UNIVERSITÄT ZU LÜBECK

From the Department of Evolutionary Theory,
Max Planck Institute for Evolutionary Biology, Plön
Director: Prof. Dr. Arne Traulsen

Models for processes of somatic evolution on multiple scales

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

From the Department of Computer Sciences

Submitted by
Marvin A. Böttcher
from Hamburg

Lübeck, 2018

First referee: Prof. Dr. Arne Traulsen

Second referee: Prof. Dr. Hauke Busch

Date of oral examination: 31.8.2018

Approved for printing. Lübeck, 11.9.2018

Abstract

One of the main conditions for multicellularity is the limitation of long-term evolutionary success of the somatic cells within an organism. The somatic cells often constitute the vast majority of a multicellular organism. However, by definition they have no chance to give their genome to the next generation of the organism and merely act as a container for the successful reproduction of the few germ line cells. Since nutrients and space are limited within a multicellular organism, a careful coordination of the somatic cells in the tissue, along with a limitation of proliferation and strict control over cell function, is paramount for the evolutionary success of the organism. However, due to somatic mutations this balance can be disturbed such that somatic cells regain proliferative ability and outcompete other cells within the organism. This somatic evolution can ultimately lead to cancer or cancer-like phenomena with unlimited growth of specific clones, which can be lethal for the multicellular organism.

In this work I will present four different mathematical models showcasing processes of somatic evolution. Both main ingredients of evolution — variation and selection — will be shown in different examples in systems of different scales.

Firstly, on the molecular level, the basis for evolutionary processes of cells is the long term storage of genetic information and the process of how this information is converted into functional proteins. Slight changes in this process can have major consequences for the affected cells, also and especially for somatic cells embedded into multicellular organisms. Here, I will present a model for the translation of RNA into functional protein which takes circularization of RNA into account. This reveals a potential mechanisms for a fitness effect of synonymous mutations, that is mutations that do not alter the protein produced.

Secondly, on a larger scale, the strict coordination of often billions of somatic cells requires sophisticated tissue structures. These tissue structures can have a strong impact on the somatic evolution of cells within the tissue. Since mutations mainly arise on cell divisions, replicative age of cells in these structures is tightly controlled and mutation accumulation is highly irregular across different cell types. I will show how a measurable replicative age distribution might provide insight into specific tissue dynamics of a hierarchically structured tissue such as blood.

Thirdly, tissue structure also has a strong impact on selection of previously emerged mutants, as mutations can have varying fitness effects in different cell types. In this

work I will examine this for the example of chronic myeloid leukemia, a malignancy which is caused by a known somatic mutation for which nowadays specific targeted treatment is available.

Finally, evolution, and specifically evolution of somatic cells, has to be seen within the environment and the ecological niche of the evolving cells, since successful traits always depend on the momentary environment. In the context of cancer this environment is often set by the treatment of the disease which shapes the growth and extinction of cancer cells, which I will explore in more detail in this work.

Overall, the models presented in this work demonstrate mechanisms for observable patterns in biological systems in the context of somatic evolution. This also shows how experimental observations or data from clinical studies can be combined with insights from theoretical models for a deeper understanding of somatic evolution on the molecular scale, intercellular scale, or to support the treatment of cancer.

Zusammenfassung

Eine notwenwendige Bedingung für evolutionären Erfolg multizellulärer Lebewesen ist die Begrenzung des evolutionären Erfolgs der somatischen Zellen, also aller Zellen, die keine Keimzellen sind und die daher ihr Genom nicht an die nächste Generation weitergeben können. Häufig besteht ein multizellulärer Organismus zum Großteil aus somatischen Zellen, deren Aufgabe lediglich darin besteht, ein Vehikel für die erfolgreiche Fortpflanzung der Keimzellen bereitzustellen. Eine sorgfältige Koordination der somatischen Zellen inklusive einer strikten Kontrolle über Stoffwechsel und Zellteilung ist daher vorrangig für den evolutionären Erfolg des multizellulären Organismus, da Nährstoffe und Platzangebot in diesem begrenzt sind. Durch somatische Mutationen kann dieses Gleichgewicht jedoch empfindlich gestört werden, wenn einzelne somatische Zellen ihre Fähigkeit zur ungehemmten Proliferation zurückerlangen. In diesem Fall kann somatische Evolution zu Krebs oder krebsartigen Phänomenen führen, die tödlich für das multizelluläre Lebewesen sein können.

In dieser Arbeit stelle ich vier verschiedene mathematische Modelle vor, die Prozesse der somatischen Evolution veranschaulichen. Dabei werden beide Hauptbestandteile der Evolution — Variation und Selektion — anhand verschiedener Beispiele in den Fokus gerückt.

Die Grundlage aller evolutionärer Prozesse auf der molekularen Ebene ist die dauerhafte Speicherung genetischer Informationen und der Prozess, um diese Informationen in funktionale Einheiten, die Proteine, zu verarbeiten. Schon kleine Abweichungen in diesem Prozess können ernste Konsequenzen für eine Zelle nach sich ziehen, auch und insbesondere für somatische Zellen in einem multizellulären Organismus. In dieser Arbeit stelle ich ein Modell für die Translation von RNA in Proteine vor, welches sich mit den Konsequenzen einer Zirkularisierung der RNA befasst. Dabei wird auch ein Mechanismus aufgezeigt, der möglicherweise einen evolutionären Fitnesseffekt synonymer Mutationen impliziert.

Um die oftmals Milliarden von Zellen innerhalb multizellulärer Lebewesen zu koordinieren, sind häufig äußerst komplexe Strukturen der Zellorganisation notwendig. Diese Strukturen können einen starken Einfluss auf die somatische Evolution innerhalb des Organismus haben. Da Mutationen zumeist bei der Zellteilung entstehen, ist das replikative Altern somatischer Zellen in den Gewebestrukturen stark kontrolliert und die Akkumulierung von Mutationen ist extrem ungleich verteilt zwischen

verschiedenen Zelltypen. Ich werde zeigen, wie die messbaren Verteilungen replikativen Alters helfen können, Erkenntnisse über die dynamischen Parameter von stark strukturiertem Gewebe wie Blut zu gewinnen.

Die Zellorganisation hat jedoch nicht nur Einfluss auf die Ansammlung von Mutationen, sondern auch auf die Selektion von Mutanten und die Fitness einer mutierten Zelle hängt stark von der Position innerhalb der Zellstruktur ab. Diesen Zusammenhang betrachte ich anhand eines Modells für chronische myeloische Leukämie, einer Erkrankung des Bluts, für die eine bekannte Mutation verantwortlich ist und für die es inzwischen auch eine sehr gezielte Behandlung gibt.

Des Weiteren muss somatische Evolution immer im Zusammenhang mit den Umweltbedingungen und der ökologischen Nische der Zellen gesehen werden. Insbesondere bei Krebs werden diese Umweltbedingungen häufig von der Behandlung mitbestimmt, welche darüber wiederum Wachstum oder Aussterben der Krebszellen und auch den evolutionären Wettbewerb zwischen verschiedenen Zelltypen entscheidend beeinflusst, wie ich ebenfalls in dieser Arbeit aufzeigen werde.

Insgesamt beschreiben die Modelle dieser Arbeit mögliche Mechanismen für die Erklärung experimenteller Beobachtungen im Bereich der somatischen Evolution. Dabei werde ich theoretische Modelle mit experimentellen Beobachtungen oder den Daten klinischer Studien verbinden, um ein tieferes Verständnis der zugrunde liegenden Prozesse zu erhalten und möglicherweise die Behandlung von Krebs zu unterstützen.

Contents

1. Introduction

All life on earth is bound to a constant turnover - organisms constantly die and get reborn. A consequence of this turnover is a mechanism that leads to a continuous change of the organisms: evolution. In essence, evolution requires two components that are given in any living organism: variability in (heritable) traits and (some form of) selection on these traits (Darwin 1859). Hence, all organisms adapt to their environment by evolution while simultaneously shaping the environment themselves.

This also applies to the somatic cells within a single multicellular organism: to keep up the function of the organism, constantly new cells need to be formed and old ones die (Pellettieri and Alvarado 2007). As above, the consequence of this constant turnover is somatic evolution, which is the process of evolution within a multicellular organism. Somatic evolution can generally lead to cancer or cancer like phenomena in all observed multicellular organisms (Aktipis et al. 2015) — characterized by unlimited growth of somatic cells — thereby contrasting the success in somatic evolution of individual cells to the evolutionary success of the whole multicellular organism.

Commonly, all cell functions of somatic cells like growth, proliferation, and metabolism are tightly controlled in a complex biochemical network, mediated by signaling molecules, proteins and mechanical cues (Tyson et al. 2003; Hynes and Naba 2012). A central part of these networks is the regulation of gene expression (Davidson 2006). Slight variations in the gene or only the expression of a gene, i.e. via mutations, can therefore lead to drastic consequences in the behavior of a cell, causing selective advantage over other cells within the tissue and therefore enable somatic evolution which is detrimental to the multicellular organism.

For this reason many multicellular organisms have developed sophisticated mechanisms to detect and prevent somatic evolution in their cells, including apoptosis, DNA repair, or other integrity checks of cellular processes on the molecular level (Ellis et al. 1991; Sung and Klein 2006; Hanawalt and Spivak 2008). Another important mechanism to prevent somatic evolution in multicellular organisms is the specific organization of cells within the tissue (Michor et al. 2003a; Nowak et al. 2003), which affects both somatic variation and somatic selection. Only highly conserved and slowly dividing stem cells stay in the multicellular organism throughout its lifetime, while the bulk of somatic cells, which accumulates most of the somatic mutations, is built from transient cells that eventually die and get replaced.

This work looks at mathematical and computational models for several aspects of somatic evolution in different systems, from RNA translation on the molecular level

to the impact of fluctuating environments on intra-cellular competition. For this I employ various theoretical frameworks, including the totally asymmetric simple exclusion process (TASEP) (MacDonald et al. 1968), the Moran process (Moran 1958) and the replicator dynamics (Taylor and Jonker 1978).

Structure of this thesis In total there are six chapters in this thesis, including this introduction and a final conclusion. In the introduction I present the basic concepts that will be applied throughout this thesis. The following chapters 2-5 showcase somatic evolution in various systems and will have a short pre-introduction illustrating their relevance in the context of somatic evolution.

Chapter 2 is about the circularization of mRNA transcripts, and how it can potentially lead to selection on mutations that leave the final protein product intact. In chapter 3 I will show how distributions of replicative age change in hierarchically structured tissues of multicellular organisms — replicative age is strongly connected to the somatic variation due to the risk of mutations on cell divisions. Following the replicative age in structured tissue, in chapter 4 I will describe the emergence and treatment and eventually relapse of chronic myeloid leukemia, which is an example for a mutation with a fitness effect dependent on the differentiation stage within the hierarchical tissue structure. Chapter 5 describes the impact of fluctuating environments on the competitive fitness of two cancer cell phenotypes under treatment, a dormant and a rapidly proliferating phenotype. Finally, I will conclude in chapter 6 with a discussion of the complete thesis and an outlook.

1.1. Somatic evolution

One striking fact in life on earth is the cooperation of living cells into multicellular organisms which requires individual cells to partially or completely sacrifice their own reproductive success. In large mammals billions of cells constantly proliferate and die, but only a tiny fraction of cells, the germ line cells, are involved in the reproduction of an individual. Only these germ line cells have the chance to pass on their genome to the next generation and participate in the long-term evolutionary success of a lineage, while the somatic cells only serve as a carrier for passing on these germ line cells.

However, somatic cells are not excluded from evolution, as they also undergo a constant turnover. In fact, mutations in somatic cells are even more likely than in germ line cells (Milholland et al. 2017) leading to a larger variation in these cells. Therefore they also undergo somatic evolution. By definition, successful mutants inside

the soma are trapped in evolutionary dead-ends, as their only possible fate is to eventually die within their host organism[1].

As such, the basic principle of somatic evolution is not different from evolution between organisms (Darwin 1859) and is caused by variation in heritable traits and selection on these: As soon as there is variation between traits of individuals, they will have different chances for survival and in producing offspring. Since there is competition for space and nutrients, or in the case of multicellular organisms also for growth factors and other signalling molecules, individual cells struggle for existence and not every cell will be able to survive and reproduce. In the long run those traits leading to more offspring or a higher chance for survival under the given environmental conditions — the advantageous traits — prevail and displace those traits which are less advantageous.

As somatic evolution promotes the success of highly proliferative cells, it often leads to the breakdown of multicellular cooperation and can severely limit the reproductive success of the affected organism. Therefore, successful somatic evolution can lead to cancer or cancer like phenomena (Aktipis et al. 2015), which is potentially lethal to the organism and hence decreases the long term evolutionary success of the germ line cells of the organism. Hence, multicellular organisms spend a lot of energy to avoid or to limit the extent of somatic evolution in the first place. In rare exceptions, however, somatic evolution is actively promoted for certain cell types within the multicellular organism, as for example in humans in cells of the adaptive immune system (Odegard and Schatz 2006).

1.1.1. Molecular basis of evolution - heritability

The storage of information across generations in the molecule deoxyribonucleic acid (DNA) is the molecular basis of heritability of traits in all living cells (Watson and Crick 1953). This molecule stores the instructions for building anything in the cell in the form of a code, written in sequences of four different letters (A, T, C, G; see chapter 2 for details). This code acts as a blueprint to build the functional units of cells, the proteins; the code for a single protein is called gene.

Processing of the genetic information is basically a two step process (Crick 1958): In a process called transcription, a gene is copied from DNA into ribonucleic acid (RNA), which is a short-term storage of genetic information and contains the code for one or only few proteins. In the next step, called translation, the gene will be converted from RNA into protein. These two steps constantly occur in almost every cell of any organism, allowing for metabolism, growth and reproduction. Due to strong selection (see below section 1.1.3) on transcription and especially translation the genes involved

[1] Somatic evolution is not always trapped in an evolutionary dead-end, as there are very few transmissible cancers known to affect specific species such as Tasmanian devils (Ujvari et al. 2016).

in these processes are usually highly conserved across living organisms (Isenbarger et al. 2008; Eigen et al. 1989). Details of RNA translation will be shown in chapter 2.

Whenever a cell reproduces, the whole genome will be copied so that both emerging cells have a complete set of genetic information and can continue to proliferate and/or fulfill their purpose within the multicellular organism. Even though the process of DNA duplication is very precise and errors occur only with a very low probability (Kunkel and Bebenek 2000), changes in the DNA can occur due to the large size of the genetic code — mutations. These mutations are the major cause for heritable variation of cells, see below.

1.1.2. Variation

Variation in cellular traits is largely caused by mutations of the DNA. Mutations - changes in the genetic code - are commonly caused by errors while replicating the DNA (Kunkel and Bebenek 2000). By altering the genetic code of proteins they can have an influence on the function of the affected cell, which most often is disadvantageous (Wloch et al. 2001; Sanjuan 2010) as it disrupts the functioning of the involved protein and renders it unusable.

For multicellular organisms both advantageous and disadvantageous somatic variation can be detrimental, as in the former case the function of a cell for the organism is disrupted, while in the latter case the cell proliferates too much, thereby breaking the multicellular cooperation. Many multicellular organisms therefore employ a variety of mechanism to avoid or correct somatic variation in the first place (Ellis et al. 1991; Sung and Klein 2006). However, these mechanisms can be energetically costly (Breivik and Gaudernack 2004), and a multicellular organism is confronted with the trade-off between allowing somatic variation to a certain degree against wasting too much energy on the prevention of variation.

Since the primary source for somatic variation are cell divisions, the replicative age of cells, that is the number of cell divisions a cell has undergone, is an important aspect to control within the tissue of a multicellular organism (Rodriguez-Brenes et al. 2013). How exactly the replicative age structure changes within a hierarchical tissue structure and how this potentially changes under disease conditions will be shown in chapter 3.

1.1.3. Selection

Selection is the other main component for any evolutionary process. Each individual struggles for survival as not every individual of a population will survive and be able to reproduce. If there is a trait that gives an advantage to the affected individual

compared to the others, the lineage carrying this trait has higher chances to survive and will potentially take over the population. If, on the other hand, a trait is deleterious, the affected lineage will go extinct with a higher probability. This preservation or loss of traits in a population over time based on their effect on reproductive success of the affected individual is (natural) selection (Darwin 1859). In a more general sense, selection is the mechanism that leads to fixation or extinction of a trait within a population. Those types reproducing faster will build up a large population quickly and will constitute a large part of the population, therefore being able to acquire more of the resources and to eventually displace the slower growing type. The dependency of selection on environmental conditions will be highlighted in more detail in chapter 5.

However, neutral traits, that is traits that give no advantage or disadvantage to the individuals carrying them, can also fixate in a population by a process called drift. In this process, the stochastic reproduction and death events of individuals lead to the fixation or extinction of the trait within the population. However, in large populations relative stochastic fluctuation due to these random processes becomes vanishingly small (Gardiner 2009) and neutral traits will take up a constant share within the population according to their initial fraction in the population. This can also be observed in some cancers, where a large part of the mutations might in fact be neutral (Sottoriva et al. 2015; Williams et al. 2016).

1.1.4. Tissue structure and somatic evolution

Since multicellular organisms predominantly contain cells with different functions, they are highly controlled for proliferation, death and cell function. This control, typically mediated by signalling molecules or mechanical stress (Hynes and Naba 2012; Lu et al. 2011), is the reason why multicellular organisms can function with a high degree of cooperation between the individual cells.

Since all cells of the soma arise from a single cell, they all have more or less the same genome; with the exception of somatic mutations that occur during development and in the lifetime of the single organism. The key for cell specialization is the process of differentiation: by changing their gene expression pattern, cells obtain a different set of traits, acquiring different functions and abilities (Gurdon 1992; Gurdon and Melton 2008). Differentiation is often thought to be an irreversible process (Wang et al. 2010; Mojtahedi et al. 2016), and for several mature cell types the process of differentiation includes the removal of the nucleus including the genome, making further reproduction of the cells impossible, e.g. red blood cells in mammals.

On the root of cellular organization by differentiation are the stem cells: these cells are the basic cells that have the ability to differentiate into many other cell types, therefore rebuilding the complete tissue if necessary (Weissman 2000), but they often

fulfil no other tissue functions. Typically, there are relatively few stem cells in each tissue and these cells divide rarely.

Differentiated cells emerging from these stem cells can either already be used for specific tissue function or they are themselves precursors for even further differentiated cells. Accordingly, multicellular organisms can have a hierarchical structures of differentiation in tissue: Few stem cells stay in the tissue permanently, while a cascade of subsequent differentiation stages amplifies the number of cells such that enough functional cells get produced for the regular turnover of the tissue (homeostasis) (Gage 2000; Marshman et al. 2002). Most notably the blood (hematopoietic) system in mammals has an enormous daily production of cells, which is achieved by only very few long lasting stem cells and a pyramid of cell number amplifications (Passegué et al. 2005; Dingli et al. 2007; Busch et al. 2015).

Generally, somatic evolution can be strongly influenced by tissue structure. Chapters 3 and 4 will show examples for both aspects of somatic evolution — variation and selection — within hierarchically structured tissue in more detail.

1.1.5. Cancer

Cancer is a disease of multicellular organisms that is characterized by abnormal behavior (usually too much proliferation) of cells and is often lethal; it is the second most common cause of death in humans (Ferlay et al. 2015). In essence, cancer is the result of somatic evolution: due to possibly several mutations, a cell acquires the ability for unlimited proliferation despite the careful checks and control in the multicellular organism.

In humans, a single mutation is in most cases not sufficient to evade the control mechanisms of the organism and to cause cancer, and cells need to acquire many mutations over time to become cancerous (Armitage and Doll 2004; Nowell 1976). Therefore, cancer cells normally carry several somatic driver mutations which increase their selective advantage over the other cells of a tissue (Vogelstein et al. 2013).

Many cancers can be characterized by a certain set of phenotypic changes which are called the hallmarks of cancer (Hanahan and Weinberg 2011). Most cancer cells exhibit some or all of these phenotypic changes, such as evasion from cell death, ability to recruit blood vessels, or genetic instability.

Except for physical removal of cancerous tissue or destruction with radiation, common cancer treatment is the use of chemotherapy, which mainly kills proliferating cells (Corrie 2008). By this, chemotherapy predominantly kills cancer cells as these grow faster and are therefore more susceptible than healthy non-cancerous cells within a patient. However, in some cases few tumor cells survive the chemotherapeutic treatment and the cancer will relapse after some time. As these regrown cells might have

acquired resistance to the previous treatment, it is potentially much harder or impossible to treat a cancer in the case of relapse (Callaghan et al. 2014). Most often the lethality of cancer in humans arises from dispersal and subsequent colonization of cancer cells in distant tissues, leading to metastasis (Chaffer and Weinberg 2011).

Two chapters of this thesis present models of cancer in two different systems: Chapter 4 shows the emergence, treatment, or relapse of chronic myeloid leukia (CML), a malignancy of the blood, with a drug that targets a specific mutation and is remarkably successful. Chapter 5 shows the impact of cyclic treatment with chemotherapy at the example of glioblastoma, a highly malignant brain tumor.

1.2. Mathematical models for evolution

Evolutionary processes can often be analyzed quantitatively with mathematical models which describe the change of populations over time (Nowak 2006; Hofbauer and Sigmund 1998; Smith 1986). Dependent on the specific application different models can be useful; in the following I will present the basic mathematical frameworks that will be employed in this thesis.

1.2.1. Fitness

Fitness is a concept to quantitatively describe evolutionary success of individuals within populations. It can therefore be seen as a measure for the strength of selection for a specific trait (Darwin 1859). For mathematical models of evolution fitness is a very central concept as it allows to describe the change in a population based on the momentary population, which makes the quantitate description of evolution and future predictions for evolutionary processes possible in the first place. Since for sexually reproducing organisms only parts of an individual's genome will be carried to the next generation and the characteristics of the trait in the next generation depend on the genetic background, in general the mathematical definition of fitness is not straightforward (Kempthorne and Pollak 1970). However, since somatic reproduction is exclusively asexual a simple definition is in this case sufficient:

In the context of this thesis fitness of an individual carrying a specific trait is defined as the number of (surviving) offspring this individual can produce. It therefore measures the potential long term success of a lineage carrying a trait - higher fitness means more offspring, which due to the inherited trait again produce more offspring than other individuals in the population and will eventually fixate in the population. Additionally it is often useful to normalize fitnesses of all individuals within a population to one, such that the fitness value gives the fraction of offspring compared to the total offspring of the population.

1.2.2. Replicator dynamics

The replicator dynamics describes the change of species or type frequencies within an infinitely large population (Taylor and Jonker 1978; Hofbauer et al. 1979; Zeeman 1980). Since the population is assumed to be infinitely large, changes in clone frequency are always caused by differences in fitness of the types and there is no stochastic fluctuation or drift. Each clone or type i has a frequency in the population of x_i and all frequencies have to sum up to one $\sum_i x_i = 1$. The fitness of each type is given by f_i and can generally depend on population composition and/or time $f_i(\vec{x}, t)$ for the vector of all clone frequencies $\vec{x} = (x_0\ x_1\ \dots)$. Since changes in frequency of the types are given by the competitive fitness, also the average fitness $\phi = \sum_i x_i f_i$ is required. The frequency of each type changes proportional to its frequency and to its fitness compared to the average fitness:

$$\frac{\mathrm{d}x_i}{\mathrm{d}t} = x_i(f_i - \phi), \tag{1.1}$$

which is called the replicator equation.

For frequency dependent fitnesses, that is $f_i = f_i(\vec{x})$, this equation is nonlinear as $\frac{\mathrm{d}x_i}{\mathrm{d}t} \propto x_i^2$ and analytic solutions are often difficult to obtain. In this case stability analysis can help to gain insight into the model system (Strogatz 2014). Stability analysis investigates the behavior of a system close to its fixed points, by observing the response to very small perturbations to these fixed points.

In the following I show this for the example of a simple replicator equation of only two types, where I only have to calculate the fixed point for one of the type with frequency x and fitness $f(x)$, as the frequencies of types always add up to one. From the replicator equation shown above the fixed points are obtained by solving

$$0 = \frac{\mathrm{d}x}{\mathrm{d}t} = x(f - \phi) =: g(x).$$

The obtained points fixed points are commonly denoted by a star x^*, so that $g(x^*) = 0$. The next step is to investigate the behavior of the system in close vicinity to the fixed points. For this one is interested in how the system behaves at the point $x = x^* + \varepsilon$ which is only a small perturbation ε away from the fixed point

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \frac{\mathrm{d}(x^* + \varepsilon)}{\mathrm{d}t} = \frac{\mathrm{d}\varepsilon}{\mathrm{d}t} = g(x^* + \varepsilon).$$

Next step is to do a Taylor expansion of the last term up to first order of ε

$$\frac{\mathrm{d}\varepsilon}{\mathrm{d}t} = g(x) = g(x^* + \varepsilon) = g(x^*) + g'(x^*)\varepsilon + O(\varepsilon^2).$$

This equation shows how the system behaves close to the fixed point: If the first derivative of the replicator equation at the fixed point is smaller than zero, $g'(x^*) < 0$,

the change of perturbation over time is in opposite direction as the perturbation itself and the fixed point is stable. If, on the other hand, $g'(x^*) > 0$, the perturbation will grow and the fixed point is unstable. For the intermediate case of $g'(x^*) = 0$, linear stability analysis cannot make any predictions and one needs look at higher order terms of the Taylor expansion of $g(x)$.

The method for linear stability analysis presented here can be extend for more dimensional dynamical systems in a straightforward manner: By calculating the eigenvalues of the Jacobian of the function $\vec{g}(\vec{x}^*)$ representing the multi dimensional replicator equation. If the real parts of all eigenvalues are smaller than zero, the corresponding fixed point is stable. If, on the other hand, only one of the eigenvalues is larger than zero, the fixed point is unstable.

In this thesis the replicator equation will occur in chapter 5, where it is used to calculate the influence of varying treatment conditions on a population with two types, implemented as fitness values f_i that change over time. As the replicator equation in this case is nonlinear, stability analysis as well as numerical integration will be applied to gain insights into the dynamical system.

1.2.3. Moran process

By contrast to the replicator equation above, the Moran process (Moran 1958) describes the stochastic change of clone sizes in a population with a finite and fixed number of individuals and also works in small population sizes. Similar to the replicator dynamics, it can be used to explore frequency changes of a mutant within this population in relation to the fitness effect of the mutation; however, one is often more interested in the fixation or extinction probabilities of the types within the population. The Moran process is useful to model the evolutionary dynamics of stem cells, as the number of stem cells is usually small and can be assumed to be constant over time. Accordingly, in chapter 4, I will describe a model where the dynamics on the stem cell level is equivalent to a neutral Moran process (Klein and Simons 2011).

In essence, the Moran process describes the fate of two types A and B within a population of fixed size N. On each update step in the population, one individual is randomly chosen for reproduction with a probability proportional to its fitness while another individual is randomly chosen for death and will be replaced by the offspring of the reproduced individual[2].

Mathematically, the Moran process can be described with transition probabilities and analytical solutions for both fixation probabilities and the average fixation time can be calculated. In the following I will exemplarily calculate the fixation probability

[2]In an unstructured, well-mixed, population it makes no difference whether or not an individual if first chosen for reproduction (birth-death) or the other way around (death-birth). In a structured population, however, this is generally not the case (Hindersin and Traulsen 2015).

S_i for a neutral Moran process of i individuals of type A within a total population of N individuals. The transition probabilities T_+ for a transition from i to $i+1$ individuals of type A, T_- from i to $i-1$, or T_0 for an unchanged population are given by multiplying the individual probabilities for birth and death

$$\begin{aligned} T_+ &= \frac{i}{N}\frac{(N-i)}{N} \\ T_- &= \frac{(N-i)}{N}\frac{i}{N} \\ T_0 &= 1 - T_+ - T_-. \end{aligned}$$

From this the fixation probability S_i can be written as a recursion:

$$\begin{aligned} S_i &= T_+ S_{i+1} - T_- S_{i-1} + T_0 S_i \\ &= T_+ S_{i+1} - T_- S_{i-1} + (1 - T_+ - T_-) S_i \\ &= T_+(S_{i+1} - S_i) - T_-(S_i - S_{i-1}) =: T_+ k_{i+1} - T_- k_i, \end{aligned}$$

where I defined the auxiliary quantity $k_i = S_i - S_{i-1}$. Since the fixation probability for zero individuals of type A is $S_0 = 0$ and therefore $k_1 = S_1$, one can directly

$$k_i = \left(\frac{T_-}{T_+}\right) k_{i-1} = \cdots = \left(\frac{T_-}{T_+}\right)^{i-1} S_1 = S_1.$$

To calculate the fixation probability from k_i we again use the fact that $S_0 = 0$

$$S_i = \sum_{j=0}^{i} k_j = i S_1.$$

The fixation probability for $i = N$ individuals of type A is $S_N = 1 = N S_1$ and the general fixation probability is given by

$$S_i = i S_1 = \frac{i}{N}. \tag{1.2}$$

As expected, for a neutral mutation, i.e. when A and B have the same fitness, the probability for fixation for one of the types is equal to the fraction of this type in the population.

1.3. Stochasticity in biological systems

Stochasticity is unavoidable in any system above zero temperature due to thermal fluctuations (Schwabl 2006). Small particles like pollen or dust submerged in water will undergo a erratic jittery movement, which was for example observed 1827 by Robert Brown and subsequently termed Brownian motion. The physical explanation

for this is in fact the thermal movement of the water molecules (Einstein 1905; von Smoluchowski 1906): The seemingly random observed motion of the larger particles is caused by countless collisions with the smaller water molecules, each collision transferring a different momentum in a different direction to the larger particle. The movement of the larger particle can then be described in the framework of stochastic dynamics (van Kampen 2007; Gardiner 2009). In the same way as Brownian particles in water, stochasticity emerging from the smallest scale has a strong influence in many systems (Chandrasekhar 1943).

Often, in large systems, the relative fluctuations caused by these thermal fluctuations are vanishingly small for the observables of interest and can safely be neglected. However, on the scale of molecular processes in a cell this is generally not the case and stochasticity plays a significant role in these systems (Ladbury and Arold 2012; Elowitz et al. 2002; Cai et al. 2006). The primary impact of stochasticity in evolution are mutations, as they provide the necessary variation for evolution to occur in the first place (discussed in section 1.1.2). However, also selection is heavily influenced by stochasticity, as individuals within a population often get born or die due to seemingly erratic circumstances (viewed from a larger scale), much like the Brownian motion mentioned above. This can lead to genetic drift within populations and mutations can fixate in a population due to this stochasticity (Kimura 1968).

Within multicellular organisms there is often a tight control of stochasticity and specifically on mutation and selection. Often there are multiple mechanisms which prevent cells that have acquired mutations to expand and take over the population. For cancerous cells to emerge, there are therefore often multiple mutations and therefore multiple stochastic hits on the same cell necessary (Iwasa et al. 2004; Ashcroft et al. 2015).

In general, there exist many elaborate frameworks for analysing stochastic systems and describing the properties of these systems (van Kampen 2007). These framework allow to investigate some models analytically, for example by using a master equation approach: The probabilities for all possible states of the system and the transition rates between these states are put into a system of ordinary differential equations which describes the time evolution for the probabilities of the states of the system.

1.3.1. Stochastic simulations

Often it is not possible to obtain general solutions for stochastic models with analytical frameworks, as the underlying model is too complex. This is for example the case in chapter 2 and chapter 4, as in both cases the number of states in the system and the number of possible transitions between these states are too large for an analytical treatment (Gillespie 2007). In this case stochastic simulations can be used to generate many trajectories for the process under consideration which can then be analyzed statistically. The main component of any stochastic simulation

is the selection of random events with a random number generator (RNG). These are algorithms that produce very long sequences of seemingly randomly distributed, but in fact purely deterministically calculated, numbers and are commonly the most computationally intensive part of stochastic simulations.

Creating trajectories for time-discrete processes like the Moran process (Moran 1958) is relatively straightforward to implement, as one only has to randomly pick one individual for birth and one for death according to their respective probabilities on each time step. For time-continuous processes, however, the Gillespie algorithm is a popular choice (see below).

Gillespie Algorithm A commonly used algorithm to create trajectories for time-continuous stochastic processes is the Gillespie algorithm, first described in 1977 for the simulation of chemical reactions (Gillespie 1977). Each possible reaction has a reaction rate which describes the average waiting time until this reaction occurs for the next time, but the exact timing of any reaction is subject to random fluctuations. In general, the Gillespie algorithm consists of the following steps:

1. Calculate the waiting time until the next reaction occurs τ based on the sum of all the reaction rates of the system $r = \sum_i r_i$ by assuming an exponentially distributed waiting time, such that $\tau = -\frac{1}{r}\log(\phi)$ with a uniformly distributed random number between zero and one $\phi \in (0, 1)$.

2. Pick one of the reactions with a probability proportional to the single reaction rate r_i.

3. Repeat from step one.

This effectively splits the stochastic process in a sequence of (random) waiting times and a selection process as in the time discrete stochastic simulation.

Compared to a discretization of a time continuous random process, this algorithm has the advantage that it is exact and does not rely on a discretization time step which is generally problematic. If a discrete time step is chosen too small to there will be many steps without any reaction, basically wasting several calculations of random numbers; if it is chosen too large, the error introduced by the discretization can strongly influence the results.

Since the development of the original Gillespie algorithm there have been numerous improvements for computational efficiency. With one of the improved algorithms, the next reaction method (Gibson and Bruck 2000), the number of random numbers that are required per reaction can be reduced to one which significantly decreases the computational cost per simulation time. In the initial step a next reaction tree and a dependency graph are initialized: The next reaction tree stores all individual reactions with their waiting times τ, sorted such that the next reaction is at the stem

of the tree. The dependency graph stores the necessary updates for the dependent reactions after a reaction occurred, e.g. the number of reactants changed.

The algorithm starts by executing the reaction at the stem of the next reaction tree and the simulation time t updated $t \leftarrow t + \tau$. Only for this reaction a new waiting time will be generated by generating a new random number. The waiting times of all dependent reactions, as stored in the dependency graph, will be updated. The updated waiting time for a dependent reaction τ_n can be calculated from the old waiting time of the dependent reaction τ_o, the old and new reaction rates r_o and r_n, the old and new number of reactants n_o and n_n, and the simulation time t (Cao et al. 2004):

$$\tau_n = \frac{n_o r_o}{n_n r_n}(\tau_o - t) + t.$$

The updated reaction times are then sorted back into the tree at the correct positions according to their new waiting times and the next reaction at the stem of the tree will be executed.

However, the next reaction method is significantly more complex to implement, as it involves specific data structures for efficient access and sorting of the relevant data. This is why it is usually implemented only for larger systems with frequent reactions that have to be simulated over a long time, as for example the model presented in chapter 4 of this thesis.

Random number generation Computers (and of course also humans) generally have difficulties producing real random numbers. Therefore, especially in the context of stochastic simulations, algorithms are used to produce seemingly randomly distributed numbers, so called pseudo random number generators (RNG). Simple random number generators often produce correlated numbers or numbers that lie on a plane in multi-dimensional space which renders them basically useless for the use in scientific applications (Marsaglia 1968). Accordingly, a multitude of statistical tests has been developed to ensure the suitability of a specific random number generator (Marsaglia and Tsang 2002).

Nowadays powerful random number generators are implemented in most programming languages or available as libraries. In this work I exclusively use the Mersenne Twister (Matsumoto and Nishimura 1998) which passes the relevant criteria for stochastic simulations in scientific applications and is also reasonably fast. For the stochastic simulations presented in this thesis the Mersenne Twister as defined in the C++11 standard (ISO/IEC 2011) is used.

2. RNA translation

RNA translation and somatic evolution

Production of functioning proteins from the genome is one of the fundamental processes in all living organisms and is essential for survival, growth and reproduction of any cell. It basically consists of two key processes, which is often called the central dogma of molecular biology (Crick 1958): First, the transcription of the genetic code from long lasting and stable double stranded deoxyribonucleic acid (DNA) into single stranded "messenger"-ribonucleic acid (mRNA). Secondly, the translation of this mRNA into functional units, the proteins.

Proteins are the major player involved in virtually any process occurring in the cell (Crick 1958). Each cell constantly produces proteins to metabolize, grow, reproduce, or basically to do anything. Therefore even slight changes in the production of proteins can have a drastic influence on the fitness of a cell, which results in a strong selection pressure on the optimization of protein production. Additionally, the process of translation is energetically costly and the associated cellular machinery is extremely complex to produce and maintain, which adds to the selection pressure for the efficiency of RNA translation (Gingold and Pilpel 2011). This is especially true for any multicellular organism where cell function and proliferation have to be tightly controlled (Gurdon 1992; Kellis et al. 2014), to allow for the complex coordination of multiple cells. Generally, cells react to their environment by expression of proteins embedded in complex biochemical reaction networks (Davidson 2006), so disturbances in protein production can potentially disturb coordination of somatic cells (Polymenis and Aramayo 2015).

Since RNA translation is one of the key steps in producing the proteins, the general fitness effect of protein production applies even more so to RNA translation. Accordingly, in multicellular organisms cancerous mutations can in some cases be connected to dysregulation of RNA translation (Blagden and Willis 2011).

Basics of RNA translation The general principle and many specifics of translation of RNA transcripts into proteins is well understood (Alberts 2015): Huge protein complexes, called ribosomes, move along the RNA and decode the genetic information of the transcript into the protein. Information on RNA transcripts is stored as in DNA in the form of four different nucleotides - the "letters" of the genetic alphabet:

Adenine (A), Uracil (U, instead of Thymine as in DNA), Cytosine (C) and Guanine (G) (Rich and Davies 1956). Three adjacent nucleotides constitute the basic translation unit of a codon; each codon codes for a specific amino acid that finally makes up the whole protein. In the actual translation process a ribosome will move along the codons of the RNA sequence and append the appropriate amino acid for each codon to the amino acid sequence. When a stop codon (UAA,UAG,UGA) is reached, the protein chain is released into the cytoplasm.

Generally, the efficiency of translation is determined the rate of protein production against the use of resources or energy used in the process (Gingold and Pilpel 2011). Accordingly, the efficiency of translation determines (i) how many proteins a cell gets from a transcript, (ii) how much energy the cell has to spend for these proteins, and (iii) how much time the cell needs to wait for these proteins. Since translation is a very complex process, efficiency of translation can be influenced by many factors, e.g. the structure of the ribosome or the mRNA (Gingold and Pilpel 2011). This chapter will explore how reinitiation of ribosomes — after a ribosome reached the end of the transcript it starts another round of translation on the same mRNA molecule — can affect the efficiency of translation.

Synonymous mutations Each codon consists of three nucleotides and there are four different nucleotides. Therefore there are in total $4^3 = 64$ different codons that can in principle code for genetic instructions[1]. However, there exist only 20 amino acids, the basic building blocks of proteins (Crick 1958). This means that there is a lot of redundancy in the genetic code and some amino acids can have multiple different codons coding for them.

Consequently, a mutation can be classified based on its functional impact: If a mutation does not change the sequence of the amino-acids, it is thought to have no or only a very low impact on the functional properties of the resulting protein. This is called a synonymous mutation. By contrast, a mutation which changes the amino acid sequence of the protein is non-synonymous and changes the chemical properties and function of the resulting protein. Often non-synonymous mutations render the involved protein completely dysfunctional and therefore have a strong impact on the fitness of the individual cell.

However, it is becoming increasingly clear that synonymous mutations can also have an impact on the fitness of a cell, especially in the context of somatic mutations and cancer (Supek et al. 2014; Gotea et al. 2015). Changes in codon usage can change the efficiency of translation and affect the amount of protein that is produced per transcript (Sørensen et al. 1989; Kudla et al. 2009). Since protein abundance is so important for cell function, especially in the setting of tightly regulated multicellular organisms, there is strong selection for optimal synonymous codons in translation.

[1] Three of these 64 codons are stop codons which tell the ribosome to stop translation, so in total only 61 codons code for amino acids.

The model presented in this chapter provides a possible explanation for these changes in translation efficiency caused by codon usage, that is under certain conditions it predicts substantial changes in protein abundance by introducing a synonymous mutation. It is therefore a mechanism which can potentially explain the above mentioned fitness effect of synonymous mutations in somatic cells.

Strength of selection Measuring the strength of selection of somatic mutations within cancer is especially important for their treatment: Often cancer treatments target specific mutations of the cancer cells (Sawyers 2004). Ideally, to have a negative effect of treatment on the fitness of cancer cells, these drugs should target mutations that are functionally important for the cancer and therefore under strong selection. However, the fitness effect of mutations in cancer is strongly debated (Sottoriva et al. 2015; Williams et al. 2016; Martincorena et al. 2017; Williams et al. 2018)

One common method to gain insight into the strength of selection acting on certain genes or parts of the genome — also in the context of somatic evolution of cancer (Ostrow et al. 2014; Martincorena et al. 2017) — is to compare the genetic code of multiple individuals and calculate the number of synonymous (supposedly neutral) mutations dS and non-synonymous mutations dN per site. The ratio $\omega = dN/dS$ then informs about the strength of selection acting on the genome (Yang and Bielawski 2000; Hurst 2002): If there are the same number of synonymous and non-synonymous substitutions ($\omega = 1$), the gene under consideration is neutral, as none of the non-synonymous mutations are suppressed or enhanced compared over the baseline mutation rate given by the synonymous mutations. If, on the other hand $\omega \ll 1$, then there is a strong deleterious fitness effect of a mutation at that site which is also termed purifying selection. Finally, if $\omega \gg 1$, there is strong positive selection at that particular site.

However, along with other potential biases this measure critically relies on the assumption of neutrality of synonymous mutations to measure the baseline neutral mutation rate dS. As outlined above and in the following, this neutrality of synonymous mutations might generally not be given, and measuring the strength of selection from sequencing data will become more complex (Zhou et al. 2010).

Publication information The rest of this chapter is published in the same words as:
Rogers, D. W., M. A. Böttcher, A. Traulsen, and D. Greig (2017). "Ribosome reinitiation can explain length-dependent translation of messenger RNA". *PLOS Computational Biology* **13**.6
(shared first authorship DWR and MAB, see author contributions for details).

Ribosome reinitiation can explain length-dependent translation of messenger RNA

David W. Rogers[1,2¶*], Marvin A. Böttcher[2,¶], Arne Traulsen[2], Duncan Greig[1,3]

[1]Experimental Evolution Research Group and
[2]Department of Evolutionary Theory, Max Planck Institute for Evolutionary Biology, Plön, Germany
[3]Department of Genetics, Evolution, and Environment, University College London, London, WC1E 6BT, United Kingdom

[¶]These authors contributed equally to this work.
*Corresponding author: rogers@evolbio.mpg.de

Abstract Models of mRNA translation usually presume that transcripts are linear; upon reaching the end of a transcript each terminating ribosome returns to the cytoplasmic pool before initiating anew on a different transcript. A consequence of linear models is that faster translation of a given mRNA is unlikely to generate more of the encoded protein, particularly at low ribosome availability. Recent evidence indicates that eukaryotic mRNAs are circularized, potentially allowing terminating ribosomes to preferentially reinitiate on the same transcript. Here we model the effect of ribosome reinitiation on translation and show that, at high levels of reinitiation, protein synthesis rates are dominated by the time required to translate a given transcript. Our model provides a simple mechanistic explanation for many previously enigmatic features of eukaryotic translation, including the negative correlation of both ribosome densities and protein abundance on transcript length, the importance of codon usage in determining protein synthesis rates, and the negative correlation between transcript length and both codon adaptation and 5' mRNA folding energies. In contrast to linear models where translation is largely limited by initiation rates, our model reveals that all three stages of translation—initiation, elongation, and termination/reinitiation—determine protein synthesis rates even at low ribosome availability.

Author Summary Recent advances in proteomics show that translation is strongly dependent on transcript length, but current theoretical models fail to capture this relationship. Here, we propose that the high initiation rates and protein yields of short transcripts result from terminating ribosomes reinitiating on the same transcript. The frequency of reinitiation depends on the time required to complete one full transit of a transcript, coupling transcript lengths and elongation rates to protein yield. Any slow step reduces the protein yield of shorter transcripts more than the yield of longer transcripts, generating stronger selective pressure to eliminate slow steps in shorter transcripts and explaining the widespread negative correlations in eukaryotes between transcript length and both 5' mRNA folding energy and codon adaptation. Our reinitiation-based model reconciles conflicting results from previous initiation-limited models with recent advances in biotechnology and identifies the mechanism underlying length-dependent translation, allowing powerful prediction of translational regulation across eukaryotes.

2.1. Introduction

The physiological state of a cell is largely determined by the identity and abundance of the proteins encoded by its genome. Understanding how genetic information is first transcribed into messenger RNA and then translated into protein is therefore fundamental to our understanding of biological systems. A wide variety of technologies has allowed detailed investigations of transcription, but—until very recently—a lack of similar tools for empirical research on translation has meant that the study of post-transcriptional regulation has been largely restricted to mathematical models with little opportunity for parameterization or evaluation. Recent advances in both sequencing technology and mass spectrometry have now produced large amounts of data on the translation of eukaryotic mRNA, revealing how transcript features, RNA-binding proteins, and non-coding RNAs influence translation (Lackner and Bähler 2008; Kuersten et al. 2013). While many of the determinants of translation rates revealed by these empirical studies were predicted by existing models, some remain difficult to explain. Perhaps the most striking correlate of translation rate is the length of the transcript itself. Multiple experimental studies, across a wide range of eukaryotic organisms, have demonstrated a steep negative correlation between the length of a given coding sequence (CDS) and three different measures of translation: translation initiation rates (Ciandrini et al. 2013; Shah et al. 2013; Weinberg et al. 2016), the density of ribosomes on a transcript (Weinberg et al. 2016; Arava et al. 2003; MacKay et al. 2004; Arava 2005; Lackner et al. 2007; Qin et al. 2007; Hendrickson et al. 2009; Ingolia et al. 2009; Park et al. 2011; Lauria et al. 2015; Thompson et al. 2016), and the abundance of the encoded protein (Ghaemmaghami et al. 2003; Gunaratne et al. 2013; Vogel et al. 2010; Wang et al. 2013).

Ribosome and polysome profiling experiments have shown a positive relationship between ribosome density and protein abundance, leading to the conclusion that transcripts with higher ribosome densities have higher translation rates (Lackner et al. 2007; Hendrickson et al. 2009; Plotkin and Kudla 2011). A positive relationship between ribosome density and translation rate can occur when translation is limited by low initiation rates. In traditional models of translation, initiation can be limiting when other steps in translation, such as elongation, occur quickly enough to prevent collisions between ribosomes (Plotkin and Kudla 2011). Consistent with this key role of initiation rates in determining translation rates, Arava et al. 2003 found that the higher densities of ribosomes on shorter transcripts was most consistent with shorter transcripts having exponentially higher initiation rates than longer transcripts, estimating a halving of the initiation rate with every 400-codon-increase in CDS length. More recent analyses (Ciandrini et al. 2013; Shah et al. 2013) have revealed that the relationship between CDS length and initiation rates is better described by a power law: the initiation rate is roughly halved for every doubling of CDS length (i.e. a log-log slope of approximately -1). However, the assumption of initiation-limitation leaves little room for variation in elongation rates to influence translation rates, which is at odds with recent work demonstrating that codon usage can be an important determinant of protein yields (Chu et al. 2014a; Tarrant and von der Haar 2014).

