

Cancer Evolution and the Emergence of Resistance to Targeted Cancer Therapy

Inaugural-Dissertation
zur
Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultät
der Universität zu Köln

vorgelegt von

Nina Müller

aus Unna

Köln, 2018

Berichterstatter: Prof. Dr. Johannes Berg
Prof. Dr. Joachim Krug

Tag der mündlichen Prüfung: 25.09.2018

ABSTRACT

Targeted therapy to cancer acts on the molecular abnormalities driving a specific tumor. In clinical use, targeted therapies can lead to an impressive, but ultimately short-lived regression of solid tumors: In virtually all cases, a therapy-resistant tumor arises during targeted therapy, thus limiting its long-time efficacy. From a population dynamics perspective, the failure of targeted mono-therapy is inevitable: a sufficiently large population of tumor cells contains therapy-resistant mutants as part of its standing genetic heterogeneity. Therapy then selects resistant mutants leading to a tumor consisting of resistant cells. As resistance to targeted therapy can be caused by diverse molecular mechanisms, targeting one particular resistance mechanism only leads to the emergence of resistance via another. If targeted therapy is to achieve a long-term tumor remission, it needs to address all resistance mechanisms present in a population of cancer cells.

As a proof of principle, we systematically derive cell lines resistant to combinations of targeted agents from PC9 cells, a well-studied lung cancer cell line. By characterizing the respective resistant lines, we show that several distinct resistance mechanisms exist simultaneously in the cell population prior to treatment. We derive four cell lines driven by different resistance mechanisms. A drug combination targeting all these mechanisms prevents indefinitely the expansion of resistant cells. Our findings explain why, for solid tumors, long-term control of the disease with targeted therapy has proven elusive so far and point to a treatment strategy differing from current clinical practice: Instead of keeping the treatment fixed until a relapse occurs, tumor evolution has to be anticipated by targeting a broad spectrum of possible resistance mechanisms as early as possible. Our iterative protocol offers a generic approach to explore the spectrum for resistance mutations and can be applied to cell lines driven by different molecular mechanisms.

However, large populations also contain non-dividing cells in a drug tolerant state, which a curative treatment would need to eradicate. To study this type of phenotypic heterogeneity, we develop a statistical model to infer growth rates in a population from single-cell lifetime measurements. Our method offers a way

to study the dynamics of drug tolerant cells and, if applied to cancer data, could clarify to what extent the drug tolerant state is heritable.

KURZZUSAMMENFASSUNG

Zielgerichtete Krebstherapie wirkt auf die molekularen Anomalien, die das Wachstum eines bestimmten Tumors antreiben. Klinisch kann zielgerichtete Krebstherapie zu einem beeindruckenden Rückgang von Tumoren führen, welcher aber nur von kurzer Dauer ist: In praktisch jedem Fall tritt während der Therapie Resistenz auf, was den langfristigen Nutzen der Therapie begrenzt. Aus Sicht der Populationsdynamik ist das Versagen von zielgerichteter Krebstherapie mit nur einem Medikament unausweichlich: Eine genügend große Population an Tumorzellen enthält als Teil ihrer genetischen Heterogenität Mutanten, welche auf das Medikament resistent sind. Die Therapie selektiert somit resistente Zellen, was zu einem neuen, resistenten Tumor führt. Weil Resistenz auf zielgerichtete Therapie normalerweise durch verschiedene molekulare Mechanismen auftreten kann, führt eine Therapie, welche auf nur einen bestimmten Resistenzmechanismus zugeschnitten ist, lediglich zum Auftreten von Resistenz über einen anderen Mechanismus. Daher müssten, wenn zielgerichtete Therapien zu einem langfristigen Tumorrückgang führen sollen, alle Resistenzmechanismen, welche in einer Krebszellpopulation existieren, berücksichtigt werden.

Für einen Grundsatzbeweis isolieren wir systematisch Zelllinien, welche resistent auf Kombinationen von zielgerichteten Medikamenten sind. Als Modellsystem nutzen wir PC9 Zellen, eine bekannte Lungenkrebszelllinie. Indem wir die jeweiligen resistenten Linien charakterisieren, zeigen wir, dass bereits vor Beginn der Therapie verschiedene Resistenzmechanismen gleichzeitig in einer Population existieren. Wir leiten vier resistente Linien ab, die alle einen unterschiedlichen Mechanismus aufweisen. Eine Medikamentenkombination, die auf all diese Mechanismen abzielt, verhindert das Auftreten weiterer Resistenzen. Unsere Ergebnisse erklären, warum die langfristige Kontrolle von Tumoren mit zielgerichteter Therapie bislang nicht erfolgreich war und legen eine neue Behandlungsstrategie nahe: Anstatt bei konstanter Therapie auf einen Rückfall zu warten, sollte der Evolution des Tumors zuvorgekommen werden, indem die Therapie so früh wie möglich neben dem Wildtyp auch auf ein breites Spektrum an möglichen Resistenzmechanismen abzielt. Unser iteratives Protokoll bietet eine generelle

Herangehensweise, mit welcher man das Resistenzspektrum verschiedener Zelllinien untersuchen kann.

Dennoch enthalten große Populationen neben resistenten Mutanten auch Zellen, die sich nicht oder nur sehr langsam teilen, aber die Therapie überleben. Diese "schlafenden Zellen" müssten durch eine kurative Therapie ebenfalls beseitigt werden. Um diese Art der phänotypischen Heterogenität zu untersuchen, entwickeln wir ein statistisches Modell, mit dem wir Wachstumsraten aus der Messung von Lebenszeiten einzelner Zellen inferieren können. Unsere Methode kann dazu genutzt werden, die Dynamik der schlafenden Zellen zu charakterisieren und festzulegen, ob dieser Zellzustand vererbt werden kann.

CONTENTS

1	A SHORT INTRODUCTION TO THE BIOLOGY OF CANCER	1
1.1	Tumors in the human body	2
1.1.1	Cancer is a genetic disease	3
1.1.2	The hallmarks of cancer	4
1.1.3	<i>Example:</i> Overactive EGFR signaling	7
1.1.4	Standard treatment options	9
1.2	Targeted therapies: a novel treatment approach	10
1.2.1	Gleevec can cure chronic myelogenous leukemia (CML)	11
1.2.2	The occurrence of resistance limits the long-term efficacy	13
1.3	Physics and theoretical models in cancer	14
1.4	Outline of this thesis	15
2	RESISTANT MUTANTS IN EVOLVING POPULATIONS	17
2.1	Introduction	17
2.1.1	Growing cancer cells in cell culture	17
2.1.2	Handling cancer cell lines	19
2.1.3	Mathematical models of cancer resistance	21
2.1.4	Outline of this chapter	23
2.2	Semi-stochastic birth-death process with logistic growth	24
2.2.1	Growth dynamics of the sensitive cells	24
2.2.2	Stochastic emergence of resistant mutants	25
2.2.3	The number of resistant colonies	33
2.3	Splitting dynamics	39
2.3.1	Competition of two types of cells	39
2.3.2	Splitting under Luria-Delbrück dynamics	41
3	ITERATIVE EXHAUSTION OF RESISTANCE MECHANISMS TO TARGETED THER- APY	45
3.1	Introduction	45
3.1.1	Can targeted therapy be curative?	45
3.1.2	Experimental protocol	49

3.2	Results	52
3.2.1	Different resistance mechanisms pre-exist in a large cell population and respond to different compounds	52
3.2.2	Different resistance mechanisms show distinct disruptions in signaling pathways	53
3.2.3	No further resistance mechanisms arise under a combination of compounds targeting R1-R4	57
3.2.4	Sleeper cells can survive the combination treatment	59
3.3	Discussion	62
4	INFERENCE OF CELLULAR STATES FROM SINGLE-CELL LIFETIMES	67
4.1	Motivation: Persistence as a way to survive treatment	67
4.2	Lifetimes on a hidden Markov tree	68
4.3	The backward problem: Learning the parameters	71
4.3.1	The expectation maximization algorithm	71
4.3.2	Calculating the marginal probabilities using belief propagation	75
4.4	Results and discussion	77
5	CONCLUSIONS AND OUTLOOK	81
	GLOSSARY	87
	BIBLIOGRAPHY	91
	LIST OF FIGURES	99
	ACKNOWLEDGEMENTS	101

1

A SHORT INTRODUCTION TO THE BIOLOGY OF CANCER

The human body consists of around 10^{13} cells [1] that are organized in tissues which make up the different organs. Even though each cell stems from the same fertilized egg and thus (except for gametes) carries the same genetic information, cells specialize and form tissues that exhibit different functions. This is achieved via gene expression patterns that vary from cell type to cell type, meaning that not all genes that code for proteins are transcribed in all cells: Whereas a brain cell and a skin cell of the same organism are genetically nearly identical, very different proteins are made in either cell. Obviously, tissues in different parts of the body have distinct functions, but they all need to fulfill some basic requirements. Most importantly, they have to be stable and conserve their function during the lifetime of the organism.

Usually, the lifetime of a cell is smaller than the lifetime of the whole organism, which implies that all tissues have to be regularly renewed. This renewal happens at very different time scales: Whereas nerve cells last nearly a lifetime without replacement, cells lining the intestine live in a hostile environment and need to be replaced every few days. So in a living organism, cells divide and die all the time. This process needs to be strictly regulated to preserve the organization and thus function of the tissue.

There are three major contributions how well-organized tissue renewal is achieved [2]. First, there exists a *cell memory*. Each cell has stable patterns of gene expression characteristic for its specific type. This pattern has been inherited from ancestral cells and has already been established during embryonic development. This preserves the diversity of cell types in the tissue. Second, cells perform *selective cell-to-cell adhesion*, meaning that cells do not stick equally well to all other cell types. They rather adhere specifically to some types, which prevents chaotic mixture of cells in tissues. Finally, *cell communication* is the last major contribution to organized tissue renewal. Cells are covered with receptors that constantly sense the surroundings for proliferative or anti-proliferative signals: They only reproduce if

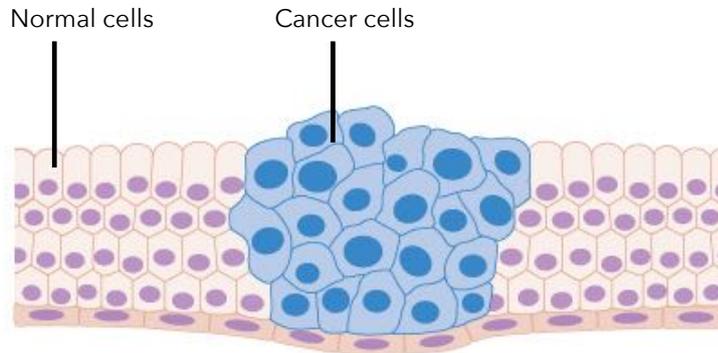


Figure 1: Each tumor is initiated from a single mutated and quickly dividing cell which is no longer subject to the strict controls of tissue formation and renewal. Over the time, a large amount of cancerous cells develops that disturb normal tissue function (sketch adapted from www.cancerresearchuk.org).

they receive the respective signals. They enter apoptosis if they receive signals to die. In this way, new cells are only created when and where they are needed.

If single cells occasionally disrespect the strict rules, the organism does not experience major problems. Only if cells systematically reproduce in places where they should not and the control mechanisms fail at stopping this behavior, first the initial tissue and later the whole organism can be affected. In that case the organism experiences a disease that is called *cancer*.

1.1 TUMORS IN THE HUMAN BODY

Cancer is a term for a broad class of related diseases in which uncontrolled cell growth leads to the emergence of abnormal or *neoplastic* tissue that forms *tumors* (see Fig. 1). Depending on where a tumor grows and how big it is, its mere physical presence is a threat to the organism. In the brain, for instance, tumors repress healthy tissue, which leads to neurological failures. In the lung, tumorous tissue disturbs oxygen intake. In the digestive system, tumors block routes for nutrient intake. In later stages, tumors even acquire the ability to invade into other tissues, a process called *metastasis*. The invasion leads to many tumors that are growing in the body at the same time. If metastatic cancers are left untreated, they usually lead to the death of the organism.

1.1.1 Cancer is a genetic disease

The reason for the abnormal and harmful behavior of cancer cells can be found in their genomes. Cancer cells accumulate thousands of mutations, some of which occur in genes that control cell proliferation. This leads to the transcription of malfunctioning proteins.

There are different ways in which mutations can make cells cancerous, and they often involve genes whose products participate in cellular signaling. A mutation can affect a *cellular receptor* such that it permanently folds into a state which signals a growth signal, even in absence of an external stimulus from a ligand. The corresponding receptor is an example of an *oncogene*, and the mutation is termed a *gain-of-function* mutation. Alternatively, a mutation can interfere with the proper function of a protein involved in the control of cellular proliferation. Such a gene is called a *tumor suppressor* and the mutation is an example of a *loss-of-function* mutation. In both cases it is mutations in specific genes that are key to the emergence of cancer.

Mutations or sets of mutations that make cells cancerous can be acquired in several ways. Sometimes, a predisposition to cancer is inherited: about 5 – 10% of all human cancers are caused by *germline mutations* that are passed from one generation to the next [3]. This explains why in some families there is an accumulation of specific cancer types, such as breast or ovarian cancer. More frequently however, cancer is triggered by *somatic mutations*, mutations that occur during the lifetime of an organism. Often, environmental factors increase the risk of cancer: Carcinogenic substances such as tobacco smoke [4] or UV radiation [5] induce DNA damage leading to cancerous mutations, causing for instance lung or skin cancer. About 12% of all human cancers are due to *oncoviruses* [6], viruses such as the Epstein–Barr (EBV) virus or the human papillomavirus (HPV) that can insert oncogenic genes into infected cells.

In all other cases, cancerous mutations occur just by chance due to unavoidable errors in DNA replication. The body usually has a very efficient machinery to check for errors and malfunctioning cells, but this mechanism does not work perfectly. During the lifetime the human organism experiences approximately 10^{16} cell divisions [1]. Any error detection rate larger than 10^{-16} inevitably leads to the occurrence of mutations, some of which can cause cancer. Most unambiguously, pediatric cancers that occur very early in the life of the patients and that are not

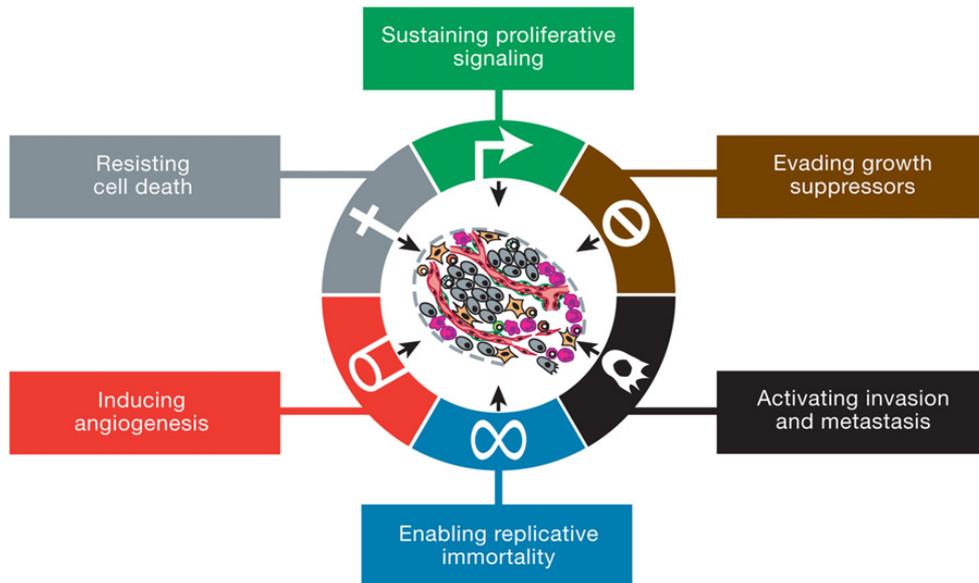


Figure 2: With their hallmarks of cancer Hanahan and Weinberg identified general principles necessary for tumor formation valid for all tumor types. Their analysis made clear that several genetic alterations have to occur before a cell becomes fully malignant (sketch taken from [7]).

driven by any germline mutations are caused by random mutations without any link to environmental influences.

1.1.2 The hallmarks of cancer

Over decades of cancer research, a broad and very detailed knowledge about different cancer types has been accumulated. Tumorigenesis is a multi-step process, where several genetic alterations have to occur in a cell before it becomes fully malignant. In an attempt to simplify and structure the known facts, Hanahan and Weinberg [8] collected general observations that are true for all cancers, independent of their tissue of origin (see Fig. 2). In their original formulation they identified six of such *hallmarks of cancer*, which were complemented by four additional characteristics in a second review article that was published eleven years later [7]. In the following, we briefly summarize these unifying features describing the nature of cancer. A much more detailed discussion can be found in the original publications and the references therein.

SUSTAINING PROLIFERATIVE SIGNALS. Healthy cells need external stimulus to divide, for instance induced by growth factors that bind to cell surface receptors. Cancerous cells do not depend on external signals but permanently proliferate: They are self-sufficient in their growth signals.

EVADING GROWTH SUPPRESSORS. Closely related to the first hallmark is the fact that cancerous cells also ignore anti-proliferative signals: Signaling circuits that stop healthy cells from excessive growth through binding of growth suppressor are modified in cancers such that the negative regulation of cell proliferation is disrupted.

RESISTING CELL DEATH. Usually, there exists a control machinery in cells sensing abnormalities, for example the tumor suppressor TP53. If errors such as DNA damage are detected, the cell is forced to enter a program of controlled death which is called *apoptosis*. Tumors can escape these mechanisms, for instance by the deletion or loss-of-function mutations in important tumor suppressors such as TP53.

INDUCING ANGIOGENESIS. All cells require an adequate supply of oxygen and of nutrients. Tumors that grow excessively cannot be sufficiently provided by the normal blood vessels of the tissue. Usually, *angiogenesis* (meaning blood vessel formation) extending this supply is strictly regulated. But tumors acquire the ability to actively induce angiogenesis to build their own nutrient supply and facilitate fast growth.

ENABLING REPLICATIVE IMMORTALITY. Normal cells cannot divide indefinitely. During each cell division, nucleotide sequences protecting the ends of the chromosomes (the so-called *telomeres*) shorten. In this way, healthy cells age, a process that inhibits excessive growth independently of cell-to-cell signaling. Tumor cells on the other hand can be immortalized by activating mechanisms to maintain their telomeres above the critical length.

ACTIVATING INVASION AND METASTASIS. Healthy cells stay in the tissue where they belong. Cancer cells by contrast have the ability to travel through the blood or lymph system to invade tissues all over the body. This process is called *metastasis*.

Apart from the six classical hallmarks, two further characteristics seem to be particularly important for tumor formation. Among the cellular alterations allowing for excessive growth, also adjustments in the energy metabolism are widely observed in cancer cells. This *metabolic switch* has already been described in 1923 by Otto Warburg [9] (and is for this reason termed *Warburg effect*). Cancer cells gain energy by processing glucose in a way healthy cells only do under anaerobic (and thus stressful) conditions. As the altered metabolism is very inefficient, energy metabolism was long thought to be a weakness of cancer cells. By contrast, more recent findings indicate that the altered metabolism facilitates the biosynthesis of components (such as nucleotides, amino acids, and lipids) that are necessary for cell replication [10] and thus actively contributes to tumorigenesis. The exact mechanism though remains elusive so far.

Another important issue is the interaction between tumors and the immune system. The immune system is able to recognize and destroy many cancer cells even before larger tumors form: Experiments in immunodeficient mice have shown that they develop more tumors than mice with an intact immune system when exposed to carcinogenic substances [11]. Consequently, solid tumors that actually form in hosts with a functioning immune system somehow *evade immune destruction*. The current understanding is that *immunogenic tumors* (tumors that provoke a strong immune response) can be attacked by the immune system and disappear before they are even diagnosed, whereas weakly immunogenic tumors are not detected by the immune system and can continue to grow. So it seems to be a necessary condition to efficiently hide from the immune system to enable tumor formation.

Altogether, many independent steps are involved in the formation of a tumor. Healthy cells are usually well protected from major damage through a powerful genomic maintenance machinery and a surveillance system that detects malfunctioning behavior and forces affected cells to die. Therefore, it is surprising that cells can acquire all of the alterations that are necessary for tumor formation. The reason why it is still possible are two *enabling hallmarks* that simplify the required changes in tumorigenesis.

GENOME INSTABILITY AND MUTATION. Most importantly, tumors are *genomically unstable*. This means that the rate at which point mutations and also major chromosomal rearrangements occur is much higher than in non-cancerous cells [12]. The loss of the tumor suppressor TP53, for instance, enables cells to resist cell death. Consequently, cells with compromised genomes are able to further

proliferate without intervening systems that monitor genomic integrity. Hence, defects in the genome maintenance and repair systems offer selective growth advantages: The more severe the DNA supervision machinery is affected, the easier and faster mutated cells can expand.

TUMOR-PROMOTING INFLAMMATION. Increasing evidence is also pointing to the important role of the tumor micro-environment during tumor formation. Virtual every neoplastic lesion is densely infiltrated with cells of the immune system, indicating an inflammation of the surrounding tissue. Historically, physicians and pathologists thought that this behavior was the attempt of the immune system to fight cancerous cells. Whereas it is indeed true that the immune system can detect tumor cells and trigger an anti-tumor defense, the immune response also contributes to the acquisition of the hallmark capabilities: It has been shown that the acute inflammation supplies bioactive molecules such as growth factors or pro-angiogenic factors to the tumor environment which is exploited by cancerous cells [13]. Moreover, chemicals that are actively mutagenic are released by inflammatory cells [14], which further accelerates the rate at which new mutations are acquired. As such, tumor-promoting inflammation is another enabling hallmark.

1.1.3 *Example: Overactive EGFR signaling*

Whereas all of the different hallmarks of cancer are necessary for tumor formation and survival, they are acquired at different times and via different mechanisms in each tumor. To give a concrete example how cells can become self-sufficient in growth signaling, we briefly discuss the disruption of cell signaling by an overactive growth factor receptor called EGFR.

The *epidermal growth factor receptor* (EGFR or ERBB1) is part of the ERBB family of cell surface receptor tyrosine kinases. The ERBB receptors are trans-membrane molecules that mediate signals between the *epithelium* (the tissue lining the outer surface of organs and blood vessels) and the connective tissue. The receptors are part of the ERBB signaling network [15], which is a layered network: Receptor-ligand binding initiates cellular signaling cascades that produce a physiological outcome. It is structured as follows (see Fig. 3): The *input layer* consists of receptors and their ligands. Binding of a ligand to a receptor monomer induces receptor dimerization and phosphorylation of the receptor tail within the cell. The *signal-*

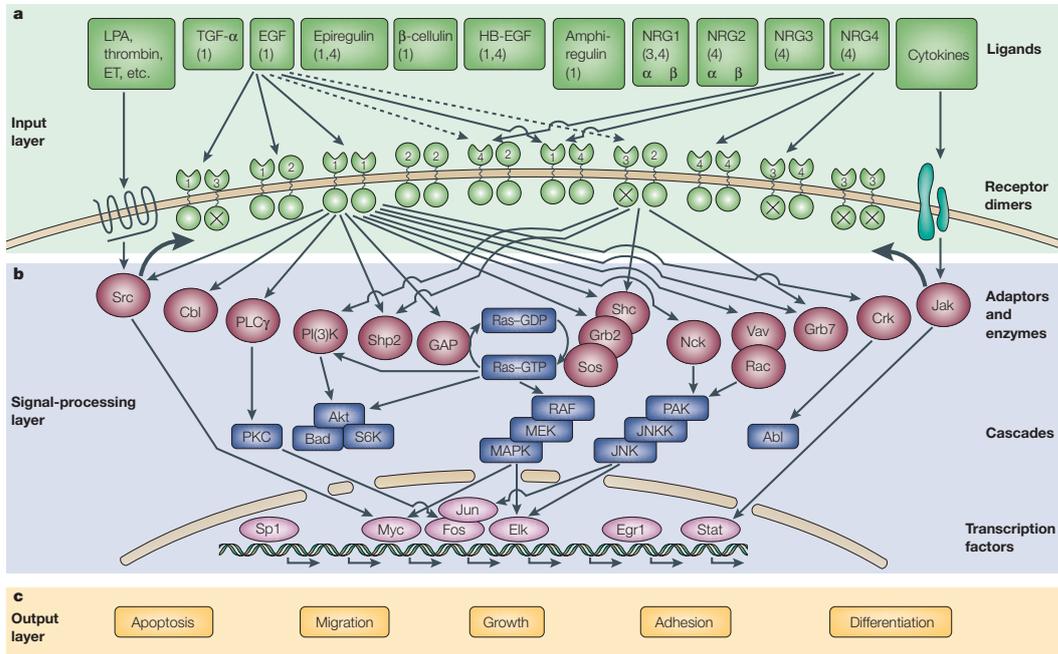


Figure 3: The ERBB signaling network is a layered network transmitting and translating extracellular signals to physiological cellular outcomes. Most important for tumorigenesis are the RAS/RAF/MEK/MAPK and PI3K/AKT cascades, that lead to cell proliferation and migration (figure taken from [15]).

processing layer consists of intracellular adapter molecules and enzymes that bind to the phosphorylated tail of the receptors triggering cellular signaling cascades such as the MAPK or AKT pathways. These signaling cascades translate the extracellular signal to specific transcriptional programs to the cell nucleus. The *output layer* finally produces a physiological outcome. Depending on which ligand was bound to which receptor, programs affecting apoptosis, adhesion or differentiation are started. Also cell division and migration can be triggered, which is important for tumorigenesis in some cancers.

If the receptor EGFR binds to a ligand, it specifically triggers the PI3K and MAPK pathways. Both lead to strong proliferative and anti-apoptotic signals. Consistent with this observation, overexpression and mutations of EGFR have been described in many cancers, mostly in *carcinomas* (tumors that arise from epithelial cells) [16]. Overexpression of EGFR leads to an increased amount of receptors that are made within a cell. As a result, small ligand concentrations can already lead to a strong cellular response [17]. Mutations, on the other hand, alter the structure of the receptor: Large deletions in the extracellular domain of the receptor

are commonly observed [18], leading to ligand-independent kinase signaling – the receptor is stuck in its active state. Both mechanisms lead to constant proliferation and make cells insensitive to other extracellular signals.

1.1.4 Standard treatment options

The current treatment options that are routinely available for solid tumors are surgery, radiotherapy and chemotherapy. Clearly, *surgery* is the oldest and most straightforward way of treating cancer. For the longest time in the history of cancer therapy, cutting out tumorous tissue was actually the only way of treating patients. Its rationale is based on two simple observations. First, small tumors eventually develop into large tumors. Second, large tumors tend to metastasize into other parts of the body, thus exacerbating the tumor burden. In principle, surgery can be a curative treatment if all cancerous cells are removed. This requirement already shows its major limitation: Tumors tend to spread into surrounding tissue, the lymph and the blood system, which is why complete elimination is hard in practice. Nevertheless, still today it is the most effective treatment if tumors are locally constrained to a tissue that is surgically accessible.

