



Cite this article: Werner B, Dingli D, Traulsen A. 2013 A deterministic model for the occurrence and dynamics of multiple mutations in hierarchically organized tissues. *J R Soc Interface* 10: 20130349. <http://dx.doi.org/10.1098/rsif.2013.0349>

Received: 17 April 2013

Accepted: 10 May 2013

Subject Areas:

biomathematics, computational biology, systems biology

Keywords:

cancer, clonal extinction and diversity, mathematical modelling

Author for correspondence:

Arne Traulsen

e-mail: traulsen@evolbio.mpg.de

A deterministic model for the occurrence and dynamics of multiple mutations in hierarchically organized tissues

Benjamin Werner¹, David Dingli² and Arne Traulsen¹

¹Evolutionary Theory Group, Max Planck Institute for Evolutionary Biology, Plön, Germany

²Division of Hematology, Mayo Clinic, College of Medicine, Rochester, MN, USA

Cancers are rarely caused by single mutations, but often develop as a result of the combined effects of multiple mutations. For most cells, the number of possible cell divisions is limited because of various biological constraints, such as progressive telomere shortening, cell senescence cascades or a hierarchically organized tissue structure. Thus, the risk of accumulating cells carrying multiple mutations is low. Nonetheless, many diseases are based on the accumulation of such multiple mutations. We model a general, hierarchically organized tissue by a multi-compartment approach, allowing any number of mutations within a cell. We derive closed solutions for the deterministic clonal dynamics and the reproductive capacity of single clones. Our results hold for the average dynamics in a hierarchical tissue characterized by an arbitrary combination of proliferation parameters. We show that hierarchically organized tissues strongly suppress cells carrying multiple mutations and derive closed solutions for the expected size and diversity of clonal populations founded by a single mutant within the hierarchy. We discuss the example of childhood acute lymphoblastic leukaemia in detail and find good agreement between our predicted results and recently observed clonal diversities in patients. This result can contribute to the explanation of very diverse mutation profiles observed by whole genome sequencing of many different cancers.

1. Introduction

The lifespan of most cells in biological organisms is limited, and usually the life expectancy of the organism exceeds this time by orders of magnitude [1,2]. As cells are continuously lost, mechanisms to replenish the cell pool have evolved in organisms, enabling sustained cell production during the lifetime [3]. Often this is realized by hierarchically organized tissue structures. At the root of the hierarchy are a few tissue-specific stem cells, combining two properties—self-renewal and differentiation potential [4]. During cell proliferation, cells differentiate and become increasingly specialized to perform specific functions within the hierarchy. After some differentiation steps, the complete spectrum of functional cells can be obtained [5–10]. A prominent example is the haematopoietic system [6–11], but other tissues such as skin [12,13] or colon [14] are also hierarchically organized.

A large number of cell divisions are indispensable to life; however, they are unavoidably accompanied by mutations. Typically, these cells are washed out of the hierarchy and thus, especially if they arise in relatively differentiated cells, the associated mutations are lost in the long run [10,15]. But cells with multiple mutational hits might persist for a long time, increasing the risk of accumulation of additional mutations during cell proliferation, which can ultimately lead to cancer. Although some cancers seem to be caused by single mutation hits, for example the BCR-ABL oncogene occurring in stem cells in chronic myeloid leukaemia [16] or the PML-RARA oncogene occurring in more differentiated cells in acute promyelocytic leukaemia [17], they are rare.

