Evolutionary dynamics of cancer

A thesis presented

by

Franziska Michor

to

The Department of Organismic and Evolutionary Biology

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Organismic and Evolutionary Biology

Harvard University

Cambridge, Massachusetts

May 2005

Thesis Advisor: Professor Martin A. Nowak Franziska Michor

Evolutionary dynamics of cancer

Evolutionary dynamics of mutation and selection can be formulated by mathematical equations. Cancer arises as a consequence of somatic evolution. Therefore, a mathematical approach can be used to understand cancer initiation and progression. Tumorigenesis is driven by genetic alterations of oncogenes, tumor suppressor genes and genetic instability genes. First, I discuss the fundamental principles that govern the dynamics of activating oncogenes and inactivating tumor suppressor genes in populations of reproducing cells. A quantitative theory of mutation and selection provides insights into the role of genetic instability in cancer initiation. Second, I analyze the dynamics of cancer progression via inactivation of one or two tumor suppressor genes. Third, I derive a mathematical framework for the complete mutational sequence of colorectal tumorigenesis (as known today) and the effect of chromosomal instability therein. Fourth, I present a new mathematical model of colon cancer initiation that assumes a linear flow from stem cells to differentiated cells to apoptosis. Finally, a mathematical model of chronic myeloid leukemia (CML) and the molecular response to imatinib therapy suggests that leukemic stem cells are spared by therapy. The model provides quantitative insights into the in vivo kinetics of this cancer and determines the probability of disease relapse due to resistance mutations.

Contents

Acknowledgments

Thanks to my committee members David Haig, Andrew Murray, Martin Nowak, and John Wakeley for their advice and support. Thanks to the Department of Organismic and Evolutionary Biology and the Faculty of Arts and Sciences of Harvard University. I thank my collaborators Yoh Iwasa, Christoph Lengauer, Bert Vogelstein, Robert May, Charles Sawyers, Tim Hughes, Neil Shah, Sue Branford, Steven Frank, Natalia Komarova, Harith Rajagopalan, and Christoph Hauert for amazing times. Thanks to the members of the Program for Evolutionary Dynamics at Harvard and the Program for Theoretical Biology at the Institute for Advanced Study; they made my years in graduate school an unforgettable experience. Finally, I thank my parents Elli and Peter, my sister Johanna, and also Mirra & Gosi for making me what I am.

1 Introduction

This thesis is about the somatic evolutionary dynamics of cancer. Cancer results from an accumulation of genetic alterations in somatic cells. This accumulation of mutations leads to a breakdown of cooperation between cells in a tissue, such that single cells revert to selfish behavior reminiscent of unicellular organisms. Mutation and selection can best be described when represented as mathematical equations. Hence, mathematical frameworks can be used to analyze the evolutionary dynamics of cancer.

Cancer susceptibility genes can be subdivided into three categories[1]: (i) gatekeepers, such as oncogenes and tumor suppressor genes, control growth and differentiation pathways of the cell - alterations of such genes lead to increased growth rates or decreased death rates as compared with neighboring cells; (ii) caretakers maintain the genomic integrity of the cell - alterations of such genes increase the rate of accumulating further mutations and cause genetic instabilities; and (iii) landscapers regulate the cell's interaction with the surrounding microenvironment - alterations of such genes contribute to the neoplastic transformation of the cell. Cancer results from an accumulation of mutations in gatekeeper genes; at the time of diagnosis, most cancers are genetically unstable. A major question and controversy in cancer biology is whether genetic instability arises early and hence is a driving force of tumorigenesis.

In Chapter 2, I present a mathematical framework designed to study the kinetics

of mutations in cancer-susceptibility genes and apply this framework to cancer initiation and progression. I discuss the evolutionary dynamics of mutations activating oncogenes and inactivating tumor suppressor genes in populations of reproducing cells. The probability of activating an oncogene in a cellular population depends on the tissue organization and population size: a population fed by a small number of stem cells can dramatically reduce the risk of accumulating mutations in oncogenes as compared to well-mixed cellular compartments. The analysis yields three different kinetic laws for the probability of inactivating a tumor suppressor gene in a population of cells: in dependence of the population size as compared to the mutation rates inactivating the two alleles, the tumor suppressor gene is homozygously inactivated in two, one, or zero rate-limiting hits. Later, this framework is extended to include mutations causing genetic instability and to analyze the role of genetic instability in human cancers. I discuss the conditions under which genetic instability is likely to initiate tumorigenesis.

In Chapter 3, the mathematical framework of tumor suppressor gene inactivation and genetic instability is extended to situations in which two tumor suppressor genes need to be inactivated to drive tumorigenesis. Once the first tumor suppressor gene has been inactivated in a cell, this cell might give rise to clonal expansion. In a moderately large population of cells, it might take only one rate-limiting hit to inactivate the second tumor suppressor gene. In this situation, genetic instability is very likely to emerge before inactivation of the first tumor suppressor gene. If clonal

expansion after inactivation of the first gene gives rise to a very large population of cells, however, then a further genetic alteration is not rate-limiting any longer and genetic instability becomes less important.

In Chapter 4, I analyze the evolutionary dynamics of the complete mutational sequence of colorectal tumorigenesis as known today. A literature review of mathematical models of colorectal tumorigenesis is provided, and the specific mutations found in colorectal cancer cell lines are outlined. Then I discuss the dynamics of colorectal tumorigenesis in well-mixed populations of (stem) cells and determine the role and importance of genetic instability in this setting.

In Chapter 5, I design a new mathematical model of colorectal tumorigenesis based on the cellular dynamics and tissue architecture of a colonic crypt. A stem cell at the base of the crypt produces a cellular lineage that populates the crypt. The differentiated cells are pushed to the top of the crypt by the continuous production and migration of such cells and undergo apoptosis on top of the crypt. All of these cells could in principle accumulate mutations that initiate tumorigenesis. I analyze the mutational dynamics of tumor suppressor genes and genetic instability genes in such a linear model and determine the mutations likely to initiate tumorigenesis.

In Chapter 6, I provide an analysis of the molecular response to imatinib (Gleevec), a chemotherapeutic agent used to treat chronic myeloid leukemia (CML), in a 169 patient dataset. This dataset establishes that imatinib leads to a bi-phasic decline of the leukemic cell burden and an overall 5000-fold decrease in the leukemic cell mass.

Upon discontinuation of therapy, however, the leukemic burden rapidly rebounds to values beyond pre-treatment baseline, suggesting that leukemic stem cells are not depleted during imatinib therapy. A mathematical model recovers all features of the in vivo kinetics and can be used to analyze the disease kinetics and parameter values. Acquired resistance mutations are similarly modeled. The probability of resistance and time to treatment failure can be estimated. This model provides the first quantitative analysis of a human cancer in vivo.

2 Dynamics of cancer progression

Michor F, Iwasa Y, Nowak MA

Dynamics of cancer progression[2]

Nature Reviews Cancer 4, 197-206 (2004)

Cancer arises from an accumulation of mutations in oncogenes, tumor suppressor genes and genes that maintain the genomic integrity of the cell. Oncogenes lead to increased net growth rates of the cell when activated by a point mutation, amplified or overexpressed. We study the evolutionary dynamics of oncogene activation in populations of reproducing cells. Tumor suppressor genes lead to increased net growth rates of the cell when inactivated in both alleles. We analyze how the kinetics of tumor suppressor gene inactivation depends on the population size of cells and the mutation rates for the first and second hit. We find three different kinetic laws: in small, intermediate and large populations, it takes, respectively, two, one and zero rate limiting steps to inactivate a tumor suppressor gene. Genetic instability refers to an increased mutation rate across the genome and arises due to mutations in genetic instability genes. A major question in cancer biology is whether genetic instability is an early event in tumorigenesis. We analyze the role of genetic instability in tumor initiation and progression and determine the

conditions under which genetic instability represents the first phenotypic alteration on the road to cancer.

2.1 Introduction

Cancer is a genetic disease[1]. Although environmental and other nongenetic factors have roles in many stages of tumorigenesis, it is widely accepted that cancer arises due to mutations in cancer-susceptibility genes. These genes belong to one of three classes[1,3]: gatekeepers, caretakers and landscapers. Gatekeepers directly regulate growth and differentiation pathways of the cell and comprise oncogenes and tumor suppressor genes. Caretakers, in contrast, promote tumorigenesis indirectly[4,5]. They function in maintaining the genomic integrity of the cell. Mutation of caretakers can lead to genetic instability, and the cell rapidly accumulates changes in other genes that directly control cell birth and death. Landscaper defects do not directly affect cellular growth, but generate an abnormal stromal environment that contributes to the neoplastic transformation of cells[6].

The alteration of one gene, however, does not suffice to give rise to full blown cancer. For progression toward malignancy and invasion, further mutational hits are necessary[7–9]. Hence the risk of cancer development does not only depend on mutations initiating tumorigenesis, but also on subsequent mutations driving tumor progression.

A quantitative understanding of cancer biology requires a mathematical frame-

work to describe the fundamental principles of population genetics and evolution that govern tumor initiation and progression[10,11]. Mutation, selection and tissue organization determine the dynamics of tumorigenesis[12–15] and should be studied quantitatively, both in terms of experiment and theory[16].

The mathematical investigation of cancer began in the 1960s, when Nordling[17], Armitage and Doll[18,19], and Fisher[20] set out to explain the age-dependent incidence curves of human cancers. These seminal studies led to the idea that multiple probabilistic events are required for the somatic evolution of cancer[21,22]. In the early 1980s, Knudson used a statistical analysis of the incidence of retinoblastoma in children to explain the role of tumor suppressor genes in sporadic and inherited cancers [10]. This work was later extended to a two-stage stochastic model for the process of cancer initiation and progression[9,11], which inspired much subsequent work[23–25]. Later on, considerable effort was devoted to the development of specific theories for drug resistance[26,27], angiogenesis[28], immune responses against tumors[29], and genetic instabilities[30–35].

Here, we address the following questions: What are the fundamental principles that determine the dynamics of activating oncogenes and inactivating tumor suppressor genes? How do mutation, selection and tissue architecture influence the rate of tumor initiation and progression? And how do quantitative approaches help to investigate the role of genetic instability in tumorigenesis?

2.2 Oncogenes

Oncogenes can contribute to tumorigenesis if one allele is mutated, amplified or inappropriately expressed[1]. In the last decades, many oncogenes have been discovered that are involved in various stages of human cancers - tumor initiation, progression, angiogenesis, and metastasis. Here, we discuss the basic aspects of evolutionary dynamics of oncogene activation and outline concepts such as selection, fixation, and random drift.

Tissues of multicellular organisms are subdivided into compartments[36], which contain populations of cells that proliferate to fulfill organ-specific tasks. Compartments are subject to homeostatic mechanisms which ensure that the cell number remains approximately constant over time. Whenever a cell divides, another cell has to die to keep the total population size the same. Cancer results if the equilibrium between cell birth and cell death is shifted toward uncontrolled proliferation. Not all cells of a compartment, however, might be at risk of becoming cancer cells. Differentiated cells, for example, might not have the capacity to divide often enough to accumulate the necessary number of mutations in cancer susceptibility genes[37]. The effective population size of a compartment describes those cells that are at risk of becoming cancer cells. In the following, compartment size will be used synonymously with effective population size within a compartment.

Consider a compartment of replicating cells. During each cell division, there is a small probability that a mistake will be made during DNA replication; in this case,

a mutated daughter cell is produced. The mutation might confer a fitness advantage to the cell by ameliorating an existing function or inducing a new function. Then the mutation is advantageous in terms of somatic selection. Alternatively, the mutation might impair an important cellular function and confer a fitness disadvantage to the cell. Then the cell proliferates more slowly or dies more quickly than its neighbors. The net reproductive rate is decreased, and the mutation is deleterious in terms of somatic selection. Finally, the mutation might not change the reproductive rate of the cell. Then the cell proliferates at the same rate as its neighbors, and the mutation is neutral in terms of somatic selection.

Let us discuss the dynamics of a particular mutation within a compartment. Initially, all cells are unmutated. What is the probability that a single mutated cell has arisen by time t? We measure time, t, in cell cycles. If the relevant cells divide once per day, then the unit of time is one day. Denote by N the number of cells in a compartment, and denote by u the mutation rate per gene per cell division. The probability that at least one mutated cell has arisen by time t is given by $P(t) = 1 - \exp(-Nut)$ (Figure 1a).

What is the fate of a single mutated cell? In the simplest scenario, there is a constant probability, q, that this cell will not die, but will initiate a neoplasia. The probability that a compartment has initiated a neoplasia by time t is given by $P(t) = 1 - \exp(-Nuqt)$.

Alternatively, consider a scenario in which the mutated cell has a relative fitness

r compared to a wild-type cell with fitness 1. If $r > 1$, the mutation is advantageous; if $r < 1$, the mutation is disadvantageous; if $r = 1$, the mutation is neutral. Normally, we expect mutations in oncogenes to cause increased net growth rates, $r > 1$; however, a mutation in an oncogene could be kept in check by apoptotic defense mechanisms, so r could be less than one.

What is the probability that such a mutation will not die out, but will take over the compartment? In order to calculate this probability, we consider a specific stochastic process known as the Moran process[38]. At each time step, a cell is chosen for reproduction at random, but proportional to fitness. If there are i mutated cells, then the probability that a mutated cell is chosen for reproduction is $ri/(ri+N-i)$. The chosen cell produces a daughter cell that replaces another randomly chosen cell that dies. The total number of cells remains strictly constant. The probability[39] that a single mutated cell with $r > 1$ or $r < 1$ takes over the whole compartment is given by $\rho = (1 - 1/r)/(1 - 1/r^N)$. For a neutral mutant[40], $r = 1$, the evolutionary dynamics are described by random drift in the frequency of the mutant in the population; the fixation probability is $\rho = 1/N$. In this stochastic process, a cell can either produce a lineage that dies out or that takes over the whole compartment. Stable coexistence of different cell types is impossible. The quantity ρ is called fixation probability. An advantageous mutation has a higher fixation probability than a neutral mutation, which has a higher fixation probability than a deleterious mutation. The events in a small compartment, however, are dominated by random

drift: if N is small, then even a deleterious mutation has a fairly high probability of reaching fixation due to chance events.

The probability that a mutation destined to be fixed has arisen by time t is given by $P(t) = 1 - \exp(-Nu\rho t)$ (Figure 1b). Here we neglect the time it takes the mutant to reach fixation in the population and therefore, $P(t)$ is also the probability that a mutant has been fixed by time t. Note that any mutation has a higher fixation probability in a small compartment than in a large compartment, but $P(t)$ is an increasing function of N for $r > 1$ and a decreasing function of N for $r < 1$. Thus, large compartments accelerate the accumulation of advantageous mutations, but slow down the accumulation of deleterious mutations. Conversely, small compartments slow down the accumulation of advantageous mutations, but accelerate the accumulation of deleterious mutations. Therefore, the compartment size is important in determining the types of mutations that are likely to occur[41,42].

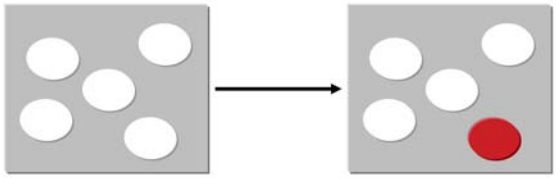
The linear process

So far, we have considered the evolutionary dynamics of a mutation that arises in a well-mixed compartment. This approach describes a tissue compartment in which all relevant cells are in equivalent positions and in direct reproductive competition with each other - there are no spatial effects. However, we can also envisage theories where cellular differentiation and spatial structure are explicitly modeled. One simple approach considers N cells in a linear array[43] (Figure 1c). Again, at each time step, a cell is chosen at random, but proportional to fitness. The cell is replaced

by two daughter cells, and all cells to its right are shifted by one place to the right. The cell at the far right undergoes apoptosis; the cell at the far left acts as a stem cell. In this approach, the fixation probability of a mutant cell is given by $\rho = 1/N$. The probability of fixation does not depend on its relative fitness r because only a mutation in the far left cell can reach fixation in the compartment. A mutation arising in an offspring cell will eventually be 'washed out' of the compartment by the continuous production of cells and their migration from the stem cell to differentiation and apoptosis. The probability that all cells of the compartment are mutated at time t is given by $P(t) = 1 - \exp(-ut)$; here we again neglect the time to fixation. Time is measured in units of stem cell divisions. If the stem cell divides more slowly than the other cells, then the accumulation of mutated cells is decelerated.

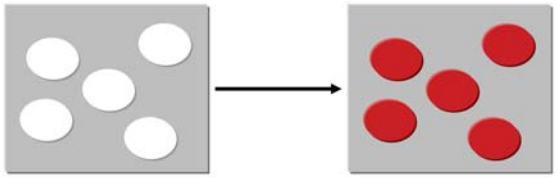
This 'linear process' of cancer initiation has the important feature of balancing out fitness differences between mutations[43]. Advantageous, deleterious and neutral mutations all have the same fixation probability, $\rho = 1/N$. This is in contrast to a well-mixed compartment, in which the fittest mutation has the highest probability of fixation. In comparison with a well-mixed compartment, a linear compartment delays the development of cancers that are initiated by advantageous mutations, such as mutations in oncogenes and tumor suppressor genes. However, it can increase the probability of cancer initiation via deleterious mutations, such as mutations in genetic instability genes[44]. Note also that the linear tissue design does not change the rate of accumulation of neutral mutations.

a



$$P(t) = 1 - e^{-Nut}$$

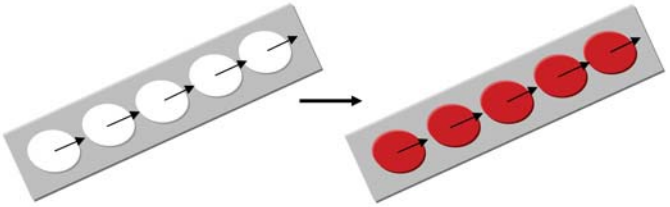
b



$$P(t) = 1 - e^{-Nu\rho t}$$

$$\rho = \frac{1 - 1/r}{1 - 1/r^{N}}$$

c



$$P(t) = 1 - e^{-ut}$$

$$\rho = 1/N$$

Figure 1: Oncogene dynamics. (a) The probability that at least one mutated cell has arisen in a compartment of N cells before time t is given by $P(t) = 1 - \exp(-Nut)$. Here u denotes the mutation rate per gene per cell division, and time is measured in units of cellular generations. (b) The probability that a compartment of N cells has been taken over by mutated cells is given by $P(t) = 1 - \exp(-Nu\rho t)$. The probability that a single mutant cell with relative fitness r reaches fixation is given by $\rho = (1 - 1/r)/(1 - 1/r^N)$. (c) N cells are arranged in a linear array. Whenever a cell divides, one daughter cell is put to its right, and all cells to the right are shifted by one place. The rightmost cell undergoes apoptosis. Here the fixation probability is $\rho = 1/N$ irrespective of r, because mutations have to arise in the leftmost cell (the stem cell) for not being 'washed out'. The probability that the compartment consists only of mutated cells at time t is given by $P(t) = 1 - \exp(-ut)$.

Numerical Examples

First, suppose an organ consists of $M = 10^7$ compartments. This is, for example, the approximate number of colonic crypts. Suppose each compartment consists of $N = 1000$ cells that divide once per day. The mutation rate per base per cell division [45] is approximately 10^{-10}. Assume a particular oncogene can be activated by any one of ten mutations. Thus, the rate of activating the oncogene per cell division is $u = 10^{-9}$. Suppose the activation of the oncogene confers a 10% growth advantage to the cell, $r = 1.1$. Then the probability of fixation is $\rho = (1-1/r)/(1-1/r^N) = 0.09$. The probability that a compartment has been taken over by mutated cells at time $t = 70$ years is $P(t) = 1 - \exp(-Nu\rho t) \approx 0.0023$. The expected number of mutated compartments at this age is $M \cdot P(t) \approx 23000$.

Second, assume a linear tissue architecture. Each compartment consists of 1000 cells, but is fed by one stem cell that divides every ten days[46,47]. Now, the probability that the compartment has been taken over by mutated cells at time $t = 70$ years is reduced to $P(t) \approx 2.6 \cdot 10^{-6}$. The expected number of mutated compartments at this age is 26.

Third, assume a population of $N = 10^7$ cells that divide every day. This population size describes, for example, a lesion that has already accumulated mutations in one or a few cancer-susceptibility genes; it does not, however, describe normal compartments in human tissues, as their population sizes are smaller. The probability that an oncogene leading to a relative fitness of $r = 1.1$ is activated within the next

year is given by $P \approx 0.28$. The time until the probability of activating the oncogene is $1/2$ is obtained as $T_{1/2} = 2.1$ years.

Conclusions

Tissue architecture and the scale of somatic selection greatly influence the rate at which mutations accumulate. Mutations activating oncogenes are thought to confer a selective advantage to the cell and are best contained by small compartments. The mutant cell is likely to reach fixation in the compartment, but its further spread is at least initially limited by the compartment boundaries. The risk of accumulating such mutations is reduced by adopting a linear tissue architecture in which the compartment is replenished by one or a few stem cells. Once one or several mutations have induced a neoplasia, however, additional mutations accumulate faster because of the increased population size.

The mathematical analysis of the dynamics of oncogene activation suggests a number of experimental investigations. What are the relative growth rates of cells bearing specific mutations in particular oncogenes as compared to wild type cells? What is the average time for the emergence and spread of an oncogenic mutation in a culture of dividing cells or in an experimental mouse tumor? How are human tissues organized? How large is the effective population size of a compartment within a particular tissue, and is it replenished asymmetrically by one or a few stem cells?

2.3 Tumor suppressor genes

So far, we have discussed genes that confer an altered phenotype to the cell when mutated in one allele, and are dominant at the cellular level. Some genes, however, are recessive at the cellular level, and have to be mutated in both alleles to cause a phenotypic change of the cell. Examples of recessive mutations are those that inactivate some tumor suppressor genes[1,48,49] (TSGs).

The concept of a TSG emerged from a statistical analysis of retinoblastoma incidence in children[10]. This study and subsequent work[1,11,50] led to the 2-hit hypothesis, which proposes that two hits in the retinoblastoma gene are the rate limiting steps of this particular cancer. In the inherited form, the first mutation is present in the germ line, whereas the second mutation emerges during somatic cell divisions. In the sporadic form, both mutations arise during somatic cell divisions. In the meanwhile, a large number of TSGs has been discovered that function in apoptosis, cell senescence, and other signaling pathways[1].

A normal cell has two alleles of a TSG. The mutation that inactivates the first allele can be neutral, disadvantageous, or advantageous, and a cell with one inactivated allele correspondingly has a normal, decreased, or increased net reproductive rate. The first hit is neutral if the TSG is strictly recessive: the remaining wild type allele has sufficient tumor suppressing function. The first hit is disadvantageous if the TSG is checked by apoptotic defense mechanisms: as soon as surveillance mechanisms discover an imbalance in the TSG product, apoptosis is triggered. The

first hit is advantageous if the TSG is haploinsufficient: the remaining wild type allele has insufficient tumor suppressing function. Here, we consider TSGs whose first hit is neutral. The mutation that inactivates the second allele is advantageous, and a cell with two inactivated alleles has an increased net reproductive rate. The mutation rates for the first and the second hit are denoted by u_1 and u_2, respectively. We assume $u_1 < u_2$ because some mutational mechanisms, such as mitotic recombination, can only constitute the second hit.

What is the probability that a single cell with two inactivated TSG alleles has arisen by time t in a population of N cells[39,51–53]? Interestingly, the answer depends on the population size, N, as compared with the mutation rates that constitute the first and second hit, u_1 and u_2. There are three different cases (Figure 2).

First, in small populations, $N < 1/\sqrt{u_2}$, a cell with one inactivated allele reaches fixation in the population before a cell with two inactivated alleles arises. The probability that at least one cell with two hits emerges before time t is

$$P(t) = 1 - \frac{Nu_2 \exp(-u_1 t) - u_1 \exp(-Nu_2 t)}{Nu_2 - u_1} \tag{1}$$

For very short times, $t < 1/Nu_2$, we can approximate $P(t) \approx Nu_1u_2t^2/2$. Therefore, this probability accumulates as second order of time: it takes two rate limiting hits to inactivate a TSG in a small population of cells.

Second, in populations of intermediate size, $1/\sqrt{u_2} < N < 1/u_1$, a cell with two inactivated alleles emerges before a cell clone with one inactivated allele has taken over the population. The population 'tunnels' from wild type directly to the second

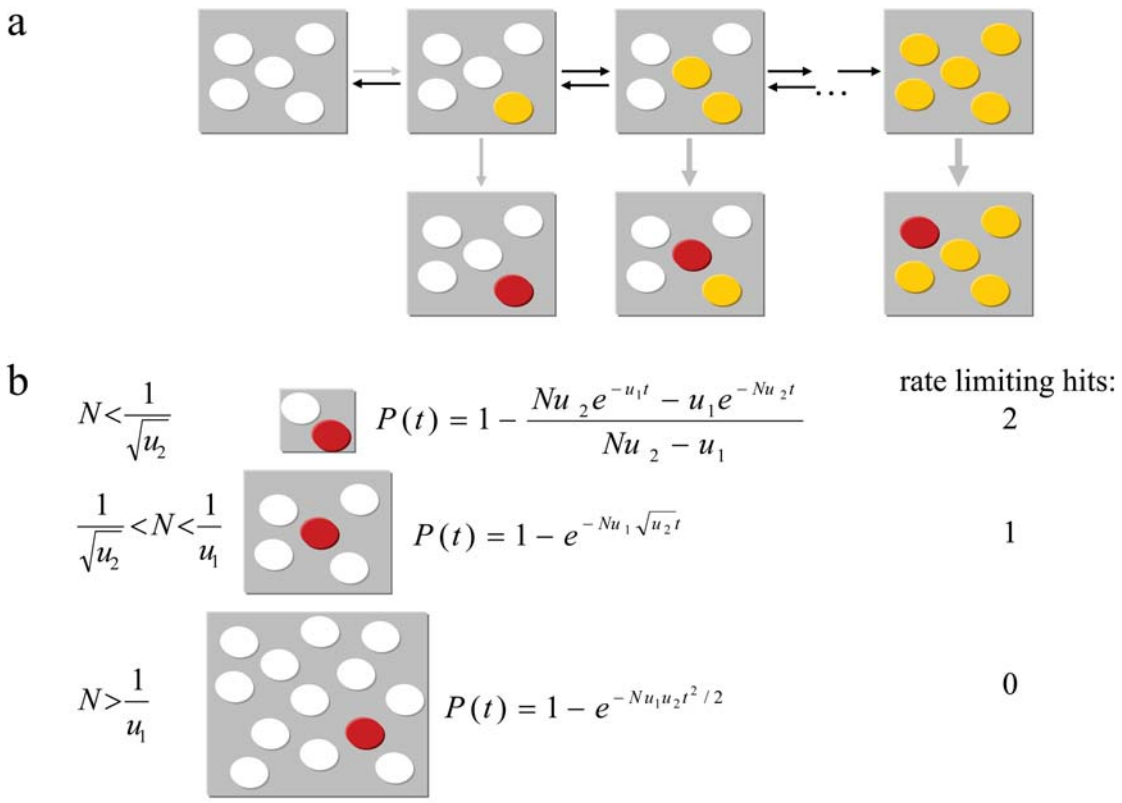


Figure 2: Tumor suppressor gene (TSG) inactivation dynamics. (a) Initially all cells are unmutated (white). Once a cell with one inactivated allele arises (grey), it gives rise to a lineage that can take over the whole compartment or become extinct again. In one of those cells, the second allele of the TSG may become inactivated (black). (b) TSG inactivation is described by three dynamic laws depending on the population size, N, and the probabilities of inactivating the first and the second allele, u_1 and u_2, per cell division. Time, t, is measured in cellular generations. In small, intermediate and large populations, a TSG is inactivated by two, one and zero rate limiting steps.

hit without ever having fixed the first hit[51,53]. The probability that at least one cell with two hits has arisen before time t is

$$P(t) = 1 - \exp(-Nu_1\sqrt{u_2}t) \tag{2}$$

This probability accumulates as a first order of time: it takes only one rate limiting hit to inactivate a TSG in a population of intermediate size.

Third, in very large populations, $N > 1/u_1$, cells with one inactivated allele arise immediately and the waiting time for a cell with two inactivated alleles dominates the dynamics. The probability that at least one cell with two hits has arisen before time t is

$$P(t) = 1 - \exp(-Nu_1u_2t^2/2) \tag{3}$$

This probability again accumulates as a second order of time. Eliminating a TSG in a large population of cells is, however, not rate limiting for the overall process of tumorigenesis. Due to the large population size, mutated cells are constantly being produced, and the inactivation of a TSG is not rate limiting (Figure 2b).

These three dynamic laws provide a complete description of TSG inactivation. In a normal tissue consisting of small compartments of cells, a TSG is eliminated by two rate limiting hits. The overall rate of inactivation is proportional to the second order of time. In small neoplasias, only one rate limiting hit is needed to inactivate a TSG. The rate of inactivation is proportional to the first order of time. In large tumors, it again takes two hits to inactivate a TSG, but neither of them is rate

limiting for the overall process of tumorigenesis. Therefore, as the population size increases, a TSG is inactivated in two, one or zero rate limiting steps.

Numerical Examples

Suppose an organ consists of $M = 10^7$ compartments. Each compartment contains about 1000 cells, but is replenished by only $N = 4$ stem cells that divide once per week. Assume a TSG is 10kb long and can be inactivated by a point mutation in any one of 500 bases, each occurring with probability 10^{-10} per cell division [45]. Then, the rate of inactivating the first of two TSG alleles is $u_1 = 2 \times 500 \times 10^{-10} = 10^{-7}$. Suppose the rate of inactivating the second TSG allele, including mitotic recombination, chromosome non-disjunction and other mechanisms of loss of heterozygosity (LOH), is $u_2 = 10^{-6}$. Then, using equation 1, the probability that a cell with two inactivated TSG alleles has arisen in a compartment at time $t = 70$ years is $P \approx 2.7 \cdot 10^{-6}$. The expected number of compartments containing at least one cell with two hits at time $t = 70$ years is $P \cdot M = 27$.

Now consider a small lesion of $N = 10^4$ cells that divide once per day. What is the probability that a TSG is inactivated in this lesion within the next twenty years? Again let $u_1 = 10^{-7}$ and $u_2 = 10^{-6}$. Using equation 2, we find $P \approx 0.0073$. What is the time, $T_{1/2}$, until the probability of having produced a cell with an inactivated TSG is 1/2? We find $T_{1/2} \approx 1900$ years. However, if a tumor contains $N = 10^9$ cells dividing once per day, the half life of the TSG reduces to $T_{1/2} \approx 120$ days.

Conclusions

The dynamics of TSG inactivation are described by three laws concerning the dependence of the population size and the mutation rates constituting the first and the second hit. In a small compartment, a TSG is inactivated by two rate limiting hits. In a small lesion, it is inactivated by one, and in a large tumor, by zero rate limiting hits.

Mutations inactivating TSGs, like mutations activating oncogenes, are best contained by small compartments that are replenished by a small number of stem cells[43]. Once a compartment has accumulated one or a few mutations in cancer-susceptibility genes, a neoplasia develops in which other TSGs might have to be inactivated for further tumor progression. Small neoplasias, however, are unlikely to succeed in inactivating a further TSG within a human lifespan under the assumption of normal mutation rates. Increased mutation rates due to genetic instability might be necessary for the further progression of some small lesions.

The mathematical analysis of TSG inactivation suggests several new experimental studies. How does the inactivation of one or two alleles of a TSG change the net growth rate of the cell? Is the first step indeed neutral or does it slightly modify the fitness of the cell? Such fitness differences can be measured in cell cultures or animal models. Furthermore, how long does it take for a population of cells to inactivate a TSG? How does the time until inactivation of a TSG depend on the population size and the mutation rates? Can the three dynamical regimes that we predict from the

theoretical analysis be verified? What are the relative rates of the various mechanisms that contribute to LOH? Precise kinetic measurements are needed in order to obtain quantitative insights into cancer progression.

2.4 Genetic instability

Genetic instability is a defining characteristic of most human cancers and represents one of the most active research areas in cancer biology[4]. Two major types of genetic instabilities have been identified[54]. In a small fraction of colorectal and some other cancers, a defect in mismatch repair results in an increased rate of point mutations and consequent widespread microsatellite instability. The majority of colorectal and most other cancers, however, have chromosomal instability (CIN), which refers to an increased rate of losing or gaining whole chromosomes or large parts of chromosomes during cell division. The consequence of CIN is an imbalance in chromosome number (aneuploidy) and an increased rate of LOH, which is an important property of CIN because it accelerates the rate of TSG inactivation.

The molecular basis for CIN is just beginning to be understood. A large number of genes that trigger CIN when mutated have been discovered in Saccharomyces cerevisiae[55–57]. These so-called 'CIN genes' are involved in chromosome condensation, sister-chromatid cohesion, kinetochore structure and function, microtubule formation and cell cycle checkpoints. By comparison with yeast, we expect several hundred human CIN genes, but only a few have been identified so far. These genes

include hBUB1, MAD2, BRCA2, and hCDC4[58]. The classification of CIN genes is based on the mutational events that are required to trigger CIN[4]. Class I CIN genes, such as MAD2, trigger CIN if one allele of the gene is mutated or lost. Class II CIN genes, such as hBUB1, trigger CIN if one allele is mutated in a dominant-negative fashion. Class III CIN genes, such as BRCA2, trigger CIN if both alleles are mutated.

A number of hereditary syndromes are known that stem from germ line mutations in what might be CIN genes. Inherited mutations in the RECQ-like helicases BLM and WRT give rise to the Bloom and Werner Syndromes, respectively. An inherited deficiency in nucleotide excision repair causes Xeroderma pigmentosum. Germ line mutations in the gene ATM give rise to Ataxia telangiectasia. These syndromes are all characterized by a high incidence of multiple types of cancer, but the mechanistic connection between these genes and CIN is still somewhat unclear.