The second main pillar of cancer management is *radiotherapy* [19]. Ionizing radiation causes DNA damage and by that leads to cell death. The field of radiation oncology was initiated after the discovery of X-rays by Wilhelm Conrad Röntgen in 1895 [20] and the discovery of radium by Pierre and Marie Curie in 1898 [21]. Both forms of radiation were observed to cause strong skin burn, and first reports on successful medical use in cancer treatment followed around 1900. At the time, the major limitations were severe radiation toxicities (radium) and extensive tissue damage of normal cells (X-rays) due to the limited depth of the low energy beams that were available. The invention of linear accelerators in the middle of the 20th century made it possible to use high energy beams that could be confined to small areas, thus sparing healthy tissue from radiation. Today, advances in imaging and the possibility to control beam shapes facilitate treatment planning and minimize side effects. Radiation therapy is often used in combination with surgery to either shrink the tumor before the operation or decrease the risk of tumor recurrence afterwards. Nevertheless, side effects such as nausea, infertility or even secondary cancers due to radiation damage remain major limitations.

The term *chemotherapy* originally means the treatment of any disease with chemical drugs. In the context of cancer, chemotherapy usually refers to cytotoxic drugs killing cells during cell division. The active development of chemotherapeutic drugs against cancer started after the observation that soldiers who have been exposed to sulfur mustard gas during World War I suffered from a depletion of the bone marrow and lymph nodes [22]. Subsequent experiments in mice with lymphoid tumors showed that the delivery of mustard gas resulted in tumor regression. Since then, different classes of chemotherapeutic drugs have been developed, for instance alkylating agents (such as mustard gas), antimetabolites or cytotoxic antibiotics. Some cancers such as acute childhood leukemia and Hodgkin's disease can be cured with chemotherapy [22]. As compared to surgery or radiation, chemotherapy has the big advantage of not acting on local parts of the body only: Mostly delivered intravenously, it targets both the primary tumor and possible metastases. Today, chemotherapy is often used in combination with surgery or radiation as *adjuvant chemotherapy* to minimize the risk of recurrence from cells surviving the local treatment. Unfortunately, most chemotherapies have severe side effects ranging from immunosuppression, anemia and hair loss to major organ damages. This is why, even if the use of chemotherapeutic agents can be very effective, it often drastically reduces the quality of life [23].

1.2 TARGETED THERAPIES: A NOVEL TREATMENT APPROACH

Driven by more efficient and affordable sequencing methods, many mechanisms enabling cancer cells to proliferate are known today. The development was accompanied by the discovery of *oncogene addiction*, a term that was first introduced by Weinstein [24, 25]. Oncogene addiction means that cancer cells, even though they require several steps before becoming fully malignant, do depend on growth signals transmitted by single oncogenes. Together with the finding that inactivation of those oncogenes in healthy cells is often well tolerated, *molecular drivers* (genetic alterations cancer cells depend on) were identified as possible vulnerabilities [26]. As advances in drug development allowed to design drugs targeting molecular drivers, in the last two decades the paradigm in cancer treatment shifted from cytotoxic chemotherapy to *targeted therapy*. Whereas classical chemotherapy interferes with all quickly dividing cells, targeted therapy kills cancerous cells of a specific molecular type only [27], implying significantly less side effects. Targeting

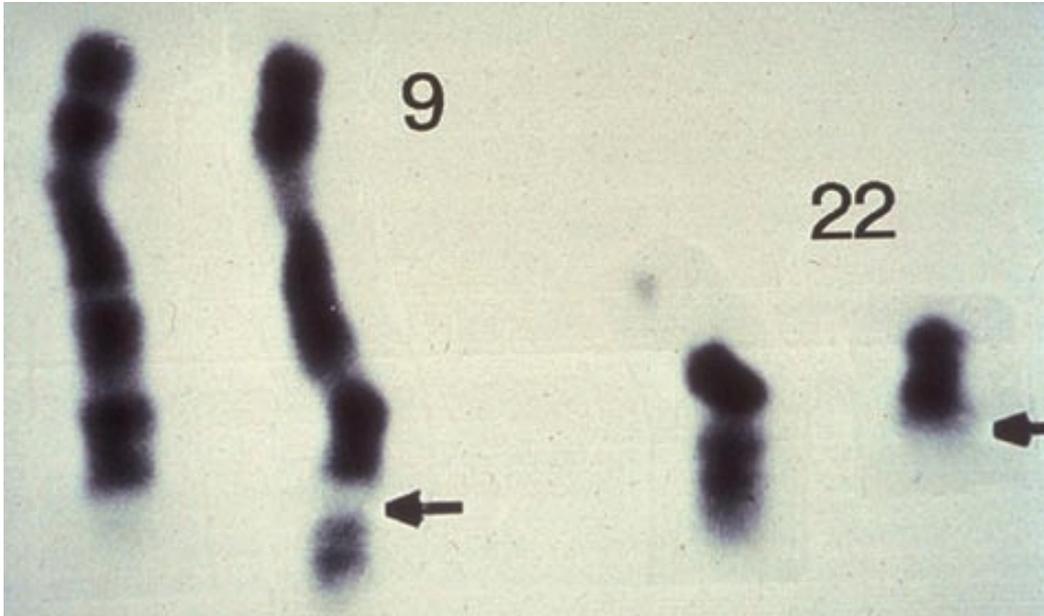


Figure 4: The unusual short Philadelphia chromosome was discovered in CML patients in 1960. The breakpoints of chromosome 9 and 22 are indicated by the black arrows (picture taken from [29]).

cancer at the molecular level also means a step towards *personalized medicine*: As tumors in different patients are not necessarily driven by the same genetic alteration (a fact that is usually referred to as *intertumor heterogeneity*), tumor classification needs to take into account the *genetic subtype* and not only the tissue of origin: For non-small cell lung cancer (NSCLC) for instance, more than 15 genetic subtypes have been identified [28]. Most targeted therapies approved today are either small molecules that can enter the cell membrane or monoclonal antibodies binding to cell surface proteins.

1.2.1 Gleevec can cure chronic myelogenous leukemia (CML)

The story of Gleevec is the biggest success of targeted therapies so far and was its major breakthrough: It transformed CML, which is a specific kind of blood cancer, from a deadly disease to a manageable condition such as diabetes.

Everything started with the discovery of an abnormally short chromosome in CML patients (see Fig. 4) [30], which was later called *Philadelphia chromosome*, after the city where it was discovered. It became clear that the abnormal chromosome

was caused by the fusion of parts of chromosome 9 and chromosome 22 [31]. More specifically, the two genes BCR and ABL which are usually not located next to each other are connected by the fusion [32, 33]. This leads to permanent kinase activity resulting in the accumulation of immature white blood cells [34]. The big hope at this point was that, if it is possible to inhibit the overactive signal, cells harboring the Philadelphia chromosome would die. Over the course of 30 years of research from initial discovery to understanding the function of the BCR-ABL gene fusion, a very clear drug target had thus been identified: The overactive kinase domain in the BCR-ABL fusion gene.

The requirement to develop a kinase inhibitor was a completely new approach in drug discovery. So far, kinase inhibitors were considered to be inefficient due to the large number of different kinases existing in cells: An unspecific kinase inhibitor would disrupt cell function in an uncontrolled way, and was thus expected to cause severe side effects in patients. Fortunately, enough variation in the ATP binding pockets of different kinases was found such that a compound specifically targeting the BCR-ABL kinase could be developed [35]. Imatinib, what the compound was called later, showed to be lethal for cells with the Philadelphia chromosome and completely harmless in all others. Clinically, the relevant observable to quantify the efficacy of a new treatment is the fraction of patients that respond to therapy and experience a significant remission. This fraction is called *response rate*. The first clinical trials investigating the safety and tolerability of Imatinib in CML patients treated with Imatinib followed quickly and showed impressive response rates of close to 100%, with very mild side effects [36]. In 2001, Imatinib was approved under the trade name *Gleevec*, and follow-up studies on patients treated with Gleevec showed an overall survival of 89% after five years [37]. Today, most CML patients have a very good prognosis: After experiencing two years of remission, patients have the same life expectancy as people who never had cancer.

Gleevec was the first targeted drug to be approved and used clinically; its overwhelming success promoted the complete field. Since then, hundreds of small molecules targeting different parts of cellular signaling pathways have been developed and tested in patients. However, in all cases other than Gleevec, even if a subset of patients benefits greatly from specific therapies, the response is typically short-lived, as we will discuss below.

1.2.2 The occurrence of resistance limits the long-term efficacy

One challenge in delivering targeted therapies is the identification of patients that benefit from a specific drug. Actually, CML turned out to be the one example where the molecular mechanism causing continuous growth is always the overactive kinase in the Philadelphia chromosome. Most other cancers however, even if classified as the same disease, are genetically much more diverse. This explains the low response rates in clinical trials on various targeted drugs that followed after Gleevec: In lung cancer for instance, EGFR is often highly expressed. But the treatment with an EGFR kinase inhibitor only led to low response rates of approximately 10%. Only later, mutations in the EGFR kinase domain were found in the subset of patients responding most strongly to therapy, implying that not the high expression of EGFR but the specific activating mutations are driving the tumor [38, 39].

Unfortunately, there is another problem limiting the efficacy of targeted therapies. Even after an initial response (which can be dramatic in some cases, with macroscopic tumors shrinking to the point where they can no longer be detected), patients in almost all cases experience a recurrence of the disease within one year: Tumors that are resistant to therapy grow despite continued treatment. Sequencing of biopsies from the relapsed tumor usually show genetic alterations such as point mutations or gene amplifications as compared to the genotype of the initial tumor which cause resistance [40].

There exist different ways how a tumor can become resistant to targeted therapy. First, the drug target can be modified by a secondary mutation. In this case, the drug can no longer bind and the initial growth signal persists. Second, a mutation can permanently activate a part of the downstream signaling cascade. This means that the drug binds correctly, but has no growth suppressing effect. Finally, a bypass track can be activated. The resistant tumor does not depend on the initial growth signal anymore, which is why it does not respond to therapy. For a review on resistance mechanisms to targeted EGFR inhibition, for instance, see [41].

While the high specificity of targeted therapies accounts for manageable side effects and is thus its biggest advantage, it is also the reason why resistance arises so frequently in practice. Often, a single nucleotide change is enough to confer resistance. Due to unavoidable errors in DNA replication, some genetic heterogeneity is found in large tumors. Cells carrying resistance mutations therefore inevitably exist in very low frequencies prior to any kind of treatment if the population size

is big enough. Targeted therapy then imposes a strong selective pressure favoring resistant mutants which grow while the sensitive majority of cells dies. Resistant cells eventually take over the population and cause a relapse. The type of mutations that confer resistance, the stochastic manner in which they arise, and their statistics is the topic of this thesis.

1.3 PHYSICS AND THEORETICAL MODELS IN CANCER

There are many examples where the application of concepts originally used in the physical sciences contributed valuable insights to different fields of cancer research. For the longest time, physics and medicine have been entangled in the context of radiation and imaging. Medical physicists run and control the accelerators used to generate radiation. A detailed understanding of the specific resonance of atoms with magnetic dipoles has led to the development of magnetic resonance imaging (MRI), which is nowadays routinely used for diagnosis and staging of tumors. Another field is classical biophysics: Cancer cells do not only have disturbed biochemical signaling networks, but also differ in shape, size and elasticity from healthy cells [42]. In this context, physicists study how mechanical features contribute to tumor development. The role of physical interactions in the formation of metastases, for instance, has been investigated extensively [43]. More recently, the accumulation of large data sets obtained by sequencing poses the challenge to interpret the data correctly and extract useful information. For gene expression, the number of measured genes usually exceeds the sample size by far, leading to noisy data and results that are not robust with respect to the incorporation of new data [44]. One possible way to overcome this *curse of dimensionality* is to strive for a coarse grained picture. This separation of scales is a concept deeply rooted in theoretical physics and has been applied to cancer data, for example by including biological knowledge on the level of signaling pathways for improved disease classification and prognosis [45].

A completely different connection between physics and cancer emerges from the *dynamics* of cancer. Starting from initial mutations allowing the establishment of a small set of rapidly dividing cells, a tumor evolves under random mutation, selection for faster growth, and competition for shared resources. As such, tumors are interesting dynamical systems where a large number of cells responds to environmental challenges. Due to the complex interactions (between tumor cells, with

the micro-environment or also with delivered drugs), the emerging macroscopic behavior is impossible to understand by characterizing each single constituent on a molecular level. Different aspects of this dynamics have been investigated: A mathematical model and extensive simulations, for instance, have been used to study how new mutations migrate within the dense, three-dimensional architecture of a solid tumor [46]. Regarding the interplay with the immune system, a fitness model for neoantigens is able to predict if patients respond to immunotherapy [47]. And in the context of drug resistance, an evolutionary model was used to determine from circulating tumor DNA if resistant mutants were present in tumors prior to treatment [48]. There are many other examples, and all of these approaches have complemented pure biomedical research. In the best case, these interdisciplinary efforts can lead to improved treatment outcomes.

1.4 OUTLINE OF THIS THESIS

The focus of this thesis is on the specific mutations which cause resistance to targeted treatment, and on their stochastic emergence in a large population of tumor cells. It involves both experimental work on cell cultures and the analysis of stochastic population genetics models. Key point is that cancer cell populations are genetically heterogeneous and already small subpopulations of a specific genotype can lead to the failure of therapy. Therefore, a detailed understanding of the underlying dynamics is crucial for rational treatment choices. We study tumor heterogeneity from different perspectives:

- In chapter 2, we use classical population dynamics to model the occurrence and growth of resistant mutants in a much bigger population of cells sensitive to a particular treatment. We do not only investigate the growth of individual mutants, but also study the dynamics of colonies or clades – a subset of mutants initiated by the same mutational event. Specifically, we are interested in the probability of resistance as a function of population size, mutation rate and growth rate of the resistant mutants.
- In chapter 3, we introduce an experimental approach based on artificial evolution: For the prospect of cure under targeted therapy, it is indispensable to know in how many ways a specific cancer cell population of given size can develop resistance to therapy over a given timescale. In a well studied

cancer cell line, we iteratively amplify all resistance mechanisms to a specific targeted therapy in a population of given size. We isolate four different resistant lines and find a set of four compounds to be sufficient to suppress all of them. By that, we give a proof of principle that at least in our model system the number of accessible resistance mechanisms is limited. We characterize the different resistant lines and discuss implications for treatment schedules.

- In chapter 4, we turn to phenotypic heterogeneity. It has been shown that subpopulations of different growth rates exist simultaneously in cancer cell populations: Besides quickly proliferating cells, slowly dividing cells have been observed, that, under therapy, neither grow nor die. The exact role of these *sleeper cells* in the development of resistance to therapy is unknown to date. We develop a statistical model to infer growth rates from single cell lifetime measurements. If applied to cancer data, our approach could offer valuable insights in the dynamics of the sleepers.

We close the thesis with a short chapter discussing future research directions.

2

RESISTANT MUTANTS IN EVOLVING POPULATIONS

2.1 INTRODUCTION

2.1.1 Growing cancer cells in cell culture

Cancer cell lines offer useful model systems to investigate tumor evolution under controlled laboratory conditions. A human cancer cell line is derived from a tumor sample of a patient, obtained for instance from a biopsy or after surgery. Simply put, the tumor tissue is cut into pieces, cells are separated with the use of enzymes, transferred to a culture flask and supplied with proper growth medium. If everything goes well, the cells adapt to the new environment and start to grow.

Compared to non-cancerous cells, cancer cells are relatively easy to cultivate. The reason is that healthy cells cannot divide indefinitely: During each cell division, the telomeres – a repetitive nucleotide sequence capping the end of each chromosome – of cells shorten. If their length falls below a critical value (which usually happens after 50-70 divisions), the cells enter programmed cell death. This natural bound is called the *Hayflick limit*, named after the microbiologist Leonard Hayflick, who experimentally showed that normal cells cannot reproduce indefinitely [49]. Conversely, most cancers have mutations in genes that control the enzyme telomerase, which, if activated, adds bases to the end of the



Figure 5: Henrietta Lacks, the woman whose cells were taken to create the first ever immortal cell line HeLa

telomeres and therefore prevents telomere shortening. Indeed, it is one key feature of cancer cells that they are immortal, in the sense that they can divide beyond the Hayflick limit. Non-cancerous cells can also be immortalized, for instance by artificially expressing proteins that prevent degradation of the telomeres. But not surprisingly, the first ever cultured immortal cell line was a cancer cell line called HeLa¹, which was established in 1952.

Over the last 50 years, thousands of cancer cell lines have been established and used for research. Of course, there are important differences between tumors and cell lines. First of all, tumors compete for resources and space with surrounding tissue. So if a tumor grows, it has to also form new blood vessels which connect the inner parts of the tumor to the blood supply. Cell cultures, on the contrary, are well fed with growth medium and can expand freely. Also, a tumor in a living organism interacts with the immune system. The immune system can recognize and fight cancerous cells – so what one observes in cell culture without selection pressure from the immune system might not resemble how the same cells would have evolved in an organism. To name yet another example, tumor heterogeneity might not be fully accounted for by cell cultures. As mentioned above, cell lines originate from small tissue samples. Tumors, however, are spatially heterogeneous, so a sample taken from one end of the tumor might behave differently compared to a sample taken from another end or even a metastasis.

Despite these limitations, cancer cell lines can be seen as molecular models for tumors: Especially in the 1990s, when high throughput techniques in sequencing were developed and sequencing whole exomes or even genomes started to be affordable, cancer cell lines helped to shed light onto genetic mechanisms driving different cancers. As cancer cell lines are easy and cheap to grow, the impact of different drugs could be studied systematically and sensitivities could be mapped to genetic alterations. By that, many driver mutations have been identified and it

¹ The name HeLa stands for Henrietta Lacks, who was an American woman suffering from cervical cancer. She died in 1951 at the age of 31. During a biopsy, tumor tissue was taken from her and George Gey succeeded in creating an immortal cell line from the sample, which he called HeLa. Over the years, HeLa cells became very popular: they were for instance used to find a polio vaccine and were the first human cells ever to be cloned. Still today, HeLa cells are used around the world. Shockingly, the cell line was created without consent or knowledge of Henrietta Lacks and her family. The Lacks family learned by chance about the existence of the cell line roughly 25 years later which raised a huge discussion about privacy and patient rights. The stunning story of Henrietta Lacks and HeLa cells can be found in the book "The immortal life of Henrietta Lacks" by Rebecca Skloot.

became clear that tumors have to be classified not only by their tissue of origin but also by their genotype.

2.1.2 Handling cancer cell lines

To perform a series of experiments on a cell line, one needs a constant supply of cells. To this end, cells need to be cultured over several weeks or months. Of course, one aims at comparing results of experiments that have been conducted at different time points. Therefore it is important to keep the growth conditions as constant as possible. To minimize effects that alter the population, for instance strong selection induced by heavy competition for shared resources, cells are kept permanently in an exponential growth phase, where they have enough space and nutrients to grow. This means that the population needs to be split regularly and provided with new growth medium (one to two times per week, depending on how fast the cells grow).

Cells that are cultured *in vitro* can either grow in suspension, which means that they float in their culture medium, or adherently, which means that viable cells attach to the bottom layer of the culture flask. Which type of growth is observed depends on the specific cell line that is chosen for culture (HeLa cells for instance grow in suspension, whereas PC9 cells, which are used in the experiments done for this thesis, grow adherently). For suspension cells, splitting is easily done by simply removing a part of the growth medium (including cells) and refilling the culture flask with fresh growth medium. For adherent cultures, one first needs to detach the cells, which is usually done with the help of specific enzymes. After that, the cell suspension is mixed, a part is taken out and the remaining population is provided with fresh growth medium. Over night, the cells adhere again to the bottom of the culture flask. The part of the population that has been separated can now be used for experiments. For instance, the response to drugs can be examined with various assays, or DNA of the cells can be sequenced to determine their genotype. One cycle of seeding a culture, exponential expansion and then splitting is called a *passage*.

If one passages cancer cells over long times, the population evolves and genetic mutations occur by random errors in DNA replications. If something vital is hit, the cells grow much slower so that over time, the mutation is removed from the population. Of course, also a beneficial mutation can occur; in this case, the

mutated cells will grow faster than the wild type and over time take over the population. But most mutations tend not to have a huge impact on the fitness of the cells. This means, that over time, a certain degree of genetic variation within the population is created. This is important when it comes to drug sensitivity. Some mutations can – just by chance – confer resistance to a treatment which the parental type is susceptible to. Naturally one assumes that resistance comes with some sort of fitness disadvantage. To give a concrete example why this assumption is reasonable: PC9 cells are driven by an overactive receptor which is stuck in an active state even without external stimulus. The cells are sensitive to a drug that binds to the ATP binding pocket of the receptor and thus silencing it. The best studied resistance mechanism for this cell line is a single nucleotide change that alters the binding pocket a little bit, massively increasing the binding affinity for ATP as compared to the one for the drug. The drug thus no longer binds to that pocket. As the ATP metabolism is disrupted, cells harboring the mutation can still grow, but divide slower. If one has a cell line that is susceptible to some kind of drug, one can think of it as a system in which two types of cells coexist: A sensitive bulk (everything without a resistance mutation) and a resistant minority (everything with some resistance mutation), which usually grows a bit slower than the parental cell type.

We are interested in characterizing the dynamics of resistant mutants in a much bigger sensitive population, so we aim at determining the number of mutant cells. Unfortunately, this number is hard to measure experimentally: One can visually not discriminate between sensitive and resistant cells. Only after treatment it becomes clear which cells were resistant, since only these survive. Moreover cells can only be seen and counted under the microscope. But with a microscope it is hard to visualize the whole growth area. Practically, it is therefore not feasible to monitor the total number of resistant cells in a large population. So it is hard to compare theoretical predictions with experimental data. But there is another interesting value which is directly accessible: The number of resistant colonies.

The cell line we use in the experimental part of this work grows adherently. If a resistance mutation occurs in one cell at some point in time, the cell is spatially fixed. The mutation will only migrate via cell division, and cells carrying the mutation form a colony. As resistant cells typically grow more slowly than the sensitive wild type, most of the colonies will die out over time because they are out-competed by the sensitive cells. The situation changes if one starts to treat the population. Most of the cells, the sensitive majority, dies but the resistant colonies

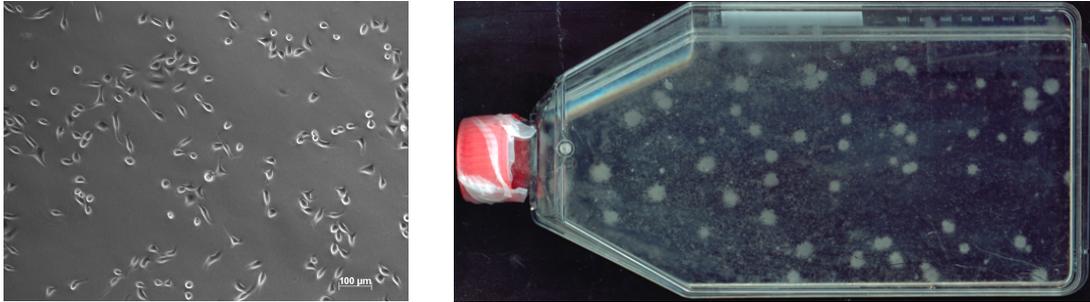


Figure 6: (left) A population of parental PC9 cells as seen under the microscope. As the scale bar implies, each cell has a diameter of approximately 10-20 μm . At this point it is not possible to discriminate between sensitive and resistant cells. (right) After treatment pre-existing resistant cells survive and expand. The resulting colonies can be detected as gray spots by the naked eye. Details of the experiments leading to these pictures are given in chapter 3.

existing at that time survive. The small colonies can expand exponentially because there is no longer competition from surrounding cells. If one waits long enough, cell density will be very high in the vicinity of resistant cells existing prior to treatment and very low everywhere else. As shown in Fig. 6 resistant colonies are visible to the naked eye as gray spots. By counting those spots one can measure the number of resistant colonies existing prior to treatment ². Therefore, investigating the number of colonies instead of the total number of resistant cells has the crucial advantage of being easily observable.

2.1.3 Mathematical models of cancer resistance

Initially, the evolution of resistant mutants in expanding populations has been investigated in the context of bacterial resistance to antibiotics. Salvatore Luria and Max Delbrück performed a classic experiment to answer the question whether bacteria acquired phage-resistant mutations spontaneously at any time during the expansion (H1, random mutation hypothesis) or due to adaptation when being exposed to the phage (H2, adaptive mutation hypothesis) [50]. Their big insight was that the two hypotheses define very different stochastic processes: Whereas H2 describes a simple binomial process, H1 implies a much broader distribution

² If the local density becomes very high cells are likely to detach from the container wall. In this case they spread quickly through the liquid growth medium and initiate new colonies quite distant from their original position – this introduces a new source of noise in the number of colonies. The number of colonies therefore cannot be monitored reliably for long times.

in which the variance exceeds the mean by a factor that grows exponentially with the expansion time. The reason is that at the beginning of the expansion the population size is small and an early mutation is very unlikely to occur. But if it happens, mutant cells expand exponentially – so the number of mutants observed at later times is very high. Therefore the expectation value of the number of mutants is dominated by the rare event of an early mutation.

In their original formulation, Luria and Delbrück considered the growth of sensitive cells and resistant cells deterministically, but treated the mutational process stochastically. The resulting distribution that describes the number of resistant bacteria in an exponentially growing population of sensitive bacteria is today known as the Luria-Delbrück distribution. As their work did not only settle the fundamental question (H_1 is the correct hypothesis), but also offered a means to estimate mutation rates, the model has initiated a long history of research ever since. Important extensions have been made by Lea and Coulson [51] who treated the growth of the resistant mutants stochastically, and by Armitage [52] who derived the exact probability generating function. These and many other contributions are discussed in reviews, for instance in [53].

More recently, the Luria-Delbrück model has been utilized to study the evolution of resistance in cancer. One of the first surveys on the issue was done by Coldman and Goldie [54], who considered a stochastic model of resistant and sensitive mutants where they included cell death. Coldman and Goldie observed that, for a given total population size, a large death rate leads to the accumulation of resistant mutants as compared to growth with zero death rate. The reason is that for large death rates, it takes longer to reach a fixed population size and there is more time for resistant mutants to emerge. Iwasa *et al.* [55] considered the growth of both the sensitive and the resistant cells as a linear birth-death process where sensitive cells could mutate with a fixed mutation rate per cell division. Importantly, the authors found that the *probability of resistance*, meaning the probability of observing at least one resistant mutant in the population, increases with the tumor size at detection and with the mutation rate.

A major limitation of the models mentioned so far is the assumption of exponential population growth. Even if the premise is valid for the early stage of tumor initiation, growth eventually slows down due to competition for shared resources when the tumor exceeds a certain size. For this reason, many models have been developed to study how non-exponential growth influences the mutant dynamics. One interesting extension for the Luria-Delbrück model was given by Dewanji *et*

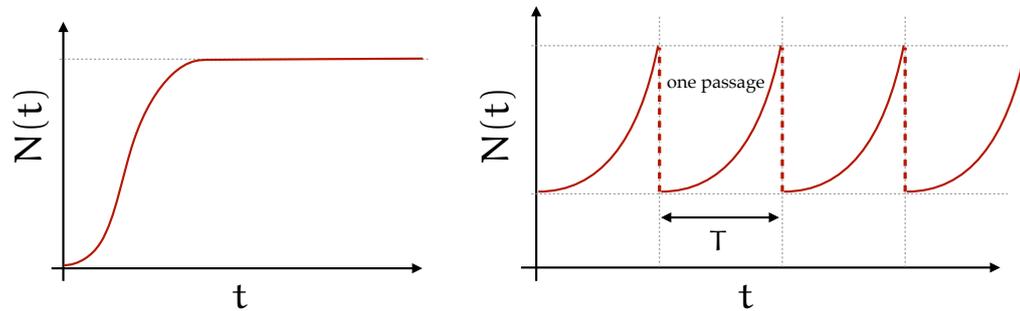


Figure 7: Growth dynamics in two different settings: In a tumor (left), the population cannot expand exponentially indefinitely since the maximum number of cells the environment can host is limited by competition for shared resources. In cell culture however (right), the population is kept in an exponential phase to constantly supply new cells for use in experiments.

al. [56], who considered arbitrary but deterministic growth of the sensitive population. Against this background, they considered the growth of the mutants as a stochastic birth death process where the mean of the number of mutants is either described by an exponential or Gompertzian growth law. Tomasetti [57] showed, in an even more general approach, that the probability of resistance for a given tumor size is independent of what type of growth curves is assumed for the mean size of the tumor. A more complete collection of models concerned with cancer resistance is given in [58].