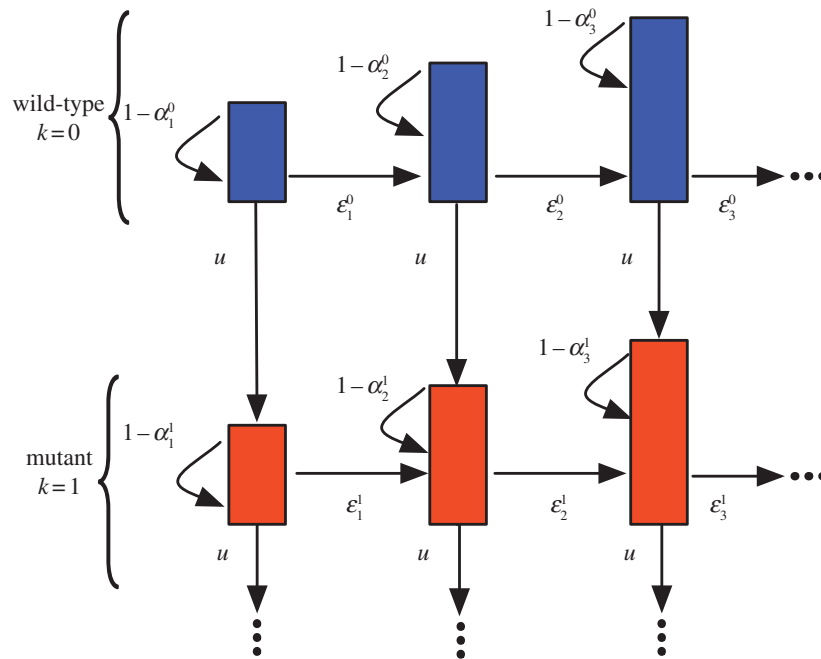


Figure 1. Schematic of the compartment structure of multiple mutations and the corresponding transition rates. The top compartments contain cells carrying no mutation. The bottom compartments contain cells carrying one mutation. Compartments to the right represent more specialized cell stages and arrows represent transition probabilities, where ε denotes the differentiation probability, u denotes the mutation rate of cells and $\alpha_i^k = \varepsilon_i^k + u$. Initially, no mutated cells are present in the hierarchy. We then determine, how many cells are acquired from the founder compartment (top left) and investigate how many cells with k mutations are on average expected at any stage of the hierarchy. (Online version in colour.)

The majority of cancers are triggered by at least a handful of mutations [18–20]. The recent progress in genome sequencing techniques has allowed, in some cases, the classification of cancer-initiating mutations; in other cases, the underlying mutations remain unknown [21]. However, many of these studies reveal a very diverse mutation landscape, indicating the existence of several cancer-initiating driver mutations and additional alterations that have a small or even no impact on cancer development, the so-called passenger mutations [22–29]. The precise impact of passenger mutations on cancer progression is still under discussion, but the typical assumption is that they are neutral and do not affect the proliferation properties of cells [30,31]. In particular, this holds for synonymous mutations that do not have any consequences for protein structure or function [32].

In mathematical and computational approaches, compartment models are frequently used to describe cell dynamics in hierarchically organized tissue structures. Many of these studies investigate the effects of stem cell mutations and the related clinical implications [6,8,33–36]. Also the stochastic effects of tissue homeostasis are analysed [37–39], highlighting that cancer-driving mutations can in principle disappear by chance (stochastic extinction). The interplay of stem cell and progenitor cell mutations and their impact on cancer initiation are discussed [40], and game theoretical approaches allow modelling of the evolutionary aspects of tissue homeostasis and intercell competition [41,42]. Often, these studies investigate the effects of cells carrying one or a very few specific mutations and assume either constant population size or only minimal hierarchies.

Here, we focus on the presence of cells carrying multiple mutations within a hierarchically organized tissue. We show mathematically that the hierarchical organization strongly suppresses cells carrying multiple mutations and thus reduces the risk of cancer initiation. Closed solutions for the total cell population that arises from a single (mutant) cell are derived,

and from this the expected diversity of the mutation landscape and the clonal size can be described. This enables a better understanding of the expected diversity in mutation landscapes that are observed in both healthy and cancerous tissues.