An important question in cancer genetics is to what extent CIN, or any genetic instability, is an early event and driving force of tumorigenesis[5,22,59]. The investigation of the role of genetic instability requires a quantitative theory of how mutation and selection of gatekeeper and caretaker genes contribute to cancer initiation and progression[51,60–62]. Here, we discuss the role of CIN in cancers that are initiated by inactivation of one or two TSGs[51,63,64].

CIN before one TSG

Consider a case where tumorigenesis is initiated by inactivating a TSG, A, in a small compartment of cells[51]. An appropriate example is the inactivation of APC in a colonic crypt[1]. If the mutation rate is less than the inverse of the compartment size, then the compartment almost always consists of a single type of cells - a mutated cell either reaches fixation or goes extinct before another mutated cell arises. Initially, all cells are wild type, $A^{+/+}$. The compartment evolves from $A^{+/+}$, via $A^{+/-}$, to $A^{-/-}$. A mutation that triggers CIN can arise at any stage of this process. The crucial effect of CIN is to increase the rate of LOH, thereby accelerating the transition from $A^{+/-}$ to $A^{-/-}$. Stochastic tunneling can lead from $A^{+/-}$ without CIN directly to $A^{-/-}$ with CIN[53].

Figure 3a shows the rates of evolution of TSG inactivation with or without CIN in a small compartment of cells. There are two rate limiting hits for inactivating the TSG without CIN. Interestingly, there are also two rate limiting hits for inactivating the TSG with CIN[51]: one rate limiting step is needed for inactivating the first allele of the TSG, another rate limiting step is needed for the CIN mutation. The inactivation of the second TSG allele is greatly accelerated in the presence of CIN and is therefore not rate limiting[51].

There are three evolutionary trajectories, all of which contain two rate limiting steps (Figure 3b): the TSG can be inactivated without CIN; the CIN mutation can occur first, followed by the inactivation of the two TSG alleles; one TSG allele can be

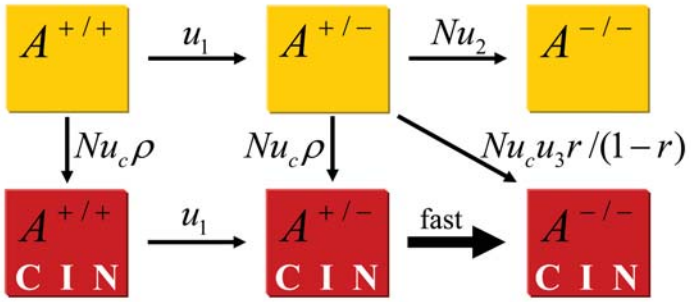
mutated first, followed by a CIN mutation, followed by the inactivation of the second TSG allele. The third trajectory contains a tunnel if there is a significant cost of CIN, i.e. if a CIN cell has a high probability of undergoing apoptosis due to deleterious and lethal mutations. Such a cell might not reach fixation in the compartment, but might nevertheless produce a cell with two inactivated TSG alleles that reaches fixation. The relative rates of these trajectories depend on the compartment size, N, the mutation rates, the number of CIN genes, and the cost of CIN (Figure 3b). We can use these rates to calculate the fraction of mutated compartments with or without CIN. We can also estimate the minimum number of CIN genes in the genome that are required to ensure that CIN precedes the inactivation of the first or second TSG allele[51,63].

Numerical Examples of CIN before one TSG

Again, consider an organ that consists of $M = 10^7$ compartments. Suppose each compartment contains 1000 cells, but is replenished by a small pool of $N = 4$ stem cells. Assume that the probabilities of inactivating the first and the second allele of a TSG per cell division are $u_1 = 10^{-7}$ and $u_2 = 10^{-6}$, respectively. We assume that the inactivation of both TSG alleles confers a big selective advantage to the cell, such that the probability of fixation is 1.

Assume there is one class I CIN gene in the genome that does not change the fitness of the cell when mutated, $r = 1$. What is the ratio of the probabilities of inactivating the TSG with CIN versus without CIN? We find that 50% of all

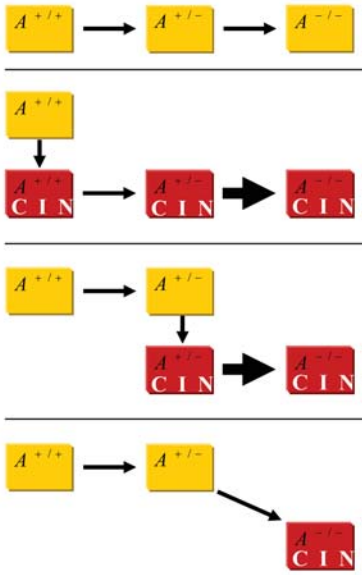
a



$$X(t) = Nu_1u_2t^2/2$$

$$Y(t) = Nu_1u_c\left(\rho + \frac{u_3}{2}\frac{r}{1-r}\right)t^2$$

b



	relative weight for $r \approx 1$	$r < 1$
	Nu_2	u_2
	u_c	$u_c\rho$
	u_c	$u_c\rho$
	0	$u_c u_3 \frac{r}{1-r}$

Figure 3: Emergence of chromosomal instability (CIN) during inactivation of one tumor suppressor gene (TSG). (a) The alleles of TSG A are inactivated at rates u_1 and u_2, respectively, per cell division. The probability of mutating a CIN gene is given by u_c. A CIN cell has relative fitness r and reaches fixation in a compartment of N cells with probability $\rho = (1 - 1/r)/(1 - 1/r^N)$. CIN increases the rate of inactivating the second TSG allele to $u_3 >> u_2$. The probabilities that the compartment is in state $A^{-/-}$ without and with CIN at time t are given by $X(t)$ and $Y(t)$, respectively. (b) The compartment can evolve along three evolutionary trajectories, all of which contain two rate limiting hits: the TSG can be inactivated without CIN; CIN can arise before the inactivation of the first TSG allele; or CIN can arise between the inactivation of the first and the second TSG allele. The relative weights of each trajectory for neutral CIN, $r \approx 1$, and costly CIN, $r < 1$, allow us to calculate their respective contributions to tumorigenesis.

mutated compartments have inactivated the TSG in a cell with CIN. In 25%, the CIN mutation occurred before the first TSG mutation, whereas in 25%, the CIN mutation occurred between the first and the second TSG mutation. In 50% of mutated compartments, there was no CIN mutation prior to the inactivation of the TSG. These results are largely independent of time.

Instead, assume that there are two class I CIN genes in the genome, each of which is sufficient to trigger CIN when mutated. Mutation of either gene reduces the fitness of the cell to $r = 0.7$. In this case, 54% of all mutated compartments have inactivated the TSG in a cell with CIN. In 25%, the CIN mutation occurred before the first TSG mutation. In 25%, the CIN mutation occurred between the first and the second TSG mutation without a tunnel, in 4% with a tunnel. In 46% of mutated compartments, there was no CIN mutation before the TSG inactivation.

Alternatively, we can calculate the critical number of CIN genes in the genome that are needed to ensure that CIN arises before the inactivation of the TSG. If CIN is neutral, $r = 1$, then one class I CIN gene or 11 class II CIN genes in the genome are needed. If CIN has a selective disadvantage of $r = 0.7$, then two class I CIN genes or 19 class II CIN genes in the genome are needed to make sure that CIN arises first. Note again that in yeast, more than hundred CIN genes are known, and it is possible that an even larger number exists in the human genome.

CIN in inherited cancers

In some genetic diseases, one allele of a TSG is mutated in the germ line. For example, patients with familial adenomatous polyposis (FAP) inherit a mutation in one allele of the tumor suppressor gene APC. By their teens, they harbor hundreds to thousands of colorectal polyps. In terms of the model, the mutational path starts with a compartment that consists only of $A^{+/-}$ cells without CIN. There are two evolutionary trajectories: the somatic mutation that inactivates the second TSG allele occurs in a cell without CIN, or a CIN mutation precedes the inactivation of the second TSG allele. Both trajectories require one rate limiting step[63]. Again, we can calculate the relative rates of these two possibilities. We can also calculate the critical number of CIN genes in the genome that are needed to ensure that CIN arises before the inactivation of the second allele. If CIN has a negligible cost, then twice as many CIN genes as before are needed to make sure that CIN comes first. The factor two comes from the fact that during the inactivation of two TSG alleles, CIN can arise either before the first or before the second hit. If the first hit is already present in the germ line, CIN can arise only before the second hit. This reduces the probability that CIN arises before the inactivation of the TSG by a factor 1/2. If CIN has a substantial cost, then the same number of CIN genes is needed as in the sporadic case with neutral CIN. Hence, there is only a small difference in the relative importance of CIN in those cases where both TSG alleles have to be inactivated somatically and where one TSG allele is already inactivated

in the germ line[63].

CIN before two TSGs

Consider a path to cancer in which two TSGs, A and B, have to be eliminated in rate limiting steps. Initially, the compartment consists of N_0 wild type cells, $A^{+/+}B^{+/+}$, that divide every τ_0 days. Suppose gene A has to be inactivated first, so the evolutionary pathway proceeds from $A^{+/+}B^{+/+}$ via $A^{+/-}B^{+/+}$ to $A^{-/-}B^{+/+}$, and subsequently to $A^{-/-}B^{+/-}$ and $A^{-/-}B^{-/-}$ (Figure 4). CIN can emerge at any stage of this pathway; once arisen, CIN accelerates the transitions from $A^{+/-}$ to $A^{-/-}$ and from $B^{+/-}$ to $B^{-/-}$. Inactivation of the first TSG induces neoplastic growth. We assume that the $A^{-/-}$ compartment gives rise to a small lesion of N_1 cells that divide every τ_1 days. In this lesion, the second TSG has to be inactivated for further tumor progression. Due to the increased compartment size, the evolutionary trajectory tunnels[53] from $A^{-/-}B^{+/+}$ directly to $A^{-/-}B^{-/-}$ (Figure 4).

Without CIN, inactivation of two TSGs requires three rate limiting hits. It takes two hits to inactivate the first TSG - if this leads to a moderate clonal expansion, then the second TSG can be inactivated in one rate limiting step (Equation 2). If the inactivation of the first TSG leads to a vast clonal expansion, then the inactivation of a further TSG is not rate limiting, and CIN becomes less important. With CIN, inactivation of two TSGs also requires three rate limiting hits. It takes two hits to inactivate the first TSG (one in a TSG allele and one in a CIN gene), and one further hit to inactivate the second TSG.

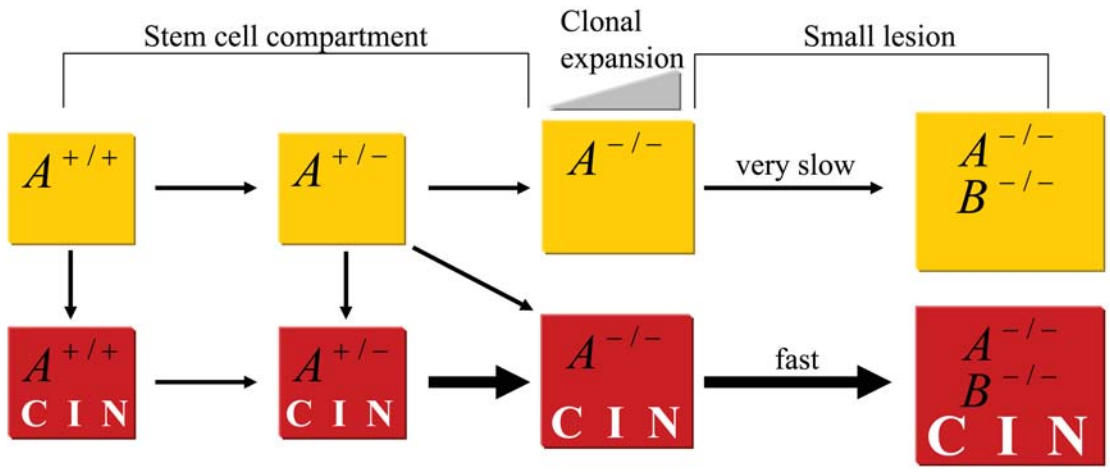


Figure 4: Emergence of chromosomal instability (CIN) during inactivation of two tumor suppressor genes (TSGs). Inactivation of TSG A in a compartment of N_0 cells causes a clonal expansion to N_1 cells in which TSG B is inactivated. Due to the increased compartment size, the evolutionary trajectory tunnels from $A^{-/-}B^{+/+}$ directly to $A^{-/-}B^{-/-}$. CIN can arise at any stage of tumorigenesis and increases the rate of TSG inactivation. It takes three rate-limiting hits to inactivate two TSGs both with and without CIN.

Again, we can calculate the minimum number of CIN genes in the genome that are needed to ensure that CIN arises before the inactivation of the first TSG in paths to cancer in which two TSGs have to be eliminated in rate limiting steps. We find that very few CIN genes in the genome are necessary to make sure that CIN arises early (Figure 5). The cost of CIN is compensated by an acceleration of every successive TSG inactivation. It is possible that the first TSG, A, is predominantly inactivated in cells without CIN. Thus, most lesions which are caused by inactivation of TSG A do not have CIN, but only the small fraction of lesions with CIN will eliminate TSG B within the time scale of a human life. In such a situation all (or almost all) cancers will derive from lesions in which a CIN mutation preceded inactivation of the first TSG.

Conclusions

The role of CIN in tumorigenesis is one of the most active and controversial fields of all of biology. In order to evaluate hypotheses and interpret experimental observations, a quantitative model is required for understanding how tissue organization, mutation rates, and selection determine the role of CIN in tumorigenesis. Here we have outlined a mathematical approach for evaluating the importance of early CIN in tumorigenesis initiated by inactivation of 'half', one or two TSGs in rate limiting situations. We find that one or a few neutral CIN genes in the genome are sufficient to ensure the emergence of CIN before the inactivation of one TSG. One or a few costly CIN genes in the genome are sufficient to ensure the emergence of CIN before

CIN before 1 TSG				CIN before 2 TSGs					
independent of N_1				$N_1 = 10^4$			$N_1 = 10^5$		
r	n_1	n_2	n_3	n_1	n_2	n_3	n_1	n_2	n_3
1.0	2	12	89	1	1	16	1	2	37
0.9	2	14	96	1	1	17	1	3	41
0.7	2	21	119	1	1	23	1	4	51
0.5	4	42	169	1	2	35	1	8	74

Figure 5: Minimum number of CIN genes needed to ensure that CIN arises before the inactivation of one or two TSGs. The relative selective fitness of CIN cells is denoted by r. Class I CIN genes, n_1, trigger CIN if one allele is mutated or lost. Class II CIN genes, n_2, trigger CIN if one allele is mutated in a dominant negative fashion. Class III CIN genes, n_3, trigger CIN if both alleles are mutated. Parameter values are $u_1 = 10^{-7}$, $u_2 = 10^{-6}$, $u_3 = 10^{-2}$, $N_0 = 4$, and $t = 80$ years.

the inactivation of the first TSG if two TSGs have to be eliminated in rate limiting steps.

The theoretical investigation of CIN suggests a number of new experiments. Can one demonstrate in mouse that the two rate limiting steps consist of inactivating one allele of the TSG and one CIN mutation? What is the selective cost of a CIN mutation? Can CIN be neutral or even advantageous[65]? Is it possible to find CIN mutations in dysplastic crypts, which represent the first stage of colon cancer[58]? What is the number of CIN genes in the human genome? How many of them fall into classes I, II or III? Is it possible to show that CIN is essential for inactivating two TSGs in small populations of cells?

2.5 Conclusions

We have outlined some fundamental principles of somatic evolutionary dynamics. We have discussed the rates at which mutations in oncogenes and tumor suppressor genes (TSGs) accumulate in small compartments of cells, early lesions, and large cancers. We have explained how the relevant time scales depend on population size, mutation rates, and fitness differences. We have studied the role of tissue architecture and genetic instability.

Tissue architecture and compartment size determine the rates at which different types of mutations accumulate. Cells with mutations in oncogenes or TSGs can have an increased somatic fitness. Such mutations are best contained when the tissue is

organized into small compartments and each compartment is fed by one (or a few) stem cells, as in the linear process[43]. Cells with mutations in genetic instability genes are likely to have a reduced somatic fitness. Such mutations accumulate faster in small compartments than in large compartments. Therefore, the optimum compartment size is a tradeoff between preventing mutations in oncogenes, tumor suppressor genes and genetic instability genes[41,42].

Chromosomal instability (CIN) confers an increased probability of gaining or losing whole chromosomes or arms of chromosomes. CIN accelerates the rate of inactivating TSGs. A major debate in cancer genetics is to what extent CIN is an early event and hence a driving force of cancer progression. We have shown that in a small compartment of cells, it takes two hits to inactivate a TSG both with and without CIN. Whether or not CIN occurs before the first TSG depends on the cost of CIN, the mutation rate, the rate of LOH, the number of CIN genes in the genome, and the population size. For a wide range of parameters, one (or a few) neutral CIN genes are enough to ensure that CIN initiates tumor formation in a pathway where one TSG needs to be eliminated in a rate limiting situation. One (or a few) costly CIN genes are enough to ensure that CIN initiates tumor formation in a pathway where two TSGs need to be eliminated in rate limiting situations.

We have reviewed models of somatic evolution with constant selection, which are based on the assumption that the fitness of mutants neither depends on their own relative abundance nor on the relative abundance of other mutants. Cancer

progression, however, might often include a complex interplay of various types of cells. Mutations in landscaper genes, alterations promoting angiogenesis and mutations enabling the cell to fight off the immune system might be examples. These complex dependencies require a game theoretical analysis[51], and we are still only at the beginning of a long endeavor.

Mathematical models of cancer evolution are essential for quantifying the effects of mutation, selection, and tissue architecture. The dynamical laws of tumor suppressor gene inactivation, oncogene activation, and emergence of genetic instability provide fundamental new insights into tumorigenesis. Theory can help to clarify concepts, interpret experimental data, and suggest new experiments. Mathematical models describing tumor initiation, progression, invasion and metastasis contribute toward a quantitative understanding of cancer biology.

3 Can chromosomal instability initiate cancer?

Michor F, Iwasa Y, Lengauer C, Vogelstein B, Nowak MA

Can chromosomal instability initiate tumorigenesis?[63]

Seminars in Cancer Biology 15, 43-49 (2005)

Cancers result from the accumulation of inherited and somatic mutations in oncogenes and tumor suppressor genes. These genes encode proteins that function in growth regulatory and differentiation pathways. Mutations in those genes increase the net reproductive rate of cells. Chromosomal instability (CIN) is a feature of most human cancers. Mutations in CIN genes increase the rate at which whole chromosomes or large parts of chromosomes are lost or gained during cell division. CIN causes an imbalance in chromosome number (aneuploidy) and an enhanced rate of loss of heterozygosity, which is an important mechanism of inactivating tumor suppressor genes. A crucial question of cancer biology is whether CIN is an early event and thus a driving force of tumorigenesis. Here we discuss mathematical models of situations where inactivation of one or two tumor suppressor genes is required for tumorigenesis. If two tumor suppressor genes have to be inactivated in rate-limiting steps, then CIN is likely to emerge before the inactivation of the first tumor suppressor gene.

3.1 Introduction

Tumor suppressor genes (TSGs) are negative regulators of cellular growth[1,48,50,66,67]. Loss of their function contributes to tumorigenesis[10,23,68]. The concept of a TSG emerged from a statistical analysis of retinoblastoma incidence in children[10]. This study and subsequent work led to the 2-hit hypothesis, which proposes that two mutations in the retinoblastoma gene are the rate limiting steps of this particular cancer[1]. In the inherited form, the first mutation is present in the germ line, whereas the second mutation emerges during somatic cell divisions. In the sporadic form, both mutations arise during somatic cell divisions. Inactivation of the first allele of a TSG does not normally change the phenotype of the cell and is considered an (almost) neutral mutation. Inactivation of the second allele confers a fitness advantage to the cell and can promote neoplastic growth. Human cancers result from the accumulation of mutations in several TSGs, oncogenes, and genes that are involved in maintaining genomic stability. Here, we discuss mathematical models of cancers that require the inactivation of one or two TSGs in rate limiting steps.

Genetic instability is a defining characteristic of most human cancers[4,69]. Two types of genetic instability have been identified[54]. A defect in mismatch repair (MMR) results in an elevated mutation rate at the nucleotide level leading to microsatellite instability[70,71] (MIN). MIN is only found in a small fraction of colorectal and some other cancers. The majority of colorectal and most other solid cancers, however, have chromosomal instability (CIN). CIN refers to an increased rate of los-

ing or gaining whole chromosomes or large parts of chromosomes during cell division. CIN causes an imbalance in chromosome number (aneuploidy) and an increased rate of loss of heterozygosity (LOH). An elevated rate of LOH is an important property of CIN, because it accelerates the inactivation of TSGs. The potential cost of CIN is discussed in section 3.2

Note that CIN refers to an enhanced rate of accumulating gross chromosomal changes during cell division; it is not synonymous with the state of aneuploidy that is observed in a static image of the cell's chromosomal content. While CIN is a process that drives most cells to aneuploidy, aneuploidy *per se* does not prove the existence of CIN. There are several ways in which a cell can become aneuploid in the absence of CIN. The cell could have divided many more times than other cells in the tissue; as gross chromosomal changes (rarely) occur in normal cells, this could lead to aneuploidy without CIN. Alternatively, aneuploidy could be due to exposure to endogenous or exogenous agents that disrupt spindle formation. It is also possible that cancer cells accumulate gross chromosomal changes at the same rate as normal cells, but that these changes are only lethal in normal cells. This hypothesis is strengthened by the knowledge that mutations in cancer cells disrupt the capacity of cells to undergo apoptosis.

Although CIN is defined as an increased rate of losing or gaining parts of chromosomes or whole chromosomes during cell division, CIN has only been formally demonstrated for whole chromosome losses[54]. There is no assay at present that can

reliably measure the rate of formation of changes at the subchromosomal level, such as deletions, insertions, inversions, rearrangements, amplifications, unequal sister chromatid exchange, and gene conversion. These changes are at least as common as losses or gains of whole chromosomes.

A large number of genes that trigger CIN when mutated have been identified in the yeast *Saccharomyces cerevisiae*[55–57]. These so-called 'CIN genes' are involved in chromosome condensation, sister-chromatid cohesion, kinetochore structure and function, and microtubule formation as well as in cell cycle checkpoints. Similarly, several genes can cause CIN in *Drosophila melanogaster* when genetically altered [72,73]. By analogy, we expect several hundred human CIN genes, but only few have been identified so far[74]. These genes include hBUB1, MAD2, BRCA1, BRCA2 and hCDC4[58,75–78]. The genes hBUB1 and MAD2 encode proteins that are required for function of the spindle assembly checkpoint[75,76]. This checkpoint modulates the timing of anaphase initiation in mitotic cells containing improperly aligned chromosomes and increases the probability of successful delivery of a correct chromosome set to each daughter cell. The genes BRCA1 and BRCA2 encode proteins that are implicated in DNA repair and recombination, checkpoint control of the cell cycle, and transcription[76,78]. hCDC4 is an E3 ubiquitin ligase that is thought to be involved in regulating the G1-S cell-cycle checkpoint by targeting proteins for destruction [58]. Inactivation of hCDC4 leads to increased levels of cyclin E, the formation of micronuclei, defects in the execution of anaphase, and CIN.

The classification of CIN genes is based on the mutational events required to engage CIN[4]. Class I CIN genes, such as MAD2, trigger CIN if one allele of the gene is mutated or lost. Class II CIN genes, such as hBUB1, trigger CIN if one allele is mutated in a dominant negative fashion. Both class I and II are 'single hit' CIN genes. Class III CIN genes, such as BRCA1 and BRCA2, trigger CIN if both alleles are mutated.

A major question in cancer genetics is to what extent CIN, or any genetic instability, is an early event and thus a driving force in tumorigenesis[2,39,41,51,59,60,79]. Experimental evidence shows that early colorectal adenomas have allelic imbalance [80] (Figure 6), but this provides no proof of underlying CIN. hCDC4 mutations have been shown to occur early in colorectal tumorigenesis[58]. Here, we discuss a mathematical framework for the evolutionary dynamics of tumorigenesis offering insights into the causal relations between CIN and cancers initiated by inactivation of one or two TSGs.

3.2 Chromosomal instability before one tumor suppressor gene

Consider a path to cancer where both alleles of a TSG, A, have to be inactivated in a single cell (Figure 7a). Initially, the cell is wild type, $A^{+/+}$. The cell evolves from $A^{+/+}$ via $A^{+/-}$ to $A^{-/-}$. The first allele of a TSG is typically inactivated by a point mutation. The second allele can be inactivated by one of several possible mechanisms

Chromosome arm	Adenomas with allelic imbalance (%)
1p	10
5q	55
8p	19
15q	28
18q	28
any of these 5	90

Figure 6: Experimental evidence of aneuploidy in early tumors. The table shows the percentage of early colorectal adenomas (size 1-3 mm, total number 32) with allelic imbalance in five different chromosome arms. Ninety percent of the adenomas have allelic imbalance in any one of the five chromosome arms. Data from Shih *et al.*[80].

(Figure 8). It can be inactivated by a second point mutation, leading to two distinct point mutations in the two heterozygous alleles. Alternatively, the chromosome on which the second allele resides can be lost; this chromosome loss might be deleterious or lethal. The two alleles at the TSG locus can recombine mitotically, leading to two identical point mutations on partly heterozygous alleles. Finally, the chromosome harboring the wild type allele can be lost and the chromosome harboring the first point mutation duplicated. This mechanism may require two steps, occurring in two distinct cell divisions, but could also occur in a single division through a process broadly related to mitotic recombination.

CIN can emerge at any stage of tumorigenesis. First, assume that CIN is dominant, i.e. triggered by mutation of a class I or II CIN gene[4,51]. The crucial effect of CIN is to increase the rate of LOH, thereby accelerating the transition from $A^{+/-}$ to $A^{-/-}$. CIN can have a cost for the cell by increasing the rate of accumulating lethal mutations and triggering apoptosis. CIN, however, can also be advantageous by reducing the duration of the cell cycle (if certain checkpoints have been eliminated) or by providing increased evolvability in detrimental environments[65,81]. Therefore, it makes sense to consider possibilities of costly, neutral, and advantageous CIN phenotypes.

Tissues are organized into small compartments of cells[12,36,46,47]. Not all cells of a compartment might be at risk of becoming cancer cells, however, because certain mutations might only have an effect if they occur in stem cells[37]. Suppose N_0 cells

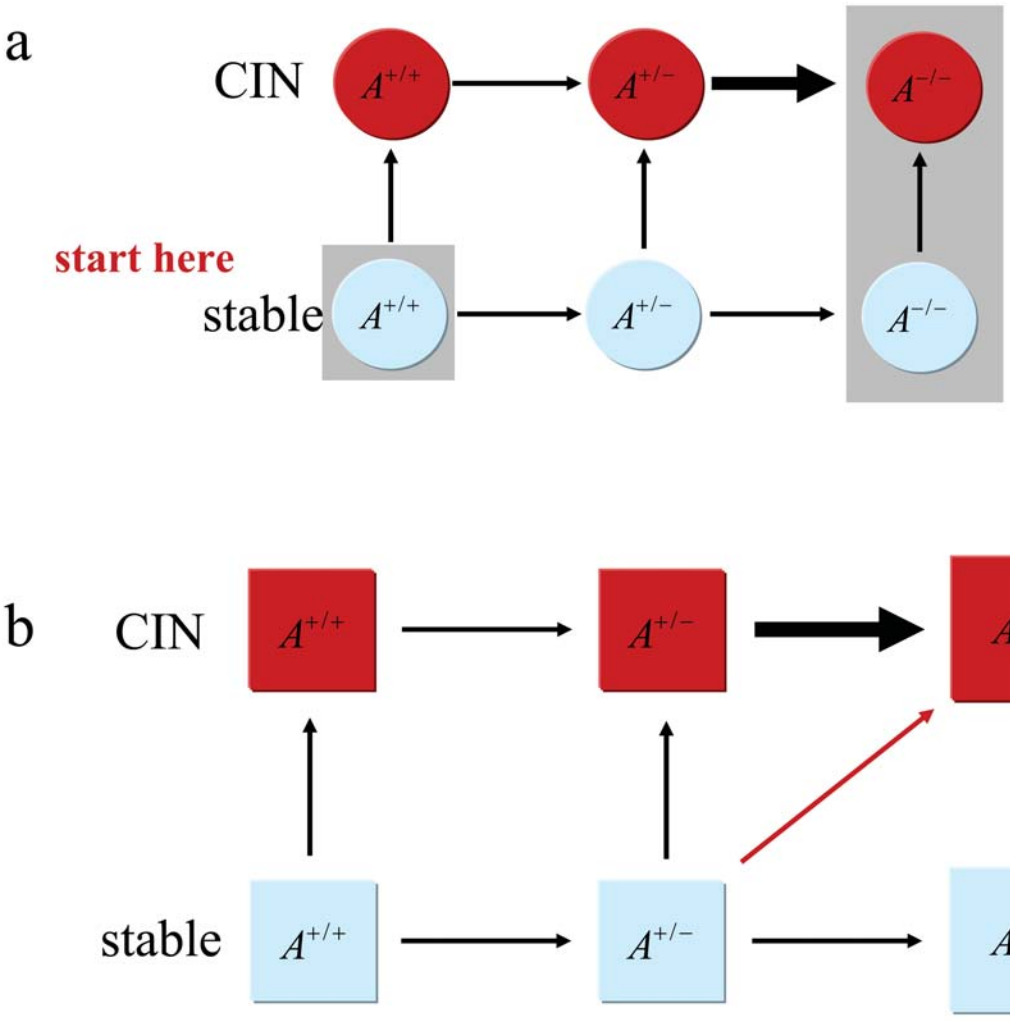

a
CIN
$A^{+/+}$
$A^{+/-}$
$A^{-/-}$
start here
stable
$A^{+/+}$
$A^{+/-}$
$A^{-/-}$
b
CIN
$A^{+/+}$
$A^{+/-}$
$A^{-/-}$
stable
$A^{+/+}$
$A^{+/-}$
$A^{-/-}$

Figure 7: Inactivation dynamics of a tumor suppressor gene (TSG), A, with and without chromosomal instability (CIN). (a) The wild type cell ('start') has two unmutated alleles of TSG A, $A^{+/+}$. The cell evolves from $A^{+/+}$ via $A^{+/-}$ to $A^{-/-}$. CIN can arise at any stage of TSG inactivation and causes very fast LOH. (b) A compartment of cells evolves along the evolutionary trajectory. If an $A^{+/-}$ cell clone with CIN produces an $A^{-/-}$ cell before taking over the compartment, a stochastic tunnel arises (diagonal arrow); the compartment evolves from $A^{+/-}$ without CIN to $A^{-/-}$ with CIN without ever visiting $A^{+/-}$ with CIN.

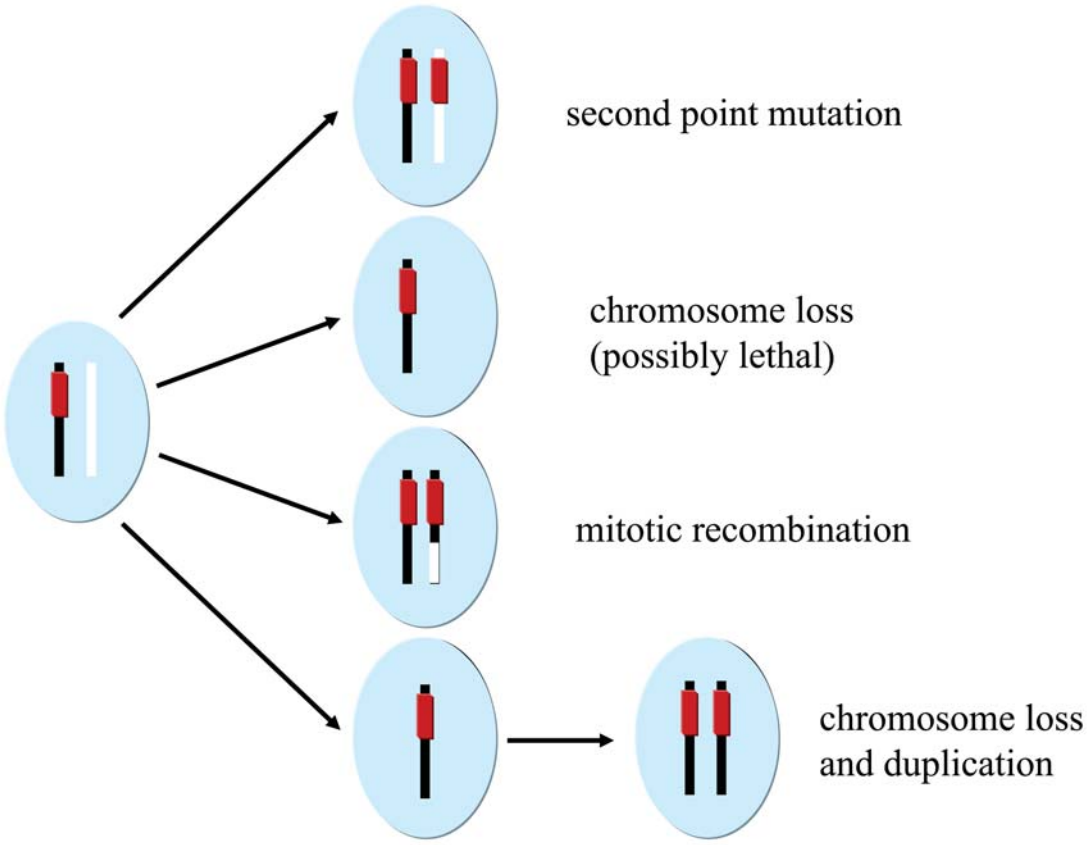


Figure 8: Mechanisms of tumor suppressor gene (TSG) inactivation. The first TSG allele is typically inactivated by a point mutation. The second allele can be inactivated by a second point mutation, by chromosome loss, mitotic recombination, or chromosome loss followed by duplication of the mutated allele.

are at risk in any one compartment. If the mutation rate is smaller than the inverse of N_0, then the approximation of homogeneous compartments holds: a mutated cell will either take over the compartment or go extinct before the next mutation arises (Figure 7b). If an $A^{+/-}$ cell clone with CIN produces an $A^{-/-}$ cell before taking over the compartment, then the phenomenon of 'stochastic tunneling' arises[51,53]: the compartment moves from $A^{+/-}$ without CIN to $A^{-/-}$ with CIN without ever visiting $A^{+/-}$ with CIN. If CIN is neutral or advantageous, then the tunnel does not arise. Mathematical procedures are outlined in Section 3.5.