2.1.4 Outline of this chapter

In the following, we investigate the population dynamics of resistant mutants in a population of sensitive cells. In section 2.2 we introduce competition for shared resources, see Fig. 7, left panel. As a minimal model incorporating competition, we study a population that is confined by a *carrying capacity* (the maximum number of individuals the environment can host) which slows down the growth at large population sizes. We consider two types of cells, sensitive and resistant to a particular treatment. The majority of the population is sensitive and we assume this part to grow according to a deterministic law. Against this background, we consider the stochastic emergence of resistant mutants – a similar approach as Lea and Coulson [51] used to describe the Luria-Delbrück distribution, but with a different growth

law for the number of sensitive cells. In section 2.3, we focus on the situation of culturing cells in vitro. To perform consecutive experiments, cells are cultivated for long times. As shown in Fig. 7 in the right panel, the population experiences multiple cycles of exponential expansion followed by randomly sampling a subset of the population and discarding the rest. We first neglect mutations and study how unfit cell types are lost from the population over time. In a second step, we survey how resistant cells that are introduced by mutations evolve and can get fixed in the population.

2.2 SEMI-STOCHASTIC BIRTH-DEATH PROCESS WITH LOGISTIC GROWTH

In the following, we assume that, without selection pressure, sensitive cells divide at rate b_s and die at rate d_s . Analogously, resistant cells divide at rate b_r and die at rate d_r . At each cell division, there is a small but finite probability μ that a cell mutates and initiates a resistant sub-population. As explained before, these mutations occur just by chance and the mutation rate per cell division is extremely low (approximately $10^{-7} - 10^{-9}$ per nucleotide and cell division). We assume that resistant mutants are less fit than sensitive cells, meaning that resistant mutants typically grow (at least a little) slower than the sensitive cells. These choices imply that mutant cells are far outnumbered by the sensitive population. But once under treatment, only the resistant cells will survive; this is why an understanding of their dynamics plays a key role for tumor evolution under treatment.

2.2.1 Growth dynamics of the sensitive cells

To model growth constrained by competition, we reduce the rate of cell division as the population size increases. We focus on the number of sensitive cells N_s , which are thought to make up the majority of the population. We describe the growth dynamics of the sensitive population by the following deterministic law

$$\frac{dN_s}{dt} = b_s \cdot (1 - \mu) \cdot \left(1 - \frac{N_s}{K}\right) \cdot N_s - d_s \cdot N_s, \quad (1)$$

where N_s denotes the number of sensitive cells, b_s and d_s the birth and death rates per unit time, μ the mutation rate per cell division, and K the carrying capacity. This differential equation (1) can easily be solved giving

$$N_s(t) = \frac{N_{\max}}{1 + \omega \exp(-g_s t)}, \quad (2)$$

where we introduced the following definitions:

$$\begin{aligned} g_s &= b_s(1 - \mu) - d_s, \\ N_{\max} &= K \left(1 - \frac{d_s}{b_s(1 - \mu)} \right), \\ \omega &= \frac{N_{\max}}{N_0} - 1, \quad \text{with } N_0 = N_s(t = 0). \end{aligned}$$

For positive net birth rate $g_s > 0$, there are two growth phases: At the beginning, the influence of the carrying capacity is small and the population size increases nearly exponentially. As soon as the number of cells is in the order of magnitude of the carrying capacity, $N_s(t) \approx K$, growth slows down and the number of cells eventually converges to N_{\max} . The growth function is shown in Fig. 8 for three different values of K .

2.2.2 Stochastic emergence of resistant mutants

We assume that the number of resistant mutants is typically orders of magnitude smaller than the number of sensitive cells. Hence the impact of fluctuations in the number of resistant cells may not be negligible. Against a background of deterministically growing sensitive cells we treat the growth of the mutant population within a stochastic framework. We consider the following events:

- Resistant cells reproduce at rate $b_r \left(1 - \frac{N_r(t)}{K} \right)$ and die at rate d_r .
- New resistant mutants arrive through mutation of sensitive cells at rate $\mu b_s \left(1 - \frac{N_s(t)}{K} \right)$.

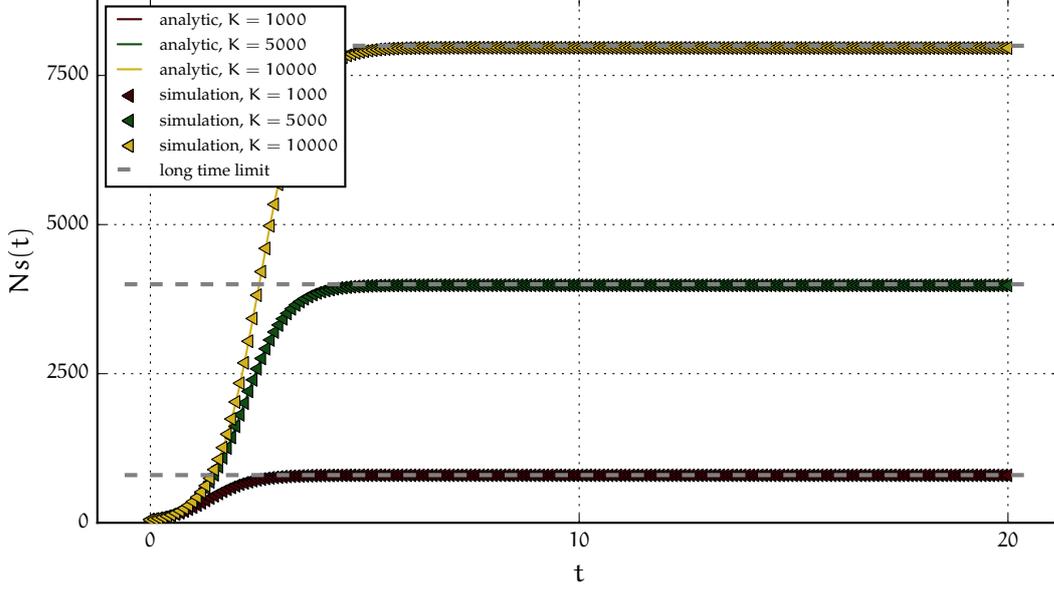


Figure 8: Growth dynamics of the sensitive cells for different carrying capacities. The gray lines indicate the limiting value N_{\max} . As parameter set we chose $b_s = 2.5$, $b_r = 2.0$, $d_s = d_r = 0.5$, $N_0 = 50$ and $\mu = 10^{-3}$; the respective values for the carrying capacities are indicated in the plot.

Under these assumptions, the probability density $p_k \equiv p_k(t)$ of having k mutants at time t can be described by the following master equation

$$\begin{aligned} \frac{dp_k}{dt} = & b_r(t) \cdot (k-1) \cdot p_{k-1} + d_r \cdot (k+1) \cdot p_{k+1} \\ & - (b_r(t) + d_r) \cdot k \cdot p_k + b_s(t) \cdot \mu \cdot N_s(t) \cdot (p_{k-1} - p_k), \end{aligned} \quad (3)$$

with the time-dependent growth rates

$$\begin{aligned} b_r(t) &= b_r \left(1 - \frac{N_s(t)}{K} \right), \\ b_s(t) &= b_s \left(1 - \frac{N_s(t)}{K} \right). \end{aligned}$$

Note that the birth rate of both sensitive and resistant cells is coupled to the number of sensitive cells. In reality, both rates couple to the total number of cells since both cell types compete for resources. But as explained before, we assume that the number of sensitive cells is much larger than the number of resistant cells. Thus, we use $N_s(t) \approx N_{\text{total}}(t)$ to simplify the model.

We can solve the master equation employing generating function techniques. The probability generating function is defined as

$$G(x, t) = \sum_{k=0}^{\infty} x^k p_k(t). \quad (4)$$

Inserting this definition, equation (3) can be written as

$$\begin{aligned} \frac{\partial G(x, t)}{\partial t} &= (b_r(t) \cdot x - d_r) \cdot (x - 1) \cdot \partial_x G(x, t) \\ &+ v(t) \cdot (x - 1) \cdot G(x, t), \end{aligned} \quad (5)$$

where

$$v(t) = \mu \cdot N_s(t) \cdot b_s(t) \quad (6)$$

denotes the time dependent production rate of new mutants. The solution of (5) is subject to the initial condition

$$G(x, 0) = \sum_{k=0}^{\infty} x^k p_k(0) = x^{R_0} \quad (7)$$

with R_0 defined as the number of resistant cells at $t = 0$. We can solve the differential equation (5) with the method of characteristics and obtain

$$G(x, t) = (1 - Q(x, t))^{R_0} \cdot \exp \left(\int_0^t ds v(s) \cdot \frac{Q(x, t)\psi(s)}{Q(x, t)\phi(s) - 1} \right), \quad (8)$$

where we defined the functions

$$\begin{aligned} \psi(t) &= \exp \left(- \int_0^t ds (b_r(s) - d_r) \right), \\ \phi(t) &= \int_0^t ds b_r(s) \psi(s), \\ Q(x, t) &= \frac{1 - x}{\psi(t) + (1 - x)\phi(t)}. \end{aligned}$$

After performing the integration in the definition, we find an easier expression for $\psi(t)$:

$$\begin{aligned}\psi(t) &= \left(\frac{N_0}{N_s(t)} \right)^\beta e^{g_s(\beta-\gamma)t} \\ &= \left(\frac{N_0}{N_{\max}} \right)^\beta (e^{g_s t} + \omega)^\beta e^{-g_s t \cdot \gamma},\end{aligned}\quad (9)$$

where

$$\beta = \frac{b_r}{b_s} \quad \text{and} \quad \gamma = \frac{g_r}{g_s} = \frac{b_r - d_r}{b_s(1-\mu) - d_s}$$

denote the relative birth and growth rate, respectively. Also the integration in the definition of $\phi(t)$ can be carried out which leads to

$$\phi(t) = \frac{b_r N_{\max}}{g_r \omega} \left[\frac{\mathcal{F}(0,1)}{N_0} - \frac{g_r}{\Delta g} \frac{\mathcal{F}(0,2)}{K} + e^{g_s t} \psi(t) \left\{ \frac{g_r}{\Delta g} \frac{\mathcal{F}(t,2)}{K} - \frac{\mathcal{F}(t,1)}{N_s(t)} \right\} \right] \quad (10)$$

with $\Delta g = g_r - g_s$ and

$$\mathcal{F}(t, c) = {}_2F_1(1, 1 + \beta - \gamma, c - \gamma, -e^{g_s t}/\omega),$$

where ${}_2F_1$ is the Gauss' hypergeometric function.

Having derived a closed form solution (8) for the probability generating function we can now determine the moments of the probability distribution in order to characterize the number of resistant cells as a function of time. In the following we assume that we start with no resistant cells at time $t = 0$, which means $R_0 = 0$.

Mean number of resistant cells

From the probability generating function, the expectation value follows as

$$\begin{aligned}R(t) \equiv E(k) &= \sum_{k=0}^{\infty} k \cdot p_k(t) \\ &= \left. \frac{\partial G(x, t)}{\partial x} \right|_{x=1}.\end{aligned}\quad (11)$$

We find

$$\begin{aligned}
 R(t) &= \frac{1}{\psi(t)} \int_0^t ds \nu(s) \psi(s) \\
 &= \frac{b_s \mu N_{\max}}{\omega} \left[\frac{1}{\psi(t)} \left\{ \frac{\mathcal{F}(0,2)}{\Delta g} - \frac{N_0}{K} \frac{\mathcal{F}(0,3)}{\Delta g - g_s} \right\} \right. \\
 &\quad \left. - e^{g_s t} \left\{ \frac{\mathcal{F}(t,2)}{\Delta g} - \frac{N_s(t)}{K} \frac{\mathcal{F}(t,3)}{\Delta g - g_s} \right\} \right]
 \end{aligned} \tag{12}$$

with $\mathcal{F}(c, t)$ as defined above. From equation (12) we can compute the time evolution of the mean,

$$\frac{dR(t)}{dt} = \nu(t) + (b_r(t) - d_r) \cdot R(t). \tag{13}$$

It is easy to see that two different processes contribute to the temporal evolution of the mean: On the one hand, new resistant mutants arrive with rate $\nu(t)$. On the other hand, already existing resistant cells expand with net growth rate $b_r(t) - d_r$. Only the second term can become negative; since the birth rate decreases as the number of sensitive cells approach the carrying capacity, the death rate is eventually greater than the birth rate. This means that two different types of growth are possible depending on how the parameters are chosen: If the net growth rate multiplied with the number of resistant cells exceeds the production rate at some point in time, a maximum can exist. If not, the mean increases monotonically. In the left panel of Fig. 9 the mean number of resistant cells is displayed for different values for different choices of K .

Variance of the number of resistant cells

From the generating function the variance is given as

$$\begin{aligned}
 V(t) \equiv E(k^2) - E(k)^2 &= \sum_{k=0}^{\infty} k^2 \cdot p_k(t) - R(t)^2 \\
 &= \frac{\partial^2 G(x, t)}{\partial x^2} \Big|_{x=1} + \frac{\partial G(x, t)}{\partial x} \Big|_{x=1} - \left(\frac{\partial G(x, t)}{\partial x} \Big|_{x=1} \right)^2, \tag{14}
 \end{aligned}$$

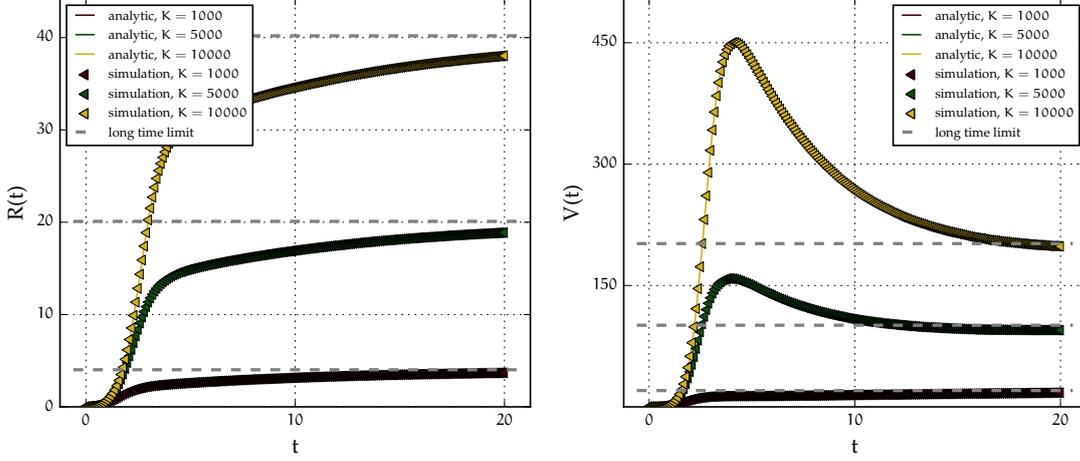


Figure 9: Mean and variance of the number of resistant cells as obtained analytically and by stochastic simulations. We use the same parameter set as defined in Fig. 8.

which leads to

$$V(t) = R(t) \left[1 + \frac{2\phi(t)}{\psi(t)} \right] - \frac{2}{\psi(t)^2} \int_0^t ds \nu(s) \psi(s) \phi(s). \quad (15)$$

As for the mean, we can derive the time evolution of the variance,

$$\frac{dV}{dt} = \nu(t) + (b_r(t) + d_r)R(t) + 2(b_r(t) - d_r)V(t). \quad (16)$$

Again, only the last term in the equation becomes negative when the death rate exceeds the birth rate. In this case, the variance can have a maximum; otherwise it increases monotonically.

Given a sufficiently large carrying capacity, the variance increases rapidly for small times as depicted in the right panel of Fig. 9. This is due to the fact that during the exponential phase the population size is small and an early mutation is very unlikely to occur. But if it happens, the cells can expand freely until the population size reaches the carrying capacity – so the number of mutants is relatively high. This means that if the cells can grow in an exponential way sufficiently long, the variance is dominated for small times by the rare event of an early mutation. These large fluctuations are the same as observed for the Luria-Delbrück distribution. When the number of sensitive cells reaches its steady state value N_{\max} ,

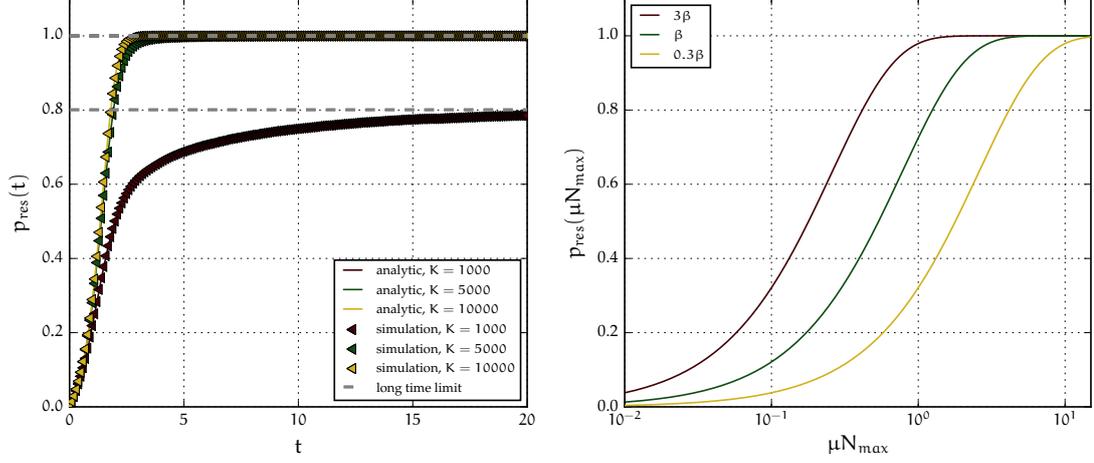


Figure 10: Probability of resistance as a function of time (left) and, in the long-time limit, as a function of population size times mutation rate (right). It becomes clear that the probability of resistance is mainly determined by $\mu \cdot N_{\text{max}}$ – other parameters, as the relative birthrate, only have a minor effect. One can always increase the population size (or the mutation rate) so that in the steady state the probability of finding at least one resistant cell is one. We use the same parameter set as defined in Fig. 8.

the variance decays slowly and eventually saturates. As a result the long-time fluctuations are much smaller compared to a pure exponential growth law.

Probability of resistance

A patient can only be cured by employing a compound the sensitive cells are susceptible to provided not a single cell in the tumor carries any resistance mutations. If, as we assume, cells are either sensitive to treatment or resistant and that this feature is set at cell birth (which means that cells do not acquire resistance during their lifetime), the key variable for the prospect of cure is the probability of resistance, which is to say the probability that at least one single resistant cells exists in the system. We can calculate this probability from the probability generating function as

$$\begin{aligned}
 p_{\text{res}}(t) &= 1 - p_{k=0}(t) = 1 - G(0, t) \\
 &= 1 - \exp\left(-\int_0^t ds \frac{\nu(s)\psi(s)}{\phi(t) - \phi(s) + \psi(t)}\right). \tag{17}
 \end{aligned}$$

The time evolution of p_{res} is depicted in the left panel of Fig. 10. As expected, it increases with time as the population size increases, and, for larger carrying capacities, converges to higher steady-state values.

Long-time limit

For long times, the number of sensitive cells approaches its maximum value

$$\bar{N}_s \equiv \lim_{t \rightarrow \infty} N_s(t) = N_{\text{max}}, \quad (18)$$

which implies that also the birth rates of both cell types converge to

$$\begin{aligned} \bar{b}_s &= \lim_{t \rightarrow \infty} b_s(t) = \frac{d_s}{1 - \mu} \\ \bar{b}_r &= \frac{b_r}{b_s} \bar{b}_s = \beta \cdot \frac{d_s}{1 - \mu} \\ \bar{v} &= \bar{b}_s \cdot \mu \cdot N_{\text{max}}. \end{aligned} \quad (19)$$

A steady-state solution for the number of mutants can only exist if resistant cells do not expand exponentially in the long-time limit. To ensure limited growth we therefore require that the long-time net birth rate is smaller than zero,

$$\bar{b}_r - d_r < 0 \Rightarrow \beta < \delta,$$

where $\delta = \frac{d_r}{d_s}$ denotes the relative death rate. If this requirement is met, we can rewrite the master equation (3) by setting the left hand side to zero and replacing the time dependent rates by their respective steady state values

$$0 = \bar{v}(x - 1)G(x) + (\bar{b}_r x - d_r)(x - 1)\partial_x G(x).$$

This equation is solved by

$$G(x) = \left(\frac{1 - \frac{\beta}{\delta}}{1 - x \cdot \frac{\beta}{\delta}} \right)^{\frac{\bar{v}}{\bar{b}_r}}. \quad (20)$$

From equation (20) we can now derive the mean number of resistant cells as

$$\bar{R} \equiv \lim_{t \rightarrow \infty} \langle R(t) \rangle = \frac{\mu}{1 - \mu} N_{\text{max}} (\delta - \beta)^{-1}. \quad (21)$$

For the variance we find

$$\bar{V} \equiv \lim_{t \rightarrow \infty} V(t) = \bar{R} \cdot \left(1 - \frac{\beta}{\delta}\right)^{-1}. \quad (22)$$

The probability of resistance is thus given by

$$\begin{aligned} p_{\text{res}} = 1 - p_0 &\equiv 1 - G(0) = 1 - \left(1 - \frac{\beta}{\delta}\right)^{\frac{\bar{v}}{b_r}} \\ &= 1 - \left(1 - \frac{\beta}{\delta}\right)^{\frac{\mu N_{\text{max}}}{(1-\mu)\beta}}. \end{aligned} \quad (23)$$

It becomes clear that the steady-state probability of resistance is crucially dependent on the *mutation supply rate*, which is defined as maximum number of sensitive cells N_{max} times mutation rate μ : No matter how small the mutation rate is, or how slowly the mutants grow, if the population size is big enough a mutant cell will exist with probability one. This is shown in Fig. 10 where the steady state probability of resistance is shown as a function of $\mu \cdot N_{\text{max}}$ for different choices of β .

2.2.3 The number of resistant colonies

As outlined in section 2.1.2, the number of resistant colonies is of interest as it can be measured more easily than the number of resistant cells. By *resistant colony* we mean a subpopulation of cells that carry a resistance mutation that is introduced to the population at one specific mutational event. In genetics, this is usually referred to as *clade*. As the PC9 cell line used in the experimental part of this work grows adherently, cells belonging to one clade are attached next to each other at the bottom layer of the culture flask. When the population is treated, only the resistant clades survive and expand – this leads to colony formation where each colony corresponds to one clade.

It is not straight-forward to formulate a model describing the dynamics of colonies. The reason is that the microscopic parameters governing birth and death of colonies need to be defined when switching from a description based on individual cells to a coarse-grained description based on colonies. Whereas the birth rate of the sensitive cells times the mutation rate gives the rate at which new mutations (and thus new colonies) are introduced to the population, describing the

death process of colonies is much harder. The colony death rate depends on how big colonies can grow during the exponential phase, and this depends on when the colony is introduced to the population. This means that colony death is a barrier crossing problem, where death of a colony is achieved if the colony size fluctuates to zero. Hence, we do not know the functional form of the colony death rate. Moreover, all colonies compete against each other for resources. Therefore, individual colonies are in principle not independent. But as we assume that the number of mutants is small compared to the number of sensitive cells, the number of colonies is even smaller. We therefore neglect the interaction between different colonies and treat them as being independent.

In the following, we assume that colonies die at some time-dependent death rate $d_c(t)$ and that new colonies arise at rate $\nu(t)$ as defined in equation (6). We first show that, independent of the functional form of the colony death rate, the number of colonies observed in a population follows a Poisson distribution. We then neglect the time dependence of the colony death rate and use previous results to calculate its steady-state value.

Consider a large number of sensitive cells $N_s(t)$, growing according to (1). Since the number of colonies is small, we need to treat it stochastically. The master equation for the probability $p_M(t)$ of having M colonies at time t is given as

$$\partial_t p_M(t) = \nu(t) \cdot p_{M-1}(t) + d_c(t)(M+1) \cdot p_{M+1}(t) - [\nu(t) + d_c(t)M] \cdot p_M(t), \quad (24)$$

with initial condition $p_M(0) = \delta_{M0}$.

Full distribution

Again, we employ the technique of generating function to solve the master equation. By introducing $G(x, t) = \sum_{k=0}^{\infty} p_k(t)x^k$, equation (24) transforms to

$$\partial_t G(x, t) = \nu(t)(x-1)G(x, t) - d_c(t)(x-1)\partial_x G(x, t). \quad (25)$$

The solution is given by

$$\begin{aligned} G(x, t) &= \exp(\eta(t)(x - 1)) \\ &= \exp(-\eta(t)) \cdot \sum_{k=0}^{\infty} \frac{[x \cdot \eta(t)]^k}{k!}, \end{aligned} \quad (26)$$

where we defined

$$\begin{aligned} \eta(t) &= \exp(-\Delta(t)) \cdot \int_0^t ds \nu(s) \exp(\Delta(s)) \\ \Delta(t) &= \int_0^t ds d_c(s) \end{aligned}$$

and plugged in the series representation of the exponential function in the last step. One can read of that the number of colonies is Poisson distributed

$$p_M(t) = \exp(-\eta(t)) \frac{\eta(t)^M}{M!}. \quad (27)$$

This directly implies

$$\langle M(t) \rangle = \text{Var}(M(t)) = \eta(t). \quad (28)$$

In Fig. 11 we compare the histogram of the number of colonies calculated from simulated data to the values expected from a Poisson distribution. We set the parameter of the distribution to the mean of the respective data set. One can see that the the simulated data are very well described by a Poisson distribution for all times and choices of parameters shown here. The difficulty is that we do not know the functional form of $d_c(t)$, which is why we cannot evaluate the parameter $\eta(t)$ of the distribution explicitly.