If translation is limited by the ability of transcripts to capture ribosomes from the cytoplasmic pool (the *de novo* initiation rate), mechanisms that allow transcripts to retain terminating ribosomes for subsequent rounds of translation should improve translation rates. The closed-loop model of translation was first proposed as a hypothetical mechanism to improve translation efficiency through intrapolysomal ribosome reinitiation (Philipps 1965; Baglioni et al. 1969). By bringing the sites of termination and initiation into close proximity through circularization of the mRNA, the closed-loop complex allows ribosomes that have finished translating to reinitiate translation on the same mRNA molecule rather than returning to the cytoplasmic pool. The closed-loop model was initially based on the appearance of many polysomes in electron micrographs as circular, rather than linear, structures (detailed high resolution tomographic analyses of circular polysomes are now available Afonina et al. 2015). Recent theoretical and experimental studies have shown that secondary structures in single stranded RNAs bring the 5' and 3' ends close together (equivalent to the distance spanned by 9–16 nucleotides) meaning that mRNAs are effectively circularized (Yoffe et al. 2011; Leija-Martínez et al. 2014). Interactions between initiation factors bound to the 5' end, and proteins associated with the 3' end including release and recycling factors, and the poly(A) binding proteins, are thought to facilitate translation, possibly by stabilizing the closed-loop structure or by actively promoting reinitiation (Mazumder et al. 2003; Wilkie et al. 2003).

The importance of reinitiation of ribosomes on circular transcripts in determining protein yield is well established *in vitro* (Philipps 1965; Baglioni et al. 1969; Nelson and Winkler 1987; Kopeina et al. 2008; Alekhina et al. 2007). Measuring translation of the luminescent protein luciferase in a eukaryotic cell-free system, Kopeina et al.

2008 showed that circular polysomes rarely exchanged ribosomes with the free pool or lost ribosomes to other transcripts, but linear polysomes did so frequently. On circular polysomes, most terminating ribosomes immediately reinitiated on the same mRNA molecule (see also Nelson and Winkler 1987). Alekhina et al. 2007 found that protein production in a similar cell-free system does not rapidly reach a steady state, as would be expected under a linear model of translation, but rather accelerates over the lifetime of the transcript, consistent with reinitiation on the same transcript. They proposed that the translation rate initially depends on slow *de novo* initiation of ribosomes from the free pool but soon becomes dominated by the much faster process of reinitiation.

Here, we use a minimal computational model to investigate the consequences of ribosome reinitiation on translation, with particular focus on transcript length and codon usage. We find that reinitiation causes ribosome densities, overall initiation rates, and protein yields to decrease with increasing transcript length. Furthermore, higher levels of reinitiation increase the importance of codon usage in determining translation rates in a length-dependent manner, even at low ribosome densities or low *de novo* initiation rates. Reinitiation therefore provides a potential mechanistic explanation for multiple previously-enigmatic patterns observed in empirical studies of translation.

2.2. Model

We use a totally asymmetric simple exclusion process (TASEP, reviewed in Chou et al. 2011) to investigate the closed loop model of translation. The TASEP (Fig. 2.1) models each transcript as a one-dimensional lattice consisting of a number of sites equal to the number of codons in the CDS: each site represents a single codon. Each site can be either free or occupied by a ribosome. Ribosomes move along the transcript in the 5' to 3' direction and cannot occupy the same codon(s) as any other ribosome. In our model, the transcript is circularized, meaning that terminating ribosomes can not only be released into the cytoplasmic pool (as in a linear TASEP) but can also move to the initiation site of the same transcript (reinitiation).

Four different types of reactions can take place in the TASEP: (i) *de novo* initiation: a free ribosome can be placed onto the 5' end of the transcript (the initiation site) at the *de novo* initiation rate; (ii) elongation: ribosomes at any codon on the transcript (except the termination site) can move forward one codon in the 3' direction at the elongation rate; if a ribosome occupies the termination site, it can either (iii) leave the transcript at the release rate or (iv) it can move to the initiation site at the reinitiation rate.

We model ribosomes as extended particles that occupy ten codons each: the A-sites (where each codon is translated) of adjacent ribosomes must be spaced apart by at

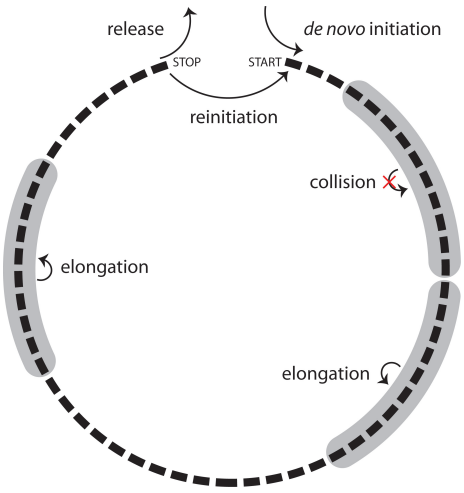

Figure 2.1: The closed-loop model of translation. Ribosomes (shaded in grey), modeled as extended particles that occupy 10 codons (black boxes), join the transcript at the start codon from the cytoplasmic pool at the *de novo* initiation rate, and hop to the next codon (in the 5' to 3' direction) at the elongation rate. Upon reaching the stop codon, ribosomes either return to the cytoplasmic pool at the release rate or return to the start codon at the reinitiation rate. The reinitiation level is determined by the reinitiation rate divided by the sum of the reinitiation and release rates. If the initiation site is occupied (i.e. any of the first 10 codons is being decoded), new ribosomes fail to join the transcript and reinitiating ribosomes either remain at the termination site or return to the cytoplasmic pool at the release rate. An elongating ribosome fails to step forward if the distance between its center and that of the ribosome in front is 10 codons (collision). Stochastic simulations were performed using the Gillespie algorithm. The Gillespie algorithm consists of multiple steps: (1) Initialization: the simulation time is set to zero; at this point in our simulations, all transcripts are empty and the only possible reaction is *de novo* initiation; (2) List all possible reactions: all possible reactions are determined and their rates are used to calculate the total rate of possible reactions; (3) Monte Carlo step: two random numbers are generated, the first determines the waiting time until the next reaction based on the total rate of possible reactions, and the second determines which reaction occurs using each reaction rate as a probabilistic weight; (4) Update: the time is increased by the randomly generated waiting time from step 3 and the chosen reaction is performed; (5) Iteration: repeat from step 2 unless the transcript lifetime has been reached.

least 10 codons. Thus, the elongation reaction is only possible when the A-site of the next ribosome in the 3' direction is > 10 codons downstream. Similarly, neither *de novo* initiation nor reinitiation is possible if any of the first 10 codons is occupied by an A-site.

Analytical solutions of the TASEP are possible, but currently can only be applied to the steady state. Consequently, most TASEP models, including a recent study of reinitiation (Marshall et al. 2014), investigate translation at the steady state, where the rate at which ribosomes join a transcript equals the rate at which they leave, and the translation rate is constant. However, every real transcript spends some proportion of its lifetime outside of the steady state, where these solutions do not apply; the assumption of a perpetual steady state is therefore an approximation. A new transcript does not instantly acquire ribosomes distributed along its length. Instead, ribosomes join at the 5' end and gradually progress towards the stop codon, where they can be released. In the absence of reinitiation, the steady state can be reached once the first ribosome to join a transcript is released. The duration of this "pioneer round" (Maquat et al. 2010) increases with transcript length, but generally represents a small proportion of eukaryotic transcript lifetimes. In the absence of reinitiation the steady state is therefore a good approximation (although it can be inappropriate for prokaryotes with short-lived transcripts Nagar et al. 2011; Gorissen and Vanderzande 2012). However under reinitiation, ribosomes do not

necessarily leave the transcript upon termination, which causes the effective initiation rate (and the translation rate) to increase over time (Alekhina et al. 2007). The time to reach the steady state therefore increases with both transcript length and reinitiation probability, and the time spent outside of the steady state thus represents a greater proportion of transcript lifetimes (Fig. 2.S1A). The steady state assumption consequently becomes a much worse approximation of translation at higher levels of reinitiation, overestimating translation rates on long transcripts and underestimating translation on short transcripts (Fig. 2.S1B). It is therefore impossible to make a fair comparison of translation at different reinitiation levels using the steady state approximation, particularly for transcripts of different lengths.

Since we do not assume that translation on any given transcript is always at the steady state, we cannot use the steady state analytical solutions of the TASEP. Instead, we perform stochastic simulations using the Gillespie algorithm (Gillespie 1977), which capture both the steady state and the non-steady state. In models that assume the steady state, all translation that occurs in simulations prior to the steady state is ignored. For example, in a recent reinitiation-based model of translation in yeast, the first 10^5s of simulations was discarded (Marshall et al. 2014). Given that the average lifetime of yeast transcripts is on the scale of 10^3-10^4 s (Pelechano et al. 2010), this means that all translation occurring over biologically plausible lifetimes was excluded from the analysis. Here, we make no assumptions about the steady state; we simply account for all translation that occurs during the lifetime of a transcript (both before and after the steady state is achieved). We simulated translation on each transcript independently. Each run generated a time evolution of the ribosome occupancy at each codon on a given transcript. We computed three measures of translation: ribosome density (the average number of ribosomes on a transcript over its lifetime divided by one tenth of CDS length, because each ribosome occupies 10 codons), effective initiation rate (the total number of initiations occurring through either *de novo* initiation or reinitiation divided by transcript lifetime) and protein yield (the total number of ribosomes reaching the stop codon of a transcript). We averaged the results of 1000 runs to produce results that are not subject to large stochastic fluctuations. We do not consider untranslated regions and our transcripts therefore represent only the CDS. The code for our TASEP is available at: `https://github.com/marvinboe/reTASEP`

Changing any transcriptome-wide parameter can dramatically alter global ribosome usage. For instance, at a given *de novo* initiation rate, increasing the reinitiation probability increases the total number of actively translating ribosomes. While this effect may be true, given that reinitiation is expected to allow more efficient use of ribosomes (see Discussion), it makes parameterizing the model difficult because the actual level of reinitiation is unknown. To keep all simulations consistent with empirical values, we have adjusted the *de novo* initiation rate to maintain the empirically observed average ribosome density. For simplicity, we have kept the number of ribosomes on a 400-codon-long transcript constant (at 6 ribosomes) for all

transcriptome-wide reinitiation probabilities and elongation rates. See 2.S1 Text for details on parameter estimates used in each simulation.

2.3. Results

High levels of reinitiation generate length-dependent translation Our model captures the negative correlation between ribosome density and CDS length observed in empirical studies, but only if the probability of reinitiation is high (Fig. 2.2). This result is intuitive; if reinitiation were perfect, all ribosomes that initiate would continue to reinitiate and translate, never leaving a transcript until it degrades. The density of ribosomes on a transcript of a given length and age would therefore be determined exclusively by the *de novo* initiation rate. If the *de novo* initiation rate is the same for all transcripts, then all transcripts of a given age should carry the same number of ribosomes and ribosome density will be the inverse of CDS length (with a log-log slope of -1). At a given elongation rate, the time required for a ribosome to complete one cycle (travel from the start codon to the stop codon) is less for short transcripts than for long transcripts. This means that, prior to the steady state, reinitiation occurs more frequently on shorter transcripts resulting in higher protein yields for short transcripts than long transcripts. When all or nearly all terminating ribosomes reinitiate, the effective initiation rate is much higher for shorter transcripts—providing a simple mechanism that could explain the length-dependence of initiation rates predicted by recent studies of translation (Ciandrini et al. 2013; Shah et al. 2013; Weinberg et al. 2016; Arava et al. 2003). The higher ribosome densities on shorter transcripts increase the likelihood of collisions between ribosomes, resulting in deviations from the expected power law relationship between measures of translation and CDS length (Fig. 2.2) through two related mechanisms. First, more frequent collisions between elongating ribosomes on shorter transcripts slow down translation, generating less steep slopes for the effective initiation rate and protein yield at low CDS lengths. Second, the initiation site is more likely to be occupied on a short transcript than on a long transcript, resulting in higher levels of initiation interference (Chu et al. 2014a) on shorter transcripts, further flattening length-dependence for all measures of translation on short transcripts.

When reinitiation is not perfect, ribosomes can return to the cytoplasmic pool after termination, and the effect of CDS length on ribosome density, effective initiation rate, and protein yield is diminished. Even small reductions in reinitiation probability greatly weaken length-dependence (Fig. 2.2). This is because short transcripts have more opportunities to lose ribosomes than do long transcripts. While a successful reinitiation event only guarantees that a ribosome remains associated with the transcript until the next termination event, ribosome loss is permanent. In the complete absence of reinitiation, length-dependence is therefore abolished.

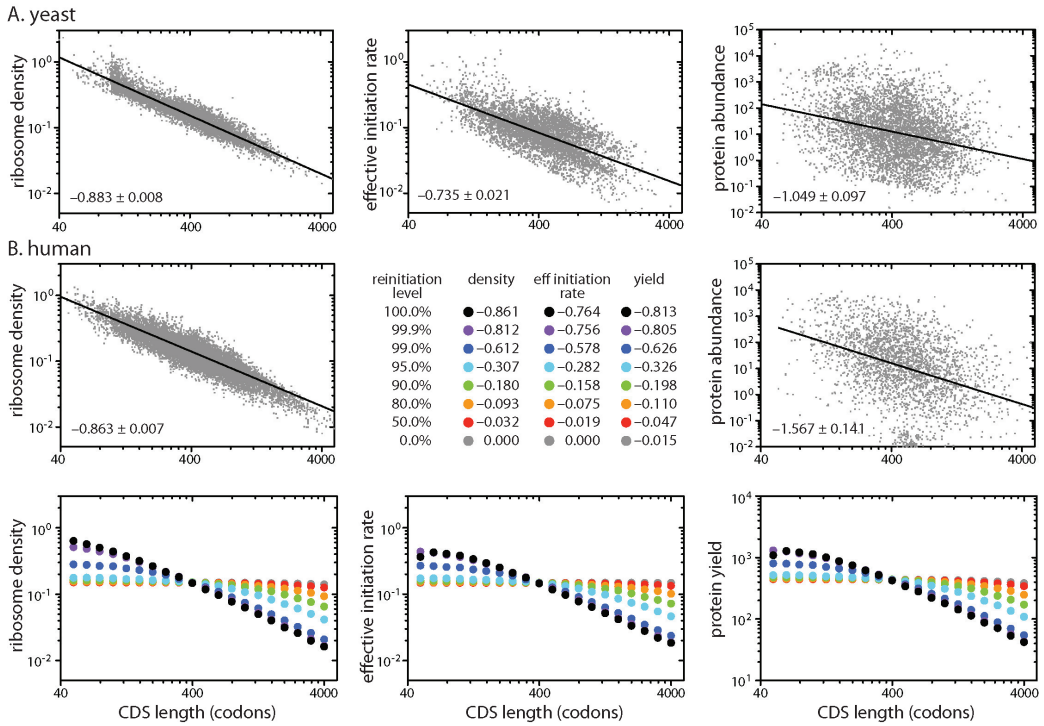

Figure 2.2: Reinitiation of post-termination ribosomes causes length-dependent translation. Ribosome density is the average number of ribosomes occupying a transcript during its lifetime divided by one tenth of the CDS length (since each ribosome occupies 10 codons). The effective initiation rate is the total number of initiation events (initiation and reinitiation) divided by transcript lifetime. Protein yield is the total number of ribosomes reaching the stop codon during the lifetime of a transcript. Slopes (95% confidence intervals) are indicated in the bottom left corner of each panel. *De novo* initiation rates were adjusted at each reinitiation level so that a 400-codon long transcript carried 6 ribosomes. **Top**: experimentally observed relationships between CDS length and ribosome density (left), initiation rate (center) or protein abundance (right) in the budding yeast *Saccharomyces cerevisiae*. **Middle**: experimentally observed relationship between CDS length and ribosome density (left) and protein abundance (right) in the human HEK293T cell line. Estimates of the initiation rate are not currently available for this cell line, so we have used this space to list the slopes from our simulations. **Bottom:** predicted relationships between ribosome density (left), the effective initiation rate (center), and protein yield (right) and CDS length at different reinitiation levels (different colours) from our simulations. Simulations were performed using a fixed elongation rate of $10s^{-1}$. See Fig. 2.S2 for simulations at other parameter values. **Data sources**: Yeast densities are weighted averages of the signals in polysomal fractions for 6071 transcripts from (Arava et al. 2003); initiation rates for 5348 transcripts were calculated by Ciandrini et al. 2013 based on ribosome density data from (MacKay et al. 2004); protein abundances of 4694 proteins included in the Peptide Atlas 2013 dataset from PaxDb (Wang et al. 2012) normalized against the total number of proteins (expressed as parts per million). HEK293T densities were calculated from mean ribosome numbers (across 3 replicates) reported by Hendrickson et al. 2009; protein abundances of 2636 proteins with identified CDS lengths included in the Geiger MCP 2012 data set (based on spectral counting) from PaxDb (Wang et al. 2012).

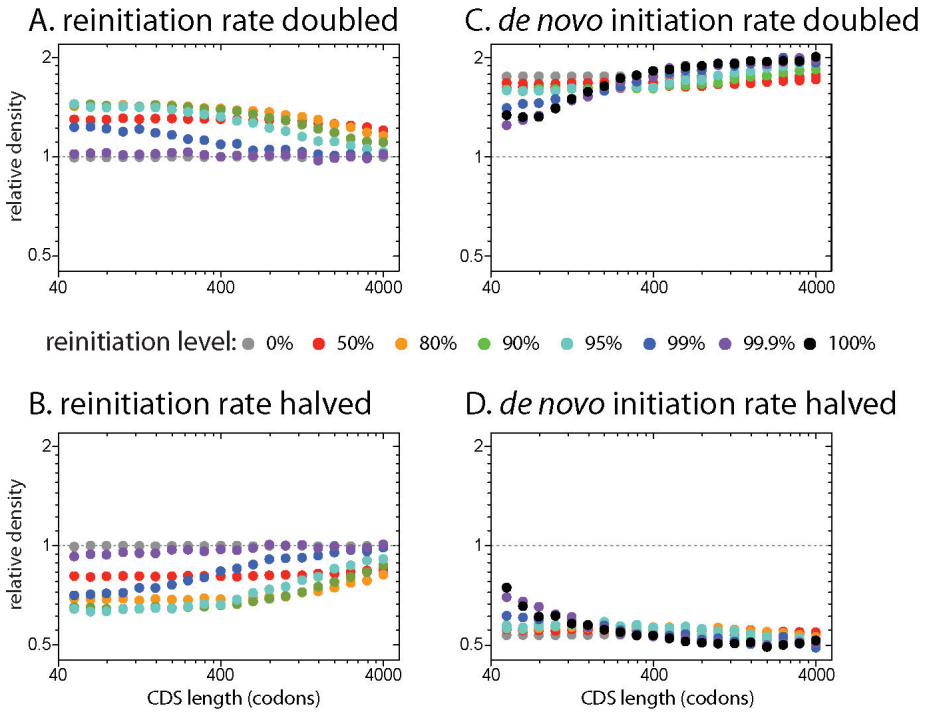

Figure 2.3: Transcript-specific change of the reinitiation rate, but not the *de novo* initiation rate, has larger effects on short transcripts than long transcripts. We simulated the effects of changing the reinitiation rate (A, B) or the *de novo* initiation rate (C, D) of a single transcript species by 2-fold or 0.5-fold at different transcriptome-wide reinitiation levels. For each transcriptome-wide reinitiation level (different colours), doubling (or halving) the reinitiation rate shifted the reinitiation level to: 99% to 99.5% (98.0%); 95% to 97.5% (90.5%); 90% to 94.7% (81.8%); 80% to 88.9% (66.7%); 50% to 66.7% (33.3%); 0% to 0% (0%). Doubling or halving the reinitiation rate at very high transcriptome-wide reinitiation levels (e.g. 99.9%) has little effect on translation since ribosomes rarely leave transcripts. Y-axes (log 2 scaled) show the ribosome density of altered transcripts relative to an equivalent transcript at the transcriptome wide reinitiation level. The effects of changing either the reinitiation rate or the *de novo* initiation rate on the effective initiation rate and protein yield were nearly identical to the effects on ribosome density. Transcriptome-wide *de novo* initiation rates were adjusted at each reinitiation level so that a 400-codon long CDS at the transcriptome-wide reinitiation level carried 6 ribosomes.

Reinitiation, but not *de novo* initiation, has a larger effect on short transcripts than long transcripts While changing transcriptome-wide parameters can dramatically affect global ribosome usage (see Model), altering parameters of transcripts encoded by a single gene will have little effect on global ribosome usage. This is because nearly all endogenous genes are expressed at low levels, so changing the translation parameters of the transcripts produced by a single gene will have a negligible effect on global ribosome availability (Shah et al. 2013; Plotkin and Kudla 2011; Charneski and Hurst 2013). By studying transcripts of individual genes, we can therefore investigate the consequences of changing a single parameter while holding all other values constant. We first tested the effects of altering the reinitiation rate of transcripts encoded by a single gene (Fig. 2.3A and B). Doubling the reinitiation rate results in

an extremely similar increase in all three measures of translation (ribosome density, effective initiation rate, and protein yield; results are therefore only shown for ribosome density), but the effects are greater for short transcripts than long transcripts. These effects are mirrored by a length-dependent decrease in translation when the reinitiation rate is halved (Fig. 2.3B). Furthermore, the length-dependent effects of changing the reinitiation rate of a single transcript species are generally stronger at higher transcriptome-wide reinitiation probabilities, except when reinitiation is so high that ribosomes rarely leave the transcript (e.g. 99.9%).

We next tested the effects of altering the *de novo* initiation rate of a single transcript species (Fig. 2.3C and D). In the absence of reinitiation, doubling the *de novo* initiation rate had an equal effect on ribosome density for transcripts of all lengths. However, at higher levels of reinitiation, doubling the *de novo* initiation rate resulted in a smaller increase in ribosome density on short transcripts than on long transcripts, caused by increased initiation interference; the higher density of ribosomes on short transcripts under reinitiation increases the probability that the initiation site is blocked, preventing successful *de novo* initiation. The effects of altering the de novo initiation rate on the effective initiation rate and protein yield are very similar to the effects on ribosome density.

High levels of reinitiation couple effective initiation rates and protein yields to the elongation rate So far, we have assumed that all transcripts have identical elongation rates, but in reality the elongation rate varies between transcripts encoded by different genes (Yu et al. 2015). We therefore investigated the consequences of changing the elongation rate of a single CDS from $10s^{-1}$ to either $20s^{-1}$ or $5s^{-1}$ (Fig. 2.4). Increasing the elongation rate reduces the amount of time between initiation and termination. In the absence of reinitiation, this causes ribosomes to spend less time on the altered transcript resulting in decreased ribosome density, but has little effect on the initiation rate or protein yield since these elongation rates are generally not limiting. Altered elongation rates do affect how long it takes to clear the initiation site and therefore the amount of initiation interference, explaining the relatively small differences in initiation rates and protein yields seen at 0% reinitiation (Chu et al. 2014a).

Under perfect reinitiation, terminating ribosomes explicitly reinitiate on the same transcript. Changing the elongation rate of a single gene therefore has no effect on the density of ribosomes on the altered transcript. However, by altering the time between reinitiation events, changing the elongation rate results in an equal change in the effective initiation rate of the altered transcript (Fig. 2.4). The protein yield of any endogenous gene is therefore exquisitely sensitive to changes in elongation rate under perfect reinitiation. Under perfect reinitiation, this effect is seen at all CDS lengths. The importance of the elongation rate decreases dramatically when reinitiation levels are reduced: faster elongation results in more opportunities to lose ribosomes, particularly on short transcripts.

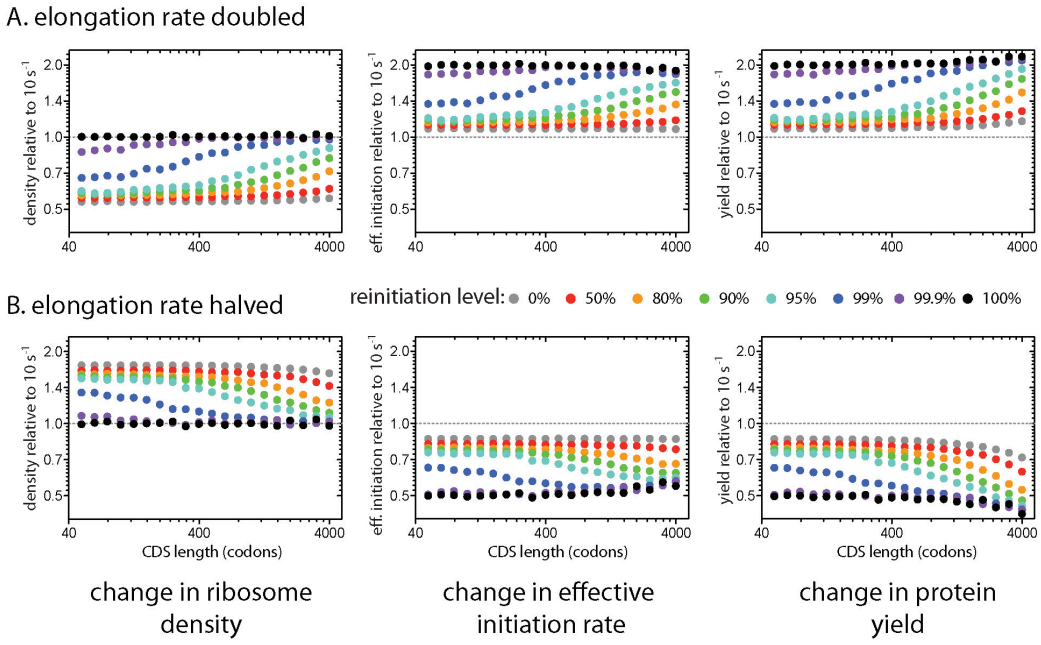

Figure 2.4: Transcript-specific change in translation caused by altering the average elongation rate of a single coding sequence. We simulated the effects of changing the average elongation rate of a single transcript species from $10s^{-1}$ to either (A) $20s^{-1}$ or (B) $5s^{-1}$ at different reinitiation levels (different colours). CDS length refers to the length of the altered coding sequence. Y-axes (log 2 scaled) show the effect of altering the elongation rate on the ribosome density (left), effective initiation rate (center), and protein yield (left) of the altered transcript at the higher or lower elongation rate relative to $10s^{-1}$ (dotted line). *De novo* initiation rates were adjusted at each reinitiation level so that a 400-codon long transcript with a fixed elongation rate of $10s^{-1}$ carried 6 ribosomes.

Length-dependent consequences of a single slow step on translation So far, we have only considered the effects of changing the average elongation rate of a transcript. However, it is difficult to imagine a mechanism that could simultaneously alter the elongation rate of all codons in a single transcript species without affecting the global elongation rate. Instead, transcripts are likely altered by mutations affecting a single codon at a time. Codon usage can affect elongation by determining the stability of secondary structures in the mRNA, but different codons are also decoded at different rates depending on the cellular availability of the appropriate tRNA. Most amino acids are encoded by multiple codons, and some codons (including synonymous codons that code for the same amino acid) are decoded faster than others (Yu et al. 2015; Presnyak et al. 2015). We therefore investigated the consequences of a single slow step on translation of transcript species of different lengths (Fig. 2.5). Here, we only examined translation at 99.9% reinitiation; similar results would be expected for other models of length-dependent translation including those that omit reinitiation. Introducing a single slow step into any transcript reduces its effective initiation rate and protein yield, but the effects are much larger for short transcripts than for long transcripts (Fig. 2.5). The length-dependence of a single slow step

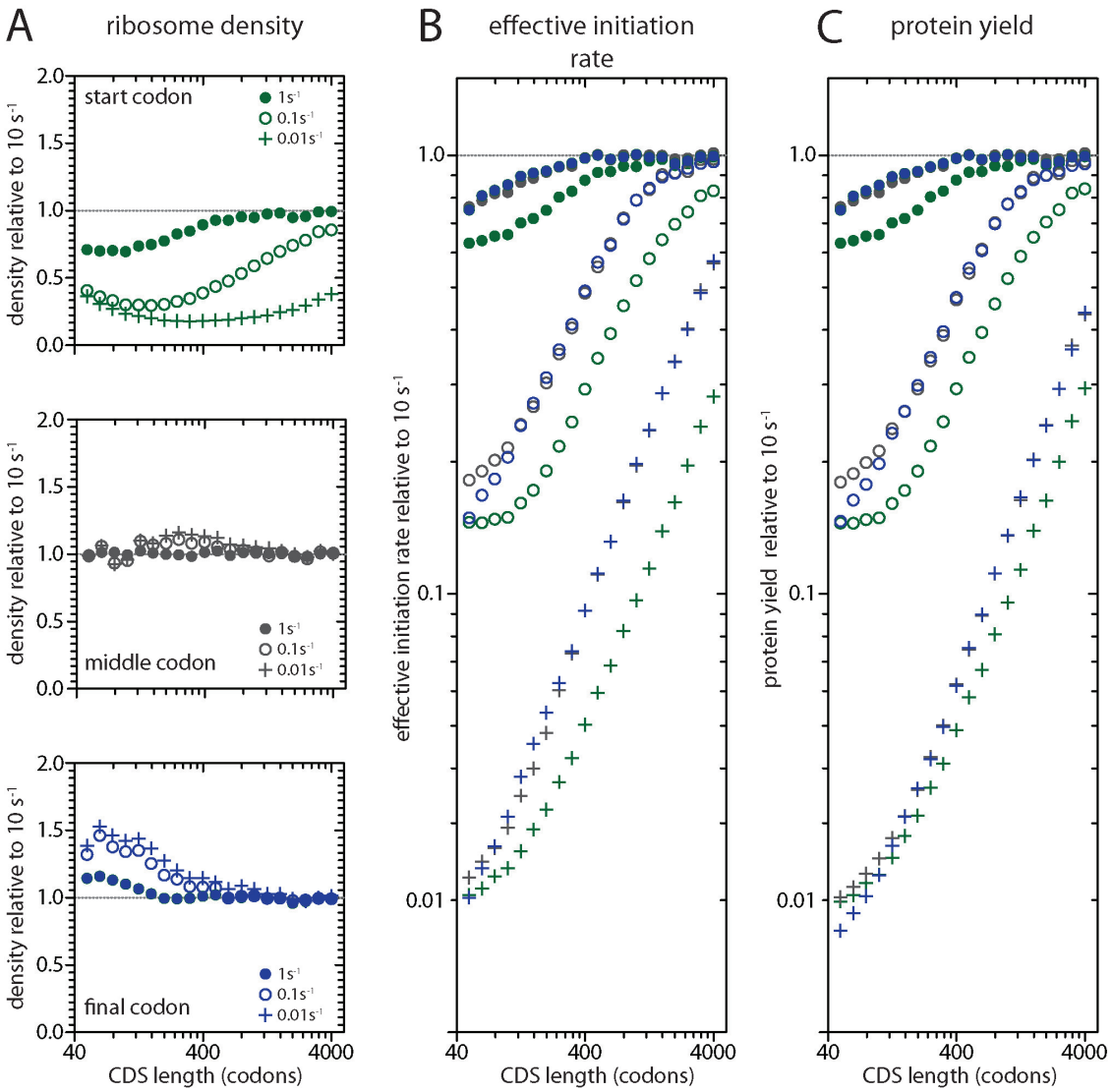

Figure 2.5: The consequences of a single slow step under length-dependent translation. We investigated the consequences of introducing a slow elongation step at either the start codon (green), the middle codon (grey), or the final codon (immediately before the stop codon, blue) under 99.9% reinitiation. CDS length refers to the length of the altered coding sequence. Slow codons were translated at $1s^{-1}$ (filled circles), $0.1s^{-1}$ (open circles), or $0.01s^{-1}$ (plus signs). Y-axes show the effect of a single slow step on the ribosome density (A), effective initiation rate (B), and protein yield (C) at the lower elongation rate relative to $10s^{-1}$ (dotted line).

arises from two sources. First, a single site represents a larger proportion of a short transcript than a long transcript and consequently results in a greater decrease in the average elongation rate (Comeron et al. 1999). Second, short transcripts have higher ribosome densities and are therefore more prone to collisions or "traffic jams" than are long transcripts. Effective initiation rates and protein yields are particularly sensitive to single slow steps near the start codon, with larger effects on shorter transcripts: slow clearance of the initiation site delays reinitiation and blocks *de novo* initiation resulting in lower ribosome densities on affected transcripts.

A yeast-specific model of translation with reinitiation Given the importance of variation in elongation rates to translation under reinitiation, we used our model to simulate translation in *S. cerevisiae* using codon-specific decoding rates. We used decoding rates (see Table 2.S1) estimated by Gilchrist and Wagner 2006 which are based on tRNA availability and wobble pairing rules and scaled so that the average decoding rate is $10s^{-1}$; they are related to measures of codon occupancy reported in (Hendrickson et al. 2009) ($r = 0.494$, $n = 61$, $P < 0.0001$). Since efficient reinitiation couples protein production to elongation rates, synonymous codon usage should have detectable consequences for protein yield at high levels of reinitiation. We tested the effects of synonymous codon usage at 99.9% reinitiation by predicting the yields of nine different synthetic GFP constructs (Lanza et al. 2014) that differ only in their synonymous codon usage (Fig. 2.6A). We compared these predictions to observed protein abundances measured in *S. cerevisiae* expressing each construct, and found a strong positive correlation between predicted yields and observed abundances ($r = 0.750$, $n = 9$, $P = 0.020$); our model predicted approximately half of the observed effect of using different synonymous codons (relative expression of highest vs. lowest construct, model = 2.4-fold, observed = 5.4-fold). Thus, efficient reinitiation correctly predicts a role for synonymous codon usage in determining yield.

Having established that Gilchrist & Wagner's (Gilchrist and Wagner 2006) codon-specific elongation rates are realistic, we used them to simulate the entire budding yeast translatome. The results of our simulations at 99.9% reinitiation are strongly correlated (Fig. 2.6B) with experimental measures of ribosome densities ($r = 0.932$, $n = 5542$, using data from Arava et al. 2003) and calculated initiation rates ($r = 0.742$, $n = 5348$, using estimates from Ciandrini et al. 2013; $r = 0.618$, $n = 3728$, using estimates from Shah et al. 2013). Our yield predictions are less strongly correlated with measured protein abundances ($r = 0.478$, $n = 4686$, data from Peptide Atlas 2013). This weaker correlation is unsurprising as our predictions of yield omit many important determinants of protein abundance including transcript abundance and protein stability. Results of simulations at other reinitiation levels are included in Fig. 2.S3 (fixed transcript lifetime) and Fig. 2.S4 (experimentally measured transcript lifetimes).

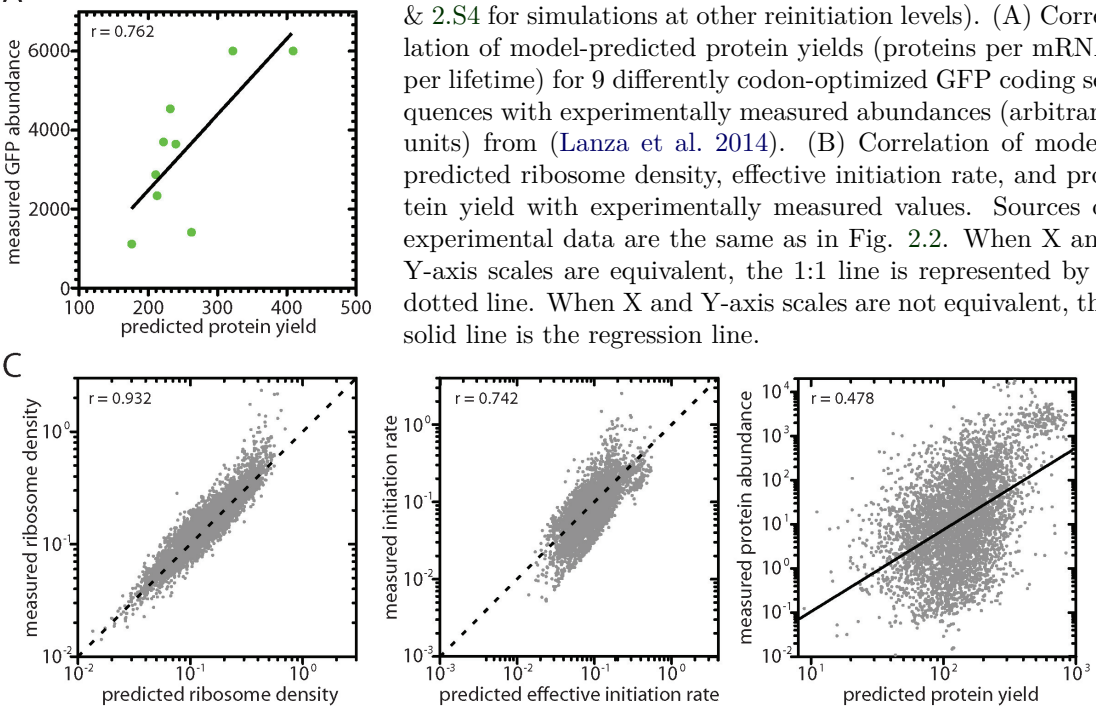

Figure 2.6: Simulating translation in the budding yeast *S. cerevisiae*. Our simulations were run using yeast-specific decoding rates (Gilchrist and Wagner 2006) and a transcript lifetime of 1553s (Pelechano et al. 2010). Results are shown for 99.9% reinitiation (see Figs. 2.S3 & 2.S4 for simulations at other reinitiation levels). (A) Correlation of model-predicted protein yields (proteins per mRNA per lifetime) for 9 differently codon-optimized GFP coding sequences with experimentally measured abundances (arbitrary units) from (Lanza et al. 2014). (B) Correlation of model-predicted ribosome density, effective initiation rate, and protein yield with experimentally measured values. Sources of experimental data are the same as in Fig. 2.2. When X and Y-axis scales are equivalent, the 1:1 line is represented by a dotted line. When X and Y-axis scales are not equivalent, the solid line is the regression line.

2.4. Discussion

We have shown that a fixed transcriptome-wide level of ribosome reinitiation can generate both length-dependent translation and a powerful transcript-specific role for codon usage, but only when reinitiation is extremely efficient. The level of reinitiation in live cells is unknown, but multiple studies have established that reinitiation is much more frequent than *de novo* initiation in cell-free systems. Furthermore, if reinitiation benefits the cell, we would expect it to evolve to become highly efficient. Maintaining a large pool of ribosomes consumes a substantial part of a cell's energy budget and selection will favor mechanisms that allow ribosomes to be used efficiently (Warner 1999). If, as reported by Nelson and Winkler 1987 and Kopeina et al. 2008, reinitiation of post-termination ribosomes is faster than *de novo* initiation from the free ribosome pool, then efficient reinitiation reduces the amount of "dead time" ribosomes spend in the pool waiting to be recruited (Marshall et al. 2014; Chu et al. 2014b). Each ribosome in a cell is therefore able to complete more rounds of translation in a given time interval under high levels of reinitiation compared to low levels of reinitiation. Reinitiation levels should be very closely associated with fitness: the translation initiation rate is thought to be the principal determinant of the rate of cell division (Soifer and Barkai 2014; Polymenis and Aramayo 2015). Consequently,

if reinitiation does occur in living cells, it is hard to imagine why it would not work very efficiently. Direct measurement of the level of reinitiation *in vivo* may soon be possible thanks to recent technological advancements enabling selective labeling of ribosomes (Jan et al. 2014) and the visualization of translation on individual mRNAs (Wang et al. 2016; Yan et al. 2016).

A single fixed level of reinitiation is not necessary to explain length-dependent translation; efficient reinitiation is only required on short transcripts (Fig. 2.S5). Studies in living cells have shown that some transcripts are more likely to be associated with translation factors required to form the closed-loop complex than others (Archer et al. 2015). If the closed-loop complex is required for efficient reinitiation, then reinitiation levels likely vary between transcripts. More specifically, shorter transcripts likely experience higher levels of reinitiation since they are both more likely to be enriched with closed-loop factors (Thompson et al. 2016; Costello et al. 2015), form more stable closed-loop complexes (Amrani et al. 2008), and may exhibit shorter end-to-end distances allowing increased levels of reinitiation by passive diffusion (Fernandes et al. 2017). Additionally, cellular depletion of both the closed loop factor eIF4G and the translational regulator Asc1/RACK1 has also been shown to have a greater effect on the translation of short transcripts than on long transcripts (Park et al. 2011; Thompson et al. 2016). Using length-dependent reinitiation levels in our simulations allows the empirical relationship between CDS length and ribosome density, effective initiation rate, and protein yield to be captured at an average reinitiation level orders of magnitude smaller (~90%; Fig. 2.S5) than does a fixed reinitiation level (99.9%; Fig. 2.S5).

Beyond acting on global mechanisms, natural selection also operates to maximize the protein yield of transcripts encoded by individual genes (translational efficiency Comeron et al. 1999). Selection for increased translational efficiency can not only increase the abundance of a given protein in a cell, but can also maintain protein levels while minimizing the cost of transcription, which has been shown to be an important determinant of fitness in yeast (Kafri et al. 2016). The strength of selection depends on the magnitude of the effect of a given mutation on translational efficiency; mutations with larger effects are subject to stronger selection. We have shown that the magnitude of the effect on translational efficiency of altering a given parameter by an equal amount can vary with the length of the altered transcript species. Thus, the strength of selection on mutations that affect a given parameter can be length-dependent (Comeron et al. 1999). For instance, doubling the reinitiation rate of a single transcript species results in a bigger increase in translational efficiency for shorter transcripts (Fig. 2.3). Mutations affecting the reinitiation rate of short transcripts are therefore more likely to be selected than are those than occur on long transcripts, potentially contributing to higher levels of reinitiation on shorter transcripts as discussed above (Fig. 2.S5). In contrast, doubling the *de novo* initiation rate does not result in higher translational efficiency on shorter transcripts and, under reinitiation, can actually have smaller effects on shorter transcripts due to increased initiation interference (Fig. 2.3). Selection for increased translational

efficiency on individual transcript species is therefore not predicted to result in higher *de novo* initiation rates on shorter transcripts. Instead, selection under reinitiation will be more effective at reducing initiation interference on shorter transcripts.

At high levels of reinitiation, we have shown that a single slow step in translation causes a greater reduction in the translational efficiency of short transcripts than that of long transcripts (Fig. 2.5). Eliminating slow steps has larger effects on the translation of short transcripts compared to long transcripts and therefore selection to eliminate slow steps will be most effective in genes encoding short transcripts. Length-dependent selection against slow steps under reinitiation therefore offers an explanation for the negative correlation between codon adaptation and CDS length observed across eukaryotes (Shah et al. 2013; Comeron et al. 1999; Ding et al. 2012; Duret and Mouchiroud 1999; Marais and Duret 2001; Kliman et al. 2003; Waldman et al. 2010, but see also Coghlan and Wolfe 2000). Translational efficiency is particularly sensitive to slow sites near the start codon (Fig. 2.5, see also (Chu et al. 2014a)): slow clearance of the initiation site delays reinitiation (promoting ribosome loss) and blocks *de novo* initiation resulting in lower ribosome densities on affected transcripts. Multiple mechanisms can determine how slowly ribosomes vacate the initiation site including the presence of one or more slow codons (Chu et al. 2014a) or the presence of stable 5' secondary structures in the transcript (Gu et al. 2010). As both features reduce yield to a greater extent on short transcripts compared to long transcripts (Fig. 2.5), selection should be more efficient at eliminating them on shorter transcripts, consistent with the negative correlations between CDS length and both 5' mRNA folding energy and 5' codon adaptation (Ding et al. 2012). Thus, length-dependent translation generated by high levels of reinitiation will generate length-dependent selection against slow steps (Comeron et al. 1999), which will in turn reinforce patterns of length-dependent translation.

Reinitiation provides a simple mechanistic explanation for empirically observed patterns of length-dependent translation including negative correlations between CDS length and ribosome density, effective initiation rate, protein yield, transcript codon adaptation, 5' codon adaptation, 5' folding energy, and association with closed-loop factors. Under reinitiation, these patterns are expected to emerge through selection for efficient ribosome usage, maximizing protein yield, and translational efficiency on individual transcript species. This is in sharp contrast to linear models in which, at low ribosome availability, length-dependence arises through direct selection for higher *de novo* initiation rates on shorter transcripts (Ciandrini et al. 2013; Shah et al. 2013). Our model is consistent with the emerging view that translation is controlled not only by initiation, but also by elongation and termination/reinitiation (Chu et al. 2014a; Tarrant and von der Haar 2014; Merrick and Harris 2014). This conceptual shift makes clear that manipulating any these stages can have profound consequences on translation, and presents factors associated with elongation, release, and recycling as new targets for therapeutic intervention (cf. Richter and Coller 2015).

2.S. Supporting information

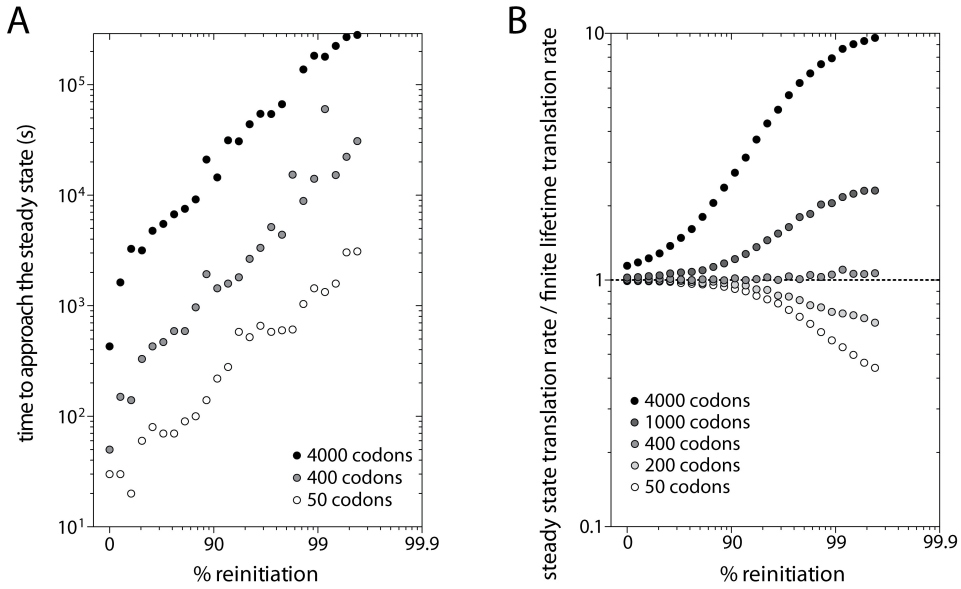

Figure 2.S1: The steady state is a poor approximation of translation at high reinitiation levels for transcripts with finite lifetimes. (A) Time to approach the steady state ribosome density at different reinitiation levels for transcripts of different lengths. Simulations were run for 3×10^5 s using the parameter values described for our general model. *De novo* initiation rates were adjusted at each reinitiation level so that a 400-codon transcript carried an average of 6 ribosomes at the steady state. The steady state ribosome density was calculated as the average density over the final 10^4 s of 1000 iterations (if this value showed no directional change over time). The first passage time represents the earliest time point that the average density of 1000 iterations equaled or exceeded the steady state density. Long transcripts (4000 codons) failed to reach the steady state during the 3×10^5 s run time at reinitiation levels above 99.6%. Data points at higher reinitiation levels were therefore excluded. (B) Steady state translation rate (yield s^{-1}) relative to the average translation rate (yield s^{-1}) on equivalent transcripts with a finite 3000 s lifetime. Deviations from a ratio of 1 represent the magnitude of the misestimation of the translation rate caused by assuming a perpetual steady state on transcripts with finite lifetimes. *De novo* initiation rates were adjusted at each reinitiation level such that the average 400 codon-transcript carried 6 ribosomes either at the steady state or on average over its 3000 s lifetime. Consequently, all 400 codon transcripts have the same average ribosome density allowing fair comparison. At high reinitiation levels, the steady state approximation overestimates ribosome density on long transcripts and underestimates ribosome density on short transcripts.