1.1. Mathematical model

The hierarchical tissue organization is typically modelled by a multi-compartment approach [6,10]. Each compartment represents a certain differentiation in the stage of cells. At the root of the hierarchy are stem cells ensuring a continuous influx of cells. A proliferating cell in compartment i divides and the two daughter cells differentiate and migrate into the next downstream compartment ($i + 1$) with probability ε , increasing the downstream compartment by two cells, mutates with probability u or self-renews within its own compartment with probability $1 - \varepsilon - u$. Mutated cells stay in the hierarchy. If a mutated cell proliferates, it differentiates with probability ε into the next downstream compartment, it self-renews with probability $1 - \varepsilon - u$, or it mutates with probability u again, leading to a cell with two (or more) mutations. All possible outcomes of a cell proliferation are depicted in figure 1. The directions of the arrows point towards the accessible cell states and the labels give the transition probabilities. We allow arbitrary parameters and introduce ε_i^k as the differentiation probability of cells in compartment i carrying k mutations. Asymmetric cell divisions are not explicitly implemented, as they can be absorbed in the differentiation probabilities on the population level. The fate of a cell's offspring is determined based on the probabilities ε_i^k . Cells proliferate with a rate r_i in each compartment i . Usually, cells in upstream compartments proliferate slowly and cell proliferation speeds up in downstream compartments (i.e. $r_i < r_{i+1}$). This general framework is very flexible and different tissue structures can be represented.

1.2. Stochastic individual-based simulations

We implement individual-based stochastic simulations of the cell dynamics in hierarchically organized tissue structures. We use an implementation of the Gillespie algorithm [43,44]. Originally introduced to simulate chemical reactions, it allows us to reproduce exact stochastic trajectories of the system. Each cell has an individual representation. Thus, the complete clonal history of cells within the hierarchy can be recorded. If a cell is chosen for reproduction (determined by the Gillespie method), it differentiates, self-renews or mutates according to the probabilities ε_i^k and u . The parameters of the simulated system are described later and chosen to represent human haematopoiesis.

1.3. The haematopoietic system

In the following, we focus on the haematopoietic system. There, approximately 400 stem cells replenish the haematopoietic cell pool [45,46]. Each stem cell divides approximately once a year [45,47]. Cell proliferation is assumed to increase exponentially with compartment number i , $r_i = \gamma^i r_0$, with $\gamma = 1.26$ and r_0 corresponds to the proliferation rate of stem cells. The differentiation probability is assumed to be constant, $\varepsilon = 0.85$, for all non-stem cell compartments, and in total $i = 31$ compartments are needed to ensure a daily bone marrow output of approximately 3.5×10^{11} cells [6,10].

2. Results

2.1. Time continuous dynamics of multiple mutations

We describe the deterministic dynamics of a cell population within a hierarchically organized tissue structure, which initially carries no mutation. A cell may commit further into the hierarchy (differentiate), mutate or self-renew. This occurs with probability ε , u and $1 - \varepsilon - u$, respectively. In figure 1, a schematic of the resulting hierarchical structure is shown. Compartments to the right represent downstream compartments of more specialized (differentiated) cells, while compartments to the bottom represent states of cells which accumulated an additional mutation. During one cell division, a cell either mutates and moves one compartment to the bottom, differentiates and produces two cells in the next downstream compartment to the right or self-renews and produces an additional cell within its original compartment. This leads to an expansion of clonal populations within the hierarchy that potentially accumulates several (distinct) mutations during the differentiation process. This is schematically shown in figure 2.

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

Assume that in compartment i there are $N_i^k(t)$ cells carrying k mutations at time t . The number of these cells increases as a result of influx from the upstream compartment at a rate $2r_{i-1}\varepsilon_{i-1}^k$, mutations at a rate $r_i u$ and self-renewal at a rate