Mean number of resistant colonies

We can derive the mean number of resistant colonies from (26) by

$$C(t) \equiv \sum_{M=0}^{\infty} M \cdot p_M(t) = \left. \frac{\partial G(x, t)}{\partial x} \right|_{x=1} = \eta(t), \quad (29)$$

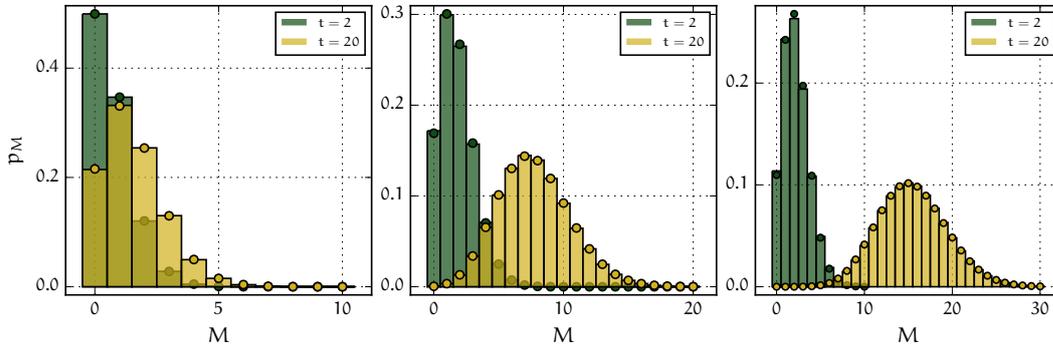


Figure 11: Distribution of the number of resistant colonies as observed in numerical simulations. As a parameter set we chose, as before, $b_s = 2.5$, $b_r = 2.0$, $d_s = d_r = 0.5$, $N_0 = 50$ and $\mu = 10^{-3}$. We compare different carrying capacities: $K = 10^3$ (left), $K = 5 \cdot 10^3$ (middle) and $K = 10^4$ (right). Data are shown from two different time points. The histogram is calculated from the data whereas the dots denote the values belonging to a Poisson distribution in which we set the parameter to the mean of the respective observations. One can see that for all depicted conditions the number of colonies follows an Poisson distribution.

from which we can directly deduce the time evolution of the mean:

$$\frac{dC(t)}{dt} = \nu(t) - d_c(t) \cdot C(t). \quad (30)$$

As outlined before, it is not clear how to write down a functional form of the colony death rate $d_c(t)$. At the beginning, the population size is small and only few colonies are produced due to mutations. Those early colonies can expand freely since there is not much competition. On the contrary, if the population size is close to the carrying capacity more colonies are produced per unit time. These late mutants expand more slowly due to competition and are therefore more prone to extinction. On the other hand, also the production rate $\nu(t)$ of new resistant colonies varies over time. But it is proportional to $N_s(t)$ which is why it changes much faster than $d_c(t)$. It is therefore plausible to neglect the time dependence of

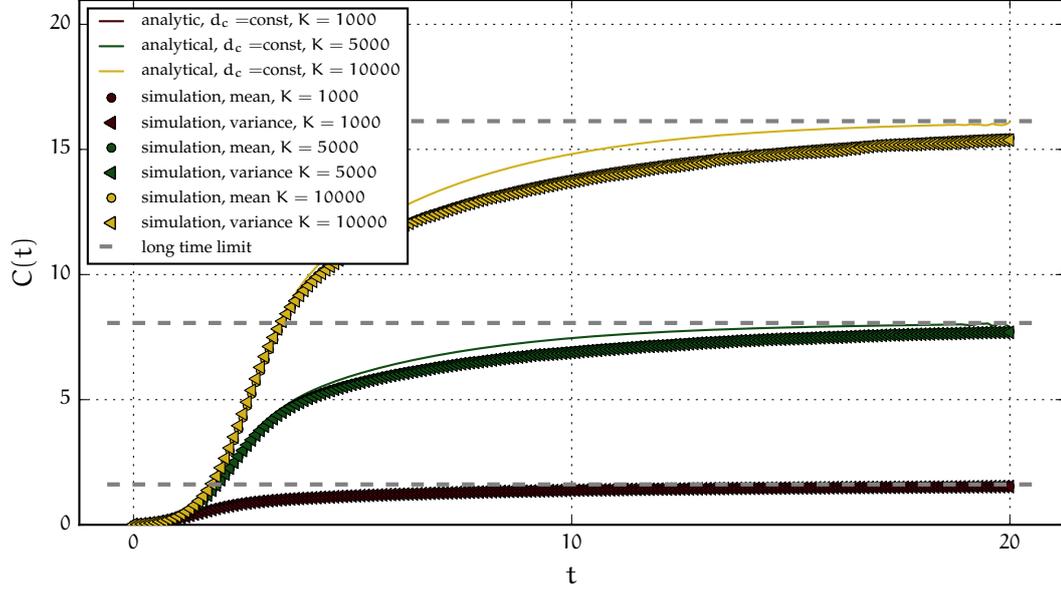


Figure 12: Mean number and variance of resistant colonies as obtained by equation (31) and by numerical simulations. We use the same parameter set as defined in Fig. 11.

the death rate and approximate it by its steady state value $d_c(t) = \bar{d}_c$. Under this assumption we can write the mean number of colonies as

$$\begin{aligned}
 C(t) &= \left[\int_0^t ds \nu(s) \exp(\bar{d}_c \cdot s) \right] \cdot \exp(-\bar{d}_c \cdot t) \\
 &= \frac{\mu}{1-\mu} \left[N_s(t) - e^{-\bar{d}_c \cdot t} \left(N_0 + \frac{N_{\max}}{\omega} \frac{\bar{d}_c - d_s}{\bar{d}_c + g_s} \left[\mathcal{G}(0) - \mathcal{G}(t) e^{(g_s + \bar{d}_c) \cdot t} \right] \right) \right].
 \end{aligned} \tag{31}$$

Here we defined

$$\mathcal{G}(t) = {}_2F_1\left(1, 1 + \frac{\bar{d}_c}{g_s}, 2 + \frac{\bar{d}_c}{g_s}, -\frac{\exp(g_s t)}{\omega}\right),$$

where ${}_2F_1$ is the Gauss' hypergeometric function.

In Fig. 12, we show the mean number of colonies as a function of time for the parameter set defined above and for three different values K . Although we neglect the time dependence of the colony death rate, the analytical solution reflects the temporal evolution reasonably well. We also show the variance of the simulated

data which, as expected from a Poisson distributed random variable, equals the mean.

Long time limit.

We can find an analytical expression for the number of colonies in the long time limit. To calculate the stationary solution of (24) we consider the stationary limits of all quantities:

$$\begin{aligned} p_M(t) &\rightarrow p_M \\ v(t) &\rightarrow \bar{v} = \mu \cdot N_{\max} \cdot \bar{b}_s \\ d_c(t) &\rightarrow \bar{d}_c \\ G(x, t) &\rightarrow G(x). \end{aligned}$$

For this case we find

$$p_M = p_0 \frac{(\bar{v}/\bar{d}_c)^M}{M!}. \quad (32)$$

The probability of having zero colonies follows from normalization as

$$p_0 = \exp\left(-\frac{\bar{v}}{\bar{d}_c}\right). \quad (33)$$

Obviously, the probability of having zero resistant colonies equals the probability of having zero resistant cells – a quantity we have calculated above:

$$p_0 = 1 - p_{\text{res}}.$$

We can use this result to determine the steady-state colony death rate

$$\bar{d}_c = -\frac{\bar{v}}{\log(1 - p_{\text{res}})} = -d_s \beta \log\left(1 - \frac{\beta}{\delta}\right)^{-1}, \quad (34)$$

and from this give an analytical expression for the mean number of colonies in the steady state

$$\bar{C} = \sum_k k \cdot p_k = \frac{\bar{v}}{\bar{d}_c} = -\frac{\mu}{1 - \mu} \frac{N_{\max}}{\beta} \log\left(1 - \frac{\beta}{\delta}\right). \quad (35)$$

Interestingly, both the relative birth rate and death rate occur in this expression. This means that it is not enough to know the net growth rates of both cell types to characterize the number of resistant colonies.

2.3 SPLITTING DYNAMICS

The model investigated so far describes a population of cells whose net growth stops at some point due to competition for shared resources. In cell culture however, populations are kept over long times in an exponential growth phase. This is achieved by regularly extracting a (random) sample of cells and discarding the rest. In the following we study how this splitting dynamics affects the evolution of resistance. We first neglect mutations and study how unfit cell types are lost from the population over time. In a second step, we survey how resistant cells that are introduced by mutations evolve and can get fixed in the population.

2.3.1 Competition of two types of cells

We investigate how two type of cells growing at different growth rates evolve under a dynamics that models the splitting procedure. Consider the following process: We start with a certain number of cells of each type, $s \cdot N_0$ for type 1 and $(1 - s) \cdot N_0$ for type 2, and the population is allowed to expand for a fixed time T . At time T , N_0 cells are randomly drawn from the population to be further cultivated, the rest is discarded. Note that this splitting does not take into account the fitness of the respective type. The two steps are iterated and we are interested in the question how long it takes until one type is lost from the population. Since, for the moment, we do not introduce new mutations, one of the two types is always fixed for long times.

As a minimal model we choose a deterministic picture: Let $N_1(m)$ and $N_2(m)$ denote the number of cells of the respective type after the m^{th} exponential expansion. The cells grow with rates g_1 and g_2 , respectively. Given that at the beginning

of the expansion there exist $N_1^{(0)}(m)$ and $N_2^{(0)}(m)$ cells, the cell numbers after the expansion are given by

$$\begin{aligned} N_1(m) &= N_1^{(0)}(m) \exp(g_1 T) \\ N_2(m) &= N_2^{(0)}(m) \exp(g_2 T). \end{aligned}$$

The probability of sampling type 1 follows as

$$p(m) = \frac{N_1(m)}{N_1(m) + N_2(m)} \quad (36)$$

As binomial sampling leaves the expected fraction of type one cells unchanged, the initial number is connected to the sampling probability after the $(m-1)^{\text{st}}$ expansion as

$$\begin{aligned} N_1^{(0)}(m) &= N_0 p(m-1) \\ N_2^{(0)}(m) &= N_0 [1 - p(m-1)]. \end{aligned}$$

With this, we find a recursive relation for the sampling probability,

$$p(m) = \frac{1}{1 + \frac{1-p(m-1)}{p(m-1)} \exp(-\Delta g T)}, \quad (37)$$

which has to be solved subject to the initial condition

$$p(0) = s.$$

The solution is given by

$$p(m) = \frac{1}{1 + \frac{1-s}{s} \exp(-m \Delta g T)}, \quad (38)$$

as can be easily shown by induction:

As a base case, we consider $m = 0$. From equation (38), the correct initial condition $p(0) = s$ follows directly. So if $p(m)$ holds, we need to show that

$$p(m+1) = \frac{1}{1 + \frac{1-s}{s} \exp(-(m+1) \Delta g T)}.$$

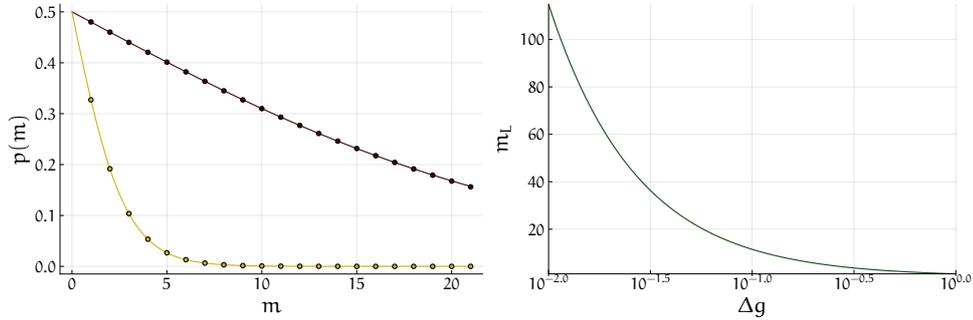


Figure 13: (left) Mean fraction of type 1 cells under splitting dynamics with two competing types for $s = 0.5$, $T = 4$, $\Delta g = -0.02$ (red) and $\Delta g = -0.18$ (yellow). The dots indicate numerical simulation and the lines are given by (38). (right) Time (in number of passages) until unfit type is lost from the population, where *lost* means the passage m_L where the a fraction of the fitter type makes up 0.99% of the population.

Starting from (37), we find

$$\begin{aligned}
 p(m+1) &= \left[1 + \frac{1-p(m)}{p(m)} \exp(-\Delta g T) \right]^{-1} \\
 &= \left[1 + \frac{1 - \frac{1}{1 + \frac{1-s}{s} \exp(-m \Delta g T)}}{\frac{1}{1 + \frac{1-s}{s} \exp(-m \Delta g T)}} \exp(-\Delta g T) \right]^{-1} \\
 &= \frac{1}{1 + \frac{1-s}{s} \exp(-(m+1) \Delta g T)},
 \end{aligned}$$

where, in the second step, we used the induction assumption.

For large m , the probability $p(m)$ will, depending on the sign of Δg , always converge to either 0 or 1. This is to be expected: As we did not include mutations, cells cannot change their type and the faster growing type will take over the population. The time (in number of passages) until fixation depends on the fitness difference, and occurs faster the greater the fitness difference is (see Fig. 13).

2.3.2 Splitting under Luria-Delbrück dynamics

Next we consider mutations. Apart from dividing, sensitive cells can mutate at rate μ per generation. The dynamics of an exponentially expanding population in which sensitive individuals can mutate to become resistant has originally been investigated by Luria and Delbrück who considered growth of both cell types to

be deterministic; on top of the expansion process, mutations occur stochastically. We adapt a more general formulation, originally given by Lea and Coulson [51], who consider the growth of the mutant cells as a stochastic birth process. Let $p_R \equiv p_R(t)$ denote the probability of having R mutants at time t . We assume that sensitive cells grow at rate g_S , resistant cells grow at rate g_R and sensitive cells mutate to resistant cells at rate $\mu \cdot g_S$. Then the dynamics of the mutant cells is described by the following master equation:

$$\partial_t p_R = g_R[(R-1)p_{R-1} - R \cdot p_R] + \mu \cdot S(t) \cdot [p_{R-1} - p_R], \quad (39)$$

where $S(t) = N_0 \exp(g_1 t)$. As we are interested in the expectation value of the resistant mutants, we multiply both sides of the equation with R and sum over all R . We find

$$\partial_t \langle R \rangle = g_2 \langle R \rangle + \mu S(t), \quad (40)$$

where $\langle R \rangle \equiv \sum_{R=0}^{\infty} R \cdot p_R(t)$. This differential equation can easily be solved subject to the initial condition $p_R(t=0) = \delta_{R,0}$ and we find

$$\langle R \rangle = N_0 \frac{\mu}{\Delta g} [1 - \exp(-\Delta g t)] \cdot \exp(g_S t), \quad (41)$$

where $\Delta g = g_S - g_R$.

As a minimal model describing the splitting dynamics, we again work on the level of expectation values. For every expansion of length T , we have to take into account two contributions: First, already existing resistant mutants will expand. Second, sensitive cells introduce new resistant cells through mutation. Using equation (41), the mean number of resistant mutants after the m^{th} expansion, which we call $R_T(m)$, is given by

$$\begin{aligned} R_T(m) &= \underbrace{R_0(m) \cdot \exp(g_R T)}_{\text{expansion of existing mutants}} + \underbrace{(N_0 - R_0(m)) \cdot \chi \cdot \exp(g_S T)}_{\text{introduction of new mutants}} \\ &= R_0(m) [\exp(g_2 T) - \chi \exp(g_1 T)] + R_T(1) \\ &= N_0 \cdot p^{\text{LD}}(m-1) [\exp(g_2 T) - \chi \exp(g_1 T)] + R_T(1), \end{aligned} \quad (42)$$

where we introduced $\chi \equiv \chi(\Delta g, \mu) = \frac{\mu}{\Delta g} [1 - \exp(-\Delta g T)]$. Furthermore, $R_0(m)$ denotes the expectation value of resistant cells at the begin of the m^{th} expansion and

$p^{\text{LD}}(m)$ is the mean sampling probability after the m^{th} expansion. Analogously, the expected number of sensitive cells is given by

$$\begin{aligned} S(m) &= (N_0 - R_0(m)) \exp(g_s T) \\ &= N_0(1 - p^{\text{LD}}(m-1)) \exp(g_s T). \end{aligned}$$

The two equations define a recursive relation for the sampling probability,

$$\begin{aligned} p^{\text{LD}}(m) &= \frac{R_T(m)}{R_T(m) + S(m)} \\ &= \frac{p^{\text{LD}}(m-1)[y-x] + x}{p^{\text{LD}}(m-1)[y-x-1] + 1+x}, \end{aligned} \quad (43)$$

where we used $y = \exp(-\Delta g T)$ as a shorthand. Equation (43) is solved by

$$p^{\text{LD}}(m) = \frac{1 - y^m}{\frac{\Delta g}{\mu} + 1 - y^{m-1}}, \quad (44)$$

which again can be easily shown by induction. As we assume that resistant cells grow slower than sensitive cells, $\Delta g > 0$ and thus $y < 1$. So, for large m , equation (44) converges to

$$\bar{p}^{\text{LD}} \equiv \lim_{m \rightarrow \infty} p^{\text{LD}}(m) = \left(\frac{\Delta g}{\mu} + 1 \right)^{-1}, \quad (45)$$

which yields, for finite μ , the limits

$$\bar{p}^{\text{LD}} = \begin{cases} 1 & \text{for } \Delta g \rightarrow 0 \\ 0 & \text{for } \Delta g \rightarrow \infty. \end{cases} \quad (46)$$

The two limits describe extreme situations: If both cell types grow at the same rate, in the long-time limit the mutants will always take over the population. This is due to the fact that we did not include back mutations – resistant cells cannot produce sensitive cells. So if the fitness difference is small, resistant mutants will accumulate. On the other hand, if the fitness difference is large, resistant mutants, even if they are introduced to the population, tend to be lost from the population due to their fitness disadvantage. Hence mutation and selection can balance so that resistance mutants occur at some fixed sampling probability in the steady

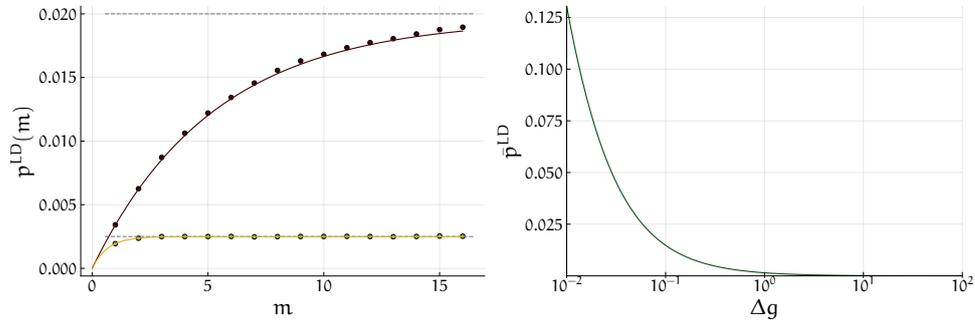


Figure 14: (left) Mean fraction of resistant cells under LD splitting dynamics with two competing types for $s = 0.5$, $T = 2.5$, $\mu = 0.0015$, $\Delta g = 0.075$ (red) and $\Delta g = 0.6$ (yellow). The dots indicate numerical simulation and the lines are given by (44). (right) Steady state fraction of resistant cells as a function of growth rate as given by equation (45).

state. This is also shown in Fig. 14, where we simulated the splitting dynamics with mutation for two different choices of Δg .

3

ITERATIVE EXHAUSTION OF RESISTANCE MECHANISMS TO TARGETED THERAPY

3.1 INTRODUCTION

3.1.1 Can targeted therapy be curative?

When the first small molecules targeting molecular cancer abnormalities were approved for clinical use, scientists and clinicians were enthusiastic about the prospects of cure under targeted therapy. If a targeted drug (for example a kinase inhibitor) is delivered to a patient with the right mutational profile, the response is often dramatic and the tumor vanishes within few weeks – often with manageable side effects. It seemed as if the remaining task in drug discovery was more of an engineering than of a conceptual problem: Finding a molecule against each possible driver mutation would be enough to cure cancer. Unfortunately, today it is a well known fact that for solid tumors the efficacy of targeted therapy is limited by the occurrence of resistance in very nearly every case¹. Within few months, the tumors grow back, with changed mutational profile so that the tumor grows despite continued therapy. At first glance, this seems surprising: How can a tumor be so smart to evolve in exactly the right way to overcome a specific targeted drug? But when looking at the numbers governing tumor evolution, it becomes clear why resistance to targeted therapy is so prevalent.

Each tumor is initiated by a single mutated and quickly dividing cell. Due to unavoidable errors in DNA replication, at each cell division there is a small but non-zero probability for some mutation to occur. Generally, mutation rates

¹ There is one single example in which targeted therapy also works long-time, namely in chronic myeloid leukemia (CML) that can be treated with Imatinib. CML is driven by a gene fusion of the two genes BCR and ABL, which leads to an unusually short chromosome causing uncontrolled cell proliferation. Before the advent of targeted therapies, people inevitably died from CML within 4 – 6 years. Imatinib is a tyrosine kinase inhibitor specifically targeting the BCR-ABL fusion gene, and has shown to be impressively successful in patients: People who experienced a two year remission under Imatinib do not have a decreased life expectancy as compared to the general population [59].

are estimated to be in the order of magnitude of 10^{-8} [60] per nucleotide and cell division ². So a single cell is very unlikely to acquire a specific mutation in one division event. However, in a tumor the population size increases over time and the probability that a mutation in a given position arises in some cell rises concomitantly. While most mutations do not affect cell function, very specific mutations can confer resistance to targeted therapy. Thus the specificity of targeted therapy is both a blessing and a curse.

One gram (or approximately 1 cm^3) of tumor tissue contains not less than 10^8 individual cells [61]. Even though tumors as small as 4 mm can be theoretically detected in a CT scan, most patients are diagnosed with cancer in much later stages. For non-small cell lung cancers (NSCLC) for instance, a study on patients that have undergone surgery with curative intent reported that in their cohort less than 20% of all tumors were smaller than 2 cm (corresponding to about 10^8 cells), and more than 55% had a diameter of 3 cm (corresponding to 10^9) or more [62]. As patients with very large or metastasized tumors were excluded from the study, it is likely that lung tumors are often even bigger. This example shows that tumors at diagnosis usually consist of $10^8 - 10^9$ cells, and that, on average, $\mu N \approx 1 - 10$ mutants with each possible point mutation are created per generation. As a result, at the time treatment begins on a macroscopically-sized tumor, the population inevitably contains cells carrying a resistance mutation if a single nucleotide resistance mutation exists. These resistant cells expand under treatment while non-resistant cells are killed. Eventually, resistant cells repopulate the tumor, leading to the acquired resistance phenotype. Typically, there are different genetic alterations (point mutations, gene amplifications and deletions, ...) which confer resistance to a given targeted therapy. For example, tumors driven by an activating mutation in epidermal growth factor receptor (EGFR) respond to the EGFR tyrosine kinase inhibitor Erlotinib. Mutations conferring resistance to Erlotinib treatment include the well-studied *gatekeeper mutation* EGFR T790M [63, 64], but also point mutations in BRAF and PIK3CA, as well as copy number alterations in MET and HER2. A similar picture holds for targeted therapies acting on other mechanisms, like ALK-rearranged tumors where several ALK mutations as well as increased EGFR signaling and KIT amplification have been identified as possible resistance mechanisms [41]. In a population of cancer cells, the different resistance muta-

² This rate was measured for normal, non-cancerous human dermal fibroblasts. Usually, cancers are genetically unstable, which means that the mutation rate is elevated and that this estimate should be understood as a lower bound.

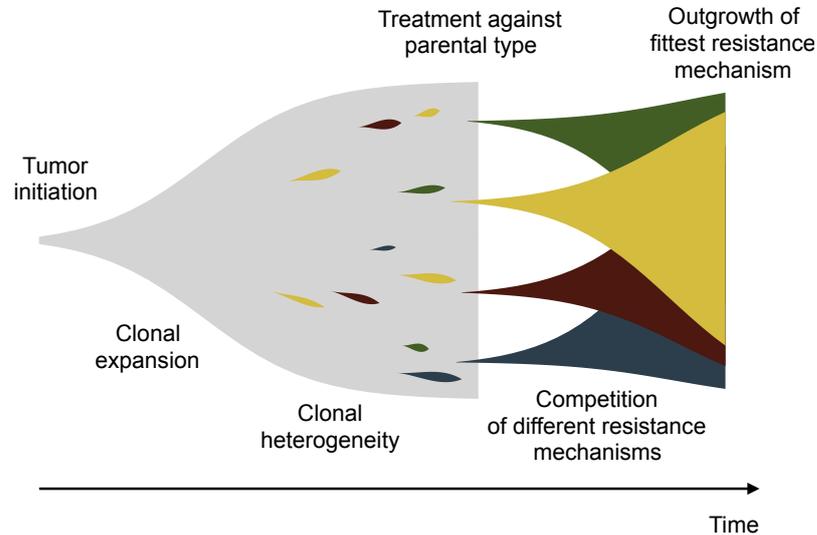


Figure 15: A Muller plot schematically shows the population dynamics of different genotypes with time running along the x-axis. Parental cancer cells, which respond to targeted therapy are indicated in gray. Different resistant mutants, indicated in red, yellow, green, and blue, appear through random mutations at different times and disappear again from the population as they typically grow slower than the parental cell type. If parental tumor cells are eradicated at some point in time due to targeted therapy, the remaining resistant cells expand and compete with each other.

tions arise independently, so mutant cells with different resistance mechanisms can co-exist in a large population of tumor cells. The different mechanisms then compete with each other once non-resistant cells have been eliminated by therapy, see Fig. 15. These considerations indicate that targeted mono-therapy can only be curative when applied to a tumor that does not contain any resistant mutants – a scenario that is very unlikely for macroscopically detectable tumors, see Fig. 16. Instead, if targeted therapy is to achieve a long-term remission, it needs to address all resistance mechanisms present in a population of cancer cells.

In the following, we use artificial evolution to study the problem of resistance to targeted therapy arising from pre-existing mutants. We iteratively isolate different resistance mechanisms existing in a large cell culture by amplifying resistant mutants. As a proof of principle, we apply our method to PC9 cells, a human adenocarcinoma-derived cell line with an activating mutation in the growth factor receptor EGFR. To analyze rare mutations, we culture large cell populations of size 5×10^8 cells. This number is far beyond the small pools containing thousands of

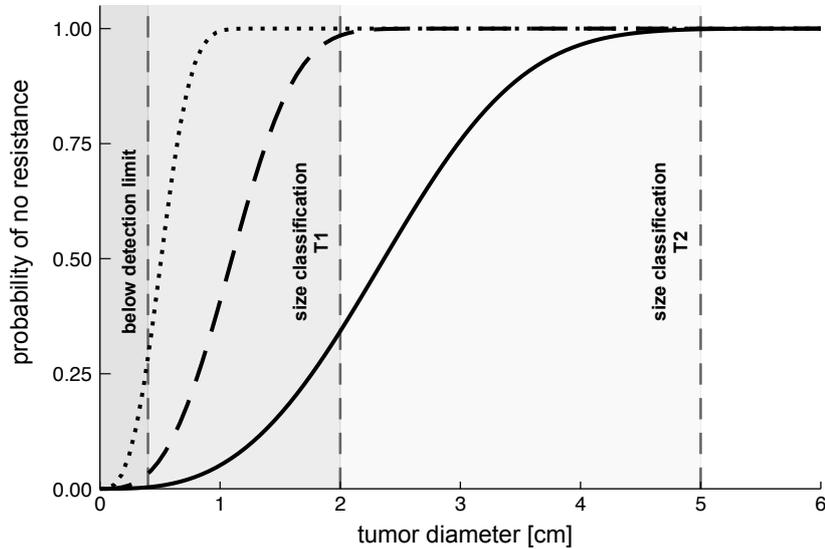


Figure 16: We show the probability of having at least one resistant mutant in the population as a function of tumor size for different mutation rates (solid line $\mu = 10^{-9}$, dashed line $\mu = 10^{-8}$ and dotted line $\mu = 10^{-7}$; roughly corresponding to one, ten or hundred resistance mechanisms mediated by point mutations occurring at the background point mutation rate; or to the background mutation rate, a tenfold or a hundredfold increased mutation rate, for instance due to genetic instability). The plot is based on equation (17) in the limit $\beta \rightarrow 0$. To map the population size N_{\max} to tumor size we assume a spherical tumor and a cell density of 10^8 cells per cm^3 . Different characteristic tumor sizes and stages are indicated along the x-axis. The plot shows that, if the mutation rate exceeds 10^{-7} , a detectable tumor inevitably contains therapy resistant mutants.

cells used previously [65] and is of the order of magnitude of detectable tumors. By characterizing the respective resistant lines, we show that several distinct resistance mechanisms exist simultaneously in the cell population prior to treatment and that their fitness determines which one emerges under treatment. Our key finding is that, for PC9 cells, a set of four compounds is sufficient to suppress all the resistance mechanisms we identified in such a population, and this holds even when the mutation rate is enhanced through a chemical mutagen. Our findings explain why, for solid tumors, long-term control of the disease with targeted therapy has proven elusive so far and point to a treatment strategy differing from current clinical practice: Instead of keeping the treatment fixed until a relapse occurs, tumor evolution needs to be anticipated by targeting a broad spectrum of possible resistance mechanisms as early as possible.