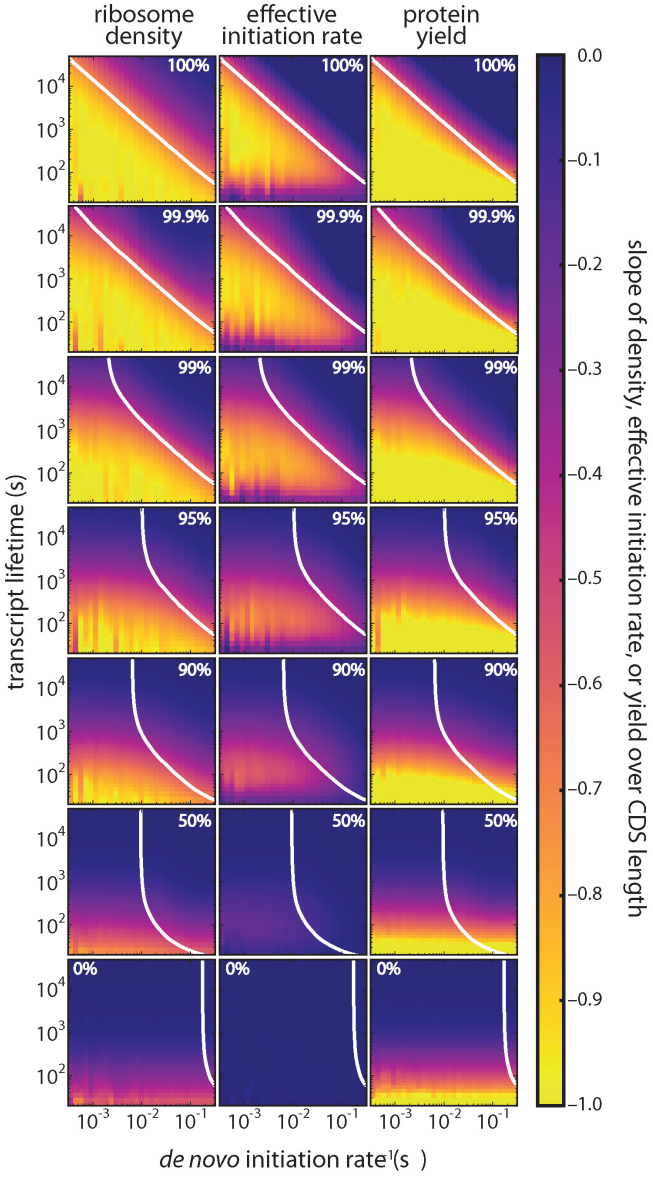

Figure 2.S2: Full model predictions of ribosome density, effective initiation rate, and protein yield at different reinitiation levels. We simulated translation for a wide range of transcript lifetimes and *de novo* initiation rates. For any given combination of transcript lifetime and *de novo* initiation rate, we simulated translation for transcripts with different CDS lengths and then calculated the slope of ribosome density, effective initiation rate, and protein yield over CDS length. Slopes are indicated in different colours (see colour bar), reflecting the degree of length-dependence. The white line in each panel shows the *de novo* initiation rate required at each lifetime such that a 400-codon long transcript carries 6 ribosomes. Our model predicts that high levels of reinitiation can cause length-dependent translation across a wide range of transcript lifetimes and *de novo* initiation rates. At high levels of reinitiation, length-dependence is strongest for short transcript lifetimes and low *de novo* initiation rates; since ribosomes are unlikely to leave the transcript, at long lifetimes or high *de novo* initiation rates, all transcripts eventually become saturated with ribosomes. As a result, the number of ribosomes per transcript becomes proportional to CDS length (since longer transcripts can carry more ribosomes) and ribosome density, the effective initiation rate, and protein yield become similar for all CDS lengths. As the reinitiation level falls, the effective initiation rate becomes dominated by the *de novo* initiation rate, which is the same for all CDS lengths, and ribosome density, the effective initiation rate, and protein yield all become independent of CDS length. In a linear model of translation (0% reinitiation), negative slopes for density and yield are only seen for extremely short transcript lifetimes simply because the first initiating ribosome fails to reach the stop codon on long transcripts before the maximum lifetime is reached. This effect disappears when transcript lifetimes exceed the time required for the first ribosome to reach the stop codon of the longest transcript (~400 seconds).

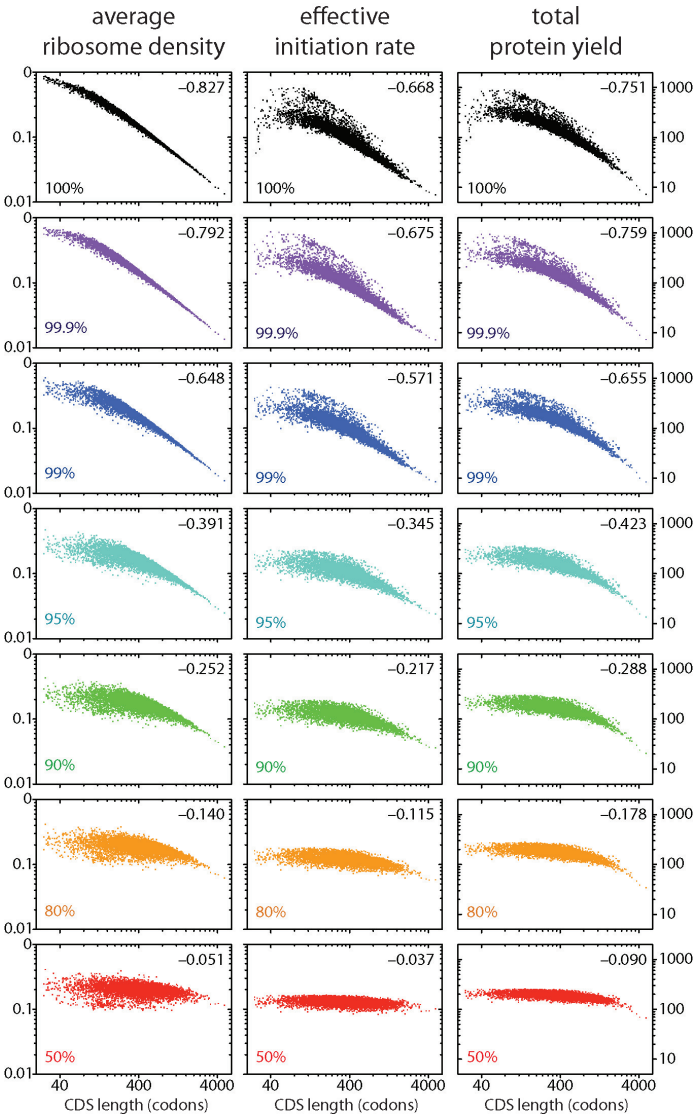

Figure 2.S3: Simulating translation in the budding yeast *S. cerevisiae* at different reinitiation levels. Predicted ribosome density, effective initiation rate, and protein yield for all 5888 budding yeast transcripts simulated at different reinitiation levels using codon-specific decoding rates from (Gilchrist and Wagner 2006). The left y-axis scale applies to ribosome density (left) and effective initiation rates (center); the right y-axis scale applies to protein yield (right). Slopes are indicated in the top-right corner. As in Fig. 2.6, all transcripts had a fixed lifetime of 1553 seconds. *De novo* initiation rates were adjusted at each reinitiation level so that a 400-codon long transcript with a fixed decoding rate of $10\,s^{-1}$ carried an average of 6 ribosomes. Pearson correlation coefficients between simulated results and empirical measures for each reinitiation level are: (i) ribosome density (n = 5542): 100% = 0.935, 99.9% = 0.932, 99% = 0.908, 95% = 0.784, 90% = 0.619, 80% = 0.377, 50% = 0.083; (ii) effective initiation rate (n = 5348): 100% = 0.742, 99.9% = 0.742, 99% = 0.729, 95% = 0.671, 90% = 0.591, 80% = 0.468, 50% = 0.289; (iii) protein yield (n = 4933): 100% = 0.480, 99.9% = 0.478, 99% = 0.475, 95% = 0.458, 90% = 0.449, 80% = 0.442, 50% = 0.438. Empirical data sources as in Fig. 2.2.

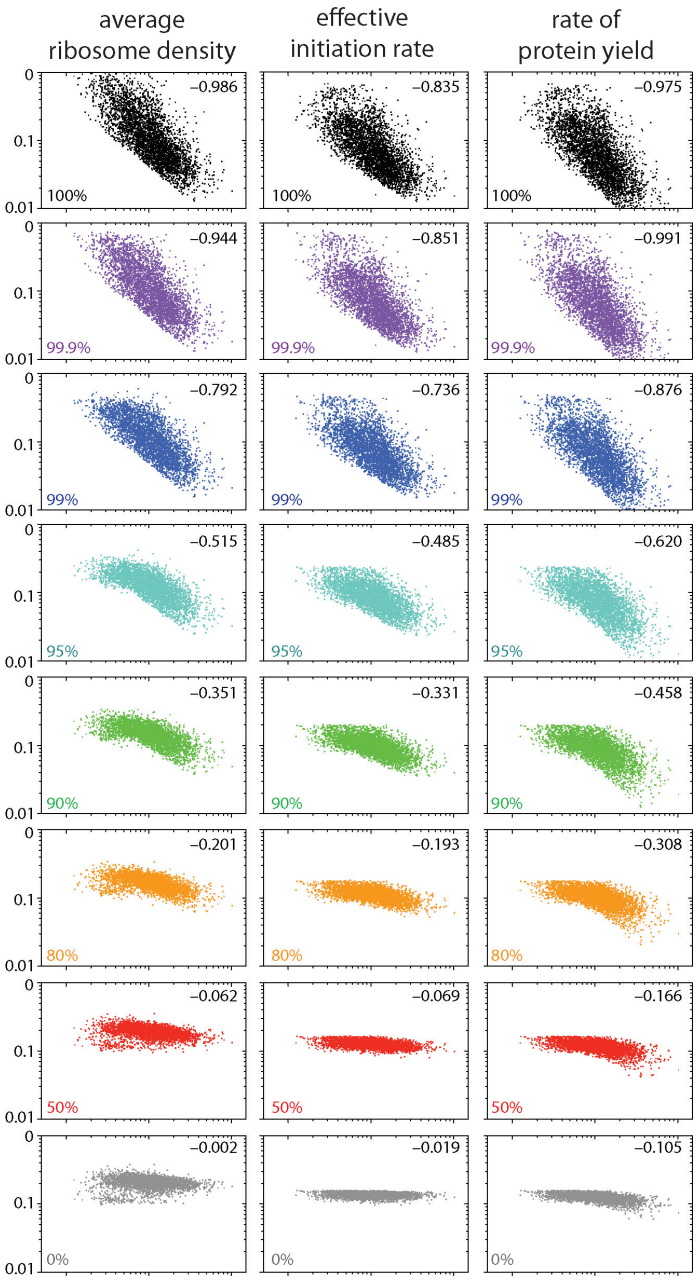

Figure 2.S4: Simulating translation in the budding yeast *S. cerevisiae* at different reinitiation levels using empirical estimates of transcript lifetime. The simulations shown in Fig. 2.S3 were repeated using empirical estimates of transcript lifetimes. Transcript lifetimes were calculated using relative abundances from (Lipson et al. 2009) (measured using single-molecule sequencing digital gene expression which eliminates the length-bias associated with RNA-Seq), a total of 36,000 transcripts per cell (Miura et al. 2008), and nascent transcription rates from (Pelechano et al. 2010). For each transcript, absolute abundance was divided by the transcription rate to obtain the average lifetime. Transcripts with average lifetimes below 400s were excluded to prevent bias towards increased length-dependence (see Fig. 2.S2). Pearson correlation coefficients between simulated results and empirical measures for each reinitiation level are: (i) ribosome density (n = 3829): 100% = 0.767, 99.9% = 0.771, 99% = 0.773, 95% = 0.741, 90% = 0.673, 80% = 0.512, 50% = 0.151, 0% = -0.072; (ii) effective initiation rate (n = 3752): 100% = 0.668, 99.9% = 0.668, 99% = 0.668, 95% = 0.654, 90% = 0.627, 80% = 0.567, 50% = 0.394, 0% = 0.182; (iii) protein yield (n = 3451): 100% = 0.615, 99.9% = 0.620, 99% = 0.609, 95% = 0.578, 90% = 0.561, 80% = 0.546, 50% = 0.535, 0% = 0.527. Empirical data sources as in Fig. 2.2.

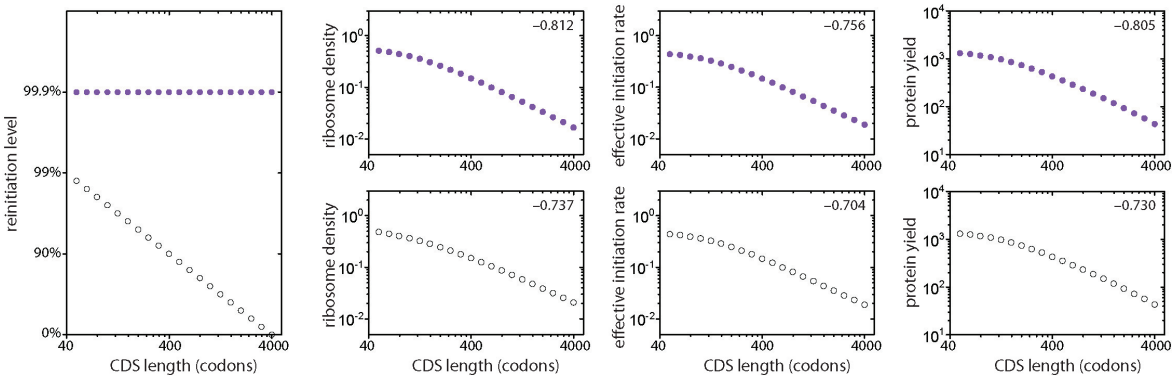

Figure 2.S5: Length-dependent translation only requires high levels of reinitiation on short transcripts. Decreasing reinitiation levels with transcript length (open circles) produces very similar relationships between CDS length and ribosome density, effective initiation rate, and protein yield as does a fixed reinitiation level of 99.9% (purple circles). Slopes of each relationship are shown in the upper right corner of each panel. Length-dependent reinitiation levels capture length dependent translation at much lower reinitiation levels than that required using a fixed reinitiation level. We have arbitrarily chosen to decrease the reinitiation level as a function of CDS length according to the formula $100\% \times (1 - \text{CDS length}/4000)$. The *de novo* initiation rate was set such that a 400-codon transcript carried an average of 6 ribosomes (99.9% reinitiation = $0.004\,58\,\text{s}^{-1}$, length-dependent reinitiation = $0.022\,85\,\text{s}^{-1}$).

Amino acid	Codon	Decoding rate	Amino acid	Codon	Decoding rate	Amino acid	Codon	Decoding rate
ALA	GCU	18.2	GLY	GGC	27.0	PRO	CCA	15.6
ALA	GCC	11.7	GLY	GGA	4.69	PRO	CCG	15.6
ALA	GCA	7.82	GLY	GGG	3.13	SER	UCU	22.4
ALA	GCG	7.82	HIS	CAU	8.43	SER	UCC	14.4
ARG	CGU	4.21	HIS	CAC	13.8	SER	UCA	7.28
ARG	CGC	2.70	ILE	AUU	20.3	SER	UCG	1.56
ARG	CGA	2.70	ILE	AUC	13.0	SER	AGU	3.82
ARG	CGG	1.56	ILE	AUA	3.13	SER	AGC	6.26
ARG	AGA	17.2	LEU	UUA	10.9	THR	ACU	17.4
ARG	AGG	1.72	LEU	UUG	19.2	THR	ACC	11.2
ASN	AAU	9.56	LEU	CUU	0.956	THR	ACA	6.26
ASN	AAC	15.6	LEU	CUC	1.56	THR	ACG	1.56
ASP	GAU	15.5	LEU	CUA	9.00	TRP	UGG	11.7
ASP	GAC	25.3	LEU	CUG	9.00	TYR	UAU	10.4
CYS	UGU	4.56	LYS	AAA	6.70	TYR	UAC	17.0
CYS	UGC	7.47	LYS	AAG	15.1	VAL	GUU	19.3
GLN	CAA	14.1	MET	AUG	7.82	VAL	GUC	12.4
GLN	CAG	1.56	PHE	UUU	8.9	VAL	GUA	3.13
GLN	GAA	18.4	PHE	UUC	14.6	VAL	GUG	2.87
GLN	GAG	3.13	PRO	CCU	3.13			
GLY	GGU	16.5	PRO	CCC	2.01			

Table 2.S1: *S. cerevisiae* codon-specific elongation rates from (Gilchrist and Wagner 2006).

2.S.1. Parameter estimates and justification

Transcript lifetime, age distribution and mode of decay The decay of eukaryotic transcripts is initiated by the stepwise removal of adenine residues from the poly(A) tail (Cao and Parker 2001). Deadenylation is the slowest step in mRNA decay, and during this process there is no apparent decay of the transcribed portion of the mRNA (Dreyfus and Régnier 2002; Chen et al. 2008). Enzymatic digestion of the poly(A) tail is distributive: the deadenylase binds to a poly(A) tail, cleaves a small number of residues, then dissociates from the transcript before binding to a different poly(A) tail (Chen and Shyu 2011). Thus deadenylation consists of a series of sequential first-order reactions, resulting in a hypo-exponential distribution of full-length transcript lifetimes (Cao and Parker 2001; Decker and Parker 1993). Hypo-exponential distributions, by definition, have lower variances than an exponential function with the same mean lifetime and age-matched populations of eukaryotic transcripts are expected to show very little degradation for long periods then decay rapidly. When the number of sequential deadenylation steps is high (~30), the distribution of transcript lifetimes becomes symmetrical and approximates the normal distribution (Kuo et al. 2006). We have therefore, for computational simplicity, assumed that all transcripts in a given simulation have the same fixed lifetime.

In our model, all translation ceases as soon as the transcript lifetime is reached; ribosomes that have a new round of translation are not allowed to finish and therefore do not contribute to the protein yield. Consequently, protein yield in our model does not exactly match the effective initiation rate. There is evidence that eukaryotic mRNA decay is co-translational: transcript degradation in the 5'-3' direction follows the last translating ribosome allowing all ribosomes to complete their final round (Pelechano et al. 2015). If this is the sole mechanism of eukaryotic transcript decay, then the true yield becomes the effective initiation rate multiplied by the transcript lifetime (a constant in our model).

De novo initiation rates

The average number of ribosomes per transcript is remarkably constant across eukaryotes. Polysome profiling studies estimate that the median *S. cerevisiae* transcript contains 6.0 ribosomes (based on weighted averages for all measured transcripts (Arava et al. 2003); an identical average was calculated in (MacKay et al. 2004)) while the average human embryonic kidney HEK293T cell contains 5.6-6.1 ribosomes (Hendrickson et al. 2009). Direct counts of the numbers of ribosomes in polysomes (Lauria et al. 2015) using atomic force microscopy provide similar values (*S. cerevisiae* = 6.5 ribosomes per transcript, HEK293T = 8.7 ribosomes per transcript, human MCF-7 cells = 8.3 ribosomes per transcript). Direct polysome counts are almost certainly over-estimates of the average number of ribosomes per transcript as they ignore unoccupied transcripts (approximately 15-30% of mRNAs (Arava et al. 2003; Hendrickson et al. 2009)) and mRNAs occupied by a single ribosome. To be consistent with these counts, in all of our models, for any transcript lifetime and reinitiation probability,

we have adjusted the *de novo* initiation rate such that the average transcript (CDS length = 400 codons) carries 6 ribosomes (white line on each heatmap in Fig. S2). This causes the *de novo* initiation rate to decrease with increasing reinitiation level and decreasing transcript lifetimes.

Maintaining six ribosomes on a 400-codon-long transcript for transcripts with different lifetimes requires much larger changes to the *de novo* initiation rate at high reinitiation probabilities compared to low reinitiation levels (Fig. S2). Under perfect reinitiation, the lifetime protein yields of transcripts of a given CDS length will be similar, but the rate of protein production over time (proteins per mRNA per unit time) will be higher for short-lived transcripts compared to long-lived transcripts. Thus, organisms with short-lived transcripts, such as yeast, will exhibit much higher protein synthesis rates than organisms with long-lived transcripts, such as mammals. Current estimates suggest that protein production rates in mammals are considerably lower than in yeast, consistent with the much higher doubling rates and much lower protein stabilities in yeast compared to mammals (Jackson et al. 2000; von der Haar 2008; Schwanhäusser et al. 2011). In contrast, assuming similar elongation rates across species, linear models predict that yeast and mammals should have similar protein production rates since they have similar average ribosome densities.

Model parameters

Full model. We explored the consequences of different reinitiation levels on the average ribosome density, effective initiation rate and protein yield for transcripts with different CDS lengths using a wide range of transcript lifetimes and *de novo* initiation rates (Fig. S2). For each combination of transcript lifetime and *de novo* initiation rate, we simulated translation for transcripts of the following log-uniform distributed CDS lengths (in codons): 50, 63, 79, 100, 126, 158, 200, 251, 316, 399, 502, 632, 796, 1002, 1263, 1590, 2002, 2522, 3176, and 4000, and then calculated the slope of the resulting estimates of ribosome density, effective initiation rate, and protein yield over CDS length. All codons were decoded at a rate of $10s^{-1}$ based on the average level in yeast (Waldron et al. 1974) and similar to the average rate observed in a mouse embryonic cell line ($5.6s^{-1}$ (Ingolia et al. 2011)). Termination rates (the sum of the release rate and the reinitiation rate) are set equal to the elongation rate at $10s^{-1}$. Different reinitiation levels are achieved by setting the reinitiation rate as the corresponding proportion of the termination rate. Translation of each transcript was averaged over 1000 runs.

General model. To explore the consequences of different reinitiation levels in more detail, we present a model using an arbitrary lifetime of 3000s (50 minutes) for all transcripts (Figs. 2-5). This value is intermediate between estimates of median transcript half-lives in yeast (10-30 minutes Geisberg et al. 2014) and mammalian cell lines (300-600 minutes Geisberg et al. 2014). Simulations were performed with a constant elongation rate of $10s^{-1}$ (except for Fig. S3 where we perform the same simulation at $5s^{-1}$ and $20s^{-1}$). The model is otherwise the same as the full model, except that a single *de novo* initiation rate is used at each reinitiation level. *De*

novo initiation rates are adjusted for each reinitiation level such that a 400-codon-long transcript carries an average of 6 ribosomes. The *de novo* initiation rates used with a transcriptome-wide elongation rate of $10s^{-1}$ were: 100% = $0.00438s^{-1}$, 99.9% = $0.00458s^{-1}$, 99% = $0.00586s^{-1}$, 95% = $0.01289s^{-1}$, 90% = $0.02285s^{-1}$, 80% = $0.04199s^{-1}$, 50% = $0.09570s^{-1}$, 0% = $0.17578s^{-1}$.

Yeast-specific model. We computed the average ribosome density, effective initiation rate, and protein yield (Fig. 6) for 5888 *S. cerevisiae* transcripts ranging in CDS length from 16 to 4910 codons (median length = 405 codons) in our model. We used the codon-specific elongation rates calculated by Gilchrist and Wagner 2006; these rates are scaled such that the average elongation rate is $10s^{-1}$. As above, we adjusted *de novo* initiation rates for each reinitiation level such that a 400-codon-long transcript (ignoring variation in decoding rates) contained an average of 6 ribosomes. The exact *de novo* initiation rates used were: 100% = $0.00859s^{-1}$, 99.9% = $0.00869s^{-1}$, 99% = $0.01016s^{-1}$, 95% = $0.01641s^{-1}$, 90% = $0.02568s^{-1}$, 80% = $0.04492s^{-1}$, 50% = $0.09766s^{-1}$, 0% = $0.17578s^{-1}$.

Most studies of mRNA stability report transcript half-lives. If eukaryotic transcripts decay with biphasic (slow-then-fast) kinetics, then transcript lifetimes cannot be calculated from observed half-lives by assuming first-order kinetics (Chen et al. 2008). We have therefore based our estimate of transcript lifetime on a study of nascent transcription rates in *S. cerevisiae* which estimated that the entire set of mRNAs in a cell turns over more than four times per 6780s (113 minute) cell cycle (Pelechano et al. 2010), resulting in an average transcript lifetime of 1553s (26 minutes). Although most yeast studies predict fairly similar median transcript half-lives, gene-specific estimates show little correlation across studies (Geisberg et al. 2014). Consequently, we have made the simplifying assumption that all transcripts have the same 1553s lifetime.

2.S.2. Author contributions

Conceptualization: DWR MAB AT DG.
Data curation: DWR MAB.
Formal analysis: MAB AT DWR.
Funding acquisition: AT DG.
Investigation: DWR MAB.
Methodology: DWR MAB AT.
Project administration: DWR MAB AT DG.
Resources: AT DG.
Software: MAB AT.
Supervision: AT DG.
Validation: DWR MAB.
Visualization: DWR MAB AT DG.
Writing – original draft: DWR MAB.
Writing – review & editing: DWR MAB AT DG.

3. Cell replicative ageing

Somatic evolution and replicative ageing

In most tissues of multicellular organisms there is a cellular turnover, which means that cells constantly die and get replenished by the birth of new cells via cell divisions. Since on each cell division the whole genome is copied, this confers a large risk for mutations. Very often these mutations are insignificant as they occur in regions of the genome which are not processed into proteins in general (introns, e.g. Berget et al. 1977), are not expressed in the specific tissue (Gurdon 1992) or render the cell completely dysfunctional which in turn often leads to cell death. In rare occasions, however, these mutations enhance the proliferative potential of the cell which causes the expansion of abundance of the clone and can ultimately lead to cancer.

The distribution of replicative age in a cell population is strongly correlated to the genetic variation inside a cellular subpopulation (see section 1.1.2), which itself is one of the main components of (somatic) evolution. One way to reduce variation and mutational load on tissue and to suppress somatic evolution is to limit the number of cell divisions. A central function of tissue structure is therefore to control the replicative age of cells and to reduce the number of cell divisions in the longest lasting stem cells. A hierarchical tissue structure with only few long lasting stem cells, but increasingly more cells in each progenitor compartment, excels at this task (Michor et al. 2003b; Derényi and Szöllősi 2017), as in the long term only the slowly dividing stem cells will prevail in the organism.

Since replicative ageing of stem cells is thought to be one of the major determinants of organismal ageing (Rando 2006) and also for the accumulation of somatic mutations, a lot of models and experimental work concentrates on stem cells. By contrast, in this chapter I will present how differentiation steps after the stem cell compartment can change the replicative age distributions away from the stem cell compartment. Additionally, I will show how measuring the replicative age distributions in differentiated tissue can give insights into the dynamical properties of a hierarchical tissue structure.

Publication information The rest of this chapter is currently under revision. A preprint of an earlier version is available: Böttcher, M. A., B. Werner, D. Dingli, and

A. Traulsen (2018b). "Replicative cellular age distributions in compartmentalised tissues". *bioRxiv preprint*, pp. 1–27.

Replicative cellular age distributions in compartmentalised tissues

Marvin A. Böttcher[1], David Dingli[2], Benjamin Werner[3*], Arne Traulsen[1]

[1]Department of Evolutionary Theory, Max Planck Institute for Evolutionary Biology, Plön, Germany
[2]Division of Hematology, Department of Internal Medicine, Mayo Clinic, Rochester, MN, USA
[3]Centre for Evolution and Cancer, The Institute of Cancer Research, London, UK

*Corresponding author: benjamin.werner@icr.ac.uk

Abstract The cellular age distribution of hierarchically organized tissues can reveal important insights into the dynamics of cell differentiation and self-renewal and associated cancer risks. Here, we examine the effect of progenitor compartments with varying differentiation and self-renewal capacities on the resulting observable distributions of replicative cellular ages. We find that strongly amplifying progenitor compartments, i.e. compartments with high self-renewal capacities, substantially broaden the age distributions which become skewed towards younger cells with a long tail of few old cells. In contrast, if tissues are organised into many downstream compartments with low self-renewal capacity, the shape of the replicative cell distributions in more differentiated compartments are dominated by stem cell dynamics with little added variation. In the limiting case of a strict binary differentiation tree without self-renewal, the shape of the output distribution becomes indistinguishable from the shape of the input distribution. Our results suggest that a comparison of cellular age distributions between healthy and cancerous tissues may inform about dynamical changes within the hierarchical tissue structure, i.e. an acquired increased self-renewal capacity in certain tumours. Furthermore, we compare our theoretical results to telomere length distributions in granulocyte populations of ten healthy individuals across different ages, highlighting that our theoretical expectations agree with experimental observations.

3.1. Introduction

Many tissues in multicellular organisms resemble a compartmentalised structure with a hierarchy of cells at different stages of differentiation and function. This hierarchy

is usually fuelled by a few stem cells that ideally can self-renew indefinitely, whereas the majority of the tissue consists of shorter-lived differentiated cells that emerge from these stem cells (Dingli et al. 2007; Tumbar et al. 2004; Blanpain and Fuchs 2009).

In most tissues it is thought that stem cells divide infrequently, while their progenitors and further differentiated cells divide more frequently to ensure tissue function under homeostasis (Busch et al. 2015). Such structures allow both the production of many cells in a short time and the reduction of the risk for the accumulation of somatic mutations within the stem cell compartment (Michor et al. 2003a; Nowak et al. 2003; Michor et al. 2003b; Dingli et al. 2007; Werner et al. 2011; Werner et al. 2013; Derényi and Szöllősi 2017).

Due to these pronounced dynamical disparities in hierarchical tissues, replicative age — the number of divisions a cell has undergone — can be an important observable providing information about the structure and cellular dynamics within these tissues. Since many somatic mutations are acquired during cell divisions (Ju et al. 2017; Williams et al. 2016), we would expect replicative age also to be strongly correlated with different cancer risks in different hierarchical tissues (Tomasetti and Vogelstein 2015; Adams et al. 2015; Noble et al. 2015). In the context of ageing, the focus is typically on changes within the stem cell compartment, since stem cells have the ability to self-renew and persist long enough to become relevant for organismal ageing (Rando 2006; Rossi et al. 2008). It is generally assumed that replicative cell age in downstream compartments is a good proxy for replicative stem cell age. For example, some of us previously developed and tested a mathematical model for human hematopoietic stem cell ageing based on replicative ages in human lymphocytes and granulocytes (Werner et al. 2015). Here, we do not model the detailed dynamics on the stem cell level. Instead, we regard the cellular age distribution on the stem cell level as a steady influx of progenitor cells into the differentiation hierarchy to keep up homeostasis of the tissue.

Cellular dynamics in hierarchically organised tissue structures can be hard to explore experimentally due to the large scaling difference between differentiation levels (Bystrykh et al. 2012) and the challenges to correctly assign cells to specific differentiation stages. One possibility to assess the age distribution of a cell population is to measure the telomere length of the cellular chromosomes. Telomeres are the protective, non-coding ends of chromosomes, consisting of the same short DNA sequence repeated thousands of times. Telomeres typically shorten with each cell division (Harley et al. 1990; Blackburn 1991; Rodriguez-Brenes and Peskin 2010). Cells with critically short telomeres enter replicative senescence, which is thought to be a cancer suppression mechanism (Collado et al. 2007). Moreover, critically short telomeres are often associated with genome instability and corresponding increased risk of cancer (Blasco 2007). For our purpose, telomere length distributions can be thought of as a measure for the cellular replicative age distribution. These can be assessed in tissue samples (Nussey et al. 2014; Baerlocher et al. 2006) which are for example especially

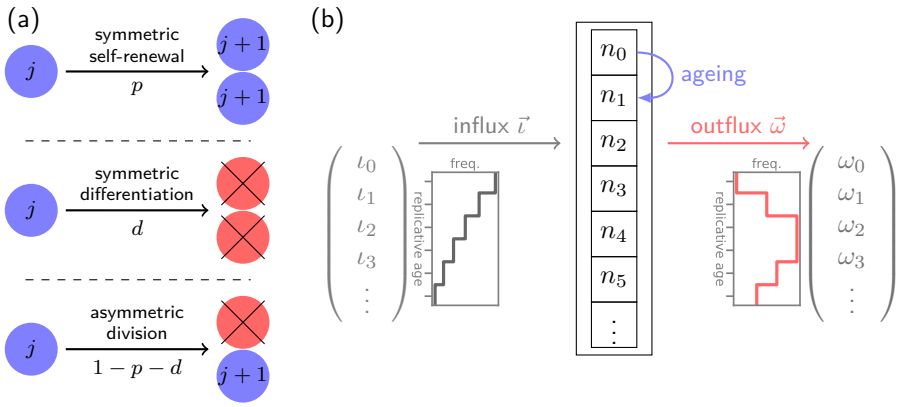

Figure 3.1: Sketch of the basic model. (a) Three different modes of cell division in the focal progenitor compartment. Blue cells are cells within the compartment, red cells differentiated and leave the compartment. The replicative age of a cell in the specific compartment is j, increasing by one in each cell division. (b) Full model for ageing in progenitor compartment. The number of cell in each age class is n_j, and all cells age according to the modes of cell division (a). The cell influx $\vec{\iota}$ into the compartment has a certain distribution of replicative age ι_j. The cell outflux from the compartment includes all differentiating cells and is denoted by the distribution $\vec{\omega}$.

accessible in differentiated tissue in the hematopoietic system and thus in principle would also allow for some time resolution within healthy human individuals (Werner et al. 2015). For simplicity, we concentrate solely on replicative ageing, that is the number of cell divisions a cell has undergone, by contrast to temporal ageing which is also commonly explored in models (Fonseca and Voit 2015).

However, it remains unclear if cellular age distributions in hierarchically organized tissues are dominated by stem cell dynamics or alternatively are determined by the possibility of a multi-step differentiation process with strong intermediate self-renewal of progenitor cells. Here, we develop a mathematical framework that allows us to describe the distribution of replicative cellular ages across several hierarchical levels of differentiation. Thereby, we demonstrate under which conditions the distribution of replicative ages in differentiated cell populations can provide insights into the properties of the dynamics within the underlying tissue.

3.2. Model

In the following, we present a mathematical description for the replicative age distributions within compartmentalized tissue structures (fig. 3.1). First, we discuss the simplest case of only two compartments - one stem cell compartment and the focal downstream progenitor compartment. We then ask what is the distribution of replicative ages of cells in the progenitor compartment provided a continuous influx

of cells from the stem cell compartment. For example, it is estimated that in hematopoiesis of mice there is a constant production of early progenitor cells from stem cells with a rate of around 150 cells per day (Busch et al. 2015). However, we do not discuss the time dynamics on the stem cell level explicitly. The temporal change of replicative age distributions in stem cell compartments and the resulting potential influx distributions for progenitor compartments are discussed in detail in Werner et al. 2015.

We assume that in the progenitor compartment there are n_j cells of each replicative age class j. Progenitor cells divide with proliferation rate r and after each division the replicative age of both daughter cells increases by one $j \to j+1$. Each daughter cell can in principle take a different cell fate that contributes differently to the distribution of replicative ages (fig. 3.1A). In general, the following outcomes are possible after a single cell division

(i) With probability p a cell self-renews symmetrically, both daughter cells stay in the same compartment and increase their cellular age by one ($n_j \to n_j - 1, n_{j+1} \to n_{j+1} + 2$)

(ii) With probability d a cell differentiates symmetrically, effectively removing it from the compartment of differentiated cells ($n_j \to n_j - 1$).

(iii) With probability $1-p-d$ a cells divides asymmetrically, with one cell staying in the pool of differentiated cells while the other cell leaves the compartment (Knoblich 2008) ($n_j \to n_j - 1, n_{j+1} \to n_{j+1} + 1$).

We choose the influx of cells from the stem cell compartment to be a constant rate ι_j that might differ for each cellular age j. Below we give explicit examples for different distributions of ι_j. We assume the dynamics on the stem cell level to be much slower compared to downstream compartments and hence consider the influx ι_j into the progenitor compartment to be constant over time.

Using the above, we can formulate differential equations for the change of the number of cells in each age class n_j. Thereby, we account for the loss of cells due to proliferation and subsequent differentiation and gain of cells due to symmetric self-renewal and cell influx from the stem cell compartment,

$$\frac{\partial n_j}{\partial t} = \begin{cases} \iota_0 - r n_0 & j = 0 \\ \iota_j - r n_j + \underbrace{(1+p-d)}_{\alpha} r n_{j-1} & j \geq 1 \end{cases}, \tag{3.1}$$

where we set $\alpha = 1+p-d$ to be the self-renewal parameter which critically determines the most relevant results of our model. Since p and d are probabilities with $p+d \leq 1$, the self-renewal parameter can be in the range $0 \leq \alpha \leq 2$. However, since we are interested in homeostasis and not an exponentially growing tissue, the symmetric division probability p in our case must be smaller than the symmetric differentiation probability d and therefore $0 \leq \alpha < 1$.

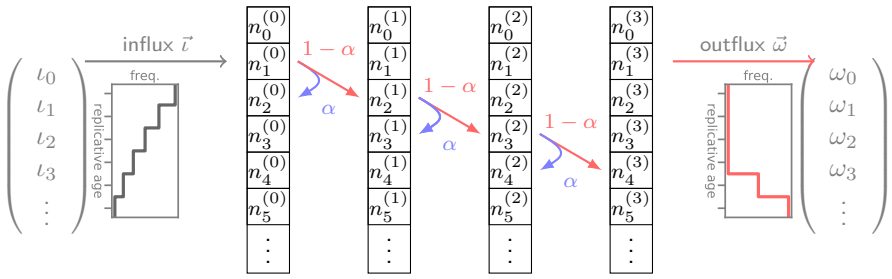

Figure 3.2: Several downstream compartments amplifying the rate of cell production from influx to outflux. In each compartment there are self-renewal or differentiation processes as described in figure 3.1. Each cell division thereby leads to an increase of replicative age and changes the age distribution of the corresponding compartment. Self renewal occurs proportional to the self-renewal parameter α, whereas differentiated cells are produced with $1-\alpha$ and go into the next downstream compartment. The compartment number c is shown as superscript, the total number of compartments is $C = 4$.

The above system of ordinary differential equations can be solved analytically (see appendix 3.E). However, as we assume that the dynamics on the level of stem cells is much slower compared to progenitor compartments we can investigate the equilibrium solutions n_j^* to equation 3.1 for each age class j. The equilibrium solutions can be obtained via recursion by setting $\frac{\partial n_j}{\partial t} = 0$ (see appendix 3.A). The general solution is

$$n_j^* = \sum_{k=0}^{j} \frac{\iota_k}{r} \alpha^{j-k} \tag{3.2}$$

which is equivalent to a convolution sum of the influx ι_k and $\frac{\alpha^k}{r}$ between zero and j.

3.2.1. Multiple compartments

In reality, most tissues will consist of multiple progenitor cell compartments. It is thus natural to ask how multiple downstream compartments affect cellular age distributions. To answer this question, we can generalise our previous framework. Differentiated cells in a downstream compartment are either produced by symmetric differentiation with probability d or by asymmetric division with probability $1-p-d$. If we denote the output of cells per unit of time for each age class as ω_j, we can write

$$\begin{aligned} \omega_j &= (1-p-d) r n_{j-1} + 2 d r n_{j-1} \\ &= (2-\alpha) r n_{j-1}. \end{aligned}$$

To allow for multiple compartments, we can identify the output distribution of a compartment c and the input distribution of the next downstream compartment

$c+1$,

$$\iota_j^{(c+1)} = \omega_j^{(c)}. \tag{3.3}$$

Total cell outflux For our purpose it is desirable to compare the effect of different tissue structures, that is a different number of total compartments C, but with the same tissue function, that is the same total outflux of fully differentiated cells. In our model, the total outflux of differentiated cells $\Omega = \sum_j \omega_j$ is determined by the total influx of cells $I = \sum_j \iota_j$, the number of compartments C and the self-renewal parameter α. We therefore choose α such that the total output of cells remains constant, i.e. assuring certain replenishing needs of a specific tissue. For this we formulate differential equations for the change of the total number of cells $N^{(c)} = \sum_j n_j^{(c)}$ in each of the compartments c with a compartment specific proliferation rate for each cell $r^{(c)}$ by collecting all influx and outflux terms:

$$\begin{aligned}
\frac{\partial N^{(0)}}{\partial t} &= (\alpha - 1) r^{(0)} N^{(0)} + I \\
\frac{\partial N^{(c)}}{\partial t} &= (\alpha - 1) r^{(c)} N^{(c)} + (2 - \alpha) r^{(c-1)} N^{(c-1)}.
\end{aligned}$$

Here, I is the total influx into the first compartment ($c = 0$) (i.e. the sum of all direct stem cell derived progenitors per time unit). The total outflux Ω is related to the number of cells in the last compartment $N^{(C-1)}$ via

$$\Omega = (2 - \alpha) r^{(C-1)} N^{(C-1)}.$$

Under steady state conditions, the above equations can be solved explicitly for the self-renewal parameter α (see appendix 3.B):

$$\alpha = \frac{\sqrt[C]{\frac{\Omega}{I}} - 2}{\sqrt[C]{\frac{\Omega}{I}} - 1}. \tag{3.4}$$

This allows us to adjust the self-renewal parameter α such that the outflux Ω remains constant given an influx I for any number of compartments C. However, since the self-renewal parameter is constrained $0 \leq \alpha < 1$ (see above section) the minimum amplification of cell production is given by $\left(\frac{\Omega}{I}\right)_{min} = 2^C$ corresponding to $\alpha = 0$.

3.2.2. Properties of the replicative age distribution

Mean and variance The mean and variance of the replicative age distribution under steady state conditions can be calculated analytically, see appendix 3.C. The

mean μ of the replicative age distribution in the progenitor compartment increases compared to the influx based on the self-renewal α to

$$\mu = \langle j \rangle_{n^*} = \langle j \rangle_\iota + \frac{\alpha}{1-\alpha} = \mu_\iota + \frac{\alpha}{1-\alpha},$$

where $\langle j \rangle_{n^*}$ is the first moment of the replicative age distribution in the focal progenitor compartment and $\langle j \rangle_\iota = \mu_\iota$ is the average replicative age of the influx. Note that the average replicative age of the outflux $\mu_\omega = \langle j \rangle_\omega$ is increased by one to account for the extra differentiation step

$$\mu_\omega = \langle j \rangle_\omega = \mu + 1 = \frac{1}{1-\alpha}. \tag{3.5}$$

The minimal increase of the mean between influx and outflux for no self-renewal ($\alpha = 0$) is therefore equal to one.

The variance σ^2 of the replicative age distribution increases similarly as the mean above

$$\sigma^2 = \langle j^2 \rangle_\iota - \langle j \rangle_\iota^2 + \frac{\alpha}{(1-\alpha)^2} = \sigma_\iota^2 + \frac{\alpha}{(1-\alpha)^2}. \tag{3.6}$$

Here, σ_ι^2 denotes the variance of the replicative age distribution of the influx.

Generally, also the higher moments $\langle j^\gamma \rangle_{n^*}$ of the replicative age distribution can be calculated based on the moments of the influx distribution $\langle j^\beta \rangle_\iota$ with $\beta \leq \gamma$. The corresponding calculations and results are shown in the appendix section 3.C.

Limiting behavior For very low self-renewal, $\alpha \ll 1$, the only age class of influx that significantly contributes to the age distribution n_j^* in equation 3.2 is $\iota_{k=j}$, as it is in zeroth order of α. The influx of all other age classes is of higher order of self-renewal α and will therefore vanish for $\alpha \ll 1$ such that

$$n_j^* \approx \frac{\iota_j}{r}.$$

Hence, the outflux distribution will look approximately like the influx distribution.

To evaluate the impact of the progenitor compartment on the replicative age distribution in the limit of high self-renewal $1 - \alpha \ll 1$, we rewrite equation 3.2 to

$$n_j^* = \frac{\alpha^j}{r} \sum_{k=0}^{j} \frac{\iota_k}{\alpha^k}.$$

The limiting behavior therefore strongly depends on the age distribution of the influx ι_k. If the influx has an upper bound K on replicative age, such that for all $k \geq K$

holds $\iota_k \ll \alpha^k$, the sum in the above equation is constant and the distribution of replicative age will decline exponentially

$$n_j^* \propto \frac{\alpha^j}{r} \quad \text{for} \quad j \geq K.$$

If, on the other hand, the influx distribution ι_k is not declining fast enough and is in the same order as α^k ($\iota_k \geq \alpha^k$), we cannot make a general prediction for this limit.

3.3. Results

It seems natural to suspect that the specific distribution of replicative ages in downstream compartments strongly depends on the distribution of cellular ages within the stem cell compartment. In the following, we present the resulting age distributions for various different influx distributions. Additionally we will compare tissue structures with many subsequent downstream compartments and a low probability for self-renewal against having only very few compartments with a high probability for self-renewal.

An important parameter for the age distribution in the progenitor compartment is $\alpha = 1 + p - d$ which depends on both the probability for symmetric splitting and symmetric differentiation and critically defines the total size of the compartment as well as the amount of cells produced (appendix 3.B). For a compartment model of hematopoiesis with many differentiation steps as for example in (Dingli et al. 2007; Lenaerts et al. 2010), α would be around 0.3, whereas for other models with fewer compartments α would need to be higher to allow for sufficient output of fully differentiated cells per unit time (Michor et al. 2005; Marciniak-Czochra et al. 2009a; Dingli and Michor 2006).

3.3.1. A single progenitor compartment

Here we discuss the distributions of replicative age in the special case of a single progenitor compartment given four different influx distributions from the stem cell compartment. All distributions are calculated analytically and the corresponding calculations can be found in the appendix 3.D.1 and 3.3.1. Realisations of the resulting replicative age distributions are shown in figure 3.3.