3.3 Chromosomal instability before two tumor suppressor genes

Consider a path to cancer where two TSGs, A and B, have to be inactivated in a single cell (Figure 9). Initially, the cell is wild type, $A^{+/+}B^{+/+}$. Suppose gene A has to be inactivated first. Hence the cell evolves from $A^{+/+}B^{+/+}$ via $A^{+/-}B^{+/+}$ to $A^{-/-}B^{+/+}$, and subsequently to $A^{-/-}B^{+/-}$ and $A^{-/-}B^{-/-}$. CIN can emerge at any stage of tumorigenesis due to mutations of class I, II or III CIN genes. Once CIN has emerged, it accelerates the transitions from $A^{+/-}$ to $A^{-/-}$ and from $B^{+/-}$ to $B^{-/-}$.

Evolutionary dynamics within a compartment of cells are illustrated in Figure 10. CIN might emerge before gene A has been inactivated. A stochastic tunnel arises if an $A^{+/-}B^{+/+}$ cell clone with CIN produces an $A^{-/-}B^{+/+}$ cell before taking

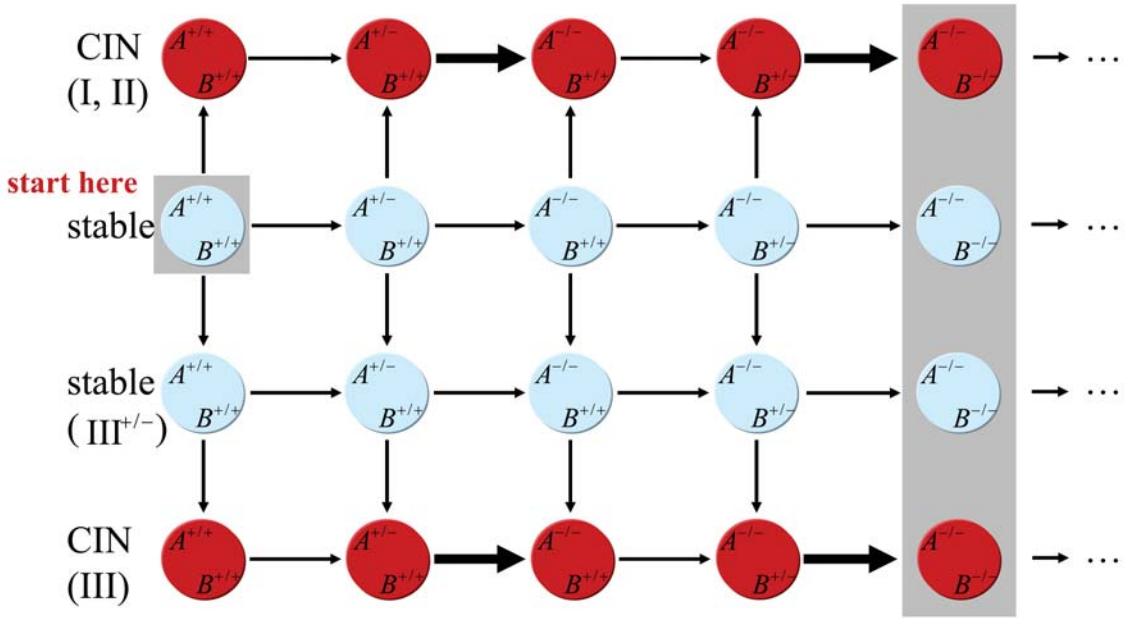


Figure 9: Mutational network of inactivating two tumor suppressor genes (TSGs), A and B, with and without chromosomal instability (CIN). The wild type cell ('start') has two unmutated alleles of both TSGs, $A^{+/+}B^{+/+}$. Mutations lead via $A^{+/-}B^{+/+}$, $A^{-/-}B^{+/+}$ and $A^{-/-}B^{+/-}$ to $A^{-/-}B^{-/-}$. Without CIN, the cell maintains a stable genotype (second row). CIN can arise at any stage of tumorigenesis due to mutations in a class I, II, or III CIN gene. Class I and II CIN genes require one hit to trigger CIN (first row). Class III CIN genes require two hits to trigger CIN, one in each allele (third and forth row). Once arisen, CIN accelerates the inactivation of the second allele of each following TSG. CIN, however, has a cost for the cell by increasing the chance of lethal mutations and apoptosis.

over the compartment[51,53]. Then the compartment moves from $A^{+/-}B^{+/+}$ without CIN to $A^{-/-}B^{+/+}$ with CIN without ever visiting $A^{+/-}B^{+/+}$ with CIN.

Inactivation of the first TSG can induce neoplastic growth. We assume that the $A^{-/-}$ compartment gives rise to a small lesion of N_1 cells. In this lesion the second TSG has to be inactivated for further tumor progression. Due to the increased compartment size, the evolutionary pathway might tunnel from $A^{-/-}B^{+/+}$ directly to $A^{-/-}B^{-/-}$. This means that the $A^{-/-}B^{+/-}$ cell does not reach fixation before the $A^{-/-}B^{-/-}$ cell arises. Mathematical procedures are outlined in Section 3.5.

The importance of early CIN in tumorigenesis depends on the number of possible CIN mutations, which in turn depends on the number of CIN genes in the human genome. We can calculate the minimum number of CIN genes in the genome that are needed to ensure that a CIN mutation precedes the inactivation of one or two TSGs. Let us discuss some plausible parameter choices (Figure 11). The mutation rate[45] per base per cell division is 10^{-10} to 10^{-11}. If an average TSG can be inactivated by any one of 1000 point mutations, the mutation rate per gene per cell division is $u = 10^{-7}$. Estimates of the rate of LOH in non-CIN cells range from $p_0 = 10^{-7}$ to $p_0 = 10^{-5}$. In our opinion, the most likely scenario is that p_0 has the same order of magnitude as u. In this case and in the absence of CIN, the hit inactivating the second TSG allele is sometimes LOH and sometimes a point mutation. If $p_0 >> u$, then two distinct point mutations should never be observed in the two TSG alleles. The rate of LOH in CIN cells[82] is $p = 10^{-2}$. Consider a compartment size of $N_0 = 4$.

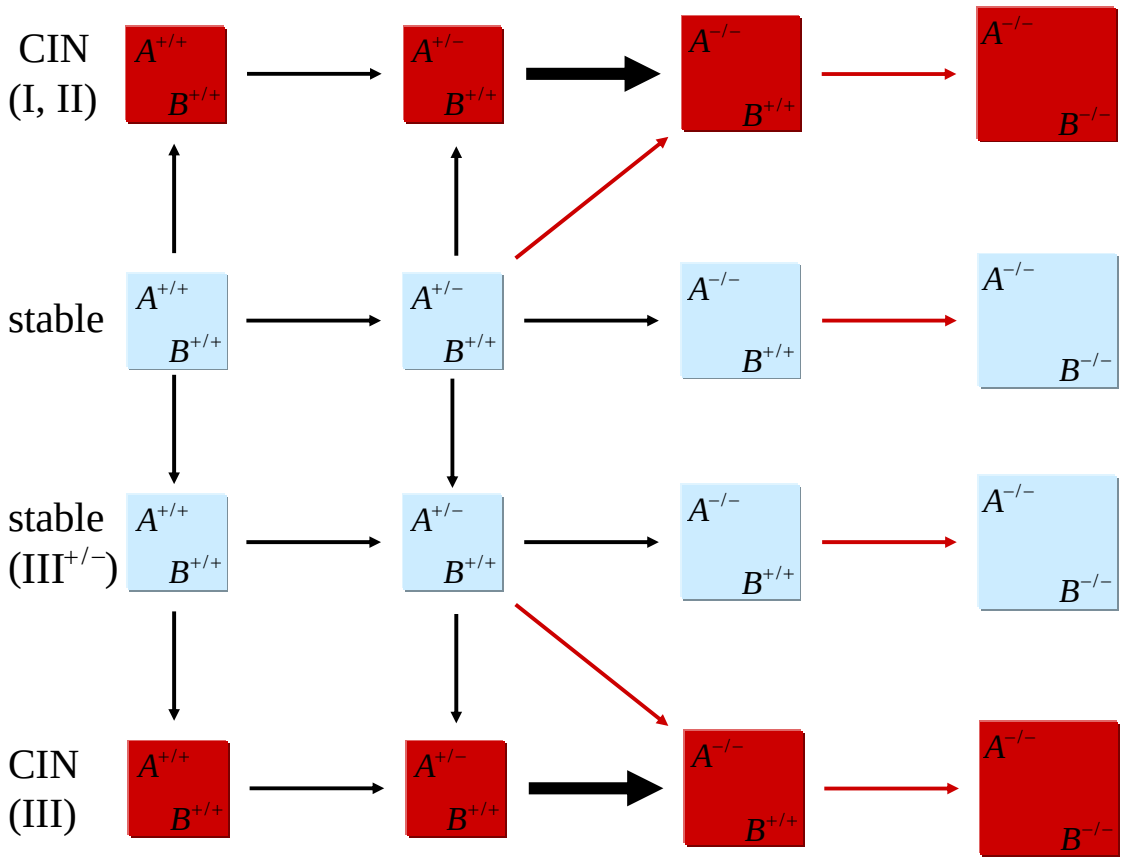


Figure 10: Inactivation dynamics starting with a population of N_0 wild type cells. CIN can emerge at any stage of tumorigenesis. If an $A^{+/-}B^{+/+}$ cell clone without CIN produces an $A^{-/-}B^{+/+}$ cell with CIN before reaching fixation, the compartment tunnels to $A^{-/-}B^{+/+}$ with CIN. Inactivation of TSG A leads to a small lesion of N_1 cells in which TSG B is eliminated. Due to the increased compartment size, the evolutionary pathway tunnels directly to $A^{-/-}B^{-/-}$.

This choice is motivated by the geometry of the colonic crypt[47]. Suppose the size of the lesion after clonal expansion of an $A^{-/-}$ cell clone is of the order of $N_1 = 10^4$ actively dividing cells. CIN can be disadvantageous, neutral, or advantageous[74,81]. The most substantial cost CIN can possibly have is $r = (1-p)^{45} \approx 0.6$; this means that loss of any chromosome other than the one containing the TSG locus is lethal. CIN is neutral if $r = 1$ and advantageous if $r > 1$. Given these parameter choices and $r = 0.6$, we find that 3 class I CIN genes, or 29 class II CIN genes, or more than 100 class III CIN genes are required to ensure that CIN precedes inactivation of the first TSG in a pathway where no other TSG has to be eliminated in a rate limiting step. If two TSGs have to be inactivated in rate limiting steps, then 1 class I CIN gene, or 1 class II CIN gene, or 30 class III CIN genes are sufficient for CIN to precede inactivation of the first TSG. Figure 11 provides further examples for other parameter choices.

The parameters have different effects on the importance of early CIN. An increase in N_0, N_1 and p_0 as well as a decrease in r and p make the early emergence of CIN less likely. Instead of calculating the inactivation dynamics of one specific TSG, we can also study the possibility that one gene out of a family of TSGs can be inactivated. Suppose gene A is any one of ten TSGs and likewise gene B is any one of ten TSGs. This is an extreme example. In this scenario and using the same parameter values as above, the same number of CIN genes in the human genome as before is required to ensure that CIN precedes inactivation of the first TSG in a

	CIN before 1 TSG				CIN before 2 TSGs					
	independent of N_1				$N_1 = 10^4$			$N_1 = 10^5$		
	r	n_1	n_2	n_3	n_1	n_2	n_3	n_1	n_2	n_3
a	1.4	1	2	95	1	1	8	1	1	21
	1.0	2	3	>100	1	1	11	1	1	26
	0.6	3	6	>100	1	1	20	1	1	42
b	1.4	1	2	>100	1	1	10	1	1	23
	1.0	3	6	>100	1	1	17	1	1	40
	0.6	27	54	>100	1	1	59	2	4	>100
c	1.4	1	8	>100	1	1	13	1	2	31
	1.0	2	12	>100	1	1	16	1	2	38
	0.6	3	29	>100	1	1	30	1	5	61
d	1.4	1	66	>100	1	5	24	1	27	65
	1.0	2	>100	>100	1	8	31	1	40	76
	0.6	3	>100	>100	1	25	54	1	100	>100

Figure 11: Minimum number of CIN genes needed to ensure that CIN emerges before one or two TSGs. Class I CIN genes, n_1, trigger CIN if one allele is mutated or lost. Class II CIN genes, n_2, trigger CIN if one allele is mutated in a dominant negative fashion. Class III CIN genes, n_3, trigger CIN if both alleles are mutated. The relative somatic fitness of CIN cells is denoted by r, and the compartment size after clonal expansion by N_1. Results are obtained by numerical simulation of Equation 6. Parameter values are $u = 10^{-7}$, $p = 10^{-2}$, $t = 80$ years, and $p_0 = 10^{-7}$ and $N_0 = 4$ in (a), $p_0 = 10^{-7}$ and $N_0 = 10$ in (b), $p_0 = 10^{-6}$ and $N_0 = 4$ in (c), and $p_0 = 10^{-5}$ and $N_0 = 4$ in (d).

pathway where no other TSG has to be eliminated in a rate limiting step. If two TSGs have to be inactivated in rate limiting steps, then 1 class I CIN gene, or 5 class II CIN gene, or 62 class III CIN genes are sufficient for CIN to precede inactivation of the first TSG. Hence, a large number of alternative TSGs increases the number of class II and III CIN genes needed, but does not significantly alter the number of class I CIN genes required for CIN to arise early.

3.4 Conclusions

Even if many alternative CIN genes are needed to ensure that costly CIN emerges before one TSG, only very few CIN genes suffice for CIN to precede two TSGs. This effect is especially strong if inactivation of the first TSG leads to a moderate clonal expansion. In this case, the second TSG must be inactivated in a rate limiting fashion. If, on the other hand, the inactivation of the first TSG causes a rapid clonal expansion ($N_1 >> 10^5$), then the inactivation of the second TSG does not occur in a rate limiting step, and CIN can only accelerate inactivation of the first TSG. A wide range of plausible parameter values, which are conservatively biased against CIN, all give the same message: of the order of 1 (to 10) neutral CIN genes are needed to ensure that CIN emerges before the inactivation of one TSG; of the order of 1 (to 10) costly CIN genes are needed to ensure that CIN emerges before the inactivation of the first of two TSGs. By analogy with yeast, we expect several hundreds of CIN genes in the human genome. Therefore it is likely that in any one human tissue a

large number of genes will lead to CIN when mutated. Thus, in any pathway of cancer progression where at least two TSGs need to be eliminated in rate limiting steps, CIN will arise before inactivation of the first TSG and therefore initiate the mutational sequence that leads to cancer.

3.5 Methods

The stochastic process illustrated in Figures 7b and 10 can be described by differential equations. Denote by X_0, X_1, X_2, and X_4 the probabilities that a compartment is in state $A^{+/+}B^{+/+}$, $A^{+/-}B^{+/+}$, $A^{-/-}B^{+/+}$, and $A^{-/-}B^{-/-}$, respectively, without chromosomal instability (CIN). Denote by Y_0, Y_1, Y_2, and Y_4 the probabilities that a compartment is in state $A^{+/+}B^{+/+}$, $A^{+/-}B^{+/+}$, $A^{-/-}B^{+/+}$, and $A^{-/-}B^{-/-}$, respectively, with CIN. The differential equations are given by

$$\begin{aligned}
\dot{X}_0 &= X_0(-r_0 - c_0) & \dot{Y}_0 &= c_0 X_0 - s_0 Y_0 \\
\dot{X}_1 &= r_0 X_0 - X_1(r_1 + c_0 + c_1) & \dot{Y}_1 &= s_0 Y_0 + c_0 X_1 - s_1 Y_1 \\
\dot{X}_2 &= r_1 X_1 - r_2 X_2 & \dot{Y}_2 &= s_1 Y_1 + c_1 X_1 - s_2 Y_2 \\
\dot{X}_4 &= r_2 X_2 & \dot{Y}_4 &= s_2 Y_2
\end{aligned} \tag{4}$$

The transition rates between the states are given by $r_0 = 2u/\tau_0$, $r_1 = N_0(u+p_0)/\tau_0$, $r_2 = N_1 2u\sqrt{u+p_0}/\tau_1$, $c_0 = N_0 u_c \rho/\tau_0$, $c_1 = [N_0 u_c r/(1-r)][p/\tau_0]$, $s_0 = r_0$, $s_1 = N_0 p/\tau_0$, and $s_2 = r N_1 2u\sqrt{p}/\tau_1$. Cells with genotype $A^{+/+}B^{+/+}$ and $A^{+/-}B^{+/+}$ divide every τ_0 days, while cells with genotype $A^{-/-}B^{+/+}$ and $A^{-/-}B^{+/-}$ divide

every τ_1 days. Let us first discuss the mutation-selection network leading from X_0 via X_1 and X_2 to X_4. The rate of change, r_0, is equal to the rate that a mutation occurs in the first allele of TSG A times the probability that the mutant cell will take over the compartment. The mutation occurs at rate $2N_0u/\tau_0$, because (i) the mutation can arise in any one of N_0 cells, (ii) the cells divide once every τ_0 days, (iii) the mutation rate per gene per cell division is given by u and (iv) there are 2 alleles. The probability that the mutated cell takes over the compartment (i.e., reaches fixation) is given by $1/N_0$, if we assume that the mutation is neutral[40]. Thus, the population size, N_0, cancels in the product, and we simply obtain $2u/\tau_0$.

Once the first allele of TSG A has been inactivated, the second allele can be inactivated either by another point mutation or an LOH event. This process occurs at rate $r_1 = N_0(u + p_0)/\tau_0$, where p_0 is the rate of LOH in normal cells. We assume that $A^{-/-}$ cells have a large fitness advantage, which means the probability of fixation of such a cell is close to one. Inactivation of both alleles of TSG A leads to a small lesion of N_1 cells dividing every τ_1 days. In this lesion, TSG B is inactivated at rate $r_2 = N_1 2u\sqrt{u + p_0}/\tau_1$.

CIN can arise at any stage of tumorigenesis and causes very fast LOH[82]. Denote the rate of LOH in CIN cells by p; we have $p >> p_0$. The rate of change, c_0, is equal to the rate at which a mutation triggering CIN occurs, $N_0u_c\tau_0$, times the probability that the mutant cell will take over the compartment. The mutation rate triggering CIN, u_c, depends on the number of available class I, II and III CIN genes in the

human genome, n_1, n_2 and n_3, respectively. The mutation rates of class I and II CIN genes are $u_c = 2n_1(u + p_0)$ and $u_c = 2n_2u$, respectively. The mutation rate of the first allele of class III CIN genes is $2n_3u$, and the mutation rate of the second allele of class III CIN genes is $u + p_0$. The probability of fixation of a CIN cell depends on its somatic fitness, r, and is given by $\rho = (1 - 1/r)/(1 - 1/r^{N_0})$. This is the standard fixation probability of a Moran process[38,43]. Therefore, the rate of change is given by $c_0 = N_0u_c\rho/\tau_0$. A stochastic tunnel arises if an $A^{-/-}B^{+/+}$ cell with CIN reaches fixation before an $A^{+/-}B^{+/+}$ cell does. The rate of tunneling is given by $c_1 = [N_0u_cr/(1 - r)][p/\tau_0]$.

Let us now discuss the mutation-selection network leading from Y_0 via Y_1 and Y_2 to Y_4. The rate of change, s_0, is equal to the rate r_0, because CIN does not change the probability of point mutations. The rate at which the second allele of TSG A is inactivated is given by $s_1 = N_0p/\tau_0$. The rate at which TSG B is inactivated is given by $s_2 = rN_1 2u\sqrt{p}/\tau_1$.

Equation 4 is a system of linear differential equations which can be solved analytically using standard techniques.

4 Dynamics of Colorectal Cancer

Michor F, Iwasa Y, Lengauer C, Nowak MA

Dynamics of Colorectal Cancer[83]

Seminars in Cancer Biology, in press (2005)

Colorectal cancer results from an accumulation of mutations in tumor suppressor genes and oncogenes. An additional defining characteristic of colorectal cancer is its genetic instability. Two main types of genetic instability have been identified. Microsatellite instability leads to an increased point mutation rate, whereas chromosomal instability refers to an enhanced rate of accumulating gross chromosomal aberrations. All colon cancer cell lines are genetically unstable. An interesting question is whether genetic instability arises early in tumorigenesis. An early emergence of genetic instability could drive most of the somatic evolution of cancer. Here, we review mathematical models of colorectal tumorigenesis and discuss the role of genetic instability.

4.1 Introduction

'I wish I had the voice of Homer, to sing of rectal carcinoma' are the opening lines of a poem by British mathematical biologist J. B. S. Haldane, who was diagnosed with this disease in 1964 and died a few years later. Haldane, together with Ronald Fisher and Sewell Wright, was one of the founding fathers of population genetics. He unified Darwinian evolution with Mendelian genetics and brought about the current understanding of evolutionary biology. Here, we apply population genetics to the evolutionary dynamics of colorectal cancer.

Colorectal cancer is the second leading cause of cancer death in the United States. In 2004, there were an estimated 100,000 new cases and 57,000 deaths from this disease[84]. The incidence of colorectal cancer, however, could be reduced dramatically by preventative methods such as colonoscopy and detection of mutations in fecal DNA[85,86].

Colorectal cancer progresses through a series of clinical and histopathological stages ranging from single crypt lesions through small benign tumors (adenomatous polyps) to malignant cancers (carcinomas)[1]. The model of colorectal tumorigenesis includes several genetic changes that are required for cancer initiation and progression[87] (Figure 12). The earliest and most prevalent genetic event yet identified in colorectal tumorigenesis is genetic disruption of the APC (adenomatous polyposis coli) pathway[66]. This pathway is documented to be altered in approx. 95% of colorectal tumors[88]. A critical function of APC is the inhibition of β-catenin/Tcf-mediated

transcription[89]. Inactivation of APC leads to increased β-catenin/Tcf-mediated transcription of growth-promoting genes[90]. In tumors with wild-type APC, mutations of β-catenin have been observed that render it resistant to the inhibitory effects of APC[91]. Loss of APC function or gain of β-catenin function leads to clonal expansion of the mutated epithelial cell, giving rise to a small adenoma[92]. Mutations in the RAS/RAF pathway, the p53 pathway, and several other genes/pathways drive tumor progression towards malignancy and metastasis[1]. Figure 13 shows the age-specific incidence of colorectal cancer.

Two types of genetic instability have been identified in colorectal cancer[54]. In about 15% of all sporadic colorectal cancers, mismatch-repair deficiency leads to microsatellite instability (MIN) at the nucleotide level[93]. The remaining 85% of sporadic colorectal cancers have chromosomal instability (CIN), which refers to an increased rate of losing or gaining whole chromosomes or parts of chromosomes during cell division[82]. CIN can contribute to an increased rate of loss of heterozygosity (LOH), which is considered an important mechanism of tumor suppressor gene inactivation. More than a hundred genes can cause CIN when mutated in yeast cells, many of which have several homologues in humans[94]. These include genes that are involved in chromosome metabolism, spindle assembly and dynamics, cell-cycle regulation and mitotic checkpoint control[74].

The classification of CIN genes is based on the mutational events that are required to trigger CIN[4]. Class I CIN genes, such as MAD2, trigger CIN if one allele

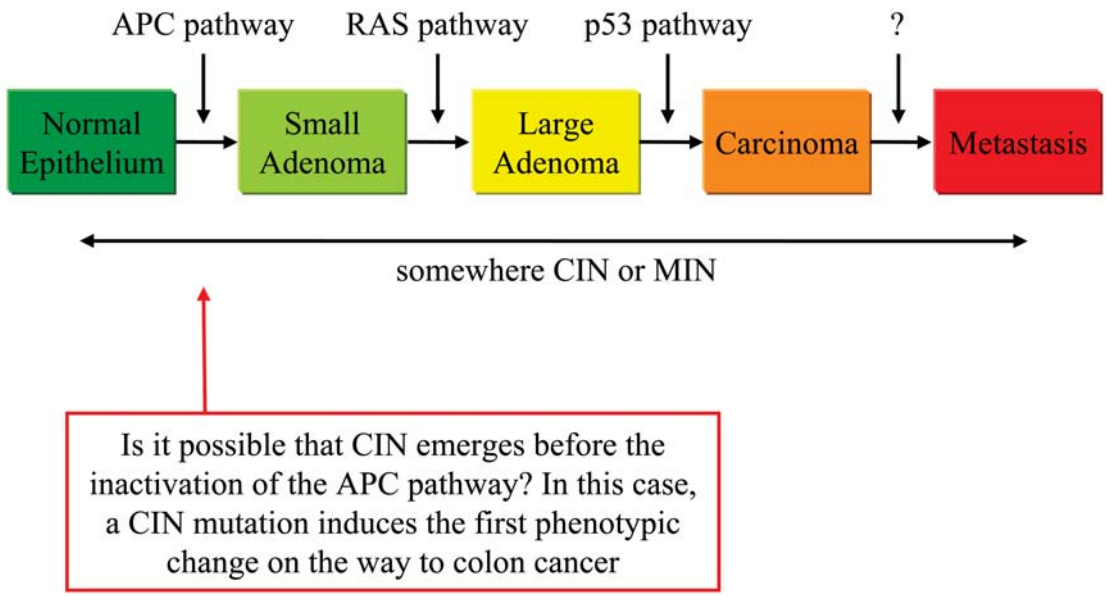


Figure 12: The genetic model of colorectal tumorigenesis. Colorectal cancers develop over the course of 20-40 years due to genetic disruption of the APC, RAS and p53 pathways. Genetic instability arises somewhere during the process of colorectal tumorigenesis, but whether it is the first event and therefore initiates and drives the neoplastic transformation is still a matter of much debate.

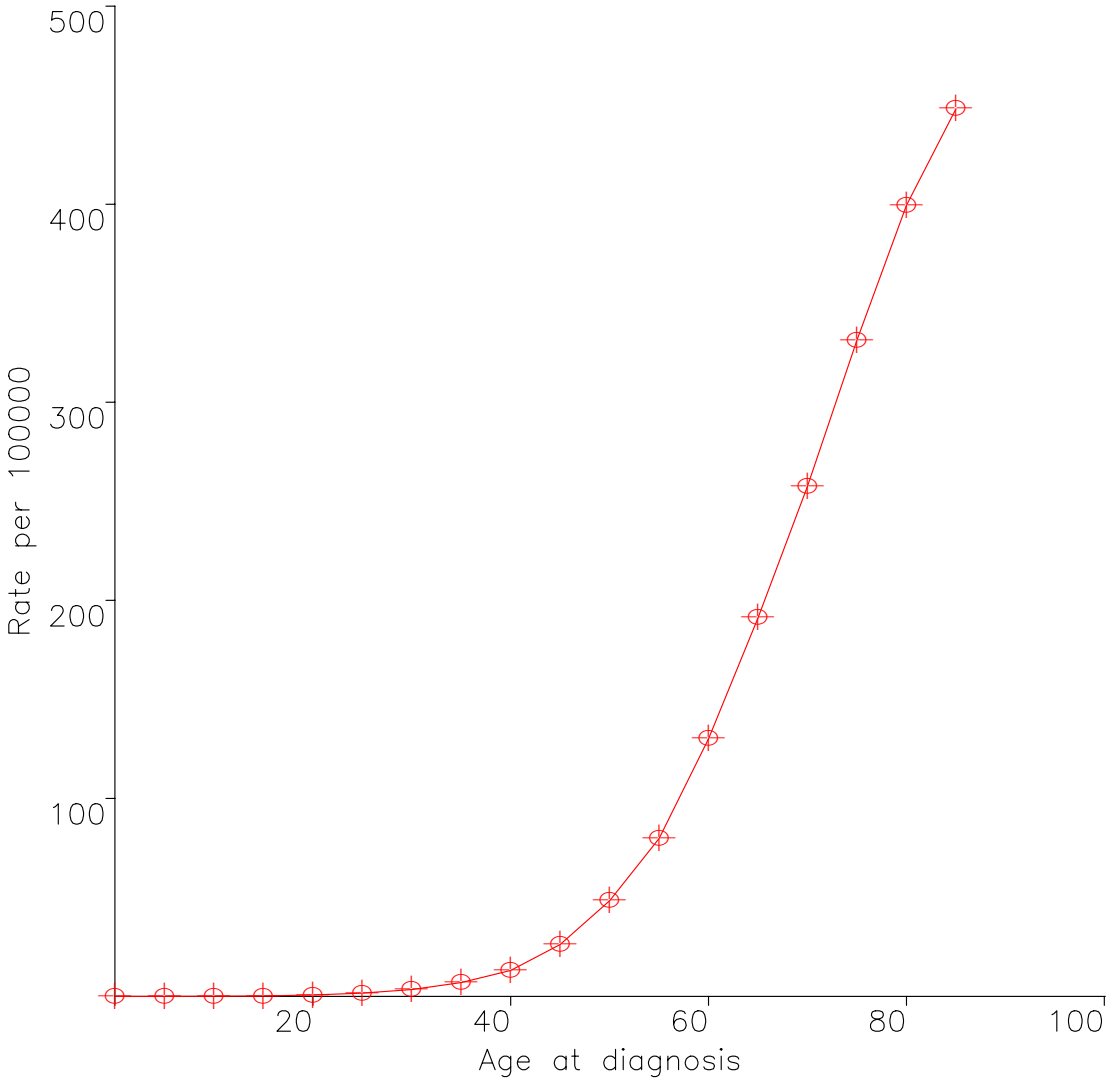


Figure 13: Age-specific incidence of colorectal cancer in the US. Statistics were generated from malignant cases only, and rates are expressed as cases per 100,000. Data from Surveillance, Epidemiology, and End Results (SEER) Program (www.seer.cancer.gov), National Cancer Institute, released April 2004, based on the November 2003 submission.

of the gene is inactivated or lost. Class II CIN genes, such as hBUB1, trigger CIN if one allele is mutated, and this mutation exerts an effect. Class III CIN genes, such as BRCA2, trigger CIN only if both alleles are mutated.

The role of genetic instability in colorectal tumorigenesis is an active research area of cancer biology[4,5]. Loeb was the first to propose that cancer cells acquire a 'mutator phenotype' early in tumor progression[69,95]. Several lines of evidence, experimental and theoretical, indicate that genetic instability can indeed arise early in colorectal tumorigenesis and subsequently drive tumor progression[4]. This view is widely, however not universally accepted[5,96].

Here, we address the following questions: How can the dynamics of colorectal tumorigenesis be described with mathematical models? And what is the role of genetic instability in tumor initiation?

4.2 Mutations in colorectal cancer cell lines

The genetic model of colorectal tumorigenesis (Figure 12) is based on decades of investigation into the mutational sequence of colorectal tumorigenesis[87]. The fraction of tumors with mutations in the APC pathway is approximately 95%; the same number holds for the fraction of tumors with mutations in the p53 pathway. Somatic mutations that lead to alteration of the Ras/Raf pathway have so far been found in about 70% of tumors. There are additional genes with somatic mutations in colorectal cancers, but their respective roles and pathways are less well understood

and examined.

Mutational studies require both tumor and normal tissue material for comparison. The attainment of tumor material poses a significant technical difficulty, as contamination with normal tissue renders signal and noise indistinguishable. Only a few colorectal cancers and their corresponding normal tissue counterparts have been analyzed systematically to determine the spectrum of specific mutations in a series of genes, i.e. all APC pathway genes (APC, β-catenin, and axin), p53 pathway genes (p53, MDM2, and BAX), and RAS pathway genes (K-Ras and B-Raf) in the same tumors. Though such data are sparse, profound differences between the mutations found in CIN and MIN tumors have been reported. At least a third of MIN cancers contain mutations in β-catenin instead of APC, whereas β-catenin mutations are extremely rare in non-MIN cancers[97]. This is indirect evidence suggesting that MIN arises before inactivation of the APC pathway. Additionally, the spectrum of APC mutations that occur in most MIN cancers without β-catenin mutations is different from that in non-MIN cancers[98]. In particular, there is a higher frequency of alterations in simple repeat sequences in the MIN than in the non-MIN cancers.

4.3 Mathematical modeling

Mathematical modeling of colorectal tumorigenesis has provided insights into the dynamics and statistics of the disease. Luebeck and Moolgavkar[23] subdivided the multi-step model into two stages; during the first stage, accumulation of neutral mutations does not lead to any phenotypic changes; during the second phase, advantageous mutations drive clonal expansion of the colorectal tissue. By fitting the age-specific incidence data of colorectal cancer, they found that two rare events (inactivation of both APC alleles) are needed for tumor initiation and only one more event is necessary for malignant transformation. Luebeck and Moolgavkar conclude that genetic instability is not necessary for fitting the incidence curve of colorectal cancer.

The difference between their approach and ours is that our model includes specific genetic mutations and the effective population sizes of cells where these mutations have to occur. We discuss possible parameter values for the rates of cell division in various stages of tumor progression and the selective advantages or disadvantages of mutated cells. We argue that age incidence curves of cancers can be fitted by many different models with or without genetic instability. For example, we have shown that two rate limiting hits for inactivation of a tumor suppressor gene can be interpreted as two stochastic events inactivating each of the two alleles, or as one stochastic event inactivating the first allele followed by a CIN mutation followed by very fast inactivation (LOH) of the second allele[51]. Both evolutionary trajectories

have two rate limiting steps. Therefore, age incidence curves alone do not provide enough information to decide whether genetic instability is early or not.