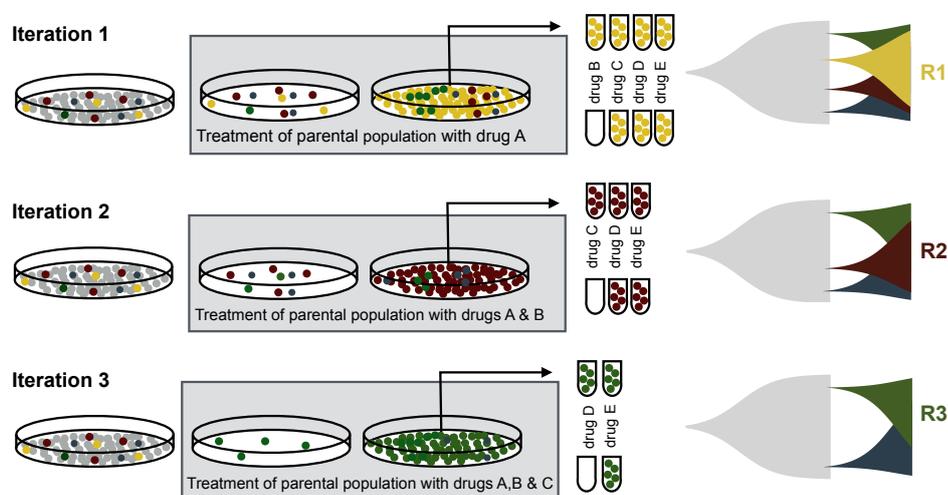


Figure 17: Our experimental scheme aims to isolate the fastest growing resistance mechanism by treating a large population of cells, most of which are susceptible to targeted therapy (gray points). The fastest growing resistant mutant (yellow) takes over the population and is screened against a panel of compounds covering a wide range of resistance mechanisms. In the next step, we again treat a large population, but with the original, first-line compound, as well as the compound active against the first resistance mechanism (yellow cells). This leaves a second resistance mechanism, which we isolate for screening and analysis. Each step generates a new resistance mechanism, as well as a compound acting on that mechanism.

3.1.2 Experimental protocol

Many resistance mechanisms can be targeted by specific drugs. If all resistance mechanisms present in a cancer cell population were known, and if we knew how to selectively target these resistance mechanisms, we could eliminate not only the bulk of therapy-susceptible cancer cells, but also eradicate the cells carrying resistance mutations. Key obstacle is that it may not be possible to identify resistant mutants by sequencing a large population of cells, as the frequency of mutants may be far below the detection threshold. To address this problem, we developed an experimental protocol to iteratively amplify all resistance mechanisms in a population of given size.

The idea is the following: On the one hand, we take a well defined model system driven by a known mutation that makes a cell cancerous (a so-called *driver mutation*). In the following we call this parental cell line P. On the other hand, we define a drug panel that covers drugs acting on that driver mutation and on

compound	targets	compound	targets
Erlotinib	EGFR	AZ628	RAF
Osimertinib	EGFR (T790M)	Imatinib	c-ABL/c-KIT/PDGFR
Lapatinib	EGFR/HER2	Sorafenib	RAF-1/B-RAF/VEGFR-2
Tepotinib	c-MET	Sunitinib	VEGFR-2/PDGFR β
Crizotinib	c-MET/ALK	Regorafenib	VEGFR/PDGFR β /KIT/RET/RAF-1
Trametinib	MEK	Dactolisib	PI ₃ K/mTOR
Dabrafenib	BRAF (V600)	Dasatinib	ABL/SRC/c-KIT
Apitolisib	PI ₃ K/mTOR	SCH77298	ERK
BGJ398	FGFR		

Table 1: Compounds and respective targets for the drug panel as employed in our experimental protocol. The drugs were chosen such that they cover all known and putative resistance mechanisms to targeted EGFR inhibition.

all known resistance mechanisms to this kind of treatment. In an initial round, we determine which drug works best on the P cell line. We call this compound drug A. We then iterate the following steps: We expand P up to a population size of 5×10^8 cells. We treat with drug A until resistant cells emerge. Several resistance mechanisms will survive the treatment, but the fittest one will usually take over the population. We call this new resistant cell line R₁. We screen R₁ against the drug panel to identify a drug that eliminates R₁, which we call drug B. We then take one step back and again expand P to obtain a population size of 5×10^8 cells. But now we treat with both drug A and drug B: Drug A will kill the parental type, whereas drug B works against R₁ such that the next fittest mutant type can take over the population (cell line R₂). We screen R₂ against the drug panel to determine the most effective compound for this cell type. For a sketch of the protocol see Fig. 17. These steps are iterated until a) we derive a cell line that is resistant against all drugs in our panel or b) we cannot derive a cell line resistant to some drug combination. Both possible outcomes of course have severe implications for the prospects of cure under targeted therapy.

To realize tumor-sized populations of 5×10^8 cells we use *hyperflasks*, special culture bottles that consist of ten intermediate layers on which cells can grow. The

crucial advantage is that these bottles are much more time and space effective than standard culture bottles: The whole bottle is filled with medium only once, whereas standard culture flask would require ten individual handling steps to reach the same population size. This cuts down the time that is used for each medium change nearly by a factor of ten. Moreover the hyperflask is only approximately twice as large as ordinary bottles, which saves valuable incubator space. An important disadvantage is that only the top and the bottom layer can be observed with a microscope, and even this is hard if the bottle is completely filled with cells as the density becomes too high for enough light to pass through all the layers. To circumvent this limitation, we transfer the surviving population in each iteration after two weeks of treatment to an ordinary culture flask with growth area 75 cm^2 (so called T75 growth bottle). This enables us to further monitor the population growth and to determine if cells are actually expanding under treatment. As soon as the population size is large enough, a sample of cells is frozen as a backup. The next time the population size reaches confluence, a sample is taken for DNA sequencing. Only after those two steps, the cell line is screened against the drug panel.

As a model system, we use PC9 cells, a human adenocarcinoma-derived NSCLC cell line driven by an EGFR mutation that has been studied extensively. PC9 cells are sensitive to the first generation EGFR tyrosine kinase inhibitor Erlotinib with a GI_{50} (concentration at which cell viability drops at 50% of control viability without treatment) of approximately 30 nM. When treated with Erlotinib, PC9 cells usually develop resistance via a secondary mutation in EGFR, the so called T790M gatekeeper mutation. Mutants carrying T790M can in turn be targeted with third generation EGFR inhibitors such as Osimertinib. The advantage of such a well established model system is that we already know what to expect for the first experimental iteration, which offers a possibility to check if our method works properly.

As Erlotinib has been approved for clinical use already in 2004, many resistance mechanisms have been described since then (for a review see [41]). Based on that, our drug panel covers all known and putative resistance mechanisms to EGFR inhibition. The full panel including drug targets is shown in Table 1.

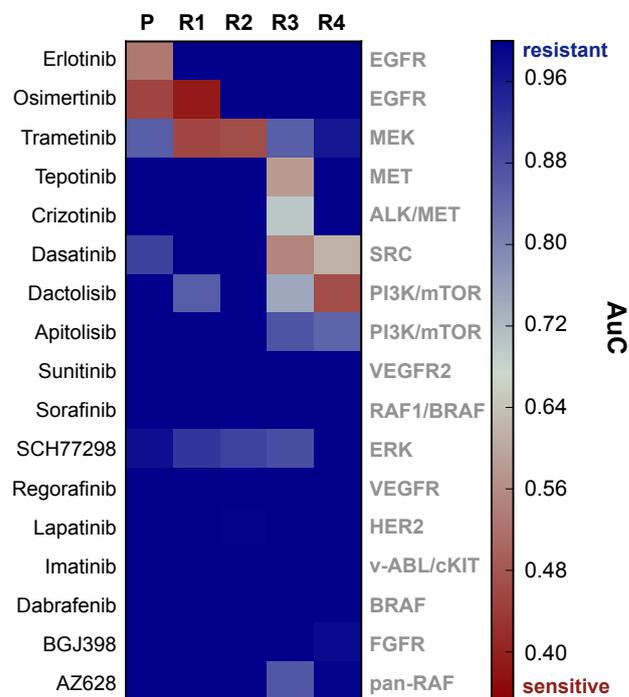


Figure 18: Drug responses of different compounds against the parental and the four different resistant cell lines. The main drug targets are annotated in gray. As a measure for sensitivity we use the area under the curve (AuC). High susceptibility, corresponding to a low AuC, is indicated in red. The responses are always measured as the formerly identified compounds plus the drug given in the table (meaning P is cultivated in pure growth medium plus the respective drug, R1 in Erlotinib plus the respective drug, R2 in Erlotinib, Osimertinib plus the respective drug, etc.). The resistance of successive cell lines to compounds previous lines are still susceptible to leads to the diagonal structure at the top right of the compound-susceptibility matrix.

3.2 RESULTS

3.2.1 Different resistance mechanisms pre-exist in a large cell population and respond to different compounds

Based on the above experimental protocol, we derived four different cell lines resistant to Erlotinib. R1 exhibits the well-known EGFR mutation T790M, and responds to the third-generation EGFR inhibitor Osimertinib. R2 exhibits a point mutation in NRAS, a gene downstream of EGFR in the MAPK pathway. This mutation, NRAS Q61R, is known to confer sensitivity to MEK inhibition. Indeed, R2

responds most strongly to the joint inhibition of EGFR and MEK. In R₃, we found an up-regulation of the growth factor receptor MET (which can at least partly substitute the EGFR signal) as well as a focal amplification and up-regulation of the corresponding ligand HGF. In accordance with this observation, R₃ responds most strongly to the joint inhibition of EGFR, MEK and MET. In R₄ we did not find an obvious genetic alteration responsible for therapy resistance, but found R₄ to be sensitive to Pi3K/mTOR inhibition. The responses of the full drug panel against the parental and the four different resistant cell lines R₁-R₄ is depicted in Fig. 18. Importantly, all cell lines respond specifically to compounds acting on a particular target instead of showing a broad response, which indicates that the drug combinations of three or more compounds are not generally toxic.

To determine the fitness of the different lines, we measured their growth rates, see Fig. 19. We find that the resistant cell lines do not only differ in drug sensitivity but also grow at different rates: R₁ grows at 96% percent of the rate of the parental PC9 cells, R₂ at 91%, R₃ at 74% and R₄ at 79% ($g_P = 0.82 \pm 0.02$, $g_{R1} = 0.79 \pm 0.01$, $g_{R2} = 0.75 \pm 0.01$, $g_{R3} = 0.61 \pm 0.04$ and $g_{R4} = 0.65 \pm 0.02$ cell divisions per day). Thus, the resistant cell lines grow more slowly than wild type cells, as is to be expected: If a resistance mutation offering a fitness advantage existed in the population, it would have taken over the population and would have become the wild type. Also, the fast-growing resistance mechanisms tend to be isolated first, as to be expected from our iterative scheme.

To test if the different cell lines maintain their resistance without selection pressure, we cultivated the four resistant lines without treatment and repeatedly measured their resistance to therapy. As shown in Fig. 20, no significant reduction in resistance was found after 14 weeks, which implies stable genetic or epigenetic resistance mechanisms.

3.2.2 Different resistance mechanisms show distinct disruptions in signaling pathways

To investigate if the sensitivity measurements are compatible with cell signaling, we examined the protein levels of the parental and the resistant cell lines. To this end, we performed Western blots of the most important components of the EGFR signaling pathway. As seen in Fig. 21, the delivery of Erlotinib shuts down the phosphorylated EGFR (pEGFR) and also the downstream pERK signal. Con-

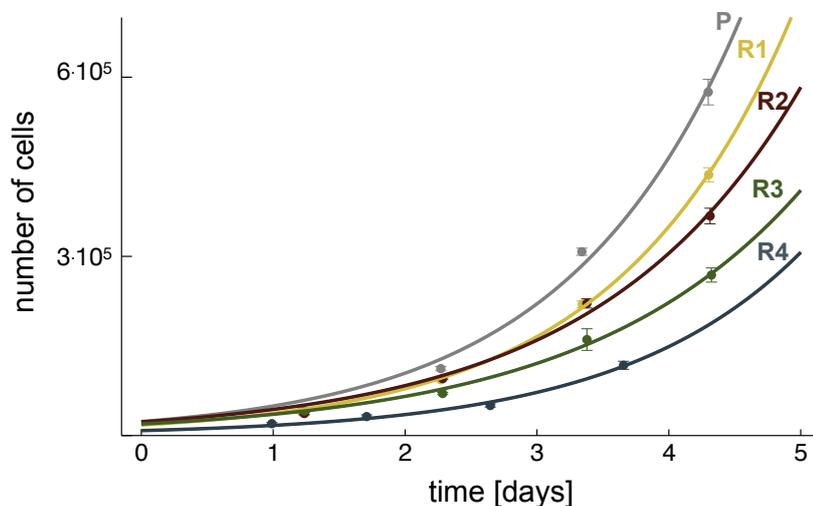


Figure 19: The different resistant cell lines differ in their growth rate, as seen by tracking their population growth over a four-day period. As expected, fast-growing resistant types are isolated first. The growth rate of R4 was measured by Carina Lorenz (lab of Martin Sos).

versely, in R1 the delivery of Erlotinib does not affect the EGFR or any downstream signal. This is to be expected: Erlotinib binds in an ATP-competitive manner to the EGFR kinase domain. The T790M mutation increases the binding affinity for ATP by an order of magnitude, which weakens the binding affinity for ATP-competitive agents such as Erlotinib [66]. This is why the inhibitor cannot effectively bind anymore and explains why the compound does not have any effect. Osimertinib in turn has a different mode of action and binds covalently, thus inhibiting EGFR even in the presence of high ATP concentrations [67]. This is why it overcomes the T790M mutation and shuts down the pEGFR and the respective downstream signal compatible with the sensitivity we measured in the viability assay. In R2, the pEGFR signal is eliminated by Erlotinib and Osimertinib, but, importantly, the downstream pERK signal remains. This means that there has to be some activation upstream of ERK that substitutes the pEGFR signal. This observation can be explained by the NRAS Q61R mutation that we found in R2. As expected, the additional delivery of Trametinib silences the downstream signal and restores sensitivity to EGFR inhibition.

In R3 we found a focal HGF amplification. It has been shown that a high concentration of the growth factor HGF in the culture medium makes cells resistant to EGFR inhibition [68]. Indeed, if we add HGF to normal growth medium, the

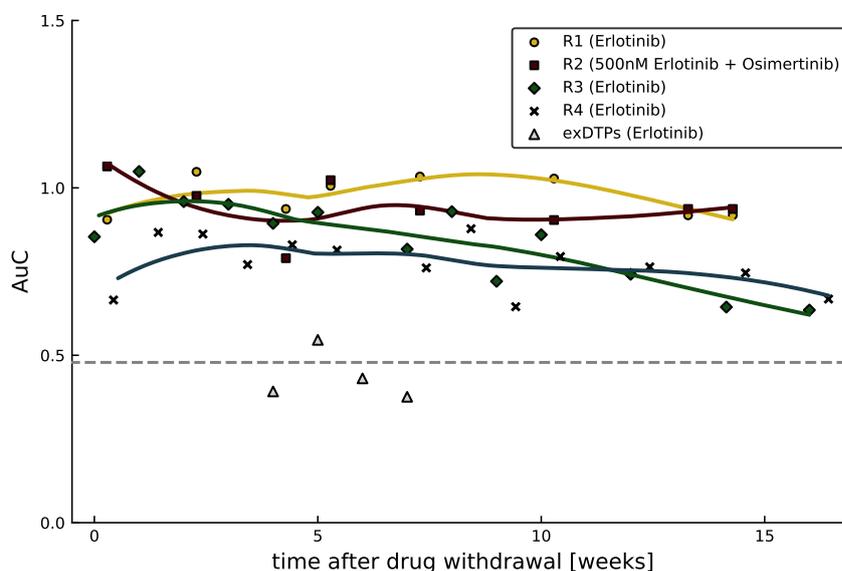


Figure 20: Stability of the resistant lines after compound withdrawal. The dashed line gives the AuC as measured in parental PC9 cells. The dots indicate individual measurements, the lines give trend lines. After 14 weeks, none of the cell lines reaches the sensitivity level of the parental line which indicates stable genetic or epigenetic resistance mechanisms. We further show the AuC as measured for cells derived from drug tolerant cells persisting despite therapy (exDTPs, see Section 3.2.4) after compound withdrawal. Once the culture is large enough for analysis, it exhibits the same sensitivity as the parental line which implies a reversible persister state instead of a stable mechanism.

GI₅₀ of the parental cells is increased by orders of magnitude, see Fig. 22, left panel. To measure the HGF concentration in the growth media of the different cell lines, we performed an HGF enzyme-linked immunosorbent assay (ELISA). For this method, samples from the growth medium of the cell line that is to be examined are transferred to a plate that is coated with antibodies that specifically bind to the target of interest (in our case HGF). After several incubation and washing steps, a substrate solution is added which develops color in proportion to the amount of bound HGF. The color intensity can be measured to quantify the amount of HGF that was secreted into the growth medium. Indeed, we find highly increased HGF levels in R3 as compared to the other cell lines. HGF is the ligand of MET, a growth factor receptor that can – among other signaling transduction pathways – activate the MAPK pathway. So the MET signal induced by the increased HGF amount offers a bypass-track for the missing EGFR signal and explains why R3 is resistant to EGFR inhibition.

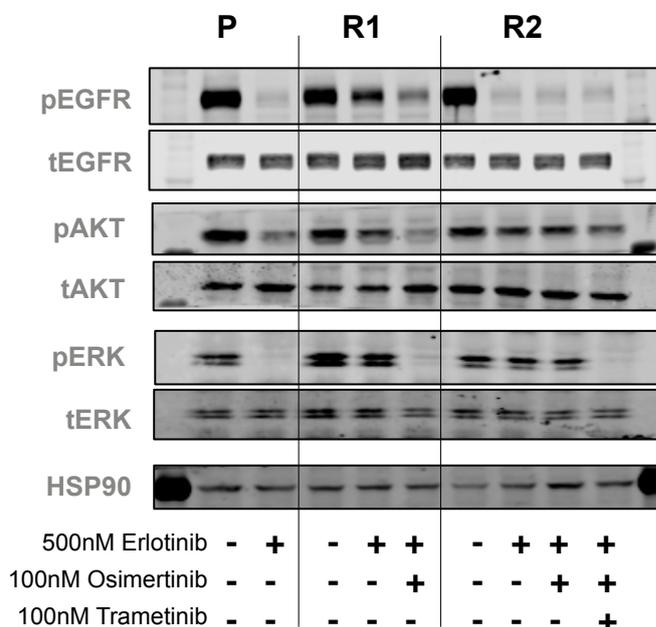


Figure 21: Protein levels of some important components of the MAPK pathway. Black bands show the presence of a specific protein, either the total amount (for instance tEGFR) or the phosphorylated version (for instance pEGFR). To ensure even protein loading in all samples, we further blotted the housekeeping protein HSP90 as a control. In the parental line P, the delivery of Erlotinib silences EGFR and the downstream signal in ERK. In R1, Erlotinib in turn does not change the EGFR signal and only the additional delivery of Osimertinib shuts down the pathway. In accordance with the observed NRAS mutation in R2, the EGFR inhibitors cancel the EGFR signal, but the ERK signal further downstream in the signaling pathway remains. Only the additional delivery of the MEK inhibitor Trametinib restores EGFR sensitivity. The western blot was performed by Carina Lorenz (lab of Martin Sos).

In R4 we could not identify a genetic alteration explaining the resistance. But as shown in Fig. 18, R4 reacts most strongly to the dual PI3K/mTOR inhibitor Dactolisib. The PI3K/AKT/mTOR pathway can have, when activated, multiple effects on cells, such as enhancing cell survival, stimulating proliferation and growth. The activation of this pathway has been reported in cancers in various contexts. To further determine which part of the pathway needs to be inhibited, we screened R4 against an extended panel of compounds, see Fig. 23, left panel. It becomes obvious that R4 is sensitive to mTOR or PI3K/mTOR dual inhibition, but not to AKT inhibition. We further checked protein levels to investigate cell signaling. The delivery of Dactolisib or Apatolisib does not affect the pERK signal, but shuts

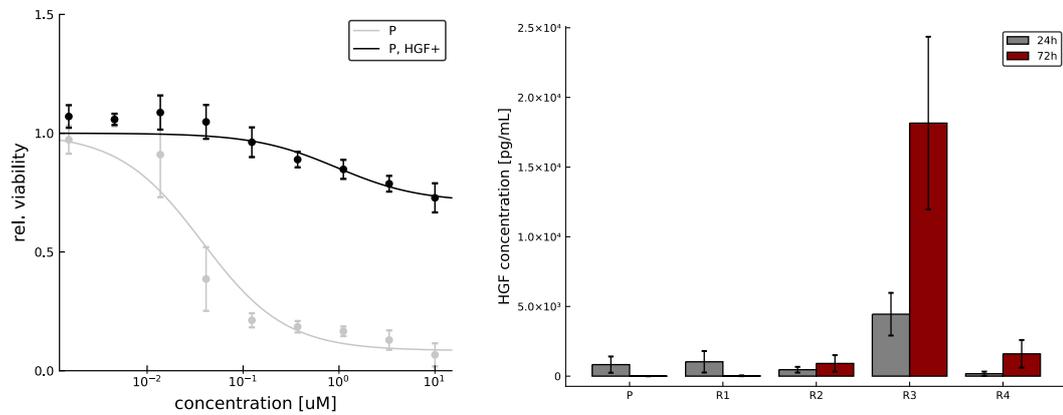


Figure 22: (left) The addition of the growth factor HGF to the culture medium make PC9 cells resistant to EGFR inhibition: Whereas the GI_{50} is usually less than 100 nM, the delivery of 50 ng/ml HGF makes the cells survive concentrations of several μM . (right) By performing an ELISA specific to HGF we measure increased HGF levels in the growth medium of the R3 cell line confirming the HGF amplification as the relevant resistance mechanism.

down p4EBP1 (a downstream target of mTOR) and pAKT. This indicates that the necessary survival signal in R4 is pAKT rather than pERK as in the parental cells.

3.2.3 No further resistance mechanisms arise under a combination of compounds targeting R1-R4

One interesting finding in the drug sensitivity matrix shown in Fig. 18 is that both R3 and R4 respond to the same drug, namely to the relatively broad inhibitor Dasatinib. This is surprising: even if we could not find a detailed explanation of the resistance mechanism in R4, it definitely differs from the HGF amplification we found in R3 (or any other MET related mechanism that would be targeted by the MET inhibitor Tepotinib). An obvious question to ask is if this is a coincidence or if there is something more general, meaning for instance that the combination Erlotinib, Osimertinib, Trametinib and Dasatinib (EOTD) is broad enough to target all possible resistance mechanisms in PC9 cells. To approach this question, we tried, according to our protocol, to derive cells resistant to the four drug combination EOTD. We failed to find any cells growing under this combination in three independent runs of the experiment. To exclude the possibility that some growing cells survived treatment but were lost in the necessary washing and centrifugation steps when transferring the surviving population from the hyperflask

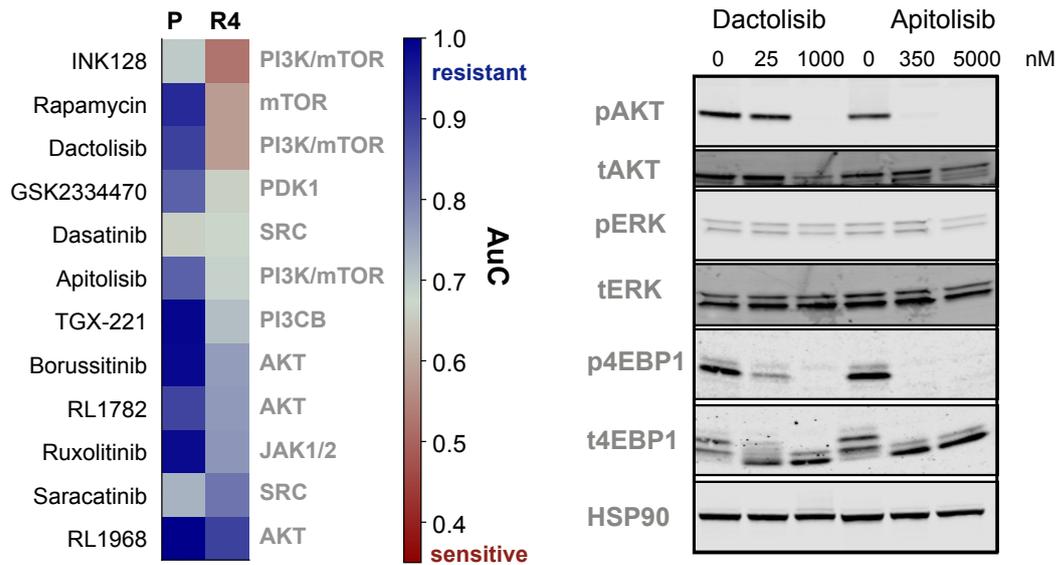


Figure 23: PI3K/mTOR activation drives resistance in R4. (left) Response to the extended drug panel targeting several parts of the PI3K/AKT/mTOR pathway. The parental cells are only treated with the indicated compounds, whereas in R4 we measure the additional effect to the standard R4 selection medium containing Erlotinib, Osimertinib, Trametinib and Tepotinib (EOTT). We find that R4 responds to mTOR and PI3K/mTOR dual inhibition, but not to AKT inhibition. (right) Western Blot of R4 cells showing some components of the PI3K/AKT/mTOR pathway. Again, the housekeeping gene HSP90 serves as a control to secure even protein loading and the treatment needs to be understood as EOTT plus the indicated compound. The delivery of Dactolisib or Apitolisib does not affect the pERK signal, but shuts down p4EBP1 (a downstream target of mTOR) and pAKT. This indicates that the necessary survival signal in R4 is pAKT rather than pERK as in the parental cells. The western blot was performed by Johannes Brägelmann (lab of Martin Sos).