Identical replicative cellular age influx

We first discuss the simplest case for a cellular age distribution on the stem cell level that is all stem cells have identical replicative age v. This results in a delta function

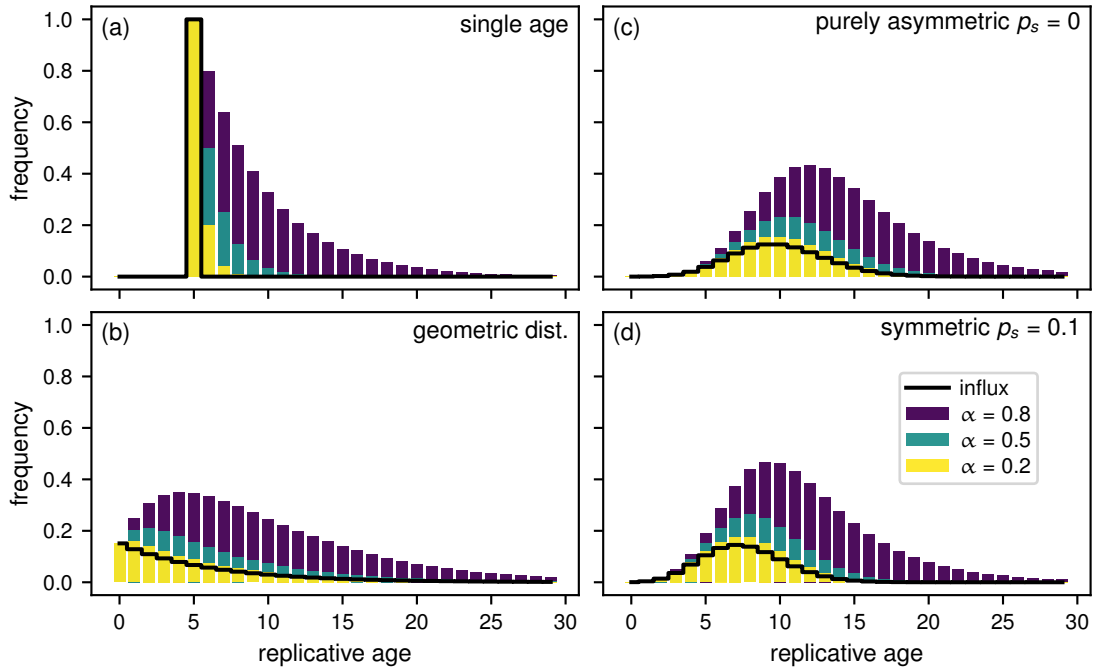

Figure 3.3: Distributions of replicative age in the first progenitor compartment for varying influx distributions from the stem cell compartment. (a) Influx of only a single replicative age $\iota_k = \delta(k - v)$ with parameter $v = 5$. (b) Influx given by a geometric distribution with many young and few old cells $\iota_k = \lambda^k(1 - \lambda)$. The distribution parameter is $\lambda = 0.85$. (c) Model based influx for purely asymmetric divisions on the stem cell level resulting in a Poisson distribution $\iota_k = \frac{\lambda^k}{k!}e^{-\lambda}$ (Werner et al. 2015). Parameter $\lambda = 10$. (d) A model based influx with symmetric divisions (probability $p_s = 0.1$) also result in differently normalized Poisson distribution $\iota_k = t^* \frac{\tilde{\lambda}^k}{k!}e^{-\tilde{\lambda}}$ with a more pronounced difference to the age distribution for purely asymmetric divisions at older ages of the stem cell pool. In (c) and (d) the underlying parameters for λ and $\tilde{\lambda}$ are the same (section 3.3.1 for details).

input $\iota_k = r_s\delta(k - v)$, where $\delta(x)$ is the Dirac delta distribution and r_s is the rate of cell production. Together with equation 3.2 this implies for the age distribution

$$\begin{aligned} n_j^* &= \frac{r_s}{r}\sum_{k=0}^{j}\delta(k - v)\alpha^{j-k} \\ &= \begin{cases} \frac{r_s}{r}\alpha^{j-v} & \text{for } j \geq v \\ 0 & \text{else} \end{cases}, \end{aligned} \tag{3.7}$$

The resulting distribution is shown in figure 3.3A. Cellular ages within the single progenitor compartment follow an exponential distribution that approaches zero faster for smaller self-renewal parameters α and always has the maximum at the influx replicative age v.

Geometrically distributed replicative cellular age influx

The former section is of course an oversimplification. We expect some form of distributed cellular ages on the stem cell level. We first discuss the possibility of a geometrically distributed replicative age $\iota_k = r_s\lambda^k(1-\lambda)$ with distribution parameter λ and total cell influx r_s as input from the stem cell level. This replicative age distribution resembles the distribution in the first progenitor compartment for an influx with identical replicative age from the stem cell compartment as shown in the previous section (Sec. 3.3.1); it would therefore correspond to the second downstream compartment for that specific influx.

The resulting age distribution within this progenitor compartment - equation 3.2 - can be solved analytically (see appendix 3.D.1):

$$n_j^* = \begin{cases} \frac{r_s}{r}(1-\lambda)\frac{\lambda^{j+1}-\alpha^{j+1}}{\lambda-\alpha} & \text{for } \alpha \neq \lambda \\ \frac{r_s}{r}(1-\lambda)\alpha^j(j+1) & \text{for } \alpha = \lambda \end{cases}$$

These age distributions are shown for different self-renewal parameters α in figure 3.3B. For low self-renewal, the shape of the replicative age in the progenitor compartment strongly resembles that of the influx distribution, i.e. a monotonically decreasing function of replicative age. For large self-renewal $\alpha \geq 0.5$, however, the distribution of replicative cellular ages in equilibrium becomes increasingly independent of the influx distribution and very similar to the age distributions resulting from other influx distributions, see e.g. figure 3.3 C and D.

Influx from stem cell pool with random stem cell divisions

We previously investigated the dynamics within the stem cell compartment given that stem cell proliferations are independent and distribution times are exponentially distributed (Werner et al. 2015). Once a stem cell is picked for division it either divides symmetrically with probability p_s, resulting in two stem cells, or asymmetrically with probability $1-p_s$, resulting in one progenitor and one stem cell. Now we ask how influx from such a stem cell pool percolates through the hierarchy.

Asymmetric stem cell divisions Exclusively asymmetric divisions ($p_s = 0$) on the stem cell level result in a Poisson distribution of replicative age (Werner et al. 2015) and the corresponding influx into the progenitor compartment is given by

$$\iota_k = r_s N_0 e^{-\frac{r_s t}{N_0}} \frac{\left(\frac{r_s t}{N_0}\right)^k}{k!}.$$

The distribution depends on age t, proliferation rate r_s, as well as the initial number of cells N_0 in the stem cell compartment. We can set $\lambda = \frac{r_s t}{N_0}$ to see that this is a

Poisson distribution multiplied by $r_s N_0$:

$$\iota_k = r_s N_0 \frac{e^{-\lambda}\lambda^k}{k!}$$

with a time dependent rate parameter $\lambda = \lambda(t)$.

The corresponding sum from equation 3.2 can be solved analytically (see appendix 3.D.2) and the distribution of replicative age becomes

$$n_j^* = \frac{r_s}{r} N_0 \frac{\alpha^j e^{\frac{\lambda}{\alpha}(1-\alpha)}}{j!} \Gamma(j+1, \tfrac{\lambda}{\alpha}),$$

where $\Gamma(a,x) = \int_x^\infty t^{a-1}e^{-t}\,\mathrm{d}t$ is the upper incomplete gamma function (Gautschi 1998).

The above distribution of replicative age is shown in figure 3.3C for various values of the self-renewal probability α. The normalization factor $r_s N_0$ is set to one, as this does not change the general shape of the underlying distribution. Similar to our previous observations, the age distribution is heavily skewed towards younger cells. This effect is more pronounced for higher values of α, corresponding to more cells in the compartment.

Symmetric stem cell divisions The age distributions for a growing stem cell compartment due to occasional symmetric stem cell self-renewals with probability $p_s > 0$ is given by (Werner et al. 2015)

$$\iota_j = r_s(1-p_s)\frac{N_0}{j!}\left(\frac{1+p_s}{p_s}\right)^j \sqrt[-p_s]{t^*}\ln^j(t^*)$$

with $t^* = \frac{r_s p_s t}{N_0} + 1$. Here the exact distribution depends explicitly on the initial number of stem cells N_0 and the ageing factor t^*, which itself depends on the relative increase of the stem cell pool size during time t given a symmetric division probability p_s and a proliferation rate r_s. However, the distribution is again a Poisson distribution with a different normalization. This becomes apparent if we substitute $\tilde{\lambda} = \frac{1+p_s}{p_s}\ln(t^*)$ and get

$$\iota_j = r_s(1-p_s)N_0 t^* \frac{e^{-\tilde{\lambda}}\tilde{\lambda}^j}{j!}$$

The solution of the convolution sum in equation 3.2 is therefore the same as for purely asymmetric stem cell divisions and the corresponding calculations are identical (if we exchange $\lambda \to \tilde{\lambda}$) (appendix 3.3.1),

$$n_j^* = \frac{r_s(1-p_s)}{r} N_0 t^* \frac{\alpha^j e^{\frac{\tilde{\lambda}}{\alpha}(1-\alpha)}}{j!} \Gamma(j+1, \tfrac{\tilde{\lambda}}{\alpha}).$$

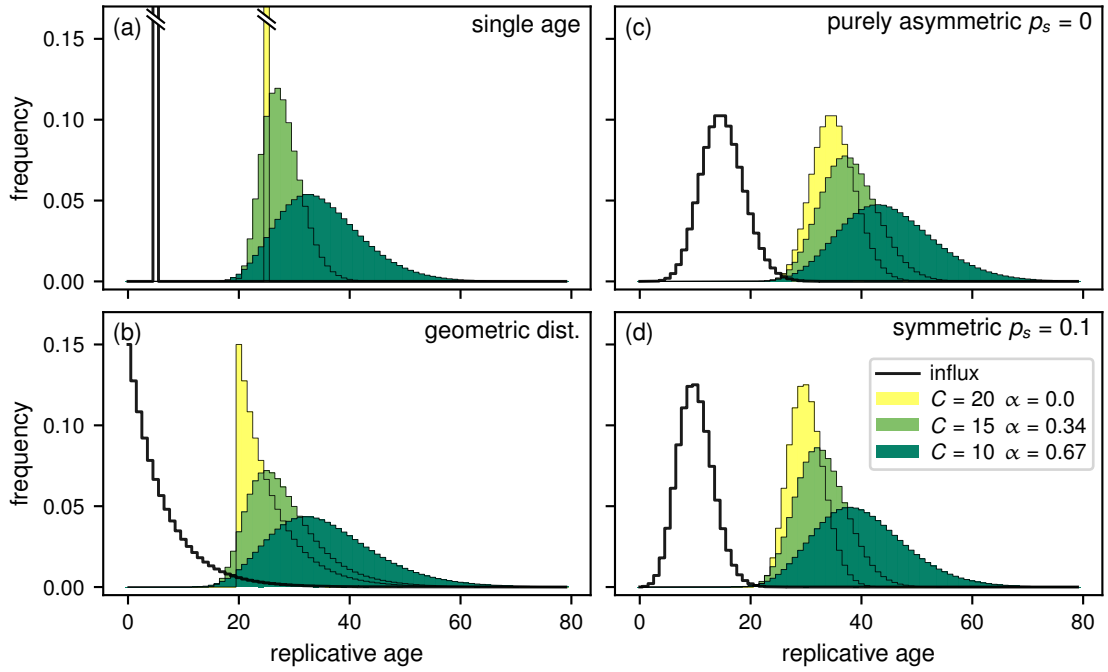

Figure 3.4: Comparison of different total number of progenitor compartments C for different influx age distributions. The self-renewal parameter α is adjusted such the total outflux Ω is the same for each C. The influx distributions are the same as in figure 3.3: (a) Influx with a single age. (b) Influx age geometrically distributed. (c) Model based influx for purely asymmetric divisions on the stem cell level. (d) Model based influx with symmetric divisions. For comparison of the influx $\vec{\iota}$ with the resulting outflux $\vec{\omega}$, the distributions are normalized (all parameters as in figure 3.3).

The shape of the resulting influx distribution therefore varies only slightly from the asymmetric case and differences in the age distribution of the progenitor compartment are minimal (fig. 3.3 C and D). However, the difference in average replicative age on the stem cell level is conserved in the progenitor compartment and still can be used to distinguish between those processes on the stem cell level (Werner et al. 2015).

3.3.2. Multiple compartments

In most organs, the maturation of functional tissue specific cells requires multiple stages of differentiation. We therefore generalize our approach above and discuss the impact of multiple subsequent non-stem cell compartments on the replicative age distribution within such hierarchical tissue organizations.

Impact of the number of compartments In order to deduce the impact of the number of compartments on the age distributions, we vary the number of compart-

ments by simultaneously keeping the final outflux of cells constant. This requires an adjustment of the self-renewal parameter α accordingly and is motivated by the idea that certain tissues might require a certain constant cell replenishment per unit time, but this could in principle be achieved in different tissue architectures. We use the same principal influx distributions from the stem cell compartment discussed above, see figure 3.3. Solutions in this section were obtained by numerically calculating the sums of equation 3.2.

Figure 3.4 shows the resulting replicative age distributions for a broad range of compartment numbers. Interestingly, the age distribution in the final compartment is very sensitive to the number of compartments, even though the total cell number amplification of the compartments is the same by construction. For a large number of compartments and corresponding small self-renewal α the shape of the influx distribution is basically conserved through all stages of the hierarchy, especially for the extreme case of a purely binary tree ($\alpha = 0$) where the shape of the distributions is unchanged but only shifts towards older replicative age.

For the other extreme case of only one or two downstream compartments ($\alpha \approx 1$), the distribution of replicative age is almost flat, such that the frequency of young cells is the same as the frequency of very old cells. Note, however, that in this case the steady state assumption might be violated as the time to reach homeostasis, i.e. the state where the system does not change anymore, might exceed realistic biological time scales. This is shown in the appendix section 3.E in more detail.

However, distributions of replicative age become similar already for intermediate, but biologically still high, values of self-renewal $\alpha \approx 0.5$. It might therefore be impossible to distinguish between age distributions on the stem cell (influx) level from measurement in the differentiated tissue alone, provided there is considerable self-renewal in non-stem cell compartments. This is especially surprising considering the extreme differences in influx distributions, for example delta distributed (fig. 3.4A) and Poisson distributed (fig. 3.4C and 3.4D), which become seemingly undistinguishable in downstream compartments (at equilibrium). This effect is reminiscent of the law of large numbers for random variables, where the sum of independent random variables tends to a normal distribution regardless the actual distribution of the random variable. In our case, though, the distribution approached is not a Gaussian.

Mean and variance through multiple compartments In a system with multiple downstream compartments it is also interesting to see how mean and variance of replicative age change from compartment to compartment. As shown in equations 3.5 and 3.6 for a single progenitor compartment, mean and variance increase linearly from compartment to compartment with a slope of $\frac{1}{1-\alpha}$ for the mean and $\frac{\alpha}{(1-\alpha)^2}$ for the variance. Strong self-renewal therefore has a more pronounced effect on the variance than the mean due to the quadratic term in the denominator.

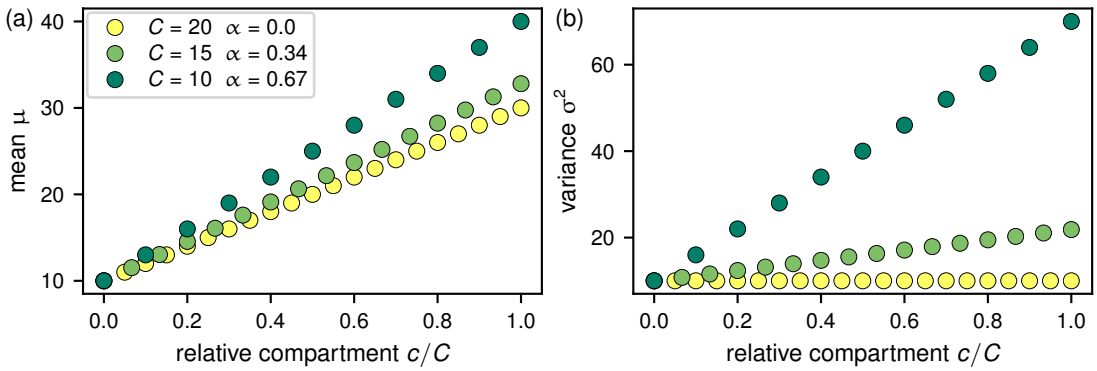

Figure 3.5: Mean μ and variance σ^2 of replicative age distributions per compartment. Influx age distribution is the Poisson distribution with a mean and variance of $\lambda = 10$ as in figure 3.3(c); the x-axis shows the progression through the compartments $\frac{c}{C}$. To compare different tissue structures, the self-renewal parameter α is adjusted for the same output of cells as in figure 3.2. (a) The mean of replicative age increases slightly faster for high self-renewal. (b) The variance of replicative age increases also linearly, however, the impact of the self-renewal parameter α is much more pronounced: for $\alpha = 0$ there is no change, but for $\alpha = 0.67$ there is a drastic increase of the standard deviation per compartment.

Figure 3.5 shows the mean and variance of replicative age for multiple subsequent compartments for different total number of compartments, but as above with the same tissue function, that is the same overall cell production. In this example, the variance for strong self-renewal, $\alpha = 0.67$, at the second out of ten compartments is already larger than in the last compartment for the case of lower self-renewal, $\alpha = 0.34$, even though there are five compartments more in the latter case. The impact on the mean of the distribution throughout the compartments is not nearly as pronounced. Since both, mean and variance, only depend on self-renewal α and the number of compartments, in principle stem cell dynamics can be inferred from comparing mean and or variance of telomere length distributions over time (Rodriguez-Brenes and Wodarz 2016; Werner et al. 2015), as long as the general tissue structure and dynamics does not change.

Telomere length data In order to compare our theoretical expectations to biological data, we use previously published telomere length distributions of human granulocyte cell populations (Werner et al. 2015) in healthy adults across different ages. Granulocytes are differentiated cells of the myeloid arm of the hematopoietic system. Differentiation from hematopoietic stem cells to fully mature granulocytes requires multiple steps, allowing us to utilise our multi-compartment model.

The telomere length distributions for ten healthy humans along with the best parameter fit are shown in Figure 3.6. We use a least-squares fit, varying the number of compartments C and adjusting the self-renewal parameter α according to equation 3.4 to maintain a constant total daily output of granulocytes. We compare

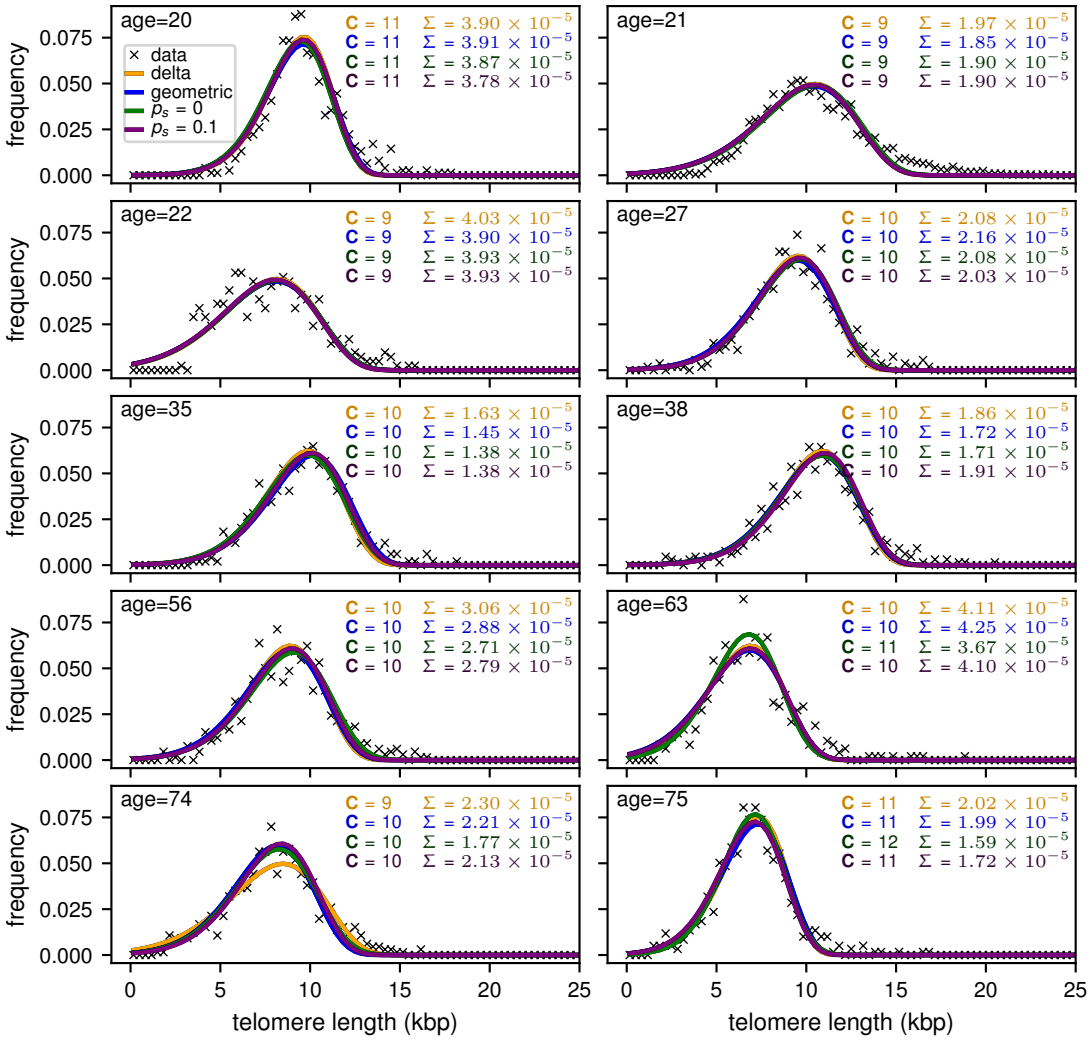

Figure 3.6: Comparison of our model to telomere length distributions from human granulocytes. Previously published telomere length distributions from (Werner et al. 2015), where details of the experimental procedures are given. We compare the fits for four different influx distributions as shown in Figure 3.3(a-d): delta influx distribution, geometric influx distribution, and influx according to the stem cell model discussed in section 3.3.1 with and without symmetric self-renewal adjusted to the age of the person. We assume a loss of 100 base pairs of telomeric DNA per cell division. We fit our model by varying the total number of compartments C and fixing the self-renewal parameter α according to equation 3.4 such that we obtain the same total outflux from the last compartment $\Omega = 2.1 \times 10^{10}\,\text{cells}\,\text{days}^{-1}$. Each panel also shows the best fit parameter C and the mean squared error Σ — which gives the quality of the fit — for each influx distribution separately.

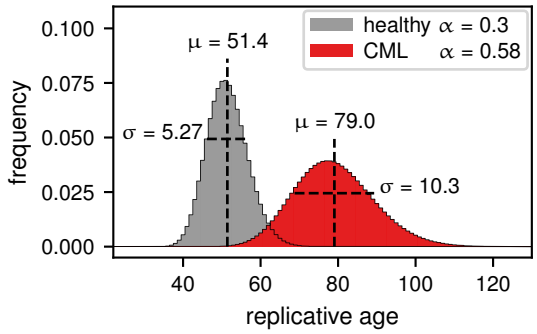

Figure 3.7: Replicative age distributions for healthy hematopoiesis and for hematopoiesis under chronic myeloid leukemia (CML). The self-renewal parameter α is the same in all 29 downstream compartments, $\alpha = 0.3$ for healthy and $\alpha = 0.56$ for cancerous hematopoiesis (Dingli et al. 2008). With CML, the replicative age distribution is much wider and shifted to a higher mean.

the fits for all four previously discussed influx distributions (Fig. 3.3), but adjust the age t for the stem cell model according to the dataset (see section 3.3.1 for details), while assuming a daily total influx of cells from the stem cell compartment of approximately $I = r_s = 1\,\text{cells}\,\text{days}^{-1}$ and $N_0 = 400$ hematopoietic stem cells (Dingli et al. 2007). The daily outflux of granulocytes is set to $\Omega = 2.1 \times 10^{10}\,\text{cells}\,\text{days}^{-1}$, as can be estimated from the total number of mature granulocytes in humans $N_{gran} \approx 2.1 \times 10^{10}$ cells (Donohue et al. 1958) and a removal rate of mature granulocytes from circulation with rate $\gamma \approx 1\,\text{days}^{-1}$ (Craig et al. 2016). Additionally we assume a fixed loss of telomeric DNA of 100 base pairs per cell division and we pick the best fitting initial telomere length for each distribution individually, as the initial telomere length is unknown.

Interestingly, by varying a single parameter, we obtain a very good agreement between our model and the granulocyte data of healthy humans, regardless the influx distribution from the stem cell compartment. The results point towards relatively high self-renewal $\alpha \approx 0.85$ with comparatively few downstream compartments and the fit result is virtually independent of the influx distribution applied. This suggests that it is difficult to infer detailed stem cell dynamics from telomere length data of differentiated tissue in this case.

Change of replicative age distribution in CML Chronic myeloid leukemia (CML) is a cancer of the hematopoietic system that can be characterized by enhanced self-renewal of cancerous cells in the progenitor compartments compared to healthy cells (Dingli et al. 2008). Here we compare the replicative age distribution for different self-renewal probabilities in the same tissue structure. For this, the tissue consists of 29 downstream compartments with either self-renewal probability $p = 0.15$ for healthy cell (Dingli et al. 2007) or self-renewal $p = 0.28$ for cancerous cells (Dingli et al. 2008) and without asymmetric division ($d = 1 - p$), leading to self-renewal parameters $\alpha_{\text{healthy}} = 0.3$ or $\alpha_{\text{CML}} = 0.56$.

The resulting distributions are shown in figure 3.7. For CML both mean and standard deviation are strongly increased compared to healthy hematopoiesis, which can be measured by comparing telomere length distributions during treatment of the disease (Brümmendorf et al. 2000). We accordingly expect that both mean and standard

deviation will decrease under successful treatment, when self-renewal in progenitor compartments normalizes again, which is consistent with available clinical data (Brümmendorf et al. 2003).

3.4. Discussion

While the age structure of cells within a tissue is driven by the age structure of the tissue specific stem cells, the progenitor compartments can substantially alter this age distribution. From a perspective of signal processing, they act as a filter that transforms an input signal (in our case a distribution) into an output signal. The properties of this filter are restricted by the biological structure of the tissue. Two limiting cases are of particular interest:

(i) Focussing on a compartment that is weakly amplifying ($\alpha \ll 1$), such that the number of output cells is approximately twice the number of input cells, the replicative age distribution in the progenitor compartment resembles that of the influx distribution. Only the average age of the cells is then increasing with the compartment number, even in tissues with many subsequent downstream compartments, such as blood (Dingli et al. 2007).

(ii) For intermediate to high self renewal (large α), the distributions of cell replicative age in a differentiated tissue with multiple progenitor compartments are virtually indistinguishable from one another, even for influxes with completely different replicative age distributions. Measuring replicative age distributions in differentiated tissue, for example via telomere lengths (Werner et al. 2015; Rodriguez-Brenes and Wodarz 2016), may therefore be more informative about the tissue structure and dynamics than the dynamics within the long lasting stem cell level.

Cellular age is explored in many experimental studies (e.g. Dykstra and de Haan 2008 gives a nice overview) and in multiple models (Ashcroft et al. 2017; Johnston et al. 2007). Some of these models also take the effect of replicative ageing into account (Glauche et al. 2009; Werner et al. 2015; Johnston et al. 2007). Furthermore, some of these models for example show that cellular age might be a critical parameter for certain diseases such as sickle cell anemia and malaria (Altrock et al. 2016; Fonseca and Voit 2015). However, replicative ageing in differentiated tissues is often overlooked, since here the cell turnover is very high and mutation accumulation as well as loss of function in these cells might not be as clinically relevant as in stem cells or early progenitor cells. On the other hand, we show that understanding the replicative ageing of differentiated cells and the resulting age distributions in the cell population can lead to a much better understanding of tissue dynamics from measurements.

Previous models of replicative ageing in a tissue hierarchy including stem cells and progenitor cells focussed strongly on the total replication limit of cells (Marciniak-Czochra et al. 2009b; Holbek et al. 2013). However, in our model such a total replication limit would most often leave the largest (and potentially measurable) portion of the replicative age distribution unchanged. However, the question becomes critical for the accumulation of mutations and the risk of cancer initiation, which we only peripherally discuss here. In reality it is not clear whether or not fully differentiated cells are close to the end of their replicative life-span in vivo, but it appears likely that they have sufficiently many cell cycles left and only a significantly increased cell turn over would lead to an exhaustion of possible cell cycles. We therefore did not include an upper replication limit explicitly here.

When comparing the distributions of replicative age between individuals or at different time points (or, for most practical purposes, their average and variance), changes of replicative age in the differentiated tissue might not always point towards changed dynamics on the stem cell level, but towards abnormal dynamics in the progenitor compartments. Accordingly, we would expect to observe these differences in replicative age distributions in certain diseases that change proliferation and differentiation characteristics in the progenitor compartments. Examples of this include chronic myeloid leukemia, acute promyelocytic leukemia and some other forms of acute myeloid leukemia where a progenitor cell in the 'middle' of the hierarchy acquires enhanced self-renewal capabilities. For example, increased self renewal would lead to a increase of average cellular age (Brümmendorf et al. 2000; Michor et al. 2005; Dingli et al. 2008; Werner et al. 2014).

In our model we focus on homogeneous population of cells that have same proliferation rates, differentiation probabilities and therefore the same fitness across multiple compartments (Traulsen et al. 2010). However, mutations or epigenetic changes can change the proliferative properties of cells within the tissue structure. Interestingly, these changes can also affect the self-renewal capacities and thus fitnesses of cells across compartments directly causing selection gradients for different lineages of cells throughout the hierarchy. Examples for both negative and positive selection are known. For example cells might die prematurely, as for example in sickle cell anemia (Altrock et al. 2016), or die later or not at all, as is observed in many cancers (Hanahan and Weinberg 2011).

An important complication that we have not considered here is that real tissues are often found in dynamical regimes that change the cellular age distribution over time. In multicellular organisms the rates for self-renewal and for symmetric differentiation or cell death are variable and tightly regulated by a variety of feedback mechanisms (Morrison and Spradling 2008). In this way, a tissue can respond to environmental conditions such as injury or infections. Also, in the context of tissue reconstitution of the hematopoietic system after stem cell transplants, the tissue structure is initially far from the steady state (Ashcroft et al. 2017). It was shown previously

that replicative age can give valuable insight into the dynamics of tissue reconstitution (Brümmendorf et al. 2003) and modelling of replicative ageing can potentially contribute towards better understanding of tissue reconstitution. It is important to point out that the steady state results presented here are not directly applicable in this situation.

In conclusion, quantitatively describing replicative age distributions of tissues in multicellular organisms can contribute to our understanding of the complex dynamical processes within such tissues and allows us to describe deviations from heathy and diseased tissue states due to changed cell proliferation properties.

Availability of source code The data for the figures in this manuscript was either calculated analytically or solved numerically by using the Scipy library for python. The scripts to create our results figures can be accessed at `https://github.com/marvinboe/DownstreamReplAge`.

Acknowledgements B.W. is supported by the Geoffrey W. Lewis Post-Doctoral Training fellowship.

Appendix

3.A. Steady state distribution

Here we show the general solution for the steady state distribution of replicative age inside a downstream compartment for any input distribution $\vec{\iota}$. The differential equations for the number of cells in each replicative age class are:

$$\dot{n}_j = \begin{cases} \iota_0 - r n_0 & j = 0 \\ \iota_j - r n_j + \underbrace{(1 + p - d)}_{=\alpha} r n_{j-1} & j \leq 1 \end{cases}$$

We then start by setting the equation for n_0 to zero, such that

$$\dot{n_0} = \iota_0 - r n_0 = 0 \quad \Rightarrow \quad n_0^* = \iota_0 / r.$$

We use this result to solve for n_1 and then continue recursively until we find the general solution for n_j:

$$\begin{aligned}
\dot{n}_1 = 0 = \iota_1 - r n_1 + \alpha r n_0 \quad &\Rightarrow n_1^* = \tfrac{1}{r}(\alpha\iota_0 + \iota_1) \\
\dot{n}_2 = 0 = \iota_2 - r n_2 + \alpha r n_1 \quad &\Rightarrow n_2^* = \alpha n_1 + \iota_2/r \\
&\qquad = \tfrac{1}{r}(\alpha^2\iota_0 + \alpha\iota_1 + \iota_2) \\
&\quad \vdots \qquad\quad \vdots \\
&\Rightarrow n_j^* = \frac{1}{r}\left(\iota_0\alpha^j + \iota_1\alpha^{j-1} + ... + \iota_{j-1}\alpha + \iota_j\right),
\end{aligned}$$

which can be written in the more compact form

$$n_j^* = \frac{1}{r}\sum_{k=0}^{j}\iota_k\alpha^{j-k}.$$

3.B. Total cell number amplification

Here we start again by writing down the differential equations for the total number of cells $N^{(c)}$ in each of the compartments c with proliferation rates $r^{(c)}$. Additionally we have the total influx I into the first compartment ($c = 0$), and similarly the total outflux Ω from the last compartment ($c = C - 1$).

$$\begin{aligned}
\frac{\mathrm{d}N^{(0)}}{\mathrm{d}t} &= (p-d)r^{(0)}N^{(0)} + I = (\alpha-1)r^{(0)}N^{(0)} + I \\
\frac{\mathrm{d}N^{(c)}}{\mathrm{d}t} &= (\alpha-1)r^{(c)}N^{(c)} + (2-\alpha)r^{(c-1)}N^{(c-1)} \\
\Omega &= (2-\alpha)r^{(C-1)}N^{(C-1)}.
\end{aligned}$$

As in the calculation above we assume our compartment to be in the steady state and set the above differential equations to zero.

$$\begin{aligned}
0 = \frac{\mathrm{d}N^{(0)}}{\mathrm{d}t} = (\alpha-1)r^{(0)}N^{(0)} + I \quad &\Rightarrow N^{(0)} = \frac{I}{(1-\alpha)r^{(0)}} \\
0 = (\alpha-1)r^{(1)}N^{(1)} + (2-\alpha)r^{(0)}N^{(0)} \quad &\Rightarrow N^{(1)} = \frac{(2-\alpha)r^{(0)}}{(1-\alpha)r^{(1)}}N^{(0)}.
\end{aligned}$$

The same calculation can be done for each compartment c:

$$\begin{aligned}
0 &= (\alpha-1)r^{(c)}N^{(c)} + (2-\alpha)r^{(c-1)}N^{(c-1)} \\
\Rightarrow N^{(c)} &= \frac{(2-\alpha)r^{(c-1)}}{(1-\alpha)r^{(c)}}N^{(c-1)} = \frac{(2-\alpha)^2 r^{(c-1)}r^{(c-2)}}{(1-\alpha)r^{(c)}r^{(c-1)}}N^{(c-2)} = ... = \frac{(2-\alpha)^c r^{(0)}}{(1-\alpha)^c r^{(c)}}N^{(0)}.
\end{aligned}$$

From this follows for the total outflux Ω

$$\Rightarrow \Omega = (2-\alpha)r^{(C-1)}N^{(C-1)} = (2-\alpha)r^{(C-1)}\frac{(2-\alpha)^{C-1}r^{(0)}}{(1-\alpha)^{C-1}r^{(C-1)}}N^{(0)} = \frac{(2-\alpha)^C}{(1-\alpha)^C}I.$$

By rearranging this for α we get

$$\alpha = \frac{\sqrt[C]{\frac{\Omega}{I}} - 2}{\sqrt[C]{\frac{\Omega}{I}} - 1}.$$

3.C. Mean and variance of replicative age distribution

To calculate the moments of the replicative age j in the progenitor compartment, we need to normalize the replicative age distribution n_j^* in the steady state by the total number of cells in the progenitor compartment $N^* = \sum_{j=0}^{\infty} n_j^*$. We can then write down the m-th moment of the age j in the progenitor compartment

$$\langle j^m \rangle_{n^\star} = \frac{\sum\limits_{j=0}^{\infty} j^m n_j^*}{\sum\limits_{j=0}^{\infty} n_j^*} = \frac{\sum\limits_{j=0}^{\infty} \frac{j^m\alpha^j}{r} \sum\limits_{k=0}^{j} \frac{\iota_k}{\alpha^k}}{\sum\limits_{j=0}^{\infty} \frac{\alpha^j}{r} \sum\limits_{k=0}^{j} \frac{\iota_k}{\alpha^k}} = \frac{\sum\limits_{j=0}^{\infty} j^m\alpha^j \sum\limits_{k=0}^{j} \frac{\iota_k}{\alpha^k}}{\sum\limits_{j=0}^{\infty} \alpha^j \sum\limits_{k=0}^{j} \frac{\iota_k}{\alpha^k}}.$$

By changing the order of summation in the denominator we get

$$\begin{aligned}\sum_{j=0}^{\infty}\alpha^j\sum_{k=0}^{j}\frac{\iota_k}{\alpha^k} &= \alpha^0\frac{\iota_0}{\alpha^0} + \alpha^1\left(\frac{\iota_0}{\alpha^0}+\frac{\iota_1}{\alpha^1}\right) + \alpha^2\left(\frac{\iota_0}{\alpha^0}+\frac{\iota_1}{\alpha^1}+\frac{\iota_2}{\alpha^2}\right)+\dots \\ &= \iota_0(1+\alpha^1+\alpha^2+\dots)+\iota_1(1+\alpha^1+\alpha^2+\dots)+\dots \\ &= \sum_{k=0}^{\infty}\iota_k\left(\sum_{j=0}^{\infty}\alpha^j\right) = \sum_{k=0}^{\infty}\iota_k\left(\frac{1}{1-\alpha}\right).\end{aligned}$$

Similarly, we change the order of summation in the nominator to

$$\begin{aligned}\sum_{j=0}^{\infty} j^m\alpha^j\sum_{k=0}^{j}\frac{\iota_k}{\alpha^k} &= 0^m\alpha^0\frac{\iota_0}{\alpha^0} + 1^m\alpha^1\left(\frac{\iota_0}{\alpha^0}+\frac{\iota_1}{\alpha^1}\right) + 2^m\alpha^2\left(\frac{\iota_0}{\alpha^0}+\frac{\iota_1}{\alpha^1}+\frac{\iota_2}{\alpha^2}\right)+\dots \\ &= \iota_0(0^m+1^m\alpha^1+2^m\alpha^2+\dots)+\iota_1(1^m+2^m\alpha^1+3^m\alpha^2+\dots)+\dots \\ &= \sum_{k=0}^{\infty}\iota_k\left(\sum_{j=0}^{\infty}(j+k)^m\alpha^j\right).\end{aligned}$$

We then rewrite the above binomial to $(j+k)^m = \sum_{i=0}^{m} \binom{m}{i} k^i j^{m-i}$ and rearrange the sums to

$$\begin{aligned}
\sum_{k=0}^{\infty} \iota_k \left(\sum_{j=0}^{\infty} (j+k)^m \alpha^j \right) &= \sum_{k=0}^{\infty} \iota_k \left(\sum_{j=0}^{\infty} \left[\sum_{i=0}^{m} \binom{m}{i} k^i j^{m-i} \right] \alpha^j \right) \\
&= \sum_{k=0}^{\infty} \iota_k \left[\sum_{i=0}^{m} \binom{m}{i} k^i \left(\sum_{j=0}^{\infty} j^{m-i} \alpha^j \right) \right] \\
&= \sum_{i=0}^{m} \binom{m}{i} \left(\sum_{k=0}^{\infty} k^i \iota_k \right) \left(\sum_{j=0}^{\infty} j^{m-i} \alpha^j \right) \\
&= \sum_{i=0}^{m} \binom{m}{i} \left(\sum_{k=0}^{\infty} k^i \iota_k \right) S_{(m-i)},
\end{aligned}$$

where in the last step we defined the sum which is independent of the influx

$$S_m = \sum_{j=0}^{\infty} j^m \alpha^j.$$

The next step is to put together the nominator and denominator and to insert the moments of the replicative age distribution of the influx $\langle j^m \rangle_\iota$:

$$\begin{aligned}
\langle j^m \rangle_{n^\star} &= \frac{\sum_{i=0}^{m} \binom{m}{i} S_{(m-i)} \left(\sum_{k=0}^{\infty} k^i \iota_k \right)}{\left(\frac{1}{1-\alpha} \right) \sum_{k=0}^{\infty} \iota_k} \\
&= (1-\alpha) \sum_{i=0}^{m} \binom{m}{i} S_{m-i} \langle j^i \rangle_\iota, \qquad (3.8)
\end{aligned}$$

which is the general solution for any moment of the replicative age distribution.

To get an expression for the mean and variance, we have to solve S_m for $m = 0, 1, 2$:

$$\begin{aligned}
S_0 &= \sum_{j=0}^{\infty} \alpha^j = \frac{1}{1-\alpha} \\
S_1 &= \sum_{j=0}^{\infty} j\alpha^j = \alpha \frac{\partial}{\partial \alpha} \sum_{j=0}^{\infty} \alpha^j \\
&= \alpha \frac{\partial}{\partial \alpha} \frac{1}{1-\alpha} = \frac{\alpha}{(1-\alpha)^2} \\
S_2 &= \sum_{j=0}^{\infty} j^2\alpha^j = \sum_{j=0}^{\infty} \left(\alpha^2 \frac{\partial^2 \alpha^j}{\partial \alpha^2} + j\alpha^j \right) \\
&= \alpha^2 \frac{\partial^2}{\partial \alpha^2} \sum_{j=0}^{\infty} \alpha^j + \alpha \frac{\partial}{\partial \alpha} \sum_{j=0}^{\infty} \alpha^j \\
&= \frac{2\alpha^2}{(1-\alpha)^3} + \frac{\alpha}{(1-\alpha)^2}
\end{aligned}$$

Generally, for $m > 0$ the sum S_m is by definition the polylogarithm $Li_{-m}(\alpha)$ (Wood 1992) with negative order m and can be written as

$$Li_{-m}(\alpha) = \frac{1}{(1-\alpha)^{m+1}} \sum_{k=0}^{m-1} E(m,k)\alpha^{m-k},$$

with the Eulerian numbers $E(n,k) = \sum_{j=0}^{k+1} (-1)^j \binom{n+1}{j} (k+1-j)^n$.

By using the general solution equation 3.8 and the above solutions for S_n for $n = 0, 1, 2$ we can calculate the mean μ and variance σ^2 of the replicative age distribution:

$$\begin{aligned}
\mu &= \langle j \rangle_{n^*} = \frac{\alpha}{1-\alpha} + \langle j \rangle_\iota = \frac{\alpha}{1-\alpha} + \mu_\iota \\
\sigma^2 &= \langle j^2 \rangle_{n^*} - \langle j \rangle_{n^*}^2 = \langle j^2 \rangle_\iota - \langle j \rangle_\iota^2 + \frac{\alpha}{(1-\alpha)^2} = \sigma_\iota^2 + \frac{\alpha}{(1-\alpha)^2},
\end{aligned}$$

where we used the mean μ_ι and the variance σ_ι^2 of the influx distribution.

3.D. Replicative age distributions for specific influx

3.D.1. Geometric influx

Here we calculate the distribution of replicative age in the steady state resulting from geometrically distributed age of the influx $\iota_k = \lambda^k(1-\lambda)$ by solving equation 3.2:

$$n_j^* = \sum_{k=0}^{j} \lambda^k(1-\lambda)\alpha^{j-k}.$$

Now we factor out all factors independent of k and substitute $x := \frac{\lambda}{\alpha}$:

$$n_j^* = (1-\lambda)\alpha^j \sum_{k=0}^{j} x^k.$$

By using the results for the geometric sum $\sum_{k=0}^{j} x^k = \frac{(x^{(j+1)}-1)}{(x-1)}$ for $x \neq 1$ we get

$$n_j^* = \begin{cases} (1-\lambda)\frac{\lambda^{j+1}-\alpha^{j+1}}{\lambda-\alpha} & \text{for } \alpha \neq \lambda \\ (1-\lambda)\alpha^j(j+1) & \text{for } \alpha = \lambda \end{cases}.$$

3.D.2. Poisson influx

To find the replicative age distribution in the progenitor compartment for Poisson distributed influx, we have to solve the following sum:

$$n_j^* = \sum_{k=0}^{j} \frac{e^{-\lambda}\lambda^k}{k!}\alpha^{j-k}.$$

By factoring out all factors independent of k, we arrive at

$$n_j^* = e^{-\lambda}\alpha^j \sum_{k=0}^{j} \frac{\lambda^k}{\alpha^k}\frac{1}{k!}.$$

Now we substitute $x = \frac{\lambda}{\alpha}$ and note that for the upper incomplete gamma function $\Gamma(a,x) = \int_x^\infty t^{a-1}e^{-t}\,\mathrm{d}t$ and integer values of j the following equality holds (Gautschi 1998):

$$\Gamma(j+1,x) = j!e^{-x}\sum_{k=0}^{j}\frac{x^k}{k!}.$$

Using this we get

$$n_j^* = \frac{\alpha^j e^{\left(\frac{\lambda}{\alpha}-\lambda\right)}}{j!}\Gamma(j+1,\tfrac{\lambda}{\alpha}).$$

as the desired result.

