpelz, M. J. Christopher, J. N. Borgerding, T. Nagasawa, and D. C. Link (2013). Cxcl12 in early mesenchymal progenitors is required for haematopoietic stem-cell maintenance. *Nature* 495, 227–231. (Cited on page 5.)
- Greenman, C., P. Stephens, R. Smith, G. L. Dalglish, C. Hunter, G. Bignell, H. Davies, J. Teague, A. Butler, C. Stevens, et al. (2007). Patterns of somatic mutation in human cancer genomes. *Nature* 446, 153–158. (Cited on pages 8 and 47.)
- Griffith, J. D., L. Comeau, S. Rosenfield, R. M. Stansel, A. Bianchi, H. Moss, and T. de Lange (1999). Mammalian telomeres end in a large duplex loop. *Cell* 97, 503–514. (Cited on page 81.)
- Griswold, I. J., M. MacPartlin, T. Bumm, V. L. Goss, T. O’Hare, K. A. Lee, A. S. Corbin, E. P. Stoffregen, C. Smith, K. Johnson, E. M. Moseson, L. J. Wood, R. D. Polakiewicz, B. J. Druker, and M. W. Deininger (2006). Kinase domain mutants of bcr-abl exhibit altered transformation potency, kinase activity, and substrate utilization, irrespective of sensitivity to imatinib. *Molecular and Cellular Biology* 26, 6082–6093. (Cited on page 80.)

- Grove, J., E. Bruscia, and D. Krause (2004). Plasticity of bone marrow-derived stem cells. *Stem Cells* 22, 487–500. (Cited on pages 4 and 5.)
- Guibal, F. C., M. Alberich-Jorda, H. Hirai, A. Ebralidze, E. Levantini, A. Di Ruscio, P. Zhang, B. A. Santana-Lemos, D. Neuberg, A. J. Wagers, E. M. Rego, and D. G. Tenen (2009). Identification of a myeloid committed progenitor as the cancer-initiating cell in acute promyelocytic leukemia. *Blood* 114, 5415–5425. (Cited on pages 44, 46, 52 and 60.)
- Haber, D. A. and J. Settleman (2007). Cancer: drivers and passengers. *Nature* 446, 145–146. (Cited on page 47.)
- Haldane, J. B. S. (1932). *The Causes of Evolution*. Longmans, Green and Co., Ltd, New York. (Cited on page 1.)
- Hamilton, W. (1967). Extraordinary sex ratios. *Science* 156, 477–488. (Cited on page 1.)
- Hamilton, W. D. (1964). The genetical evolution of social behavior I and II. *Journal of Theoretical Biology* 7, 1–16 + 17–52. (Cited on pages 1 and 21.)
- Han, T. A., A. Traulsen, and C. S. Gokhale (2012). On equilibrium properties of evolutionary multi-player games with random payoff matrices. *Theoretical Population Biology* 81, 264–72. (Cited on page 24.)
- Hanahan, D. and R. A. Weinberg (2000). The hallmarks of cancer. *Cell* 100, 57–70. (Cited on pages 2, 6 and 47.)
- Hande, M. P., E. Samper, P. Lansdorp, and M. A. Blasco (1999). Telomere length dynamics and chromosomal instability in cells derived from telomerase null mice. *The Journal of Cell Biology* 144, 589–601. (Cited on pages 11 and 81.)
- Hanfstein, B., M. C. Müller, S. Kreil, T. Ernst, T. Schenk, C. Lorentz, U. Schwindel, A. Leitner, R. Hehlmann, and A. Hochhaus (2011). Dynamics of mutant bcr-abl-positive clones after cessation of tyrosine kinase inhibitor therapy. *Haematologica* 96, 360–366. (Cited on page 80.)
- Hao, Q., A. Shah, F. Thiemann, E. Smogorzewska, and G. Crooks (1995). A functional comparison of cd34+ cd38-cells in cord blood and bone marrow. *Blood* 86, 3745–3753. (Cited on page 5.)

- Harley, C. B., A. B. Futcher, and C. W. Greider (1990). Telomeres shorten during ageing of human fibroblasts. *Nature* 345, 458–460. (Cited on pages 12 and 81.)
- Hauert, C. (1999). *The Evolution of Cooperation: The Prisoner's Dilemma and its Applications as an Example*. Aachen: Shaker. PhD thesis. (Cited on page 21.)
- Hauert, C. and M. Doebeli (2004). Spatial structure often inhibits the evolution of cooperation in the snowdrift game. *Nature* 428, 643–646. (Cited on page 21.)
- Hayflick, L., P. Moorhead, et al. (1961). The serial cultivation of human diploid cell strains. *Experimental Cell Research* 25, 585–621. (Cited on pages 4, 5, 46, 54 and 84.)
- Hicks, W. M., M. Kim, and J. E. Haber (2010). Increased mutagenesis and unique mutation signature associated with mitotic gene conversion. *Science* 329, 82–85. (Cited on page 79.)
- Hillmen, P., M. Bessler, P. J. Mason, W. M. Watkins, and L. Luzzatto (1993). Specific defect in n-acetylglucosamine incorporation in the biosynthesis of the glycosylphosphatidylinositol anchor in cloned cell lines from patients with paroxysmal nocturnal hemoglobinuria. *Proceedings of the National Academy of Sciences* 90, 5272–5276. (Cited on page 40.)
- Hines, W. G. (1980). An evolutionarily stable strategy model for randomly mating diploid populations. *Journal of Theoretical Biology* 87, 379–384. (Cited on page 115.)
- Hochhaus, A., S. Kreil, A. Corbin, P. La Rosee, M. Muller, T. Lahaye, B. Hanfstein, C. Schoch, N. Cross, U. Berger, et al. (2002). Molecular and chromosomal mechanisms of resistance to imatinib (sti571) therapy. *Leukemia* 16, 2190–2196. (Cited on page 10.)
- Hochhaus, A., P. La Rosée, M. C. Müller, T. Ernst, and N. C. Cross (2011). Impact of bcr-abl mutations on patients with chronic myeloid leukemia. *Cell Cycle* 10, 250–260. (Cited on page 67.)
- Hofbauer, J., P. Schuster, and K. Sigmund (1982). Game dynamics in mendelian populations. *Biological Cybernetics* 43, 51–57. (Cited on pages 105 and 115.)