Herrero-Jimenez et al.[99] provided a statistical analysis of colorectal cancer incidence data accounting for demographic stochasticity, i.e. human heterogeneity for inherited traits and environmental experiences. They fitted the two-stage initiation-promotion model to the incidence data and calculated adenoma growth rates, the number of initiating mutations and the rate of LOH, which was found to be significantly higher than in normal cells. Pinsky[100] adapted the multi-stage model to the problem of development and growth of adenomas, both in sporadic and inherited colorectal cancers. He fitted the model to adenoma data including adenoma prevalence by age, the size distribution of adenomas, clustering of adenomas within individuals and the correlation between proximal and distal adenomas.

Little and Wright[60] extended the multi-stage model to incorporate genomic instability and an arbitrary number of mutational stages. They found that a model with five stages and two levels of genomic destabilization fits colon cancer incidence data. Yatabe et al.[47] investigated stem cells in the human colon by using methylation patterns and statistics. They found that colonic stem cells follow a stochastic rather than deterministic model, i.e. each stem cell division produces two, one, or zero daughter stem cells and zero, one, or two differentiated cells, respectively. This leads to drift in the number of descendents of each stem cell lineage over time. Zahl [101] developed a proportional regression model for colon cancer in Norway. Mehl[102]

presented a mathematical computer simulation model to predict the natural history of colon polyp and cancer development. He incorporated different cell types and two kinetic processes, mutation and promotion, into his model.

Most mathematical analyses consist of fitting a stochastic model to colorectal cancer incidence data. The number of steps is then extracted from the fitting instead of from cell-biological and genetic assumptions of the necessary number of mutations, rate constants and fitness values. Quantitative estimates, however, are essential to correctly infer the mechanistic sequence of events. Population genetic studies have shown that the effective number of steps observed from the probability distribution of the time to cancer initiation can be smaller than the actual number[2]. This is demonstrated by stochastic tunneling in which two mutations occur in one rate-limiting process[51,53]. The effective step number also depends on the population size and the possibility of somatic selection[2,41].

Colorectal cancer incidence data can be explained by many alternative models. Most models are not easily falsifiable, as the assumptions of intermediate steps and rate constants are not explicitly stated. The identification of the number of rate-limiting steps in a particular model cannot justify the rejection of other interpretations of the incidence data. The assumptions about intermediate steps and rate constants must be identified to render mathematical models testable by experiments. A careful discussion of parameter values and their likely ranges is needed for testing hypotheses. Hence, the specification of intermediate steps and their parameter

values is essential for a full understanding of colorectal tumorigenesis.

Our group has developed a mathematical model of colorectal cancer that investigates the role of genetic instability in tumorigenesis[2,39,51,63,64]. We have studied the question whether a CIN mutation induces the first phenotypic change on the way to colon cancer. We have shown in various models that this is indeed highly probably if there are enough 'CIN genes' in the human genome. We have also shown that tissue organization tremendously influences the probability of emergence of genetic instability[41]. Furthermore, we have developed a specific model including cell-biological and mechanical assumptions of a colonic crypt to describe colorectal tumorigenesis [44].

In the following, we offer a perspective of colorectal tumorigenesis including all sequential mutations needed for tumor initiation and progression and determine the importance of genetic instabilities in colorectal cancer.

4.4 The model

Here, we discuss the evolutionary dynamics of a model of colorectal tumorigenesis including alterations of four genes (Figure 12). Inactivation of the tumor suppressor gene APC initiates tumorigenesis. Subsequent tumor progression is driven by activation of the oncogene RAS. Finally, inactivation of the tumor suppressor gene p53 gives rise to a carcinoma. Somewhere CIN or MIN emerges.

We will analyze the radical hypothesis that CIN precedes inactivation of the

APC tumor suppressor gene. Here, we do not discuss MIN because it occurs only in a minority of sporadic colorectal cancers and is rarely found in other cancers. A theoretical investigation of MIN in sporadic and inherited colorectal cancer is provided elsewhere[64].

The mammalian colon is lined with a rapidly proliferating epithelium. This epithelium is subdivided into 10^7 compartments of cells called crypts. Each crypt contains about 1000 to 4000 cells. Approximately 1 - 10 of those cells are stem cells, residing at the base of each crypt[46,47]. The progeny of stem cells migrate up the crypt, continuing to divide until they reach its mid-portion. Then they stop dividing and differentiate into mature cells. When the cells reach the top of the crypt, they undergo apoptosis and are engulfed by stromal cells or shed into the gut lumen. The cell migration from the base to the top of the crypt takes about 3 - 6 days[103].

In this setting, two scenarios of cellular dynamics within a colonic crypt are possible[2,44]. In the first scenario, only colorectal stem cells are considered to be at risk of becoming cancer cells. Differentiated cells are assumed not to live long enough to accumulate the necessary number of mutations that lead to cancer. Mutated differentiated cells are washed out of the crypt by the continuous production and migration of cells to the top of the crypt where they undergo apoptosis[37]. In this case, the effective population size of each colonic crypt is equal to the number of stem cells. The stem cell compartment is assumed to be well-mixed, i.e. all stem cells reside in equivalent positions and are in direct reproductive competition with each

other. There are no spatial effects. In the second scenario, all cells of the colonic crypt are assumed to be at risk of accumulating mutations that lead to cancer[44]. Mutations in specific genes might confer an increased probability to the cell to stick on top of the crypt instead of undergoing apoptosis. Here, we use the first scenario to discuss the evolutionary dynamics of mutations in APC, RAS, and p53. The second scenario is outlined in the next chapter.

APC inactivation

We will now describe the dynamics of APC inactivation in a well-mixed compartment of cells. Denote the number of stem cells by N_0; stem cells divide every τ_0 days. Denote the probabilities of inactivating the first and second APC allele per cell division by u_1 and u_2, respectively. We have $u_1 < u_2$ because there are more possibilities, such as LOH and mitotic recombination, for the second hit. If the mutation rate is less than the inverse of the population size, then the approximation of homogeneous compartments holds: at any time, all cells of the compartment will have 0, 1 or 2 inactivated alleles of APC. Denote the probabilities that all cells of the compartment have 0, 1 or 2 inactivated alleles of APC at time t by X_0, X_1 and X_2, respectively (Figure 14a). The transition rate between states $APC^{+/+}$ and $APC^{+/-}$ is equal to the rate at which a mutation occurs in the first APC allele times the probability that the mutant cell takes over the compartment. The mutation occurs at rate $N_0 u_1/\tau_0$, because (i) the mutation can arise in any one of N_0 cells, (ii) the rate at which the first APC allele is inactivated per cell division is u_1, and (iii) the

cells divide once every τ_0 days. The probability that the mutated cell takes over the compartment (i.e., reaches fixation) is given by $1/N_0$, if we assume that the mutation is neutral[40]. Thus, the compartment size, N_0, cancels in the product, and we obtain u_1/τ_0 as the transition rate between states $APC^{+/+}$ and $APC^{+/-}$. The transition rate between states $APC^{+/-}$ and $APC^{-/-}$ is equal to the rate that a mutation occurs in the second APC allele times the probability that the mutant cell will take over the compartment. We assume that a cell with two inactivated APC alleles has a large fitness advantage, which means the fixation probability of such a cell is close to one. Therefore, we obtain $N_0 u_2/\tau_0$ as the transition rate between states $APC^{+/-}$ and $APC^{-/-}$ (Figure 14a). The differential equation describing the stochastic process is outlined in the Appendix. The overall probability that APC has been inactivated by time t is approximately given by $X_2(t) = N_0 u_1 u_2 t^2/(2\tau_0^2)$. This probability accumulates as a second order of time, which means that it takes two rate-limiting hits to inactivate APC in the compartment of cells[2,10].

Chromosomal instability (CIN)

CIN could arise at any stage of tumorigenesis. Here we discuss the possibility that CIN emerges before inactivation of APC (Figure 14b). The crucial effect of CIN is to accelerate the inactivation rate of the second APC allele. Denote the probability of inactivating the second APC allele per CIN cell division by u_3; we have $u_3 >> u_2$. This increased rate, however, might confer a selective cost to the CIN cell, because such a cell has an increased probability of lethal mutations and

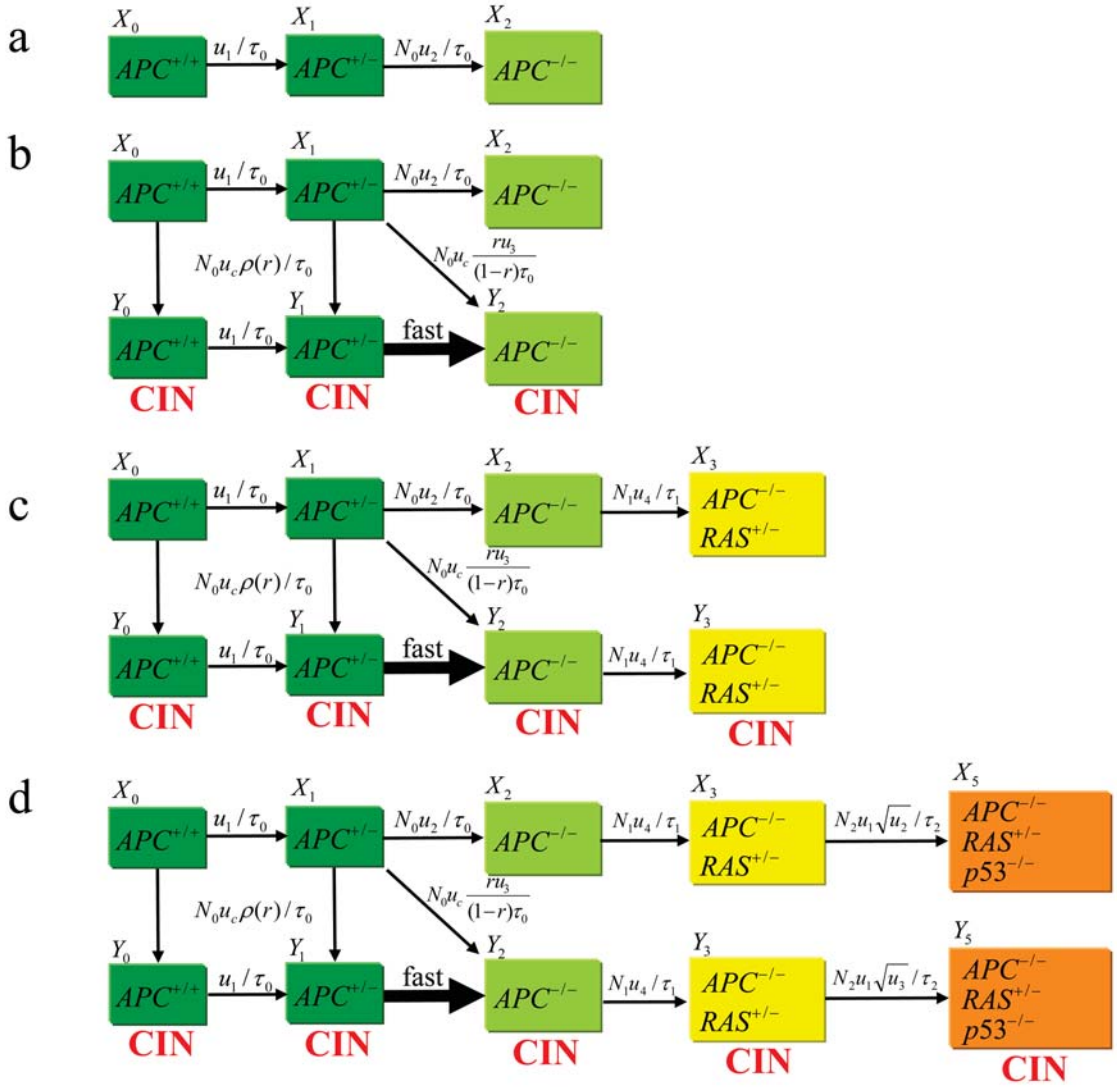

a
b
c
d
X_0
$APC^{+/+}$
u_1/τ_0
X_1
$APC^{+/-}$
N_0u_2/τ_0
X_2
$APC^{-/-}$
$N_0u_c\rho(r)/\tau_0$
$N_0u_c\frac{ru_3}{(1-r)\tau_0}$
Y_0
Y_1
Y_2
fast
CIN
N_1u_4/τ_1
X_3
$APC^{-/-}$
$RAS^{+/-}$
Y_3
$N_2u_1\sqrt{u_2}/\tau_2$
X_5
$APC^{-/-}$
$RAS^{+/-}$
$p53^{-/-}$
$N_2u_1\sqrt{u_3}/\tau_2$
Y_5

Figure 14: Dynamics of colorectal tumorigenesis. (a) Inactivation of APC in a well-mixed compartment of colonic stem cells, N_0, that divide once every τ_0 days. The probabilities of inactivating the first and second APC allele per cell division are denoted by u_1 and u_2. (b) Chromosomal instability (CIN) can arise at any stage of tumorigenesis and causes an accelerated rate of inactivation of the second APC allele; denote this rate by u_3. A CIN cell has relative fitness r, arises at rate u_c per cell division, and reaches fixation in the compartment with probability ρ. If an $APC^{+/-}$ cell clone with CIN produces an $APC^{-/-}$ cell with CIN before taking over the compartment, a stochastic tunnel arises: the compartment evolves from state $APC^{+/-}$ without CIN directly to state $APC^{-/-}$ with CIN without ever visiting state $APC^{+/-}$ with CIN (diagonal arrow). (c) APC inactivation leads to a small lesion of N_1 cells dividing every τ_1 days in which activation of RAS might be the next sequential mutation necessary for colorectal tumorigenesis. The rate at which RAS is mutated per cell division is denoted by u_4. (d) RAS activation leads to an adenoma of N_2 cells dividing every τ_2 days in which inactivation of p53 might be the next sequential mutation. Suppose the p53 alleles are inactivated at the same rates as the APC alleles. Due to the large compartment size, it takes only one rate-limiting hit to inactivate both p53 alleles.

apoptosis. Conversely, it is possible that a CIN cell has a faster cell cycling time because of avoiding certain checkpoints. Denote the relative fitness of a CIN cell by r. If $r = 1$, then CIN is neutral; if $r < 1$, it is disadvantageous, and a CIN cell has a lower growth rate than a cell with stable karyotype; if $r > 1$, then it is advantageous, and a CIN cell has a larger growth rate than a cell with stable karyotype. Denote the probabilities that all cells of the compartment have CIN and 0, 1 or 2 inactivated APC alleles at time t by $Y_0(t)$, $Y_1(t)$ and $Y_2(t)$, respectively. The transition rate between states $APC^{+/+}$ without CIN and $APC^{+/+}$ with CIN is the same as the transition rate between states $APC^{+/-}$ without CIN and $APC^{+/-}$ with CIN and equal to the rate at which a mutation triggering CIN occurs, $N_0 u_c/\tau_0$, times the probability that the mutant cell will take over the compartment. The fixation probability of a CIN cell depends on its somatic fitness, r, and the compartment size, N_0, and is given by $\rho = (1 - 1/r)/(1 - 1/r^{N_0})$. This is the standard fixation probability of a Moran process[39]. Therefore, the transition rate is given by $N_0 u_c \rho/\tau_0$.

If an $APC^{+/-}$ cell clone with CIN produces an $APC^{-/-}$ cell with CIN before taking over the compartment, a stochastic tunnel arises[2,53]: the compartment evolves from state $APC^{+/-}$ without CIN directly to state $APC^{-/-}$ with CIN without ever visiting state $APC^{+/-}$ with CIN. This tunnel occurs at rate $N_0 u_c r u_3/[(1-r)\tau_0]$ (Figure 14b). If CIN is neutral, $r = 1$, then the tunnel does not occur. The transition rate between states $APC^{+/+}$ with CIN and $APC^{+/-}$ with CIN is u_1/τ_0 as before, because CIN is not assumed to cause increased point mutation rates. The transition

rate between states $APC^{+/-}$ with CIN and $APC^{-/-}$ with CIN is given by $N_0 u_3/\tau_0$; this rate is fast in the sense that $N_0 u_3/\tau_0 >> 1$, and hence is not rate-limiting.

The overall probability that APC has been inactivated with CIN by time t is given by $Y_2(t) = N_0 u_1 u_c[2\rho + r u_3/(1-r)]t^2/(2\tau_0^2)$. This probability again accumulates as a second order of time, so it also takes two rate-limiting hits to inactivate APC in a compartment with CIN. Hence, Knudson's two-hit hypothesis[10] is compatible with tumor suppressor gene inactivation both without and with CIN[2].

RAS activation

Inactivation of APC leads to clonal expansion of the colonic stem cell compartment to a small lesion of N_1 cells dividing every τ_1 days. It is in this lesion that other genes must be mutated for further tumor progression. The next genetic event in the sequence might be activation of the oncogene RAS[87].

Let us discuss the activation dynamics of RAS. Denote the probability that a RAS mutation occurs per cell division by u_4. Denote the probabilities that a lesion consists only of $APC^{-/-}RAS^{+/-}$ cells without and with CIN at time t by $X_3(t)$ and $Y_3(t)$, respectively (Figure 14c). The transition rate between states $APC^{-/-}RAS^{+/+}$ and $APC^{-/-}RAS^{+/-}$ without CIN is given by $N_1 u_4/\tau_1$, because (i) the mutation can arise in any one of N_1 cells, (ii) the rate at which the RAS allele is mutated per cell division is u_4, (iii) the cells divide once every τ_1 days, and (iv) a cell with a RAS mutation is assumed to have a large fitness advantage, such that its fixation probability is close to 1. The transition rate between states

$APC^{-/-}RAS^{+/+}$ and $APC^{-/-}RAS^{+/-}$ with CIN is the same as the transition rate between states $APC^{-/-}RAS^{+/+}$ and $APC^{-/-}RAS^{+/-}$ without CIN, because the CIN phenotype does not alter the point mutation rate.

p53 inactivation

Activation of RAS leads to clonal expansion of the compartment to a late adenoma of N_2 cells dividing every τ_2 days. In this adenoma, other genes must be mutated for progression to a carcinoma. The next genetic event in the sequence might be inactivation of the tumor suppressor gene p53[1].

We will now discuss the inactivation dynamics of p53 in the context of our model. Denote the probability of inactivating the first p53 allele per cell division by u_1 and the probabilities of inactivating the second p53 allele per normal and CIN cell division by u_1 and u_3, respectively. Again we have $u_3 >> u_2 > u_1$. Here, we assume that the p53 alleles are inactivated at the same rates as the APC alleles. Denote the probabilities that a lesion consists only of $APC^{-/-}RAS^{+/-}p53^{-/-}$ cells without and with CIN at time t by $X_5(t)$ and $Y_5(t)$, respectively (Figure 14d). The transition rate between states $APC^{-/-}RAS^{+/-}p53^{+/+}$ and $APC^{-/-}RAS^{+/-}p53^{-/-}$ without CIN is $N_2 u_1 \sqrt{u_2}/\tau_2$, because (i) the mutation can arise in any of N_2 cells, (ii) the rate at which the two p53 alleles are inactivated per cell division is $u_1\sqrt{u_2}$ due to the increased compartment size[53], and (iii) the cells divide once every τ_2 days. The transition rate between states $APC^{-/-}RAS^{+/-}p53^{+/+}$ and $APC^{-/-}RAS^{+/-}p53^{-/-}$ with CIN is $N_2 u_1 \sqrt{u_3}/\tau_2$, because the rate at which the two p53 alleles are inac-

tivated per CIN cell division is $u_1\sqrt{u_3}$. The differential equation is given in the Appendix and can be used to calculate the probability as a function of time to develop a carcinoma. We can also determine the role of CIN in tumorigenesis. In section 4.5, we will provide numerical examples of our model.

The cost of CIN

Once a CIN phenotype has reached fixation in a pre-cancer lesion, the question remains how to quantify the further cost of CIN. Comparing two lesions, one without CIN and one with CIN, we might be tempted to argue that the rate of evolution of the CIN lesion has to be multiplied by a factor (given by a number between 0 and 1) which takes into account the increased death rate caused by CIN. Therefore, the rate of evolution of the CIN lesion is reduced by a factor describing the cost of CIN.

On the other hand, it is conceivable that the increased rate of cell death in the CIN lesion allows for an increased number of cell divisions (to make up for the cells that have died) and therefore even accelerates the rate of evolution. A variety of simple mathematical models can be constructed that demonstrate that the 'cost of CIN' can be a net selective advantage when comparing the competition between CIN and non-CIN lesions.

Since both scenarios are possible, here we assume that CIN lesions and non-CIN lesions are neutral with respect to the cost of CIN during further tumor progression. The cost of CIN is only considered when calculating the probability that a CIN cell reaches fixation in the compartment where it first arises.

4.5 Numerical examples

The human colon consists of approximately $Z = 10^7$ crypts, each of which is replenished by about 1-10 stem cells[47,63]. Let us choose $N_0 = 4$. Later, we will discuss numerical examples with $N_0 = 1$, 4 and 10 stem cells. The mutation rate[45] per base per cell division is approximately 10^{-10}. Suppose the APC tumor suppressor gene can be inactivated by point mutations in any of about 500 bases[104]; hence the rate of inactivating the first of two alleles is $u_1 = 2 \times 500 \times 10^{-10} = 10^{-7}$ per cell division both in normal and CIN cells. The rate of inactivating the second allele per cell division, u_2, is the sum of the rate of LOH and the rate of a point mutation. Assume the rate of inactivating the second allele is about $u_2 = 10^{-6}$ per cell division in normal cells. The rate of inactivating the second allele in CIN cells[82] has been measured to be $u_3 = 10^{-2}$.

Assume clonal expansion of the stem cell compartment to $N_1 = 10^3$ cells after APC inactivation and to $N_2 = 10^5$ cells after RAS activation. Suppose normal stem cells divide once every $\tau_0 = 10$ days[47]. Assume that $APC^{-/-}$ cells divide every $\tau_1 = 5$ days and $APC^{-/-}RAS^{+/-}$ cells every $\tau_2 = 1$ day. The most substantial cost CIN can possibly have is $r = (1-u_3)^{45} \approx 0.6$; this means that loss of any chromosome other than the one containing the tumor suppressor gene locus is lethal.

(i) Assume there is only one class I CIN gene that triggers CIN when mutated; then $u_c = 2u_2 = 2 \cdot 10^{-6}$, because either allele can be inactivated. The expected number of carcinomas that were initiated by inactivation of APC in a $t = 70$ year-

old is 0.0014. The expected number of carcinomas that were initiated by CIN in a $t = 70$ year-old is 0.0082 (see Appendix). Hence, 85% of carcinomas were initiated by CIN. Among 1000 people, there are 9 colorectal cancer cases, 8 of which were initiated by CIN. However, all of them might have CIN in the end.

(ii) Assume there is one class II CIN gene that triggers CIN when mutated; then $u_c = 10^{-7}$. The expected number of carcinomas that were initiated by inactivation of APC in a $t = 70$ year-old is 0.0014. The expected number of carcinomas that were initiated by CIN is 0.0004 (see Appendix). Hence, 22% of carcinomas were initiated by CIN.

(iii) Alternatively, we can calculate the minimum number of CIN genes in the human genome needed to make sure CIN arises before the inactivation of APC. Given the parameter values above, one class I CIN gene or four class II CIN genes are necessary to ensure that CIN arises early in colorectal tumorigenesis. Note that there are hundreds of CIN genes in yeast and even larger numbers are expected in humans.

Figure 15 provides further examples of parameter values.

a $N_0 = 1, N_1 = 10^4, N_2 = 10^6$

r	$n_1 = 1$	$n_2 = 10$
1.0	89%	81%
0.8	90%	81%
0.6	90%	81%

b $N_0 = 4, N_1 = 10^4, N_2 = 10^6$

r	$n_1 = 1$	$n_2 = 10$
1.0	70%	54%
0.8	64%	47%
0.6	50%	33%

c $N_0 = 4, N_1 = 10^3, N_2 = 10^5$

r	$n_1 = 1$	$n_2 = 10$
1.0	93%	88%
0.8	92%	85%
0.6	86%	75%

d $N_0 = 10, N_1 = 10^3, N_2 = 10^5$

r	$n_1 = 1$	$n_2 = 10$
1.0	85%	74%
0.8	74%	59%
0.6	40%	25%

Figure 15: Percentage of carcinomas that were initiated by chromosomal instability (CIN), if there is $n_1 = 1$ class I CIN gene or $n_2 = 10$ class II CIN genes in the human genome. Class I CIN genes, n_1, trigger CIN if one allele is mutated or lost. Class II CIN genes, n_2, trigger CIN if one allele is mutated in a dominant negative fashion. The relative fitness of CIN cells is denoted by r. Parameter values are $u_1 = 10^{-7}$, $u_2 = 10^{-6}$, $u_3 = 10^{-2}$, $u_4 = 10^{-9}$, $\tau_0 = 10$, $\tau_1 = 5$, $\tau_2 = 1$, $t = 70$ years, and (a) $N_0 = 1$, $N_1 = 10^4$ and $N_2 = 10^6$, (b) $N_0 = 4$, $N_1 = 10^4$ and $N_2 = 10^6$, (c) $N_0 = 4$, $N_1 = 10^3$ and $N_2 = 10^5$, and (d) $N_0 = 10$, $N_1 = 10^3$ and $N_2 = 10^5$.

4.6 Discussion

Here, we have discussed the mutational sequence that leads to colorectal cancer. The traditional model of colorectal tumorigenesis[87] describes several genetic changes that are required for cancer initiation and progression (Figure 12). However, no large scale sequencing of genes like APC or p53 has been done. An important goal of the field should be detailed experimental analyses of the mutations found in colorectal tumors.

We have provided a review of the literature of mathematical and statistical analyses of colon cancer. Several excellent statistical studies have been performed to analyze colorectal cancer incidence data. Stochastic multi-stage and population genetic models have been developed and been fitted to incidence data. Although this effort determined the number of stages needed for initiation and promotion of colon cancer, the biological implications are unclear; biologically and clinically useful investigation of colorectal tumorigenesis must include the exact specification of individual steps. Experimental input of parameter values such as mutation rates, fitness values and rate of population growth are required for a quantitative understanding of cancer progression.

We have reviewed our mathematical approach to colon cancer by means of a specific model of colorectal tumorigenesis. We have assumed that colorectal tumorigenesis is driven by subsequent genetic alterations of APC, RAS, and p53. Naturally, the model describes only a subset of cancers and does not attempt to provide a uni-

versally found sequence of mutations leading to colorectal cancer.

On the basis of this case study, we developed a mathematical representation of the evolutionary dynamics of colorectal tumorigenesis. We specified the dependence of the transition rates between stages on mutation rates, population sizes, fitness values and cell division times. We introduced mutation of chromosomal instability (CIN) genes that lead to an increased rate of loss of heterozygosity (LOH), thereby accelerating the inactivation of tumor suppressor genes. We investigated the role of CIN in initiating and driving tumor progression in dependence of the number of genes that can initiate CIN when mutated in colorectal tissue. Given the experimental input of rate constants and parameter values, we found that one or two CIN genes in the genome are enough to make sure CIN emerges early. Therefore, it is very likely that a CIN mutation initiates colorectal tumorigenesis.

There is the possibility of a second tumor suppressor gene that needs to be genetically altered after inactivation of APC and before activation of RAS. In that case and if both tumor suppressor genes are inactivated in rate-limiting situations, i.e. if inactivation of APC alone does not lead to a large tumor, then chromosomal instability it very likely to initiate the mutational sequence that leads to colorectal cancer.

4.7 Appendix

The stochastic process illustrated in Figure 14 can be described by differential equations. Denote the probabilities that a compartment is in state $APC^{+/+}$, $APC^{+/-}$, $APC^{-/-}$, $APC^{-/-}RAS^{+/-}$, and $APC^{-/-}RAS^{+/-}p53^{-/-}$ without CIN by X_0, X_1, X_2, X_3, and X_5, respectively. Denote the probabilities that a compartment is in the corresponding states with CIN by Y_0, Y_1, Y_2, Y_3, and Y_5, respectively. The differential equations are given by

$$\begin{aligned}
\dot{X}_0 &= -(a+b)X_0 & \dot{Y}_0 &= bX_0 - aY_0 \\
\dot{X}_1 &= aX_0 - (b+c+d)X_1 & \dot{Y}_1 &= aY_0 + bX_1 - gY_1 \\
\dot{X}_2 &= cX_1 - eX_2 & \dot{Y}_2 &= gY_1 + dX_1 - eY_2 \\
\dot{X}_3 &= eX_2 - fX_3 & \dot{Y}_3 &= eY_2 - hY_3 \\
\dot{X}_5 &= fX_3 & \dot{Y}_5 &= hY_3
\end{aligned} \tag{5}$$

Here $a = u_1/\tau_0$, $b = N_0 u_c \rho(r)$, $c = N_0 u_2/\tau_0$, $d = N_0 u_c r u_3/[(1-r)\tau_0]$, $e = N_1 u_4/\tau_1$, $f = N_2 u_1\sqrt{u_2}/\tau_2$, $g = N_0 u_3/\tau_0$, and $h = N_2 u_1\sqrt{u_3}/\tau_2$. See the main text for an explanation of the parameter values. Equation 5 is a system of linear differential equations which can be solved analytically using standard techniques. The numerical examples contain numerical simulation of system 5.

5 Linear model of colon cancer initiation

Michor F, Iwasa Y, Rajagopalan H, Lengauer C, Nowak MA

Linear model of colon cancer initiation[44]

Cell Cycle 3, 358-362 (2004)

Cancer results if regulatory mechanisms of cell birth and death are disrupted. Colorectal tumorigenesis is initiated by somatic or inherited mutations in the APC tumor suppressor gene pathway. Several additional genetic hits in other tumor suppressor genes and oncogenes drive the progression from polyps to malignant, invasive cancer. The majority of colorectal cancers present chromosomal instability, CIN, which is caused by mutations in genes that are required to maintain chromosomal stability. A major question in cancer genetics is whether CIN is an early event and thus a driving force of tumor progression. We present a new mathematical model of colon cancer initiation assuming a linear flow from stem cells to differentiated cells to apoptosis. We study the consequences of mutations in different cell types and calculate the conditions for CIN to precede APC inactivation. We find that early emergence of CIN is very likely in colorectal tumorigenesis.

5.1 Introduction

The mammalian colon is lined with a rapidly proliferating epithelium. This epithelium is organized into compartments of cells called crypts. Intestinal stem cells reside at the base of each crypt[46]. The progeny of stem cells migrate up the crypt, continuing to divide until they reach its mid-portion. Then they stop dividing and differentiate to mature cells. When the cells reach the top of the crypt, they undergo apoptosis and are engulfed by stromal cells or shed into the gut lumen. The cell migration from the base to the top of the crypt takes about 3 - 6 days[105]. In normal crypts, the rate of cellular death balances the rate of cellular proliferation. If this homeostatic mechanism is shifted toward cellular growth, neoplasia results.

Colorectal tumors progress through four distinct clinical stages that describe dysplastic crypts, small benign tumors, malignant tumors invading surrounding tissues, and finally metastatic cancer. This progression involves several genetic changes such as inactivation of tumor suppressor genes and activation of oncogenes[87]. Mutations of the adenomatous polyposis coli (APC) gene are considered the earliest[88] and most prevalent genetic changes in colorectal tumorigenesis. More than 95% of colon cancers are estimated to have a somatic mutation of APC[1]. APC is a tumor suppressor gene on chromosome 5q21. The APC gene product plays an important role in modulating the Wnt/β-catenin signal transduction pathway that regulates, among several other genes, the transcription of the oncogene c-Myc[90,91]. APC has been shown to reduce net cellular proliferation through an increased rate of apopto-

sis[106]. Most APC mutations found in cancer cells lead to a truncation of the protein, thereby preventing its regulating function. An inherited APC germ line mutation leads to the development of hundreds to thousands of benign colorectal tumors, some of which progress to malignancy (familial adenomatous polyposis, FAP)[107].

Genetic instability is a defining characteristic of most human cancers. Two forms of genetic instability have been identified: chromosomal instability, CIN, and microsatellite instability, MIN[54]. About 85% of sporadic colorectal cancers and most other solid tumors have CIN. CIN refers to an increased rate of losing or gaining whole chromosomes or large parts of chromosomes during cell division. It also increases the rate of loss of heterozygosity (LOH). LOH is thought to be an important property of CIN, because tumor suppressor genes have to be inactivated in both alleles. CIN accelerates the rate of inactivation of tumor suppressor genes.

The less prevalent form of genetic instability, MIN, is caused by alterations in mismatch repair genes[54]. These alterations lead to an approximately 1000-fold increased point mutation rate in MIN cancer cells. No increased rate of gross chromosomal changes is observed. MIN occurs in about 15% of sporadic colorectal cancers, but seldom happens in other cancer types. CIN and MIN are generally mutually exclusive.

A major question in cancer genetics is to what extent CIN, or any genetic instability, is an early event and thus a driving force of tumorigenesis[2,23,39,41,51,59,60,64,79]. The precise timing of the emergence of CIN is difficult to measure experimentally,

as the molecular basis of CIN is still poorly understood. In this paper, we develop a mathematical approach for the cellular dynamics of colon cancer initiation and determine the importance of early CIN.

In earlier work, we described each colon crypt as a well-mixed pool of (stem) cells that are at risk of receiving APC and CIN mutations[2,39,41,51,64]. All cells in this pool were assumed to be in equivalent positions and in direct reproductive competition with each other. There were no spatial effects. All cells in the pool were assumed to have the same rate of cell division and the same probability of mutation. Mutations could only occur in cells within the pool. Mutations in other cells of the colonic crypt were ignored. Here we study a model assuming a simple linear geometry of cellular arrangement, division, differentiation, and death[43]. We allow for the possibility that stem cells and differentiated cells divide at different rates. And we explore the consequences of mutations occurring in stem cells and differentiated cells.

5.2 Mechanism

The human colon is subdivided into $M \approx 10^7$ crypts. Each crypt contains about 1000 to 4000 cells; maybe 1-10 of those cells are stem cells[47]. Stem cells reside at the base of the crypt and divide asymmetrically to produce differentiated offspring (Figure 16). The differentiated cells migrate toward the top of the crypt, where they eventually undergo apoptosis and are shed into the gut[108].