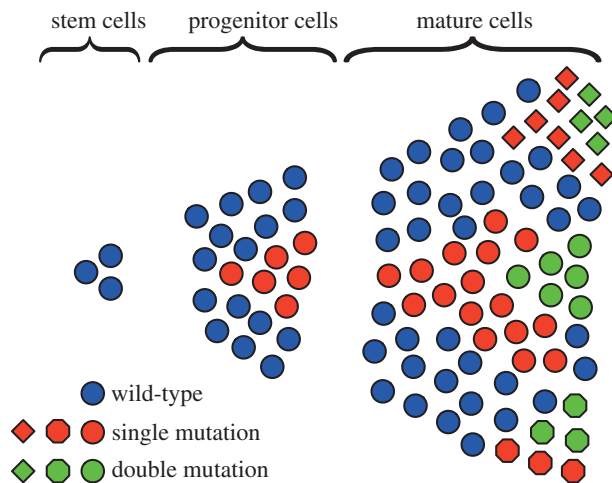


Figure 2. Clonal expansion within a hierarchically organized tissue. Cell proliferation is driven by a few slow-dividing stem cells, giving rise to faster dividing progenitor cells. After some differentiation steps, the mature tissue cells are obtained. Initially cells have no mutations, but mutants can arise and expand within the hierarchy. These cells either vanish or gain an additional mutation, which again potentially spreads within the hierarchy. Different colours code for a different number of mutations, whereas different shapes indicate different mutations. (Online version in colour.)

$r_i(1 - \varepsilon_i^k - u)$. Cells are lost either by mutation at a rate $r_i u$ or by differentiation at a rate $r_i \varepsilon_i^k$. The deterministic description of the hierarchical compartment model becomes a system of coupled differential equations [10], given by

$$\dot{N}_i^k(t) = \begin{cases} r_i(1 - 2\alpha_i^k)N_i^k(t) + 2r_{i-1}\varepsilon_{i-1}^k N_{i-1}^k(t) & k = 0 \\ r_i(1 - 2\alpha_i^k)N_i^k(t) + 2r_{i-1}\varepsilon_{i-1}^k N_{i-1}^k(t) + r_i u N_i^{k-1}(t) & k > 0. \end{cases} \quad (2.1)$$

Here, $\alpha_i^k = \varepsilon_i^k + u$ denotes the probability that a cell with k mutations leaves compartment i . Typically, α_i^k is very close to ε . A model for stochastic cell dynamics in the stem cell compartment for neutral and non-neutral mutations can be found in [39,48]. In those papers, the stochastic Moran process is used to investigate the extinction and fixation probabilities of stem cell mutations. The deterministic stem cell-driven cell replenishment in hierarchical tissues is studied in detail in [10]. However, in that prior work, the effects arising from additional mutations were neglected. Here, we focus on non-stem cell-driven clonal dynamics. We explicitly allow for an arbitrary number of mutational hits at any stage of the hierarchy, but we neglect a continuous influx of mutated cells from the stem cell level. This assumption gives the condition $N_0^k(t) = 0$. The initial condition

$$N_i^k(0) = \begin{cases} n_0 & i = 1 \quad k = 0, \\ 0 & \text{otherwise,} \end{cases} \quad (2.2)$$

corresponds to initially n_0 cells in compartment 1 carrying no mutation. One can imagine a neutral marker approach, in which one cell in the hierarchy is genetically marked, and one considers the clonal population arising from this marked cell [49]. Although we neglect a continuous influx of mutated cells from the stem cell compartment, stem cell mutations can be implemented indirectly. Our approach allows for altered cell proliferation properties of the founder cell, potentially derived by a mutation at the stem cell level. For a constant differentiation probability, i.e. the case where all α_i^k and all ε_i^k are identical, equation (2.1) can be solved recursively.