3.E. Time to steady state

Equation 3.1 can be solved analytically to get an estimate for the time until steady state is reached. In the following we solve the equations for an initially empty progenitor compartment, which gives us an estimate to the relaxation time until steady state is reached. For this we start by solving the equation for the first replicative age class $j = 0$ and then subsequently for all others:

$$\frac{\mathrm{d}n_0}{\mathrm{d}t} = \iota_0 - rn_0 \quad \Rightarrow n_0^{(h)}(t) = Ce^{-rt} \quad \text{variation of parameter} \quad C \to C(t)$$
$$\Rightarrow \frac{\partial C}{\partial t}e^{-rt} - Cre^{-rt} + rCe^{-rt} = \iota_0 \quad \Rightarrow C = \tfrac{\iota_0}{r}(e^{rt} - 1)$$
$$\Rightarrow n_0(t) = \tfrac{\iota_0}{r}(1 - e^{-rt})$$

This we plug into the differential equation for the next age class

$$\frac{\mathrm{d}n_1}{\mathrm{d}t} = \iota_1 + \alpha r n_0 - r n_1 = \iota_1 + \alpha\iota_0(1 - e^{-rt}) - rn_1$$

which can be solved in the same way

$$n_1(t) = \frac{(\iota_1 + \iota_0\alpha)}{r}(1 - e^{-rt}) + \alpha\iota_0 te^{-rt}.$$

Now we use the steady state values of the main text (equation 3.2), $n_j^* = \sum_{k=0}^{j} \frac{\iota_k \alpha^{j-k}}{r}$, which allows us to rewrite the above equation to

$$n_1(t) = n_1^*(1 - e^{-rt}) + \alpha r n_0^* t e^{-rt}.$$

For the third age class we get

$$n_2(t) = n_2^*(1 - e^{-rt}) + \alpha^2 r^2 \tfrac{t^2}{2!} n_1^* e^{-rt} + \alpha r n_0^* \tfrac{t}{1!} e^{-rt}.$$

From this we can infer the general solution

$$n_j(t) = n_j^*(1 - e^{-rt}) + \sum_{m=0}^{j-1} \tfrac{(\alpha r t)^{(j-m)}}{(j-m)!} n_m^* e^{-rt},$$

which as expected goes towards the equilibrium solutions n_j^* for the limit of time t to infinity. Figure 3.E1(a) shows the time evolution of of n_j for different age classes j. Approach to steady state for this specific value of α is relatively fast. To compare the time until steady state is reached for different values of α, we numerically calculated the time until n_j reached 99% of the equilibrium value n_j^* and the results are shown in figure 3.E1(b). As expected, for large α the time to reach steady state increases drastically, since many more downstream compartments have to be filled due to the long tail towards old ages in the distribution of replicative age. Also for higher age classes, in this case $j = 20$ for example, time to steady state is much longer as all changes have to go through the previous age classes first. However, since the steady state value in this case is close to zero and the influx into this compartment is even smaller, our assumption of a quasi-static process is still valid.

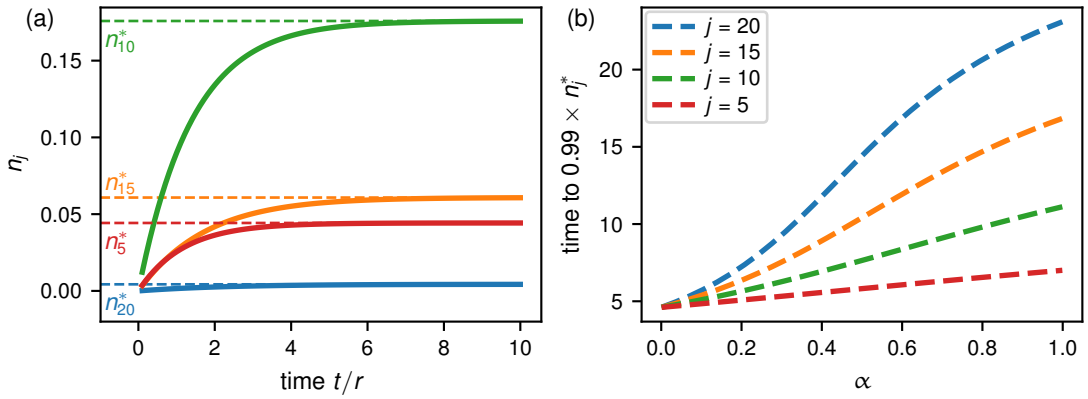

Figure 3.E1: Time to the steady state for different parameters. (a) The number of cells n_j for various age classes j increases until the steady state is reached, earlier compartments reach the steady state faster ($\alpha = 0.3$). (b) Time to approach equilibrium solutions for different age classes j in a single progenitor compartment for different self-renewal parameters α. The influx age is Poisson distributed with $\lambda = 10$ (see section 3.3.1).

4. Cancer in compartmentalized tissues

Somatic evolution in hierarchically structured tissue

In the previous chapter I have shown how a hierarchical tissue structure influences the replicative age of cells, a major source for mutations and therefore for somatic variation. However, the other main ingredient of evolution - selection - can also be strongly influenced by this tissue structure.

If an advantageous clone arises within a later stage of the differentiation hierarchy, its offspring will only transiently increase in abundance (Werner et al. 2011), unless the clone can acquire stem cell like properties to prevail in the tissue indefinitely. The total amount of cells with this mutation strongly depends on the stage within the differentiation hierarchy where it emerged: If the clone arises in a very early progenitor compartment, it will produce a large amount of offspring due to the subsequent amplification of cell numbers, potentially already causing problems for the multicellular organism. If, on the other hand, the clone arises only at a later stage of the differentiation hierarchy, the clone abundance will stay very low. Accordingly, for clones to fully fixate within a tissue they need to fixate within the stem cell population of the tissue (or acquire stem cell like behavior) (Kaveh et al. 2016). Another aspect of somatic evolution in hierarchically structured tissue is the potentially different impact of variation and selection in different positions within the structure.

In this chapter I will present the effect of a mutation which is neutral on the stem cell level, but has a positive fitness effect in later differentiation stages, at the example of chronic myeloid leukemia (CML). For this I will employ a stochastic model for the hierarchical tissue structure of full hematopoiesis from stem cells to fully maturated cells.

Manuscript status The rest of this chapter is a working draft of a manuscript with the title given below by: Marvin A. Böttcher, Benjamin Werner, David Dingli, Arne Traulsen, and Tom Lenaerts.

Stem cell dynamics is crucial for cure of chronic myeloid leukemia

Marvin A. Böttcher[1], Benjamin Werner[2], David Dingli[3], Arne Traulsen[1*], Tom Lenaerts[4,5]

[1]Department of Evolutionary Theory, Max Planck Institute for Evolutionary Biology, Plön, Germany
[2]Centre for Evolution and Cancer, The Institute of Cancer Research, London, UK
[3]Division of Hematology, Department of Internal Medicine, Mayo Clinic, Rochester, MN, USA
[4]AI lab, Computer Science Department, Vrije Universiteit Brussel, Brussels, Belgium
[5]MLG, Département d'Informatique, Université Libre de Bruxelles, Brussels, Belgium

*corresponding author: traulsen@evolbio.mpg.de

Abstract Chronic myeloid leukemia can be effectively treated with tyrosine kinase inhibitors such as Imatinib. However, it remains unclear if these drugs provide a cure to the disease, or a lifelong treatment of patients is necessary. Recent show that for some patients treatment can be stopped after several years without relapse of the disease; other patients in seemingly similar conditions experience a rapid relapse.

Here, we use a compartment based computational model of chronic myeloid leukemia to describe this disparity in outcomes. We use a multiscale approach to simulate thousands of virtual patients starting with a single mutated hematopoetic stem cell and follow each patient through clonal expansion ultimately leading to diagnosis, treatment and eventually cure or relapse. Our analysis suggests that the stochasticity of the stem cell dynamics is crucial in reproducing the clinical observations. What is more, our stochastic simulations might help to identify factors that predict if or when it is possible to safely stop the treatment.

4.1. Introduction

Chronic myeloid leukemia (CML) is a neoplasm of the hematopoietic system that leads to an increase in bone marrow output of white blood cells (Fialkow et al. 1977). CML is caused by the aberrant fusion of two chromosomes leading to the Philadelphia chromosome and the production of the fusion oncoprotein bcr/abl (Rowley 1973; Daley et al. 1990). In 2001, treatment with Tyrosine Kinase Inhibitors (TKI) became

available; these drugs target with high specificity the ABL kinase that drives the disease and result in a high and long lasting response (Hochhaus et al. 2017). Treatment with TKI, however, has to be given continuously as long as it is well tolerated or the patient does not progress to blast crisis since it supposedly does not eradicate the mutated leukemic stem cells from the bone marrow and the mutated cells regrow if treatment is stopped (Quintás-Cardama and Cortes 2009).

Despite treatment usually being applied for the foreseeable lifetime of the patient, recent studies have shown that TKI therapy can safely be stopped under certain circumstances (Kimura 2016) and guidelines for discontinuation of treatment outside of clinical studies have already been published (Hughes and Ross 2016). Clinical data for the timing and probability of relapse from several studies across the world are now available (Etienne et al. 2016; Mahon et al. 2010; Ross et al. 2013; Rousselot et al. 2014; Imagawa et al. 2015). In all these studies, treatment is only stopped when the disease burden based on bcr/abl estimation falls below some critical treshold for extended periods of time - a state that is commonly referred to as deep molecular response (DMR) (Cross et al. 2015). However, the exact depth of molecular response, the exact term used to describe this reduction limit, and the required duration this deep response has to be maintained for, varies from study to study. When the conditions for treatment cessation are fulfilled, treatment is discontinued and the level of the bcr/abl mutation in the blood is regularly monitored. Around 40-60% of patients participating in these studies experienced a detectable expansion of the bcr/abl mutation and needed to restart treatment. Interestingly, in many of these cases, relapse usually occurs within a few months after stopping treatment. Which patients experience a relapse or remain in remission appears to be entirely random. This high influence of chance on treatment outcome despite seemingly similar deep responses across patients, suggest that stochastic processes play a major role in the dynamics of CML and its therapy.

Previous work has shown that CML can be understood in terms of an evolutionary framework where the fitness of leukemic cells is altered such that normal regulatory feedback within the hematopoietic system is disturbed (Michor et al. 2005; Roeder et al. 2006; Dingli et al. 2008; MacLean et al. 2014). While the leukemic cells have a reproductive fitness advantage compared to normal cells, targeted treatment with TKI decreases the fitness of the leukemic cells (Traulsen et al. 2010; Lenaerts et al. 2011). As a result bone marrow output normalizes again and cells containing the Philadelphia chromosome gradually decrease in abundance. However, it is hypothesized that leukemic stem cells (LSC) are not affected by this targeted treatment (Graham et al. 2002; Jørgensen et al. 2007; Lenaerts et al. 2010; Hamilton et al. 2012). Stem cells harbouring the bcr/abl-mutation may therefore persist in the stem cell pool and constitute a permanent risk of CML recurrence after stopping therapy.

Here we use a hybrid stochastic and deterministic model to investigate the time dynamics and relapse probabilities before and after stopping of treatment of CML in the light of evolution. Predictions from previous work (Lenaerts et al. 2010) are

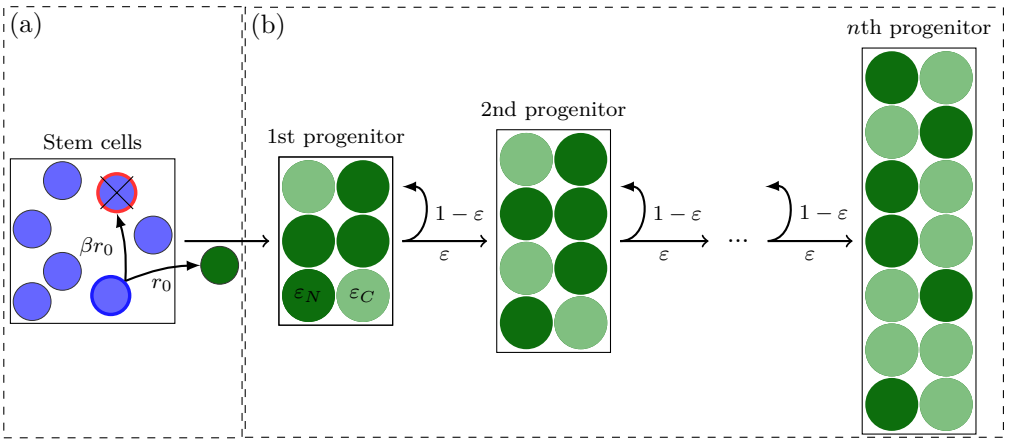

Figure 4.1: Compartment model of hematopoiesis. Full hematopoietic differentiation line including the stem cell compartment and n progenitor compartments. (a) Stem cells produce progenitor cells with a constant proliferation rate r_0. With a probability β there is a stem cell replacement event along with the production of the progenitor cell, which allows for the expansion or extinction of leukemic cells in the stem cell pool. The number of stem cells is constant and the BCR-ABL mutation has no phenotypic effect here. (b) In the downstream compartments cells either differentiate with probability ε, creating two daughter cells that move to the next compartment, or they self-renew with probability $1-\varepsilon$, leaving both daughter cells in the same compartment. Cells harbouring the bcr/abl mutation have a lower differentiation probability than normal cells $\varepsilon_C < \varepsilon_N$ which leads to the strong expansion of leukemic cells in the progenitor compartments.

validated by the above mentioned data: relapse occurs only in some cases, but rapidly due to the existing progenitor CML cells that constantly feed bcr/abl cells into the hematopoietic system. However, here we show that the crucial determinant for the properties of relapse is the stochastic nature of the leukemic stem cell dynamics.

4.2. Model

To explore the emergence, treatment, and relapse of CML we use a coupled stochastic and deterministic model of the hematopoietic system (Lenaerts et al. 2010). We model the full differentiation process of the whole hematopoietic system from the stem cell pool of around 400 cells that divide approximately once a year up to the enormous output of 10^{11} fully differentiated cells per day (Fig. 4.1). We partition the hematopoietic system into compartments, each comprising a different level of differentiation of the hematopoietic cells (Michor et al. 2005; Dingli et al. 2007).

In general, we distinguish between the dynamics within the stem cell compartment and all the progenitor compartments as they fundamentally differ: In the stem cell compartment there is no influx from an upstream compartment and we assume the size of the stem cell pool to be fixed (modelling disease in adults). For all progenitor compartments, on the other hand, there is a constant influx of cells from upstream

compartments. The dynamics of stem cells and progenitor cells are described in more detail in the following subsections.

Throughout all compartments we account for three different cell types - healthy cells H, cancer cells with bcr/abl mutation C, and cancer cells harbouring the bcr/abl mutation bound to the TKI treatment B. All three cell types have different proliferation characteristics. However, there is no interaction between the cell types apart from a transition of C cells into B cells under treatment.

Stem cell compartment We are interested in two basic processes in the stem cell compartment: Production of differentiated progenitor cells to enable tissue turnover of the hematopoietic system and self-renewal of stem cells to assure a constant number of stem cells: Firstly, we assume that progenitor cells are produced at a fixed rate r_0 to maintain regular homeostasis. Secondly, we consider stem cell replacement, where one stem cell may stochastically replace another. However, these replacement events are coupled with the proliferation rate, such that faster proliferation leads to faster replacement.

In the following we use a parameter $\beta \in [0, 1]$ to describe the fraction of stem cell replacements per produced progenitor cell. For $\beta = 0$ there are no replacement events so that the stem cell pool stays unchanged throughout the simulations and a mutation within a stem cell cannot stochastically go extinct or increase in number. For $\beta = 1$, on the other hand, there is one replacement event for each differentiated cell produced, which corresponds to a pure Moran process (Moran 1958) as was implemented in (Lenaerts et al. 2010).

However, we do not model the exact microscopic processes on the stem cell level explicitly, as the same replacement dynamics (given by β) could in principle be obtained by different microscopic stem cell proliferation patterns (Ashcroft et al. 2017; Glauche et al. 2009). For example, for $\beta = \frac{1}{2}$ there are exactly two differentiated cells produced per replacement event which could be explained by exclusively symmetric self-renewal and symmetric differentiation as microscopic processes on the stem cell level (Werner et al. 2011). Smaller values of β would point to a larger fraction of asymmetric divisions (Knoblich 2008), where only one of the two daughter cells becomes a differentiated cell while the other stays in the stem cell pool. Accordingly, no stem cell replacement ($\beta = 0$) corresponds to exclusively asymmetric divisions at the stem cell level.

As argued previously, we assume that the bcr/abl mutation does not provide a selective advantage in the long lasting stem cell compartment (Dingli et al. 2008; Huntly et al. 2004). All stem cells- healthy H and cancerous C - in our model, have the same proliferation rates and replacement characteristics.

Downstream compartments Downstream compartments experience permanent influx and outflux of cells that leads to a dynamic equilibrium of cell numbers in each compartment determined by the model parameters. This steady state is for example reached in normal unperturbed hematopoiesis (Werner et al. 2011). Similarly to stem cells, cells in a progenitor compartment can either differentiate with probability ε and both daughter cells go into the downstream compartment; or they self-renew with probability $1-\varepsilon$, and both daughter cells stay in the compartment (Fig. 4.1A). We do not take asymmetric division in the progenitor compartments into account, as it can be included into the above mentioned symmetric modes of division in the average over many cell replications (Dingli et al. 2007).

In the progenitor compartment, the different cell types have an effect on the proliferative characteristics: The differentiation probability for C cells, i.e. ε_C, is lower, which causes an expansion of cancerous cells within the hierachy. Differentiation probabilities for H and B cells, ε_H and ε_B are in comparison higher, which in the absence of cancerous cells C would allow for normal unperturbed hematopoiesis.

Treatment with TKI Treatment with TKI specifically targets cells harbouring the cancerous bcr/abl fusion gene and affects their proliferation characteristics (Quintás-Cardama and Cortes 2009). Accordingly, in our model cells change from the cancerous state C to the bound state B with some rate r_{treat}. If treatment is discontinued, B cells immediately switch back into the state C. Normal cells are not affected by the treatment and for simplicity we assume no other interactions between cells.

Coupled stochastic simulations and numeric integration: For our analysis of CML relapse, the stochasticity of cellular events plays a central role. However, the vast number of cells and corresponding cell divisions in hematopoiesis precludes a full stochastic treatment.

Therefore we use a hybrid algorithm of stochastic simulations coupled with numerical integration (Lenaerts et al. 2010): The first few compartments , with only several thousand cells and slow reaction rates, are modelled stochastically using the next-reaction-method Gillespie algorithm in continuous time (Gillespie 1977; Cao et al. 2004). In the compartments further downstream, the number of cells is usually too large for stochastic simulations to be feasible. In these compartments the cell numbers are updated numerically using the Euler algorithm in discrete time steps.

Since there is only downstream differentiation in our model, the coupling of the two methods is straightforward: Reactions in the stochastic simulations are executed until the time counter (continuous time) is larger than the time for the next deterministic update. All differentiated cells produced in the last stochastic compartment from these stochastic reactions are added to the number of cells in the first deterministic compartment and the deterministic update is executed. Afterwards the

stochastic simulations are continued again until the next deterministic time interval is reached.

Simulation observables Using our model we can simulate full hematopoiesis of virtual patients while keeping track of the complete life history of CML starting from a single mutated hematopoietic stem cell through diagnosis, reduction under treatment, and eventually eradication or relapse. To compare our results with clinical data we record the number of bcr/abl cells $N_{\text{bcr/abl}}$ and healthy cells N_{healthy} in the last progenitor compartment and calculate the burden b as is done in typical polymerase chain reaction (PCR) measurements in the clinical setting (Egan and Radich 2016); we calculate the ratio of mutated against unaltered chromosomes via

$$b = \frac{N_{\text{bcr/abl}}}{2N_{\text{healthy}} + N_{\text{bcr/abl}}},$$

as each cell contains the bcr/abl fusion gene only once (Michor et al. 2005).

In our model, diagnosis is reached when the total number of cells in the last compartment reaches 10^{12} (Lenaerts et al. 2010). From the burden at diagnosis b_{init} and the burden $b(t)$ at time t we calculate the log-reduction over time $\Lambda(t) = \log_{10}\left(\frac{b(t)}{b_{\text{init}}}\right)$.

Model parameters Only few parameters need to be adjusted to reflect treatment dynamics of clinical studies with our model.

In order to describe the dynamics of normal hematopoiesis, we use the parameters of a previously parametrised compartment based model of blood formation (Dingli et al. 2007): there are in total 32 compartments from stem cell to fully differentiated cells, a differentiation probability of healthy cells of $\varepsilon_H = 0.85$ and an exponentially increasing proliferation rate per compartment i according to $r_i = r_0\gamma^i$ with the proliferation expansion factor $\gamma = 1.26$. In the stem cell compartment a cell divides on average once per year, such that their proliferation rate is $r_0 = 1\,\text{years}^{-1}$ (Rufer et al. 1999).

In cells harbouring the bcr/abl mutation, the differentiation probabilities are altered. The same is true for cells with bcr/abl mutation bound to treatment as their fitness is reduced compared to untreated cells (Traulsen et al. 2010). Parameter values for a similar, but fully deterministic model were obtained in (Dingli et al. 2008) by fitting the model to published clinical data (Michor et al. 2005; Roeder et al. 2006) and are for the differentiation probability for cancer cells $\varepsilon_C = 0.72$, for cancer cells bound to TKI $\varepsilon_B = 0.9$, and the treatment rate is given by $z = 0.046\,\text{days}^{-1}$.

Number of stochastic compartments The number of stochastic compartment is chosen such that a further increase has no effect on the observed outcome of interest. However, this minimum number may differ for different obervables. We therefore tested the effect of changing the number of stochastic compartments for each observable presented here in this manuscript.

For the fraction of virtual patients that go into diagnosis seven stochastic compartments are sufficient (Lenaerts et al. 2010). Relapse dynamics and probability, however, strongly depend on extinction of the mutated cell type which is only possible in the stochastic compartments. In our simulations we typically used nine stochastic compartments to achieve consistent results regarding relapse dynamics (data not shown).

4.3. Results

4.3.1. Treatment dynamics

In our model treatment with TKI reduces self-renewal for cells that harbour the bcr/abl mutation except for stem cells. Thus treatment causes a drastic reduction of cancerous cells in the hematopoetic system. Interestingly, in the clinical setting the tumor burden can be inferred from blood or bone marrow samples on the molecular level, since the tumor causing mutation is well characterized (Egan and Radich 2016).

In general, an exponential decay of tumor burden with two distinct decay rates can be observed under treatment (Michor et al. 2005): An initial fast decline caused by reduction of tumor cells in the compartments with cells having fast turn over is followed by a slower second decline that is caused by the reduction of slower diving early progenitor cells [Cite: Werner et al Cancer Research 2016]. This general treatment trajectory can also be observed in all of our virtual patients, Figure 4.2(a). As expected, the initial faster decline is not influenced by the speed of stem cell replacement β.

By contrast, the long term treatment success, Figure 4.2(b), is strongly dependent on the stem cell replacement rate. For fast stem cell replacement $\beta = 1$ the leukemia causing stem cell leaves the stem cell pool quickly and a deep molecular response (DMR) of $MR^{4.5}$ (Cross et al. 2015), which means a 4.5-log reduction from the baseline burden, is reached relatively fast. In this case almost all patients reach DMR within 50 years. For slower stem cell replacement rates, on the other hand, the leukemic stem cell stays in the system and keeps providing progenitor cells which effectively prevents deep molecular response for the majority of patients. Additionally, if patients reach this threshold for these slow stem cell replacement rates at all, it takes a much larger time on average.

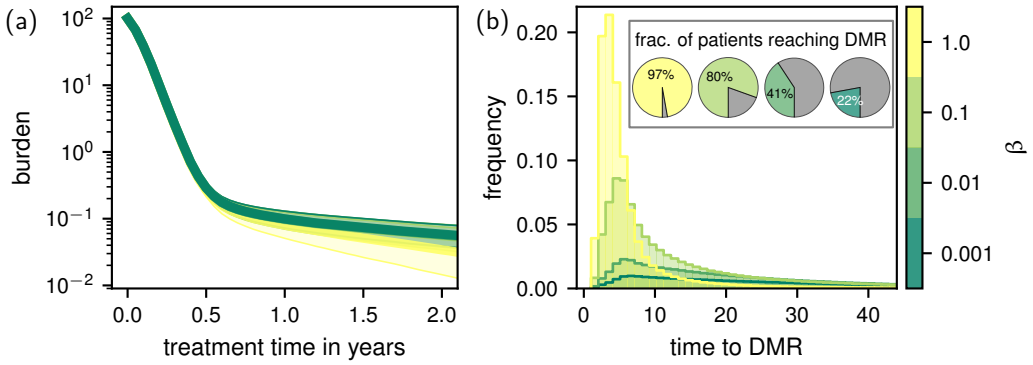

Figure 4.2: CML burden dynamics under treatment. (a) Median of initial treatment trajectories from 5000 virtual patients including upper and lower quartile. The dynamics in the first years of treatment with TKI are very similar regardless of the stem cell dynamics. (b) Time until deep molecular response (DMR) of $MR^{4.5}$ (4.5-log reduction from baseline) (Cross et al. 2015) is reached for different stem cell replacement values β. The inset shows the overall percentage of patients reaching deep molecular response within 50 years. For fast stem cell dynamics $\beta = 1$ almost all patients reach deep molecular response within 50 years. For low stem cell replacement $\beta = 0.001$, 4.5-log reduction is only achieved in a quarter of patients. For slower stem cell replacement (data not shown) the fraction of virtual patients going into reduction is even lower and the distribution over time is essentially flat over relevant time scales.

4.3.2. Stopping TKI treatment

Even though treatment with TKI is very effective and side effects are far less severe compared to other cancer treatments, stopping of TKI treatment can be necessary for various reasons. To date, several clinical studies have shown the effect of stopping the treatment (reviewed in (Kimura 2016)). However, these studies are variable regarding the selection of patients, the total duration of treatment, the required reduction of tumor burden, and the disease burden level that defines relapse before treatment is restarted.

All TKI stop studies have in common that treatment is continued for a certain amount of time after a specific level of disease burden reduction close to or below the detection limit of the cancer causing mutation is reached. For the STIM study (Mahon et al. 2010) this limit is at 5-log reduction from the baseline burden, for TWISTER at 4.5-log reduction (Ross et al. 2013), for A-STIM at undetectable level (40,000 copies of ABL, similar to 4.5-log reduction) (Rousselot et al. 2014), for ISAV between 4-log and 4.5-log reduction (Mori et al. 2015), and finally for DADI at 4.16-log reduction (Imagawa et al. 2015). Also common for all trials is a highly variable duration of treatment continuation after the above mentioned reduction threshold has been reached before treatment is stopped. Additionally, the definition for relapse varies between the studies, from just crossing the previous burden level once or twice

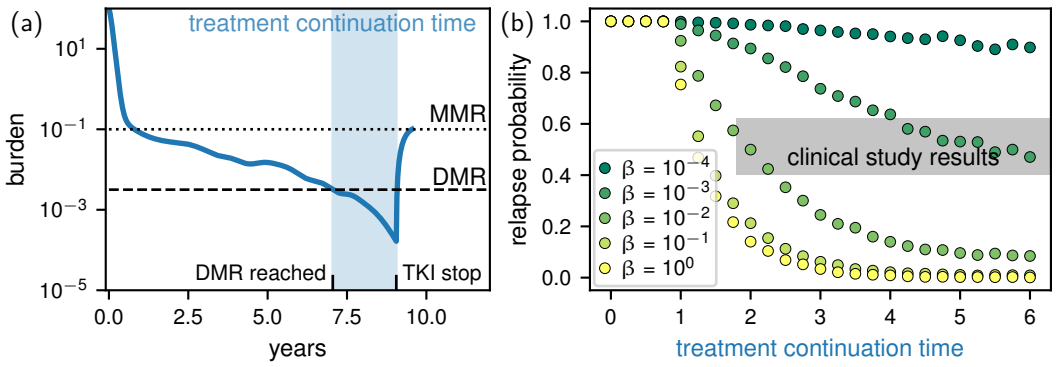

Figure 4.3: Relapse after stopping of TKI treatment. (a) Example treatment trajectory for one virtual patient exhibiting relapse. The treatment is stopped when DMR is maintained for the duration of the treatment continuation time. (b) Probability of relapse for different treatment continuation times and stem cell dynamics β. In clinical trials treatment continuation time is usually above two years and overall relapse probabilities are between 40% and 60% (Kimura 2016) (gray area). Each data point is the average of 20000 virtual patients.

again to even going back to major molecular response (MMR), which is commonly defined as 3-log reduction from the baseline burden.

Here, we universally choose a DMR of $MR^{4.5}$ as a threshold for treatment discontinuation and the reappearance of MMR as a threshold for relapse. An example treatment trajectory of a virtual patient is shown in Figure 4.3(a).

At first, we are interested in the probability for a virtual patient to go into relapse. This probability is both dependent on treatment continuation time after DMR is reached, as well as on the stem cell dynamics which critically determines the existence of the leukemic stem cell after long treatment durations. Later we will look at the time dynamics of relapse after stopping the treatment with TKI.

Probability of relapse after stopping treatment The relapse probability in our model is shown for varying treatment continuation times and under varying stem cell dynamics β in Figure 4.3(b). As expected, there is a strong negative correlation between relapse probability and treatment continuation time. For very small continuation times, below a certain threshold, there is an almost 100% probability of relapse. In this case the self-renewal of the remaining bcr/abl cells in the downstream compartments leads to a fast relapse. For longer treatment continuation times the relapse probability is strongly dependent on the underlying scenario for stem cell dynamics. For fast stem cell replacement, $\beta \approx 1$, the relapse probability quickly drops to zero with increasing treatment continuation times. For very low stem cell replacement $\beta < 0.01$, on the other hand, the relapse probability also drops, but neither as fast nor as low as for strong stem cell replacement.

In clinical trials treatment continuation times are generally very long, between 2 and 8 years, and the resulting relapse probabilities collected from all patients are around 50% (Kimura 2016). Since the exact distribution of treatment continuation times in these trials are not published, we can not directly compare our model to these data. However when accumulating results over treatment continuation times above two years it appears plausible that the stem cell dynamics within patients is consistent with a stem cell replacement probability $\beta \approx 10^{-3}$, that is one replacement event per 1000 produced early progenitor cells.

Probability of relapse and stem cell dynamics As mentioned above, the speed of stem cell replacement β also has a significant impact on the relapse probabilities. In this section we explore the relation of stem cell dynamics and relapse probability in more detail, see Figure 4.4.

For no or very low stem cell replacement ($\beta \approx 0$) the probability of relapse is close to one regardless of the treatment continuation time τ, and almost all patients go into relapse eventually if they reach reduction at all. Interestingly, even for still very small stem cell replacement rates of $\beta \approx 0.01$ the relapse probability already drops to around 50%. For larger stem cell replacement rates $\beta = 0.1$ and higher the mutated clone quickly vanishes from the stem cell pool and the relapse probability is low, dropping to values well below 25% even for short treatment continuation times. For the smallest treatment continuation time $\tau = 1y$, relapse probability is very high regardless of the stem cell dynamics which points to progenitor cell driven relapse in this case.

By contrast to real patients, in simulated patients we have full information about the dynamics on the stem cell level. In fact, the presence or absence of leukemic stem cells is strongly correlated with relapse, see Figure 4.4(b): Almost all patients that have at least one LSC left (right panel) when stopping treatment will experience a relapse. Conversely, the absence of the LSC does not guarantee relapse free survival, as around 11% of patients that do not have a LSC will experience a relapse. Similarly from the viewpoint of relapse or non-relapse (left panel), almost 100% of patients that do not experience a relapse have no LSC left at the time of stopping the treatment. By contrast, 26% of virtual patients experiencing a relapse have lost the LSC at time of stopping the treatment.

Relapse probability and treatment dynamics Since both long term treatment success and relapse after stopping of treatment are strongly connected to the dynamics of the LSC, it seems natural to ask whether we could infer relapse probabilities from the treatment trajectories of patients. Accordingly, Figure 4.5 shows individual burden trajectories under treatment for two cohorts - patients experiencing a relapse or not. As shown before (Fig. 4.2a), the CML burden decays in two slopes, the initial slope being the same for all patients regardless of the stem cell dynamics.

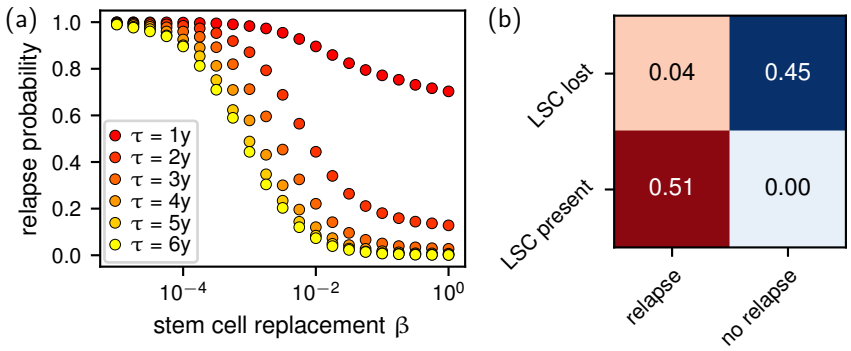

Figure 4.4: Stem cell dynamics and relapse probability. (a) Relapse probability is dependent on the underlying speed of stem cell replacement β and treatment continuation time τ (see Fig. 4.3). For very small $\beta \leq 10^{-4}$ the relapse probability is close to 100%, regardless of the treatment continuation time τ, and drops to around 50% as seen in the clinical studies for the stem cell replacement in the range of $10^{-3} \leq \beta \leq 10^{-2}$. (b) Relapse is strongly correlated to the presence of a LSC at the time of stopping the treatment; if there is a single LSC left in the hematopoietic system there will be relapse with close to 100% probability (bottom row). Almost all patients that do not experience a relapse have lost their LSC accordingly. However, there is a also substantial fraction of patients going into relapse not having a single LSC left at stopping the treatment (top left). Accumulated data from simulations with parameters shown in (a).

The second slope, however, is determined by early progenitor cells which in turn are strongly influenced by the existence of the LSC. It was shown previously that the fraction of cancer stem cells is inversely related to the offset η of the second slope under treatment (Werner et al. 2016).

In our model there is such a correlation between slope s and offset η of the second burden decay and relapse probability. For fast stem cell replacement ($\beta = 1$), the LSC usually goes extinct quickly and only in rare cases reaches relatively high abundance within the stem cell pool instead. Accordingly, the long term prevalence of a single or very few LSC does not occur in this scenario, and there is a strong distinction in both slope s and offset η between the two cohorts.

For the biologically more relevant case of slower stem cell replacement ($\beta < 0.1$), however, this distinction becomes less clear since the fraction of LSC is similar for both cohorts. For these slower stem cell dynamics, it might therefore be impossible to predict the relapse probability for patients given their previous treatment history (as in the extreme case we would need to predict a single stochastic extinction event).

4.3.3. Dynamics of relapse

In light of the broad variation of relapse probabilities shown in the previous section, it is interesting to explore the dynamics of relapse in more detail. The available clinical studies for stopping TKI treatment also contain information about the recurrence

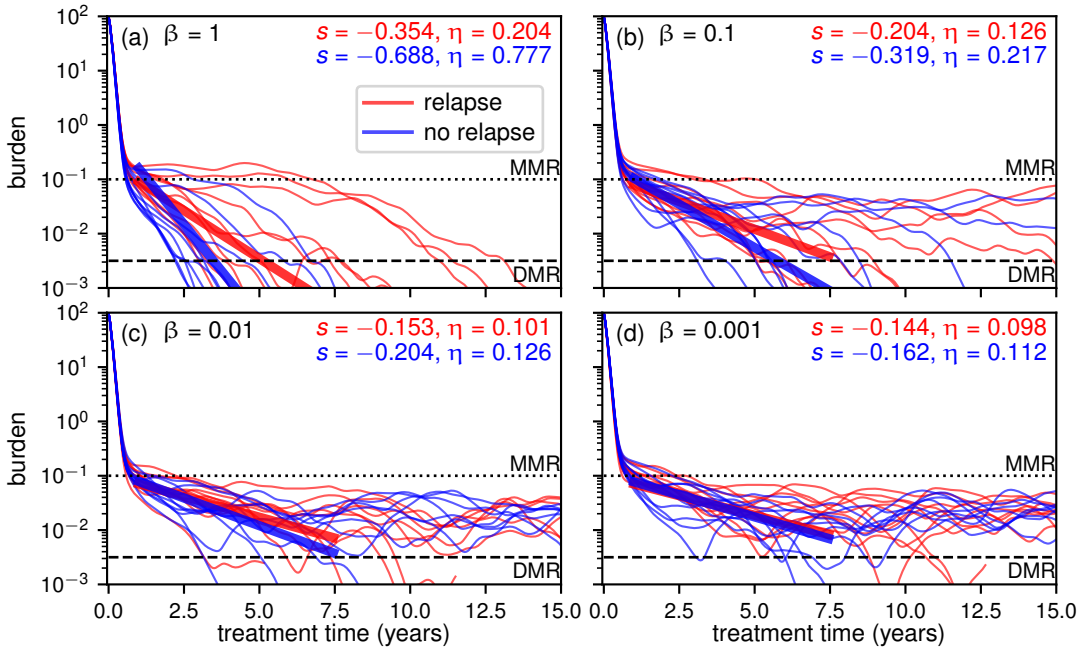

Figure 4.5: Treatment dynamics of individual patients. 10 representative burden trajectories under treatment for patients experiencing a relapse after stopping of treatment or not along with a linear regression of the second burden decay slope over 5000 virtual patients. After the initial decay, which is virtually indistinguishable across all patients, the slope of the decay s is steeper and the offset η larger for the non-relapse cohort (thick blue bars). This effect is much stronger for fast stem cell dynamics $\beta = 1$ (a). For slower - and biologically more plausible - stem cell dynamics (b-d) the two cohorts become more similar and the second burden decay is less informative about prognosis after stopping the treatment. Treatment continuation time before stopping treatment is two years.

dynamics of CML (Kimura 2016). In the clinical setting, patient relapse is usually very fast: most often within a year the clinical threshold for relapse is reached, if it is reached at all.

Individual patient dynamics Figures 4.6 (a)-(c) show the burden trajectories after stopping treatment for 75 virtual patients in the relapse and non-relapse cohort each. For both cohorts the CML burden increases significantly after stopping of treatment which already leads to a relapse in many patients, dependent on their initial burden before stopping treatment. However, after this initial increase the burden steadily decreases for all patient groups. Late relapse only occurs sporadically in few patients and only for low stem cell replacement probabilities β. In this case, the sudden increase of tumor burden is the result of a stochastic proliferation event of the leukemic stem cell or a very early progenitor cell, which is still present in the virtual patient.

Interestingly, in some cases of relapse the tumor burden will decline again after crossing the burden threshold for relapse. This potentially corresponds to all cases

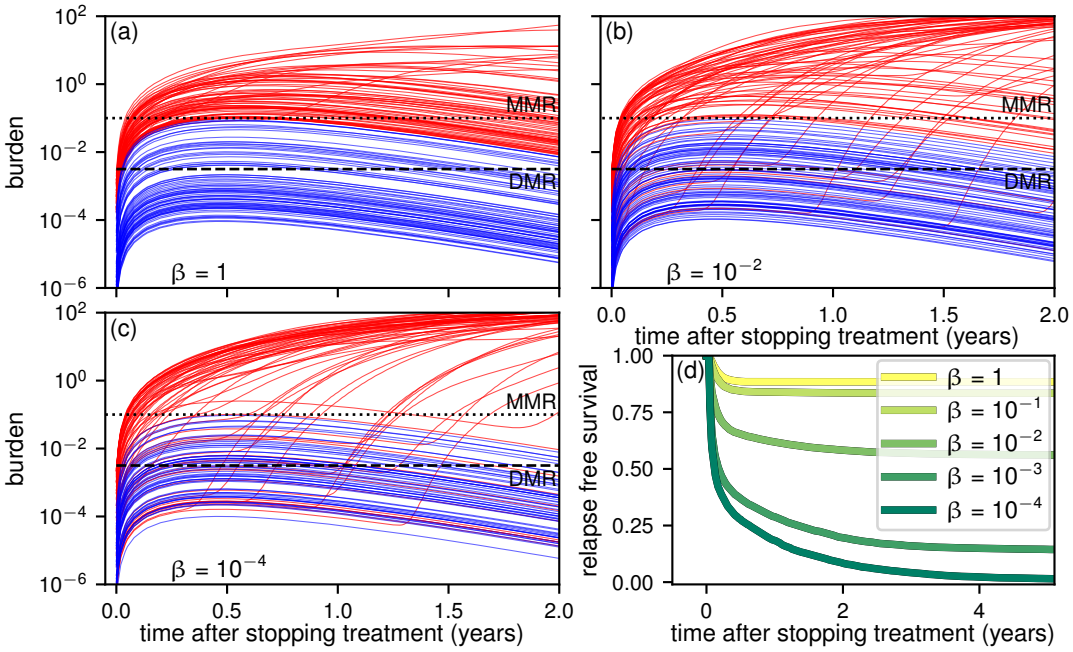

Figure 4.6: Relapse dynamics. Treatment is stopped when DMR is maintained for two years. (a)-(c) Examples of relapse trajectories for different speed of stem cell replacement β. After stopping of treatment the burden in all patients will go up, but only few reach MMR,the limit for relapse in this case. The burden for most patients will decline steadily after this initial increase, but in some cases for $\beta < 1$ patients will go into relapse spontaneously later. Each plot shows 75 virtual patients eventually going into relapse (red) and 75 virtual patients not going into relapse (blue). (d) If relapse occurs, it appears relatively fast. The fraction of virtual patients without relapse declines quickly in the first few months after stopping the treatment, due to the deterministic increase of tumor burden shown in (a)-(c).

where relapse occurs even though the leukemic stem cell is lost and the relapse is driven solely by progenitor cells, compare Figure 4.4. If the leukemic stem cell is still present in the patient, on the other hand, tumor burden will increase back to the baseline level.

Time to relapse The fraction of virtual patients which that do not relapse for various rates of stem cell replacement rates β is depicted in Figure 4.6(d). The overall relapse dynamics is very similar to published clinical data, as for example (Etienne et al. 2016), regardless of the stem cell dynamics and the final overall probability for relapse. These results strengthen our previous analysis of individual patient trajectories above: The timing of the initial relapse is hardly affected by the stem cell dynamics and is driven by intermediate progenitor cells in the hierarchy, which are already present at time of stopping treatment.

4.4. Discussion

Here we have shown how the long-term success of treatment of CML with TKI crucially depends on the dynamics of the disease causing leukemic stem cell in the stem cell compartment. This holds for both reduction of the tumor burden to very low levels as well as for the relapse probabilities after stopping treatment. However, relapse probabilities as observed in the clinical setting (Kimura 2016) are in our model only obtained for very slow stem cell dynamics with around one stem cell replacement event per hundred or more progenitor cells produced ($\beta \leq 0.01$). If phenotypic changes only occur at cell divisions, this could be a sign for predominantly asymmetric cell divisions at the stem cell level (Knoblich 2008; Werner et al. 2015). Surprisingly, the first of 31 compartments downstream of the long lasting stem cells, which still have a very low proliferation rate, have almost no relevance on the long term relapse probability of CML in our model.

The exact nature of the evolutionary processes within the hematopoetic stem cell compartment are hard to explore experimentally. Recent ideas include measuring the telomere length distributions — markers for replicative age — over time (Werner et al. 2015) or by fluorescent staining of cells expressing certain stem cell markers (Takizawa et al. 2011; Busch et al. 2015). However, even when specifically modelling the dynamics on the stem cell level in more detail (for example (Ashcroft et al. 2017)), the process often resembles a Moran process (Moran 1958) with stochastic replacement of cells in an approximately constant cell population. The advantage of our approach of modeling the dynamics on the stem cell dynamics with a single parameter β is that we can capture a wide variety of possible scenarios without modeling the process in detail. In fact, β can also be seen as a survival parameter for the LSC: for low beta the extinction time of the LSC is massively prolonged and, under treatment with TKI, stays in the stem cell pool in a seemingly dormant state (Zhang et al. 2018).

The dynamics of CML is strongly influenced by the hierarchical differentiation structure of hematopoiesis, which is usually thought to reduce proliferation in the long lasting stem cells and hence minimizes their mutational load and cancer risk (Michor et al. 2003a; Nowak et al. 2003; Michor et al. 2003b; Werner et al. 2011; Werner et al. 2013; Derényi and Szöllősi 2017). This is consistent with the overall rare occurrence of the disease (Noone et al. 2018). However, if the mutation already occurred and the disease appeared, this structure with the slow extinction of stem cell clones along with the neutrality of the mutation on the stem cell level (Huntly et al. 2004) can cause the persistence of the cancer in the hematopoietic system, making a lifelong treatment often necessary.

The basic requirement for successful discontinuation of treatment without fast relapse is the loss of the (single) long lasting leukemic stem cell which cannot be measured directly. We show that very low tumor burden, i.e. 4.5-log reduction (Branford et al. 2007), can be achieved when either this leukemic stem cell is lost or when it did not

proliferate for an extended duration due to stochastic effects. Therefore, maintaining a specific reduction threshold as in the clinical studies may be a necessary, but not a sufficient condition for the loss of the leukemic stem cells under current detection limits. However, for relatively fast stem cell dynamics ($\beta \geq 0.01$) slope and offset of the long term burden decay under treatment are correlated with relapse probability. These observables could therefore potentially be used to support estimating the risk of relapse or the need for longer treatment continuation time after DMR is reached.

In contrast to the importance of stem cells for the long-term outcome of the disease (cure versus non cure), it is the downstream progenitor cells in the differentiation hierarchy that are responsible for rapid relapse once therapy is stopped (Marley and Gordon 2005; Werner et al. 2011). These leukemic progenitor cells are constantly fed into the hierarchy by the still present leukemic stem cell, even under treatment. However, when stopping treatment these cells regain their proliferative ability and their abundance is quickly amplified in the downstream progenitor compartments with high proliferation rates. If the treatment is immediately stopped after reaching the required reduction level, the prevailing progenitor cells with the bcr/abl mutation will cause a quick increase of the tumor burden, possibly above the burden threshold for relapse, regardless of whether the leukemic stem cell is lost. If, on the other hand, treatment is further continued for a long enough duration after reaching the required reduction — in our simulations dependent on the relapse burden threshold more than three years —, relapse predominantly occurs in cases where the leukemic stem cell is still present in the stem cell pool.

Other models for CML either use deterministic equations, which do not capture stochastic effects on the stem cell level (Fokas et al. 1991; Michor et al. 2005; MacLean et al. 2014), or downscale the model of stem cell dynamics and subsequent hematopoiesis to allow for computational efficiency (Roeder et al. 2006). By contrast, the main advantage of our modeling approach is that we take into account both the stochastic effects in small stem cell and early progenitor populations and also the large scale dynamics of fully differentiated cells.

In conclusion we have shown the underlying mechanisms of relapse in CML after discontinuation of treatment that are observed in many clinical studies. Our model of neutral evolution in a small population of stem cells successfully captures major finding from these clinical studies such as probabilistic occurrence of relapse and relapse dynamics over time. This helps to establish a clearer view on the evolutionary processes involved in treatment with TKI, stopping treatment, and relapse of CML.

Availability of source code All the data in this work was produced with our software Stochtreat. The complete source code can be accessed and downloaded at `https://github.com/marvinboe/stochtreat`.

Acknowledgments B.W. is supported by the Geoffrey W. Lewis Post-Doctoral Training fellowship.

5. Cancer in fluctuating environments

As previously shown for hierarchically structured tissues, cells with the same genome can have different phenotypes, that is they have a different set of traits based on their gene expression. This is also the case for cells within a cancer which can exhibit a variety of heterogeneous phenotypes (Heppner 1984). Accordingly, the cancer cell phenotypes can be described as different species with in an ecosystem, where each species has a certain set of traits and occupies its own niche (Gatenby et al. 2011; Maley et al. 2017). Since selection generally acts on a set of traits (see section 1.1.3), the outcome of somatic selection is therefore dependent on these different phenotypes within the cancer.