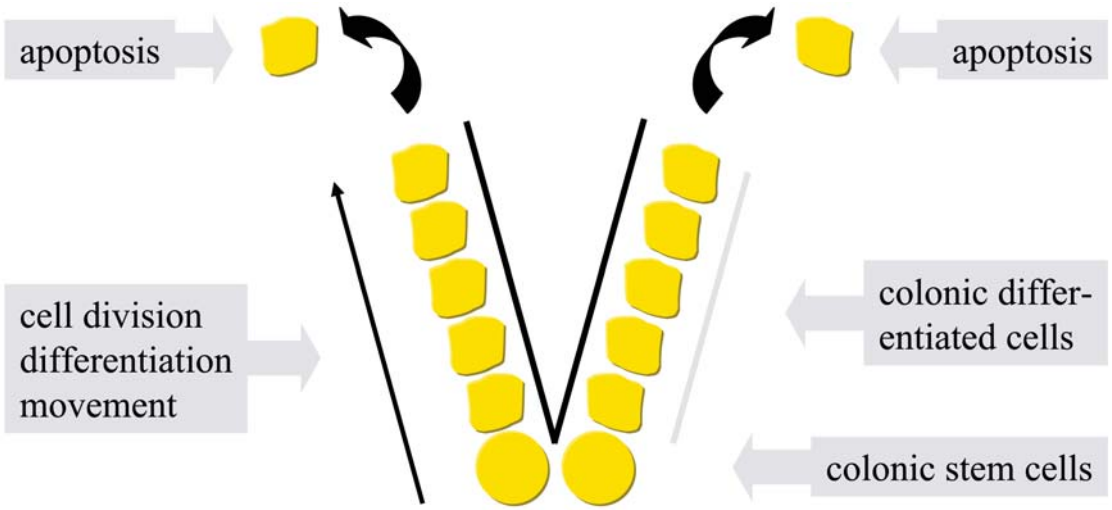


Figure 16: A colonic crypt. Colonic stem cells reside at the base of the crypt and divide asymmetrically to produce differentiated progeny. The progeny migrate toward the top of the crypt where they undergo apoptosis and are shed into the gut lumen or engulfed by stromal cells.

In this setting, different scenarios of tumorigenesis are plausible (Figures 17 to 19). The 'traditional' view of colorectal tumor initiation does not involve CIN (Figure 17). Initially, the crypt is unmutated and contains only wild type cells, $APC^{+/+}$ without CIN. The inactivation of the first APC allele has to happen in the stem cell in order to take over the whole crypt. If the mutation happens in a differentiated cell, the mutated cell will be pushed to the top of the crypt by the continuous production and migration of differentiated cells and will be 'washed out' of the crypt[12,14,37].

Once the mutation has happened in the stem cell, the stem cell produces mutated progeny that eventually populates the whole crypt. The crypt moves from state $APC^{+/+}$ to state $APC^{+/-}$. The inactivation of the remaining APC allele might take place in any cell. The inactivation of both APC alleles confers a certain probability to the cell to stick on top of the crypt instead of undergoing apoptosis. There are two possibilities: (i) the cell fails to stick on top of the crypt, undergoes apoptosis, and the crypt stays in state $APC^{+/-}$, or (ii) the cell succeeds to stick on top of the crypt and initiates dysplastic growth. We call this state $APC^{-/-}$.

CIN can emerge before the inactivation of either APC allele. Figures 18 and 19 show two different mutational pathways in which the CIN mutation precedes the inactivation of the second APC allele. The cell lineage is in state $APC^{+/-}$ without CIN. Assume the mutation triggering CIN has to happen in the stem cell (Figure 18). Then the mutated stem cell produces progeny that eventually populate the whole

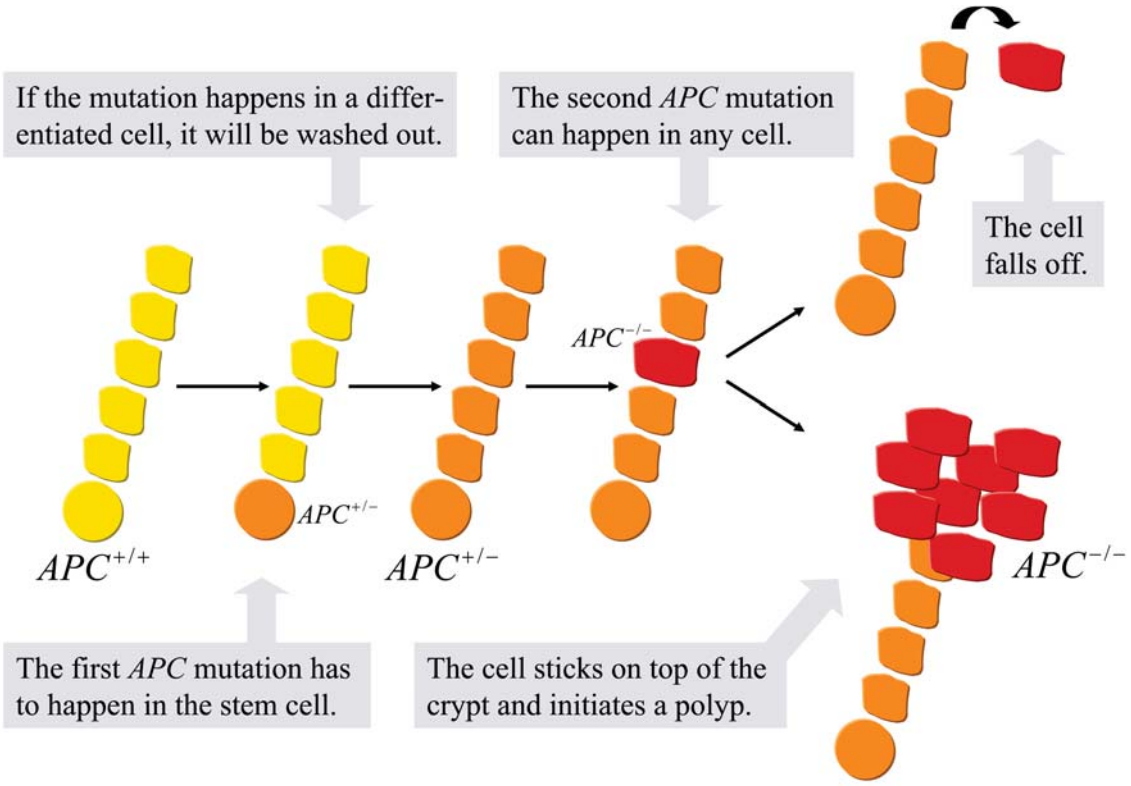


Figure 17: Colorectal tumorigenesis is thought to be initiated by the inactivation of the tumor suppressor gene APC. Initially, all cells are wild type, $APC^{+/+}$. The inactivation of the first APC allele has to happen in the stem cell; otherwise the mutated cells are washed out of the crypt. The stem cell produces mutated progeny that eventually populate the whole crypt. The crypt moves to state $APC^{+/-}$. The inactivation of the second APC allele can happen in any cell. Inactivation of both APC alleles might enable the cell to stick on top of the crypt, evade apoptosis, and initiate dysplastic growth. This state is called $APC^{-/-}$.

crypt, and the cell lineage moves to state $APC^{+/-}$ with CIN. The crucial effect of CIN is to increase the rate of LOH, thereby greatly accelerating the inactivation of the second APC allele. Thus, the inactivation of the second allele can happen in any cell. If a cell with two inactivated alleles of APC succeeds to stick on top of the crypt, dysplastic growth is initiated.

Assume the CIN mutation happens in a differentiated cell (Figure 19). The CIN cell produces mutated progeny that populates only part of the crypt. The CIN cells will eventually be washed out of the crypt, unless the second APC allele is inactivated in a CIN cell on its way up the crypt. If this cell then succeeds to stick on top of the crypt, it initiates dysplastic growth.

Figure 20 shows all pathways of how a crypt consisting of wild type cells can mutate into a dysplastic crypt. A mathematical analysis of the model reveals the transition rates between different mutational states and the probabilities of APC inactivation without and with CIN. These probabilities specify the importance of early CIN and determine the minimum number of CIN genes in the human genome needed to ensure that CIN arises before the inactivation of APC.

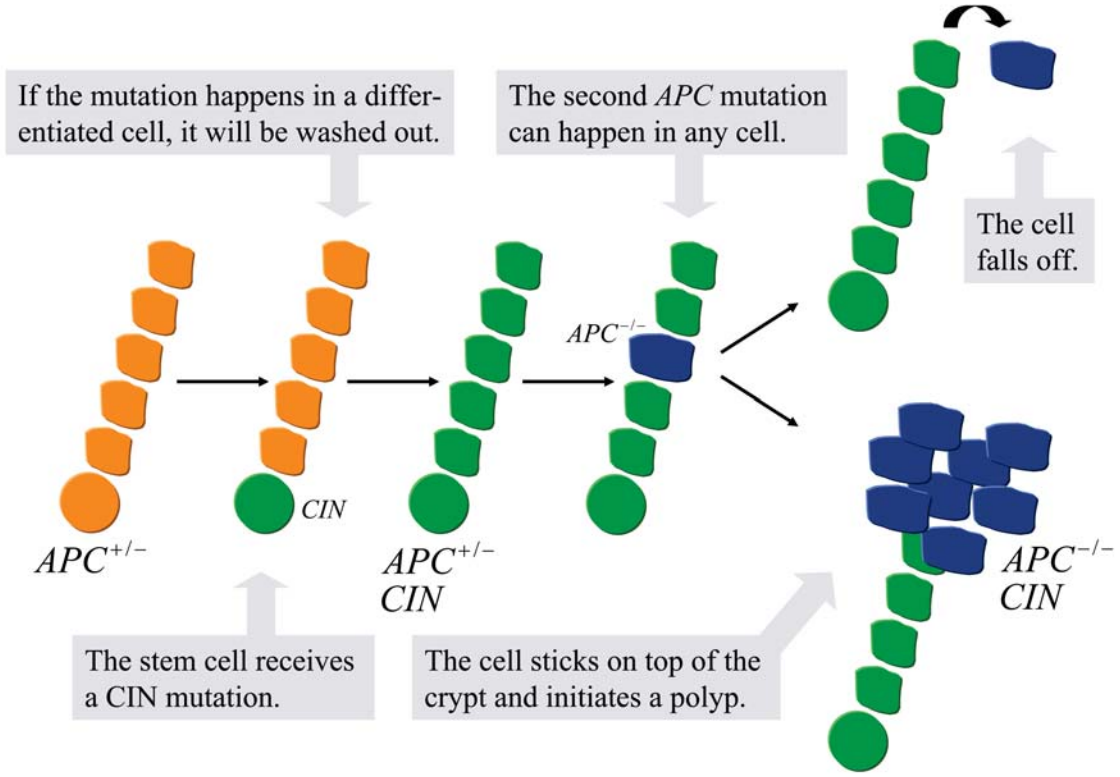


Figure 18: Colorectal tumorigenesis with CIN emerging in the stem cell of an $APC^{+/-}$ cell lineage. The stem cell produces mutated progeny that eventually populate the whole crypt. The crypt moves to state $APC^{+/-}$ with CIN. The inactivation of the second APC allele can happen in any cell. This cell might succeed to stick on top of the crypt and initiate a polyp. This state is called $APC^{-/-}$ with CIN.

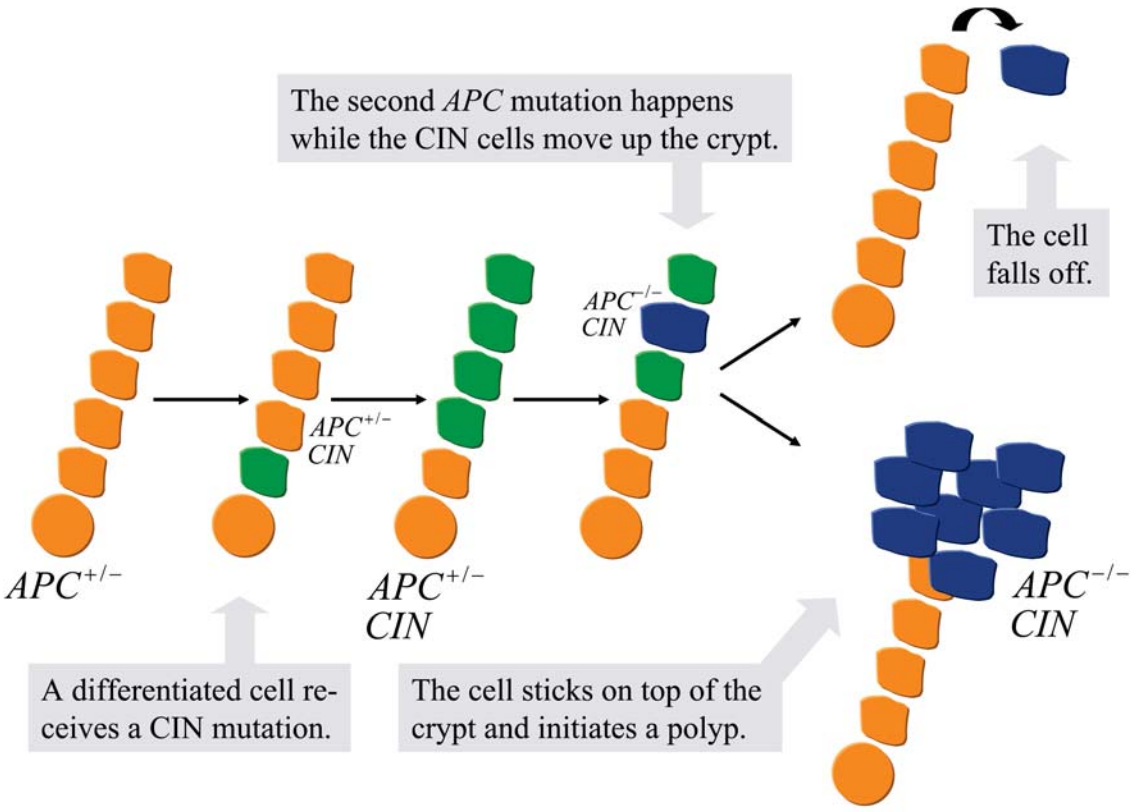


Figure 19: Colorectal tumorigenesis with CIN emerging in a differentiated cell of an $APC^{+/-}$ cell lineage. The mutated cell produces progeny that populate part of the crypt. The CIN cells will eventually be washed out of the crypt, unless the inactivation of the second APC allele happens in a CIN cell while it moves up the crypt. This cell might succeed to stick on top of the crypt and initiate a polyp. The rest of the crypt is repopulated by $APC^{+/-}$ cells without CIN. This state is called $APC^{-/-}$ partly with CIN.

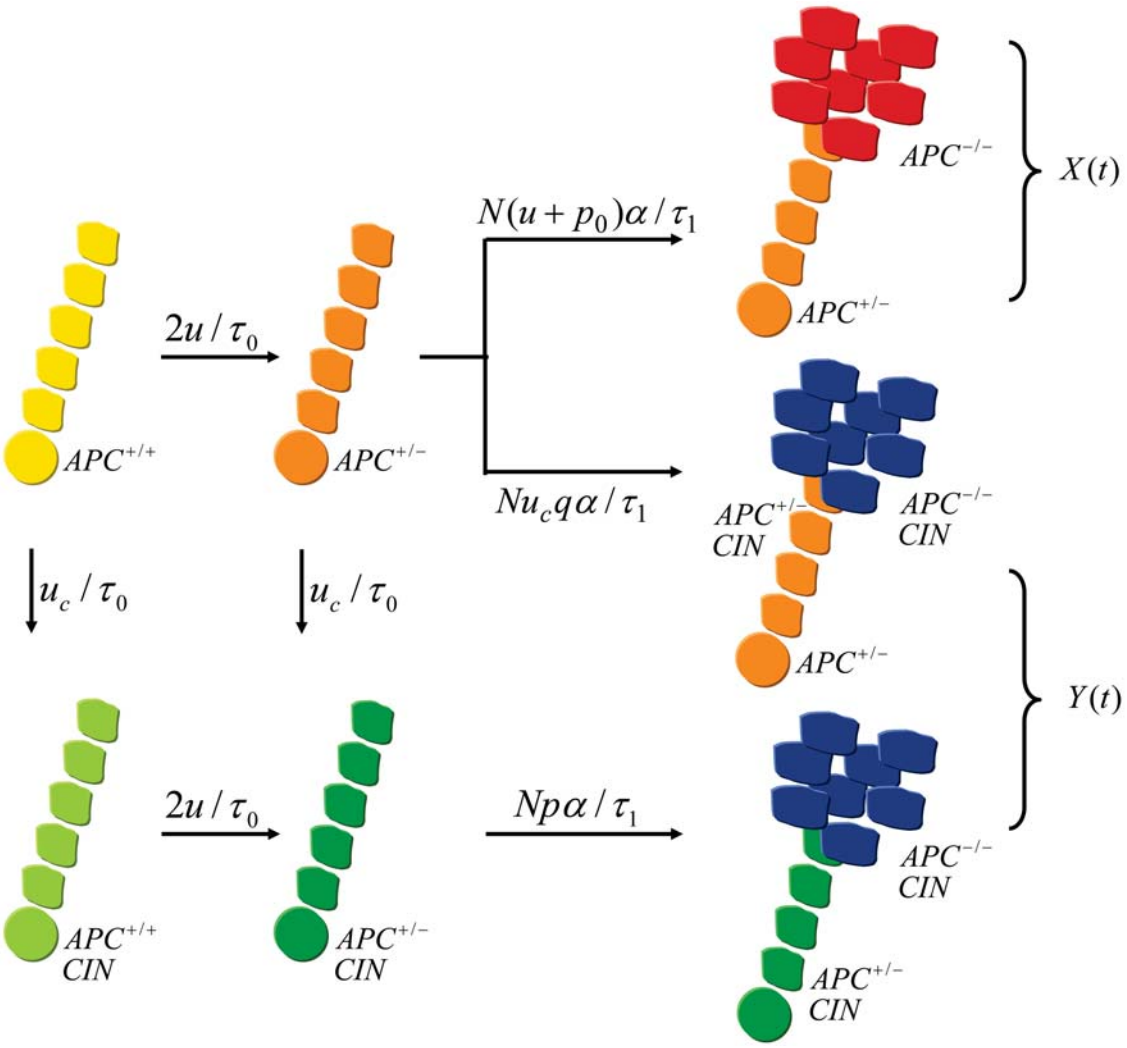

$APC^{+/+}$
$2u/\tau_0$
$APC^{+/-}$
$N(u+p_0)\alpha/\tau_1$
$Nu_cq\alpha/\tau_1$
$APC^{-/-}$
$APC^{+/-}$
$X(t)$
u_c/τ_0
u_c/τ_0
$APC^{+/-}$ CIN
$APC^{-/-}$ CIN
$APC^{+/-}$
$Y(t)$
$APC^{+/+}$ CIN
$2u/\tau_0$
$APC^{+/-}$ CIN
$Np\alpha/\tau_1$
$APC^{-/-}$ CIN
$APC^{+/-}$ CIN

Figure 20: Pathways to dysplasia. Colorectal tumorigenesis might be initiated without CIN. Then the crypt evolves from $APC^{+/+}$ via $APC^{+/-}$ to $APC^{-/-}$ without CIN. CIN can emerge before the inactivation of either APC allele. If CIN arises in the stem cell, the crypt evolves from $APC^{+/+}$ either via $APC^{+/+}$ with CIN or via $APC^{+/-}$ without CIN to $APC^{+/-}$ with CIN, and subsequently to $APC^{-/-}$ with CIN. If CIN arises in a differentiated cell, the crypt evolves from $APC^{+/+}$ via $APC^{+/-}$ without CIN to $APC^{-/-}$ partly with CIN. The transition rates are outlined in the text and specify the probabilities of tumor initiation without and with CIN.

5.3 Mathematical Analysis

The mathematical and statistical analysis of cancer progression using stochastic processes has a long tradition[11,18,23]. Here we develop a new approach for studying cancer initiation by describing mutational events and cellular reproduction in well-defined populations of stem cells and differentiated cells.

The mutational network of colorectal tumorigenesis is illustrated in Figure 20. First, consider tumor initiation without CIN. The inactivation of the first APC allele has to happen in the stem cell. The crypt moves at rate $2u/\tau_0$ from state $APC^{+/+}$ to state $APC^{+/-}$: the stem cell divides every τ_0 days; the mutation rate per APC allele per cell division is denoted by u; and there are two APC alleles. The inactivation of the second APC allele can happen either in the stem cell or in a differentiated cell. The crypt moves at rate $(u + p_0)[1/\tau_0 + N\alpha/\tau_1]$ from state $APC^{+/-}$ to state $APC^{-/-}$: the rate of LOH is denoted by p_0; there are N differentiated cells in the crypt; they divide every τ_1 days; and the probability that a cell with two inactivated APC alleles sticks on top of the crypt and initiates dysplastic growth is denoted by α. For cell-biological reasons, the second APC mutation might have to occur in the stem cell; in that case, we can set $\alpha = 0$. The probability that a crypt is dysplastic without CIN at time t is given by

$$X(t) = \frac{2u}{\tau_0}\Big(\frac{u + p_0}{\tau_0} + \frac{N(u + p_0)\alpha}{\tau_1}\Big)\frac{t^2}{2} \tag{6}$$

CIN can emerge before the inactivation of either APC allele. Suppose the mutation triggering CIN has to occur in the stem cell. Then the crypt moves at rate

u_c/τ_0 from state $APC^{+/+}$ or $APC^{+/-}$ without CIN to state $APC^{+/+}$ or $APC^{+/-}$ with CIN: the probability that a CIN mutation arises per cell division is denoted by u_c. Both normal and CIN crypts move at rate $2u/\tau_0$ from state $APC^{+/+}$ to state $APC^{+/-}$. The inactivation of the second APC allele can happen in any cell, either in the stem cell or in a differentiated cell. The crypt moves at rate $p(1/\tau_0 + N\alpha/\tau_1)$ from state $APC^{+/-}$ to state $APC^{-/-}$: the rate of LOH in CIN cells is denoted by p; we have $p >> p_0$. This rate is fast in the sense that $p(1/\tau_0 + N\alpha/\tau_1) >> 1$. If the second APC mutation has to occur in the stem cell, we again set $\alpha = 0$. The probability that a crypt is dysplastic and contains only CIN cells at time t is given by $Y_1(t) = 2uu_ct^2/\tau_0^2$. This probability contains both the case in which the CIN mutation precedes the first APC mutation and the case in which the CIN mutation precedes the second APC mutation.

Suppose the mutation triggering CIN occurs in a differentiated cell. Then the crypt moves at rate $Nu_cq\alpha/\tau_1$ from state $APC^{+/-}$ without CIN to state $APC^{-/-}$ partly with CIN: the expected probability that a CIN cell encounters an LOH event on its way up the crypt is $q = 1-[1-2^n(1-p)^n]/[(2^n-1)(1-2(1-p))]$; the maximum number of cell divisions a differentiated cell undergoes before being shed off the top of the crypt is denoted by n. If the cell with two inactivated APC alleles succeeds to stick on top of the crypt, the $APC^{+/-}$ stem cell repopulates the remainder of the crypt. The probability that a crypt is dysplastic and partly consists of CIN cells at time t is given by $Y_2 = Nuu_cq\alpha t^2/(\tau_0\tau_1)$. The overall probability that a crypt is

dysplastic with CIN at time t is given by

$$Y(t) = Y_1(t) + Y_2(t) = \frac{2u}{\tau_0}\Big(\frac{2u_c}{\tau_0} + \frac{Nu_cq\alpha}{\tau_1}\Big)\frac{t^2}{2} \tag{7}$$

CIN precedes the inactivation of APC if $Y(t) > X(t)$. The inequality specifies the minimum number of CIN genes needed to make sure CIN emerges before the inactivation of APC. Suppose there are n_1 class I genes that trigger CIN if one allele is mutated or lost, and n_2 class II genes that trigger CIN if one allele is mutated; then the probability that a CIN mutation arises per cell division is given by $u_c = 2n_1(u + p_0) + 2n_2u$. Either $n_1 > (N\alpha\tau_0 + \tau_1)[2Nq\alpha\tau_0 + 4\tau_1]^{-1}$ class I CIN genes or $n_2 > (u+p_0)(N\alpha\tau_0+\tau_1)[2u(Nq\alpha\tau_0+2\tau_1]^{-1}$ class II CIN genes are sufficient to make sure that CIN arises before the inactivation of APC.

5.4 Numerical Examples

The human colon contains about $M = 10^7$ crypts. Each crypt consists of approximately $N = 1000$ differentiated cells[47]. The mutation rate per APC allele per cell division is $u = 10^{-7}$: the mutation rate per base per cell division[45] is about 10^{-10}, and we assume an APC allele can be inactivated by point mutations in approximately 1000 bases. The rate of LOH in normal cells, p_0, might be of the order of u. The rate of LOH in CIN cells[82] is $p = 0.01$. Assume stem cells divide every $\tau_0 = 10$ days and differentiated cells divide every $\tau_1 = 1$ day. The maximum number of cell divisions a colonic cell undergoes is $n \approx 10$, because ten cell divisions are sufficient to populate a crypt of about $N = 1000$ cells (Figure 21). Hence, the probability

that a CIN cell with one inactivated APC allele encounters an LOH event in the second APC allele on its way up the crypt is $q \approx 0.01$. Note that the probability α that a cell with two inactivated APC alleles sticks on top of the crypt cannot be too high, otherwise too many dysplastic crypts develop.

(i) Assume $\alpha = 0.001$. Then a 70-year old has about 14 dysplastic crypts without CIN and 5 dysplastic crypts with CIN if there is $n_1 = 1$ class I CIN gene, or 3 dysplastic crypts with CIN if there is $n_2 = 1$ class II CIN gene.

(ii) Assume $\alpha = 0.01$. Then a 70-year old has about 132 dysplastic crypts without CIN and 8 dysplastic crypts with CIN if there is $n_1 = 1$ class I CIN gene, or 4 dysplastic crypts with CIN if there is $n_2 = 1$ class II CIN gene.

(iii) We can calculate the minimum number of CIN genes needed to ensure that CIN emerges before the inactivation of APC. If $\alpha = 0.01$, then $n_1 = 17$ class I CIN genes or $n_2 = 34$ class II CIN genes are needed. If $\alpha = 0.01$, then $n_1 = 3$ class I CIN genes or $n_1 = 6$ class II CIN genes are needed. If $\alpha = 0$, i.e. if the inactivation of the second APC allele has to happen in the stem cell, then $n_1 = 1$ class I CIN gene or $n_2 = 1$ class II CIN gene is needed to ensure that CIN emerges before the inactivation of APC. Figure 22 shows the number of class I CIN genes, n_1, needed for a 10%, 50%, and 90% probability that CIN emerges before the inactivation of APC in dependence of α.

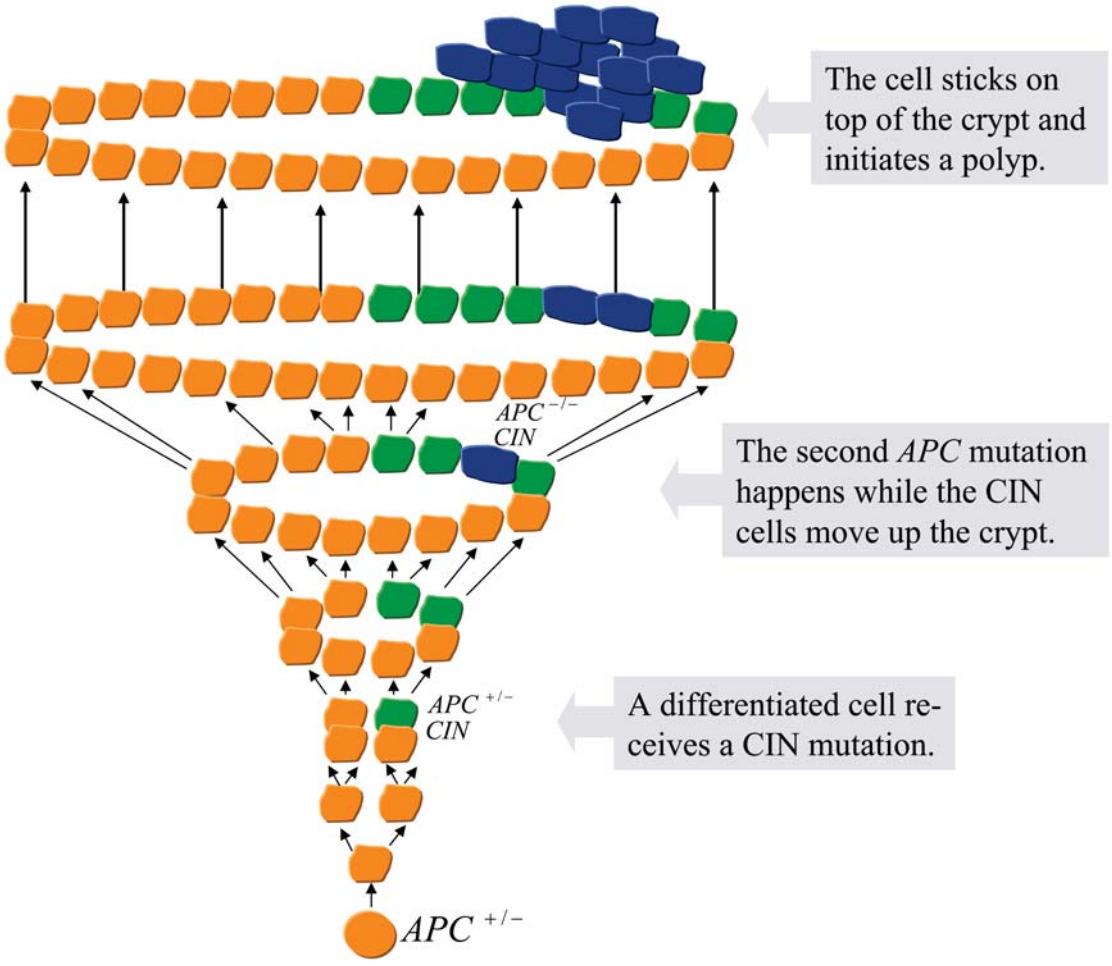


Figure 21: Cell division dynamics in the colonic crypt. Ten cell divisions suffice to populate the whole crypt of about 1000 differentiated cells. After having completed ten cell divisions, differentiated cells cease to proliferate, are pushed to the top of the crypt, and undergo apoptosis. The cell division dynamics specify the probability that a differentiated CIN cell receives the second APC mutation while moving up the crypt.

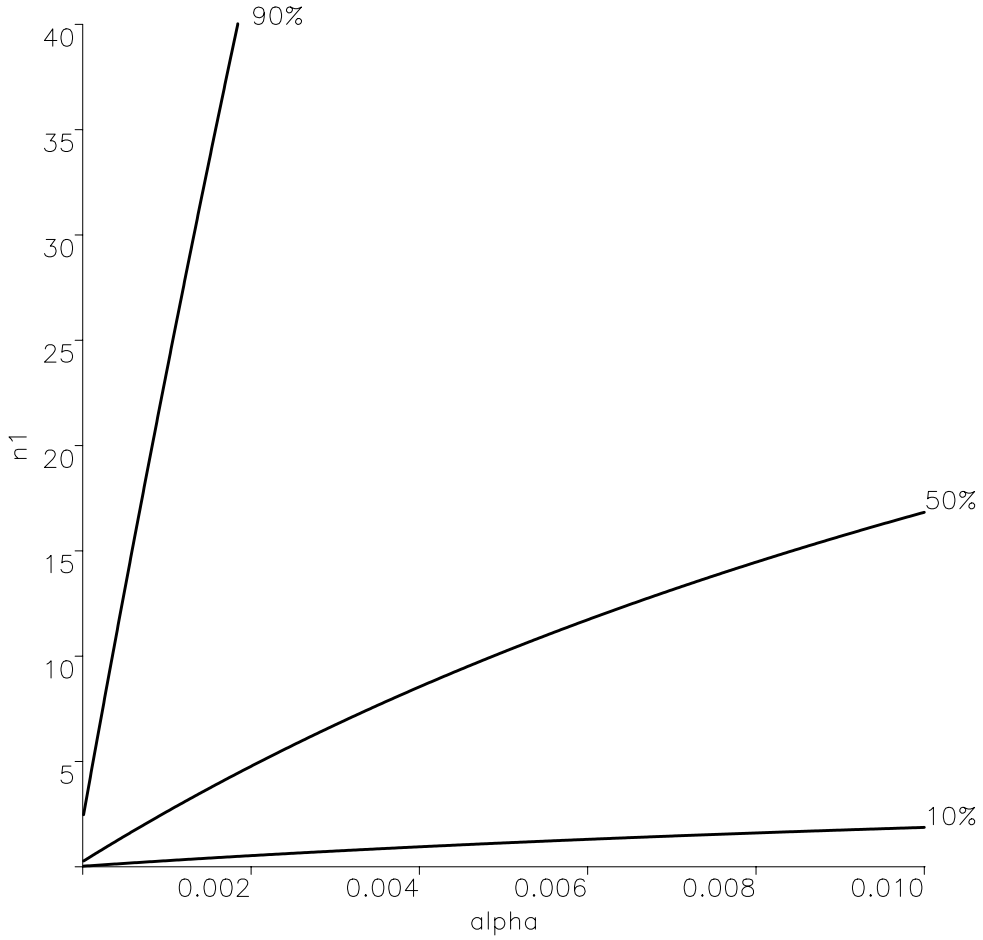


Figure 22: Number of class I CIN genes, n_1, needed for a 10%, 50%, and 90% probability that CIN emerges before the inactivation of APC, in dependence of the probability α that a cell with two inactivated APC alleles, $APC^{-/-}$, sticks on top of the crypt and initiates dysplastic growth. Parameter values are $N = 1000$, $\tau_0 = 10$, $\tau_1 = 1$, $q = 0.01$, and $t = 70$ years.