- Hofbauer, J. and K. Sigmund (1998). *Evolutionary Games and Population Dynamics*. Cambridge University Press, Cambridge. (Cited on pages 102 and 109.)
- Hong, D., R. Gupta, P. Ancliff, A. Atzberger, J. Brown, S. Soneji, J. Green, S. Colman, W. Piacibello, V. Buckle, et al. (2008). Initiating and cancer-propagating cells in tel-aml1-associated childhood leukemia. *Science* 319, 336–339. (Cited on page 58.)
- Hornsby, C., K. M. Page, and I. P. Tomlinson (2007). What can we learn from the population incidence of cancer? armitage and doll revisited. *The Lancet Oncology* 8, 1030–1038. (Cited on page 7.)
- Huang, W., B. Haubold, C. Hauert, and A. Traulsen (2012). Emergence of stable polymorphism driven by evolutionary games between mutants. *Nature Communications* 3, 919. (Cited on pages 24, 99, 111 and 116.)
- Huang, W. and A. Traulsen (2010). Fixation probabilities of random mutants under frequency dependent selection. *Journal of Theoretical Biology* 263, 262–268. (Cited on pages 99 and 103.)
- Huang, W., B. Werner, and A. Traulsen (2012). The impact of random frequency-dependent mutations on the average population fitness. *BMC Evolutionary Biology* 12, 160. (Cited on pages 99, 123 and 158.)
- Huntly, B. J., H. Shigematsu, K. Deguchi, B. H. Lee, S. Mizuno, N. Duclos, R. Rowan, S. Amaral, D. Curley, I. R. Williams, K. Akashi, and D. G. Gilliland (2004). Moz-tif2, but not bcr-abl, confers properties of leukemic stem cells to committed murine hematopoietic progenitors. *Cancer Cell* 6, 587–596. (Cited on page 44.)
- Jackson, J. D. and R. F. Fox (1999). Classical electrodynamics. *American Journal of Physics* 67, 841. (Cited on page 1.)
- Jackson, T. L. (2003). Intracellular accumulation and mechanism of action of doxorubicin in a spatio-temporal tumor model. *Journal of theoretical biology* 220, 201–213. (Cited on page 29.)
- Jiang, X., K. M. Saw, A. Eaves, and C. Eaves (2007). Instability of bcr-abl gene in primary and cultured chronic myeloid leukemia stem cells. *Journal of The National Cancer Institute* 99, 680–693. (Cited on page 80.)

- Jiang, Y., B. Jahagirdar, R. Reinhardt, R. Schwartz, C. Keene, X. Ortiz-Gonzalez, M. Reyes, T. Lenvik, T. Lund, M. Blackstad, et al. (2002). Pluripotency of mesenchymal stem cells derived from adult marrow. *Nature* 418, 41–49. (Cited on page 5.)
- Johnson, D. D. P., P. Stopka, and J. Bell (2002). Individual variation evades the prisoner’s dilemma. *BMC Evolutionary Biology* 2, 15. (Cited on page 116.)
- Johnston, M. D., C. M. Edwards, W. F. Bodmer, P. K. Maini, and S. J. Chapman (2007). Mathematical modeling of cell population dynamics in the colonic crypt and in colorectal cancer. *Proceedings of the National Academy of Sciences* 104, 4008–4013. (Cited on pages 28, 29 and 35.)
- Jones, S., W.-D. Chen, G. Parmigiani, F. Diehl, N. Beerenwinkel, T. Antal, A. Traulsen, M. A. Nowak, C. Siegel, V. Velculescu, K. W. Kinzler, B. Vogelstein, J. Willis, and S. Markowitz (2008). Comparative lesion sequencing provides insights into tumor evolution. *Proceedings of the National Academy of Sciences USA* 105, 4283–4288. (Cited on page 47.)
- Jordan, C. T., M. L. Guzman, and M. Noble (2006). Cancer stem cells. *New England Journal of Medicine* 355, 1253–1261. (Cited on page 9.)
- Kampen, N. G. v. (1997). *Stochastic Processes in Physics and Chemistry* (2 ed.). Amsterdam: Elsevier. (Cited on pages 22, 23 and 30.)
- Kantarjian, H., N. P. Shah, A. Hochhaus, J. Cortes, S. Shah, M. Ayala, B. Moiraghi, Z. Shen, J. Mayer, R. Pasquini, H. Nakamae, F. Huguet, C. Boqué, C. Chuah, E. Bleickardt, B. Bradley-Garelik, C. Zhu, T. Szatrowski, D. Shapiro, and M. Bacarani (2010). Dasatinib versus imatinib in newly diagnosed chronic-phase chronic myeloid leukemia. *New England Journal of Medicine* 326, 2260–2270. (Cited on page 66.)
- Kantarjian, H. M., F. Giles, N. Gattermann, K. Bhalla, G. Alimena, F. Palandri, G. J. Ossenkoppele, F.-E. Nicolini, S. G. O’Brien, M. Litzow, et al. (2007). Nilotinib (formerly amn107), a highly selective bcr-abl tyrosine kinase inhibitor, is effective in patients with philadelphia chromosome–positive chronic myelogenous leukemia in chronic phase following imatinib resistance and intolerance. *Blood* 110, 3540–3546. (Cited on page 10.)

- Katouli, A. A. and N. L. Komarova (2010). Optimizing combination therapies with existing and future cml drugs. *PLoS One* 5, e12300. (Cited on pages 67 and 80.)
- Kelly, P. N., A. Dakic, J. M. Adams, S. L. Nutt, and A. Strasser (2007). Tumor growth need not be driven by rare cancer stem cells. *Science* 317, 337–337. (Cited on page 9.)
- Kemeny, J. G. and J. L. Snell (1960). *Finite Markov Chains*. Princeton, NJ: Van Nostrand. (Cited on page 16.)
- Kimura, M. (1983). *The Neutral Theory of Molecular Evolution*. Cambridge, U.K.: Cambridge University Press. (Cited on pages 17 and 18.)
- Kimura, M. (1991). Recent development of the neutral theory viewed from the wrightian tradition of theoretical population genetics. *Proceedings of the National Academy of Sciences* 88, 5969–5973. (Cited on page 18.)
- King, K. Y. and M. A. Goodell (2011). Direct conversion of skin into blood: Alchemy or science? *Molecular Therapy* 19, 277–228. (Cited on page 4.)
- Klawitter, J., D. J. Kominsky, J. L. Brown, J. Klawitter, U. Christians, D. Leibfritz, J. V. Melo, S. G. Eckhardt, and N. J. Serkova (2009). Metabolic characteristics of imatinib resistance in chronic myeloid leukaemia cells. *British Journal of Pharmacology* 158, 588–600. (Cited on page 79.)
- Knudson, A. G. (1971). Mutation and cancer: statistical study of retinoblastoma. *Proceedings of the National Academy of Sciences* 68, 820–823. (Cited on pages 7 and 8.)
- Komarova, N. L. (2011). Mathematical modeling of cyclic treatments of chronic myeloid leukemia. *Mathematical Biosciences and Engineering* 8, 289–306. (Cited on page 80.)
- Komarova, N. L. and P. Cheng (2006). Epithelial tissue architecture protects against cancer. *Mathematical Biosciences* 200, 90–117. (Cited on page 28.)
- Komarova, N. L., A. A. Katouli, and D. Wodarz (2009). Combination of two but not three current targeted drugs can improve therapy of chronic myeloid leukemia. *PLoS One* 4, e4423. (Cited on pages 67 and 80.)