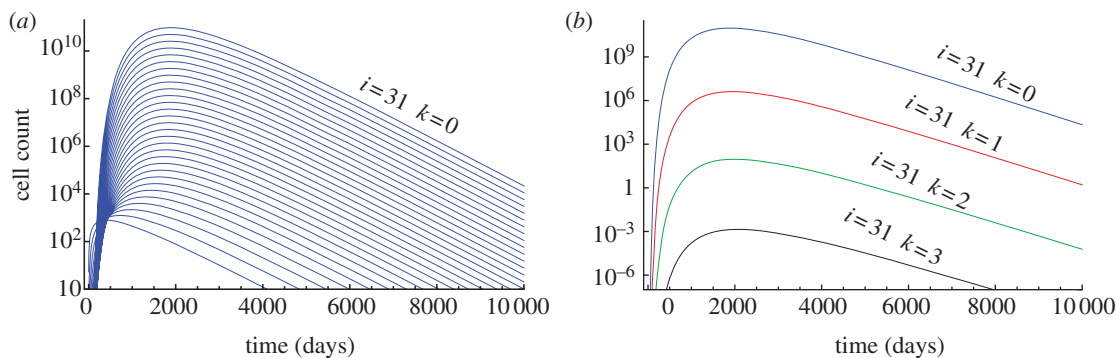


Figure 3. (a) Number of cells carrying no mutation in compartments 1–31 arising from compartment 1 containing 1000 cells. Lines show equation (2.3), with parameters $n_0 = 1000$, $\epsilon = 0.85$, $\gamma = 1.26$, $u = 10^{-6}$ and $r_0 = 1/400$. Cells are more likely to differentiate than to self-renew and thus progressively travel into more committed compartments. Initially, the cell count increases, but cells get washed out in the long run. The time scale is determined by the number of stem cell divisions. A stem cell is assumed to divide once a year, thus after 400 stem cell divisions a year has passed. (b) Count of cells carrying 0–3 mutations in compartment 31, given by equation (2.4). We used the same parameters as in (a). Cells carrying multiple mutations are exponentially suppressed. (Online version in colour.)

The number of cells in compartment i carrying no mutation changes with time as

$$N_i^0(t) = n_0 \frac{(2\epsilon)^{i-1}}{(2\alpha - 1)^{i-1}} \left(\prod_{j=1}^{i-1} r_j \right) \sum_{l=1}^i \frac{e^{\eta(1-2\alpha)t}}{\prod_{h=1}^l (r_h - r_l)}. \quad (2.3)$$

The solution can also be derived recursively for cells carrying k mutations in compartment i and becomes

$$N_i^k(t) = n_0 \sum_{h=0}^k \frac{u^{k-h} g_i^h}{k!} \frac{(2\epsilon)^{i-1}}{(2\alpha - 1)^{i+h-1}} \left(\prod_{j=1}^{i-1} r_j \right) \times \sum_{l=1}^i \frac{(r_l t)^{k-h} e^{\eta(1-2\alpha)t}}{\prod_{h=1}^l (r_h - r_l)}, \quad (2.4)$$

where g_i^h is a combinatoric parameter which is given for cells carrying up to three mutations by

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

(2.5)

If $\alpha < 0.5$, non-stem cells will continuously accumulate in downstream compartments. The probability of self-renewal in this case is larger than the probability of differentiation. This scenario seems to be realized in certain blood cancers. For example, in acute promyelocytic leukaemia an abnormal increase in immature granulocytes and promyelocytes is observed, resulting from a block of cell differentiation at a late progenitor cell stage [17,50]. However, these cases are rare.

For $\alpha > 0.5$, the solution becomes a clonal wave, travelling through the hierarchy in time. In this case, the probability of differentiation is larger than the probability of self-renewal and thus cells progressively travel downstream (figure 3). The cell population founded by a single non-stem cell expands within the hierarchy initially, but gets washed out and vanishes in the long run. This is believed to be true for healthy homeostasis. For example, for the haematopoietic system, the differentiation probability was estimated to be $\epsilon = 0.85$ [6]. As by far the most cell proliferations occur at the progenitor and more committed differentiation stages, this provides a natural protection for the organism against the accumulation of multiple mutations, as the survival time of most (non-stem cell-like) mutations is finite.

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

$$t_{\max}^k = \frac{k}{(2\alpha - 1)r_1}. \quad (2.6)$$

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

$$N_1^k(t_{\max}) = \frac{(uk)^k}{k!(2\alpha - 1)^k} e^{-k} \approx \frac{u^k}{(2\alpha - 1)^k} \frac{1}{\sqrt{2\pi k}}, \quad (2.7)$$

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

$$m_i^k(t) = r_i \alpha_i^k \int_0^t ds N_i^k(s), \quad (2.8)$$

and the reproductive capacity can be derived by taking the time limit to infinity. The general solution (2.4) allows us to carry out the integral exactly by integration by parts. However, the problem can be tackled from a different perspective, leading to a more transparent solution of (2.8) that is easier to handle.