5.5 Discussion

In this paper, we develop a new mathematical approach for cellular dynamics, tissue architecture, and differentiation patterns of colonic crypts. The model is based on simple genetic and mechanical assumptions. A colonic stem cell gives rise to a lineage of differentiated cells that move up the crypt and undergo apoptosis on top of the crypt. Stem cells are assumed not to be subject to somatic selection; each stem cell and its offspring can be described as an independent cell lineage. All of these cells can receive mutations inactivating one or both alleles of the tumor suppressor gene APC, which is thought to initiate tumorigenesis. The cells can also receive mutations in genes that give rise to chromosomal instability, CIN. CIN might have a cost as it increases the chance of lethal mutations and apoptosis. If CIN causes the stem cell to divide more slowly, the stem cell still populates the crypt with mutated progeny. If CIN causes the stem cell to undergo apoptosis, the crypt dies out and might be replaced by a neighbor crypt.

We study the effect of various mutations that arise in stem cells or differentiated cells. Some mutations, such as the inactivation of the first APC allele, have to happen in the stem cell; otherwise the mutated cells are pushed to the top of the crypt and 'washed out' by the continuous production and migration of wild type differentiated cells. Other mutations, such as the inactivation of the second APC allele, can also happen in a differentiated cell. The inactivation of both APC alleles enables the cell to stick on top of the crypt, evade apoptosis, and initiate dysplastic

growth.

A major question in cancer genetics is whether genetic instability is an early event and thus a driving force of tumorigenesis. The present paper shows how to calculate the rate of colon cancer initiation and the conditions for chromosomal instability to precede the inactivation of APC. For plausible parameter values, we obtain that a small number of CIN genes is sufficient to ensure that CIN emerges before the inactivation of APC. CIN, however, accelerates the rate of inactivation of every successive tumor suppressor gene needed for further tumor progression. If colorectal tumorigenesis requires the inactivation of two or more tumor suppressor genes in rate limiting situations, then even fewer CIN genes are needed to ensure that CIN arises before the inactivation of the first tumor suppressor gene[63]. It is therefore very likely that a CIN mutation initiates colorectal tumorigenesis.

6 Dynamics of chronic myeloid leukemia

Michor F, Hughes TP, Iwasa Y, Branford S,

Shah NP, Sawyers CL, Nowak MA

Dynamics of chronic myeloid leukemia[109]

Nature, in press (2005)

Chronic myeloid leukemia (CML) is associated with the oncogenic fusion gene BCR-ABL. The clinical success of the ABL tyrosine kinase inhibitor imatinib in CML serves as a model for molecularly targeted therapy of cancer, but at least two critical questions remain. Can imatinib eradicate leukemic stem cells? What are the dynamics of relapse due to imatinib resistance, which is caused by mutations in the ABL kinase domain? The precise understanding of how imatinib exerts its therapeutic effect in CML and the ability to measure disease burden by quantitative PCR provide an opportunity to develop a mathematical approach. We find that a 4-compartment model, based on the known biology of hematopoietic differentiation, can explain the kinetics of the molecular response to imatinib in a 169 patient dataset. Successful therapy leads to a bi-phasic exponential decline of leukemic cells. The first slope of 0.05 per day represents the turnover rate of differentiated leukemic cells,

while the second slope of 0.008 per day represents the turnover rate of leukemic progenitors. The model suggests that imatinib is a potent inhibitor of the production of differentiated leukemic cells, but does not deplete leukemic stem cells. We calculate the probability of developing imatinib resistance mutations and estimate the time until detection of resistance. Our model provides the first quantitative insights into the in-vivo kinetics of a human cancer.

6.1 Introduction

Peripheral blood consists of terminally differentiated hematopoietic cells such as erythrocytes, myelocytes, lymphocytes, and thrombocytes[110]. These cells have half lives of hours (granulocytes) to months (erythrocytes) and thus need to be continuously replenished. Their production is organized into the multiple differentiation and expansion steps that make up the hierarchy of the hematopoietic system (Figure 23).

The most primitive cells, long-term hematopoietic stem cells (LT-HSCs), are responsible for life-long blood cell production[111–113]. LT-HSCs are (i) multipotential - they give rise to all cells found in the blood[114,115] and can also transdifferentiate to produce non-hematopoietic tissues[116] such as muscle cells[117], and (ii) self-renewing - during normal stem cell proliferation, lineage commitment and differentiation are balanced by the production of additional stem cells, such that their total num-

ber remains constant[118,119]. Throughout adult life, the LT-HSC population resides predominantly in the bone marrow and is associated with a highly organized microenvironment[120–123]. Low numbers of stem cells, however, are found migrating in peripheral blood[124].

LT-HSCs give rise to short-term HSCs (ST-HSCs), which give rise to multipotent progenitors (MPPs)[125]. Dedifferentiation cannot be detected[126]. As HSCs mature from the long-term self-renewing pool to multipotent progenitors, they progressively lose their potential to self-renew, but become more mitotically active. In mice, it has been showsn that LT-HSCs produce mature hematopoietic cells for the lifetime of the mouse. ST-HSCs and MPPs, however, reconstitute lethally irradiated mice for less than eight weeks[126].

Common lymphoid progenitors (CLPs) and common myeloid progenitors (CMPs) are oligolineage-restricted progeny of MPPs and precursors of all lymphoid and myeloid cells, respectively. CMPs give rise to megakaryocyte/erythrocyte progenitors (MEPs) or granulocyte/monocyte progenitors (GMPs). These cells eventually produce all mature myelocytes found in the blood. CLPs give rise to further precursors which in turn produce all mature lymphoid cells.

Hematopoietic cell numbers are proportional to their differentiation stage. The most primitive stem cells are the least abundant, and it has been estimated that only one in 10^4 to 10^5 bone marrow cells is a stem cell[127,128]. Terminally differentiated blood cells are the most abundant and account for 10^{12} cells in total.

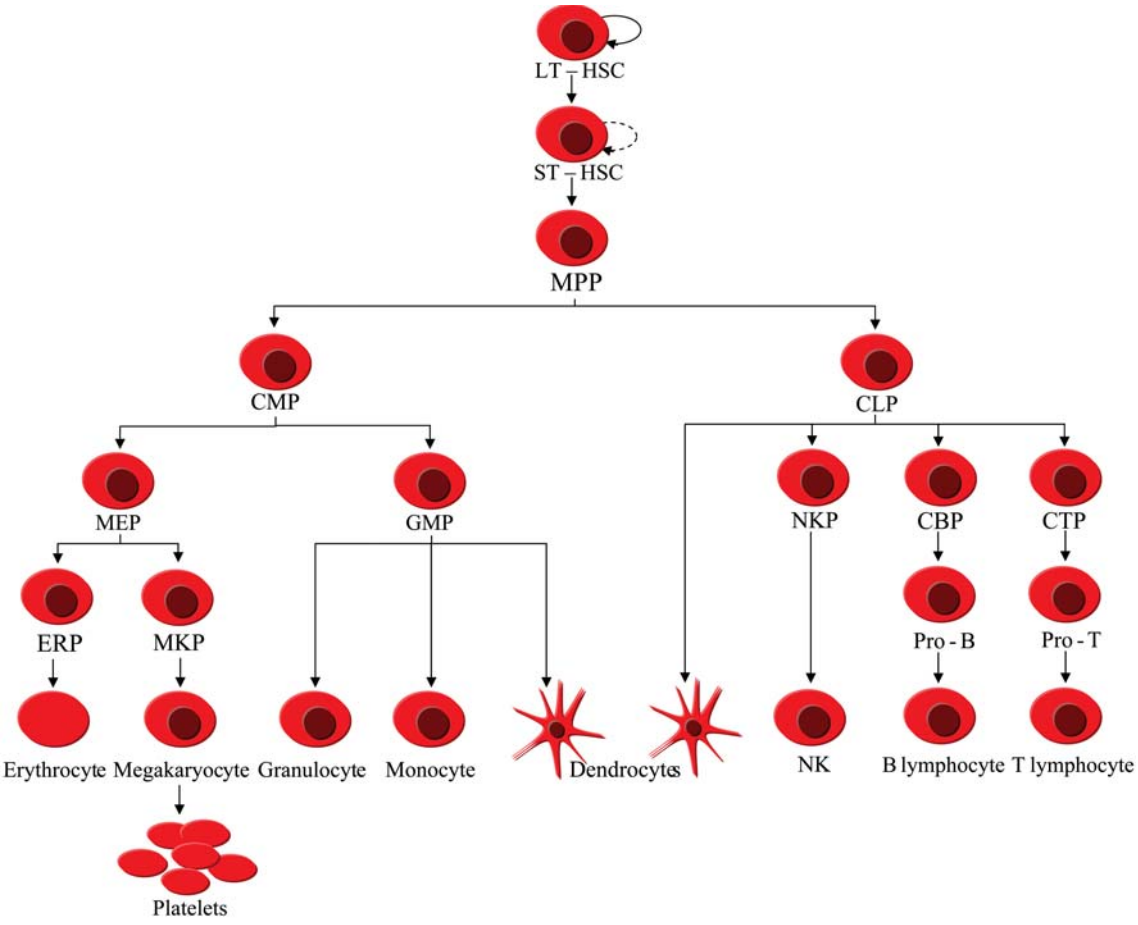

LT – HSC
ST – HSC
MPP
CMP
CLP
MEP
GMP
NKP
CBP
CTP
ERP
MKP
Pro - B
Pro - T
Erythrocyte
Megakaryocyte
Granulocyte
Monocyte
Dendrocytes
NK
B lymphocyte
T lymphocyte
Platelets

Figure 23: The hematopoietic system. The blood-producing system is organized hierarchically. The most primitive and least abundant cells are the long-term hematopoietic stem cells (LT-HSCs). LT-HSCs produce short-term hematopoietic stem cells (ST-HSCs) which in turn give rise to multipotent progenitors (MPPs). MPPs produce common lymphoid progenitors (CLPs; the precursors of all lymphoid cells) and common myeloid progenitors (CMPs; the precursors of all myeloid cells). Both CMPs/GMPs (granulocyte macrophage precursors) and CLPs can give rise to all known mouse dendritic cells. ERP, erythrocyte precursor; MEP, megakaryocyte erythrocyte precursor; MKP, megakaryocyte precursor; NKP, natural killer precursor; NK, natural killer; CBP, common B lymphocyte precursor; CTP, common T lymphocyte precursor.

Chronic myeloid leukemia (CML) is a malignant clonal disorder of hematopoietic stem cells that leads to increased numbers of myelocytes, erythrocytes and thrombocytes in peripheral blood[129,130]. The molecular hallmark of CML is the Philadelphia (Ph) chromosome, first described as shortened chromosome 22 in 1960[131] and then as a t(9;22) translocation in 1973[132]. This abnormal chromosome is found in cells from the myeloid, erythroid, megakaryocytic, and B lymphoid lineages, indicating that CML is a stem cell disease[130]. The Ph chromosome is present in 95% of patients. The remaining 5% have complex or variant translocations that have the same result: fusion of the BCR (breakpoint cluster region) gene on chromosome 22 to the ABL (Ableson leukemia virus) gene on chromosome 9 (Figure 24a). The chimeric oncogene BCR-ABL encodes a constitutively active cytoplasmic tyrosine kinase. This protein activates growth and differentiation pathways in hematopoietic cells; its effectors include RAS, RAF, MYC, JUN, STAT and PI3K[129]. BCR-ABL initiates a process to transform hematopoietic cells such that their growth and survival become independent of cytokines[133]. Transgenic expression of BCR-ABL in mice causes acute leukemia at birth[134].

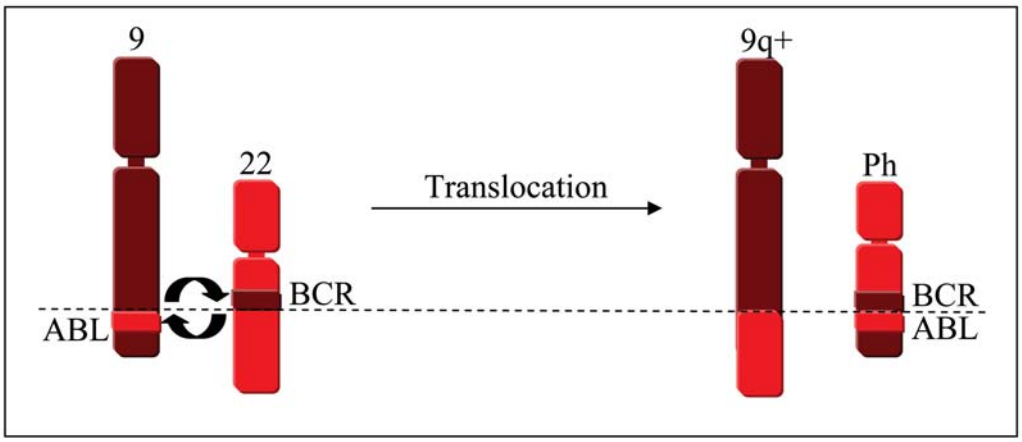
CML is recognized when the BCR-ABL clone has expanded to an extent that allows about 10^{12} leukemic cells to be produced daily. This occurs only after a prolonged latent period - on average 6 to 7 years, as calculated from the atomic bomb survivor data[135]. The disease then progresses from an initial chronic phase via an accelerated phase to the final blast crisis within three to five years. During

chronic phase, leukemic myeloid progenitors, so-called blasts, increase in number to constitute up to five percent of the total blood cell population. Progression of the disease is marked by accelerated growth of immature myelocytes that are incapable of differentiation; this progression might be due to further genetic alterations such as duplication of the Ph chromosome or inactivation of tumor suppressor genes[136–138]. During accelerated phase, blasts accumulate to abundances of five to thirty percent. CML then progresses to the rapidly fatal blast crisis which often transforms into an acute leukemia[139].

Traditional treatment of CML includes killing of stem cells by chemotherapy and subsequent bone marrow transplantation[129]. Due to age limitations of patients and the lack of appropriate bone marrow donors, only 15 to 20% of CML patients are candidates for transplantation. The survival rates of these patients are 50 to 60% [140]. Patients not eligible for transplantations are treated with interferon alfa, alone or in combination with other drugs, but this treatment fails to induce long-term remissions in most patients.

The revolution in CML treatment came about with the discovery of the tyrosine kinase inhibitor imatinib (STI-571, Gleevec)[141]. This compound was demonstrated to be effective in all stages of the disease[142]. Therefore, CML represents the first human cancer in which molecularly targeted therapy leads to a dramatic clinical response[142–148]. Imatinib binds competitively with ATP to BCR-ABL and blocks its abnormal signaling (Figure 24b). It selectively inhibits the proliferation of BCR-

a


9
22
Translocation
9q+
Ph
BCR
ABL
BCR
ABL

b

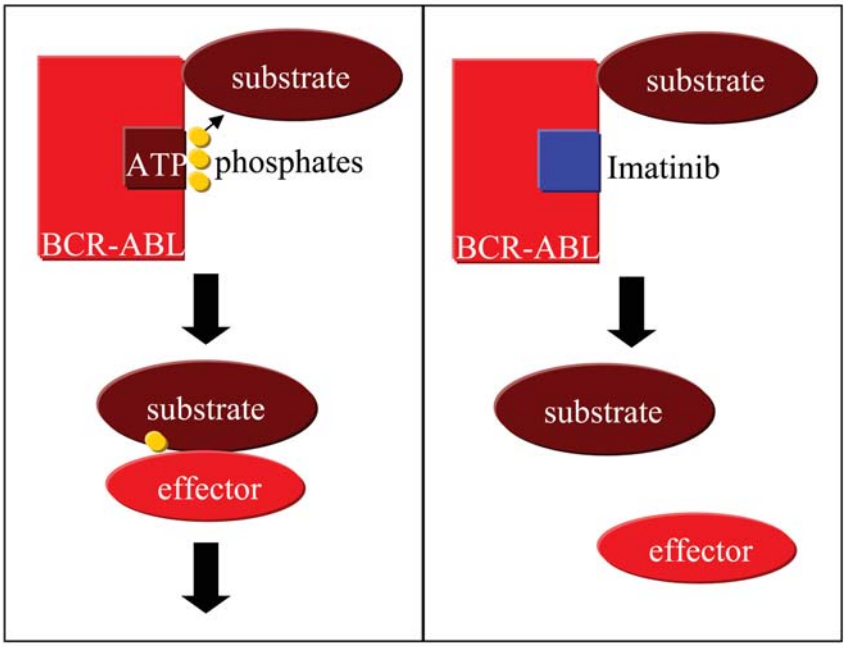

substrate
ATP
phosphates
BCR-ABL
substrate
effector
substrate
Imatinib
BCR-ABL
substrate
effector

Figure 24: The Philadelphia chromosome and imatinib (Gleevec). (a) The Philadelphia chromosome results from a t(9;22) translocation that creates the BCR-ABL fusion gene. The BCR-ABL gene encodes a constitutively active cytoplasmic tyrosine kinase. This protein activates growth and differentiation pathways in hematopoietic cells and causes chronic myeloid leukemia (CML). (b) Imatinib (Gleevec) binds competitively with ATP to BCR-ABL and blocks its abnormal signaling.

ABL positive cell lines. In most patients, however, the rapid decline of the leukemic cell burden induced by the ABL tyrosine kinase inhibitor imatinib fails to eliminate residual disease. Bone marrow studies have shown that the residual cells are part of the stem cell compartment[149,150]. This observation raises the question whether imatinib is capable of impairing the proliferation of leukemic stem cells. Moreover, a substantial fraction of patients develops acquired resistance to imatinib. Mutations in the ABL kinase domain are the main mechanism of resistance and account for 70 - 80% of cases with treatment failure[151–159] (Figure 25). Sometimes, resistance mutations are present in leukemic cells prior to imatinib therapy[159–161].

6.2 Results

We design a mathematical model which describes four layers of the differentiation hierarchy of the hematopoietic system (see Section 6.3). Stem cells give rise to progenitors, which produce differentiated cells, which produce terminally differentiated cells. This hierarchy applies both to normal and leukemic cells. The BCR-ABL oncogene is present in all leukemic cells. It leads to a slow clonal expansion of leukemic stem cells and accelerates the rate at which these cells produce progenitors and differentiated cells.

We analyze 169 CML patients. The levels of BCR-ABL transcripts in the blood of the patients is measured by a quantitative real-time polymerase-chain-reaction (RQ-PCR) assay[154,162]. BCR is used as the control gene and BCR-ABL values are

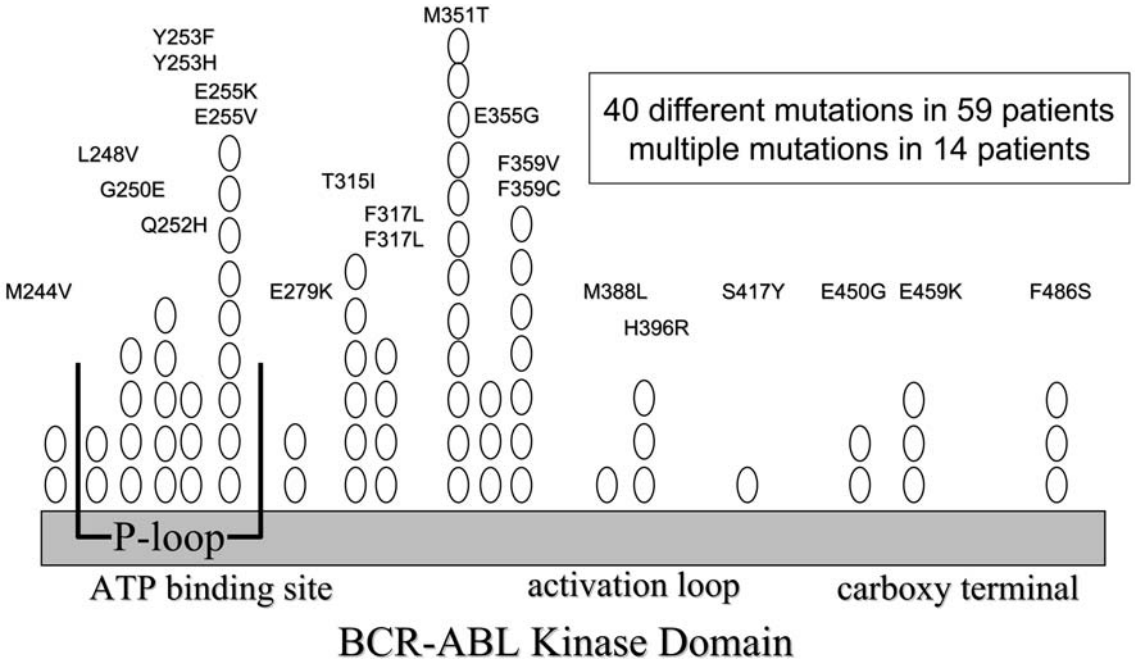


Figure 25: Resistance mutations. Acquired resistance to imatinib usually develops due to single point mutations in the BCR-ABL kinase domain. Each point mutation is sufficient to confer resistance to imatinib; the degree of resistance, however, differs between the point mutations. About 40 different resistance mutations have been discovered so far. In some patients, multiple resistance mutations were detected. The labels denote the individual mutations that are detected; for example, E255K means that at position 255 the amino acid glutamic acid was replaced by lysine. The number of circles represent the number of patients in which the respective mutation was found. Data from T. P. Hughes.

expressed as a percentage of the BCR transcript levels to compensate for variations in the RNA quality and efficiency of reverse transcription. As the blood predominantly contains terminally differentiated cells, the obtained values give an estimate of the fraction of terminally differentiated leukemic cells.

Successful therapy leads to a bi-phasic exponential decline of leukemic cells (Figure 26). The first slope is determined by calculating the exponential decline between 0 and 3 months; a mean value of 0.05 ± 0.02 per day is obtained, which corresponds to a decline of 5% per day. The second slope is determined by calculating the exponential decline between 6 and 12 months; a mean value of 0.008 ± 0.004 per day is obtained, which corresponds to a decline of 0.8% per day. This analysis considers only patients who do not have any rise in the leukemic cell burden during the first 12 months of therapy to exclude the effect of acquired resistance.

Fitting our model to the data, we conclude that the first slope represents the turnover rate of differentiated leukemic cells. Therefore, these cells have an average life-span of $1/0.05 = 20$ days. Upon reaching a steady-state with the leukemic progenitors, the number of differentiated leukemic cells decreases at the same rate as the number of leukemic progenitors. The second slope represents the turnover rate of leukemic progenitors. Hence, these cells have an average life-span of $1/0.008 = 125$ days. Both estimates denote average life-spans during imatinib therapy. Any process that removes cells from the corresponding subpopulation (including further differentiation) contributes to the average life-span.

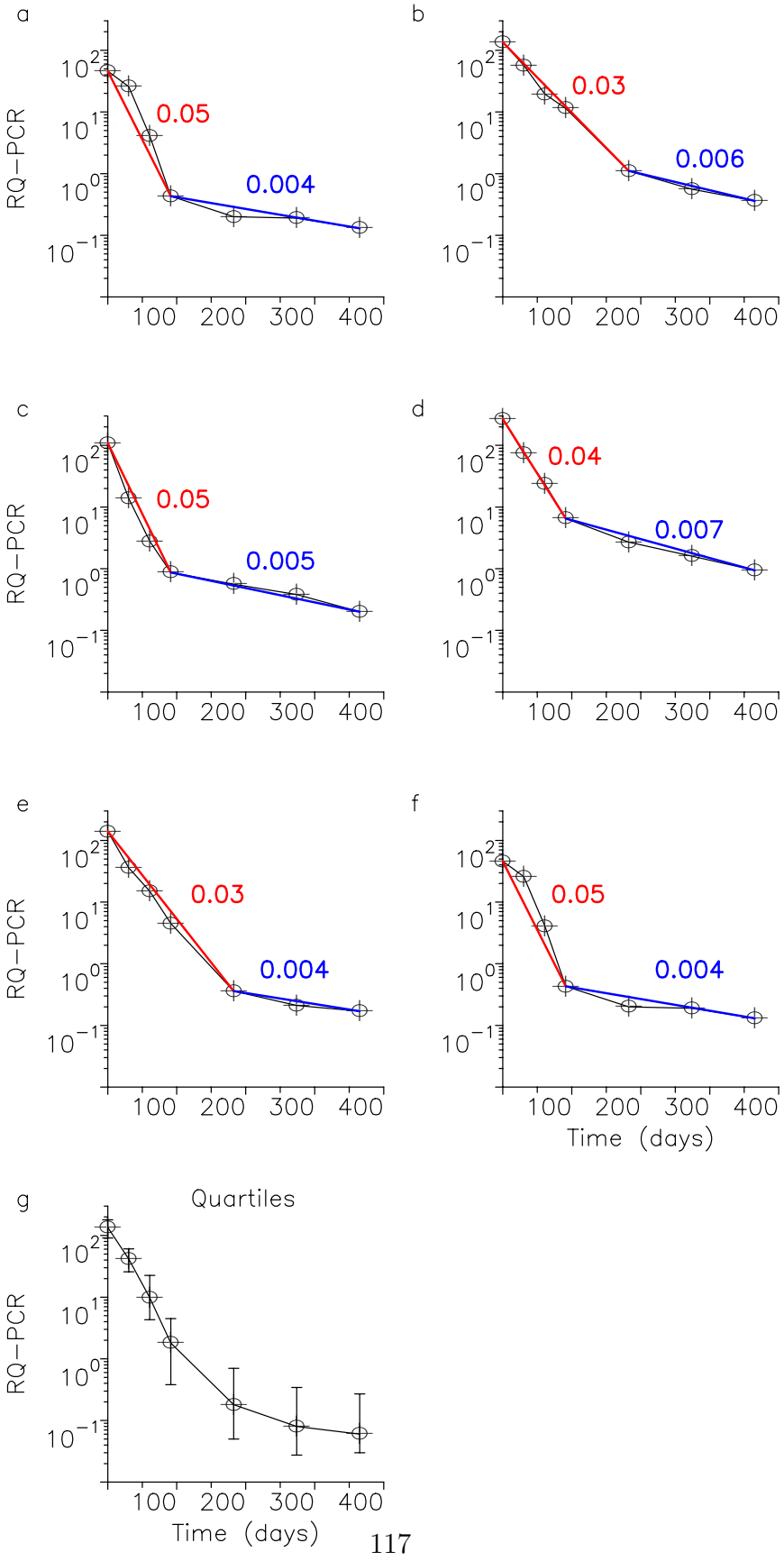

a
b
c
d
e
f
g
RQ−PCR
0.05
0.004
0.03
0.006
0.05
0.005
0.04
0.007
0.03
0.004
0.05
0.004
Quartiles
100
200
300
400
Time (days)

Figure 26: Imatinib leads to a bi-phasic decline of leukemic cells. (a-f) The levels of BCR-ABL transcripts in the blood of 6 patients are shown during 12 months of therapy starting at day 0. The BCR-ABL transcript levels are determined by quantitative PCR (RQ-PCR) and are expressed as percentage of the values of BCR. Every leukemic cell has one copy of the normal BCR gene and at least one copy of BCR-ABL. The values might exceed 100% because BCR-ABL expression can be upregulated in leukemic cells. In the 6 patients shown here the first slope ranges from 0.03 to 0.05 per day. The second slope ranges from 0.004 to 0.007 per day. On average, we find that the first slope is 0.05 ± 0.02 per day while the second slope is 0.008 ± 0.004 per day. In the context of our model, the first slope represents the death rate of leukemic differentiated cells and the second slope the death rate of leukemic progenitors during imatinib therapy. On average, the first slope leads to a 1000-fold decline in the leukemic cell burden. Panel (g) shows the median with quartiles taken over all patients who do not have a rise in the leukemic cell burden during the first 12 months of therapy. This approach attempts to exclude the effect of acquired resistance.

The first slope leads to an approximately 1000-fold decline in the leukemic cell burden. Therefore, imatinib reduces the rate at which leukemic differentiated cells arise from leukemic progenitors about 1000-fold. This effect is as if imatinib prevented about 10 rounds of cell division of leukemic cells ($2^{10} = 1024$), either by increasing their death rate or by reducing their division rate.

Some patients ceased imatinib therapy due to complications or side effects (Figure 27). Even if imatinib had been administered for many months (up to three years), the numbers of BCR-ABL transcripts in those patients rose within 3 months after discontinuation of therapy to levels at pre-treatment baseline or above. Other studies of patients who ceased imatinib led to similar findings[163,164]. We conclude that long-term imatinib treatment does not deplete the cell population that drives this disease. This conclusion is consistent with the hypothesis that leukemic stem cells are insensitive to chemotherapy[165]. The average rate of the exponential increase after stop of treatment is 0.09 ± 0.05 per day, corresponding to a doubling time of 8 days. This time scale indicates the rate at which terminally differentiated leukemic cells arise from leukemic stem cells.

A comparison between model and data suggests the following two concepts: (i) Imatinib treatment leads to a competitive disadvantage of leukemic progenitors and differentiated cells; their production rates are dramatically reduced. The consequence is a bi-phasic decline of the abundance of the BCR-ABL transcript in response to therapy. (ii) Leukemic stem cells, however, are not depleted during

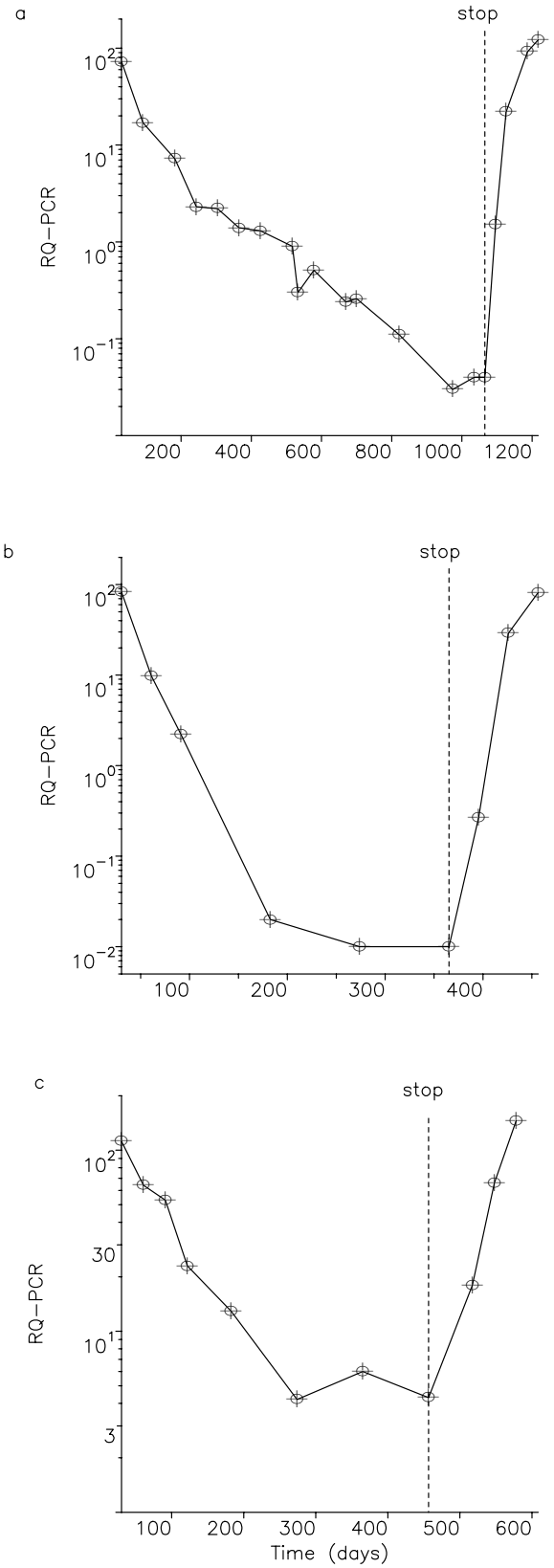

a
stop
RQ−PCR
10^2
10^1
10^0
10^{-1}
200
400
600
800
1000
1200
b
stop
RQ−PCR
10^2
10^1
10^0
10^{-1}
10^{-2}
100
200
300
400
c
stop
RQ−PCR
10^2
30
10^1
3
100
200
300
400
500
600
Time (days)

Figure 27: Discontinuation of imatinib therapy after 1-3 years leads to a rapid increase of leukemic cells to levels at or beyond pre-treatment baseline. We conclude that leukemic stem cells, which drive CML disease, are not depleted by imatinib therapy. The up-slope of 0.09 ± 0.05 per day corresponds to a doubling time of roughly 8 days, which characterizes the rate at which differentiated leukemic cells are regenerated from leukemic stem cells.

imatinib treatment. The total leukemic cell burden rapidly returns to the baseline value (or beyond) when imatinib is discontinued.

Acquired imatinib resistance usually develops due to mutations in the ABL kinase domain[151–159]. Resistant leukemic cells emerge after an initially successful response to imatinib therapy and lead to a relapse of disease (Figure 28). The average slope was determined by calculating the exponential increase after the first appearance of resistance mutations in 30 patients; a value of 0.02 ± 0.01 per day was obtained. Of those patients who start imatinib in the early chronic, late chronic, and accelerated phase of CML, respectively, 12%, 32%, and 62% develop detectable resistance mutations within two years of treatment[166] (Figure 30a).

Our basic model can be extended to include the stochastic evolution of resistance. Consider an exponentially growing leukemia continually producing resistant cells at rate u per cell division. The probability of having resistance mutations once leukemic stem cells have reached a certain abundance, y_0, is given by $P = 1 - \exp(-uy_0\sigma)$ where $\sigma = (1 + s)\log(1 + 1/s)$. The parameter s denotes the excess reproductive ratio of the exponentially growing leukemia, which is the relative difference between birth and death rates, and is given by $s = (r_y - d_0)/d_0$. This calculation assumes that resistance mutations are neutral prior to therapy.