However, selection is also shaped by the momentary environmental and ecological conditions (Darwin 1859). In the context of somatic evolution and cancer, many factors influence the ecological niche of a cell, including nutrient availability, external signals as for example for proliferation or apoptosis, or predation by immune cells (Tyson et al. 2003; Hynes and Naba 2012; Lu et al. 2011). This external microenvironment determines the somatic selection on the different phenotypes within a cancer and therefore also their evolution (Mintz and Illmensee 1975; Dolberg and Bissell 1984; Gatenby and Gillies 2008). Accordingly, interference with the tumor microenvironment could be a promising target for cancer therapies (Anderson et al. 2006).

Combining the two arguments from above, in the following I will show how fluctuating environments can influence the selection in a population consisting of two phenotypes — a dormant and a rapidly proliferating type — using the example of a highly malignant brain tumor. Here, the environmental fluctuations are caused by the chemotherapeutic treatment which has a different effect on selection for dormant or rapidly proliferating cells. This elucidates the role of the environment in somatic evolution. Additionally, it demonstrates how an already established cancer with different phenotypes survives in adverse conditions exerted by the treatment.

Publication information The rest of this chapter is published in the same words as:
Böttcher, M. A., J. Held-Feindt, M. Synowitz, R. Lucius, A. Traulsen, and K. Hattermann (2018a). "Modeling treatment-dependent glioma growth including a dormant tumor cell subpopulation". *BMC Cancer* **18**.1, p. 376

(shared first authorship MAB and JH, shared senior authorship AT and KH, see author's contributions for details).

Modeling treatment-dependent glioma growth including a dormant tumor cell subpopulation

Marvin A. Böttcher*[1], Janka Held-Feindt*[2], Michael Synowitz[2], Ralph Lucius[3], Arne Traulsen#[1], Kirsten Hattermann#§[3]

[1]Max Planck Institute for Evolutionary Biology, Department Evolutionary Theory, 24306 Plön, Germany
[2]Department of Neurosurgery, University Medical Center Schleswig-Holstein UKSH, Campus Kiel, 24105 Kiel, Germany
[3]Department of Anatomy, University of Kiel, 24098 Kiel, Germany

* shared first authorship

shared senior authorship

§ corresponding author: k.hattermann@anat.uni-kiel.de

Abstract

Background: Tumors comprise a variety of specialized cell phenotypes adapted to different ecological niches that massively influence the tumor growth and its response to treatment.

Methods: In the background of *glioblastoma multiforme*, a highly malignant brain tumor, we consider a rapid *proliferating* phenotype that appears susceptible to treatment, and a *dormant* phenotype which lacks this pronounced proliferative ability and is not affected by standard therapeutic strategies. To gain insight in the dynamically changing proportions of different tumor cell phenotypes under different treatment conditions, we develop a mathematical model and underline our assumptions with experimental data.

Results: We show that both cell phenotypes contribute to the distinct composition of the tumor, especially in cycling low and high dose treatment, and therefore may influence the tumor growth in a phenotype specific way.

Conclusion: Our model of the dynamic proportions of dormant and rapidly growing glioblastoma cells in different therapy settings suggests that phenotypically different cells should be considered to plan dose and duration of treatment schedules.

5.1. Background

Gliomas are the most common type of primary brain tumors including their highly malignant form, the *glioblastoma multiforme* (GBM), which accounts for about 15% of all brain tumors (Ohgaki and Kleihues 2005). Despite current standard treatment of GBM by surgical resection and adjuvant radio- and chemotherapy, the median survival time for GBM patients is still poor, approximating 12-15 months (Stupp et al. 2005), mostly due to unsatisfactory response of the tumor to treatment strategies. Additionally, combined aggressive radio-/chemotherapy is causing severe side effects frequently necessitating interruptions of the therapy due to e.g. blood toxicity (Niewald et al. 2011). GBMs and also many other tumors are heterogeneous tumors, being composed of cells with different, partly specialized phenotypes (Gatenby et al. 2011). Besides e.g. rapidly proliferating tumors cells, invading immune cells, endothelial cells and (tumor) stem cells, also a subpopulation of so called *dormant* tumor cells exists in the heterogeneous tumor mass. These cells enter a quiescent state driven by cell-intrinsic or extrinsic factors, including permanent competition for nutrients, oxygen, and space ("cellular dormancy") (Almog 2010; Wikman et al. 2008; Bragado et al. 2012; Yeh and Ramaswamy 2015). In several tumors and metastases, dormant cells have been shown to be not proliferative or only very slowly cycling (Zhang et al. 2013; Sosnoski et al. 2015; Linde et al. 2016; Sertil 2014). Linking dormancy and effects of chemotherapy, studies on glioma cells showed that cells underwent a prolonged cell cycle arrest upon treatment with temozolomide (TMZ), the most common chemotherapeutic in GBM therapy (Hirose et al. 2001).

Evolutionary forces, such as competition and selection, shape the growth of the tumor and therefore the progression of the cancer. These forces create different ecological niches within the tumor encouraging the adaption of specialized tumor cell phenotypes. Accordingly, the proportional balance between different tumor cellular phenotypes can drastically change with treatment conditions. Indeed, compared to rapidly proliferating tumor cells, especially dormant cells exhibit a much higher robustness against chemotherapeutic drugs (Almog 2010). This dormant state seems to be reversible (Hirose et al. 2001), so that the conversion to dormancy and the exit from dormancy may be a mechanism that facilitates tumor survival and progression even upon adverse or changing conditions. Hence, a better understanding of the proportional dynamics of different cell phenotypes within gliomas under chemotherapeutic treatment may improve further therapeutic approaches.

Mathematical models are beneficial resources to gain insight into key mechanisms of cancer development, growth, and evolution and to help identifying potential therapeutic targets (Altrock et al. 2015). Among these approaches, evolutionary game theory (Nowak 2006; Hofbauer and Sigmund 1998) models the interactions between different individuals as a game between agents playing different strategies and relates the payoff from this game to the reproductive fitness of the corresponding agent (Bas-

anta and Anderson 2013; Basanta and Deutsch 2008; Dingli et al. 2009; Kaznatcheev et al. 2015; Orlando et al. 2012).

Here, we use evolutionary game theory to model the proportions of two different phenotypes of GBM cells in a variety of different treatment conditions, see Basanta and Deutsch 2008 for a related approach in GBM. Defining the fitness of the different cell types as growth rate in comparison to cells of the respective other phenotype, we focus especially on the balance between the rapidly proliferating and the cellular dormant phenotype and describe the corresponding payoffs in a payoff matrix which also includes the effect of treatment. Then, we use a special form of the replicator-mutator equation (Bomze and Bürger 1995; Page 2002), which takes into account that conversion from dormant to rapidly proliferating phenotype and *vice versa* is possible. To strengthen our theoretical assumptions, we analyzed cell numbers and the cellular expression of a dormancy marker under different chemotherapy dosages and the phenotypic conversion modalities in cultured GBM cells *in vitro*. Taken together, the aim of our study was to develop a simple theoretical model which describes the dynamically changing proportions of two different GBM cell phenotypes, rapidly proliferating and dormant cells, under different treatment conditions. Showing this, we suggest that different properties of cell phenotypes should be taken into account for the development of more efficient, less toxic treatment schedules in order to improve patient's prognosis and quality of life.

5.2. Methods

Theoretical model We analyze the proportions of two different GBM cell phenotypes, dormant (D) and rapidly proliferating (P) cells, in a mathematical model including the influence of different treatment conditions. In the following, we characterize the cells in terms of their fitness, which we define as the growth rate in comparison to cells of the other phenotype. Dormant cells always have a very low or even zero growth rate ε, which we assume to be independent of the exact composition of the population and the treatment condition. Rapidly proliferating P cells, on the other hand, have a very large fitness advantage compared to dormant cells, which means they proliferate much faster, but they also compete with each other for space and resources. Facing another P cell, a focal P cell has an intermediate fitness, which we assume to be still much larger than the growth rate of D cells ε. Their fitness therefore depends on the relative fraction of D vs. P cells. Due to the very slow growth of D cells, P cells will represent the vast majority of glioblastoma cells in the absence of treatment.

Under treatment conditions, however, the population composition changes. Even though D cells still have the same very low (or zero) growth rate ε, P cells experience a fitness cost λ due to treatment. The reduction of the fitness due to treatment only applies to P cells, because cytotoxic drugs mostly affect rapidly dividing cells.

The fitness cost parameter λ can be adjusted to account for the strength of the applied treatment. In principle we can continuously vary this parameter. However, for simplicity we focus on two different treatment strategies: In high dosage (HD) chemotherapy the treatment strength parameter λ is large compared to the growth rate of the P type. Since high dosage chemotherapy has strong side effects for the whole organism (for GBM: (Niewald et al. 2011)), in reality this treatment strategy cannot be maintained for extended time periods. Therefore, strong treatment needs to be applied in turns with weaker or no treatment. For low dosage (LD) chemotherapy, λ means only a small reduction of the growth rate of the P cells. As the side-effect stress to the organism should also be lower, this treatment regime could be applied for longer time spans.

Dormant (D) and rapidly proliferating (P) phenotypes in glioblastoma and their aforementioned interactions can be described by the following payoff matrix (Basanta and Deutsch 2008):

$$\begin{array}{c} \\ \mathrm{D} \\ \mathrm{RP} \end{array} \begin{array}{c} \begin{array}{cc} \mathrm{D} & \mathrm{RP} \end{array} \\ \begin{pmatrix} \epsilon & \epsilon \\ 1-\epsilon-\lambda & 1/2-\lambda \end{pmatrix} \end{array} \tag{5.1}$$

This matrix gives the fitness for each type if confronted with any of the two other types. Here, we find for example that the fitness of a focal P cell interacting with a D cell is$1-\varepsilon-\lambda$, which includes both the small or zero growth rate of D cells ε and the fitness cost for P cells under treatment λ.

As the phenomenon of dormancy is presumably a reversible process that also occurs without any treatment, we assume that conversion between both phenotypes is possible with a small rate σ. Thus, P cells may enter a dormant phenotype, and D cells may exit from their quiescent state, converting into a P phenotype at any time point.

In the following, we include these fitness effects and phenotypic conversion into a set of ordinary differential equations. In general, the growth of a whole cell population can be explained in terms of a differential equation that describes the change in the number of individuals over time

$$\frac{\mathrm{d}n}{\mathrm{d}t} = r(n,t)n$$

Here n is the number of individuals, t is the time and $r(n,t)$ is the growth rate, which can itself depend on the number of cells and the time.

At first, we focus on the number of D cells, n_D, in the population over time, which have a very small but constant growth rate ε,

$$\frac{\mathrm{d}n_D}{\mathrm{d}t} = \varepsilon n_D.$$

For P cells on the other hand, the growth rate of n_P, given by the average fitness from the payoff matrix (weighted to the cell fractions), changes with the composition of the population

$$\frac{dn_P}{\mathrm{dt}} = n_P \left((1-\varepsilon-\lambda)\frac{n_D}{n_D+n_P} + \left(\frac{1}{2}-\lambda\right)\frac{n_P}{n_D+n_P} \right).$$

Since the system under consideration is constrained, both in terms of nutrients and space, in reality the cell population only grows exponentially as indicated by the growth equations in the very beginning of the process where the constraints regarding space or nutrients are negligible. However, we are mainly interested in the fraction of D cells $x_D = \frac{n_D}{n_D+n_P}$ in the population and vice versa the fraction of P cells $x_P = 1 - x_D = \frac{n_P}{n_D+n_P}$. To obtain the change in fractions for both types, we subtract the average growth rate $\overline{f}$ of the population from both individual growth rates,

$$\overline{f} = \varepsilon x_D + \left[(1-\varepsilon-\lambda)x_D + \left(\frac{1}{2}-\lambda\right) x_P \right] x_P.$$

From this we obtain two differential equations for the fractions of D and P cells,

$$\begin{aligned} \dot{x}_D &= x_D(\varepsilon - \overline{f}) \\ \dot{x}_P &= x_P \left(\left[(1-\varepsilon-\lambda)x_D + \left(\frac{1}{2}-\lambda\right) x_P \right] - \overline{f} \right). \end{aligned}$$

Next, we include the spontaneous conversion between phenotypes with a constant rate σ, which is independent of the cellular growth. This leads to an additional term to the differential equation of both phenotypes

$$\begin{aligned} \dot{x}_D &= \left[\varepsilon - \overline{f}\right] x_D + \sigma(x_P - x_D) \\ \dot{x}_P &= \left[(1-\varepsilon-\lambda)\, x_D + \left(\frac{1}{2}-\lambda\right) x_P - \overline{f} \right] x_P + \sigma(x_D - x_P). \end{aligned} \tag{5.2}$$

n_X	number of cells of type X
x_X	ratio of cells of type X in population
D, P	index for dormant or rapidly proliferating cell type, respectively
ε	Fitness of dormant (D) cells
λ	treatment cost on normally growing cells
σ	probability for spontaneous conversion between types
$\overline{f}$	total average fitness of all cell types in the population

Table 5.1: Overview of all symbols used in the model.

These equations have the important difference to the usual replicator-mutator equation (Nowak 2006) that phenotype conversion is a spontaneous process with a constant rate and is independent of the growth in the population. This allows conversion from D to P even if D cells do not grow at all.

Using these equations, we model different therapy schedules combining different treatment strengths in different cycling time plans. Since the equations are nonlinear, we use numerical integration with *Odeint* of the Python library Scipy[1] to examine the temporal dynamics of the system under different treatment regimes. Additionally we analytically determine the fixed points of the system and their stability.

5.2.1. Experimental model

Cell culture and cell number determination The GBM cell line LN229 was purchased from the European Collection of Authenticated Cell Cultures (ECACC, Salisbury, UK) and cultured in Dulbecco's modified eagle medium (DMEM) plus 10% fetal calf serum (FCS, PAN Biotech, Aidenbach, Germany). *Mycoplasma* contaminations were routinely excluded by bisbenzimide staining. The GBM cell line identity was proven routinely by STR (Short Tandem Repeat) profiling at the Department of Forensic Medicine (Kiel, Germany) using the Powerplex HS Genotyping Kit (Promega, Madison, WC). Briefly, DNA was amplified with a STR multiplex PCR, electrophoretic separation was performed with the 3500 Genetic Analyser (Thermo Fisher Scientific, Waltham, MA, USA), and evaluated using the Software GeneMapper ID-X (Thermo Fisher Scientific). For determination of cell numbers after low and high dose chemotherapy treatment, 25,000 cells/well were seeded in 6 well plates (Greiner Bio-one, Frickenhausen, Germany). Cells were grown for 24 h, then washed

[1] https://www.scipy.org/

with phosphate buffered saline (PBS), supplemented with fresh DMEM + 10 % FCS and temozolomide concentrations (Sigma-Aldrich, St. Louis, MO, USA; dissolved in dimethyl sulfoxide DMSO) as indicated in Figure 2A (5, 50 or 100 µg/ml for 10 days). Temozolomide (TMZ) is a DNA alkylating drug causing apoptotic cell death and the most commonly used chemotherapeutic in GBM therapy. Control cells were supplemented with 0.5 % DMSO, which corresponds to the solvent concentrations of each TMZ stimulated sample. Cells were stimulated for 10 days with TMZ, while media were changed every 2-3 days. After 10 days, cells were detached by trypsination and total cell numbers per well counted using trypan blue exclusion and a Neubauer chamber (Brand, Wertheim, Germany). DMSO stimulated control cells were already detached after 6 days of stimulation, split 1:10 and seeded again to exclude limitations of growth due to space and nutrient limitations. This splitting factor (1:10) was considered when relative cell numbers of TMZ treated samples in comparison to DMSO controls were determined for n = 5-6 independent experiments.

Immunocytochemistry For immunocytochemistry, 50,000 cells were seeded onto poly-D-lysine coated glass cover slips, grown for 24 h and supplemented with indicated TMZ or DMSO concentrations as described above. From day 6, growth media were additionally supplemented with 10 µM 5-bromo-2'-deoxyuridine (BrdU, Sigma-Aldrich, St. Louis, MO) to allow for incorporation in the DNA in the S phase of the cell cycle. After 10 days, cover slips were fixed with an ice-cold mixture of methanol and acetone (1:1) for 10 min, rinsed with 0.1% Tween / PBS (3x5 min), incubated with 1 M HCl for 30 min, neutralized with 0.1 M sodium borate buffer (pH 8.5), and rinsed again with 0.1% Tween/PBS. Afterwards, cells were blocked for unspecific bindings with 0.5% bovine serum albumin (BSA) / 0.5% glycine in PBS (1 h) and incubated over night with the primary antibody against H2BK (1:300, Biorbyt, Cambridge, UK), a marker of glioma dormancy (Almog et al. 2009; Adamski et al. 2017a) and the primary antibody against BrdU (1:200, Abcam, Cambridge, UK). Then cover slips were incubated with the secondary antibodies (donkey anti-rabbit IgG, labelled with Alexa Fluor 488, and donkey anti-sheep labelled with Alexa Fluor 555, both Invitrogen, Carlsbad, CA, USA) for 1 h at 37°, and 4´, 6-diamidino-2-phenylindole (DAPI; Sigma Aldrich, St. Louis, MO, USA; 1 mg/ml, 1:30,000, 30 min at room temperature) to stain nuclei. Cover slips were embedded using Immu-Mount (Thermo Fisher Scientific, Rockford, IL, USA), and analysed with equal exposure times using an Axiovert microscope and digital camera (Zeiss, Jena, Germany). H2BK-immunopositive, BrdU-positive and double positive cells were counted and normalized to total cell numbers in 6 (DMSO controls) to 10 (TMZ samples) fields of view for n = 4 independent experiments.

DiO retention and cell countings on phenotype conversion To monitor the conversion to and from dormancy we used the green fluorescent vital dye DiO (Invitrogen), as rapidly proliferating cells lose the dye due to repeated divisions, while

resting, dormant (or very slowly cycling) cells retain the dye and can be detected by fluorescence microscopy. Investigating the conversion to dormancy, 150,000 LN229 cells were seeded into 6-well-plates, stained with Vybrant® DiO Cell-Labeling Solution (Thermo Fisher Scientific, Waltham, MA, USA) following the manufacturer's instructions and stimulated with 100 µg/ml TMZ (or equal volume of the solvent DMSO) for 10-12 days. Cells were photographed combining transmitted-light microscopy and fluorescence microscopy with equal exposure times for TMZ and control treated cells, and green fluorescent cell portions were determined in comparison to total cell counts. To determine the influence of different cell densities on the incidence of conversion, 50,000 and 150,000 cells were seeded, respectively, into 6-well-plates and treated with 100 µg/ml TMZ (or equal volumes of the solvent DMSO) for 10 days. As the DMSO control treated cells rapidly proliferate, cells were detached at day 6 (50,000) or day 3 and 6 (150,000), cell numbers counted using a Neubauer chamber to determine the growth rate over this time period, and seeded again at initial density, to allow for cell growth without limitation of space and nutrients. After 10 days, TMZ and control treated cells were detached and counted. To extrapolate the total cell numbers of control cells, growth rates determined at day 3, 6 and 10 were used, and TMZ surviving cells were calculated as percentage of extrapolated total cells.

Statistical analysis Statistical analysis and graphical presentation of experimental data were performed with Graph Pad Prism using a two-tailed t-test (*** $p < 0.001$).

5.3. Results

5.3.1. Modelling the dynamics of cell frequencies

The temporal dynamics of the proportion of D against P cells in GBM strongly depends on the treatment conditions. Therefore, we first analyze the fixed points of the dynamical system and how they change for different treatment strengths λ, without considering possible conversions of phenotypes. The fixed points mark a stable equilibrium between the portions of P and D cells under certain, predefined conditions and are found by setting equation 5.2 to zero. Of particular interest are *stable* fixed points, as the system returns into this state again after a small perturbation (Strogatz 2014).

For our system, there is only one stable fixed point for each treatment condition (Fig. 1A). If we consider the case of no phenotype conversion, $\sigma = 0$, we can give the exact

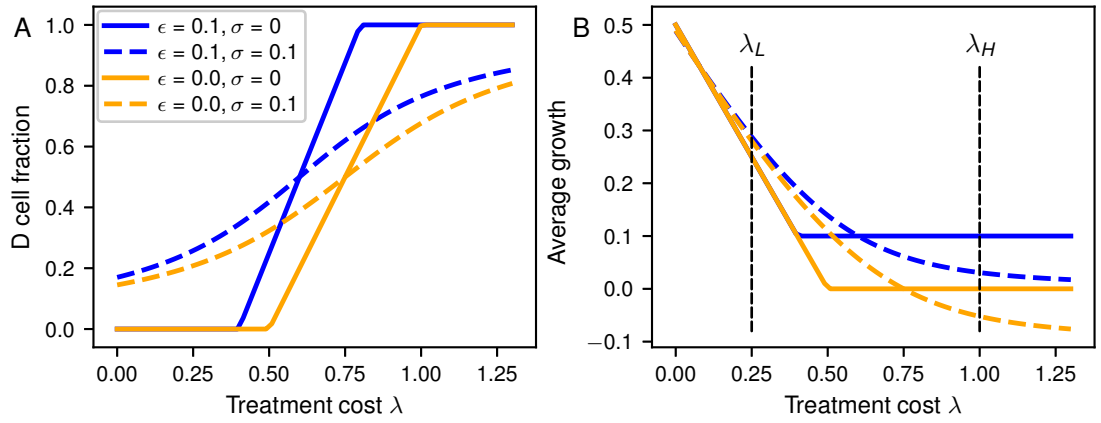

Figure 5.1: A Equilibrium fraction of dormant cells depending on treatment cost λ. **B** Average growth rate at the fixed point depending on treatment cost λ. Blue lines indicate a growth rate of D cells of $\varepsilon = 0$ and orange lines a growth rate $\varepsilon = 0.1$. Solid lines are for the absence of phenotype conversion ($\sigma = 0$) and dashed lines with phenotype conversion ($\sigma = 0.1$).

position of this point for each treatment condition λ. The fraction of D cells at the fixed point is then given by

$$x_D = \begin{cases} 0 & \frac{2\lambda}{1-2\varepsilon} \leq 1 \\ \frac{2\lambda}{1-2\varepsilon} - 1 & 1 \leq \frac{2\lambda}{1-2\varepsilon} \leq 2 \\ 1 & 2 \leq \frac{2\lambda}{1-2\varepsilon}. \end{cases}$$

For small treatment strengths λ the fraction of D cells in the population at the stable fixed point is zero, but after reaching a threshold, the fraction of dormant cells increases linearly with λ until the whole population consists of dormant cells at very high treatment strengths.

With conversion between the two phenotypes ($\sigma > 0$), the analytical calculation for the stable fixed points is more difficult and the result is not instructive. By contrast to the previous results without phenotype conversion, there is always a small proportion of dormant cells in the population, even at very low treatment strengths. The proportion of dormant cells at the fixed points increases immediately with increasing treatment strength until it approaches a maximum at high treatment strengths. For both D cell growth rates ε ($\varepsilon = 0$, orange lines; and $\varepsilon = 0.1$, blue lines) the population composition is very similar or even the same without phenotype conversion at very small or large treatment strengths . In contrast, t intermediate values of ame without phenotype conversionnt strengths is very very similar. eas e dynamics.fore decided to onlythe largest effect of ε on the population composition is at intermediate values of treatment strength.

The average fitness $\overline{f}$ for the whole tumor cell population including P and D cells decreases linearly from the maximum at treatment strength $\lambda = 0$ until it reaches the

minimum of $\overline{f} = \varepsilon$ at the point where the fraction of dormant cells in the population starts to increase (Fig. 1B). Interestingly, with spontaneous conversion $\sigma > 0$, the average fitness at the fixed point can become smaller than ε and even negative for high treatment strengths, potentially leading to a shrinking tumor. This is caused by conversion of D cells into P cells which are then susceptible to treatment.

5.3.2. Comparison to experimental data

To test our mathematical model of phenotype composition upon treatment, we used LN229 cells as an experimental *in vitro* model. We treated these cells for 10 days with temozolomide (TMZ), the most commonly used cytotoxic drug in glioma therapy. In a first step, we focused on different treatment strength and analysed the portions of surviving cells in comparison to control cultures and the percentage of cells expressing H2BK (histone cluster 1), a marker of glioma dormancy (Almog et al. 2009; Adamski et al. 2017a), alongside with incorporation of BrdU in the late treatment phase (day 6-10). In general, after 10 days of treatment, samples stimulated with 5, 50 and 100 µg/ml TMZ had significantly less total cell numbers than control treated cells (Figure 5.2A). By immunocytochemistry of H2BK, we could detect and quantify the fraction of dormant cells within the cultures, and by adding BrdU to the cells from day 6 of treatment and immunocytochemical staining of BrdU, we could in parallel mark cells that incorporate BrdU in the DNA (examples of microscopic pictures in Figure 5.2B). While DMSO-treated control cells showed a low fraction of H2BK-positive cells (mean: 9.7% ± 3.5), TMZ treatment yielded increased numbers of dormant cells reaching a plateau at high concentrations (5 µg/ml: mean 26.8% ± 9.0, 50 µg/ml: mean: 82.8% ± 5.3, 100 µg/ml: 87.7% ± 8.0, compare Figure 5.2B, grey graph portions). In parallel, we investigated the incorporation of BrdU in the DNA and determined a high portion (66.0% ± 7.8) of BrdU positive cells in the control cultures and lower portions upon TMZ treatment (5 µg/ml: 53.6% ± 14.5; 50 µg/ml: 33.4% ± 5.3; 100 µg/ml: 33.7% ± 10.1, compare Figure 5.2B, hatched graph portions). Interestingly, BrdU incorporation also took place in TMZ treated cultures, so that staining for the dormancy marker H2BK and for BrdU could be observed in the very same cells (compare examples of microscopic photographs in Figure 5.2B) indicating that cell cycle arrest may occur after the S phase of the cell cycle. Together with our experiments described in the following section and Figure 5.2C, showing that dormant cells hardly divide within our experimental time frame, these observations suggest that dormant glioma cells are not or only very slowly cycling. Furthermore, taking into account that we use a clonal cell line, the occurrence of dormant cells needs to be a phenotypic adaption to the environmental conditions as all cells are genetically homogenous (as proven by routinely STR profiling, compare Materials and Methods section).

To investigate if the conversion to a dormant phenotype depended on the cell density, initially, we determined in a DiO retention assay that nearly all cells (98.3% ± 1.2)

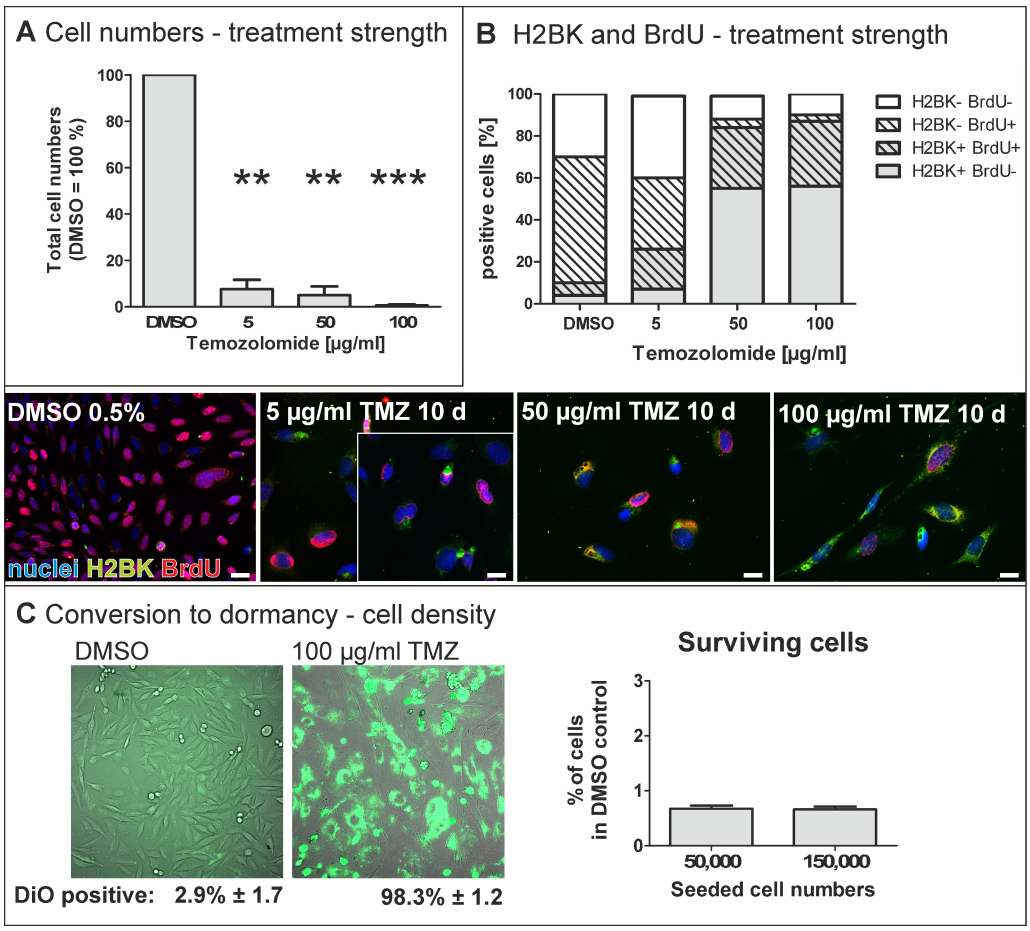

Figure 5.2: A Decrease of total cell numbers upon different temozolomide (TMZ) treatment strength. LN229 glioma cells were treated with different TMZ concentrations for 10 days, control cells were treated with the solvent DMSO (0.5%). Total cell numbers strongly decreased in a TMZ concentration dependent manner. Given are mean values of cell counting ± SD from n = 3 independent experiments. **B** Increase of the H2BK positive dormant cell portion upon different TMZ treatment strength, and incorporation of BrdU. The fraction of dormant cells as determined by immunoreactivity for the glioma dormancy marker H2BK and counting of the positively stained cells was remarkably increased in a concentration dependent manner (grey portions of the graph). The fraction of cells with BrdU incorporation in turn decreased (hatched portions of the columns), but, remarkably, in higher TMZ concentrations, H2BK and BrdU double positive cells were frequently observed (hatched, grey portions of the columns). Microscopic pictures exemplarily show cells expressing the dormancy marker H2BK (green) and the incorporation of BrdU (red) upon stimulation with different concentrations of TMZ for 10 days. The pictures are representative examples from 6-10 fields of view that were analyzed for n = 4 independent experiments and summarized in the graphs in the upper part; the bars indicate 20 µm. **C** Influence of cell density on portions of dormant cells. Left: When stained with the vital dye DiO and treated with 100 µg/ml TMZ for 10-12 days, nearly all cells (about 98%) cease from dividing which is indicated by the retention of the green fluorescent dye. Meanwhile, in control treated cells (DMSO), nearly all cells lose the green fluorescent dye due to repeated cell divisions. Right: Graphs show surviving dormant cells after 10 days of TMZ treatment (100 µg/ml) in dependence of the initially seeded cell numbers. Portions of surviving cells are very low in comparison to total (extrapolated) cell numbers in DMSO control cultures, and do not depend on initially seeded cell numbers. Given are mean values of cell counting ± SD from n = 6 independent experiments.

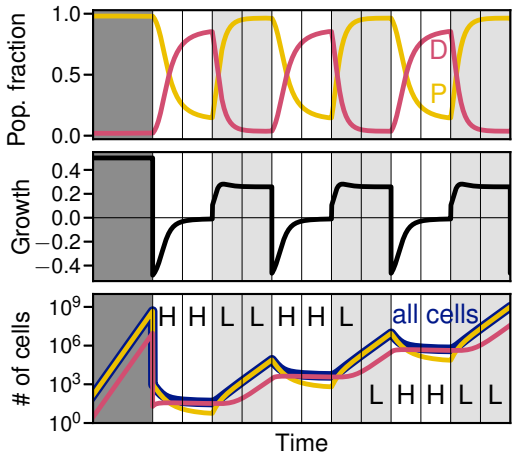

Figure 5.3: Impact of cyclic treatment on the cell population. In each treatment interval, either high dosage (H) or low dosage (L) treatment is applied.
Top: Population composition between dormant (D) and rapidly proliferating (P) cells displayed by the relative fraction of both phenotypes under changing conditions.
Middle panel: Average growth rate of the whole population under changing treatment condition. A negative growth rate only occurs in phases of strong treatment.
Bottom panel: Absolute number of all tumor cells, P and D phenotype, assuming exponential growth with the given average growth rate (parameters: dormant cell growth rate $\varepsilon = 0$, conversion rate $\sigma = 0.01$, high dosage effect $\lambda_H = 1$, low dosage effect $\lambda_L = 0.25$).

retain the green fluorescent dye when treated with 100 µg/ml ("high dose") TMZ for 10-12 days, while in control cultures (treated with equal volumes of the solvent DMSO) only 2.9% ± 1.7 retained the dye (Figure 2C, left part). The vital dye is included in every cell at the moment of staining, and is transferred to every daughter cell upon cell division. However, this means the staining is diminishing after several divisions of rapidly proliferating cells, but retained in non-proliferative or very slowly cycling dormant cells. Thus, assuming that nearly all cells that survive treatment with 100 µg/ml TMZ are dormant in our particular setting, we determined the relative incidence of phenotype conversion and the influence of the cell density on this conversion factor by determination of TMZ surviving cells in relation to (extrapolated) total cell numbers of control (DMSO treated) cultures. In our experimental setting, a portion of 0.68% ± 0.13 cells of initially seeded 50,000 LN229 glioblastoma cells survived this high dose treatment, while in cultures of initially seeded 150,000 cells, the portion of surviving cells was nearly similar (0.66% ± 0.13; Figure 2C, right part) underlining the assumptions of our theoretical model.

Thus, treatment with TMZ significantly reduced total cell numbers of LN229 cells, while the share of dormant cells within the culture, as detected by the dormancy marker H2BK, was drastically elevated. The incidence of conversion to dormancy did not depend on cell densities in our particular experimental setting.

5.3.3. Treatment schedules

Next, we use our model to analyze the dynamics of the population composition for periodically changing treatment conditions. One example trajectory for a growth

rate for D cells $\varepsilon = 0$ and a conversion rate between phenotypes σ =0.1 is depicted in Figure 5.3. The fraction of D and P cells in the population alternates between the fixed points corresponding to the momentary treatment condition. The trajectory starts with a phase of no treatment, which is characterized by a high average growth rate and a cell population composition of mostly P cells and only very few D cells. After the first phase of unconstrained growth large parts of the tumor are removed (e.g. by surgery), leaving only a small number of cancer cells. Under the following high dosage treatment conditions, the dormant phenotype has the highest fitness and takes over the population. The relative fraction of D cells will increase until the steady state under high treatment conditions is reached. The impact of treatment on P cells leads to a strong initial decline in average growth rate, until the population has a significant proportion of dormant cells and the growth rate starts to recover slightly.

Under the following low dosage treatment conditions, P cells (making up a small fraction of the whole population at the end of the high dosage treatment) are less affected by the treatment and now grow faster. The average growth rate will have a maximum when then relative fraction of P cells in the population is still low, since they have a competitive advantage over D cells, and then declines afterwards towards an equilibrium well above the high dosage growth rate. Accordingly, the total number of cells increases strongly again in this regime.

Switching the order of high dosage and low dosage treatment only has a small effect on total number of cells: If treatment starts with low dosage, the system will go into a state with a slightly higher fraction of dormant cells, which makes it less susceptible to the following high dosage treatment. Starting with low dosage therefore does not help to reduce the tumor size.

In Figure 5.4 we compare three different treatment schedules: just one switch from initial high (H) dose to low (L) dose treatment (HHHHHHLLLLLL, Fig. 4A, each instance of the letter H or L corresponds to the same time interval), slow cyclic switching (HHHLLL, Fig 4B), and fast cyclic switching (HLHL, Fig. 4C) for two different growth rates of D cells (left panels $\varepsilon = 0$ and right panels $\varepsilon = 0.1$). In case of only switching once, the fixed points for each treatment are quickly reached. At high dosage treatment the number of cells increases very slowly or even decreases. In the following low treatment phase, however, P cells take over growing particularly fast and jeopardizing any positive effect from the previous strong treatment. This is true for both treatment strengths of the high dosage phase.

For the fast switching treatment schedule (HLHL, Fig. 4C), the fixed point of the population proportion is not reached before the treatment changes again. Therefore the population dynamics stays between the two stable fixed points for the two treatment regimes, but does not reach them. By contrast, in the slow treatment switching regime (HHHLLL, Fig. 4B) the fixed points for both high and low dosage treatment are reached such that the composition of the cell population essentially resembles the

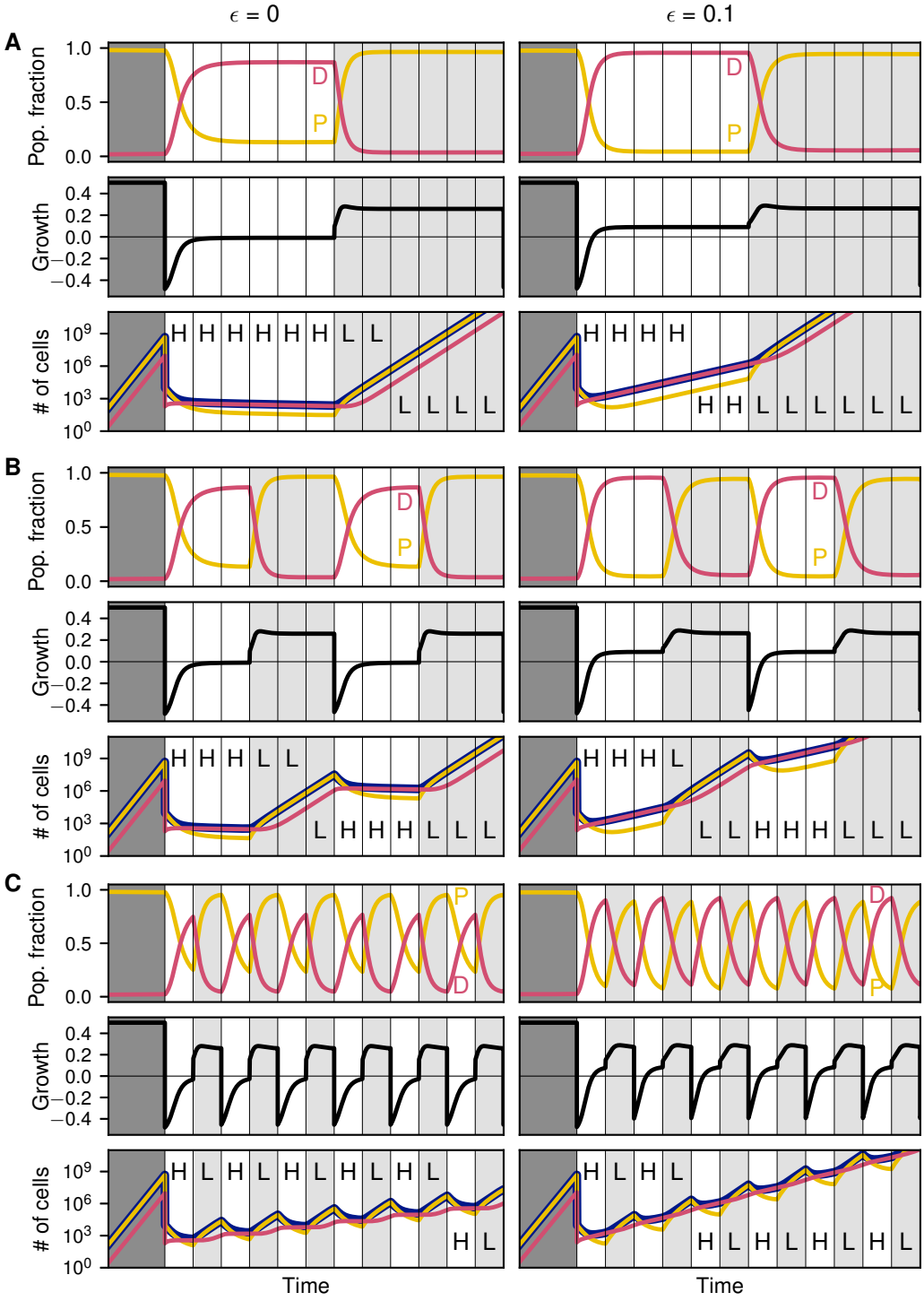

Figure 5.4: Comparison of the effect of different treatment cycle lengths on population composition, average growth rate and number of cells, similarly to Figure 5.3. All panels on the left have a D cell growth rate of $\varepsilon = 0$, whereas all panels on the right have $\varepsilon = 0.1$. **A** The top row shows the case of high dosage treatment followed by a sustained low dosage treatment. **B** The middle panels use a relatively slow switching between high dosage and low dosage treatment, whereas the bottom panel **C** shows very fast switching. All other parameters as in Figure 5.3.

case of just one switching event (Fig. 4A). However the time spent at these fixed points is still significantly reduced compared to only a single switch.

The bottom panels of Figure 4 A, B and C show the total number of cells based on the average fitness of the population under the assumption of exponential growth. When the growth rate is positive the cell population grows, otherwise it shrinks. Interestingly, the average growth rate of the population is well below zero only for a short period during the high dosage treatment and only if the share of P cells is still very high and the fraction of D cells in the population is small. However, in this regime the fitness recovers fast and approaches equilibrium with an average fitness close to zero, such that the total number of cells does not change anymore. The strongly negative growth rate directly after switching to the high dosage treatment is therefore the reason why the number of cells for quickly changing treatment regimes is significantly smaller than for slowly changing treatment cases.

5.3.4. Population growth

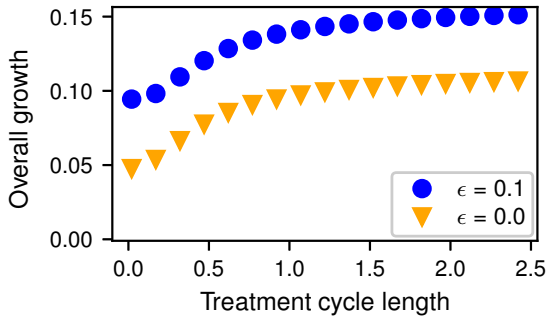

Figure 5.5: Overall growth rate for different treatment cycle length and for two different growth rates of dormant cells $\varepsilon = 0$ and $\varepsilon = 0.1$. High and low dosage phases are alternating with the given treatment cycle length for in total 30 cycles. The overall growth rate is then calculated from a linear fit to the log-plot. Other parameters as in Figure 3 and 4.

To systematically examine the effect of switching treatment cycles on the growth rate of the population, we analyze the temporal dynamics of the population size for varying treatment cycle durations and two different growth rates of D cells ε (Fig. 5). Unsurprisingly, a lower growth rate of D cells has a diminishing effect on the overall growth of the cells. For increasing treatment cycle length cancer cells have an increasing overall growth, while maintaining the same total high and low dosage treatment durations. The overall growth rate approaches a maximum with increasing treatment cycle length when the dynamics reaches equilibrium in each cycle.

Taken together, our mathematical model allows us to theoretically predict the fitness and proportions of rapidly proliferating and dormant tumor cells under different treatment conditions. Strengthening our theoretical assumptions we could exemplarily show the effect of high and low chemotherapy doses on the cell numbers and the proportion of a dormant cell phenotype in cultured GBM cells *in vitro*. Simulating different therapy schedules, we observed that fast switching of low and high treatment

doses yields a lower total tumor cell number at equal total drug dose in comparison to low switching schedules.

5.4. Discussion

In this study, we established a mathematical model and analyzed the proportions of two different cell phenotypes occurring in GBM, rapidly proliferating and dormant cells. Corroborated by experimental data obtained from *in vitro* experiments with cultured GBM cells, we observed that treatment strength influences the balance between both phenotypes which in turn influences the growth of the whole tumor population. Sequential switching of treatment strength may thus drastically influence the proportion of dormant and rapidly proliferating cells, especially if switching to the next condition takes place before the population dynamics reaches a steady state.

Dormancy in GBM has been shown by the existence of distinct fraction of temporarily non-proliferative cells in murine models (Endaya et al. 2016), as well as by the identification of clones which were able to generate indolent dormant tumors both in subcutaneous and orthotropic intracranial sites (Satchi-Fainaro et al. 2012). Additionally, dormancy seems to be characterized by specific features in GBM, such as a non-angiogenic phenotype (Almog et al. 2009; Satchi-Fainaro et al. 2012; Naumov et al. 2006), and is influenced by the (micro-)environment e.g. hypoxia (Hofstetter et al. 2012) and coagulation (Magnus et al. 2013; Magnus et al. 2014b; Magnus et al. 2014a). However, cellular dormancy in tumors is not only regarded as a state to overcome times of adverse conditions but has also been assigned to DNA repair mechanisms (Evans 2015). Interestingly, dormant GBM cells are hallmarked by the upregulation of specific genes like angiomotin, ephrin type-A receptor 5 (EphA5), insulin-like growth factor-binding protein 5 (IGFBP5), and histone cluster 1 (H2BK) (Almog et al. 2009; Adamski et al. 2017a). We used the latter as a marker to detect dormant cells in our *in vitro* experiments and show that the proportion of dormant cells increases with increasing chemotherapy concentrations.

As the fitness in the competition for space and resources depends on the proportions of phenotypically different cell subpopulations, we used evolutionary game theory as a framework for our mathematical model. Previous studies discussed game theoretic interactions with more phenotypes for many different types of cancer, including glioma (Basanta et al. 2008), prostate cancer (Basanta et al. 2012), and multiple myeloma (Dingli et al. 2009; Pacheco et al. 2014) or general tumors (Kaznatcheev et al. 2017). Also, evolutionary game theory is often used in spatially structured populations to answer questions about the effect of environmental constraints on tumor composition and invasiveness of cancer cells (Kaznatcheev et al. 2015; Gerlee and Anderson 2007; Anderson et al. 2009). However, including spatial structure in order to increase the realism of the model leads to a large number of additional assumptions and potential pitfalls (Zukewich et al. 2013; Hindersin and Traulsen 2015). Other

modeling approaches for dormancy in cancer focus on the interaction between the immune system and the tumor (Wilkie and Hahnfeldt 2013; Hahnfeldt et al. 2003; Page and Uhr 2005), the effect of angiogenesis (Kareva 2016), or spatial competition between cells (Enderling et al. 2009). However, these approaches do not explicitly model the conversion between phenotypes and its consequences under therapy with varying strength.

Thus, we decided to simplify our model on several levels: (i) We do not take spatial structure into account. (ii) We abstract from the interaction with the immune system. (iii) We concentrate on two tumor cell phenotypes – rapidly proliferating and dormant cells – although other cell phenotypes, such as fast migrating cells, cells mimicking vasculature, (cancer) stem cells and invading immune cells (e.g. Adamski et al. 2017b; Hattermann et al. 2016; Held-Feindt et al. 2010), also contribute to the whole tumor mass. (iv) We focus on the fitness of the respective phenotypes rather than the potentially underlying reasons for phenotypical changes (e.g. genetic or epigenetic changes).