- Komarova, N. L., A. Sengupta, and M. A. Nowak (2003). Mutation–selection networks of cancer initiation: tumor suppressor genes and chromosomal instability. *Journal of Theoretical Biology* 223, 433–450. (Cited on page 8.)
- Komarova, N. L. and L. Wang (2004). Initiation of colorectal cancer: where do the two hits hit? *Cell Cycle* 3, 1558–1565. (Cited on page 47.)
- Komarova, N. L. and D. Wodarz (2005). Drug resistance in cancer: principles of emergence and prevention. *Proceedings of the National Academy of Sciences* 102, 9714–9719. (Cited on page 10.)
- Komarova, N. L. and D. Wodarz (2009). Combination therapies against chronic myeloid leukemia: short-term versus long-term strategies. *Cancer Research* 69, 4904–4910. (Cited on pages 67 and 80.)
- Kominsky, D. J., J. Klawitter, J. L. Brown, L. G. Boros, J. V. Melo, S. G. Eckhardt, and N. J. Serkova (2009). Abnormalities in glucose uptake and metabolism in imatinib-resistant human bcr-abl-positive cells. *Clinical Cancer Research* 15, 3442–3450. (Cited on page 79.)
- Krivtsov, A., D. Twomey, Z. Feng, M. Stubbs, Y. Wang, J. Faber, J. Levine, J. Wang, W. Hahn, D. G. Gilliland, T. R. Golub, and S. A. Armstrong (2006). Transformation from committed progenitor to leukaemia stem cell initiated by MLL-AF9. *Nature* 442(7104), 818–822. (Cited on page 44.)
- Lálic, J., J. Cuevas, and S. F. Elena (2011). Effects of host species on the distribution of mutational fitness effects for an rna virus. *PLoS Genetics* 11, e1002378. (Cited on page 102.)
- Lange, T., B. Park, S. G. Willis, and M. W. Deininger (2005). Bcr-abl kinase domain mutations in chronic myeloid leukemia: not quite enough to cause resistance to imatinib therapy? *Cell Cycle* 4, 1761–1766. (Cited on page 79.)
- Lavker, R. and T. Sun (2000). Epidermal stem cells: properties, markers, and location. *Proceedings of the National Academy of Sciences* 97, 13473–13475. (Cited on page 5.)
- Lee, W., Z. Jiang, J. Liu, P. M. Haverty, Y. Guan, J. Stinson, P. Yue, Y. Zhang, K. P. Pant, D. Bhatt, et al. (2010). The mutation spectrum revealed by paired

- genome sequences from a lung cancer patient. *Nature* 465, 473–477. (Cited on page 8.)
- Lenaerts, T., F. Castagnetti, A. Traulsen, J. M. Pacheco, G. Rosti, and D. Dingli (2011). Explaining the in vitro and in vivo differences in leukemia therapy. *Cell Cycle* 10, 1540–1544. (Cited on pages 9, 10, 47, 57 and 80.)
- Lenaerts, T., J. M. Pacheco, A. Traulsen, and D. Dingli (2010). Tyrosine kinase inhibitor therapy can cure chronic myeloid leukemia without hitting leukemic stem cells. *Haematologica* 95, 900–907. (Cited on pages 9, 10, 28, 31, 39, 44, 47, 66 and 80.)
- Levin, S. A., B. Grenfell, A. Hastings, and A. S. Perelson (1997). Mathematical and computational challenges in population biology and ecosystems science. *Science* 275, 334–343. (Cited on page 103.)
- Levy, M. Z., R. C. Allsopp, A. B. Futcher, C. W. Greider, and C. B. Harley (1992). Telomere end-replication problem and cell aging. *Journal of Molecular Biology* 225, 951–960. (Cited on pages 11, 81 and 82.)
- Lieberman, E., C. Hauert, and M. A. Nowak (2005). Evolutionary dynamics on graphs. *Nature* 433, 312–316. (Cited on page 20.)
- Lin, H. (2002). The stem-cell niche theory: lessons from flies. *Nature Reviews Genetics* 3, 931–940. (Cited on page 5.)
- Loeffler, M. and H. E. Wichmann (1980). A comprehensive mathematical model of stem cell proliferation which reproduces most of the published experimental results. *Cell and Tissue Kinetics* 13, 543–561. (Cited on page 46.)
- Longo, L., M. Bessler, P. Beris, D. Swirsky, and L. Luzzatto (2008). Myelodysplasia in a patient with pre-existing paroxysmal nocturnal haemoglobinuria: a clonal disease originating from within a clonal disease. *British Journal of Haematology* 87, 401–403. (Cited on page 40.)
- Lugo, T. G., A.-M. Pendergast, A. J. Muller, O. N. Witte, et al. (1990). Tyrosine kinase activity and transformation potency of bcr-abl oncogene products. *Science* 247, 1079–1082. (Cited on page 8.)

- Luzzatto, L., M. Bessler, and B. Rotoli (1997). Somatic mutations in paroxysmal nocturnal hemoglobinuria: a blessing in disguise? *Cell* 88, 1–4. (Cited on page 40.)
- Lynch, M., J. Blanchard, D. Houle, T. Kibota, S. Schultz, L. Vassilieva, and J. Willis (1999). Spontaneous deleterious mutation. *Evolution* 53, 645–663. (Cited on page 102.)
- Ma, Y., S. A. Dobbins, A. L. Sherborne, D. Chubb, M. Galbiati, G. Cazzaniga, M. Concetta, T. R., A. L. Lloyd, R. Hain, M. Greaves, and R. S. Houlston (2013). Developmental timing of mutations revealed by whole-genome sequencing of twins with acute lymphoblastic leukemia. *Proceedings of the National Academy of Sciences Early Edition*. (Cited on pages 58 and 59.)
- Maciejewski, J. P., E. M. Sloand, T. Sato, S. Anderson, and N. S. Young (1997). Impaired hematopoiesis in paroxysmal nocturnal hemoglobinuria/aplastic anemia is not associated with a selective proliferative defect in the glycosylphosphatidylinositol-anchored protein-deficient clone. *Blood* 89, 1173–1181. (Cited on page 40.)
- Macy, M. W. and A. Flache (2002). Learning dynamics in social dilemmas. *Proceedings of the National Academy of Sciences USA* 99, 7229–7236. (Cited on page 21.)
- Mahon, F. X., D. Ra, J. Guilhot, F. Guilhot, F. Huguet, F. Nicolini, L. Legros, A. Charbonnier, A. Guerci, B. Varet, G. Etienne, J. Reiffers, P. Rousselot, and I. F. des Leucmies Mylodes Chroniques. (2010). Discontinuation of imatinib in patients with chronic myeloid leukaemia who have maintained complete molecular remission for at least 2 years: the prospective, multicentre stop imatinib (stim) trial. *Lancet Oncol* 11, 1029–1035. (Cited on page 66.)
- Marciniak-Czochra, A., T. Stiehl, A. D. Ho, W. Jäger, and W. Wagner (2009). Modeling of asymmetric cell division in hematopoietic stem cells—regulation of self-renewal is essential for efficient repopulation. *Stem Cells Development* 18, 377–385. (Cited on pages 29, 33, 47 and 82.)
- May, R. M. (2004). Use and abuse of mathematics in biology. *Science* 303, 790–793. (Cited on page 99.)