If the point mutation rate is about 10^{-8} per base per cell division[45], then resistance due to any one of about forty known mutations[150–160] arises at rate $u = 4 \cdot 10^{-7}$ per cell division. The abundance of leukemic stem cells at diagnosis in early chronic

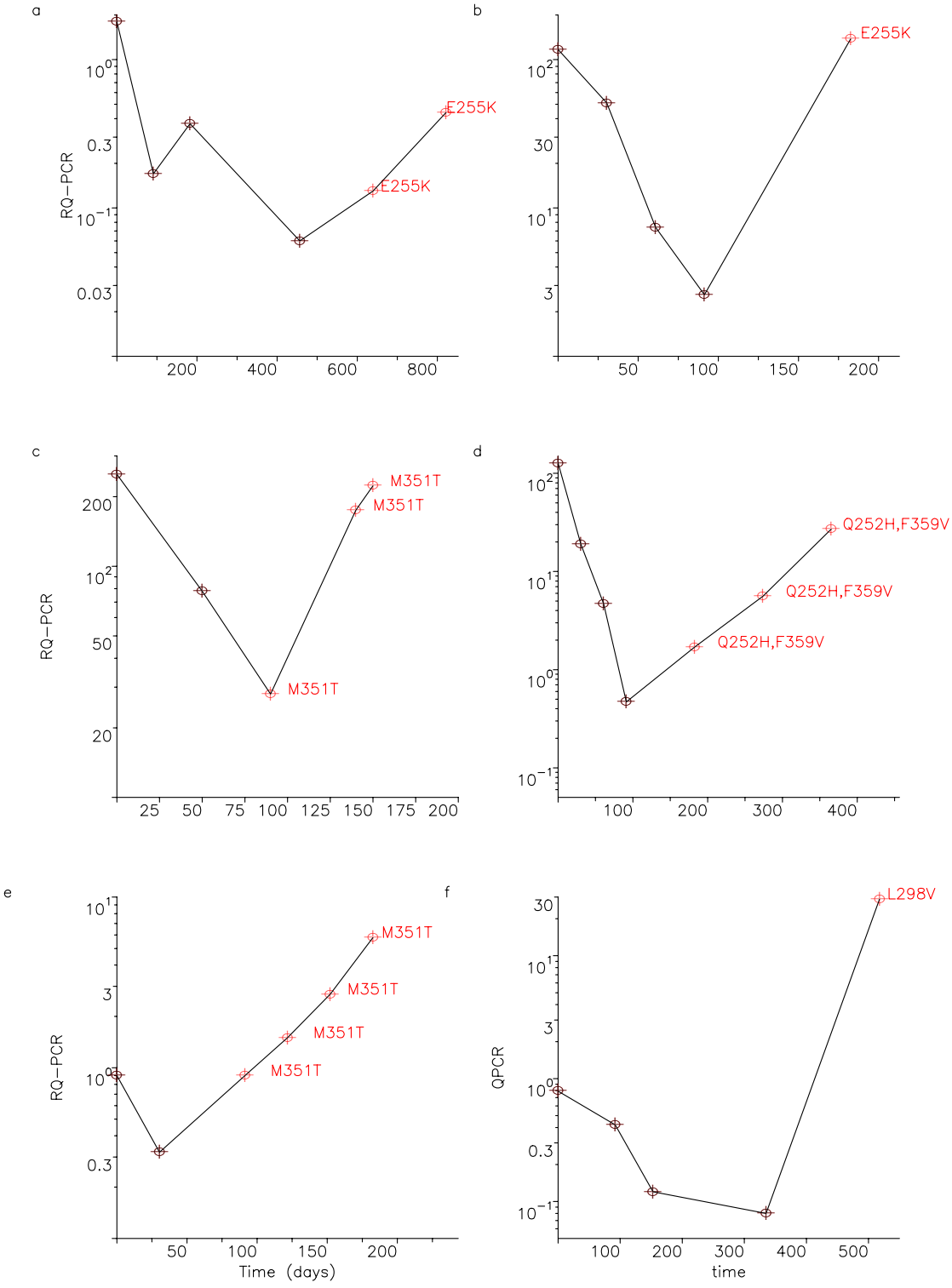

a
RQ−PCR
E255K
E255K
b
E255K
c
RQ−PCR
M351T
M351T
M351T
d
Q252H,F359V
Q252H,F359V
Q252H,F359V
e
RQ−PCR
M351T
M351T
M351T
M351T
Time (days)
f
QPCR
L298V
time

Figure 28: About 40 different point mutations in BCR-ABL have been identified that confer various degrees of resistance to imatinib therapy. We show the evolution of resistance in six patients. The labels denote the individual mutations that are detected at various time points. For example, E255K means that at position 255 the amino acid glutamic acid was replaced by lysine. Resistance mutations lead to a relapse of leukemic cells. The up-slope ranges from 0.003 to 0.06 per day in 30 patients. The average up-slope is 0.02 ± 0.01 per day. Our mathematical analysis suggests that the characteristic time scale for the rise of resistance is given by the rate at which resistant leukemic stem cells expand during therapy. Therefore, more aggressive leukemias should lead to a faster emergence of resistance, while slowly growing leukemias should allow many years of successful therapy even if resistance mutations are present in a patient at low frequencies.

phase[167–169] is estimated to be about $y_0 = 2.5 \cdot 10^5$. For these values and $s = 1$, we obtain that about 13% of patients harbor resistance mutations at the time of diagnosis. If imatinib therapy commences at a later stage in disease, when the leukemic stem cell population has expanded to 10^7 cells, for example, then 100% of those patients have some resistant leukemic stem cells. Therefore, the higher incidence of resistance in patients who start imatinib therapy in a later phase of disease can be explained by an increased leukemic cell burden (Figure 30b).

We expect that the frequency of resistance mutations at the start of imatinib therapy is below detection limit in most patients. We can use the deterministic model to calculate the time until detection of resistance mutations and treatment failure (Figure 30c). In our model, the characteristic time scale is given by the rate at which the resistant cancer stem cells are expanding during therapy. Therefore, a faster growing leukemia leads to an earlier emergence of resistance, while a slower growing leukemia might allow many years of successful therapy, even if resistant cancer cells are present at low frequencies. The reason for this unusual behavior is that leukemic stem cells, with or without resistance mutations, continue to expand during treatment. Imatinib primarily acts by reducing the rate of differentiation of those leukemic cells that do not have resistance mutations.

Figure 29 summarizes the dynamical features of our mathematical model. Normal hematopoietic cells are in a steady state. Leukemic stem cells expand exponentially at a slow rate. Imatinib reduces the rate at which leukemic stem cells produce

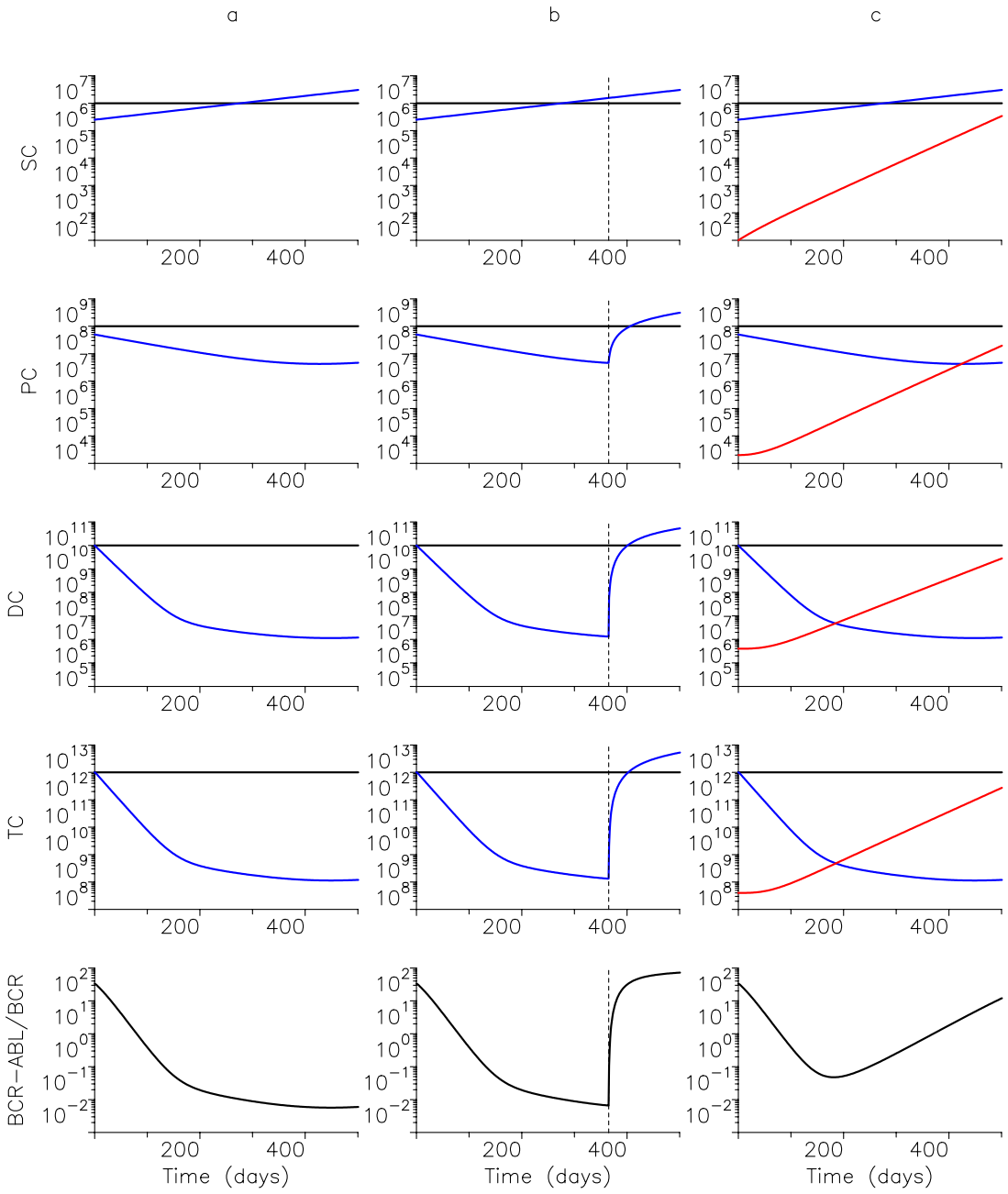

a
b
c
SC
PC
DC
TC
BCR−ABL/BCR
200
400
Time (days)

Figure 29: Model dynamics of treatment response to imatinib (a) without resistance mutations, (b) when therapy is stopped after one year, and (c) with resistance mutations. The rows show stem cells (SC), progenitor cells (PC), differentiated cells (DC), and terminally differentiated cells (TC). The bottom row shows the ratio of BCR-ABL over BCR in %. (a) Imatinib therapy is started at day 0 (when leukemic terminally differentiated cells have reached the abundance of healthy cells). Treatment leads to a bi-phasic decline of BCR-ABL/BCR ratio which reflects the bi-phasic decline of differentiated and terminally differentiated leukemic cells. Leukemic progenitors decline only at the slower rate of the second slope. Leukemic stem cells continue to expand at a slow rate. (b) Discontinuation of treatment after one year (indicated by the broken line) leads to a rapid rise of leukemic cells to levels above the pre-treatment baseline, because leukemic stem cells have not been depleted during therapy. Further disease progression shows the characteristic time scale of leukemic stem cell expansion. (c) Emergence of resistance mutations leads to an increase of the BCR-ABL to BCR ratio at a rate which is determined by the rise of resistant leukemic stem cells. Parameter values are $d_0 = 0.003$, $d_1 = 0.008$, $d_2 = 0.05$, $d_3 = 1$, $a_x = 0.8$, $b_x = 5$, $c_x = 100$, $r_y = 0.008$, $a_y = 2a_x$, $b_y = 2b_x$, $c_y = c_x$, $r_z = 0.023$, $a'_y = a_y/100$, $b'_y = b_y/750$, $c'_y = c_y$, $a_z = a'_z = a_y$, $b_z = b'_z = b_y$, and $c_z = c'_z = c_y$. In (c), the abundance of resistant leukemic stem cells at start of treatment is $z_0(0) = 10$, and $u = 4 \cdot 10^{-8}$.

a

	early chronic phase	late chronic phase	accelerated phase
1 year	5.9%	14%	38%
2 years	12%	32%	62%

b

u \ y_0	10^5	10^6	10^7	10^8
$4 \cdot 10^{-7}$	5.4%	43%	100%	100%
$4 \cdot 10^{-8}$	0.6%	5.4%	43%	100%

c

$z_0(0)$	detection	failure
10	186 d	566 d
100	141 d	470 d
1000	101 d	357 d
10 000	61 d	241 d
100 000	9.2 d	119 d

Figure 30: Observing and predicting resistance. (a) Observed percentage of patients with acquired resistance mutations during the first and second year of treatment. In 5.9% of early chronic phase patients, resistant leukemic cells are detected within the first year of treatment; in 12%, resistant leukemic cells are detected within the first two years of therapy. These numbers increase to 14% and 32% in late chronic phase patients and to 38% and 62% in accelerated phase patients. 'Early chronic phase' refers to patients who commenced imatinib within one year of diagnosis. (b) Predicted percentage of patients which harbor resistance mutations depending on the abundance of leukemic stem cells, y_0, and the mutation rate conferring resistance, u. Resistant leukemic cells can be below detection limit in some patients. We use $s = 1$. (c) Predicted time until detection of resistance mutations ($y_3 = z_3$) and until treatment failure ($z_3 = 10^{12}$) in dependence of the abundance of resistant stem cells at the start of therapy, $z_0(0)$. The parameter values are the same as in Figure 29.

progenitors. Hence, the abundance of leukemic progenitors declines once treatment is started. Similarly, imatinib reduces the rate at which leukemic progenitors produce differentiated cells. The abundance of differentiated leukemic cells shows a bi-phasic decline. The first slope is determined by the average life-span of leukemic differentiated cells (20 days), while the second slope reflects the longer life-span of leukemic progenitors (125 days). We assume that imatinib does not affect the rate at which leukemic differentiated cells produce terminally differentiated cells. This cell population has a fast turnover rate - on a time scale of one day, and simply tracks the biphasic decline of leukemic differentiated cells. If imatinib is discontinued, there is a rapid resurgence of the leukemic load because the cell population which drives the disease, the leukemic stem cells, was not depleted during therapy. Resistant leukemic stem cells might expand faster than leukemic stem cells during therapy for two reasons: (i) either they have an inherent selective advantage; or (ii) imatinib somewhat reduces the growth rate of leukemic stem cells without depleting them. In any case, resistant leukemic stem cells continue to produce large amounts of progenitors and differentiated cells during therapy. The total leukemic cell burden declines initially, but rises again once resistant differentiated cells become abundant. In this case, the time to treatment failure is determined by the rate of expansion of resistant leukemic stem cells.

In summary, the comparison between model and data provides quantitative insights into the in-vivo kinetics of CML. We obtain numerical estimates for the

turnover rates of leukemic progenitors and differentiated cells. Imatinib reduces the rate at which those cells are being produced from stem cells, but does not lead to an observable decline of leukemic stem cells. The probability of harboring resistance mutations increases with disease progression as a consequence of an increased leukemic stem cell abundance. The characteristic time to treatment failure caused by acquired resistance is given by the growth rate of the leukemia. Thus, multiple drug therapy[170] is especially important for patients who are diagnosed with advanced and rapidly growing disease.

6.3 Methods

The basic model. Our mathematical model is based on the architecture of the hematopoietic system as proposed by Weissman[112,114]. Denote by x_0, x_1, x_2, and x_3 the abundances of normal hematopoietic stem cells, progenitors, differentiated cells, and terminally differentiated cells. Their respective leukemic abundances are given by y_0, y_1, y_2, and y_3 for cells without resistance mutations and by z_0, z_1, z_2, and z_3 for cells with imatinib resistance mutations. We assume that normal hematopoietic stem cells are held at a constant level by a homeostatic mechanism. The growing leukemic cell population might eventually upset this homeostasis, but we do not consider this effect here. Leukemic stem cells escape from the homeostatic control and grow at a slow pace; it takes about 6-7 years from the first occurence of the BCR-ABL oncogene to detection of disease[135]. If imatinib resistance mutations are

neutral in the absence of treatment, then leukemic stem cells with these mutations grow at the same rate as the other leukemic stem cells. It is conceivable that some resistance mutations confer a selective advantage or disadvantage leading to different expansion rates of resistant clones. We assume, for simplicity, that the leukemia follows an exponential growth law; other dynamics are possible and need separate investigation.

Stem cells produce progenitors which produce differentiated cells which produce terminally differentiated cells. The rate constants are given by a, b, and c with subscripts x, y and z distinguishing between healthy, leukemic, and resistant cell lineages. The current understanding is that the BCR-ABL oncogene increases the rate at which leukemic stem cells proliferate and differentiate into progenitors. Furthermore, the BCR-ABL oncogene is thought to increase the rate at which progenitors proliferate and turn into differentiated leukemic cells. Therefore, we expect that $a_y > a_x$ and $b_y > b_x$. Imatinib counteracts this effect and reduces the rates to $a'_y < a_y$ and $b'_y < b_y$. The death rates of stem cells, progenitors, differentiated, and terminally differentiated cells are denoted by d_0, d_1, d_2, and d_3. The 'death' rates can include further differentiation of some cell types. Progenitors, differentiated, and terminally differentiated cells have a limited potential for cell division. They divide a certain number of times before they differentiate further or undergo apoptosis. The actual dynamics of this process is more complicated than described in our simple model, but the essential features are captured: we assume that a stem

cell leads to a progenitor lineage that divides i times thereby generating 2^i cells. The number 2^i is embedded in the rate constant a. Equally, a progenitor leads to a differentiated cell that generates 2^j cells, and this number is embedded in the rate constant b. Imatinib reduces the rate constants a and b, thereby preventing a certain number of cell divisions in each stage.

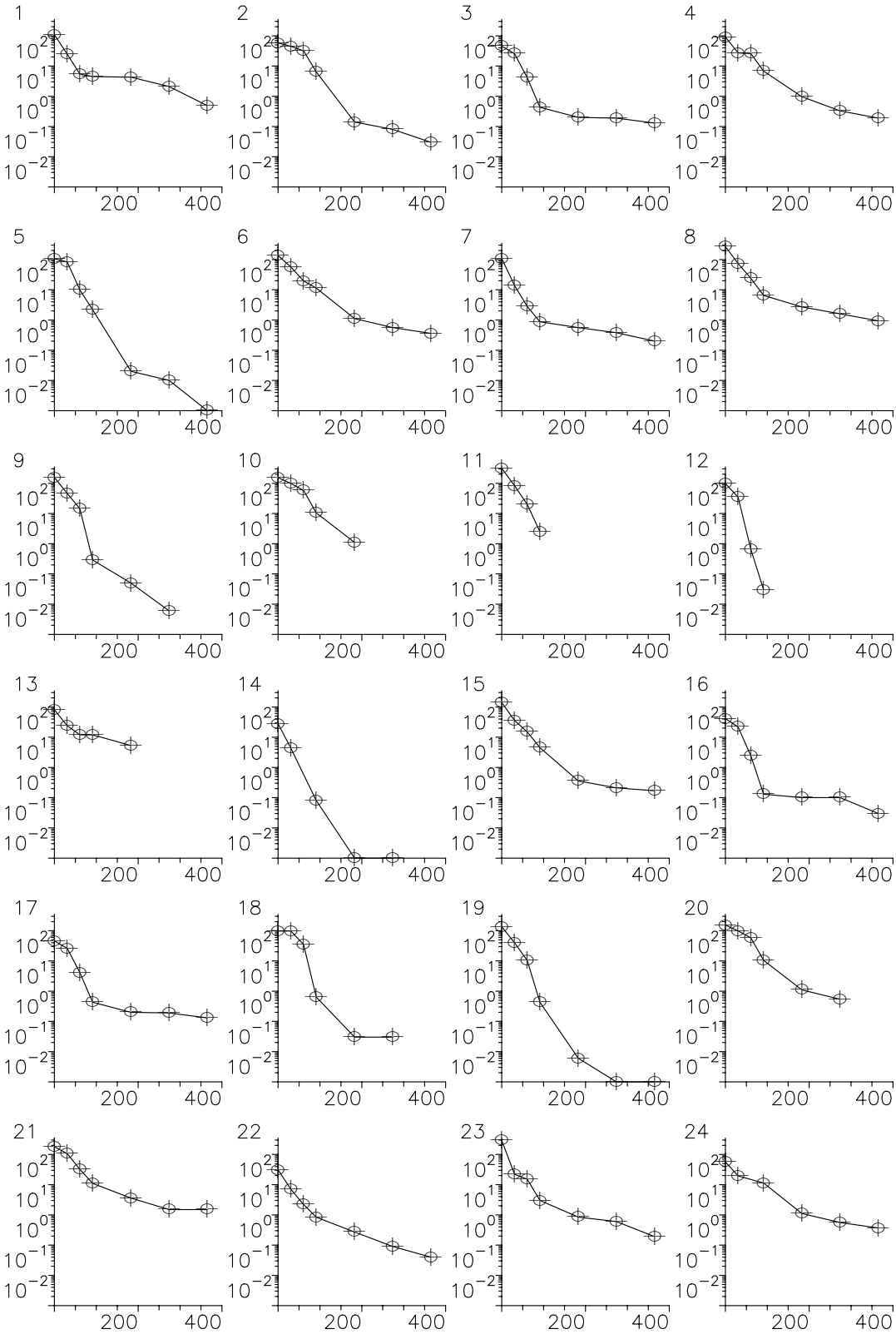
Homeostasis of normal stem cells is achieved by an appropriate declining function, λ. Leukemic stem cells divide at rate r_y and give rise to resistant stem cells at rate u per cell division. Resistant stem cells divide at rate r_z. With these assumptions, we have the following system of differential equations:

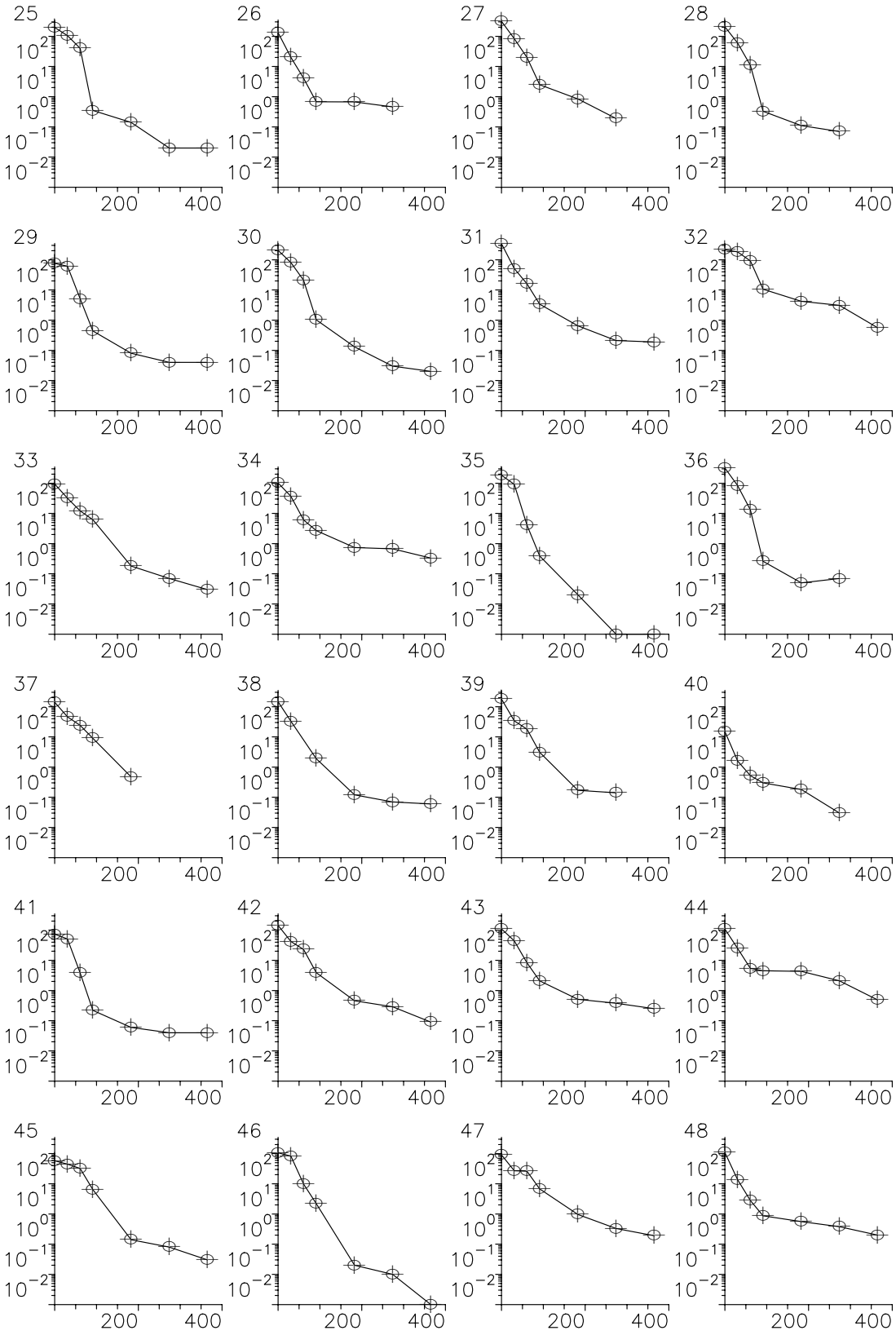
$$\begin{array}{lll} \dot{x}_0 = [\lambda(x_0) - d_0]x_0 & \dot{y}_0 = [r_y(1-u) - d_0]y_0 & \dot{z}_0 = (r_z - d_0)z_0 + r_y y_0 u \\ \dot{x}_1 = a_x x_0 - d_1 x_1 & \dot{y}_1 = a_y y_0 - d_1 y_1 & \dot{z}_1 = a_z z_0 - d_1 z_1 \\ \dot{x}_2 = b_x x_1 - d_2 x_2 & \dot{y}_2 = b_y y_1 - d_2 y_2 & \dot{z}_2 = b_z z_1 - d_2 z_2 \\ \dot{x}_3 = c_x x_2 - d_3 x_3 & \dot{y}_3 = c_y y_2 - d_3 y_3 & \dot{z}_3 = c_z z_2 - d_3 z_3 \end{array} \quad (8)$$

We assume that normal cells remain at their equilibrium values, x_0, $x_1 = a_x x_0/d_1$, $x_2 = b_x x_1/d_2$, and $x_3 = c_x x_2/d_3$. At the start of therapy, leukemic cells are in steady state ratios with their precursors: $y_3 = c_y y_2/d_3$, $y_2 = b_y y_1/d_2$, and $y_1 = a_y y_0/d_1$. The leukemic stem cell population expands as $y_0(t) = \exp[(r_y - d_0)t]$ (ignoring resistance mutations). Imatinib dramatically reduces the rate constants, a_y to a'_y and b_y to b'_y. This leads to a bi-phasic decline. The first slope describes the exponential decline of differentiated leukemic cells to their new steady state, $y_2 = b'_y y_1/d_2$. The magnitude of this decline is around 1000-fold, suggesting that b'_y is 1000 times smaller than b_y.

Thus, imatinib prevents about 10 rounds of cell division of differentiated leukemic cells. Note that $2^{10} = 1024$. The exponential decline is given by $\exp(-d_2 t)$. The observed slope establishes that the turnover rate of differentiated leukemic cells in the presence of imatinib is $d_2 = 0.05$ per day. Hence the average life-time of these cells is about $1/d_2 = 20$ days. After about 200 days of therapy, differentiated leukemic cells have reached their new steady state with leukemic progenitors and follow their decline that is given by $\exp(-d_1 t)$. The data suggest that the turnover rate of leukemic progenitors is $d_1 = 0.008$ per day corresponding to an average life-time of about $1/d_1 = 125$ days. A minimum estimate of the magnitude of this decline is about 7-fold, suggesting that a'_y is at the very least 7 times smaller than a_y. The overall 5000-fold decline is the combined effect of a'_y and b'_y. Figure 31 shows the molecular response to imatinib in all patients whose leukemic cell load was used to calculate the bi-phasic decline.

We have assumed that imatinib primarily acts by reducing the proliferation rate of leukemic cells, but it is in principle possible that imatinib also increases the death rates of those cells. In this case, we have to assign different death rates to the x, y and z populations, but all our conclusions remain the same; the numerical estimates for the decay slopes refer to the death rates of the corresponding cells during imatinib therapy. 'Death' includes any process that removes cells from the relevant subpopulation and does include cellular differentiation. It is also conceivable that the BCR-ABL oncogene changes the death rates of leukemic cells. In this case


1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
200
400


25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
200
400

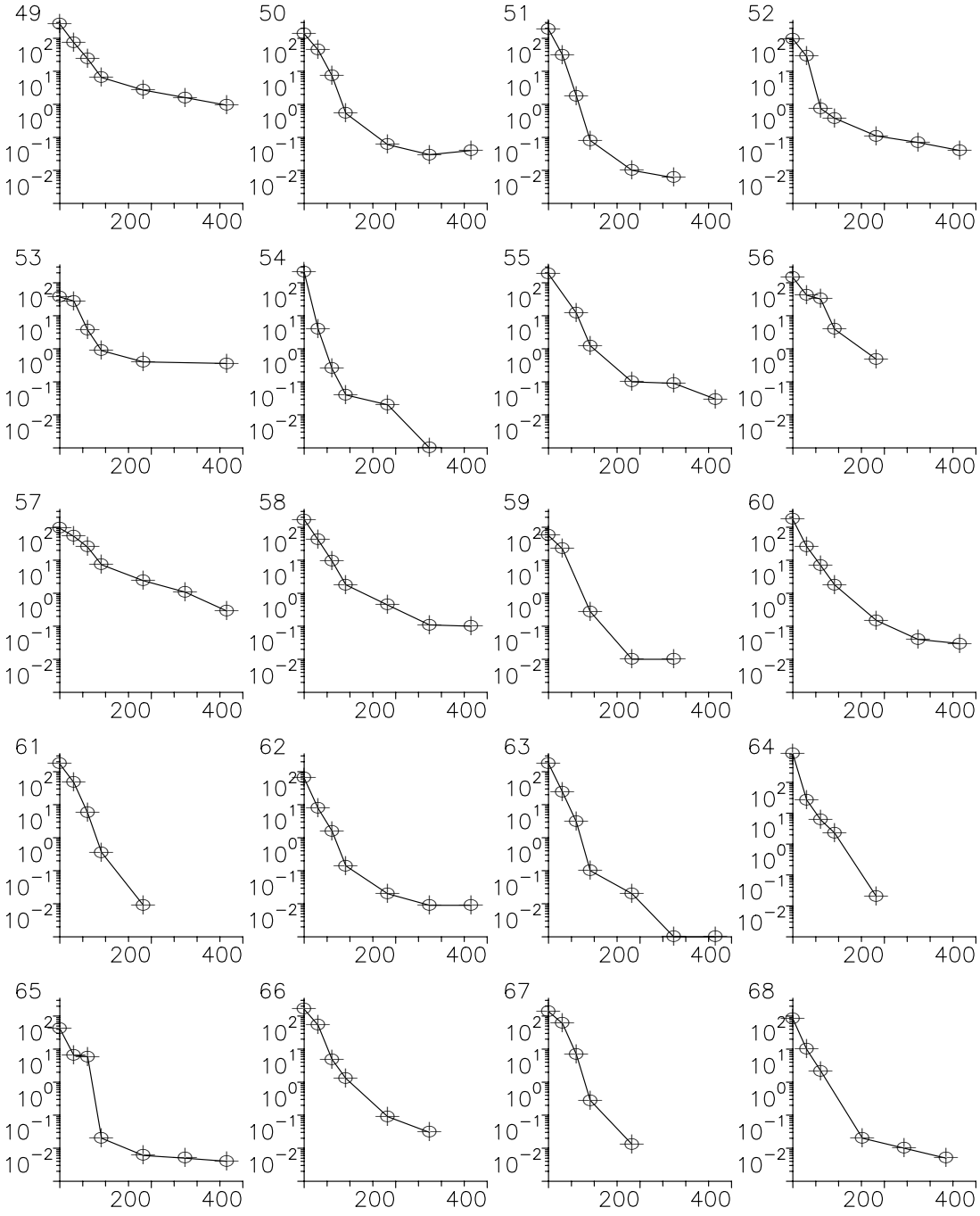

49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
200
400

Figure 31: Molecular response to imatinib. The figure shows the leukemic cell load in all patients used to calculate the bi-phasic decline. This analysis excludes patients who had a rise in their leukemic burden during the first twelve months of therapy. This approach attempts to exclude the effect of acquired resistance mutations.

again, the observed slopes indicate the turnover rates of leukemic cells during imatinib therapy. Should BCR-ABL and imatinib, however, not affect the death rates of these cells, then our results apply to the turnover rates of healthy hematopoietic cells, too.

If therapy is interrupted, the model predicts an explosive recurrence of cancer cells. Since imatinib leads to an at least 5000-fold reduction of the production of differentiated leukemic cells, stopping therapy leads to a sudden 5000-fold increase of differentiated leukemic cells at the time scale of their cell division. If imatinib led to a decline in leukemic stem cells, then the rebound after therapy should be to a level below baseline (by whatever amount leukemic stem cells have declined during therapy). If imatinib does not lead to a decline of leukemic stem cells, the rebound should lead to baseline levels or beyond. The latter is observed in all three patients who stopped therapy (Figure 27). Once terminally differentiated leukemic cells reach their new steady state, the further increase of the disease burden follows the characteristic time scale of leukemic stem cell expansion.

For comparison with the experimental PCR data, we calculate the BCR-ABL to BCR ratio as $(y_3 + z_3)/(2x_3 + y_3 + z_3)$ times 100%. A healthy cell has two copies of BCR. A leukemic cell normally has one copy of BCR and one copy of BCR-ABL. Most cells that are sampled by the PCR assay are terminally differentiated cells.