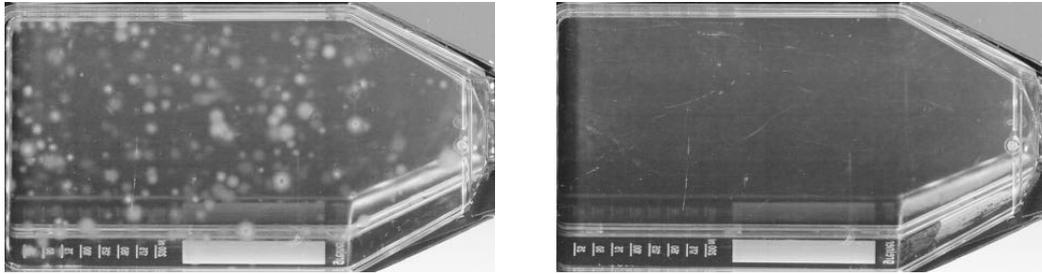


Figure 24: Culture flasks containing 5×10^7 cells are treated with ENU, such that the mutation rate is elevated by a factor of ten. After three weeks of treatment start, hundreds of small colonies (visible as gray spots) grow under Erlotinib mono-therapy (left). Under EOTD, not a single colony is observed (right).

(with a large growth area which is not completely observable) to the T75 culture flask (with a smaller growth area which is completely observable), we treated a population of 5×10^7 cells in a T175 bottle with a single dose of the chemical mutagen N-ethyl-N-nitrosourea (ENU) which enhances the mutation rate approximately tenfold [60]. Whereas in flasks treated only with Erlotinib hundreds of resistant colonies developed after two weeks of treatment, we did not observe a single colony under EOTD, see Fig. 24. Interestingly, the culture flasks were not completely empty: Even after several weeks of treatment, we could – with a microscope – find isolated cells (10 – 100 cells per T175 bottle). Importantly, those cells did not expand as they did not form small colonies over time. We conclude that, even if we could not identify further resistance mechanisms, the problem of resistance to EGFR inhibition in PC9 cells has yet another dimension: Cells that neither die nor grow under treatment.

3.2.4 Sleeper cells can survive the combination treatment

In order to probe if all proliferating cells from the original population had been eradicated, we stopped the EOTD treatment after seven weeks. Two weeks after the treatment had stopped, we observed the growth of a small number of colonies, obviously stemming from the few cells that persisted under EOTD treatment. Interestingly, these colonies responded to Erlotinib mono-therapy. To check if this behavior was reproducible, we treated a bottle of PC9 cells for one week with EOTD, stopped the treatment and waited until we had enough cells to screen them against Erlotinib. In several independent runs of the experiment, we could

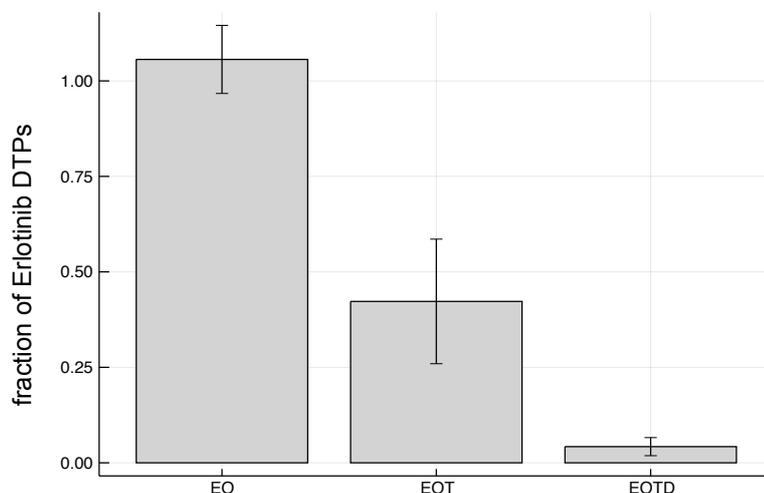


Figure 25: To quantify the population fraction that can survive under different compound combinations, we treated full bottles of parental PC9 cells with the indicated drugs. After four days, we counted the number of cells and normalized everything to the number of sleepers that survived under Erlotinib mono-therapy (Erlotinib DTPs). Here, we show the result of three independent runs of the experiment. The gray bars indicate the mean, the error bars the standard deviation. We did not observe a difference when treating with EO as compared to Erlotinib only. But when adding Trametinib, the surviving fraction drops by a factor of two, and when adding Dasatinib, even by a factor of ten.

always derive proliferating cells and did not observe a difference in Erlotinib sensitivity between the cells that emerged from sleeper cells and parental cells (see the gray triangles in Fig. 20). We concluded that the surviving cells were not truly resistant to the treatment, but a fraction of the population was in a drug tolerant persister state that was reversed after treatment had stopped.

The phenomenon of persisting cells has already been described in the context of Erlotinib resistance in PC9 cells: a part of the population, there called *drug-tolerant persisters* (DTPs), survives treatment and can eventually lead to the emergence of resistance [65, 69]. We measured that about 20% of the population survived four days of Erlotinib treatment. But under EOTD, we observed far fewer cells surviving treatment. To quantify the effect, we treated full bottles of parental cells with Erlotinib (E), Erlotinib and Osimertinib (EO), Erlotinib, Osimertinib and Trametinib (EOT) and EOTD. As shown in Fig. 25, we did not observe any difference between E and EO treatment. But when adding Trametinib, the surviving fraction drops by a factor of two, and when adding Dasatinib, even by a factor of ten. These observation could indicate that there exists a spectrum of different per-

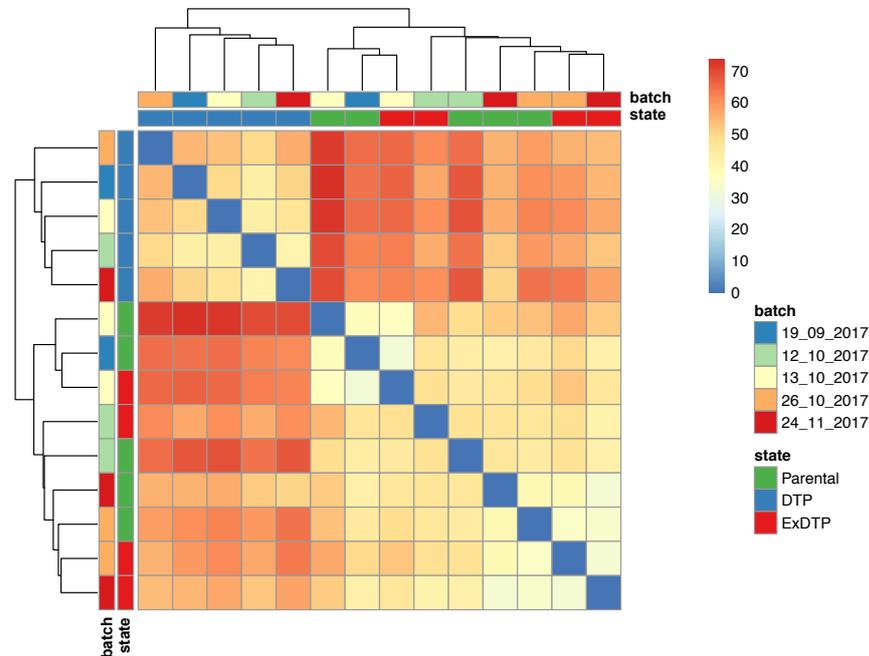


Figure 26: Cluster analysis of gene expression of sleeper cells (here called DTPs), cells that emerged from sleepers and parental cells. Blue colors indicate high similarity, red colors high differences. One can observe that two distinct clusters: One containing all sleepers and the other one containing both parental cells and cells that emerged from sleepers indicating that there are no major differences between the two types. This means that the sleeper state might be completely reversible. We also indicate the date at which we extracted RNA, to see if there are any batch effects, but we do not observe any dependence on the extraction date. The figure was created by Johannes Brägelmann (lab of Martin Sos).

sistence mechanisms (similar to the spectrum of co-existing resistance mechanism) that allows cells to survive under different compound combinations. Further work is needed to confirm this result.

To investigate the gene expression of the sleeper cells, we treated parental cells for one week with EOTD and extracted RNA from the surviving cells. As shown in Fig. 26, the expression of sleepers differs from the one of parental cell, but parental and cells that woke from the sleeper state fall into one cluster. This indicates that the persister state is completely reversible and probably not caused by genetic mutations. Previous work showed persister cells to be dependent on the lipid hydroperoxidase GPX4 for survival, thus identifying GPX4 as a possible target to inhibit [70]. While we observed a higher GPX4 expression in sleeper cells

as compared to parental cells, we did not find an increased sensitivity of sleepers to the GPX4 inhibitor RSL3.

Finally, we studied how the sleeper cells wake up from the persister state. We came up with two possible hypothesis: a) It could be an active process, meaning that sleepers stay in their drug tolerant state as long as treatment is maintained. Once the compounds are removed from the medium, cells sense the changed conditions and actively switch back to the proliferating state. This hypothesis would imply that the number of surviving cells does not depend on the treatment time. b) It could be a stochastic process, meaning that sleepers leave the persister state with certain probability all the time, independent of the treatment status. If the treatment is still on, cells immediately die. If the treatment is off, they can expand. This hypothesis implies that the number of surviving cells depends on the treatment time: the longer the population is treated, the lower the number of surviving cells. To test the two possibilities, we treated full bottles of parental cells with EOTD for different times (2, 3, 5 and 7 weeks). After the treatment stopped, we let surviving cells expand for 2.5 weeks and counted the emerging colonies. To simplify the quantification, we stained with crystal violet, two sample pictures are shown in Fig. 27. As depicted in Fig. 28, we observe an exponential decay in the number of colonies, falsifying the hypothesis that cells wake up in response to treatment stop.

3.3 DISCUSSION

Currently, the standard treatment for cancer patients who respond to some targeted therapy is to continue this treatment until a relapse occurs. This strategy implies that one waits for a macroscopically sized tumor to grow despite treatment before changing the medication. In most cases, the time until a relapse is roughly a year: A study investigating Erlotinib as first-line therapy³ compared to classical chemotherapy in EGFR driven NSCLC found a median progression free survival (PFS, the time during the treatment in which the tumor does not get bigger) of 9.7 months for Erlotinib versus 5.2 weeks for chemotherapy [71]. If a patient is lucky, the relapse is driven by a known resistance mechanism like the

³ The term *first-line therapy* is defined as the first treatment that is given for a disease. If at some point the medication is changed because the initial treatment stopped working or the patient suffers from unbearable side effects, the new treatment is called *second-line therapy*.

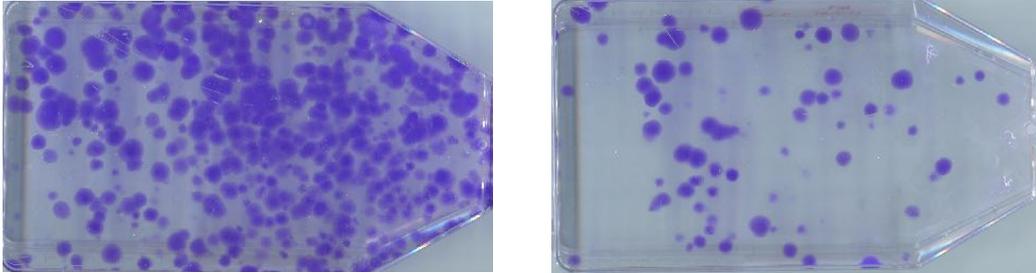


Figure 27: To study the dynamics of how cells exit the persister state, we treated full bottles of parental PC9 cells with EOTD and varied the treatment time. Each violet spot corresponds to one colony that was stained for counting. On the left, the population was treated for two weeks before cells could expand for two and a half weeks. On the right, the population as treated for five weeks and again cells could expand for two and a half week. It is obvious that the number of colonies drastically decreases over time.

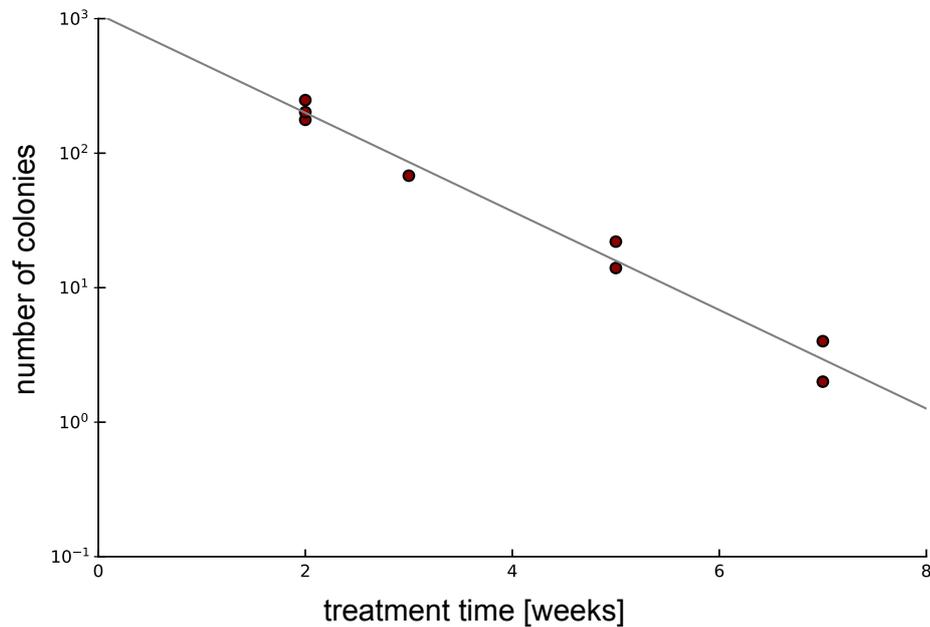


Figure 28: The number of surviving sleeper cells that initiate colonies of proliferating cells decays exponentially with treatment time. Each data point corresponds a picture as shown in Fig. 27 where we counted the number of colonies. The line corresponds to a fitted exponential curve. The data indicate that the number of colonies decays with rate of 0.844 per week.

T790M mutation that can be targeted by another drug. As long as the mutations driving the relapsed tumor can be identified and can be targeted with some drug, the procedure can be repeated. But it is clear that the approach can not lead to a cure, since macroscopic tumors inevitably contain therapy resistant mutants: In first-line therapy, resistance occurs on the background of the wild type. In second-line therapy, resistance occurs on the background of the fastest growing resistance mechanism to first-line therapy. Hence, the number of drugs a tumor is resistant to increases with each round of treatment, making it more and more difficult to find new treatment options. Nevertheless, the general strategy can lead to a life extension of several years, and due to the relatively harmless side effects of targeted therapies patients may have a high quality of life during this time. In many cases, this is much more than could be achieved with chemotherapy or radiation therapy.

However, if targeted therapy actually aimed at curing patients, it would need to address not only the wild type but all possible resistance mechanisms as early as possible. This intent requires complete knowledge about how cells can become resistant to a specific therapy. Our experimental protocol offers a generic approach to study the spectrum of resistance mechanisms in cell lines driven by different activating mutations. In our model system, we found a set of four compounds to be sufficient to target all resistance mechanisms existing in a population of given size. Further work needs to investigate if this result also holds in other EGFR driven cell lines. If so, tumor evolution of EGFR addicted cancers could be anticipated by targeting the full set of resistance mechanisms already during first-line therapy.

Apart from targeting all resistant cells, we have to find a way of dealing with the sleeper cells, which in our model system limit the long-time efficacy of the treatment. To date, it is unclear how relevant the problem of drug persistence is in patients, as a relapse usually occurs quickly and is presumably caused by pre-existing resistant mutants. This means that for now resistance and not persistence is the critical factor limiting therapy success. It might be that the sleepers are an artifact of cell cultures, and that in a patient for instance the immune system might eliminate them. But there is evidence that at least in xenograft mouse models the knockout of GPX4, a gene that is thought to be crucial for cell survival in the persister state, inhibits tumor relapse [70]. Therefore the sleeper cells might be as relevant for the long-term control of the disease in patients as they are in cell cultures, and future work should address this question.

There are several potential strategies to clinically cope with surviving sleepers. The straightforward approach would be to directly target the persister state. Previous work suggested different targets as possible vulnerabilities of sleeper cells surviving Erlotinib treatment, such as GPX4 [70], IGF-1R [69] or HDAC [72]. If these results could be confirmed in our sleepers surviving EOTD treatment, co-treatment with compounds targeting one or all of the identified vulnerabilities might eliminate the population completely. Another route to explore might be to actively wake up the sleeper cells. We showed that, once proliferating again, cells emerging from sleeper cells are sensitive to Erlotinib mono-therapy. One could exploit this behavior: There might be a messenger substance (that yet has to be identified) inducing the transition, and thus re-sensitizing the population to EGFR inhibition. Obviously, proliferating cells are much more harmful to the organism than persister cells, which is why in this approach tumor evolution had to be closely monitored. Finally, our results suggest a third treatment strategy, namely waiting until all sleepers have decayed over time. A short treatment period with EOTD would have to be followed by a long Erlotinib maintenance therapy targeting sleepers that wake up from the persister state. Assuming a tumor size of 10^9 cells, a sleeper fraction of 1% and a decay rate as measured of 0.844 per week, this would mean that after 30 weeks all sleeper cells would have been destroyed with probability 99.99% – given that the drug is fully available at the tumor site. As a single surviving persister cell could wake up, expand and cause a relapse, it is crucial to actually eradicate the population completely. Therefore the treatment should not be stopped, even if the patient does not exhibit any cancer related symptoms anymore. In any case, the biology of persistence is very complex (probably as complex the one of resistance), and future work should aim at further characterizing sleeper cells and revealing their molecular features.

4

INFERENCE OF CELLULAR STATES FROM SINGLE-CELL LIFETIMES

4.1 MOTIVATION: PERSISTENCE AS A WAY TO SURVIVE TREATMENT

In the previous chapters, we focused on cancer resistance that is due to resistant mutants existing in a population of sensitive cells prior to any kind of treatment. But there is another way in which cancer cells can escape from treatment: cells can enter a cellular state in which they divide very slowly, but are relatively drug resistant. The degree of resistance as a function of growth rate is currently unknown. The phenomenon has been shown in the context of resistance to EGFR inhibition in PC9 cells [69]. Cells in the persistent state are called drug tolerant persisters (DTPs). Under prolonged treatment, drug resistant mutants can emerge from DTPs [65]. In section 3.2.4 we showed that, also in our model system, sleeper cells persisting under therapy limit the long-time efficacy of the treatment. This means that, even after eliminating all quickly dividing resistant cell types, DTPs survive and can lead to a relapse. Therefore, persistence might be clinically relevant in the maintenance treatment of cancer patients.

To date, it is not clear what drives the transition to the persister state. We focus on a study that was conducted on PC9 cells and which points to the hypothesis that a fraction of the population is always in the slow dividing state – independent of being treated or not [73]. Using time-lapse automated imaging, the lifetimes of single cells were recorded. By plotting the histogram of the measured lifetimes, the authors observed that two peaks emerged – one at small lifetimes (≈ 15 hours) and another at larger lifetimes ($\approx 40 - 50$ hours). The authors interpreted this observation as two distinct subpopulations and fitted a model to the histogram with parameters for cell birth, death and transition to the persistent state. Whereas this method is an easy approach to determine the parameters governing the underlying process, the analysis does not exploit all of the available information. If one assumes that the average lifetime is fixed at cell birth and that the state is at least

to some degree heritable, the distribution of lifetimes along the phylogenetic tree offers useful information on the probability of switching to or from the persistent state.

In this chapter, we develop a statistical model to infer heterogeneous growth rates in a population. As an example of such heterogeneity we assume a mixture of normal cells and persisters, where the entire population is tracked by imaging and individual birth and death events are recorded. We ask how to infer the existence and heritability of states with different growth rates (lifetimes until division) for such data. We then present a framework to infer both the mean lifetimes and the switching probabilities from synthetic data and discuss the data required for applications.

4.2 LIFETIMES ON A HIDDEN MARKOV TREE

We consider the population dynamics defined by a hidden Markov model as sketched in Fig. 29. Consider N cells which live on a binary tree representing the kinship of the cells: a link between two cells connects a cell to its direct ancestor. Each cell is in one out of two discrete states, which we call active (+) for the quick-dividing and quiescent (−) for the slow dividing state. The generalization to more states is straightforward. The cellular state is a hidden or *latent* variable which cannot be observed directly, but which impacts the observed lifetime: For each element, a lifetime is drawn from some distribution which depends on the cellular state. The cellular state is determined at cell division and does not change during the lifetime of a cell. We consider a Markovian dynamics of these states $\{\sigma_1, \dots, \sigma_N\}$, with a transition probability

$$A_{jk} = p(\sigma_i = k | \text{par}(\sigma_i) = j) \quad (47)$$

between a cell i and its parent $\text{par}(\sigma_i)$ with $j, k \in \{+, -\}$. As a starting point, we assume the lifetimes to be exponentially distributed,

$$p(\tau_i | \sigma_i) = \frac{1}{T_{\sigma_i}} \exp(-\tau_i / T_{\sigma_i}), \quad \sigma_i \in \{+, -\} \quad (48)$$

with a mean expected lifetime T_{σ_i} that is much higher in the quiescent than in the active state.

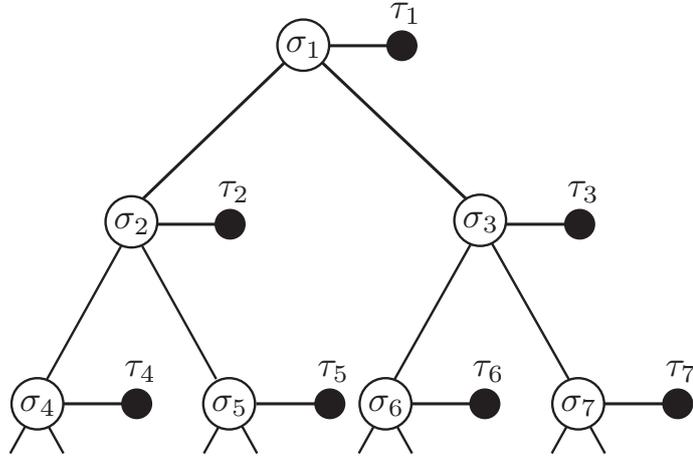


Figure 29: Single cell lifetimes on a hidden Markov tree. The states σ_i , active or quiescent, are not observable and are thus the hidden variables. But they determine from which distribution the lifetimes τ_i , which are directly observable, are drawn.

We cannot observe the hidden variables $\sigma = \{\sigma_1, \dots, \sigma_N\}$. Therefore, in the probability distribution of one specific set of lifetimes we have to sum over all configurations of the hidden variables,

$$p(\boldsymbol{\tau}) = \sum_{\sigma_1=\{+,-\}} \cdots \sum_{\sigma_N=\{+,-\}} p(\boldsymbol{\tau}, \boldsymbol{\sigma}), \quad (49)$$

with $\boldsymbol{\tau} = \{\tau_1, \dots, \tau_N\}$.

The joint distribution over the lifetimes and the hidden states $p(\boldsymbol{\tau}, \boldsymbol{\sigma})$ consists of two parts: First, there is a term describing the transition between states. This couples the state of each cell with the respective parental state. Second, there is a term describing the probabilities of the lifetimes given the states which we call emission probabilities. The observed lifetimes of one cell only depend on its own cellular state – so there is no coupling between neighboring cells. This means that the joint probability partly factorizes:

$$p(\boldsymbol{\tau}, \boldsymbol{\sigma}) = p(\sigma_1) \cdot \underbrace{\prod_{n=2}^N p(\sigma_n | \text{par}(\sigma_n))}_{\text{transition probabilities}} \cdot \underbrace{\prod_{m=1}^N p(\tau_m | \sigma_m)}_{\text{emission probabilities}}. \quad (50)$$

This statistical model is determined by five independent parameters: The two mean lifetimes T_+ and T_- govern the emission probabilities. The transition probabilities A_{+k} need to be normalized, which means $A_{+-} = 1 - A_{++}$ (analogously for A_{-k}). So the transition probabilities only have two independent parameters. There is one more parameter describing the probability of the initial state $p(\sigma_1)$, which we call π . If we make these parameter dependencies explicit, we find

$$p(\boldsymbol{\tau}, \boldsymbol{\sigma} | \pi, A_{++}, A_{--}, T_+, T_-) = p(\sigma_1 | \pi) p(\tau_1 | \sigma_1, T_+, T_-) \cdot \prod_{n=2}^N p(\sigma_n | \text{par}(\sigma_n), A_{++}, A_{--}) \cdot p(\tau_n | \sigma_n, T_+, T_-). \quad (51)$$

Our model can formally be mapped onto an asymmetric Ising problem on a tree, with states of cells corresponding to spins and rates at which these states change from one generation to the next generating couplings between these spins. Each cell can be in two different states, which means that the system of N cells can take 2^N different configurations. Therefore, the number of configurations scales exponentially in the number of cells. In order to determine the parameters of the model, we need to calculate expectation values over the configuration space which means summing over 2^N configurations. To do this efficiently, we use an iterative algorithm starting at the leaves of the tree and progressively proceeding to the root reducing the number of computation steps to polynomial order. In physics, this approach is well known as the transfer matrix method. It is also widely used in the machine learning community, where it is called belief propagation.

4.3 THE BACKWARD PROBLEM: LEARNING THE PARAMETERS

To learn the parameters of the model we use a Bayesian approach. We aim at calculating the Bayesian posterior

$$p(\theta|\tau) = \frac{p(\tau|\theta) \cdot p(\theta)}{p(\tau)}, \quad (52)$$

where $\theta = \{\pi, A_{++}, A_{--}, T_+, T_-\}$ is the set of parameters. We want to maximize $p(\theta|\tau)$ with respect to θ . From equations (49) and (51) we know how to express the probability of the data given the parameters as a function of θ . Moreover, $p(\tau)$ does not depend on θ . For the case where we have no prior information on the parameters, a reasonable assumption is that the prior probability $p(\theta)$ is flat and does not depend on θ . Under this assumption, the posterior probability distribution of the parameters conditioned on the observations is proportional to probability of the observations conditioned on the parameters,

$$p(\theta|\tau) \propto p(\tau|\theta). \quad (53)$$

The function $p(\tau|\theta)$ is called the likelihood of the parameters, and the resulting set of parameters

$$\theta^{\text{ML}} = \underset{\theta}{\operatorname{argmax}} p(\tau|\theta) \quad (54)$$

the maximum likelihood estimator. In the following, we seek for the maximum likelihood estimator given a set of observations subject to the model above.

4.3.1 The expectation maximization algorithm

The expectation maximization (EM) algorithm is a very general and broadly applicable method to calculate maximum likelihood estimates from data with underlying models that have latent variables [74]. In the following, we first give a general sketch of how EM works. We then apply it to our statistical model. The line of argument follows [75] where the calculation was performed for a linear hidden Markov chain designed to model sequential data.

Our goal is to maximize the likelihood $p(\tau|\theta)$. For convenience, we maximize the logarithm of the likelihood (called log-likelihood). To do this efficiently, we

decompose the log-likelihood in two parts. We introduce a probability distribution $q(\sigma)$ and rearrange terms

$$\begin{aligned}
\ln p(\tau|\theta) &= \sum_{\sigma} q(\sigma) \ln p(\tau|\theta) \\
&= \sum_{\sigma} q(\sigma) \ln \frac{p(\sigma, \tau|\theta)}{p(\sigma, \tau|\theta) \frac{1}{p(\tau|\theta)}} \\
&= \sum_{\sigma} q(\sigma) \ln \frac{p(\sigma, \tau|\theta)}{p(\sigma|\tau, \theta)} \\
&= \sum_{\sigma} q(\sigma) \left[\ln \frac{p(\sigma, \tau|\theta)}{q(\sigma)} - \ln \frac{p(\sigma|\tau, \theta)}{q(\sigma)} \right] \\
&= \mathcal{L}(q, \theta) + D_{\text{KL}}(q||p),
\end{aligned} \tag{55}$$

where we defined

$$\mathcal{L}(q, \theta) = \sum_{\sigma} q(\sigma) \ln \frac{p(\sigma, \tau|\theta)}{q(\sigma)} \tag{56}$$

and the Kullback-Leibler (KL) divergence between q and p

$$D_{\text{KL}}(q||p) = - \sum_{\sigma} q(\sigma) \ln \frac{p(\sigma|\tau, \theta)}{q(\sigma)}. \tag{57}$$

The KL divergence $D_{\text{KL}}(p||q)$ has the property $D_{\text{KL}}(p||q) \geq 0$, where the equality is saturated if and only if $q(\sigma) = p(\sigma|\tau, \theta)$ [76]. So $\mathcal{L}(q, \theta)$ is a lower bound to the log-likelihood, and this bound is obviously saturated if the KL divergence vanishes.