- Maynard Smith, J. (1974). Theory of games and the evolution of animal contests. *Journal of Theoretical Biology* 47, 209–221. (Cited on page 20.)
- Maynard Smith, J. (1982). *Evolution and the Theory of Games*. Cambridge University Press, Cambridge. (Cited on pages 1, 20, 102 and 116.)
- Maynard Smith, J. (1995). *The theory of evolution* (Reprint of the 3rd ed.). Cambridge: Cambridge University Press. 60. (Cited on page 99.)
- Maynard Smith, J. and G. R. Price (1973). The logic of animal conflict. *Nature* 246, 15–18. (Cited on page 102.)
- McCulloch, J. E. and J. E. Till (2005). Perspectives on the properties of stem cells. *Nature Medicine* 11, 1026–1028. (Cited on pages 4, 28 and 46.)
- McKane, A. J. and T. J. Newman (2004). Stochastic models in population biology and their deterministic analogs. *Physical Review E* 70, 041902. (Cited on page 12.)
- Michor, F., T. P. Hughes, Y. Iwasa, S. Branford, N. P. Shah, C. L. Sawyers, and M. A. Nowak (2005). Dynamics of chronic myeloid leukaemia. *Nature* 435, 1267–1270. (Cited on pages 4, 8, 10, 28, 35, 46 and 47.)
- Michor, F., M. A. Nowak, S. A. Frank, and Y. Iwasa (2003). Stochastic elimination of cancer cells. *Proceedings of the Royal Society B* 270, 2017–2024. (Cited on pages 28 and 46.)
- Moran, P. A. P. (1958). Random processes in genetics. In *Mathematical Proceedings of the Cambridge Philosophical Society*, Volume 54, pp. 60–71. Cambridge Univ Press. (Cited on page 16.)
- Moran, P. A. P. (1962). *The Statistical Processes of Evolutionary Theory*. Clarendon Press, Oxford. (Cited on pages 15 and 71.)
- Morrison, S., N. Uchida, and I. Weissman (1995). The biology of hematopoietic stem cells. *Annual review of cell and developmental biology* 11, 35–71. (Cited on page 5.)
- Morrison, S. J. and J. Kimble (2006). Asymmetric and symmetric stem-cell divisions in development and cancer. *Nature* 441, 1068–1074. (Cited on page 82.)

- Morrison, S. J., N. M. Shah, and D. J. Anderson (1997). Regulatory mechanisms in stem cell biology. *Cell* 88, 287–298. (Cited on page 82.)
- Morrison, S. J., A. M. Wandycz, K. Akashi, A. Globerson, and I. L. Weissman (1996). The aging of hematopoietic stem cells. *Nature Medicine* 2, 1011–1016. (Cited on page 11.)
- Muller, H. J. (1964). The relation of recombination to mutational advance. *Mutation Research* 106, 2–9. (Cited on page 117.)
- Nash, J. F. (1950). Equilibrium points in n-person games. *Proceedings of the National Academy of Sciences USA* 36, 48–49. (Cited on page 21.)
- Nielsen, C., H. S. Birgens, B. G. Nordestgaard, L. Kjær, and S. E. Bojesen (2011). The jak2 v617f somatic mutation, mortality and cancer risk in the general population. *Haematologica* 96, 450–453. (Cited on page 43.)
- Nowak, M. A. (2006a). *Evolutionary Dynamics*. Harvard University Press, Cambridge. (Cited on pages 2, 4, 8, 16, 17, 19, 20, 21, 24, 25, 28, 46, 71, 73 and 102.)
- Nowak, M. A. (2006b). Five rules for the evolution of cooperation. *Science* 314, 1560–1563. (Cited on pages 21 and 108.)
- Nowak, M. A. and R. M. May (1992). Evolutionary games and spatial chaos. *Nature* 359, 826–829. (Cited on page 20.)
- Nowak, M. A., F. Michor, N. L. Komarova, and Y. Iwasa (2004). Evolutionary dynamics of tumor suppressor gene inactivation. *Proceedings of the National Academy of Sciences USA* 101, 10635–10638. (Cited on page 29.)
- Nowak, M. A. and K. Sigmund (1990). The evolution of stochastic strategies in the prisoner’s dilemma. *Acta Applicandae Mathematicae* 20, 247–265. (Cited on page 113.)
- Nowak, M. A. and K. Sigmund (2004). Evolutionary dynamics of biological games. *Science* 303, 793–799. (Cited on page 102.)
- O’Hare, T., C. A. Eide, and M. W. Deininger (2007). Bcr-abl kinase domain mutations, drug resistance, and the road to a cure for chronic myeloid leukemia. *Blood* 110, 2242–2249. (Cited on page 66.)