Imatinib and stem cells. Finding a cellular mechanism for drug resistance of stem cells is a very important goal for future experimental research. Imatinib is a

substrate for the multidrug resistance protein MDR p-glycoprotein and will therefore be excluded from cells that express significant MDR levels[171]. Stem cells naturally express higher levels of MDR[172]. It is unknown if this is the actual mechanism for sparing stem cells, because it has not been possible to accurately measure imatinib concentration in stem cells nor to measure BCR-ABL inhibition selectively in this compartment. Another possibility is that leukemic stem cells are less dependent on BCR-ABL for growth and survival than committed progenitors are, and therefore BCR-ABL inhibition does not eliminate leukemic stem cells.

Oncogene addiction. There is a fundamental unanswered question at the heart of molecularly targeted cancer therapy: if drugs impair the action of an oncogene and thereby reduce the growth of tumor cells, then why do these tumor cells have a competitive disadvantage relative to normal cells? The term 'oncogene addiction' has been coined to describe the phenomenon of tumor cells becoming dependent on a single molecular driver, but its cellular basis is unknown[173]. Most hypotheses are focused on the notion that tumor cells lose the redundant signaling machinery that normal cells retain, but this remains to be proven.

The probability of resistance mutations. We use a continuous-time branching process to calculate the probability that a patient has resistance mutations at the beginning of imatinib therapy. Assume that resistance mutations are neutral prior to therapy: this means mutated and unmutated stem cells expand at the same rate in the absence of imatinib. Leukemic stem cells, y_0, follow a branching process

starting with a single cell at time $t = 0$. They go extinct with probability d_0/a_y and grow exponentially to give rise to leukemia with probability $1 - d_0/a_y$. Resistant mutants are produced from normal leukemic cells with probability u per cell division. They reproduce and die at the same rate as normal leukemic cells. For the probability of having resistance mutations once the stem cell population has reached a certain size, y_0, we obtain $P = 1 - \exp(-uy_0\sigma)$. Here $\sigma = (1 + s)\log(1 + 1/s)$, where $s = (a_y - d_0)/d_0$ denotes the excess reproductive ratio of leukemic stem cells. For a wide range of plausible values of s, there is only little variation in σ: if s changes from 0.1 to infinity, then σ changes from 2.64 to 1.

We expect that patients who are diagnosed at later stages of CML disease tend to have a larger population size, y_0, of leukemic stem cells and consequently a higher chance of harboring resistance mutations at the time when therapy is started. In addition, these patients might already have a more aggressive, faster growing leukemia and therefore the time to detection of resistance and treatment failure can be shortened (see Figure 30c). Finally, especially in blast crisis it is conceivable that the mutation rate of leukemic cells is increased, which could also contribute to a higher incidence of resistance mutations.

7 References

[1] Vogelstein, B. & Kinzler, K. W. 2002. *The genetic basis of human cancer.* New York: McGraw-Hill.

[2] Michor, F., Iwasa, Y., & Nowak, M. A. 2004 Dynamics of cancer progression. *Nat Rev Can* **4**, 197–205.

[3] Kinzler, K. W. & Vogelstein, B. 1997 Gatekeepers and caretakers. *Nature* **386**, 761–763.

[4] Rajagopalan, H., Nowak, M. A., Vogelstein, B., & Lengauer, C. 2003 The significance of unstable chromosomes in colorectal cancer. *Nat Rev Can* **3**, 675–701.

[5] Sieber, O. M., Heinimann, K., & Tomlinson, I. P. 2003 Genomic instability - the engine of tumorigenesis?. *Nat Rev Can* **3**, 701–708.

[6] Bissell, M. J. & Radisky, D. 2001 Putting tumors in context. *Nat Rev Can* **1**, 46–54.

[7] Boveri, T. 1914. *Zur Frage der Entstehung maligner Tumoren.* Jena: Gustav Fischer.

[8] Muller, H. J. 1927 Artificial transmutation of the gene. *Science* **46**, 84–87.

[9] Knudson, A. G. 2001 Two genetic hits to cancer. *Nat Rev Can* **1**, 157–162.

[10] Knudson, A. G. 1971 Mutation and cancer: statistical study of retinoblastoma. *Proc Natl Acad Sci U S A* **68**, 820–823.

[11] Moolgavkar, S. H. & Knudson, A. G. 1981 Mutation and cancer: a model for human carcinogenesis. *J Natl Can Inst* **66**, 1037–1052.

[12] Cairns, J. 1975 Mutation, selection and the natural history of cancer. *Nature* **255**, 197–200.

[13] Cairns, J. 2002 Somatic stem cells and the kinetics of mutagenesis and carcinogenesis. *Proc Natl Acad Sci U S A* **99**, 10567–10570.

[14] Frank, S. A., Iwasa, Y., & Nowak, M. A. 2003 Patterns of cell division and the risk of cancer. *Genetics* **163**, 1527–1532.

[15] Frank, S. A. & Nowak, M. A. 2003 Cell biology: developmental predisposition to cancer. *Nature* **422**, 494.

[16] Gatenby, R. A. & Maini, P. K. 2003 Mathematical oncology: cancer summed up. *Nature* **421**, 321.

[17] Nordling, C. O. 1953 A new theory on cancer-inducing mechanism. *Br J Cancer* **7**, 68–72.

[18] Armitage, P. & Doll, R. 1954 The age distribution of cancer and a multi-stage theory of carcinogenesis. *Br J Cancer* **8**, 1–12.

[19] Armitage, P. & Doll, R. 1957 A two-stage theory of carcinogenesis in relation to the age distribution of human cancer. *Br J Cancer* **11**, 161–169.

[20] Fisher, J. C. 1959 Multiple-mutation theory of carcinogenesis. *Nature* **181**, 651–652.

[21] Nunney, L. 1999 Lineage selection and the evolution of multistage carcinogenesis. *Proc Roy Soc Lond B* **266**, 493–498.

[22] Tomlinson, I., Sasieni, P., & Bodmer, W. 2002 How many mutations in a cancer?. *Am J Pathol* **160**, 755–758.

[23] Luebeck, E. G. & Moolgavkar, S. H. 2002 Multistage carcinogenesis and the incidence of colorectal cancer. *Proc Natl Acad Sci U S A* **99**, 15095–15100.

[24] Grist, S. A. & al. 1992 In vivo human somatic mutation: frequency and spectrum with age. *Mutat Res* **266**, 189–196.

[25] Gatenby, R. A. & Vincent, T. L. 2003 An evolutionary model of carcinogenesis. *Cancer Res* **63**, 6212–6220.

[26] Goldie, J. H. & Coldman, A. J. 1979 A mathematical model for relating the drug sensitivity of tumors to their spontaneous mutation rate. *Cancer Treat Rep* **63**, 1727–1733.

[27] Goldie, J. H. & Coldman, A. J. 1983 Quantitative model for multiple levels of drug resistance in clinical tumors. *Cancer Treat Rep* **67**, 923–931.

[28] Anderson, A. R. & Chaplain, M. A. 1998 Continuous and discrete mathematical models of tumor-induced angiogenesis. *Bull Math Biol* **60**, 857–899.

[29] Owen, M. R. & Sherrat, J. A. 1999 Mathematical modeling of macrophage dynamics in tumors. *Math Models Methods Appl Biol Chem* **377**, 675–684.

[30] Chang, S., Khoo, C., & DePinho, R. A. 2001 Modeling chromosomal instability and epithelial carcinogenesis in the telomerase-deficient mouse. *Semin Cancer Biol* **11**, 227–239.

[31] Chang, S., Khoo, C., Naylor, M. L., Maser, R. S., & DePinho, R. A. 2003 Telomere-based crisis: functional differences between telomerase activation and ALT in tumor progression. *Genes Dev* **17**, 88–100.

[32] Maser, R. S. & DePinho, R. A. 2002 Connecting chromosomes, crisis, and cancer. *Science* **297**, 565–569.

[33] Taddei, F. & al. 1997 Role of mutator alleles in adaptive evolution. *Nature* **387**, 700–702.

[34] Otsuka, K. & al. 2003 Analysis of the human APC mutation spectrum in a saccharomyces cerevisiae strain with a mismatch repair defect. *Int J Cancer* **103**, 624–630.

[35] Strauss, B. S. 1998 Hypermutability in carcinogenesis. *Genetics* **148**, 1619–1626.

[36] Mintz, B. 1971 Clonal basis of mammalian differentiation. *Symp Soc Exp Biol* **25**, 345–370.

[37] Michor, F., Nowak, M. A., Frank, S. A., & Iwasa, Y. 2003 Stochastic elimination of cancer cells. *Proc Roy Soc Lond B* **270**, 2017–2024.

[38] Moran, P. 1962. *The statistical processes of evolutionary theory.* : Clarendon Press.

[39] Komarova, N. L., Sengupta, A., & Nowak, M. A. 2003 Mutation-selection networks of cancer initiation: tumor suppressor genes and chromosomal instability. *Jour Theor Biol* **223**, 433–450.

[40] Kimura, M. 1985 The role of compensatory neutral mutations in molecular evolution. *Jour Genetics* **64**, 7–19.

[41] Michor, F., Iwasa, Y., Komarova, N. L., & Nowak, M. A. 2003a Local regulation of homeostasis favors chromosomal instability. *Curr Biol* **13**, 581–584.

[42] Michor, F., Frank, S. A., May, R. M., Iwasa, Y., & Nowak, M. A. 2003b Somatic selection for and against cancer. *Jour Theor Biol* **225**, 337–382.

[43] Nowak, M. A., Michor, F., & Iwasa, Y. 2003 The linear process of somatic evolution. *Proc Natl Acad Sci U S A* **100**, 14966–14969.

[44] Michor, F., Iwasa, Y., Rajagopalan, H., Lengauer, C., & Nowak, M. A. 2004 Linear model of colon cancer initiation. *Cell Cycle* **3**, 358–362.

[45] Kunkel, T. A. & Bebenek, K. 2000 DNA replication fidelity. *Annu Rev Biochem* **69**, 497–529.

[46] Bach, S. P., Renehan, A. G., & Potten, C. S. 2000 Stem cells: the intestinal stem cell as a paradigm. *Carcinogenesis* **21**, 469–476.

[47] Yatabe, Y., Tavare, S., & Shibata, D. 2001 Investigating stem cells in human colon by using methylation patterns. *Proc Natl Acad Sci U S A* **98**, 10839–10844.

[48] Weinberg, R. A. 1991 Tumor suppressor genes. *Science* **254**, 1138–1146.

[49] Levine, A. J. 1993 The tumor suppressor genes. *Annu Rev Biochem* **62**, 623–651.

[50] Friend, S. H. & al. 1986 A human DNA segment with properties of the gene that predisposes to retinoblastoma and osteosarcoma. *Nature* **323**, 643–646.

[51] Nowak, M. A. & al. 2002 The role of chromosomal instability in tumor initiation. *Proc Natl Acad Sci U S A* **99**, 16226–16231.

[52] Nowak, M. A., Michor, F., Komarova, N. L., & Iwasa, Y. 2004 Evolutionary dynamics of tumor suppressor gene inactivation. *Proc Natl Acad Sci U S A* **101**, 10635–10638.

[53] Iwasa, Y., Michor, F., & Nowak, M. A. 2004 Stochastic tunnels in evolutionary dynamics. *Genetics* **166**, 1571–1579.

[54] Lengauer, C., Kinzler, K. W., & Vogelstein, B. 1998 Genetic instabilities in human cancers. *Nature* **396**, 643–649.

[55] Nasmyth, K. 2002 Segregating sister genomes: the molecular biology of chromosome separation. *Science* **297**, 559–565.

[56] Kolodner, R. D., Putnam, C. D., & Myung, K. 2002 Maintenance of genome stability in saccharomyces cerevisiae. *Science* **297**, 552–557.

[57] Shonn, M. A., McCarroll, R., & Murray, A. W. 2000 Requirement of the spindle checkpoint for proper chromosome segregation in budding yeast meiosis. *Science* **289**, 300–303.

[58] Rajagopalan, H., Jallepalli, P. V., Rago, C., Velculescu, V. E., Kinzler, K. W., Vogelstein, B., & Lengauer, C. 2004 Inactivation of hCDC4 can cause chromosomal instability. *Nature* **428**, 77–81.

[59] Loeb, L. A. 1991 Mutator phenotype may be required for multistage carcinogenesis. *Cancer Res* **51**, 3075–3079.

[60] Little, M. P. & Wright, E. G. 2003 A stochastic carcinogenesis model incorporating genomic instability fitted to colon cancer data. *Math Biosciences* **183**, 111–134.

[61] Wodarz, D. & Krakauer, D. C. 2001 Genetic instability and the evolution of angiogenic tumor cell lines. *Oncol Rep* **8**, 1195–1201.

[62] Plotkin, J. B. & Nowak, M. A. 2002 The different effects of apoptosis and DNA repair on tumorigenesis. *J Theor Biol* **214**, 452–467.

[63] Michor, F., Iwasa, Y., Lengauer, C., Vogelstein, B., & Nowak, M. A. 2005 Can chromosomal instability initiate tumorigenesis?. *Seminars in Cancer Biology* **15**, 43–49.

[64] Komarova, N. L., Lengauer, C., Vogelstein, B., & Nowak, M. A. 2002 Dynamics of genetic instability in sporadic and familial colorectal cancer. *Cancer Biol Ther* **1**, 685–692.

[65] Bardelli, A. & al. 2001 Carcinogen-specific induction of genetic instability. *Proc Natl Acad Sci U S A* **98**, 5770–5775.

[66] Kinzler, K. W. & al. 1991 Identification of a gene located at chromosome 5q21 that is mutated in colorectal cancers. *Science* **251**, 1366–1370.

[67] Knudson, A. G. 1993 Antioncogenes and human cancer. *Proc Natl Acad Sci U S A* **90**, 10914–10921.

[68] Park, B. H. & Vogelstein, B. 2000. *Clinical Oncology. Editor: M. D. Abeloff.* : Churchill Livingston.

[69] Loeb, L. A. 2001 A mutator phenotype in cancer. *Cancer Res* **61**, 3230–3239.

[70] Kinzler, K. W. & Vogelstein, B. 1996 Lessons from hereditary colon cancer. *Cell* **87**, 159–170.

[71] Perucho, M. 1996 Cancer of the microsatellite phenotype. *Biol Chem* **377**, 675–684.

[72] Mihaylov, I. S. & al. 2002 Control of DNA replication and chromosome ploidy by geminin and cyclin A. *Mol Cell Biol* **22**, 1868–1880.

[73] Fung, S. M., Ramsay, G., & Katzen, A. L. 2002 Mutations in drosophila myb lead to centrosome amplification and genomic instability. *Development* **129**, 347–359.

[74] Cahill, D. P. & al. 1998 Mutations of mitotic checkpoint genes in human cancers. *Nature* **392**, 300–303.

[75] Li, Y. & Benezra, R. 1996 Identification of a human mitotic checkpoint gene: hsMAD2. *Science* **274**, 246–248.

[76] Milner, J. & al. 1997 Transcriptional activation functions in BRCA2. *Nature* **386**, 772–773.

[77] Pangilinan, F. & al. 1997 Mamalian BUB1 protein kinases: map positions and in vivo expression. *Genomics* **46**, 379–388.

[78] Yarden, R. I. & al. 2002 BRCA1 regulates the G2/M checkpoint by activating Chk1 kinase upon DNA damage. *Nat Genet* **30**, 265–269.

[79] Sieber, O. M. & al. 2002 Analysis of chromosomal instability in human colorec-

tal adenomas with two mutational hits at APC. *Proc Natl Acad Sci U S A* **99**, 14800–14803.

[80] Shih, I. M. & al. 2001 Evidence that genetic instability occurs at an early stage of colorectal tumorigenesis. *Cancer Res* **61**, 818–822.

[81] Breivik, J. 2001 Don't stop for repairs in a war zone: Darwinian evolution unites genes and environment in cancer development. *Proc Natl Acad Sci U S A* **98**, 5379–5381.

[82] Lengauer, C., Kinzler, K. W., & Vogelstein, B. 1997 Genetic instability in colorectal cancer. *Nature* **386**, 623–627.

[83] Michor, F., Iwasa, Y., Lengauer, C., & Nowak, M. A. 2005 Dynamics of colorectal cancer. *Seminars in Cancer Biology,* in press.

[84] Society, A. C. 2005. *Cancer Facts and Figures 2005.* Atlanta: American Cancer Society.

[85] Sidransky, D. & al. 1992 Identification of ras oncogene mutations in the stool of patients with curable colorectal tumors. *Science* **256**, 102–105.

[86] Traverso, G. & al. 2002 Detection of APC mutations in fecal DNA from patients with colorectal tumors. *N E J M* **346**, 311–320.

[87] Fearon, E. R. & Vogelstein, B. 1990 A genetic model for colorectal tumorigenesis. *Cell* **61**, 759–767.

[88] Powell, S. M. & al. 1992 APC mutations occur early during colorectal tumorigenesis. *Nature* **359**, 235–237.

[89] Su, L. K., Vogelstein, B., & Kinzler, K. W. 1993 Association of the APC tumor suppressor protein with catenins. *Science* **262**, 1734–1737.

[90] He, T. C. & al. 1998 Identification of c-MYC as a target of the APC pathway. *Science* **281**, 1509–1512.

[91] Morin, P. J. & al. 1997 Activation of beta-catenin-Tcf signaling in colon cancer by mutations in beta catenin or APC. *Science* **275**, 1787–1790.

[92] Su, L. K. & al. 1992 Multiple intestinal neoplasia caused by a mutation in the murine homolog of the APC gene. *Science* **256**, 669–670.

[93] Aaltonen, L. A. & al. 1993 Clues to the pathogenesis of familial colorectal cancer. *Science* **260**, 812–816.

[94] Jallepalli, P. V. & Lengauer, C. 2001 Chromosome segregation and cancer: cutting through the mystery. *Nat Rev Can* **1**, 109–117.

[95] Loeb, L. A., Springgate, C. F., & Battula, N. 1974 Errors in DNA replication as a basis of malignant changes. *Cancer Res* **34**, 2311–2321.

[96] Duesberg, P. H. 2003 Are cancers dependent on oncogenes or on aneuploidy?. *Cancer Genet Cytogenet* **143**, 89–91.

[97] Sparks, A. B. & al. 1998 Mutational analysis of the APC/beta-catenin/Tcf pathway in colorectal cancer. *Cancer Res* **58**, 1130–1134.

[98] Huang, J. & al. 1996 Apc mutations in colorectal tumors with mismatch repair deficiency. *P N A S* **93**, 9049–9054.

[99] Herrero-Jimenez, P. & al. 2000 Population risk and physiological rate parameters for colon cancer. the union of an explicit model for carcinogenesis with the public health records of the united states. *Mut Res* **447**, 73–116.

[100] Pinsky, P. F. 2000 A multi-stage model of adenoma development. *JTB* **207**, 129–143.

[101] Zahl, P. H. 1995 A proportional regression model for 20 year survival of colon cancer in norway. *Stat Med* **14**, 1249–1261.

[102] Mehl, L. E. 1991 A mathematical computer simulation model for the development of colonic polyps and colon cancer. *J Surg Oncol* **47**, 243–252.

[103] Lipkin, M., Shelrock, P., & Bell, B. 1962 Generation time of epithelial cells in the human colon. *Nature* **195**, 175–177.

[104] Nagase, H. & Nakamura, Y. 1993 Mutations of the APC gene. *Hum Mut* **2**, 425–434.

[105] Lipkin, M., Bell, B., & Shelrock, P. 1963 Cell proliferation kinetics in the

gastrointestinal tract of man. I. cell renewal in colon and rectum. *J Clin Invest* **42**, 767–776.

[106] Morin, P. J., Vogelstein, B., & Kinzler, K. W. 1996 Apoptosis and APC in colorectal tumorigenesis. *Proc Natl Acad Sci U S A* **93**, 7950–7954.

[107] Kinzler, K. W. & al. 1991 Identification of FAP locus genes from chromosome 5q21. *Science* **253**, 661–665.

[108] Meineke, F. A., Potten, C. S., & Loeffler, M. 2001 Cell migration and organization in the intestinal crypt using a lattice-free model. *Cell Prolif* **34**, 253–266.

[109] Michor, F., Iwasa, Y., Hughes, T. P., Branford, S., Shah, N. P., Sawyers, C. L., & Nowak, M. A. 2005 Dynamics of chronic myeloid leukemia. *Nature,* in press.

[110] McKenzie, S. 2004. *Clinical Laboratory Hematology.* : Prentice Hall.

[111] Domen, J. & Weissman, I. L. 1999 Self-renewal, differentiation or death: regulation and manipulation of hematopoietic stem cell fate. *Mol Med Today* **5**, 201–208.

[112] Spangrude, G. J., Heimfeld, S., & Weissman, I. L. 1988 Purification and characterization of mouse hematopoietic stem cells. *Science* **241**, 58–62.

[113] Reya, T., Morrison, S. J., Clarke, M. C., & Weissman, I. L. 2001 Stem cells, cancer, and cancer stem cells. *Nature* **414**, 105–111.

[114] Morrison, S. J., Uchida, N., & Weissman, I. L. 1995 The biology of hematopoietic stem cells. *Annu Rev Cell Dev Biol* **11**, 35–71.

[115] Akashi, K., Traver, D., Miyamoto, T., & Weissman, I. L. 2000 A clonogenic common myeloid progenitor that gives rise to all myeloid lineages. *Nature* **404**, 193–197.

[116] Zubair, A. C., Silberstein, L., & Ritz, J. 2002 Adult hematopoietic stem cell plasticity. *Transfusion* **42**, 1096–1101.

[117] Gussoni, E., Soneoka, Y., Strickland, C. D., Buzney, E. A., Khan, M. K., Flint, A. F., Kunkel, L. M., & Mulligan, R. C. 1999 Dystrophin expression in the mdx mouse restored by stem cell transplantation. *Nature* **401**, 390–394.

[118] Jordan, C. T. & Lemischka, I. R. 1990 Clonal and systemic analysis of long-term hematopoiesis in the mouse. *Genes Dev* **4**, 220–232.

[119] Potten, C. S. & Loeffler, M. 1990 Stem cells: attributes, cycles, spirals, pitfalls and uncertainties. lessons for and from the crypt. *Development* **110**, 1001–1020.

[120] Nilsson, S. K., Johnston, H. M., & COverdale, J. A. 2001 Spatial localization of transplanted hemopoietic stem cells: inferences for the localization of stem cell niches. *Blood* **97**, 2293–2299.

[121] Schofield, R. 1978 The relationship between the spleen colony-forming cell and the haemopoietic stem cell. *Blood Cells* **4**, 7–25.

[122] Gidali, J. & Lajtha, L. G. 1972 Regulation of haemopoietic stem cell turnover in partially irradiated mice. *Cell Tissue Kinet* **5**, 147–157.

[123] Moore, K. A., Ema, H., & Lemischka, I. R. 1997 In vitro maintenance of highly purified, transplantable hematopoietic stem cells. *Blood* **89**, 4337–4347.

[124] Wright, D. E., Wagers, A. J., Gulati, A. P., Johnson, F. L., & Weissman, I. L. 2001 Physiological migration of hematopoietic stem and progenitor cells. *Science* **294**, 1933–1936.

[125] Morrison, S. J. & Weissman, I. L. 1994 The long-term repopulating subset of hematopoietic stem cells is deterministic and isolatable by phenotype. *Immunity* **1**, 661.

[126] Morrison, S. J., Wandycz, A. M., Hemmati, H. D., Wright, D. E., & Weissman, I. L. 1997 Identification of a lineage of multipotent hematopoietic progenitors. *Development* **124**, 1929–1939.

[127] Harrison, D. E. & Astle, C. M. 1997 Short- and long-term multilineage repopulating hematopoietic stem cells in late fetal and newborn mice: models for human umbilical cord blood. *Blood* **90**, 174–181.

[128] Cheshier, S. H., Morrison, S. J., Liao, X., & Weissman, I. L. 1999 In vivo proliferation and cell cycle kinetics of long-term self-renewing hematopoietic stem cells. *Proc Natl Acad Sci USA* **96**, 3120–3125.

[129] Sawyers, C. L. 1999 Chronic myeloid leukemia. *N E J M* **340**, 1330–1340.

[130] Reya, T., Morrison, S. J., Clarke, M. F., & Weissman, I. L. 2001 Stem cells, cancer, and cancer stem cells. *Nature* **414**, 105–111.

[131] Nowell, P. C. 1997 Genetic alterations in leukemias and lymphomas: impressive progress and continuing complexity. *Cancer Genet Cytogenet* **94**, 13–19.

[132] Rowley, J. D. 1973 Letter: A new consistent chromosomal abnormality in chronic myelogenous leukaemia identified by quinacrine fluorescence and giemsa staining. *Nature* **243**, 290–293.

[133] Gishizky, M. L. & Witte, O. N. 1992 Initiation of deregulated growth of multipotent progenitor cells by BCR-ABL in vitro. *Science* **256**, 836–839.

[134] Heisterkamp, N., Jenster, G., ten Hoeve, J., Zovich, D., Pattengale, P. K., & Groffen, J. 1990 Acute leukaemia in BCR/ABL transgenic mice. *Nature* **344**, 251–253.

[135] Ichimaru, M., Ishimaru, T., Mikami, M., Yamada, Y., & Ohkita, T. 1981. *Incidence of leukemia in a fixed cohort of atomic bomn survivors and controls, Hiroshima and Nagasaki October 1950-December 1978: Technical Report RERF TR 13-81.* Hiroshima: Radiation Effects Research Foundation.

[136] Bernstein, R. 1988 Cytogenetics of chronic myelogenous leukemia. *Semin Hematol* **25**, 20–34.

[137] Sill, H., Goldman, J. M., & Cross, N. C. 1995 Homozygous deletions of the p16 tumor-suppressor gene are associated with lymphoid transformation of chronic myeloid leukemia. *Blood* **85**, 2013–2016.

[138] Ahuja, H., Bar-Eli, M., Arlin, Z., Advani, S., Allen, S. L., Goldman, J., Snyder, D., Foti, A., & Cline, M. 1991 The spectrum of molecular alterations in the evolution of chronic myelocytic leukemia. *J Clin Invest* **87**, 2042–2047.

[139] Dutcher, J. P. & Wiernik, P. H. 2000 Accelerated and blastic phase of chronic myeloid leukemia. *Curr Treat Options Oncol* **1**, 51–62.

[140] Horowitz, M. M., Rowlings, P. A., & Passweg, J. R. 1996 Allogeneic bone marrow transplantation for CML: a report from the international bone marrow transplant registry. *Bone Marrow Transplant* **17 Suppl 3**, S5–6.

[141] Druker, B. J. & al. 2001 Activity of a specific inhibitor of the BCR-ABL tyrosine kinase in the blast crisis of chronic myeloid leukemia and acute lymphoblastic leukemia with the philadelphia chromosome. *N E J M* **344**, 1038–1042.

[142] Sawyers, C. L. & al. 2002 Imatinib induces hematologic and cytogenetic responses in patients with chronic myelogenous leukemia in myeloid blast crisis: results of a phase ii study. *Blood* **99**, 3530–3539.

[143] Goldman, J. M. & Melo, J. V. 2003 Chronic myeloid leukemia - advances in bology and new approaches to treatment. *N E J M* **349**, 1451–1464.

[144] Druker, B. J., Tamura, S., Buchdunger, E., Ohno, S., Segal, G. M., Fanning, S., Zimmermann, J., & Lydon, S. B. 1996 Effects of a selective inhibitor of the ABL tyrosine kinase on the growth of BCR-ABL positive cells. *Nat Med* **2**, 561–566.

[145] Deininger, M. W., Goldman, J. M., Lydon, N., & Melo, J. V. 1997 The tyrosine kinase inhibitor CGP57148B selectively inhibits the growth of BCR-ABL positive cells. *Blood* **90**, 3691–3698.

[146] S, B, M. & al. 2000 The tyrosine kinase inhibitor STI571, like interferon-alpha, preferentially reduces the capacity for amplification of granulocyte-macrophage progenitors from patients with chronic myeloid leukemia. *Exp Hematol* **28**, 551–557.

[147] Holtz, M. S. & al. 2002 Imatinib mesylate (STI571) inhibits growth of primitive malignant progenitors in chronic myelogenous leukemia through reversal of abnormally increased proliferation. *Blood* **99**, 3792–3800.

[148] Hughes, T. P. & al. 2003 Frequency of major molecular responses to imatinib or interferon alfa plus cytarabine in newly diagnosed chronic myeloid leukemia. *N E J M* **349**, 1423–1432.

[149] Bhatia, R. & al. 2003 Persistence of malignant hematopoietic progenitors in chronic myelogenous leukemia patients in complete cytogenetic remission following imatinib mesylate treatment. *Blood* **101**, 4701–4707.

[150] Chu, S., Xu, H., Shah, N. P., Snyder, D. S., Forman, S. J., Sawyers, C. L., & Bhatia, R. 2004 Detection of BCR-ABL kinase mutations in CD34+ cells from chronic myelogenous leukemia patients in complete cytogenetic remission on imatinib mesylate treatment.. *Blood*.

[151] Gorre, M. E., Mohammed, M., Ellwood, K., Hsu, N., Paquette, R., Rao, P. N., & Sawyers, C. L. 2001 Clinical resistance to STI-571 cancer therapy caused by bcr-abl gene mutation or amplification. *Science* **293**, 876–880.

[152] Branford, S. & al. 2002 High frequency of point mutations clustered within the adenosine triphosphate-binding region of BCR/ABL in patients with chronic myeloid leukemia or ph-positive acute lymphoblastic leukemia who develop imatinib (STI571) resistance. *Blood* **99**, 3472–3475.

[153] Branford, S. & al. 2003 Detection of BCR-ABL mutations in patients with CML treated with imatinib is virtually always accompanied by clinical resistance, and mutations in the ATP phosphate-binding loop (P-loop) are associated with a poor prognosis. *Blood* 102.

[154] Branford, S. & al. 2004 Real-time quantitative PCR analysis can be used as a primary screen to identify imatinib-treated patients with CML who have BCR-ABL kinase domain mutations. *Blood*.

[155] Hochhaus, A. & al. 2002 Molecular and chromosomal mechanisms of resistance to imatinib (STI571) therapy. *Leukemia* **16**, 2190–2196.

[156] von Bubnoff, N. & al. 2002 BCR-ABL gene mutations in relation to clinical resistance of philadelphia-chromosome-positive leukaemia to STI571: a prospective study. *Lancet* **359**, 487–491.

[157] Kreill, S. & al. 2003 Management and clinical outcome of CML patients after imatinib resistance associated with ABL kinase domain mutations. *Blood* **102**, 71a.

[158] Al-Ali, H. K. & al. 2004 High incidence of BCR-ABL kinase domain mutations and absence of mutations of the pdgfr and kit activation loops in cml patients with secondary resistance to imatinib. *Hematol J* **5**, 55–60.

[159] Shah, N. P., Nicoll, J. M., Nagar, B., Gorre, M. E., Paquette, R. L., Kuriyan, J., & Sawyers, C. L. 2002 Multiple BCR-ABL kinase domain mutations confer polyclonal resistance to the tyrosine kinase inhibitor imatinib (STI571) in chronic phase and blast crisis chronic myeloid leukemia. *Cancer Cell* **2**, 117–125.

[160] Roche-Lestienne, C., Soenen-Cornu, V., Grardel-Duflos, N., Lai, J. L., Philippe, N., Facon, T., Fenaux, P., & Preudhomme, C. 2002 Several types of mutations of the ABL gene can be found in chronic myeloid leukemia patients resistant to STI571, and they can pre-exist to the onset of treatment. *Blood* **100**, 1014–1018.

[161] Roche-Lestienne, C. & al. 2003 A mutation conferring resistance to imatinib at the time of diagnosis of chronic myeloid leukemia. *N E J M* **348**, 2265–2266.

[162] Branford, S. & al. 1999 Monitoring chronic myeloid leukaemia therapy by real-time quantitative PCR in blood is a reliable alternative to bone marrow cytogenetics. *Br J Haematol* **107**, 587–599.

[163] Cortes, J., O'Brien, S., & Kantarjian, H. 2004 Discontinuation of imatinib therapy after achieving molecular response. *Blood* **104**, 2204–2205.

[164] Higashi, T. & al. 2004 Imatinib mesylate-sensitive blast crisis immediately after discontinuation of imatinib mesylate therapy in chronic myelogenous leukemia: report of two cases. *Am J Hematol* **76**, 275–278.

[165] Graham, S. M. & al. 2002 Primitive, quiescent, philadelphia-positive stem cells from patients with chronic myeloid leukemia are insensitive to STI571 in vitro. *Blood* **99**, 319–325.

[166] Branford, S. & al. 2003 The incidence of BCR-ABL kinase mutations in chronic myeloid leukemia patients is as high in the second year of imatinib therapy as the first but survival after mutation detection is significantly longer for patients with mutations detected in the second year of therapy. *Blood* **102**, 414.

[167] Holyoake, T. L., Jiang, X., Drummond, M. W., Eaves, A. C., & Eaves, C. J. 2002 Elucidating critical mechanisms of deregulated stem cell turnover in the chronic phase of chronic myeloid leukemia. *Leukemia* **16**, 549–558.

[168] Holyoake, T. L., Jiang, X., Eaves, C., & Eaves, A. 1999 Isolation of a highly

quiescent subpopulation of primitive leukemic cells in chronic myeloid leukemia. *Blood* **94**, 2056–2064.

[169] Holyoake, T. L. & al. 2001 Primitive quiescent leukemic cells from patients with chronic myeloid leukemia spontaneously initiate factor-independent growth in vitro in association with up-regulation of expression of interleukin-3. *Blood* **97**, 720–728.

[170] Shah, N. P. & al. 2004 Overriding imatinib resistance with a novel ABL kinase inhibitor. *Science* **305**, 399–401.

[171] Mahon, F. X. & al. 2003 MDR1 gene overexpression confers resistance to imatinib mesylate in leukemia cell line models. *Blood* **101**, 2368–2373.

[172] Chaudhary, P. M. & Roninson, I. B. 1991 Expression and activity of p-glycoprotein, a multidrug efflux pump, in human hematopoietic stem cells. *Cell* **66**, 85.

[173] Weinstein, I. B. 2002 Addiction to oncogenes–the achilles heal of cancer. *Science* **297**, 63–64.