The EM algorithm for maximizing the log-likelihood exploits this fact. It consists of two steps:

1. **E step:** The lower bound is maximized with respect to $q(\sigma)$, but the parameters $\theta = \theta^{\text{old}}$ remain constant. This is achieved by choosing

$$q(\sigma) = p(\sigma|\tau, \theta^{\text{old}}) \tag{58}$$

such that $D_{\text{KL}}(p||q) = 0$.

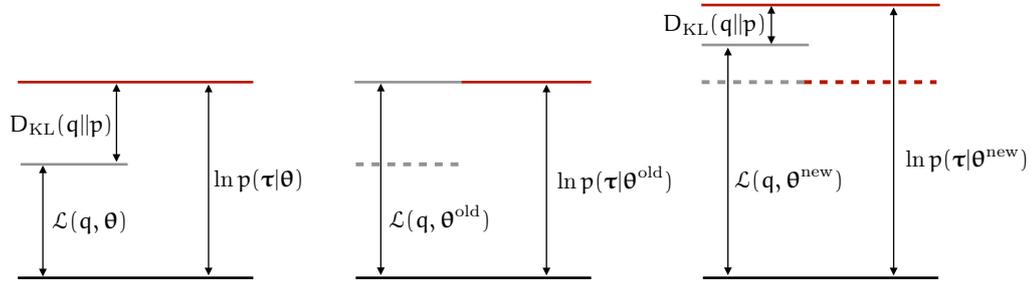


Figure 30: Sketch of how the EM algorithm works. The log-likelihood is decomposed into two parts. In the E step, the lower bound is maximized with respect to the auxiliary distribution q . In the M step, the lower bound is maximized with respect to the parameters θ . If not already at its maximum, this will lead to an increase in the log-likelihood and a non-vanishing Kullback-Leibler divergence.

2. **M step:** The lower bound is maximized with respect to θ . By rearranging terms a little further, we see that only a part of \mathcal{L} depends on θ ,

$$\begin{aligned}
 \mathcal{L}(q, \theta) &= \sum_{\sigma} q(\sigma) \ln \frac{p(\sigma, \tau|\theta)}{q(\sigma)} \\
 &\stackrel{(58)}{=} \sum_{\sigma} p(\sigma|\tau, \theta^{\text{old}}) \ln \frac{p(\sigma, \tau|\theta)}{p(\sigma|\theta^{\text{old}})} \\
 &= \underbrace{\sum_{\sigma} p(\sigma|\tau, \theta^{\text{old}}) \ln p(\sigma, \tau|\theta)}_{L(\theta, \theta^{\text{old}})} + \text{const.} \tag{59}
 \end{aligned}$$

Maximizing $L(\theta, \theta^{\text{old}})$ with respect to θ yields new parameters

$$\theta^{\text{new}} \equiv \operatorname{argmax}_{\theta} [L(\theta, \theta^{\text{old}})]$$

If these new parameters differ more than some convergence criterion from the old ones, we set $\theta^{\text{old}} = \theta^{\text{new}}$ and go back to the E step.

In Fig. 30 we show graphically how the EM algorithm works. Instead of directly maximizing $\ln p(\tau|\theta) = \ln \sum_{\sigma} p(\tau, \sigma|\theta)$, we iteratively maximize $L(\theta, \theta^{\text{old}})$. This has the crucial advantage that the summation over the hidden variables now ap-

pears outside the logarithm. As we saw earlier, the joint probability distribution of our model partly factorizes. By plugging in equation (51), we find

$$L(\theta, \theta^{\text{old}}) = \sum_{\sigma} p(\sigma|\tau, \theta^{\text{old}}) \left(\ln p(\sigma_1|\pi) + \sum_{n=2}^N \ln p(\sigma_n|\text{par}(\sigma_n), A_{++}, A_{--}) + \sum_{m=1}^N \ln p(\tau_m|\sigma_m, T_+, T_-) \right). \quad (60)$$

Crucially, not all terms depend on all hidden variables, and we can partly perform the summation the latent variables

$$\begin{aligned} L(\theta, \theta^{\text{old}}) &= \sum_{\sigma_1} p(\sigma_1|\tau, \theta^{\text{old}}) \ln p(\sigma_1|\pi) \\ &+ \sum_{n=2}^N \sum_{\sigma_n} \sum_{\sigma_{n/2}} p(\sigma_n, \sigma_{n/2}|\tau, \theta^{\text{old}}) \ln p(\sigma_n|\sigma_{n/2}, A_{++}, A_{--}) \\ &+ \sum_{m=1}^N \sum_{\sigma_m} p(\sigma_m|\tau, \theta^{\text{old}}) \ln p(\tau_m|\sigma_m, T_+, T_-), \end{aligned} \quad (61)$$

where we chose a labeling in which node σ_i is the parent of node σ_{2i} and σ_{2i+1} , and where $n/2$ refers to $\lfloor \frac{n}{2} \rfloor$. Equation (61) can be easily maximized with respect to the parameters. We find for the bias of the initial state

$$\pi = \frac{p(\sigma_1 = +|\tau, \theta)}{p(\sigma_1 = +|\tau, \theta) + p(\sigma_1 = -|\tau, \theta)}, \quad (62)$$

for the transition probabilities

$$A_{++} = \frac{\sum_{n=2}^N p(\sigma_n = +, \sigma_{n/2} = +|\tau, \theta)}{\sum_{n=2}^N p(\sigma_n = +, \sigma_{n/2} = +|\tau, \theta) + \sum_{n=2}^N p(\sigma_n = -, \sigma_{n/2} = +|\tau, \theta)}, \quad (63)$$

$$A_{--} = \frac{\sum_{n=2}^N p(\sigma_n = -, \sigma_{n/2} = -|\tau, \theta)}{\sum_{n=2}^N p(\sigma_n = +, \sigma_{n/2} = -|\tau, \theta) + \sum_{n=2}^N p(\sigma_n = -, \sigma_{n/2} = -|\tau, \theta)}, \quad (64)$$

and, assuming exponentially distributed lifetimes, for the parameters determining the emission probabilities

$$T_+ = \frac{\sum_{n=1}^N \tau_n p(\sigma_n = +|\boldsymbol{\tau}, \boldsymbol{\theta})}{\sum_{n=1}^N p(\sigma_n = +|\boldsymbol{\tau}, \boldsymbol{\theta})}, \quad (65)$$

$$T_- = \frac{\sum_{n=1}^N \tau_n p(\sigma_n = -|\boldsymbol{\tau}, \boldsymbol{\theta})}{\sum_{n=1}^N p(\sigma_n = -|\boldsymbol{\tau}, \boldsymbol{\theta})}. \quad (66)$$

The remaining task is to efficiently calculate the marginal probabilities

$$p(\sigma_n|\boldsymbol{\tau}, \boldsymbol{\theta}) = \prod_{j \neq n} \sum_{\sigma_j} \frac{p(\boldsymbol{\sigma}, \boldsymbol{\tau})}{p(\boldsymbol{\tau})} = c \cdot \prod_{j \neq n} \sum_{\sigma_j} p(\boldsymbol{\sigma}, \boldsymbol{\tau}), \quad (67)$$

where c is a constant which is set by normalization, and $p(\sigma_n, \sigma_{n/2}|\boldsymbol{\tau}, \boldsymbol{\theta})$, which is accordingly given by summing over all hidden variables except for σ_n and $\sigma_{n/2}$.

4.3.2 Calculating the marginal probabilities using belief propagation

Belief propagation (BP) is an algorithm to efficiently calculate marginal probabilities of nodes in a Bayesian network [77]. The idea is to organize the computation by introducing variables $m_{i \rightarrow j}(\sigma_j)$ for all neighboring pairs $\{i, j\}$. The variable $m_{i \rightarrow j}(\sigma_j)$ can be interpreted as a message node i sends to node j , about what state the hidden variable σ_j should be in based on the information node i has. The messages are self-consistently defined as

$$m_{i \rightarrow j}(\sigma_j) = \sum_{\sigma_i} p(\tau_i|\sigma_i) \cdot p(\sigma_i|\sigma_j) \prod_{k \in \mathcal{N}(i) \setminus j} m_{k \rightarrow i}(\sigma_i), \quad (68)$$

where $\mathcal{N}(i)$ means the set of neighbors of node i . From the messages, the marginal probabilities can be calculated as

$$p(\sigma_i|\boldsymbol{\tau}) = c \cdot p(\tau_i|\sigma_i) \prod_{j \in \mathcal{N}(i)} m_{j \rightarrow i}(\sigma_i), \quad (69)$$

where c is a normalization constant. Also, the two state marginal probabilities can be expressed in terms of the messages, namely as

$$p(\sigma_n, \sigma_{n/2}) = c \cdot p(\tau_n | \sigma_n) \cdot p(\tau_{n/2} | \sigma_{n/2}) \cdot p(\sigma_n | \sigma_{n/2}) \prod_{k \in \mathcal{N}(n) \setminus n/2} m_{k \rightarrow n}(\sigma_n) \prod_{j \in \mathcal{N}(n/2) \setminus n} m_{j \rightarrow n/2}(\sigma_{n/2}). \quad (70)$$

If one has all messages at hand, one can determine all required marginal probabilities. This means that every message only has to be calculated once. So the number of computations scales polynomially in the number of links M : as each link is occupied by two messages, the number of messages is $2M$. For loop-free graphs, BP gives the exact marginals. The reason is that in such tree-like structures, one can start by calculating the messages at the leaves and then progressively go up the tree. So writing everything in terms of messages is just a re-ordering of terms in equation (67). In physics, this approach is well known as the transfer matrix method. This line of argument breaks down once the graphs has loops; in this case, BP is only an approximate method [78].

In our problem, we have a binary tree, in which each node – except for the root and the leaves – has exactly two children and one parent. This means that each message that is sent from the bottom to the top of the tree only depends on other bottom-up messages. Messages that are sent from the top down the tree, conversely, depend both on bottom-up and top-down messages (see Fig. 31 for a sketch). By starting at the leaves, we first calculate all bottom-up messages. We then go down the tree, starting at the root and calculate all top-down messages. The message each node sends in direction of the root is given by

$$m_{i \rightarrow i/2}(\sigma_{i/2}) = \begin{cases} \sum_{\sigma_i} p(\sigma_i | \sigma_{i/2}) p(\tau_i | \sigma_i) & \text{if } i \text{ is a leaf} \\ \sum_{\sigma_i} p(\sigma_i | \sigma_{i/2}) p(\tau_i | \sigma_i) m_{2i \rightarrow i}(\sigma_i) m_{2i+1 \rightarrow i}(\sigma_i) & \text{else.} \end{cases} \quad (71)$$

Analogously, the top-down message that a node sends to its left child is given by

$$m_{i \rightarrow 2i}(\sigma_{2i}) = \begin{cases} \sum_{\sigma_1} p(\sigma_1) p(\sigma_2 | \sigma_1) p(\tau_1 | \sigma_1) m_{3 \rightarrow 1}(\sigma_1) & \text{for } i = 1 \\ \sum_{\sigma_i} p(\sigma_{2i} | \sigma_i) p(\tau_i | \sigma_i) m_{2i+1 \rightarrow i}(\sigma_i) m_{i/2 \rightarrow i}(\sigma_i) & \text{else,} \end{cases} \quad (72)$$

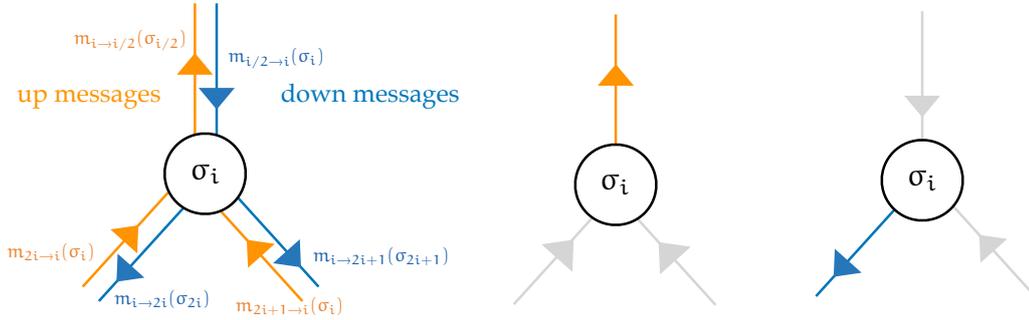


Figure 31: Sketch of messages flowing in and out of node i . As the up messages only depend on other top messages further down the tree, one first calculates all top messages. The down messages can then be calculated from the root to the leaves.

with $2i$ and $2i + 1$ interchanged for the message that is sent to the right child.

4.4 RESULTS AND DISCUSSION

We test our approach on artificially generated data. We simulate m generations of cells, which yields a complete binary tree of lifetimes (each node except for the leaves has exactly two children) of length

$$N = \sum_{i=0}^m 2^i.$$

We chose the state σ_1 of the initial cell randomly and draw a lifetime τ_1 from an exponential distribution with mean T_+ if $\sigma_1 = +$ or T_- if $\sigma_1 = -$. The state of each subsequent cell is inherited with probabilities A_{++} or A_{--} from the parental state, and the lifetimes are drawn from the respective exponential distribution.

In Fig. 32 we show the inferred parameters for different tree sizes N . Each dot is an inferred parameter set for one simulated tree of respective length. For short trees $N = 15$ and $N = 63$ (corresponding to 3 and 5 cell generations), the reconstruction works very poorly. This is because the trees are simply too short for short any transitions to occur. As we chose $A_{++} = A_{--} = 0.95$, it is very unlikely to find transitions within one tree. Hence the underlying model of two states is wrong: in each tree, lifetimes are only drawn from one distribution. This is also why many reconstructed rates fall on the boundaries 0.0 and 1.0. For longer trees,

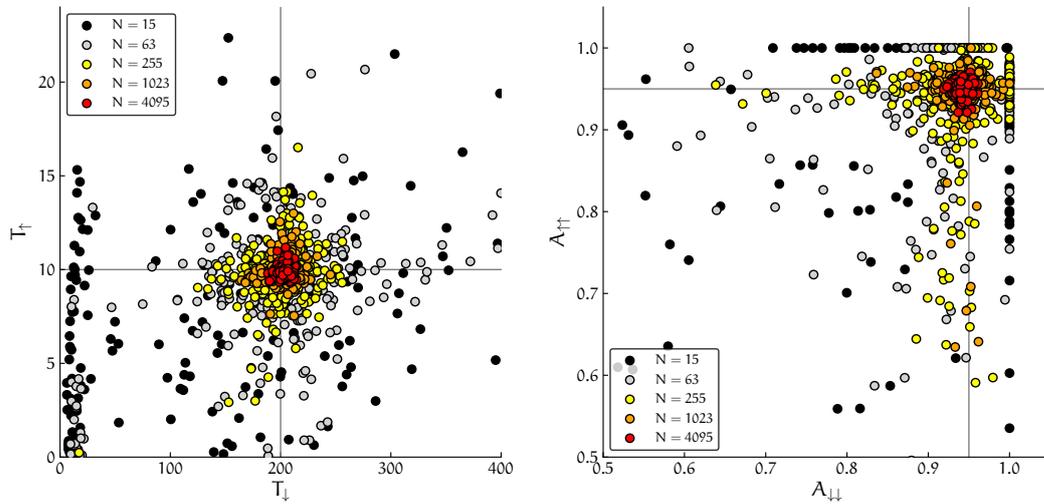


Figure 32: Inferred mean lifetimes and switching rates for different nodes N . Each dot represents a set of inferred parameters for one simulated tree consisting of N cells. The lines indicate the parameters that were used to simulate the data. As expected, the reconstruction gets more reliable the longer the trees are. For shorter trees with $N = 15$ and $N = 63$ the reconstructed rates often fall on the boundaries 0.0 and 1.0.

more transitions per tree occur. Thus, the reconstructed parameters are closer to the ones we used to generate the data. This means if the tree size is large enough to reasonably display the transition rates, our method reliably reconstructs the parameters.

For practical applications it is therefore important to record long phylogenetic trees of at least seven or eight generations instead of observing many but short trees. In the context of cancer cells, this implies a considerable experimental effort, since – on laboratory time scales – cancer cells grow slowly. They usually have a doubling time of several hours or even days. PC9 cells for example, a fast growing cancer cell line, divide approximately once per day, so it would be necessary to monitor a population for a whole week. As a comparison, bacteria divide several times per hour and a measurement time of four or five hours would be enough to record a sufficiently long tree of lifetimes. To our knowledge, long trees of cancer cell lifetimes have not been measured to date: The data that was recorded in [73] by time-lapse microscopy is only partly publicly available. The published data set consists of 163 lifetime observations, but the total measurement time was only 90 hours and single lineages were tracked for two or less generations. Therefore, our method cannot be applied to the data set.

If we had a data set with sufficiently long trees at hand, a first step would be to run our algorithm on the data to determine if two or more cellular states can be observed. If so, the next step would be to modify the distribution of lifetimes. For simplicity, we used an exponential distribution for the artificial data. In reality, cells have to pass several phases in cell cycle before they can divide again. So considering another distribution incorporating this fact, for instance the distribution of the sum of two exponentially distributed random variables, might be necessary and would lead to more reliable growth rate estimates.

5

CONCLUSIONS AND OUTLOOK

In this thesis, we studied the evolution of resistance to targeted cancer therapy from different perspectives. Within a semi-stochastic population genetics model incorporating a carrying capacity to describe confined tumor growth, we investigated how the statistics of resistant mutants depends on their fitness. We found that not only the relative birth rate but also the relative death rate determines how many resistant mutants exist in the population: if at given net growth rate more sensitive cells die per unit time (implying a larger birth rate), a higher number of resistant mutants survives on average in the population (see equation (21)). We further calculated the probability of resistance as a function of population size, mutation rate and growth rates. We showed that the probability of resistance depends on the fitness of the resistant mutants (the higher the fitness, the higher the probability of resistance at a given population size). By increasing the population size or the mutation rate, however, the fitness dependence decreases: no matter how slowly the resistant mutants grow, if only the population size or the mutation rate are large enough, resistant mutants exist in the population with probability close to one (see Fig. 10). We additionally studied how the expectation value of the number of resistant mutants changes under a splitting procedure as conducted in cell cultures, and determined how slowly growing mutants get fixed in the population under a Luria-Delbrück dynamics (see equation (45) and Fig. 14).

Our findings explain why resistance is so prevalent when treating cancer patients with targeted drugs: Resistance is often conferred by single point mutations. In realistic tumors the point mutation rate times population size is typically much greater than one, $\mu N \gg 1$, which implies a low frequency of resistant mutants existing prior to any kind of treatment. Therapy then exerts a strong selective pressure favoring resistant cells. Therefore, as long as there is no way of systematically diagnosing small tumors at early stages of the disease, targeted mono-therapy cannot be curative for macroscopic, detectable tumors.

If one aims at a therapy with curative intent, knowledge about the complete spectrum of resistance mechanisms of the respective cancer type is necessary. This

spectrum may be highly context-dependent, so cancers driven by different molecular mechanisms become resistant in different ways. We presented an experimental protocol (see Fig. 17) to systematically amplify all resistance mechanisms to a specific targeted treatment in a population of given size. Using a specific cell line as a model, we could show that a set of four compounds is enough to suppress all mechanisms (for details see section 3.2.3). In principle, one could perform the experimental protocol on different cell lines and try to identify corresponding compound combinations for those. Although the procedure is conceptually straightforward, it implies a significant experimental effort.

We further found that sleeper cells limit the long time efficacy of even the most effective treatment combination. To date, the molecular features of the sleepers only begin to be understood, and not much is known about their dynamics. We developed a statistical model to infer growth rates in heterogeneous populations from single cell lifetime measurements (Fig. 32). If applied to cancer data, our method could provide clarification to what extent the sleeper state is heritable.

The prospects of cure under targeted therapy

The long-time efficacy of targeted therapies in cancer treatment turned out to be disappointing so far. Indeed, there exist some conceptual limitations that make the treatment with targeted cancer drugs extremely challenging, even if effective compounds are available: Cancerous cells are inherent to the body and thus very similar to healthy cells. Indeed, they differ from healthy cells only in a few, specific properties. This is different from agents entirely foreign to an organism, such as bacteria or viruses. As there are not many independent mechanisms that can be attacked without harming non-cancerous tissue, tumors are naturally hard to fight.

To date, it seems as if each tumor only depends on one specific molecular driver, which complicates treatment with targeted drugs for two reasons. First, the therapy needs to be highly patient-specific. Treatment decisions have to be based on biomarkers such as point mutations in or amplifications of oncogenes, which means that samples taken from biopsies have to be sequenced to determine the genetic subtype. This implies cumbersome diagnostic procedures as well as practical limitations. Solid tumors are, for instance, spatially heterogeneous. A single biopsy does not necessarily give a representative picture of the tumor genotype.

These shortcomings can in principle be overcome by spending more time and effort.

More severely, the dependence on only one pro-proliferative signal prevents the implementation of *combination therapies targeting independent mechanisms* that effectively impede the emergence of resistance in other diseases. In HIV, for instance, combinations of at least two compounds from different drug classes are administered as first-line therapy [79]. Whereas both of the drugs alone already effectively target the viral infection, each can be overcome by point mutations that turn the virus insensitive to the treatment, preventing lasting efficacy of mono-therapy [80, 81]. Only the combinatorial delivery of drugs targeting independent mechanisms enables the long-term control of the disease. This kind of combination therapy has also been studied theoretically in the context of cancer resistance [82], and it was shown that if two such independent mechanisms exist, a combination therapy with two targeted drugs can in general lead to long-term tumor regression [83]: The possibility of a cell existing which harbors two different resistance mechanisms is very low. Unfortunately, cancer cells are not susceptible to two kinds of targeted drugs and targeting independent mechanisms, as in HIV, is impossible in cancer treatment.

As a result, the focus could be on defining sets of drugs that inhibit a molecular driver *and* all accessible resistance mechanisms. In our EGFR-driven model system, we only found resistance mechanisms which are clinically known: Alteration of the drug target (R1: EGFR T790M), activation of a downstream signal (R2: NRAS Q61R) and activation of bypass tracks (R3: HGF amplification; R4: PI3K/mTOR signaling). If this result holds in other EGFR driven cell lines, the emergence of resistance could be anticipated by delivering drugs targeting the wild type and all accessible resistance mechanism already as first-line therapy. This type of combination therapy is conceptually different from the combination treatment used in HIV or from the combinations of antibiotics delivered in bacterial infections. It may offer a way to allow targeted therapy to be less patient specific (the driver still needs to be identified, but a standardized set of resistance mechanisms could be co-treated from the start), and drastically increase the long-term efficacy of the treatment.

Immune checkpoint blockade – a promising treatment approach

In recent years immunotherapy has been established as another promising alternative to standard cancer therapy. In contrast to targeted therapies, immunotherapies do not act against tumors directly, but stimulate the anti-tumor immune response of the organism. Immune checkpoint blockade could be a treatment approach that overcomes the conceptual limitations of targeted therapy. Immune checkpoints are proteins in T-cells that act as switches to regulate the immune response and that are therefore crucial for self-tolerance [84]: If an inhibitory checkpoint protein binds to its respective ligand, the T-cell is inactivated and does not attack surrounding cells [85]. Tumors often express genes that code for inhibitory immune checkpoint ligands and thus “hide” from immune detection [86]. Immune checkpoint blockade, for instance with PD-1 (a checkpoint protein) or PD-L1 (the ligand of PD-1) inhibitors, works by targeting the checkpoint proteins on the T-cell surface (or their ligands), reactivating T-cell recognition [84]. Some patients benefit greatly from immune checkpoint inhibition, although at the moment only a subset of patients responds to this kind of therapy: large clinical trials on PD-1 and PD-L1 inhibition found response rates of 10 – 15% [87, 88]. To date, it is unclear what determines the sensitivity to immunotherapy.

T-cells do not usually attack healthy endogenous cells but only those which have been infected with viruses or bacteria. But also cancer cells can be recognized if their proteins differ sufficiently from healthy cells. Simply put, the more the protein content of a cancer cell differs from a healthy cell, the easier it is detected by T-cells. This is mirrored in the response of lung cancer patients to PD-1 blockade. Patients who regularly smoked before diagnosis – meaning that their cancers typically had more mutations (since tobacco smoke is carcinogenic) – on average respond better than non-smokers [89]. This implies a completely different situation compared to targeted therapies. The latter act on very specific genetic alterations, whereas immune checkpoint blockade can target many different mutations at a time. Therefore, it is much more robust against genetic heterogeneity in large populations which limits the long-time efficacy of targeted therapies.

Immunotherapy also is very interesting from a modeling perspective. The interaction with the immune system adds an additional layer of complexity to the process of tumor evolution. This makes the dynamics harder to model and predict. Probably not all tumors are in principle susceptible to immunotherapy. If one were able to foresee which patients respond to immunotherapy, treatment efficacy

could be drastically improved. A first approach towards this goal is based on a fitness model for neoantigens and correctly predicts the response to checkpoint blockade immunotherapy in three patient cohorts [47]. Future work in this direction will hopefully lead to a more conceptual understanding of the requirements for successful immunotherapy.

GLOSSARY

ADENOCARCINOMA Type of cancer started by glandular cells lining some organs. Adenocarcinomas are often found in the breast, lung, prostate, or colon.

(GENE) AMPLIFICATION An increased copy number of a gene in the genome compared to a normal diploid organism. The amplification of a gene can lead to more RNA and thus a higher production of proteins made from the gene (see also overexpression).

APOPTOSIS Cellular program of controlled cell death which can be triggered by extra-cellular stimuli (also called cell suicide). It is crucial for tissue organization and renewal: unwanted, DNA damaged, or superfluous cells can be eliminated in a highly ordered physiological process that does not harm surrounding cells.

CELLULAR RECEPTOR Cell surface protein that mediates the communication of a cell with its environment. A signal is transmitted by binding to specific extracellular molecules (e.g. hormones, growth factors, or nutrients). Binding activates the receptor, for instance by dimerization with another receptor or by a conformational change. This in turn changes how the intracellular part of the receptor interacts with other proteins within the cell and triggers a chemical signaling cascade. Depending on which cellular receptor binds to which ligand, different physiological outcomes occurs, for instance apoptosis, adhesion, differentiation, proliferation, or migration.

CHROMOSOMAL REARRANGEMENT Genetic alteration that changes the structure of a native chromosome. Chromosomal rearrangements are usually caused by DNA double strand breaks and consecutive rejoining of the broken strands in an abnormal way. By that, the usual order in which genes are arranged on a chromosome is disrupted. Chromosomal rearrangements can involve deletions, duplications, inversions, and translocations.

CRYSTAL VIOLET Bright violet dye that binds to proteins and DNA. It is for example used to stain cells grown in cell culture to make them easily visible.