- O'Hare, T., W. C. Shakespeare, X. Zhu, C. A. Eide, V. M. Rivera, F. Wang, L. T. Adrian, T. Zhou, W. S. Huang, Q. Xu, C. A. Metcalf, J. W. Tyner, M. M. Loriaux, A. S. Corbin, S. Wardwell, Y. Ning, J. A. Keats, Y. Wang, R. Sundaramoorthi, M. Thomas, D. Zhou, J. Snodgrass, L. Commodore, T. K. Sawyer, D. C. Dalgarno, M. W. Deininger, B. J. Druker, and T. Clackson (2009). Ap24534, a pan-bcr-abl inhibitor for chronic myeloid leukemia, potently inhibits the t315i mutant and overcomes mutation-based resistance. *Cancer Cell* 16, 401–412. (Cited on page 67.)
- Olofsson, P. and M. Kimmel (1999). Stochastic models of telomere shortening. *Mathematical Biosciences* 158, 75–92. (Cited on page 82.)
- Olovnikov, A. (1996). Telomeres, telomerase, and aging: origin of the theory. *Experimental Gerontology* 31, 443–448. (Cited on page 5.)
- Orr, H. A. (2003). The distribution of fitness effects among beneficial mutations. *Genetics* 163, 1519–1526. (Cited on page 102.)
- Orr, H. A. (2006). The distribution of fitness effects among beneficial mutations in fisher's geometric model of adaptation. *Journal of Theoretical Biology* 238, 279–285. (Cited on page 102.)
- Page, K. M. and M. A. Nowak (2002a). Unifying evolutionary dynamics. *Journal of Theoretical Biology* 219, 93–98. (Cited on pages 23 and 26.)
- Page, K. M. and M. A. Nowak (2002b). Unifying evolutionary dynamics. *Journal of Theoretical Biology* 219, 93–98. (Cited on page 73.)
- Pardee, A. B. (1989). G1 events and regulation of cell proliferation. *Science* 246, 603–608. (Cited on page 46.)
- Patterson, M. K. (1979). Measurement of growth and viability of cells in culture. *Methods Enzymol* 58, 141–152. (Cited on page 68.)
- Pene-Dumitrescu, T. and T. E. Smithgall (2010). Expression of a src family kinase in chronic myelogenous leukemia cells induces resistance to imatinib in a kinase-dependent manner. *J Biol Chem* 285, 21446–21457. (Cited on page 79.)
- Peng, C., Y. Chen, D. Li, and S. Li (2010). Role of pten in leukemia stem cells. *Oncotarget* 1, 156–160. (Cited on page 80.)

- Pittenger, M., A. Mackay, S. Beck, R. Jaiswal, R. Douglas, J. Mosca, M. Moorman, D. Simonetti, S. Craig, and D. Marshak (1999). Multilineage potential of adult human mesenchymal stem cells. *Science* 284, 143–147. (Cited on page 4.)
- Pleasance, E. D., R. K. Cheetham, P. J. Stephens, D. J. McBride, S. J. Humphray, C. D. Greenman, I. Varela, M.-L. Lin, G. R. Ordóñez, G. R. Bignell, et al. (2009). A comprehensive catalogue of somatic mutations from a human cancer genome. *Nature* 463, 191–196. (Cited on page 9.)
- Pleasance, E. D., R. K. Cheetham, P. J. Stephens, D. J. McBride, S. J. Humphray, C. D. Greenman, I. Varela, M.-L. Lin, G. R. Ordóñez, G. R. Bignell, K. Ye, J. Ali-paz, M. J. Bauer, D. Beare, A. Butler, R. J. Carter, L. Chen, A. J. Cox, S. Edkins, P. I. Kokko-Gonzales, N. A. Gormley, R. J. Grocock, C. D. Haudenschield, M. M. Hims, T. James, M. Jia, Z. Kingsbury, C. Leroy, J. Marshall, A. Menzies, L. J. Mudie, Z. Ning, T. Royce, O. B. Schulz-Trieglaff, A. Spiridou, L. A. Stebbings, L. Szajkowski, J. Teague, D. Williamson, L. Chin, M. T. Ross, P. J. Campbell, D. R. Bentley, P. A. Futreal, and M. R. Stratton (2010). A comprehensive catalogue of somatic mutations from a human cancer genome. *Nature* 463, 191–196. (Cited on page 47.)
- Popper, K. (1959). *The Logic of Scientific Discovery*. Hutchinson. (Cited on page 1.)
- Potten, C. S., R. Gandara, Y. R. Mahida, M. Loeffler, and N. A. Wright (2009). The stem cells of small intestinal crypts: where are they? *Cell Proliferation* 42, 731–750. (Cited on pages 4, 28, 46 and 82.)
- Raff, M. C. (1992). Social controls on cell survival and cell death. *Nature* 356, 397–400. (Cited on pages 4 and 46.)
- Rapoport, A. and A. M. Chammah (1965). *Prisoner's Dilemma*. University of Michigan Press, Ann Arbor. (Cited on pages 108 and 116.)
- Raymond, P., J. Warrell, T. Hugues de, Z. Y. Wang, and L. Degos (1993). Acute promyelocytic leukemia. *The New England Journal of Medicine* 329, 177–189. (Cited on page 52.)
- Remold, S. K. and R. E. Lenski (2001). Contribution of individual random mutations to genotype-by-environment interactions in eschrichia coli. *Proceedings of the National Academy of Sciences USA* 98, 11388–11393. (Cited on page 102.)

- Ridley, M. (1996). *Evolution* (Second ed.). Wiley-Blackwell. (Cited on page 99.)
- Rodriguez-Brenes, I. A., N. L. Komarova, and D. Wodarz (2011). Evolutionary dynamics of feedback escape and the development of stem-cell-driven cancers. *Proceedings of the National Academy of Sciences* 108, 18983–18988. (Cited on pages 47 and 80.)
- Roeder, I., M. Horn, I. Glauche, A. Hochhaus, M. C. Mueller, and M. Loeffler (2006). Dynamic modeling of imatinib-treated chronic myeloid leukemia: functional insights and clinical implications. *Nature Medicine* 12, 1181–1184. (Cited on pages 6 and 47.)
- Rossi, D., D. Bryder, J. Zahn, H. Ahlenius, R. Sonu, A. Wagers, and I. Weissman (2005). Cell intrinsic alterations underlie hematopoietic stem cell aging. *Proceedings of the National Academy of Sciences* 102, 9194–9199. (Cited on page 6.)
- Rossi, D., C. Jamieson, and I. Weissman (2008). Stem cells and the pathways to aging and cancer. *Cell* 132(4), 681–696. (Cited on page 6.)
- Rowley, J. D. (1973). Letter: A new consistent chromosomal abnormality in chronic myelogenous leukaemia identified by quinacrine fluorescence and giemsa staining. *Nature* 243, 290–293. (Cited on page 66.)
- Rufer, N., T. H. H. Brümmendorf, S. Kolvraa, C. Bischoff, K. Christensen, L. Wadsworth, M. Schulzer, and P. M. Lansdorp (1999). Telomere fluorescence measurements in granulocytes and T lymphocyte subsets point to a high turnover of hematopoietic stem cells and memory T cells in early childhood. *Journal of Experimental Medicine* 190, 157–167. (Cited on pages 6, 12, 49, 81, 87 and 92.)
- Saglio, G., D. W. Kim, S. Issaragrisil, P. le Coutre, G. Etienne, C. Lobo, R. Pasquini, R. E. Clark, A. Hochhaus, T. P. Hughes, N. Gallagher, A. Hoenekopp, M. Dong, A. Haque, R. A. Larson, H. M. Kantarjian, and ENESTnd Investigators (2010). Nilotinib versus imatinib for newly diagnosed chronic myeloid leukemia. *New England Journal of Medicine* 362, 2251–2259. (Cited on pages 60 and 66.)
- Sanjuan, A. F., A. Moya, and S. F. Elena (2004). The distribution of fitness effects caused by single nucleotide substitutions in an rna virus. *Proceedings of the National Academy of Sciences* 101, 8396–8401. (Cited on page 102.)