An important aspect of our model is the conversion between the different cell phenotypes. Recent studies suggest that dormant cells may originate from "normal" tumor cells by currently intensively investigated mechanisms (e.g. Hoppe-Seyler et al. 2017; Sosa et al. 2015; Ranganathan et al. 2006). As a fundamental criterion for tumor dormancy, dormant cells need to be able to reawake and start growing again, so that they then phenotypically resemble the rapidly growing cell phenotype. Thus, we introduced a conversion factor σ into our model capturing these phenotypical transformation processes. Whether such conversions occur spontaneously or can be induced specifically or randomly by extrinsic or intrinsic mechanisms is poorly understood. We thus assumed a spontaneous event which can be modeled by a constant rate.

Using our theoretical approach we showed that dependent on the applied treatment strength an equilibrium balance between rapidly proliferating cells and dormant cells is eventually reached. At this fixed point and with low dosage or no treatment, mostly rapidly proliferating cells dominate the population, similar to the findings of Basanta and colleagues (Basanta et al. 2008). At stronger treatment, the fraction of dormant cells becomes successively larger, yielding a lower growth rate of the whole tumor. However, high dosage treatment cannot be applied for longer time periods, as it causes severe side effects (for GBM: Niewald et al. 2011). Hence, we focused on alternative treatment schedules.

Several previous models discuss the effect different treatment schedules on various aspects of the cancer growth like angiogenesis (Hahnfeldt et al. 2003; Schättler et al. 2016) or evolution of resistance (Dhawan et al. 2017). Here, either the dosage and timing or the type of the chemotherapeutic drug is varied, which can have a massive effect on the growth of the tumor. Accordingly, we combine sequential cycles of low and high dose treatment with different durations. Thereby, we observed that the

total growth of the tumor is considerably lower for fast switching compared to a slow switching scheme.

5.4.1. Conclusion

In this study, we have developed a theoretical model to predict the tumor growth kinetics under different treatment strengths including a dormant cell phenotype and underlined our theoretical approach with experimental data. Using our model which allows for phenotypic conversion, we could simulate how different tumor cell phenotypes proportionally contribute to the growing tumor mass in cycling treatment schedules. Additionally, we could observe that switching between high and low dosage treatment (with equal total treatment amounts) remarkably affects tumor growth in a frequency dependent way.

Thus, the dynamic proportions between cell phenotypes should be taken into account in the optimization of treatment schedules in order to control tumor growth.

List of abbreviations

DMSO, dimethyl sulfoxide; FCS, fetal calf serum; GBM, glioblastoma *multiforme*; SD: standard deviation; TMZ, temozolomide; D, dormant cells; P, rapidly proliferating cells

5.5. Declarations

Ethics approval and consent to participate Not applicable.

Consent for publication Not applicable.

Availability of Data and material All data generated or analysed during this study are included in this published article. The python code to reproduce Figures 5.1, 5.3, 5.4, and 5.5 is open source and available under `https://github.com/marvinboe/TreatCycle`.

Competing interest The authors declare that they have no competing interests.

Funding This work was supported by a sponsorship of the Medical Faculty of the University Kiel ("Forschungsförderung 2016" to JHF and KH) and the Deutsche Forschungsgemeinschaft RTG2154 [project 7 (KH) and project 8 (JHF)]. The funding body did not have any influence in the design of the study, collection, analysis and interpretation of data and in writing of the manuscript.

Authors' contribution MB, JHF, AT and KH designed experimental and theoretical models and prepared the manuscript; MB and AT calculated and visualized the theoretical model; JHF, MS, RL and KH contributed to the experimental model; all authors read and approved the final manuscript.

Acknowledgements We thank Judith Becker, Martina Burmester, Fereshteh Ebrahim and Brigitte Rehmke for expert technical assistance.

6. Conclusion

6.1. Conclusion & Discussion

In this work I have discussed four different models in the context of somatic evolution on different scales. Both basic ingredients of any evolutionary process were involved — variation and selection — and each model put the spotlight on different aspects of somatic evolution:

1. The fitness effect of synonymous mutations on the molecular level
2. The impact of tissue structure on mutational load of cells via replicative ageing
3. A malignancy caused by a somatic mutation in a hierarchical tissue structure
4. The impact of treatment fluctuations on tumor growth

On the molecular level of somatic evolution, the mechanisms for storing and processing genetic information are the source for genetic variation and fitness differences in cells. Previous models of RNA translation do not provide an explanation for many recent experimental findings, including the strong dependency of observables of translation on the length of the translated gene as for example ribosome density and protein yield. Chapter 2 has shown a possible mechanistic explanation for these observations: When taking ribosome reinitiation into account, the time to steady state is significantly longer than previously thought and the longest part of the translation process occurs before a steady state is reached. This provides a possible explanation for the connection between speed of translation and protein yield, which has a fitness effect on the cells. This in turn could be a possible cause for the evolution of "optimal" (faster) synonymous codons in shorter transcripts compared to longer transcripts (Comeron et al. 1999). However, also other mechanisms like changes in RNA structure (Goodman et al. 2013) or changes in protein folding (Zhou et al. 2013) could potentially lead to a fitness effect of synonymous mutations.

Interestingly, the fitness effect of synonymous mutations might also play an essential role in the context of somatic evolution and cancer (Supek et al. 2014) and changes in the quantity of protein produced can have drastic consequences on cellular reaction networks (Bagowski and Ferrell 2001; Veening et al. 2008). Besides, it also implies potential complications for one of the commonly used measures to infer the strength of selection dN/dS, dividing the number of non-synonymous by the number

of supposedly neutral synonymous mutations (Yang and Bielawski 2000). For example, dN/dS is also used to determine the strength of selection in cancer mutations (Martincorena et al. 2017), which can have important therapeutic consequences for targeted treatment (Sawyers 2004). However, due to the fitness effect of synonymous mutations — predicted by the model of RNA translation presented in this work, but also shown in other studies (overview in Plotkin and Kudla 2011) — it might be necessary to further refine dN/dS (Zhou et al. 2010).

Moving away from the intracellular towards the intercellular level of somatic evolution, tissue in multicellular organisms is often structured. In most cases, this structure consists of stem cells, which stay within the organism for its lifetime, and differentiated cells, which constitute the bulk of the somatic cells and fulfill the tissue function (Tumbar et al. 2004; Blanpain and Fuchs 2009; Busch et al. 2015). Often, there are multiple intermediate steps from stem cells to mature functional cells which strongly amplify the rate of cells that are produced within the tissue (Dingli et al. 2007). This tissue structure has strong implications for both variation and selection in somatic cells.

In hierarchical tissue structures the long lasting stem cells usually divide far less often than differentiated functional cells and therefore have a smaller replicative age and a smaller risk for the accumulation of mutations. In chapter 3 I have shown how the replicative age distribution of stem cells changes through several rounds of differentiation which is strongly connected to the somatic variation in cells of the specific tissues. For strong self-renewal in the progenitor compartments, this variation will become much broader than in the stem cells. This in turn allows to infer parameters of the differentiation structure from measurements of replicative age in the differentiated tissue, for example via telomere length (Werner et al. 2015). Furthermore, control of replicative age within the tissue is important as a cancer prevention mechanism (Rodriguez-Brenes et al. 2013), as in most tissues the maximum number of cell divisions is limited (Hayflick 1965).

However, tissue structure not only has an effect on somatic variation as shown above, but also on selection (Nowak et al. 2003; Lieberman et al. 2005; Allen et al. 2017). Using the example of chronic myeloid leukemia (CML), in chapter 4 I presented a mutation which is neutral on the stem cell level, but is under strong positive selection within the differentiated tissue. This leads to significant expansion of cell number of the mutated cells (Werner et al. 2011) which finally leads to diagnosis of the disease. In the case of CML, a targeted treatment exists which basically reverses the selection effect of mutated cells (Dingli et al. 2008). However, in the model presented here, the long-term treatment success critically depends on the neutral evolutionary dynamics — or drift — on the stem cell level, which might be much slower than previously thought (Lenaerts et al. 2010).

If, as can be hypothesized from the above observation, stem cells divide mainly asymmetrically and existing stem cells rarely get replaced by others (Werner and

Sottoriva 2018), the time for successful fixation in the stem cell pool might be significantly prolonged such that the effect of somatic selection is reduced. However, since some cancer mutations initially might have deleterious fitness effects without the presence of other driver mutations (Bauer et al. 2014), it is generally not clear in which situations tissues should ideally suppress or amplify selection to reduce the risk for cancer (Michor et al. 2003b; Hindersin et al. 2016).

Since cancers arise from normal somatic tissue, their structure often also resembles a similar organization structure with slowly cycling cancer stem cells while the bulk tumor mass consists of "differentiated" tumor cells (Perez-Losada and Balmain 2003). However, often also other phenotypes are described in cancer like dormancy, which is potentially related to cancer stem cells (Adamski et al. 2017a; Li and Bhatia 2011). Dormant cells divide only rarely or not at all and therefore suffer a huge fitness disadvantage against rapidly proliferating cancer cells under normal growth conditions. However, by applying treatment, the environmental conditions for the cancer are drastically changed and accordingly the fitness effect is basically reversed (chapter 5). Since chemotherapeutic drugs mostly kill dividing cells, dormant cells are essentially immune or at least much less affected by the treatment than rapidly growing cells. By periodically cycling the treatment this competition between the phenotypes could be used to at least slow down the overall growth of the tumour.

Overall, this work shows that very different modelling approaches are necessary depending on the aspect of somatic evolution under consideration. In some cases stochasticity is central, as in this work for the totally asymmetric simple exclusion process to model RNA translation (chapter 2) or the extinction of the leukemic stem cell for the relapse of CML (chapter 4). In other cases the populations sizes are large, such that descriptions with ordinary differential equations are sufficient, as for the distribution of replicative age in downstream compartments or glioblastoma growth in fluctuating environments (chapters 3 and 5). Often observing a system at a fixed point by stability analysis (Strogatz 2014) or at a stationary state (Chou et al. 2011) can be sufficient to capture important features of the system; in some cases, though, the convenience of these stationary approaches can prevent new insights, as here specifically shown in the case of RNA translation.

The objective of the presented models is to be descriptive rather than predictive, that is they show possible underlying mechanisms in observable experimental results. Remarkably, even these rather abstract models provided specific predictions that could in principle be verified by experiments or are consistent with existing experimental observations.

6.2. Outlook

The future implications of the presented work can be divided into two categories: On the one hand gaining new insights by combining models with potentially new experimental techniques, and on the other hand application of evolutionary theory to help solving real-world problems. In the future further advances in experimental techniques and measurements are to be expected that might allow a more detailed look on some of the processes presented in this work.

In the field of RNA translation research the past 10 years have seen an incredible advancement of measurement techniques such as ribosome profiling (Ingolia et al. 2009) or even real time imaging of nascent proteins of RNA translation (Morisaki et al. 2016; Wu et al. 2016), which for example allows insights into initiation, elongation and location of translation in living cells. However, as demonstrated in chapter 2, models are still necessary to understand and interpret findings of these new experimental techniques.

A promising and relatively new technique to explore the structure of the hematopoietic system is the genetic tagging of cells - barcoding - which can be used to track clones throughout time and is already used extensively in this context (Blundell and Levy 2014; Levy et al. 2015; Bhang et al. 2015; Blundell et al. 2017). Results from these experiments help to further refine structured tissue models and to measure parameters such as division rates of hematopoietic stem cells which are otherwise difficult to obtain *in vivo* (Busch et al. 2015). However, experimental results are not straightforward to interpret, such as in the example of replicative age distributions presented in this work. Hence, knowledge frequently advances at the interplay of new experimental results along with ideas about the underlying mechanisms — the models.

Especially interesting for inferring the replicative age of cells is the measurement of telomere length, which is technically challenging due to the repetitive structure of telomeric DNA (Nussey et al. 2014). However, even though there is a general correlation between telomere length and replicative age, the exact connection is influenced by various other factors. New measurements (Zong et al. 2014) in conjunction with useful models (Rodriguez-Brenes and Peskin 2010) might therefore help to obtain more precise distributions of replicative age of cells. Eventually this information could be used to infer dynamical parameters of tissue structure within multicellular organisms.

Intriguingly, insights from studying processes of evolution can be and are currently applied in the treatment of diseases. This includes prevention of resistance in infectious diseases like bacterial or viral infections (Rosenbloom et al. 2012), or preventing the spread of infectious diseases within a population (Leventhal et al. 2012; Leventhal et al. 2015). In many cases it is desirable to manipulate evolutionary paths in

these complex fitness landscapes such that the impact of a specific pathogen will stay manageable in the long run (Nichol et al. 2015).

Furthermore, also for cancer treatment in humans it can be vital to understand somatic evolution: The initial treatment for cancer is generally successful, such that the size of the primary tumor gets significantly reduced. However, after the initial treatment there often is relapse with a resistant clone, which can severely limit the subsequent treatment options (Hanahan and Weinberg 2011). Most often the onset of metastasis, that is the dispersal of cancer cells and recolonization at distant sites throughout the body, is lethal for the organism (Chaffer and Weinberg 2011). In fact, both emergence of resistance and dispersal to distant locations are evolutionary processes and mathematical models can support their understanding (Iwasa et al. 2006; Yamamoto et al. 2015; Bozic and Nowak 2017). Broadening the quantitative knowledge of somatic evolution in the diverse tissue structures within humans can therefore potentially help to inform treatment decisions (Gatenby et al. 2011). Accordingly, evolutionary informed treatment approaches for cancer are already applied in clinical trials and yield promising results (Zhang et al. 2017).

Overall, this work has shown that insights into processes of somatic evolution can be gained by combining mathematical modeling with new experimental observations. As the cooperation between the experimental and theoretical biologists becomes more common and experimental techniques become more refined, I expect many more interesting insights in the area of somatic evolution in the near future.

Bibliography

Adams, P. D., H. Jasper, and K. L. Rudolph (2015). "Aging-Induced Stem Cell Mutations as Drivers for Disease and Cancer". *Cell Stem Cell* **16**, pp. 601–612 (page 46).

Adamski, V., A. Hempelmann, C. Flüh, R. Lucius, M. Synowitz, K. Hattermann, and J. Held-Feindt (2017a). "Dormant glioblastoma cells acquire stem cell characteristics and are differentially affected by Temozolomide and AT101 treatment". *Oncotarget* **8**, pp. 108064–108078 (pages 97, 100, 106, 113).

Adamski, V., A. D. Schmitt, C. Flüh, M. Synowitz, K. Hattermann, and J. Held-Feindt (2017b). "Isolation and Characterization of Fast-Migrating Human Glioma Cells in the Progression of Malignant Gliomas". *Oncology Research Featuring Preclinical and Clinical Cancer Therapeutics* **25**, pp. 341–353 (page 107).

Afonina, Z. A., A. G. Myasnikov, V. A. Shirokov, B. P. Klaholz, and A. S. Spirin (2015). "Conformation transitions of eukaryotic polyribosomes during multi-round translation". *Nucleic Acids Research* **43**, pp. 618–628 (page 21).

Aktipis, C. A., A. M. Boddy, G. Jansen, U. Hibner, M. E. Hochberg, C. C. Maley, and G. S. Wilkinson (2015). "Cancer across the tree of life: cooperation and cheating in multicellularity". *Philosophical Transactions of the Royal Society B: Biological Sciences* **370**, p. 20140219 (pages 1, 3).

Alberts, B. (2015). *Molecular biology of the cell*. 6th ed. Garland Science (page 15).

Alekhina, O. M., K. S. Vassilenko, and A. S. Spirin (2007). "Translation of non-capped mRNAs in a eukaryotic cell-free system: acceleration of initiation rate in the course of polysome formation". *Nucleic Acids Research* **35**, pp. 6547–6559 (pages 21, 22, 24).

Allen, B., G. Lippner, Y.-T. Chen, B. Fotouhi, N. Momeni, S.-T. Yau, and M. A. Nowak (2017). "Evolutionary dynamics on any population structure". *Nature* **544**, pp. 227–230 (page 112).

Almog, N. (2010). "Molecular mechanisms underlying tumor dormancy". *Cancer Letters* **294**, pp. 139–146 (page 92).

Almog, N., L. Ma, R. Raychowdhury, C. Schwager, R. Erber, S. Short, L. Hlatky, P. Vajkoczy, P. E. Huber, J. Folkman, and A. Abdollahi (2009). "Transcriptional Switch of Dormant Tumors to Fast-Growing Angiogenic Phenotype". *Cancer Research* **69**, pp. 836–844 (pages 97, 100, 106).

Altrock, P. M., C. Brendel, R. Renella, S. H. Orkin, D. A. Williams, and F. Michor (2016). "Mathematical modeling of erythrocyte chimerism informs genetic intervention strategies for sickle cell disease". *American Journal of Hematology* **91**, pp. 931–937 (pages 61, 62).

Altrock, P. M., L. L. Liu, and F. Michor (2015). "The mathematics of cancer: integrating quantitative models". *Nature Reviews Cancer* **15**, pp. 730–745 (page 92).

Amrani, N., S. Ghosh, D. A. Mangus, and A. Jacobson (2008). "Translation factors promote the formation of two states of the closed-loop mRNP". *Nature* **453**, pp. 1276–1280 (page 33).

Anderson, A. R. A., M. Hassanein, K. M. Branch, J. Lu, N. A. Lobdell, J. Maier, D. Basanta, B. Weidow, A. Narasanna, C. L. Arteaga, A. Reynolds, V. Quaranta, L. Estrada, and A. M. Weaver (2009). "Microenvironmental independence associated with tumor progression". *Cancer Research* **69**, pp. 8797–8806 (page 106).

Anderson, A. R., A. M. Weaver, P. T. Cummings, and V. Quaranta (2006). "Tumor Morphology and Phenotypic Evolution Driven by Selective Pressure from the Microenvironment". *Cell* **127**, pp. 905–915 (page 89).

Arava, Y. (2005). "Dissecting eukaryotic translation and its control by ribosome density mapping". *Nucleic Acids Research* **33**, pp. 2421–2432 (page 20).

Arava, Y., Y. Wang, J. D. Storey, C. L. Liu, P. O. Brown, and D. Herschlag (2003). "Genome-wide analysis of mRNA translation profiles in Saccharomyces cerevisiae". *Proceedings of the National Academy of Sciences* **100**, pp. 3889–3894 (pages 20, 21, 25, 26, 31, 40).

Archer, S. K., N. E. Shirokikh, C. V. Hallwirth, T. H. Beilharz, and T. Preiss (2015). "Probing the closed-loop model of mRNA translation in living cells". *RNA Biology* **12**, pp. 248–254 (page 33).

Armitage, P. and R. Doll (2004). "The age distribution of cancer and a multi-stage theory of carcinogenesis". *British Journal of Cancer* **91**, pp. 1983–1989 (page 6).

Ashcroft, P., M. G. Manz, and S. Bonhoeffer (2017). "Clonal dominance and transplantation dynamics in hematopoietic stem cell compartments". *PLOS Computational Biology* **13**, e1005803 (pages 61, 62, 76, 86).

Ashcroft, P., F. Michor, and T. Galla (2015). "Stochastic tunneling and metastable states during the somatic evolution of cancer". *Genetics* **199**, pp. 1213–1228 (page 11).

Baerlocher, G. M., I. Vulto, G. de Jong, and P. M. Lansdorp (2006). "Flow cytometry and FISH to measure the average length of telomeres (flow FISH)". *Nature Protocols* **1**, pp. 2365–2376 (page 46).

Baglioni, C., C. Vesco, and M. Jacobs-Lorena (1969). "The role of ribosomal subunits in mammalian cells." In: *Cold Spring Harbor symposia on quantitative biology*. Vol. 34. Cold Spring Harbor Laboratory Press, pp. 555–565 (page 21).

Bagowski, C. P. and J. E. Ferrell (2001). "Bistability in the JNK cascade". *Current Biology* **11**, pp. 1176–1182 (page 111).

Basanta, D., J. G. Scott, M. N. Fishman, G. Ayala, S. W. Hayward, and a. R. a. Anderson (2012). "Investigating prostate cancer tumour–stroma interactions: clinical and biological insights from an evolutionary game". *British Journal of Cancer* **106**, pp. 174–181 (page 106).

Basanta, D., M. Simon, H. Hatzikirou, and A. Deutsch (2008). "Evolutionary game theory elucidates the role of glycolysis in glioma progression and invasion". *Cell Proliferation* **41**, pp. 980–987 (pages 106, 107).

Basanta, D. and A. R. A. Anderson (2013). "Exploiting ecological principles to better understand cancer progression and treatment". *Interface Focus* **3**, p. 20130020 (page 92).

Basanta, D. and A. Deutsch (2008). "A Game Theoretical Perspective on the Somatic Evolution of Cancer". *Selected Topics in Cancer Modeling: Genesis, Evolution, Immune Competition, and Therapy*, pp. 97–112 (pages 93, 94).

Bauer, B., R. Siebert, and A. Traulsen (2014). "Cancer initiation with epistatic interactions between driver and passenger mutations". *Journal of Theoretical Biology* **358**, pp. 52–60 (page 113).

Berget, S. M., C. Moore, and P. A. Sharp (1977). "Spliced segments at the 5 terminus of adenovirus 2 late mRNA". *Proceedings of the National Academy of Sciences* **74**, pp. 3171–3175 (page 43).

Bhang, H.-e. C. et al. (2015). "Studying clonal dynamics in response to cancer therapy using high-complexity barcoding". *Nature Medicine* **21**, pp. 440–448 (page 114).

Blackburn, E. H. (1991). "Structure and function of telomeres." *Nature* **350**, pp. 569–573 (page 46).

Blagden, S. P. and A. E. Willis (2011). "The biological and therapeutic relevance of mRNA translation in cancer". *Nature Reviews Clinical Oncology* **8**, pp. 280–291 (page 15).

Blanpain, C. and E. Fuchs (2009). "Epidermal homeostasis: a balancing act of stem cells in the skin". *Nature Reviews Molecular Cell Biology* **10**, pp. 207–217 (pages 46, 112).

Blasco, M. A. (2007). "Telomere length, stem cells and aging." *Nature chemical biology* **3**, pp. 640–649 (page 46).

Blundell, J. R. and S. F. Levy (2014). "Beyond genome sequencing: Lineage tracking with barcodes to study the dynamics of evolution, infection, and cancer". *Genomics* **104**, pp. 1–14 (page 114).

Blundell, J. R., K. Schwartz, D. Francois, D. S. Fisher, G. J. Sherlock, and S. F. Levy (2017). "The dynamics of adaptive genetic diversity during the early stages of clonal evolution". *bioarxiv*, pp. 1–40 (page 114).

Bomze, I. M. and R. Bürger (1995). "Stability by Mutation in Evolutionary Games". *Games and Economic Behavior* **11**, pp. 146–172 (page 93).

Böttcher, M. A., J. Held-Feindt, M. Synowitz, R. Lucius, A. Traulsen, and K. Hattermann (2018a). "Modeling treatment-dependent glioma growth including a dormant tumor cell subpopulation". *BMC Cancer* **18**, p. 376 (page 89).

Böttcher, M. A., B. Werner, D. Dingli, and A. Traulsen (2018b). "Replicative cellular age distributions in compartmentalised tissues". *bioRxiv preprint*, pp. 1–27 (page 43).

Bozic, I. and M. A. Nowak (2017). "Resisting Resistance". *Annual Review of Cancer Biology* **1**, pp. 203–221 (page 115).

Bragado, P., M. S. Sosa, P. Keely, J. Condeelis, and J. A. Aguirre-Ghiso (2012). "Microenvironments Dictating Tumor Cell Dormancy". In: *Minimal Residual Disease and Circulating Tumor Cells in Breast Cancer*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 25–39 (page 92).

Branford, S., J. F. Seymour, A. Grigg, C. Arthur, Z. Rudzki, K. Lynch, and T. Hughes (2007). "BCR-ABL Messenger RNA Levels Continue to Decline in Patients with Chronic Phase Chronic Myeloid Leukemia Treated with Imatinib for More Than 5 Years and Approximately Half of All First-Line Treated Patients Have Stable Undetectable BCR-ABL Using Strict Se". *Clinical Cancer Research* **13**, pp. 7080–7085 (page 86).

Breivik, J. and G. Gaudernack (2004). "Resolving the evolutionary paradox of genetic instability: a cost-benefit analysis of DNA repair in changing environments". *FEBS Letters* **563**, pp. 7–12 (page 4).

Brümmendorf, T. H., T. L. Holyoake, N. Rufer, M. J. Barnett, M. Schulzer, C. J. Eaves, A. C. Eaves, and P. M. Lansdorp (2000). "Prognostic implications of differences in telomere length between normal and malignant cells from patients with chronic myeloid leukemia measured by flow cytometry." *Blood* **95**, pp. 1883–1890 (pages 60, 62).

Brümmendorf, T. H. et al. (2003). "Normalization of previously shortened telomere length under treatment with imatinib argues against a preexisting telomere length deficit in

normal hematopoietic stem cells from patients with chronic myeloid leukemia." *Annals of the New York Academy of Sciences* **996**, pp. 26–38 (pages 61, 63).

Busch, K., K. Klapproth, M. Barile, M. Flossdorf, T. Holland-Letz, S. M. Schlenner, M. Reth, T. Höfer, and H.-R. Rodewald (2015). "Fundamental properties of unperturbed haematopoiesis from stem cells in vivo". *Nature* **518**, pp. 542–546 (pages 6, 46, 48, 86, 112, 114).

Bystrykh, L. V., E. Verovskaya, E. Zwart, M. Broekhuis, and G. de Haan (2012). "Counting stem cells: methodological constraints". *Nature Methods* **9**, pp. 567–574 (page 46).

Cai, L., N. Friedman, and X. S. Xie (2006). "Stochastic protein expression in individual cells at the single molecule level." *Nature* **440**, pp. 358–62 (page 11).

Callaghan, R., F. Luk, and M. Bebawy (2014). "Inhibition of the Multidrug Resistance P-Glycoprotein: Time for a Change of Strategy?" *Drug Metabolism and Disposition* **42**, pp. 623–631 (page 7).

Cao, D. and R. Parker (2001). "Computational modeling of eukaryotic mRNA turnover". *RNA* **7**, pp. 1192–1212 (page 40).

Cao, Y., H. Li, and L. Petzold (2004). "Efficient formulation of the stochastic simulation algorithm for chemically reacting systems". *Journal of Chemical Physics* **121**, pp. 4059–4067 (pages 13, 77).

Chaffer, C. L. and R. A. Weinberg (2011). "A Perspective on Cancer Cell Metastasis". *Science* **331**, pp. 1559–1564 (pages 7, 115).

Chandrasekhar, S. (1943). "Stochastic Problems in Physics and Astronomy". *Reviews of Modern Physics* **15**, pp. 1–89 (page 11).

Charneski, C. A. and L. D. Hurst (2013). "Positively Charged Residues Are the Major Determinants of Ribosomal Velocity". *PLoS Biology* **11**, e1001508 (page 27).

Chen, C.-Y. A. and A.-B. Shyu (2011). "Mechanisms of deadenylation-dependent decay". *Wiley Interdisciplinary Reviews: RNA* **2**, pp. 167–183 (page 40).

Chen, C. A., N. Ezzeddine, and A. Shyu (2008). "Chapter 17 Messenger RNA Half-Life Measurements in Mammalian Cells". In: *Methods Enzymology*. Vol. 448, pp. 335–357 (pages 40, 42).

Chou, T., K. Mallick, and R. K. P. Zia (2011). "Non-equilibrium statistical mechanics: from a paradigmatic model to biological transport". *Reports on Progress in Physics* **74**, p. 116601 (pages 22, 113).

Chu, D., E. Kazana, N. Bellanger, T. Singh, M. F. Tuite, and T. von der Haar (2014a). "Translation elongation can control translation initiation on eukaryotic mRNAs". *The EMBO Journal* **33**, pp. 21–34 (pages 21, 25, 28, 34).

Chu, D., J. Thompson, and T. von der Haar (2014b). "Charting the dynamics of translation". *BioSystems* **119**, pp. 1–9 (page 32).

Ciandrini, L., I. Stansfield, and M. C. Romano (2013). "Ribosome Traffic on mRNAs Maps to Gene Ontology: Genome-wide Quantification of Translation Initiation Rates and Polysome Size Regulation". *PLoS Computational Biology* **9**, e1002866 (pages 20, 21, 25, 26, 31, 34).

Coghlan, A. and K. H. Wolfe (2000). "Relationship of codon bias to mRNA concentration and protein length inSaccharomyces cerevisiae". *Yeast* **16**, pp. 1131–1145 (page 34).

Collado, M., M. A. Blasco, and M. Serrano (2007). "Cellular Senescence in Cancer and Aging". *Cell* **130**, pp. 223–233 (page 46).

Comeron, J. M., M. Kreitman, and M. Aguadé (1999). "Natural selection on synonymous sites is correlated with gene length and recombination in Drosophila." *Genetics* **151**, pp. 239–249 (pages 31, 33, 34, 111).

Corrie, P. G. (2008). "Cytotoxic chemotherapy: clinical aspects". *Medicine* **36**, pp. 24–28 (page 6).

Costello, J., L. M. Castelli, W. Rowe, C. J. Kershaw, D. Talavera, S. S. Mohammad-Qureshi, P. F. G. Sims, C. M. Grant, G. D. Pavitt, S. J. Hubbard, and M. P. Ashe (2015). "Global mRNA selection mechanisms for translation initiation". *Genome Biology* **16**, p. 10 (page 33).

Craig, M., A. R. Humphries, and M. C. Mackey (2016). "An upper bound for the half-removal time of neutrophils from circulation". *Blood* **128**, pp. 1989–1991 (page 60).

Crick, F. H. C. (1958). "On protein synthesis." *Symposia of the Society for Experimental Biology* **12**, pp. 138–63 (pages 3, 15, 16).

Cross, N. C. P. et al. (2015). "Laboratory recommendations for scoring deep molecular responses following treatment for chronic myeloid leukemia". *Leukemia* **29**, pp. 999–1003 (pages 74, 79, 80).

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**, pp. 824–830 (page 73).

Darwin, C. (1859). *On the Origin of the Species*. Vol. 5. London: John Murray, p. 386 (pages 1, 3, 5, 7, 89).

Davidson, E. H. (2006). *The Regulatory Genome*. Elsevier (pages 1, 15).

Decker, C. J. and R. Parker (1993). "A turnover pathway for both stable and unstable mRNAs in yeast: evidence for a requirement for deadenylation." *Genes & Development* **7**, pp. 1632–1643 (page 40).

Derényi, I. and G. J. Szöllősi (2017). "Hierarchical tissue organization as a general mechanism to limit the accumulation of somatic mutations". *Nature Communications* **8**, p. 14545 (pages 43, 46, 86).

Dhawan, A., D. Nichol, F. Kinose, M. E. Abazeed, A. Marusyk, E. B. Haura, and J. G. Scott (2017). "Collateral sensitivity networks reveal evolutionary instability and novel treatment strategies in ALK mutated non-small cell lung cancer". *Scientific Reports* **7**, pp. 1–9 (page 107).

Ding, Y., P. Shah, and J. B. Plotkin (2012). "Weak 5-mRNA Secondary Structures in Short Eukaryotic Genes". *Genome Biology and Evolution* **4**, pp. 1046–1053 (page 34).

Dingli, D., F. A. C. C. Chalub, F. C. Santos, S. Van Segbroeck, and J. M. Pacheco (2009). "Cancer phenotype as the outcome of an evolutionary game between normal and malignant cells". *British Journal of Cancer* **101**, pp. 1130–1136 (pages 93, 106).

Dingli, D. and F. Michor (2006). "Successful Therapy Must Eradicate Cancer Stem Cells". *Stem Cells* **24**, pp. 2603–2610 (page 52).

Dingli, D., A. Traulsen, and J. M. Pacheco (2007). "Compartmental Architecture and Dynamics of Hematopoiesis". *PLoS ONE* **2**, e345 (pages 6, 46, 52, 60, 61, 75, 77, 78, 112).

Dingli, D., A. Traulsen, and J. M. Pacheco (2008). "Chronic Myeloid Leukemia: Origin, Development, Response to Therapy, and Relapse". *Clinical Leukemia* **2**, pp. 133–139 (pages 60, 62, 74, 76, 78, 112).

Dolberg, D. S. and M. J. Bissell (1984). "Inability of Rous sarcoma virus to cause sarcomas in the avian embryo". *Nature* **309**, pp. 552–556 (page 89).

Donohue, D. M., R. H. Reiff, M. L. Hanson, Y. Betson, and C. A. Finch (1958). "Quantitative Measurement of the Erythrocytic and Granulocytic Cells of the Marrow and Blood". *Journal of Clinical Investigation* **37**, pp. 1571–1576 (page 60).

Dreyfus, M. and P. Régnier (2002). "The Poly(A) Tail of mRNAs". *Cell* **111**, pp. 611–613 (page 40).

Duret, L. and D. Mouchiroud (1999). "Expression pattern and, surprisingly, gene length shape codon usage in Caenorhabditis, Drosophila, and Arabidopsis". *Proceedings of the National Academy of Sciences* **96**, pp. 4482–4487 (page 34).

Dykstra, B. and G. de Haan (2008). "Hematopoietic stem cell aging and self-renewal". *Cell and Tissue Research* **331**, pp. 91–101 (page 61).

Egan, D. and J. Radich (2016). "Monitoring disease burden in chronic myeloid leukemia: Past, present, and future". *American Journal of Hematology* **91**, pp. 742–746 (pages 78, 79).

Eigen, M., B. Lindemann, M. Tietze, R. Winkler-Oswatitsch, A. Dress, and A. von Haeseler (1989). "How old is the genetic code? Statistical geometry of tRNA provides an answer". *Science* **244**, pp. 673–679 (page 4).

Einstein, A. (1905). "Über die von der molekularekinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssikkeiten suspendierten Teilchen". *Annalen der Physik* **17**, pp. 549–560 (page 11).

Ellis, R. E., J. Yuan, and H. R. Horvitz (1991). "Mechanisms and Functions of Cell Death". *Annual Review of Cell Biology* **7**, pp. 663–698 (pages 1, 4).

Elowitz, M. B., A. J. Levine, E. D. Siggia, and P. S. Swain (2002). "Stochastic gene expression in a single cell." *Science (New York, N.Y.)* **297**, pp. 1183–1186 (page 11).

Endaya, B. B., P. Y. Lam, A. C. Meedeniya, and J. Neuzil (2016). "Transcriptional profiling of dividing tumor cells detects intratumor heterogeneity linked to cell proliferation in a brain tumor model". *Molecular Oncology* **10**, pp. 126–137 (page 106).

Enderling, H., A. R. Anderson, M. A. Chaplain, A. Beheshti, L. Hlatky, and P. Hahnfeldt (2009). "Paradoxical Dependencies of Tumor Dormancy and Progression on Basic Cell Kinetics". *Cancer Research* **69**, pp. 8814–8821 (page 107).

Etienne, G. et al. (2016). "Long-Term Follow-Up of the French Stop Imatinib (STIM1) Study in Patients With Chronic Myeloid Leukemia". *Journal of Clinical Oncology* **126**, p. 345 (pages 74, 85).

Evans, E. B. (2015). "New insights into tumor dormancy: Targeting DNA repair pathways". *World Journal of Clinical Oncology* **6**, p. 80 (page 106).

Ferlay, J., I. Soerjomataram, R. Dikshit, S. Eser, C. Mathers, M. Rebelo, D. M. Parkin, D. Forman, and F. Bray (2015). "Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012". *International Journal of Cancer* **136**, E359–E386 (page 6).

Fernandes, L. D., A. P. S. de Moura, and L. Ciandrini (2017). "Gene length as a regulator for ribosome recruitment and protein synthesis: theoretical insights". *Scientific Reports* **7**, p. 17409 (page 33).

Fialkow, P. J., R. J. Jacobson, and T. Papayannopoulou (1977). "Chronic myelocytic leukemia: Clonal origin in a stem cell common to the granulocyte, erythrocyte, platelet and monocyte/macrophage". *The American Journal of Medicine* **63**, pp. 125–130 (page 73).

Fokas, A. S., J. B. Keller, and B. D. Clarkson (1991). "Mathematical model of granulocytopoiesis and chronic myelogenous leukemia." *Cancer research* **51**, pp. 2084–91 (page 87).

Fonseca, L. L. and E. O. Voit (2015). "Comparison of mathematical frameworks for modeling erythropoiesis in the context of malaria infection". *Mathematical Biosciences* **270**, pp. 224–236 (pages 47, 61).

Gage, F. H. (2000). "Mammalian Neural Stem Cells". *Science* **287**, pp. 1433–1438 (page 6).

Gardiner, C. W. (2009). *Stochastic methods: A Handbook for the Natural and Social Sciences*. 4th. Springer (pages 5, 11).

Gatenby, R. a. and R. J. Gillies (2008). "A microenvironmental model of carcinogenesis". *Nature Reviews Cancer* **8**, pp. 56–61 (page 89).

Gatenby, R. a., R. J. Gillies, and J. S. Brown (2011). "Of cancer and cave fish". *Nature Reviews Cancer* **11**, pp. 237–238 (pages 89, 92, 115).

Gautschi, W. (1998). "The Incomplete Gamma Functions Since Tricomi". In: *Tricomi's Ideas and Contemporary Applied Mathematics, Atti dei Convegni Lincei, n. 147, Accademia Nazionale dei Lincei*, pp. 203–237 (pages 55, 68).

Geisberg, J. V., Z. Moqtaderi, X. Fan, F. Ozsolak, and K. Struhl (2014). "Global Analysis of mRNA Isoform Half-Lives Reveals Stabilizing and Destabilizing Elements in Yeast". *Cell* **156**, pp. 812–824 (pages 41, 42).

Gerlee, P. and A. R. A. Anderson (2007). "An evolutionary hybrid cellular automaton model of solid tumour growth". *Journal of Theoretical Biology* **246**, pp. 583–603 (page 106).

Ghaemmaghami, S., W.-K. Huh, K. Bower, R. W. Howson, A. Belle, N. Dephoure, E. K. O'Shea, and J. S. Weissman (2003). "Global analysis of protein expression in yeast". *Nature* **425**, pp. 737–741 (page 20).

Gibson, M. A. and J. Bruck (2000). "Efficient Exact Stochastic Simulation of Chemical Systems with Many Species and Many Channels". *The Journal of Physical Chemistry A* **104**, pp. 1876–1889 (page 12).

Gilchrist, M. A. and A. Wagner (2006). "A model of protein translation including codon bias, nonsense errors, and ribosome recycling". *Journal of Theoretical Biology* **239**, pp. 417–434 (pages 31, 32, 37, 39, 42).

Gillespie, D. T. (1977). "Exact stochastic simulation of coupled chemical reactions". *The Journal of Physical Chemistry* **81**, pp. 2340–2361 (pages 12, 24, 77).

Gillespie, D. T. (2007). "Stochastic Simulation of Chemical Kinetics". *Annual Review of Physical Chemistry* **58**, pp. 35–55 (page 11).

Gingold, H. and Y. Pilpel (2011). "Determinants of translation efficiency and accuracy". *Molecular Systems Biology* **7**, pp. 1–13 (pages 15, 16).

Glauche, I., K. Moore, L. Thielecke, K. Horn, M. Loeffler, and I. Roeder (2009). "Stem cell proliferation and quiescence - Two sides of the same coin". *PLoS Computational Biology* **5**, pp. 3–12 (pages 61, 76).

Goodman, D. B., G. M. Church, and S. Kosuri (2013). "Causes and Effects of N-Terminal Codon Bias in Bacterial Genes". *Science* **342**, pp. 475–479 (page 111).

Gorissen, M. and C. Vanderzande (2012). "Ribosome Dwell Times and the Protein Copy Number Distribution". *Journal of Statistical Physics* **148**, pp. 628–636 (page 23).

Gotea, V., J. J. Gartner, N. Qutob, L. Elnitski, and Y. Samuels (2015). "The functional relevance of somatic synonymous mutations in melanoma and other cancers". *Pigment Cell & Melanoma Research* **28**, pp. 673–684 (page 16).

Graham, S. M., H. G. Jørgensen, E. Allan, C. Pearson, M. J. Alcorn, L. Richmond, and T. L. Holyoake (2002). "Primitive, quiescent, Philadelphia-positive stem cells from patients with chronic myeloid leukemia are insensitive to STI571 in vitro". *Blood* **99**, pp. 319–325 (page 74).

Gu, W., T. Zhou, and C. O. Wilke (2010). "A Universal Trend of Reduced mRNA Stability near the Translation-Initiation Site in Prokaryotes and Eukaryotes". *PLoS Computational Biology* **6**, e1000664 (page 34).

Gunaratne, J., A. Schmidt, A. Quandt, S. P. Neo, Ö. S. Saraç, T. Gracia, S. Loguercio, E. Ahrné, R. L. H. Xia, K. H. Tan, C. Lössner, J. Bähler, A. Beyer, W. Blackstock, and R. Aebersold (2013). "Extensive Mass Spectrometry-based Analysis of the Fission Yeast Proteome". *Molecular & Cellular Proteomics* **12**, pp. 1741–1751 (page 20).

Gurdon, J. B. and D. A. Melton (2008). "Nuclear Reprogramming in Cells". *Science* **322**, pp. 1811–1815 (page 5).

Gurdon, J. (1992). "The generation of diversity and pattern in animal development". *Cell* **68**, pp. 185–199 (pages 5, 15, 43).

von der Haar, T. (2008). "A quantitative estimation of the global translational activity in logarithmically growing yeast cells". *BMC Systems Biology* **2**, p. 87 (page 41).

Hahnfeldt, P., J. Folkman, and L. Hlatky (2003). "Minimizing Long-Term Tumor Burden: The Logic for Metronomic Chemotherapeutic Dosing and its Antiangiogenic Basis". *Journal of Theoretical Biology* **220**, pp. 545–554 (page 107).

Hamilton, A., G. V. Helgason, M. Schemionek, B. Zhang, S. Myssina, E. K. Allan, F. E. Nicolini, C. Muller-Tidow, R. Bhatia, V. G. Brunton, S. Koschmieder, and T. L. Holyoake (2012). "Chronic myeloid leukemia stem cells are not dependent on Bcr-Abl kinase activity for their survival". *Blood* **119**, pp. 1501–1510 (page 74).

Hanahan, D. and R. A. Weinberg (2011). "Hallmarks of Cancer: The Next Generation". *Cell* **144**, pp. 646–674 (pages 6, 62, 115).

Hanawalt, P. C. and G. Spivak (2008). "Transcription-coupled DNA repair: two decades of progress and surprises". *Nature Reviews Molecular Cell Biology* **9**, pp. 958–970 (page 1).

Harley, C. B., A. B. Futcher, and C. W. Greider (1990). "Telomeres shorten during ageing of human fibroblasts". *Nature* **345**, pp. 458–460 (page 46).

Hattermann, K., C. Flüh, D. Engel, H. M. Mehdorn, M. Synowitz, R. Mentlein, and J. Held-Feindt (2016). "Stem cell markers in glioma progression and recurrence". *International Journal of Oncology* **49**, pp. 1899–1910 (page 107).

Hayflick, L. (1965). "the Limited in Vitro Lifetime of Human Diploid Cell Strains." *Experimental cell research* **37**, pp. 614–636 (page 112).

Held-Feindt, J., K. Hattermann, S. S. Müerköster, H. Wedderkopp, F. Knerlich-Lukoschus, H. Ungefroren, H. M. Mehdorn, and R. Mentlein (2010). "CX3CR1 promotes recruitment of human glioma-infiltrating microglia/macrophages (GIMs)". *Experimental Cell Research* **316**, pp. 1553–1566 (page 107).

Hendrickson, D. G., D. J. Hogan, H. L. McCullough, J. W. Myers, D. Herschlag, J. E. Ferrell, and P. O. Brown (2009). "Concordant Regulation of Translation and mRNA Abundance for Hundreds of Targets of a Human microRNA". *PLoS Biology* **7**, e1000238 (pages 20, 21, 26, 31, 40).

Heppner, G. H. (1984). "Tumor heterogeneity". *Cancer research* **44**, pp. 2259–2265 (page 89).

Hindersin, L. and A. Traulsen (2015). "Most Undirected Random Graphs Are Amplifiers of Selection for Birth-Death Dynamics, but Suppressors of Selection for Death-Birth Dynamics". *PLOS Computational Biology* **11**, e1004437 (pages 9, 106).

Hindersin, L., B. Werner, D. Dingli, and A. Traulsen (2016). "Should tissue structure suppress or amplify selection to minimize cancer risk?" *Biology Direct* **11**, p. 41 (page 113).

Hirose, Y., M. S. Berger, and R. O. Pieper (2001). "p53 Effects Both the Duration of G 2 /M Arrest and the Fate of Temozolomide-treated Human Glioblastoma Cells 1". *Cancer research* **61**, pp. 1957–1963 (page 92).

Hochhaus, A., R. A. Larson, F. Guilhot, J. P. Radich, S. Branford, T. P. Hughes, M. Baccarani, M. W. Deininger, F. Cervantes, S. Fujihara, C.-E. Ortmann, H. D. Menssen, H. Kantarjian, S. G. O'Brien, and B. J. Druker (2017). "Long-Term Outcomes of Imatinib Treatment for Chronic Myeloid Leukemia". *New England Journal of Medicine* **376**, pp. 917–927 (page 74).

Hofbauer, J., P. Schuster, and K. Sigmund (1979). "A note on evolutionary stable strategies and game dynamics". *Journal of Theoretical Biology* **81**, pp. 609–612 (page 8).

Hofbauer, J. and K. Sigmund (1998). "Evolutionary Games and Population Dynamics". In: Cambridge: Cambridge University Press (pages 7, 92).

Hofstetter, C. P., J.-K. Burkhardt, B. J. Shin, D. B. Gürsel, L. Mubita, R. Gorrepati, C. Brennan, E. C. Holland, and J. A. Boockvar (2012). "Protein Phosphatase 2A Mediates Dormancy of Glioblastoma Multiforme-Derived Tumor Stem-Like Cells during Hypoxia". *PLoS ONE* **7**, e30059 (page 106).

Holbek, S., K. M. Bendtsen, and J. Juul (2013). "Moderate stem-cell telomere shortening rate postpones cancer onset in a stochastic model". *Physical Review E* **88**, p. 042706 (page 62).

Hoppe-Seyler, K., F. Bossler, C. Lohrey, J. Bulkescher, F. Rösl, L. Jansen, A. Mayer, P. Vaupel, M. Dürst, and F. Hoppe-Seyler (2017). "Induction of dormancy in hypoxic human papillomavirus-positive cancer cells". *Proceedings of the National Academy of Sciences* **114**, E990–E998 (page 107).

Hughes, T. P. and D. M. Ross (2016). "Moving treatment-free remission into mainstream clinical practice in CML". *Blood* **128**, pp. 17–23 (page 74).

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. Gilliland (2004). "MOZ-TIF2, but not BCR-ABL, confers properties of leukemic stem cells to committed murine hematopoietic progenitors". *Cancer Cell* **6**, pp. 587–596 (pages 76, 86).