ENU (N-ETHYL-N-NITROSOUREA) Chemical mutagen that increases the mutation rate. It causes heritable mutations by transferring ethyl groups to nitrogen or oxygen sites on each of the four deoxyribonucleic acids. During the next cell division, these additional ethyl groups lead to a misreading of the bases resulting in mutations that get fixed in the next generations.

GROWTH FACTOR Protein triggering cell growth and proliferation by binding to a cellular receptor.

(PROTEIN) KINASE An enzyme modifying proteins by catalyzing the transfer of a phosphate group from a donor (such as ADP) to a substrate. In cell signaling protein kinases act as molecular switches: the addition of a phosphate group often leads to a conformational change activating the protein.

MAPK PATHWAY Example of a signaling circuit that transmits a signal from a cell surface receptor to the cell nucleus. It consists of a chain of intracellular proteins that, upon activation of a receptor such as EGFR, gradually induces kinase phosphorylation (RAS activation → RAF activation → MEK activation → ERK activation). Phosphorylated ERK activates transcription factors that are important for cell growth, proliferation, and survival.

METASTASIS A tumor consisting of cells which derive from a malignant growth somewhere else in the body. For example, if a lung cancer spreads to the liver, the cancerous cells there (forming a metastasis) are lung cells and not liver cells.

MOLECULAR DRIVER Specific genetic alteration that is necessary for the survival and the proliferation of a cancer cell. Drivers can be point mutations in genes that code for cellular receptors or other components of the cellular signaling network, but also amplifications or overexpressions of genes controlling proliferation.

NEOPLASM (NEOPLASTIC TISSUE) Tissue that grows abnormally and excessively. Depending on their behavior neoplasms are grouped in different classes. Malignant neoplasms that invade and destroy surrounding tissue are called tumors.

ONCOGENE A mutated form of a normal gene that induces cancer. Genes that upon mutation can turn to oncogenes are called proto-oncogenes.

ONCOVIRUS A virus that can cause cancer by introducing genetic material to the DNA or RNA of a host. Some well-known representatives are the human papillomavirus (increasing the risk of cervical cancer), the Epstein-Barr virus (inducing, for instance, Hodgkin's lymphoma or gastric cancer), or the hepatitis C virus (inducing liver cancer).

OVEREXPRESSION Refers to a situation where more RNA or proteins are made from a specific gene in a cell than normal. Overexpression may be caused by a genetic amplification (which means that more copies are transcribed) or by deregulation of genes that control transcription.

PASSAGE One cycle of seeding, expanding, and (when the population has filled the culture vessel) splitting a cell culture.

PHOSPHORYLATION Covalent binding of a phosphate group to an organic compound such as a protein. A prominent example is the phosphorylation of adenosine diphosphate (ADP) to adenosine triphosphate (ATP), a process of storing and transferring energy as used by all living cells. Phosphorylation often activates a protein, a process that is crucial in cell signaling.

SOMATIC MUTATION Mutation that occurs in the genome outside the germ line and that is acquired during the lifetime of an organism. By definition, somatic mutations are not passed on to the next generation.

TUMOR SUPPRESSOR A protein that negatively regulates the cell cycle or promotes apoptosis, thus constraining cell proliferation. A mutation or deletion in a gene coding for a tumor suppressor affecting its function is a crucial step in tumorigenesis.

TUMORIGENESIS The multi-step process of tumor formation involving many independent genetic alterations (see 'The hallmarks of cancer', section 1.1.2).

WESTERN BLOT Experimental method to examine the protein levels in cells. In brief, it works as follows: Cells are seeded in different culture conditions (for instance with and without treatment). After a defined treatment time, cell lysis is induced by a lysis buffer which breaks up the membrane structure of the cells. The solution containing the cell content is transferred to an agarose gel. By employing an electrical field, the molecules are separated by size and subsequently, by applying an electrical field in the orthogonal direction, transferred to a membrane. With the help of antibodies that bind specifically to the protein of interest and which are fluorescently labeled, the amount of protein can be visualized. Importantly, one uses additional antibodies, that only bind to the phosphorylated form (meaning the active form) of the protein of interest, so that one can not only monitor the total protein content (for example total EGFR or tEGFR) but also if the respective protein is in its active state (for example phosphorylated EGFR or pEGFR). The name *western blot* alludes to Edwin Southern, who invented the blotting technique for DNA fragments – a method he called southern blot. Subsequent variations of the protocol to detect RNA and proteins were called northern blot and western blot, respectively.

WHOLE EXOME SEQUENCING (WES) Technique to sequence the protein-coding part of the genome. In humans only about 1% of the genome codes for proteins (according to approximately 30 million base pairs), which is why WES is much cheaper compared to sequencing a whole genome. WES consists of two consecutive steps: At first the exonic parts of the genome are selectively enriched. After that the exonic DNA is sequenced, usually using some high-throughput sequencing technology.

XENOGRAFT Tissue taken from a donor of one species that is transferred to a recipient of another species. In cancer research human tumor cells are often introduced into mice, for instance to study the in-vivo response of a cell line to a drug. These mice are then called xenograft mouse models.

BIBLIOGRAPHY

- [1] R. A. Weinberg. *The Biology of Cancer*. Second Edition. Garland Science, 2014.
- [2] B. Alberts et al. *Essential Cell Biology: An Introduction to the Molecular Biology of the Cell*. Garland Publisher, 1998.
- [3] J. Ngeow and C. Eng. "Precision medicine in heritable cancer: when somatic tumour testing and germline mutations meet." In: *NPJ genomic medicine* 1.1 (Nov. 2016), p. 15006.
- [4] S. S. Hecht. "Tobacco Smoke Carcinogens and Lung Cancer". In: *JNCI Journal of the National Cancer Institute* 91.14 (July 1999), pp. 1194–1210.
- [5] F. R. de Gruijl. "Skin cancer and solar UV radiation." In: *European journal of cancer (Oxford, England : 1990)* 35.14 (Dec. 1999), pp. 2003–9.
- [6] E. A. Mesri, M. A. Feitelson, and K. Munger. "Human viral oncogenesis: a cancer hallmarks analysis." In: *Cell host & microbe* 15.3 (Mar. 2014), pp. 266–82.
- [7] D. Hanahan and R. A. Weinberg. "Hallmarks of cancer: the next generation." In: *Cell* 144.5 (Mar. 2011), pp. 646–74.
- [8] D. Hanahan and R. A. Weinberg. "The Hallmarks of Cancer". In: *Cell* 100.1 (Jan. 2000), pp. 57–70.
- [9] O. Warburg and S. Minami. "Versuche an überlebendem Carcinomgewebe". In: *Klinische Wochenschrift* 2.17 (1923), pp. 776–777.
- [10] M. G. Vander Heiden, L. C. Cantley, and C. B. Thompson. "Understanding the Warburg effect: the metabolic requirements of cell proliferation." In: *Science (New York, N.Y.)* 324.5930 (May 2009), pp. 1029–33.
- [11] M. W. L. Teng et al. "Immune-mediated dormancy: an equilibrium with cancer." In: *Journal of leukocyte biology* 84.4 (Oct. 2008), pp. 988–93.
- [12] S. Negrini, V. G. Gorgoulis, and T. D. Halazonetis. "Genomic instability — an evolving hallmark of cancer". In: *Nature Reviews Molecular Cell Biology* 11.3 (Mar. 2010), pp. 220–228.

- [13] D. G. DeNardo, P. Andreu, and L. M. Coussens. "Interactions between lymphocytes and myeloid cells regulate pro- versus anti-tumor immunity". In: *Cancer and Metastasis Reviews* 29.2 (June 2010), pp. 309–316.
- [14] S. I. Grivnenkov, F. R. Greten, and M. Karin. "Immunity, Inflammation, and Cancer". In: *Cell* 140.6 (Mar. 2010), pp. 883–899.
- [15] Y. Yarden and M. X. Sliwkowski. "Untangling the ErbB signalling network". In: *Nature Reviews Molecular Cell Biology* 2.2 (Feb. 2001), pp. 127–137.
- [16] D. S. Salomon et al. "Epidermal growth factor-related peptides and their receptors in human malignancies". In: *Critical Reviews in Oncology/Hematology* 19.3 (July 1995), pp. 183–232.
- [17] A. F. Gazdar et al. "Mutations and addiction to EGFR: the Achilles 'heel' of lung cancers?" In: *Trends in Molecular Medicine* 10.10 (Oct. 2004), pp. 481–486.
- [18] D. R. Emlet et al. "Targeting a Glioblastoma Cancer Stem-Cell Population Defined by EGF Receptor Variant III". In: *Cancer Research* 74.4 (Feb. 2014), pp. 1238–1249.
- [19] P. P. Connell and S. Hellman. "Advances in Radiotherapy and Implications for the Next Century: A Historical Perspective". In: *Cancer Research* 69.2 (Jan. 2009), pp. 383–392.
- [20] W. C. Röntgen. "Ueber eine neue Art von Strahlen (vorläufige Mitteilung)". In: *Aus den Sitzungsberichten der Würzburger Physik.-medic. Gesellschaft Würzburg* (1895), pp. 137–147.
- [21] P. Curie, M. P. Curie, and G. Bémont. "Sur une nouvelle substance fortement radio-active, contenue dans la pechblende". In: *Comptes Rendus Hebdomadaires des Séances de l'Académie des Sciences* 127 (1898), pp. 1215–1217.
- [22] V. T. DeVita and E. Chu. "A History of Cancer Chemotherapy". In: *Cancer Research* 68.21 (Nov. 2008), pp. 8643–8653.
- [23] A. E. Kayl and C. A. Meyers. "Side-effects of chemotherapy and quality of life in ovarian and breast cancer patients". In: *Current Opinion in Obstetrics and Gynecology* 18.1 (Feb. 2006), pp. 24–28.
- [24] I. B. Weinstein et al. "Disorders in cell circuitry associated with multistage carcinogenesis: exploitable targets for cancer prevention and therapy." In: *Clinical cancer research : an official journal of the American Association for Cancer Research* 3.12 Pt 2 (Dec. 1997), pp. 2696–702.

- [25] I. B. Weinstein. "Addiction to oncogenes—the Achilles heel of cancer." In: *Science (New York, N.Y.)* 297.5578 (July 2002), pp. 63–4.
- [26] S. V. Sharma and J. Settleman. "Oncogene addiction: setting the stage for molecularly targeted cancer therapy". In: *Genes & Development* 21.24 (Dec. 2007), pp. 3214–3231.
- [27] M. Huang et al. "Molecularly targeted cancer therapy: some lessons from the past decade". In: *Trends in Pharmacological Sciences* 35.1 (Jan. 2014), pp. 41–50.
- [28] L. A. Pikor et al. "Genetic alterations defining NSCLC subtypes and their therapeutic implications". In: *Lung Cancer* 82.2 (Nov. 2013), pp. 179–189.
- [29] P. C. Nowell. "Discovery of the Philadelphia chromosome: a personal perspective". In: *Journal of Clinical Investigation* 117.8 (Aug. 2007), pp. 2033–2035.
- [30] P. C. Nowell and D. A. Hungerford. "Chromosome studies on normal and leukemic human leukocytes." In: *Journal of the National Cancer Institute* 25. July (July 1960), pp. 85–109.
- [31] J. D. Rowley. "A New Consistent Chromosomal Abnormality in Chronic Myelogenous Leukaemia identified by Quinacrine Fluorescence and Giemsa Staining". In: *Nature* 243.5405 (June 1973), pp. 290–293.
- [32] N. Heisterkamp et al. "Localization of the c-abl oncogene adjacent to a translocation break point in chronic myelocytic leukaemia". In: *Nature* 306.5940 (Nov. 1983), pp. 239–242.
- [33] J. Groffen et al. "Philadelphia chromosomal breakpoints are clustered within a limited region, bcr, on chromosome 22." In: *Cell* 36.1 (Jan. 1984), pp. 93–9.
- [34] T. G. Lugo et al. "Tyrosine kinase activity and transformation potency of bcr-abl oncogene products". In: *Science* 247.4946 (Mar. 1990), pp. 1079–1082.
- [35] B. J. Druker et al. "Effects of a selective inhibitor of the Abl tyrosine kinase on the growth of Bcr-Abl positive cells." In: *Nature medicine* 2.5 (May 1996), pp. 561–6.
- [36] B. J. Druker et al. "Efficacy and Safety of a Specific Inhibitor of the BCR-ABL Tyrosine Kinase in Chronic Myeloid Leukemia". In: *New England Journal of Medicine* 344.14 (Apr. 2001), pp. 1031–1037.
- [37] B. J. Druker et al. "Five-Year Follow-up of Patients Receiving Imatinib for Chronic Myeloid Leukemia". In: *New England Journal of Medicine* 355.23 (Dec. 2006), pp. 2408–2417.

- [38] J. G. Paez et al. "EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy." In: *Science (New York, N.Y.)* 304.5676 (June 2004), pp. 1497–500.
- [39] R. Sordella et al. "Gefitinib-sensitizing EGFR mutations in lung cancer activate anti-apoptotic pathways." In: *Science (New York, N.Y.)* 305.5687 (Aug. 2004), pp. 1163–7.
- [40] L. M. Ellis and D. J. Hicklin. "Resistance to Targeted Therapies: Refining Anticancer Therapy in the Era of Molecular Oncology." In: *Clinical cancer research : an official journal of the American Association for Cancer Research* 15.24 (Dec. 2009), pp. 7471–7478.
- [41] D. R. Camidge, W. Pao, and L. V. Sequist. "Acquired resistance to TKIs in solid tumours: learning from lung cancer." In: *Nature reviews. Clinical oncology* 11.8 (Aug. 2014), pp. 473–81.
- [42] S. Suresh. "Biomechanics and biophysics of cancer cells." In: *Acta biomaterialia* 3.4 (July 2007), pp. 413–38.
- [43] D. Wirtz, K. Konstantopoulos, and P. C. Searson. "The physics of cancer: the role of physical interactions and mechanical forces in metastasis." In: *Nature reviews. Cancer* 11.7 (June 2011), pp. 512–22.
- [44] L. Ein-Dor, O. Zuk, and E. Domany. "Thousands of samples are needed to generate a robust gene list for predicting outcome in cancer." In: *Proceedings of the National Academy of Sciences of the United States of America* 103.15 (Apr. 2006), pp. 5923–8.
- [45] Y. Drier, M. Sheffer, and E. Domany. "Pathway-based personalized analysis of cancer." In: *Proceedings of the National Academy of Sciences of the United States of America* 110.16 (Apr. 2013), pp. 6388–93.
- [46] B. Waclaw et al. "Spatial model predicts dispersal and cell turnover cause reduced intra-tumor heterogeneity". In: (Mar. 2015), pp. 1–31.
- [47] M. Łuksza et al. "A neoantigen fitness model predicts tumour response to checkpoint blockade immunotherapy." In: *Nature* 551.7681 (2017), pp. 517–520.
- [48] L. A. Diaz Jr et al. "The molecular evolution of acquired resistance to targeted EGFR blockade in colorectal cancers." In: *Nature* 486.7404 (June 2012), pp. 537–40.

- [49] L. Hayflick and P. Moorhead. "The serial cultivation of human diploid cell strains". In: *Experimental Cell Research* 25.3 (Dec. 1961), pp. 585–621.
- [50] S. E. Luria and M. Delbrück. "Mutations of Bacteria from Virus Sensitivity to Virus Resistance." In: *Genetics* 28.6 (Nov. 1943), pp. 491–511.
- [51] D. E. Lea and C. A. Coulson. "The distribution of the numbers of mutants in bacterial populations." In: *Journal of genetics* 49.3 (Dec. 1949), pp. 264–85.
- [52] P. Armitage. "The Statistical Theory of Bacterial Populations Subject to Mutation". In: *Journal of the Royal Statistical Society. Series B (Methodological)* 14.1 (1952), pp. 1–40.
- [53] Q. Zheng. "Progress of a half century in the study of the Luria–Delbrück distribution". In: *Mathematical Biosciences* 162.1-2 (Nov. 1999), pp. 1–32.
- [54] A. J. Coldman and J. H. Goldie. "A stochastic model for the origin and treatment of tumors containing drug-resistant cells". In: *Bulletin of Mathematical Biology* 48.3-4 (May 1986), pp. 279–292.
- [55] Y. Iwasa, M. A. Nowak, and F. Michor. "Evolution of resistance during clonal expansion." In: *Genetics* 172.4 (Apr. 2006), pp. 2557–66.
- [56] A. Dewanji, E. G. Luebeck, and S. H. Moolgavkar. "A generalized Luria-Delbrück model." In: *Mathematical biosciences* 197.2 (Oct. 2005), pp. 140–52.
- [57] C. Tomasetti. "On the probability of random genetic mutations for various types of tumor growth." In: *Bulletin of mathematical biology* 74.6 (June 2012), pp. 1379–95.
- [58] J. Foo and F. Michor. "Evolution of acquired resistance to anti-cancer therapy." In: *Journal of theoretical biology* 355 (Aug. 2014), pp. 10–20.
- [59] C. Gambacorti-Passerini et al. "Multicenter independent assessment of outcomes in chronic myeloid leukemia patients treated with imatinib." In: *Journal of the National Cancer Institute* 103.7 (2011), pp. 553–61.
- [60] J. H. Bielas and L. A. Loeb. "Quantification of random genomic mutations." In: *Nature methods* 2.4 (Apr. 2005), pp. 285–90.
- [61] U. Del Monte. "Does the cell number 10^9 still really fit one gram of tumor tissue?" In: *Cell Cycle* 8.March 2015 (2009), pp. 505–506.
- [62] F. Yang et al. "Relationship between tumor size and disease stage in non-small cell lung cancer." In: *BMC cancer* 10.1 (Sept. 2010), p. 474.

- [63] S. Kobayashi et al. "EGFR mutation and resistance of non-small-cell lung cancer to gefitinib." In: *The New England journal of medicine* 352.8 (Feb. 2005), pp. 786–92.
- [64] W. Pao et al. "Acquired resistance of lung adenocarcinomas to gefitinib or erlotinib is associated with a second mutation in the EGFR kinase domain." In: *PLoS medicine* 2.3 (Mar. 2005), e73.
- [65] A. N. Hata et al. "Tumor cells can follow distinct evolutionary paths to become resistant to epidermal growth factor receptor inhibition." In: *Nature medicine* 22.3 (Mar. 2016), pp. 262–9.
- [66] C.-H. Yun et al. "The T790M mutation in EGFR kinase causes drug resistance by increasing the affinity for ATP". In: *Proceedings of the National Academy of Sciences* 105.6 (Feb. 2008), pp. 2070–2075.
- [67] D. A. E. Cross et al. "AZD9291, an irreversible EGFR TKI, overcomes T790M-mediated resistance to EGFR inhibitors in lung cancer." In: *Cancer discovery* 4.9 (Sept. 2014), pp. 1046–61.
- [68] T. R. Wilson et al. "Widespread potential for growth-factor-driven resistance to anticancer kinase inhibitors." In: *Nature* 487.7408 (July 2012), pp. 505–9.
- [69] S. V. Sharma et al. "A chromatin-mediated reversible drug-tolerant state in cancer cell subpopulations." In: *Cell* 141.1 (Apr. 2010), pp. 69–80.
- [70] M. J. Hangauer et al. "Drug-tolerant persister cancer cells are vulnerable to GPX4 inhibition." In: *Nature* 551.7679 (2017), pp. 247–250.
- [71] R. Rosell et al. "Erlotinib versus standard chemotherapy as first-line treatment for European patients with advanced EGFR mutation-positive non-small-cell lung cancer (EURTAC): a multicentre, open-label, randomised phase 3 trial." In: *The Lancet. Oncology* 13.3 (Mar. 2012), pp. 239–46.
- [72] M. Ramirez et al. "Diverse drug-resistance mechanisms can emerge from drug-tolerant cancer persister cells." In: *Nature communications* 7 (Feb. 2016), p. 10690.
- [73] D. R. Tyson et al. "Fractional proliferation: a method to deconvolve cell population dynamics from single-cell data." In: *Nature methods* 9.9 (Sept. 2012), pp. 923–8.
- [74] C. Biben et al. "Murine cerberus homologue mCer-1: a candidate anterior patterning molecule." In: *Developmental biology* 194.2 (Feb. 1998), pp. 135–51.

- [75] C. M. Bishop. *Pattern recognition and machine learning*. Springer, 2006.
- [76] T. M. Cover and J. A. Thomas. *Elements of information theory*. Second Edition. Wiley, 2006.
- [77] J. S. Yedidia, W. T. Freeman, and Y. Weiss. "Understanding belief propagation and its generalizations". In: *Exploring artificial intelligence in the New Millenium*. Morgan Kaufmann Publishers, 2003. Chap. 8, pp. 239–270.
- [78] M. Mézard and A. Montanari. *Information, Physics, and Computation*. Oxford University Press, 2009.
- [79] C. Bock and T. Lengauer. "Managing drug resistance in cancer: lessons from HIV therapy." In: *Nature reviews. Cancer* 12.7 (July 2012), pp. 494–501.
- [80] D. E. Goldberg, R. F. Siliciano, and W. R. Jacobs. "Outwitting evolution: fighting drug-resistant TB, malaria, and HIV." In: *Cell* 148.6 (Mar. 2012), pp. 1271–83.
- [81] R. M. Ribeiro and S. Bonhoeffer. "Production of resistant HIV mutants during antiretroviral therapy." In: *Proceedings of the National Academy of Sciences of the United States of America* 97.14 (July 2000), pp. 7681–6.
- [82] A. Coldman and J. Goldie. "A model for the resistance of tumor cells to cancer chemotherapeutic agents". In: *Mathematical Biosciences* 65.2 (Aug. 1983), pp. 291–307.
- [83] I. Bozic et al. "Evolutionary dynamics of cancer in response to targeted combination therapy." In: *eLife* 2 (June 2013), e00747.
- [84] D. M. Pardoll. "The blockade of immune checkpoints in cancer immunotherapy." In: *Nature reviews. Cancer* 12.4 (Mar. 2012), pp. 252–64.
- [85] G. J. Freeman et al. "Engagement of the Pd-1 Immunoinhibitory Receptor by a Novel B7 Family Member Leads to Negative Regulation of Lymphocyte Activation". In: *The Journal of Experimental Medicine* 192.7 (2000), pp. 1027–1034.
- [86] H. Dong et al. "Tumor-associated B7-H1 promotes T-cell apoptosis: A potential mechanism of immune evasion". In: *Nature Medicine* 8.8 (2002), pp. 793–800.
- [87] S. L. Topalian et al. "Safety, Activity, and Immune Correlates of Anti-PD-1 Antibody in Cancer". In: *New England Journal of Medicine* 366.26 (June 2012), pp. 2443–2454.

- [88] J. R. Brahmer et al. "Safety and activity of anti-PD-L1 antibody in patients with advanced cancer". In: *New England Journal of Medicine* 366.26 (2012), pp. 2455–2465.
- [89] B. Li, X. Huang, and L. Fu. "Impact of smoking on efficacy of PD-1/PD-L1 inhibitors in non-small cell lung cancer patients: a meta-analysis." In: *OncoTargets and therapy* 11 (June 2018), pp. 3691–3696.

LIST OF FIGURES

Figure 1	Sketch of an arising tumor	2
Figure 2	The hallmarks of cancer	4
Figure 3	ERBB signaling network	8
Figure 4	The Philadelphia chromosome	11
Figure 5	Portrait of Henrietta Lacks	17
Figure 6	Microscopic cells and macroscopic colonies	21
Figure 7	Growth dynamics of cancer cell in tumors and in cell culture	23
Figure 8	Growth dynamics of the sensitive cells for different carrying capacities	26
Figure 9	Mean and variance of the number of resistance cells	30
Figure 10	Probability of resistance.	31
Figure 11	Distribution of the number of resistant colonies	36
Figure 12	Mean number of resistant colonies	37
Figure 13	Mean fraction under competition splitting dynamics	41
Figure 14	Mean fraction under LD splitting dynamics	44
Figure 15	Muller plot of tumor evolution	47
Figure 16	Probability of resistance as a function of tumor size	48
Figure 17	Experimental protocol	49
Figure 18	Response of the parental and the resistant cell lines to the drug panel	52
Figure 19	Growth curves of the respective cell lines	54
Figure 20	Stability of the resistant lines after compound withdrawal .	55
Figure 21	Secondary mutations in the MAPK pathway cause resis- tance in R1 and R2	56
Figure 22	An amplification in HGF causes resistance in R3	57
Figure 23	PI3K/mTOR activation drives resistance in R4	58
Figure 24	Colony formation under mono- and combination-therapy .	59
Figure 25	Quantification of sleepers that survive different drug com- binations	60

Figure 26	Cluster analysis of gene expression of sleepers, cells emerging from sleepers and parental cells	61
Figure 27	Colony formation from sleeper cells after treatment stop . .	63
Figure 28	Exponential decay of the number of surviving sleeper cells .	63
Figure 29	Single cell lifetimes on a hidden Markov tree	69
Figure 30	Graphical explanation of the EM algorithm	73
Figure 31	Sketch of message flow in and out of node i	77
Figure 32	Inferred mean parameters for different nodes N	78

ACKNOWLEDGEMENTS

I am deeply grateful to my supervisor, Johannes Berg, not only for introducing me to the exciting topic of cancer resistance but also for his continuous encouragement, trust, guidance, and advice. His unwavering enthusiasm helped me through the frustrating times when experiments failed, often repeatedly, and I nearly lost faith in the project. His kindness and sense of humor shaped the atmosphere in our group and always created a pleasant working environment. All of this made my time as a PhD student a very rewarding and memorable experience.

I thank Roman Thomas and Martin Sos for giving me the opportunity to pursue my experimental project. A special thanks goes to all the people working in their labs who very patiently taught me the basics of molecular biology and experimental methods (I actually had no idea what I was doing). In this context I am particularly thankful to Carina Lorenz and Johannes Brägelmann for their time, effort, and help.

Sincere thanks to my beloved friend Vera Schaubach who illustrated the beautiful cover of this thesis, and who has been a huge support in all situations during my PhD.

Over the last years, I spent hundreds of enjoyable lunch breaks discussing scientific and non-scientific topics. Many thanks for that to my dear colleagues Arman, Chau, Christa, Daniel, Denis, Lucy, Mara, Nandita, Nico, Niklas, Prasanna, Simon, Simone, Stephan, Tom, Torsten, Ulli, and Yijing. Special thanks to Stephan for helping me with various coding problems. Also many thanks to Fabio, Fanny and Hanno for all the cheerful coffee breaks during the first half of my PhD.

Thanks and love to my family and friends who always supported me and listened to me, no matter if I was complaining about problems or enthusiastically sharing new results.

Finally, thanks to Nils – for everything in the past, the present and the future.

ERKLÄRUNG

Ich versichere, dass ich die von mir vorgelegte Dissertation selbständig angefertigt, die benutzten Quellen und Hilfsmittel vollständig angegeben und die Stellen der Arbeit - einschließlich Tabellen, Karten und Abbildungen -, die anderen Werken im Wortlaut oder dem Sinn nach entnommen sind, in jedem Einzelfall als Entlehnung kenntlich gemacht habe; dass diese Dissertation noch keiner anderen Fakultät oder Universität zur Prüfung vorgelegen hat; dass sie - abgesehen von unten angegebenen Teilpublikationen - noch nicht veröffentlicht worden ist, sowie, dass ich eine solche Veröffentlichung vor Abschluss des Promotionsverfahrens nicht vornehmen werde. Die Bestimmungen der Promotionsordnung sind mir bekannt. Die von mir vorgelegte Dissertation ist von Prof. Dr. Johannes Berg betreut worden.

Köln, Oktober 2018

Nina Müller