- Sawyers, C. L. (1999). Chronic myeloid leukemia. *New England Journal of Medicine* 340, 1330–1340. (Cited on page 10.)
- Scadden, D. (2006). The stem-cell niche as an entity of action. *Nature* 441, 1075–1079. (Cited on page 5.)
- Schwab, M., J. Ellison, M. Busch, W. Rosenau, H. E. Varmus, and J. M. Bishop (1984). Enhanced expression of the human gene n-myc consequent to amplification of dna may contribute to malignant progression of neuroblastoma. *Proceedings of the National Academy of Sciences* 81, 4940–4944. (Cited on page 8.)
- Shah, N. P., C. Tran, F. Y. Lee, P. Chen, D. Norris, and C. L. Sawyers (2004). Overriding imatinib resistance with a novel abl kinase inhibitor. *Science* 305, 399–401. (Cited on pages 10 and 66.)
- Shah, S. P., A. Roth, R. Goya, A. Oloumi, G. Ha, Y. Zhao, G. Turashvili, J. Ding, K. Tse, G. Haffari, A. Bashashati, L. M. Prentice, J. Khattra, A. Burleigh, D. Yap, V. Bernard, A. McPherson, K. Shumansky, A. Crisan, R. Giuliany, A. Heravi-Moussavi, J. Rosner, D. Lai, I. Birol, R. Varhol, A. Tam, N. Dhalla, T. Zeng, K. Ma, S. K. Chan, M. Griffith, A. Moradian, S. W. Cheng, G. B. Morin, P. Watson, K. Gelmon, S. Chia, S. F. Chin, C. Curtis, O. M. Rueda, P. D. Pharoah, S. Damaraju, J. Mackey, K. Hoon, T. Harkins, V. Tadigotla, M. Sigaroudinia, P. Gascard, T. Tlsty, J. F. Costello, I. M. Meyer, C. J. Eaves, W. W. Wasserman, S. Jones, D. Huntsman, M. Hirst, C. Caldas, M. A. Marra, and S. Aparicio (2012). The clonal and mutational evolution spectrum of primary triple-negative breast cancers. *Nature* 486, 395–399. (Cited on pages 8 and 47.)
- Sharpless, N. E., R. A. DePinho, et al. (2004). Telomeres, stem cells, senescence, and cancer. *Journal of Clinical Investigation* 113, 160–168. (Cited on page 11.)
- Shay, J. and S. Bacchetti (1997). A survey of telomerase activity in human cancer. *European Journal of Cancer* 33, 787–791. (Cited on page 11.)
- Shay, J., W. Wright, et al. (2000). Hayflick, his limit, and cellular ageing. *Nature Reviews Molecular Cell Biology* 1, 72–75. (Cited on page 5.)
- Sidorov, I. A., D. Gee, and D. S. Dimitrov (2004). A kinetic model of telomere shortening in infants and adults. *Journal of Theoretical Biology* 226, 169–175. (Cited on page 82.)

- Siegel, R., D. Naishadham, and A. Jemal (2013). Cancer statistics, 2013. *Cancer Journal for Clinicians* 63, 11–30. (Cited on pages 2 and 3.)
- Sjöblom, T., S. Jones, L. Wood, D. Parsons, J. Lin, T. Barber, D. Mandelker, R. Leary, J. Ptak, N. Silliman, S. Szabo, P. Buckhaults, C. Farrell, P. Meeh, S. Markowitz, J. Willis, D. Dawson, J. Willson, A. Gazdar, J. Hartigan, L. Wu, C. Liu, G. Parmigiani, B. Park, K. Bachman, N. Papadopoulos, B. Vogelstein, K. Kinzler, and V. Velculescu (2006). The consensus coding sequences of human breast and colorectal cancers. *Science* 314, 268–274. (Cited on pages 8 and 47.)
- Skaggs, B. J., M. E. Gorre, A. Ryzkin, M. R. Burgess, Y. Xie, Y. Han, E. Komisopoulou, L. M. Brown, J. A. Loo, E. M. Landaw, C. L. Sawyers, and T. G. Graeber (2006). Phosphorylation of the atp-binding loop directs oncogenicity of drug-resistant bcr-abl mutants. *Proceedings of the National Academy of Sciences* 103, 19466–19471. (Cited on page 80.)
- Stratton, M. R., P. J. Campbell, and P. A. Futreal (2009). The cancer genome. *Nature* 458, 719–724. (Cited on pages 9 and 46.)
- Tarnita, C. E., T. Antal, H. Ohtsuki, and M. A. Nowak (2009). Evolutionary dynamics in set structured populations. *Proceedings of the National Academy of Sciences* 106, 8601–8604. (Cited on page 20.)
- Taylor, C., Y. Iwasa, and M. A. Nowak (2006). A symmetry of fixation times in evolutionary dynamics. *Journal of Theoretical Biology* 243, 245–251. (Cited on page 18.)
- Taylor, J. R. (1997). *An Introduction to Error Analysis: The Study of Uncertainties in Physical Measurements*, Volume Second Edition. University Science Books. (Cited on page 57.)
- Taylor, P. D. and L. Jonker (1978). Evolutionary stable strategies and game dynamics. *Mathematical Biosciences* 40, 145–156. (Cited on pages 23 and 105.)
- Thompson, J. and B. Cunningham (2002). Geographic structure and dynamics of coevolutionary selection. *Nature* 417, 735–738. (Cited on page 102.)
- Thorley-Lawson, D. A. and A. Gross (2004). Persistence of the epstein–barr virus and the origins of associated lymphomas. *New England Journal of Medicine* 350, 1328–1337. (Cited on page 7.)