Hurst, L. D. (2002). "The Ka/Ks ratio: Diagnosing the form of sequence evolution". *Trends in Genetics* **18**, pp. 486–487 (page 17).

Hynes, R. O. and A. Naba (2012). "Overview of the Matrisome–An Inventory of Extracellular Matrix Constituents and Functions". *Cold Spring Harbor Perspectives in Biology* **4**, a004903 (pages 1, 5, 89).

Imagawa, J. et al. (2015). "Discontinuation of dasatinib in patients with chronic myeloid leukaemia who have maintained deep molecular response for longer than 1 year (DADI trial): a multicentre phase 2 trial". *The Lancet Haematology* **2**, e528–e535 (pages 74, 80).

Ingolia, N. T., L. F. Lareau, and J. S. Weissman (2011). "Ribosome Profiling of Mouse Embryonic Stem Cells Reveals the Complexity and Dynamics of Mammalian Proteomes". *Cell* **147**, pp. 789–802 (page 41).

Ingolia, N. T., S. Ghaemmaghami, J. R. S. Newman, and J. S. Weissman (2009). "Genome-Wide Analysis in Vivo of Translation with Nucleotide Resolution Using Ribosome Profiling". *Science* **324**, pp. 218–223 (pages 20, 114).

Isenbarger, T. A., C. E. Carr, S. S. Johnson, M. Finney, G. M. Church, W. Gilbert, M. T. Zuber, and G. Ruvkun (2008). "The Most Conserved Genome Segments for Life Detection on Earth and Other Planets". *Origins of Life and Evolution of Biospheres* **38**, pp. 517–533 (page 4).

ISO/IEC (2011). *ISO International Standard 14882:2011(E) – Programming Language C++*. Geneva, Switzerland (page 13).

Iwasa, Y., F. Michor, and M. A. Nowak (2004). "Stochastic Tunnels in Evolutionary Dynamics". *Genetics* **166**, pp. 1571–1579 (page 11).

Iwasa, Y., M. A. Nowak, and F. Michor (2006). "Evolution of Resistance During Clonal Expansion". *Genetics* **172**, pp. 2557–2566 (page 115).

Jackson, D. A., A. Pombo, and F. Iborra (2000). "The balance sheet for transcription: an analysis of nuclear RNA metabolism in mammalian cells". *The FASEB Journal* **14**, pp. 242–254 (page 41).

Jan, C. H., C. C. Williams, and J. S. Weissman (2014). "Principles of ER cotranslational translocation revealed by proximity-specific ribosome profiling". *Science* **346**, p. 1257521 (page 33).

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**, pp. 4008–4013 (page 61).

Jørgensen, H. G., E. K. Allan, N. E. Jordanides, J. C. Mountford, and T. L. Holyoake (2007). "Nilotinib exerts equipotent antiproliferative effects to imatinib and does not induce apoptosis in CD34+ CML cells". *Blood* **109**, pp. 4016–4019 (page 74).

Ju, Y. S. et al. (2017). "Somatic mutations reveal asymmetric cellular dynamics in the early human embryo". *Nature* **543**, pp. 714–718 (page 46).

Kafri, M., E. Metzl-Raz, G. Jona, and N. Barkai (2016). "The Cost of Protein Production". *Cell Reports* **14**, pp. 22–31 (page 33).

van Kampen, N. G. (2007). *Stochastic Processes in Physics and Chemistry (Third Edition)*. Elsevier (page 11).

Kareva, I. (2016). "Escape from tumor dormancy and time to angiogenic switch as mitigated by tumor-induced stimulation of stroma". *Journal of Theoretical Biology* **395**, pp. 11–22 (page 107).

Kaveh, K., M. Kohandel, and S. Sivaloganathan (2016). "Replicator dynamics of cancer stem cell: Selection in the presence of differentiation and plasticity". *Mathematical Biosciences* **272**, pp. 64–75 (page 71).

Kaznatcheev, A., J. G. Scott, and D. Basanta (2015). "Edge effects in game-theoretic dynamics of spatially structured tumours". *Journal of The Royal Society Interface* **12**, p. 20150154 (pages 93, 106).

Kaznatcheev, A., R. Vander Velde, J. G. Scott, and D. Basanta (2017). "Cancer treatment scheduling and dynamic heterogeneity in social dilemmas of tumour acidity and vasculature". *British Journal of Cancer*, pp. 1–24 (page 106).

Kellis, M. et al. (2014). "Defining functional DNA elements in the human genome". *Proceedings of the National Academy of Sciences* **111**, pp. 6131–6138 (page 15).

Kempthorne, O. and E. Pollak (1970). "Concepts of fitness in mendelian populations." *Genetics* **64**, pp. 125–45 (page 7).

Kimura, M. (1968). "Evolutionary Rate at the Molecular Level". *Nature* **217**, pp. 624–626 (page 11).

Kimura, S. (2016). "Current status of ABL tyrosine kinase inhibitors stop studies for chronic myeloid leukemia". *Stem Cell Investigation* **3**, pp. 36–36 (pages 74, 80–82, 84, 86).

Klein, A. M. and B. D. Simons (2011). "Universal patterns of stem cell fate in cycling adult tissues". *Development* **138**, pp. 3103–3111 (page 9).

Kliman, R. M., N. Irving, and M. Santiago (2003). "Selection Conflicts, Gene Expression, and Codon Usage Trends in Yeast". *Journal of Molecular Evolution* **57**, pp. 98–109 (page 34).

Knoblich, J. A. (2008). "Mechanisms of Asymmetric Stem Cell Division". *Cell* **132**, pp. 583–597 (pages 48, 76, 86).

Kopeina, G. S., Z. A. Afonina, K. V. Gromova, V. A. Shirokov, V. D. Vasiliev, and A. S. Spirin (2008). "Step-wise formation of eukaryotic double-row polyribosomes and circular translation of polysomal mRNA". *Nucleic Acids Research* **36**, pp. 2476–2488 (pages 21, 32).

Kudla, G., A. W. Murray, D. Tollervey, and J. B. Plotkin (2009). "Coding-Sequence Determinants of Gene Expression in Escherichia coli". *Science* **324**, pp. 255–258 (page 16).

Kuersten, S., A. Radek, C. Vogel, and L. O. F. Penalva (2013). "Translation regulation gets its 'omics' moment". *Wiley Interdisciplinary Reviews: RNA* **4**, pp. 617–630 (page 20).

Kunkel, T. a. and K. Bebenek (2000). "DNA Replication Fidelity". *Annual Review of Biochemistry* **69**, pp. 497–529 (page 4).

Kuo, T. C., W. C. Huang, S. C. Wu, and P. L. Cheng (2006). "A case study of inter-arrival time distributions of container ships". *Journal of Marine Science and Technology* **14**, pp. 155–164 (page 40).

Lackner, D. H. and J. Bähler (2008). "Translational Control of Gene Expression: From Transcripts to Transcriptomes". *International Review of Cell and Molecular Biology* **271**, pp. 199–251 (page 20).

Lackner, D. H., T. H. Beilharz, S. Marguerat, J. Mata, S. Watt, F. Schubert, T. Preiss, and J. Bähler (2007). "A Network of Multiple Regulatory Layers Shapes Gene Expression in Fission Yeast". *Molecular Cell* **26**, pp. 145–155 (pages 20, 21).

Ladbury, J. E. and S. T. Arold (2012). "Noise in cellular signaling pathways: Causes and effects". *Trends in Biochemical Sciences* **37**, pp. 173–178 (page 11).

Lanza, A. M., K. A. Curran, L. G. Rey, and H. S. Alper (2014). "A condition-specific codon optimization approach for improved heterologous gene expression in Saccharomyces cerevisiae". *BMC Systems Biology* **8**, p. 33 (pages 31, 32).

Lauria, F., T. Tebaldi, L. Lunelli, P. Struffi, P. Gatto, A. Pugliese, M. Brigotti, L. Montanaro, Y. Ciribilli, A. Inga, A. Quattrone, G. Sanguinetti, and G. Viero (2015). "RiboAbacus: a model trained on polyribosome images predicts ribosome density and translational efficiency from mammalian transcriptomes". *Nucleic Acids Research* **43**, e153 (pages 20, 40).

Leija-Martínez, N., S. Casas-Flores, R. D. Cadena-Nava, J. A. Roca, J. A. Mendez-Cabañas, E. Gomez, and J. Ruiz-Garcia (2014). "The separation between the 5-3 ends in long RNA molecules is short and nearly constant". *Nucleic Acids Research* **42**, pp. 13963–13968 (page 21).

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**, pp. 1540–1544 (page 74).

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**, pp. 900–907 (pages 52, 74–79, 112).

Leventhal, G. E., A. L. Hill, M. A. Nowak, and S. Bonhoeffer (2015). "Evolution and emergence of infectious diseases in theoretical and real-world networks". *Nature Communications* **6**, p. 6101 (page 114).

Leventhal, G. E., R. Kouyos, T. Stadler, V. von Wyl, S. Yerly, J. Böni, C. Cellerai, T. Klimkait, H. F. Günthard, and S. Bonhoeffer (2012). "Inferring Epidemic Contact Structure from Phylogenetic Trees". *PLoS Computational Biology* **8**, e1002413 (page 114).

Levy, S. F., J. R. Blundell, S. Venkataram, D. a. Petrov, D. S. Fisher, and G. Sherlock (2015). "Quantitative evolutionary dynamics using high-resolution lineage tracking". *Nature* **519**, pp. 181–186 (page 114).

Li, L. and R. Bhatia (2011). "Stem Cell Quiescence". *Clinical Cancer Research* **17**, pp. 4936–4941 (page 113).

Lieberman, E., C. Hauert, and M. A. Nowak (2005). "Evolutionary dynamics on graphs". *Nature* **433**, pp. 312–316 (page 112).

Linde, N., G. Fluegen, and J. A. Aguirre-Ghiso (2016). "Chapter Two - The Relationship Between Dormant Cancer Cells and Their Microenvironment". In: *Molecular and Cellular Basis of Metastasis: Road to Therapy*. Vol. 132. Advances in Cancer Research. Academic Press, pp. 45–71 (page 92).

Lipson, D., T. Raz, A. Kieu, D. R. Jones, E. Giladi, E. Thayer, J. F. Thompson, S. Letovsky, P. Milos, and M. Causey (2009). "Quantification of the yeast transcriptome by single-molecule sequencing". *Nature Biotechnology* **27**, pp. 652–658 (page 38).

Lu, P., K. Takai, V. M. Weaver, and Z. Werb (2011). "Extracellular Matrix Degradation and Remodeling in Development and Disease". *Cold Spring Harbor Perspectives in Biology* **3**, a005058 (pages 5, 89).

MacDonald, C. T., J. H. Gibbs, and A. C. Pipkin (1968). "Kinetics of biopolymerization on nucleic acid templates". *Biopolymers* **6**, pp. 1–25 (page 2).

MacKay, V. L., X. Li, M. R. Flory, E. Turcott, G. L. Law, K. A. Serikawa, X. L. Xu, H. Lee, D. R. Goodlett, R. Aebersold, L. P. Zhao, and D. R. Morris (2004). "Gene Expression Analyzed by High-resolution State Array Analysis and Quantitative Proteomics". *Molecular & Cellular Proteomics* **3**, pp. 478–489 (pages 20, 26, 40).

MacLean, A. L., S. Filippi, and M. P. H. Stumpf (2014). "The ecology in the hematopoietic stem cell niche determines the clinical outcome in chronic myeloid leukemia". *Proceedings of the National Academy of Sciences* **111**, pp. 3883–3888 (pages 74, 87).

Magnus, N., N. Gerges, N. Jabado, and J. Rak (2013). "Coagulation-related gene expression profile in glioblastoma is defined by molecular disease subtype". *Journal of Thrombosis and Haemostasis* **11**, pp. 1197–1200 (page 106).

Magnus, N., E. D'Asti, B. Meehan, D. Garnier, and J. Rak (2014a). "Oncogenes and the coagulation system – forces that modulate dormant and aggressive states in cancer". *Thrombosis Research* **133**, S1–S9 (page 106).

Magnus, N., D. Garnier, B. Meehan, S. McGraw, T. H. Lee, M. Caron, G. Bourque, C. Milsom, N. Jabado, J. Trasler, R. Pawlinski, N. Mackman, and J. Rak (2014b). "Tissue fac-

tor expression provokes escape from tumor dormancy and leads to genomic alterations". *Proceedings of the National Academy of Sciences* **111**, pp. 3544–3549 (page 106).

Mahon, F.-X., D. Réa, J. Guilhot, F. Guilhot, F. Huguet, F. Nicolini, L. Legros, A. Charbonnier, A. Guerci, B. Varet, G. Etienne, J. Reiffers, and P. Rousselot (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". *The Lancet Oncology* **11**, pp. 1029–1035 (pages 74, 80).

Maley, C. C. et al. (2017). "Classifying the evolutionary and ecological features of neoplasms". *Nature Reviews Cancer* **17**, pp. 605–619 (page 89).

Maquat, L. E., W.-Y. Tarn, and O. Isken (2010). "The Pioneer Round of Translation: Features and Functions". *Cell* **142**, pp. 368–374 (page 23).

Marais, G. and L. Duret (2001). "Synonymous Codon Usage, Accuracy of Translation, and Gene Length in Caenorhabditis elegans". *Journal of Molecular Evolution* **52**, pp. 275–280 (page 34).

Marciniak-Czochra, A., T. Stiehl, A. D. Ho, W. Jäger, and W. Wagner (2009a). "Modeling of Asymmetric Cell Division in Hematopoietic Stem Cells — Regulation of Self-Renewal Is Essential for Efficient Repopulation". *Stem Cells and Development* **18**, pp. 377–386 (page 52).

Marciniak-Czochra, A., T. Stiehl, and W. Wagner (2009b). "Modeling of replicative senescence in hematopoietic development". *Aging* **1**, pp. 723–732 (page 62).

Marley, S. B. and M. Y. Gordon (2005). "Chronic myeloid leukaemia: stem cell derived but progenitor cell driven". *Clinical Science* **109**, pp. 13–25 (page 87).

Marsaglia, G. (1968). "Random numbers fall mainly in the planes." *Proceedings of the National Academy of Sciences of the United States of America* **61**, pp. 25–28 (page 13).

Marsaglia, G. and W. W. Tsang (2002). "Some Difficult-to-Pass Tests of Randomness". *Journal of Statistical Software* **7**, pp. 1–9 (page 13).

Marshall, E., I. Stansfield, and M. C. Romano (2014). "Ribosome recycling induces optimal translation rate at low ribosomal availability". *Journal of The Royal Society Interface* **11**, p. 20140589 (pages 23, 24, 32).

Marshman, E., C. Booth, and C. S. Potten (2002). "The intestinal epithelial stem cell". *BioEssays* **24**, pp. 91–98 (page 6).

Martincorena, I., K. M. Raine, M. Gerstung, K. J. Dawson, K. Haase, P. Van Loo, H. Davies, M. R. Stratton, and P. J. Campbell (2017). "Universal Patterns of Selection in Cancer and Somatic Tissues". *Cell* **171**, pp. 1029–1041 (pages 17, 112).

Matsumoto, M. and T. Nishimura (1998). "Mersenne twister: a 623-dimensionally equidistributed uniform pseudo-random number generator". *ACM Transactions on Modeling and Computer Simulation* **8**, pp. 3–30 (page 13).

Mazumder, B., V. Seshadri, and P. L. Fox (2003). "Translational control by the 3-UTR: the ends specify the means". *Trends in Biochemical Sciences* **28**, pp. 91–98 (page 21).

Merrick, W. C. and M. E. Harris (2014). "Control not at initiation? Bah, humbug!" *The EMBO Journal* **33**, pp. 3–4 (page 34).

Michor, F., M. A. Nowak, S. A. Frank, and Y. Iwasa (2003a). "Stochastic elimination of cancer cells". *Proceedings of the Royal Society B: Biological Sciences* **270**, pp. 2017–2024 (pages 1, 46, 86).

Michor, F., S. A. Frank, R. M. May, Y. Iwasa, and M. A. Nowak (2003b). "Somatic selection for and against cancer". *Journal of Theoretical Biology* **225**, pp. 377–382 (pages 43, 46, 86, 113).

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**, pp. 1267–1270 (pages 52, 62, 74, 75, 78, 79, 87).

Milholland, B., X. Dong, L. Zhang, X. Hao, Y. Suh, and J. Vijg (2017). "Differences between germline and somatic mutation rates in humans and mice". *Nature Communications* **8**, p. 15183 (page 2).

Mintz, B. and K. Illmensee (1975). "Normal genetically mosaic mice produced from malignant teratocarcinoma cells." *Proceedings of the National Academy of Sciences* **72**, pp. 3585–3589 (page 89).

Miura, F., N. Kawaguchi, M. Yoshida, C. Uematsu, K. Kito, Y. Sakaki, and T. Ito (2008). "Absolute quantification of the budding yeast transcriptome by means of competitive PCR between genomic and complementary DNAs". *BMC Genomics* **9**, p. 574 (page 38).

Mojtahedi, M., A. Skupin, J. Zhou, I. G. Castaño, R. Y. Y. Leong-Quong, H. Chang, K. Trachana, A. Giuliani, and S. Huang (2016). "Cell Fate Decision as High-Dimensional Critical State Transition". *PLOS Biology* **14**, e2000640 (page 5).

Moran, P. A. P. (1958). "Random processes in genetics". *Mathematical Proceedings of the Cambridge Philosophical Society* **54**, p. 60 (pages 2, 9, 12, 76, 86).

Mori, S. et al. (2015). "Age and dPCR can predict relapse in CML patients who discontinued imatinib: The ISAV study". *American Journal of Hematology* **90**, pp. 910–914 (page 80).

Morisaki, T., K. Lyon, K. F. DeLuca, J. G. DeLuca, B. P. English, Z. Zhang, L. D. Lavis, J. B. Grimm, S. Viswanathan, L. L. Looger, T. Lionnet, and T. J. Stasevich (2016). "Real-time quantification of single RNA translation dynamics in living cells". *Science* **352**, pp. 1425–1429 (page 114).

Morrison, S. J. and A. C. Spradling (2008). "Stem Cells and Niches: Mechanisms That Promote Stem Cell Maintenance throughout Life". *Cell* **132**, pp. 598–611 (page 62).

Nagar, A., A. Valleriani, and R. Lipowsky (2011). "Translation by Ribosomes with mRNA Degradation: Exclusion Processes on Aging Tracks". *Journal of Statistical Physics* **145**, pp. 1385–1404 (page 23).

Naumov, G. N., E. Bender, D. Zurakowski, S.-Y. Kang, D. Sampson, E. Flynn, R. S. Watnick, O. Straume, L. A. Akslen, J. Folkman, and N. Almog (2006). "A Model of Human Tumor Dormancy: An Angiogenic Switch From the Nonangiogenic Phenotype". *JNCI: Journal of the National Cancer Institute* **98**, pp. 316–325 (page 106).

Nelson, E. M. and M. M. Winkler (1987). "Regulation of mRNA entry into polysomes. Parameters affecting polysome size and the fraction of mRNA in polysomes." *The Journal of biological chemistry* **262**, pp. 11501–11506 (pages 21, 22, 32).

Nichol, D., P. Jeavons, A. G. Fletcher, R. a. Bonomo, P. K. Maini, J. L. Paul, R. a. Gatenby, A. R. Anderson, and J. G. Scott (2015). "Steering Evolution with Sequential Therapy to Prevent the Emergence of Bacterial Antibiotic Resistance". *PLOS Computational Biology* **11**, e1004493 (page 115).

Niewald, M., C. Berdel, J. Fleckenstein, N. Licht, R. Ketter, and C. Rübe (2011). "Toxicity after radiochemotherapy for glioblastoma using temozolomide - a retrospective evaluation". *Radiation Oncology* **6**, p. 141 (pages 92, 94, 107).

Noble, R., O. Kaltz, and M. E. Hochberg (2015). "Peto's paradox and human cancers". *Philosophical Transactions of the Royal Society B: Biological Sciences* **370**, p. 20150104 (page 46).

Noone, A. M., N. Howlader, M. Krapcho, D. Miller, A. Brest, M. Yu, J. Ruhl, Z. Tatalovich, A. Mariotto, D. R. Lewis, H. S. Chen, E. J. Feuer, and K. A. (Cronin (2018). *SEER Cancer Statistics Review, 1975-2015, National Cancer Institute, Bethesda, MD. Based on November 2017 SEER data submission, posted to the SEER web site.* (Page 86).

Nowak, M. A., F. Michor, and Y. Iwasa (2003). "The linear process of somatic evolution". *Proceedings of the National Academy of Sciences* **100**, pp. 14966–14969 (pages 1, 46, 86, 112).

Nowak, M. A. (2006). *Evolutionary dynamics*. Harvard University Press (pages 7, 92, 96).

Nowell, P. (1976). "The clonal evolution of tumor cell populations". *Science* **194**, pp. 23–28 (page 6).

Nussey, D. H., D. Baird, E. Barrett, W. Boner, J. Fairlie, N. Gemmell, N. Hartmann, T. Horn, M. Haussmann, M. Olsson, C. Turbill, S. Verhulst, S. Zahn, and P. Monaghan (2014). "Measuring telomere length and telomere dynamics in evolutionary biology and ecology". *Methods in Ecology and Evolution* **5**, pp. 299–310 (pages 46, 114).

Odegard, V. H. and D. G. Schatz (2006). "Targeting of somatic hypermutation". *Nature Reviews Immunology* **6**, pp. 573–583 (page 3).

Ohgaki, H. and P. Kleihues (2005). "Epidemiology and etiology of gliomas". *Acta Neuropathologica* **109**, pp. 93–108 (page 92).

Orlando, P. A., R. A. Gatenby, and J. S. Brown (2012). "Cancer treatment as a game: integrating evolutionary game theory into the optimal control of chemotherapy". *Physical Biology* **9**, p. 065007 (page 93).

Ostrow, S. L., R. Barshir, J. DeGregori, E. Yeger-Lotem, and R. Hershberg (2014). "Cancer Evolution Is Associated with Pervasive Positive Selection on Globally Expressed Genes". *PLoS Genetics* **10**, e1004239 (page 17).

Pacheco, J. M., F. C. Santos, and D. Dingli (2014). "The ecology of cancer from an evolutionary game theory perspective". *Interface Focus* **4**, p. 20140019 (page 106).

Page, K. M. (2002). "Unifying Evolutionary Dynamics", pp. 93–98 (page 93).

Page, K. and J. Uhr (2005). "Mathematical models of cancer dormancy". *Leukemia & Lymphoma* **46**, pp. 313–327 (page 107).

Park, E.-H., F. Zhang, J. Warringer, P. Sunnerhagen, and A. G. Hinnebusch (2011). "Depletion of eIF4G from yeast cells narrows the range of translational efficiencies genome-wide". *BMC Genomics* **12**, p. 68 (pages 20, 33).

Passegué, E., A. J. Wagers, S. Giuriato, W. C. Anderson, and I. L. Weissman (2005). "Global analysis of proliferation and cell cycle gene expression in the regulation of hematopoietic stem and progenitor cell fates". *The Journal of Experimental Medicine* **202**, pp. 1599–1611 (page 6).

Pelechano, V., S. Chávez, and J. E. Pérez-Ortín (2010). "A Complete Set of Nascent Transcription Rates for Yeast Genes". *PLoS ONE* **5**, e15442 (pages 24, 32, 38, 42).

Pelechano, V., W. Wei, and L. M. Steinmetz (2015). "Widespread Co-translational RNA Decay Reveals Ribosome Dynamics". *Cell* **161**, pp. 1400–1412 (page 40).

Pellettieri, J. and A. S. Alvarado (2007). "Cell Turnover and Adult Tissue Homeostasis: From Humans to Planarians". *Annual Review of Genetics* **41**, pp. 83–105 (page 1).

Perez-Losada, J. and A. Balmain (2003). “Stem-cell hierarchy in skin cancer”. *Nature Reviews Cancer* **3**, pp. 434–443 (page 113).

Philipps, G. R. (1965). “Haemoglobin Synthesis and Polysomes in Intact Reticulocytes”. *Nature* **205**, pp. 567–570 (page 21).

Plotkin, J. B. and G. Kudla (2011). “Synonymous but not the same: the causes and consequences of codon bias”. *Nature Reviews Genetics* **12**, pp. 32–42 (pages 21, 27, 112).

Polymenis, M. and R. Aramayo (2015). “Translate to divide: ontrol of the cell cycle by protein synthesis”. *Microbial Cell* **2**, pp. 94–104 (pages 15, 32).

Presnyak, V., N. Alhusaini, Y.-H. Chen, S. Martin, N. Morris, N. Kline, S. Olson, D. Weinberg, K. E. Baker, B. R. Graveley, and J. Coller (2015). “Codon Optimality Is a Major Determinant of mRNA Stability”. *Cell* **160**, pp. 1111–1124 (page 29).

Qin, X., S. Ahn, T. P. Speed, and G. M. Rubin (2007). “Global analyses of mRNA translational control during early Drosophila embryogenesis”. *Genome Biology* **8**, R63 (page 20).

Quintás-Cardama, A. and J. Cortes (2009). “Molecular biology of bcr-abl1 – positive chronic myeloid leukemia”. *Blood* **113**, pp. 1619–1630 (pages 74, 77).

Rando, T. A. (2006). “Stem cells, ageing and the quest for immortality”. *Nature* **441**, pp. 1080–1086 (pages 43, 46).

Ranganathan, A. C., A. P. Adam, and J. A. Aguirre-Ghiso (2006). “Opposing Roles of Mitogenic and Stress Signaling Pathways in the Induction of Cancer Dormancy”. *Cell Cycle* **5**, pp. 1799–1807 (page 107).

Rich, A. and D. R. Davies (1956). “A new two stranded helical structure: polyadenylic acid and polyuridylic acid”. *Journal of the American Chemical Society* **78**, pp. 3548–3549 (page 16).

Richter, J. D. and J. Coller (2015). “Pausing on Polyribosomes: Make Way for Elongation in Translational Control”. *Cell* **163**, pp. 292–300 (page 34).

Rodriguez-Brenes, I. A., D. Wodarz, and N. L. Komarova (2013). “Minimizing the risk of cancer: tissue architecture and cellular replication limits”. *Journal of The Royal Society Interface* **10**, p. 20130410 (pages 4, 112).

Rodriguez-Brenes, I. A. and C. S. Peskin (2010). “Quantitative theory of telomere length regulation and cellular senescence.” *Proceedings of the National Academy of Sciences of the United States of America* **107**, pp. 5387–5392 (pages 46, 114).

Rodriguez-Brenes, I. A. and D. Wodarz (2016). “Telomeres open a window on stem cell division”. *eLife* **5**, e12481 (pages 58, 61).

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**, pp. 1181–1184 (pages 74, 78, 87).

Rogers, D. W., M. A. Böttcher, A. Traulsen, and D. Greig (2017). “Ribosome reinitiation can explain length-dependent translation of messenger RNA”. *PLOS Computational Biology* **13** (page 17).

Rosenbloom, D. I. S., A. L. Hill, S. A. Rabi, R. F. Siliciano, and M. A. Nowak (2012). “Antiretroviral dynamics determines HIV evolution and predicts therapy outcome”. *Nature Medicine* **18**, pp. 1378–1385 (page 114).

Ross, D. M., S. Branford, J. F. Seymour, A. P. Schwarer, C. Arthur, D. T. Yeung, P. Dang, J. M. Goyne, C. Slader, R. J. Filshie, A. K. Mills, J. V. Melo, D. L. White, A. P. Grigg, and T. P. Hughes (2013). “Safety and efficacy of imatinib cessation for CML patients

with stable undetectable minimal residual disease: results from the TWISTER study". *Blood* **122**, pp. 515–522 (pages 74, 80).

Rossi, D. J., C. H. M. Jamieson, and I. L. Weissman (2008). "Stems Cells and the Pathways to Aging and Cancer". *Cell* **132**, pp. 681–696 (page 46).

Rousselot, P. et al. (2014). "Loss of Major Molecular Response As a Trigger for Restarting Tyrosine Kinase Inhibitor Therapy in Patients With Chronic-Phase Chronic Myelogenous Leukemia Who Have Stopped Imatinib After Durable Undetectable Disease". *Journal of Clinical Oncology* **32**, pp. 424–430 (pages 74, 80).

Rowley, J. D. (1973). "A New Consistent Chromosomal Abnormality in Chronic Myelogenous Leukaemia identified by Quinacrine Fluorescence and Giemsa Staining". *Nature* **243**, pp. 290–293 (page 73).

Rufer, N., T. 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". *The Journal of Experimental Medicine* **190**, pp. 157–168 (page 78).

Sanjuan, R. (2010). "Mutational fitness effects in RNA and single-stranded DNA viruses: common patterns revealed by site-directed mutagenesis studies". *Philosophical Transactions of the Royal Society B: Biological Sciences* **365**, pp. 1975–1982 (page 4).

Satchi-Fainaro, R., S. Ferber, E. Segal, L. Ma, N. Dixit, A. Ijaz, L. Hlatky, A. Abdollahi, and N. Almog (2012). "Prospective Identification of Glioblastoma Cells Generating Dormant Tumors". *PLoS ONE* **7**, e44395 (page 106).

Sawyers, C. (2004). "Targeted cancer therapy". *Nature* **432**, pp. 294–297 (pages 17, 112).

Schättler, H., U. Ledzewicz, and B. Amini (2016). "Dynamical properties of a minimally parameterized mathematical model for metronomic chemotherapy". *Journal of Mathematical Biology* **72**, pp. 1255–1280 (page 107).

Schwabl, F. (2006). *Statistische Mechanik*. 3rd ed. Springer-Lehrbuch. Berlin/Heidelberg: Springer-Verlag (page 10).

Schwanhäusser, B., D. Busse, N. Li, G. Dittmar, J. Schuchhardt, J. Wolf, W. Chen, and M. Selbach (2011). "Global quantification of mammalian gene expression control". *Nature* **473**, pp. 337–342 (page 41).

Sertil, A. R. (2014). *Hypoxia and Tumor Dormancy: Can the Two Tango?* Vol. 3. Dordrecht: Springer Netherlands, pp. 13–24 (page 92).

Shah, P., Y. Ding, M. Niemczyk, G. Kudla, and J. B. Plotkin (2013). "Rate-Limiting Steps in Yeast Protein Translation". *Cell* **153**, pp. 1589–1601 (pages 20, 21, 25, 27, 31, 34).

Smith, J. M. (1986). "Evolutionary game theory". *Physica D: Nonlinear Phenomena* **22**, pp. 43–49 (page 7).

von Smoluchowski, M. (1906). "Zur kinetischen Theorie der Brownschen Molekularbewegung und der Suspensionen". *Annalen der Physik* **326**, pp. 756–780 (page 11).

Soifer, I. and N. Barkai (2014). "Systematic identification of cell size regulators in budding yeast". *Molecular Systems Biology* **10**, p. 761 (page 32).

Sørensen, M. A., C. Kurland, and S. Pedersen (1989). "Codon usage determines translation rate in Escherichia coli". *Journal of Molecular Biology* **207**, pp. 365–377 (page 16).

Sosa, M. S., F. Parikh, A. G. Maia, Y. Estrada, A. Bosch, P. Bragado, E. Ekpin, A. George, Y. Zheng, H.-M. Lam, C. Morrissey, C.-Y. Chung, E. F. Farias, E. Bernstein, and J. A.

Aguirre-Ghiso (2015). "NR2F1 controls tumour cell dormancy via SOX9- and RARβ-driven quiescence programmes". *Nature Communications* **6**, p. 6170 (page 107).

Sosnoski, D. M., R. J. Norgard, C. D. Grove, S. J. Foster, and A. M. Mastro (2015). "Dormancy and growth of metastatic breast cancer cells in a bone-like microenvironment". *Clinical and Experimental Metastasis* **32**, pp. 335–344 (page 92).

Sottoriva, A., H. Kang, Z. Ma, T. A. Graham, M. P. Salomon, J. Zhao, P. Marjoram, K. Siegmund, M. F. Press, D. Shibata, and C. Curtis (2015). "A big bang model of human colorectal tumor growth". *Nature Genetics* **47**, pp. 209–216 (pages 5, 17).

Strogatz, S. H. (2014). *Nonlinear Dynamics and Chaos* (pages 8, 98, 113).

Stupp, R. et al. (2005). "Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma." *The New England journal of medicine* **352**, pp. 987–996 (page 92).

Sung, P. and H. Klein (2006). "Mechanism of homologous recombination: mediators and helicases take on regulatory functions". *Nature Reviews Molecular Cell Biology* **7**, pp. 739–750 (pages 1, 4).

Supek, F., B. Miñana, J. Valcárcel, T. Gabaldón, and B. Lehner (2014). "Synonymous Mutations Frequently Act as Driver Mutations in Human Cancers". *Cell* **156**, pp. 1324–1335 (pages 16, 111).

Takizawa, H., R. R. Regoes, C. S. Boddupalli, S. Bonhoeffer, and M. G. Manz (2011). "Dynamic variation in cycling of hematopoietic stem cells in steady state and inflammation". *The Journal of Experimental Medicine* **208**, pp. 273–284 (page 86).

Tarrant, D. and T. von der Haar (2014). "Synonymous codons, ribosome speed, and eukaryotic gene expression regulation". *Cellular and Molecular Life Sciences* **71**, pp. 4195–4206 (pages 21, 34).

Taylor, P. D. and L. B. Jonker (1978). "Evolutionary stable strategies and game dynamics". *Mathematical Biosciences* **40**, pp. 145–156 (pages 2, 8).

Thompson, M. K., M. F. Rojas-Duran, P. Gangaramani, and W. V. Gilbert (2016). "The ribosomal protein Asc1/RACK1 is required for efficient translation of short mRNAs". *eLife* **5**, pp. 1–22 (pages 20, 33).

Tomasetti, C. and B. Vogelstein (2015). "Variation in cancer risk among tissues can be explained by the number of stem cell divisions". *Science* **347**, pp. 78–81 (page 46).

Traulsen, A., J. M. Pacheco, and D. Dingli (2010). "Reproductive fitness advantage of BCR–ABL expressing leukemia cells". *Cancer Letters* **294**, pp. 43–48 (pages 62, 74, 78).

Tumbar, T., G. Guasch, V. Greco, C. Blanpain, W. E. Lowry, M. Rendl, and E. Fuchs (2004). "Defining the Epithelial Stem Cell". *Science* **303**, pp. 359–364 (pages 46, 112).

Tyson, J. J., K. C. Chen, and B. Novak (2003). "Sniffers, buzzers, toggles and blinkers: dynamics of regulatory and signaling pathways in the cell". *Current Opinion in Cell Biology* **15**, pp. 221–231 (pages 1, 89).

Ujvari, B., R. A. Gatenby, and F. Thomas (2016). "The evolutionary ecology of transmissible cancers". *Infection, Genetics and Evolution* **39**, pp. 293–303 (page 3).

Veening, J.-W., W. K. Smits, and O. P. Kuipers (2008). "Bistability, Epigenetics, and Bet-Hedging in Bacteria". *Annual Review of Microbiology* **62**, pp. 193–210 (page 111).

Vogel, C., R. de Sousa Abreu, D. Ko, S.-Y. Le, B. A. Shapiro, S. C. Burns, D. Sandhu, D. R. Boutz, E. M. Marcotte, and L. O. Penalva (2010). "Sequence signatures and mRNA concentration can explain two-thirds of protein abundance variation in a human cell line". *Molecular Systems Biology* **6**, pp. 1–9 (page 20).

Vogelstein, B., N. Papadopoulos, V. E. Velculescu, S. Zhou, L. A. Diaz, and K. W. Kinzler (2013). "Cancer Genome Landscapes". *Science* **339**, pp. 1546–1558 (page 6).

Waldman, Y. Y., T. Tuller, T. Shlomi, R. Sharan, and E. Ruppin (2010). "Translation efficiency in humans: tissue specificity, global optimization and differences between developmental stages". *Nucleic Acids Research* **38**, pp. 2964–2974 (page 34).

Waldron, C., R. Jund, and F. Lacroute (1974). "The elongation rate of proteins of different molecular weight classes in yeast." *FEBS letters* **46**, pp. 11–16 (page 41).

Wang, C., B. Han, R. Zhou, and X. Zhuang (2016). "Real-Time Imaging of Translation on Single mRNA Transcripts in Live Cells". *Cell* **165**, pp. 990–1001 (page 33).

Wang, J., L. Xu, E. Wang, and S. Huang (2010). "The Potential Landscape of Genetic Circuits Imposes the Arrow of Time in Stem Cell Differentiation". *Biophysical Journal* **99**, pp. 29–39 (page 5).

Wang, M., M. Weiss, M. Simonovic, G. Haertinger, S. P. Schrimpf, M. O. Hengartner, and C. von Mering (2012). "PaxDb, a Database of Protein Abundance Averages Across All Three Domains of Life". *Molecular & Cellular Proteomics* **11**, pp. 492–500 (page 26).

Wang, T., Y. Cui, J. Jin, J. Guo, G. Wang, X. Yin, Q.-Y. He, and G. Zhang (2013). "Translating mRNAs strongly correlate to proteins in a multivariate manner and their translation ratios are phenotype specific". *Nucleic Acids Research* **41**, pp. 4743–4754 (page 20).

Warner, J. R. (1999). "The economics of ribosome biosynthesis in yeast". *Trends in Biochemical Sciences* **24**, pp. 437–440 (page 32).

Watson, J. D. and F. H. C. Crick (1953). "Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid". *Nature* **171**, pp. 737–738 (page 3).

Weinberg, D. E., P. Shah, S. W. Eichhorn, J. A. Hussmann, J. B. Plotkin, and D. P. Bartel (2016). "Improved Ribosome-Footprint and mRNA Measurements Provide Insights into Dynamics and Regulation of Yeast Translation". *Cell Reports* **14**, pp. 1787–1799 (pages 20, 25).

Weissman, I. L. (2000). "Stem Cells". *Cell* **100**, pp. 157–168 (page 5).

Werner, B., F. Beier, S. Hummel, S. Balabanov, L. Lassay, T. Orlikowsky, D. Dingli, T. H. Brümmendorf, and A. Traulsen (2015). "Reconstructing the in vivo dynamics of hematopoietic stem cells from telomere length distributions". *eLife* **4**, e08687 (pages 46–48, 53–56, 58, 59, 61, 86, 112).

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 (pages 46, 71, 76, 77, 86, 87, 112).

Werner, B., D. Dingli, and A. Traulsen (2013). "A deterministic model for the occurrence and dynamics of multiple mutations in hierarchically organized tissues". *Journal of The Royal Society Interface* **10**, p. 20130349 (pages 46, 86).

Werner, B., R. E. Gallagher, E. M. Paietta, M. R. Litzow, M. S. Tallman, P. H. Wiernik, J. L. Slack, C. L. Willman, Z. Sun, A. Traulsen, and D. Dingli (2014). "Dynamics of Leukemia Stem-like Cell Extinction in Acute Promyelocytic Leukemia". *Cancer Research* **74**, pp. 5386–5396 (page 62).

Werner, B., J. G. Scott, A. Sottoriva, A. R. Anderson, A. Traulsen, and P. M. Altrock (2016). "The Cancer Stem Cell Fraction in Hierarchically Organized Tumors Can Be Estimated Using Mathematical Modeling and Patient-Specific Treatment Trajectories". *Cancer Research* **76**, pp. 1705–1713 (page 83).

Werner, B. and A. Sottoriva (2018). "Variation of mutational burden in healthy human tissues suggests non-random strand segregation and allows measuring somatic mutation rates". *PLOS Computational Biology* **14**, e1006233 (page 112).

Wikman, H., R. Vessela, and K. Pantel (2008). "Cancer micrometastasis and tumour dormancy". *APMIS* **116**, pp. 754–770 (page 92).

Wilkie, G. S., K. S. Dickson, and N. K. Gray (2003). "Regulation of mRNA translation by 5- and 3-UTR-binding factors". *Trends in Biochemical Sciences* **28**, pp. 182–188 (page 21).

Wilkie, K. P. and P. Hahnfeldt (2013). "Tumor-immune dynamics regulated in the microenvironment inform the transient nature of immune-induced tumor dormancy". *Cancer Research* **73**, pp. 3534–3544 (page 107).

Williams, M. J., B. Werner, T. Heide, C. Curtis, C. P. Barnes, A. Sottoriva, and T. A. Graham (2018). "Quantification of subclonal selection in cancer from bulk sequencing data". *Nature Genetics* **50**, pp. 895–903 (page 17).

Williams, M. J., B. Werner, C. P. Barnes, T. A. Graham, and A. Sottoriva (2016). "Identification of neutral tumor evolution across cancer types". *Nature Genetics* **48**, pp. 238–244 (pages 5, 17, 46).

Wloch, D. M., K. Szafraniec, R. H. Borts, and R. Korona (2001). "Direct estimate of the mutation rate and the distribution of fitness effects in the yeast Saccharomyces cerevisiae." *Genetics* **159**, pp. 441–452 (page 4).

Wood, D. (1992). *The computation of polylogarithms*. Tech. rep. Canterbury, UK: University of Kent, Computing Laboratory, pp. 182–196 (page 67).

Wu, B., C. Eliscovich, Y. J. Yoon, and R. H. Singer (2016). "Translation dynamics of single mRNAs in live cells and neurons". *Science* **352**, pp. 1430–1435 (page 114).

Yamamoto, K. N., A. Nakamura, and H. Haeno (2015). "The evolution of tumor metastasis during clonal expansion with alterations in metastasis driver genes". *Scientific Reports* **5**, p. 15886 (page 115).

Yan, X., T. A. Hoek, R. D. Vale, and M. E. Tanenbaum (2016). "Dynamics of Translation of Single mRNA Molecules In Vivo". *Cell* **165**, pp. 976–989 (page 33).

Yang, Z. and J. P. Bielawski (2000). "Statistical methods for detecting molecular adaptation". *Trends in Ecology & Evolution* **15**, pp. 496–503 (pages 17, 112).

Yeh, A. C. and S. Ramaswamy (2015). "Mechanisms of cancer cell dormancy-another hallmark of cancer?" *Cancer Research* **75**, pp. 5014–5022 (page 92).

Yoffe, A. M., P. Prinsen, W. M. Gelbart, and A. Ben-Shaul (2011). "The ends of a large RNA molecule are necessarily close". *Nucleic Acids Research* **39**, pp. 292–299 (page 21).

Yu, C.-H., Y. Dang, Z. Zhou, C. Wu, F. Zhao, M. S. Sachs, and Y. Liu (2015). "Codon Usage Influences the Local Rate of Translation Elongation to Regulate Co-translational Protein Folding". *Molecular Cell* **59**, pp. 744–754 (pages 28, 29).

Zeeman, E. C. (1980). "Population dynamics from game theory". In: *Global theory of dynamical systems, Lecture Notes in Mathematics*. Vol. 819. Berlin, Heidelberg: Springer, pp. 471–497 (page 8).

Zhang, B. et al. (2018). "Bone marrow niche trafficking of miR-126 controls the self-renewal of leukemia stem cells in chronic myelogenous leukemia". *Nature Medicine* **24**, pp. 450–462 (page 86).

Zhang, J., J. J. Cunningham, J. S. Brown, and R. A. Gatenby (2017). "Integrating evolutionary dynamics into treatment of metastatic castrate-resistant prostate cancer". *Nature Communications* **8**, p. 1816 (page 115).

Zhang, X. H.-F., M. Giuliano, M. V. Trivedi, R. Schiff, and C. K. Osborne (2013). "Metastasis Dormancy in Estrogen Receptor-Positive Breast Cancer". *Clinical Cancer Research* **19**, pp. 6389–6397 (page 92).

Zhou, M., J. Guo, J. Cha, M. Chae, S. Chen, J. M. Barral, M. S. Sachs, and Y. Liu (2013). "Non-optimal codon usage affects expression, structure and function of clock protein FRQ". *Nature* **495**, pp. 111–115 (page 111).

Zhou, T., W. Gu, and C. O. Wilke (2010). "Detecting Positive and Purifying Selection at Synonymous Sites in Yeast and Worm". *Molecular Biology and Evolution* **27**, pp. 1912–1922 (pages 17, 112).

Zong, S., Z. Wang, H. Chen, and Y. Cui (2014). "Assessing Telomere Length Using Surface Enhanced Raman Scattering". *Scientific Reports* **4**, p. 6977 (page 114).

Zukewich, J., V. Kurella, M. Doebeli, and C. Hauert (2013). "Consolidating Birth-Death and Death-Birth Processes in Structured Populations". *PLoS ONE* **8**, e54639 (page 106).

List of Figures

Acknowledgements

Many people supported me on my erratic and seemingly random paths (see section 1.3) leading to the completion of this thesis. Here I would like to name a significant subset of them:

First and foremost I thank my supervisor Arne Traulsen for the extraordinary support and care in the time leading to this thesis. He made it possible for me to follow the very diverse research interests throughout my PhD which is invaluable to me.

I also thank all who supported me in the different projects: David Dingli, Kirsten Hattermann, Janka Held-Feindt, Tom Lenaerts, David Rogers, and Benjamin Werner. I am grateful for their patience in all our discussions and Skype meetings as well as for the great feedback I always received from them. Out of these people I want to emphasize David Rogers, who not only sparked my interest in RNA translation, but also guided me in countless discussions with his vast knowledge of the world of biology and evolution and his ability to bring this knowledge to life. Tom Lenaerts gave me the opportunity to visit him in Brussels — I learned a lot from him scientifically and thanks to him and his very welcoming research group, especially Elias Fernández Domingos, Sofia Papadimitriou, and Jelena Grujic, I had an extremely enriching time that I will never forget. Also, I'd like to thank the people in the Evoltheo department in Plön who created a unique and positive atmosphere to work in; in particular the two co-inhabitants of my office, Luka Opašić and Yuriy Pichugin.

I am grateful for the endless conversations about basically anything and sometimes even science with Benedikt Bauer, Pete Czuppon, Laura Hindersin, and Hanna Schenk on the daily train rides and when relaxing, going climbing or running together. They made the past three and half years pass by in no time.

Most importantly, however, I want to thank my partner Jana and my family Meike, Monika, and Manfred. Without their support, in all areas of life, I would not be nowhere close to handing in this thesis.

Curriculum Vitae

2014-2018	PhD candidate at the Max Planck Institute for Evolutionary Biology, Plön, Germany Focus: Theoretical Evolutionary Biology Supervisor: Prof. Dr. Arne Traulsen
2011-2013	M.Sc. of Physics at the Ruprecht-Karls-Universität Heidelberg, Germany Focus: Theoretical Biophysics, Stochastic Dynamics Overall average grade 1.2
2008-2011	B.Sc. of Physics at the Georg-August-Universität Göttingen, Germany Focus: Nonlinear dynamics, biophysics Overall average grade 1.4
2007-2008	Civilian service (Zivildienst)
2007	Abitur at Gelehrtenschule des Johanneums zu Hamburg, Germany