2.2. Cell reproductive capacity

We call the cell subpopulation within a compartment i , which is derived by a single founder cell in an upstream compartment, the reproductive capacity of this founder cell. This idea directly corresponds to the method of neutral markers. We imagine a genetically marked cell somewhere in the hierarchy and count the offspring of this cell at any stage of the hierarchy. This corresponds to the total count of cells with the same colour in figure 2.

Assume a single cell carrying no mutation in compartment 1. This cell differentiates with probability ϵ_1^0 into the next downstream compartment, mutates with probability u or produces an additional cell in compartment 1 with

probability $1 - \varepsilon_1^0 - u$. We first discuss the probability that a cell leaves compartment 1 after exactly l cell divisions. A cell can leave a compartment either by mutation or by differentiation, before which the cell has to undergo $l - 1$ self-renewals. Thus, this probability becomes $(\varepsilon + u)(1 - \varepsilon - u)^{l-1}$. During this time, the cell population in compartment 1 derived from this single cell increases to 2^{l-1} cells, if all daughter cells share the same proliferation probabilities. With this, the reproductive capacity of a single cell in compartment 1 is on average

$$m_1^0 = \sum_{l=0}^{\infty} \alpha_1^0 2^l (1 - \alpha_1^0)^l = \begin{cases} \frac{\alpha_1^0}{2\alpha_1^0 - 1} & \alpha_1^0 > \frac{1}{2} \\ \infty & \alpha_1^0 \leq \frac{1}{2}. \end{cases} \quad (2.9)$$

The sum becomes infinite if $\alpha \leq 0.5$, as the probability of producing offspring in the founder compartment is higher than the probability of leaving the compartment. Thus, the cell population continuously increases. Of course, under normal conditions, cells do not have an unlimited capacity to divide and serial telomere erosion, among others, will impose a physical limit on the number of divisions a cell can undergo [1,51]. The biologically more relevant case is $\alpha > 0.5$ and cells tend to differentiate into more committed compartments. In this case, the total number of offspring cells that arise from a single cell (i.e. a clone) is finite and given by (2.9). The number of cells m_i^0 in compartment i carrying no mutation increases because of the influx of cells via differentiation from compartment $i - 1$ and the expansion of these cells owing to self-renewal in compartment i . Thus, we can write for the reproductive capacity of cells in compartment i without mutation

$$m_i^0 = 2 \frac{\varepsilon_{i-1}^0}{\alpha_{i-1}^0} m_{i-1}^0 \frac{\alpha_i^0}{2\alpha_i^0 - 1} = \frac{\alpha_i^0}{2\alpha_i^0 - 1} \prod_{l=1}^{i-1} \frac{2\varepsilon_l^0}{2\alpha_l^0 - 1}. \quad (2.10)$$

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

$$m_i^k = \left(2 \frac{\varepsilon_{i-1}^k}{\alpha_{i-1}^k} m_{i-1}^k + \frac{u}{\alpha_i^{k-1}} m_i^{k-1} \right) \frac{\alpha_i^k}{2\alpha_i^k - 1}, \quad (2.11)$$

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

$$m_i^k = \frac{\alpha_i^k}{2\alpha_i^k - 1} \sum_{l=1}^i 2^{i-l} \frac{u}{\alpha_l^{k-1}} m_l^{k-1} \prod_{h=l}^{i-1} \frac{\varepsilon_h^k}{2\alpha_h^k - 1}. \quad (2.12)$$

As m_i^0 is given by (2.10), one can construct the explicit solution iteratively. Equation (2.12) allows for arbitrary parameters α_i^k and thus incorporates any mutation-induced change in the cell proliferation parameters. However, if $\alpha_i^k \leq 0.5$ the sum diverges. Cells with at least k mutations will accumulate in all compartments downstream of i . Note that equation (2.12) captures the general deterministic dynamics of a cell lineage founded by a single cell somewhere in the hierarchy. We are especially interested in the case where this founder cell carries critical, potentially cancer-driving mutations that will

allow us to address the expected number of mutations arising in this mutant clone. In addition, the probability of obtaining such a critical mutational hit can be investigated. But even if a cell accumulated a critical number of mutations, it might still become extinct because of stochastic effects [19,37,39]. In the following, we discuss the general solution of equation (2.12) for mutations that are neutral relative to the founder cell.