- Traulsen, A., J. C. Claussen, and C. Hauert (2005). Coevolutionary dynamics: from finite to infinite populations. *Physical Review Letters* 95, 238701. (Cited on page 22.)
- Traulsen, A., J. C. Claussen, and C. Hauert (2006). Coevolutionary dynamics in large, but finite populations. *Physical Review E* 74, 011901. (Cited on page 23.)
- Traulsen, A. and C. Hauert (2009). Stochastic evolutionary game dynamics. In H. G. Schuster (Ed.), *Reviews of Nonlinear Dynamics and Complexity*, Volume II, pp. 25–61. Weinheim: Wiley-VCH. (Cited on pages 15, 16, 18 and 22.)
- Traulsen, A., T. Lenaerts, J. M. Pacheco, and D. Dingli (2013). On the dynamics of neutral mutations in a mathematical model for a homogeneous stem cell population. *Journal of the Royal Society Interface* 10, 20120810. (Cited on pages 47 and 51.)
- Traulsen, A., J. M. Pacheco, and D. Dingli (2007). On the origin of multiple mutant clones in paroxysmal nocturnal globinuria. *Stem Cells* 25, 3081–3084. (Cited on page 42.)
- Traulsen, A., J. M. Pacheco, and D. Dingli (2010). Reproductive fitness advantage of bcr-abl expressing leukemia cells. *Cancer Letters* 294, 43–48. (Cited on pages 10 and 118.)
- Traulsen, A., J. M. Pacheco, L. Luzzatto, and D. Dingli (2010). Somatic mutations and the hierarchy of hematopoiesis. *BioEssays* 32, 1003–1008. (Cited on pages 36 and 43.)
- Traulsen, A. and F. A. Reed (2012). From genes to games: Cooperation and cyclic dominance in meiotic drive. *Journal of Theoretical Biology* 299, 120–125. (Cited on page 109.)
- Tumbar, T., G. Guasch, V. Greco, C. Blanpain, W. E. Lowry, M. Rendl, and E. Fuchs (2004). Defining the epithelial stem cell niche in skin. *Science* 303, 359–363. (Cited on pages 4, 5, 28, 46 and 82.)
- Turner, P. E. and L. Chao (1999). Prisoner’s Dilemma in an RNA virus. *Nature* 398, 441–443. (Cited on page 116.)

- van Valen, L. (1973). A new evolutionary law. *Evolutionary Theory* 1, 1–30. (Cited on page 1.)
- Varmus, H. E. (1984). The molecular genetics of cellular oncogenes. *Annual Review of Genetics* 18, 553–612. (Cited on page 8.)
- Vaziri, H., W. Dragowska, R. C. Allsopp, T. E. Thomas, C. B. Harley, and P. M. Lansdorp (1994). Evidence for a mitotic clock in human hematopoietic stem cells: loss of telomeric dna with age. *Proceedings of the National Academy of Sciences* 91, 9857–9860. (Cited on pages 11, 12, 40 and 81.)
- Vicente-Dueñas, C., F. Abollo-Jiménez, L. Ruiz-Roca, E. Alonso-Escudero, R. Jiménez, M. B. Cenador, F. J. Criado, C. Cobaleda, and I. Sánchez-García (2010). The age of the target cell affects b-cell leukaemia malignancy. *Aging* 12, 908–913. (Cited on page 80.)
- Vogelstein, B. and K. Kinzler (2004). Cancer genes and the pathways they control. *Nature Medicine* 10, 789–799. (Cited on pages 7 and 47.)
- von Bubnoff, N., C. Peschel, and J. Duyster (2003). Resistance of philadelphia-chromosome positive leukemia towards the kinase inhibitor imatinib (sti571, glivec): a targeted oncoprotein strikes back. *Leukemia* 17, 829–838. (Cited on page 78.)
- von Neumann, J. and O. Morgenstern (1944). *Theory of Games and Economic Behavior*. Princeton: Princeton University Press. (Cited on page 20.)
- von Zglinicki, T. (2002). Oxidative stress shortens telomeres. *Trends in Biochemical Sciences* 27, 339–344. (Cited on page 11.)
- Wald, R. M. (1984). *General relativity*. University of Chicago press. (Cited on page 1.)
- Walter, M. J., D. Shen, L. Ding, J. Shao, D. C. Koboldt, K. Chen, D. E. Larson, M. D. McLellan, D. Dooling, R. Abbott, R. Fulton, V. Magrini, H. Schmidt, J. Kalicki-Veizer, M. O’Laughlin, X. Fan, M. Grillot, S. Witowski, S. Heath, J. L. Frater, W. Eades, M. Tomasson, P. Westervelt, J. F. DiPersio, D. C. Link, E. R. Mardis, T. J. Ley, R. K. Wilson, and T. A. Graubert (2012). Clonal architecture of secondary acute myeloid leukemia. *New England Journal of Medicine* 366, 1090–1098. (Cited on page 47.)

- Weinberg, R. A. et al. (1991). Tumor suppressor genes. *Science* 254, 1138–1146. (Cited on page 8.)
- Weisberg, E., P. W. Manley, W. Breitenstein, J. Brügger, S. W. Cowan-Jacob, A. Ray, B. Huntly, D. Fabbro, G. Fendrich, E. Hall-Meyers, A. L. Kung, J. Mes-tan, G. Q. Daley, L. Callahan, L. Catley, C. Cavazza, M. Azam, A. Mohammed, D. Neuberg, R. D. Wright, D. G. Gilliland, and J. D. Griffin (2005). Characterization of amn107, a selective inhibitor of native and mutant bcr-abl. *Cancer Cell* 7, 129–141. (Cited on pages 10 and 66.)
- Welch, J. S., T. J. Ley, D. C. Link, C. A. Miller, D. E. Larson, D. C. Koboldt, L. D. Wartman, T. L. Lamprecht, F. Liu, J. Xia, C. Kandoth, R. S. Fulton, M. D. McLellan, D. J. Dooling, J. W. Wallis, K. Chen, C. C. Harris, H. K. Schmidt, J. M. Kalicki-Veizer, C. Lu, Q. Zhang, L. Lin, M. D. O’Laughlin, J. F. McMichael, K. D. Delehaunty, L. A. Fulton, V. J. Magrini, S. D. McGrath, R. T. Demeter, T. L. Vickery, J. Hundal, L. L. Cook, G. W. Swift, J. P. Reed, P. A. Alldredge, T. N. Wylie, J. R. Walker, M. A. Watson, S. E. Heath, W. D. Shannon, N. Varghese, R. Nagarajan, J. E. Payton, J. D. Baty, S. Kulkarni, J. M. Klcó, M. H. Tomasson, P. Westervelt, M. J. Walter, T. A. Graubert, J. F. DiPersio, L. Ding, E. R. Mardis, and R. K. Wilson (2012). The origin and evolution of mutations in acute myeloid leukemia. *Cell* 150, 264–278. (Cited on pages 46 and 61.)
- Werner, B., D. Dingli, T. Lenaerts, J. M. Pacheco, and A. Traulsen (2011). Dynamics of mutant cells in hierarchical organized tissues. *PLoS Computational Biology* 7, e1002290. (Cited on pages 4, 9, 13, 27, 46, 47, 49, 51, 80, 82, 121 and 158.)
- Werner, B., D. Lutz, T. H. Brümmendorf, A. Traulsen, and S. Balabanov (2011). Dynamics of resistance development to imatinib under increasing selection pressure: A combination of mathematical models and in vitro data. *PLoS One* 6, e28955. (Cited on pages 10, 60, 63, 122 and 158.)
- Wetzler, M., M. Talpaz, Z. Estrov, and R. Kurzrock (1993). Cml: mechanisms of disease initiation and progression. *Leukemia* 11, 47–50. (Cited on page 10.)
- Wild, G. and A. Traulsen (2007). The different limits of weak selection and the evolutionary dynamics of finite populations. *Journal of Theoretical Biology* 247, 382–390. (Cited on pages 21 and 113.)

- Wodarz, D. and N. Komarova (2005). *Computational biology of cancer: Lecture notes and mathematical modeling*. World Scientific Publishing. (Cited on pages 28 and 46.)
- Wright, S. (1932). The roles of mutation, inbreeding, crossbreeding and selection in evolution. In *Proc of the 6th International Congress of Genetics*, Volume 1, pp. 356–366. (Cited on page 1.)
- Wright, W. E., M. A. Piatyszek, W. E. Rainey, W. Byrd, and J. W. Shay (1996). Telomerase activity in human germline and embryonic tissues and cells. *Dev Genet* 18, 173–179. (Cited on page 81.)
- Wu, B., P. M. Altrock, L. Wang, and A. Traulsen (2010). Universality of weak selection. *Physical Review E* 82, 046106. (Cited on pages 20 and 21.)
- Wu, J., F. Meng, L. Y. Kong, Z. Peng, Y. Ying, W. G. Bornmann, B. G. Darnay, B. Lamothe, H. Sun, M. Talpaz, and N. J. Donato (2008). Association between imatinib-resistant bcr-abl mutation-negative leukemia and persistent activation of lyn kinase. *Journal of The National Cancer Institute* 100, 926–939. (Cited on page 79.)
- Xu, H., J. Yan, and S. T. Ildstad (2007). Xenotransplantation: a step closer to reality? *Blood* 110, 3815–3815. (Cited on page 43.)
- Yu, J., M. Vodyanik, K. Smuga-Otto, J. Antosiewicz-Bourget, J. Frane, S. Tian, J. Nie, G. Jonsdottir, V. Ruotti, R. Stewart, et al. (2007). Induced pluripotent stem cell lines derived from human somatic cells. *Science* 318, 1917–1920. (Cited on page 5.)
- Zeeman, E. C. (1980). Population dynamics from game theory. *Lecture Notes in Mathematics* 819, 471–497. (Cited on pages 23 and 105.)
- Zeyl, C. and J. A. DeVisser (2001). Estimates of the rate and distribution of fitness effects of spontaneous mutation in *Saccharomyces cerevisiae*. *Genetics* 157, 53–61. (Cited on page 102.)
- Zhang, J., C. Niu, L. Ye, H. Huang, X. He, W. Tong, J. Ross, J. Haug, T. Johnson, J. Feng, et al. (2003). Identification of the haematopoietic stem cell niche and control of the niche size. *Nature* 425, 836–841. (Cited on page 5.)

Zur Hausen, H. (2002). Papillomaviruses and cancer: from basic studies to clinical application. *Nature Reviews Cancer* 2, 342–350. (Cited on page [6](#).)

Acknowledgments

I'd like to express my sincere gratitude to my supervisor Arne Traulsen. He taught me the basics of scientific working and thinking. His patients and extraordinary support helped me to develop further as a person and a scientist alike. I appreciate the trust he showed and the responsibilities he equally distributes among all group members. I can not think of a better environment for my Phd than here. I further like to thank all the current and former members of the theory group, that are Weini, Philipp, Chaitanya, Benedikt, Christian, Julian, Maria, Bin, Yixian and Ursula for their daily happiness and support. Everyone is searching for success, but no one creates a competitive environment. Thanks for all the nice discussions and all the provided help.

I further like to thank all coauthors that worked and participated in my work during the last three years, especially I like to thank David Dingli for his immense support and his enduring positive feedbacks, but no lesser gratitude goes to Tom Lenaerts, Jorge M. Pacheco, Stefan Balabanov, Tim H. Brümmendorf and Fabian Beier. Thanks to Weini, Chaitanya and Micha for reading parts of the thesis.

Ich möchte besonders meiner Familie danken, die mich uneingeschränkt unterstützt und mir immer ein großer Halt sind und waren. Ohne euch hätte ich dieses Ziel sicher nicht erreicht!

However, the most important change in my life is certainly due to Weini. I'd like to thank you for all the support, all the happy hours and all the motivation during my work. I am looking forward to everything that will come in our both future. I also promise that I will improve my chinese in future :) !

Curriculum vitae

- Born September 15th 1984 in Leipzig
- 2003 Abitur at the gymnasium Veit-Ludwig van Seckendorff in Meuselwitz Thuringia
- September 2003 to June 2004 civil service in Berlin
- September 2004 to July 2006 basic studies in physics at the University of Leipzig
- August 2006 intermediate Diploma in Physics
- September 2006 to october 2008 advanced studies in Physics at the University of Leipzig
- December 2009 to February 2010 Diploma thesis under the guidance of Prof. Dr. Ulrich Behn with the title:"Idiotypische Netzwerke: Gedächtnis und Selbsttoleranz."
- May 2010 to May 2013 Phd student at the Max Planck Institute for Evolutionary Biology in Plön under the supervision of Prof. Dr. Arne Traulsen

Declaration

This thesis is a presentation of my original work, apart from my supervisor, Arne Traulsen's guidance. This thesis has not been submitted partly or wholly as a part of a doctoral degree to any other examining body. This thesis has been prepared according to the rules of Good Scientific Practice of the German Research Foundation.

Three papers published separately in the Journals PLoS Computational Biology, PLoS ONE, BMC Evolutionary Biology, one paper under revision in the Journal Interface of the Royal Society London and one chapter based on work in progress are included in this thesis.

- For the publication in PLoS Computational Biology ([Werner et al., 2011](#)), Benjamin Werner (BW), Arne Traulsen (AT), David Dingli (DD), Tom Lenaerts (TL) and Jorge M. Pacheco (JMP) designed the model, BW and AT performed the simulations and calculations and BW, AT, DD, TL, and JMP wrote the manuscript.
- For the publication in PLoS ONE ([Werner et al., 2011](#)), Stefan Balabanov (SB) and Tim H. Brümmendorf (THB) designed the experiment, David Lutz and SB performed the experiment, BW and AT designed the mathematical model performed the simulations and calculations, BW, AT, SB and THB wrote the manuscript.
- For the publication in BMC Evolutionary Biology ([Huang et al., 2012](#)), Weini Huang (WH) and AT designed the model, WH, BW and AT evaluated the model and wrote the manuscript.
- For the paper under revision in the Journal Interface of the Royal Society, BW and AT designed and evaluated the model, BW, AT and DD wrote the manuscript.
- For the work in progress, BW and AT designed and evaluated the model, BW and AT wrote the manuscript, SB THB and Fabian Beier designed and performed the experiments.

Benjamin Werner