2.3. Reproductive capacity of neutral mutants

We call a mutation neutral if the reproductive capacity of the mutant and the founder cell is equal. In §2.2, we have shown that the reproductive capacity of a cell depends on its differentiation probability ε and its mutation rate u , but interestingly it is independent of the reproduction rate r . Therefore, the clonal lineage and the number and type of mutations that arise from a single founder cell do not depend on the proliferation rates of the founder cell. However, the time to reach those states of course depends on r . Therefore, although two mutations lead to the same outcome, this might occur on distinct time scales, with observable differences in the progression of diseases. Nonetheless, our definition of neutral mutations only requires constant differentiation probabilities and mutation rates relative to the founder cell. This assumption allows us to write $\varepsilon_i^k = \varepsilon_i$ and thus the number of parameters is reduced from $(k + 1) i + 1$ for the general case to $i + 1$ for the neutral case. This number can be reduced to two parameters, u and ε , if a constant differentiation probability for all non-stem cell stages is assumed, $\varepsilon_i = \varepsilon$. This simplifies the evaluation of the recurrence relation (2.12) significantly. The reproductive capacity m_i^k of neutral mutations in compartment i carrying k mutations becomes

$$\begin{aligned} m_i^k &= \alpha \frac{u^k}{k!} \frac{(2\varepsilon)^{i-1}}{(2\alpha - 1)^{i+k}} \prod_{l=1}^k (i + l - 1) \\ &= \alpha u^k \frac{(2\varepsilon)^{i-1}}{(2\alpha - 1)^{i+k}} \binom{i + k - 1}{k}. \end{aligned} \quad (2.13)$$

Mutants carrying k mutations are suppressed by a factor u^k and thus are rare in the early differentiation stages. The number increases exponentially for downstream compartments, and a significant load of cells carrying few mutations can be observed in the late differentiation stages (figure 4).

Equation (2.13) reveals interesting properties of hierarchical tissue structures. The ratio of cells carrying k mutations to cells carrying $k - 1$ mutations in compartment i is

$$\frac{m_i^k}{m_i^{k-1}} = \frac{u}{2\alpha - 1} \left(1 + \frac{i - 1}{k} \right). \quad (2.14)$$

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

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

$$\frac{m_{i+1}^k}{m_i^k} = \frac{2\varepsilon}{2\alpha - 1} \left(1 + \frac{k}{i} \right). \quad (2.15)$$

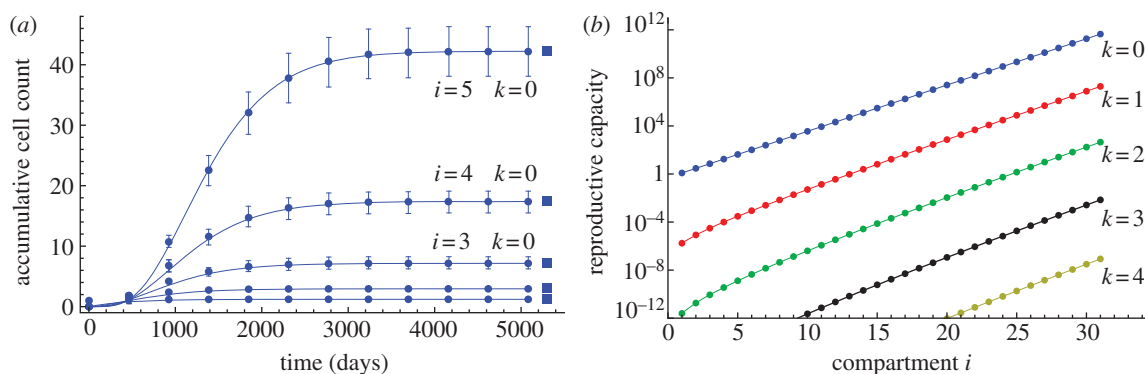


Figure 4. (a) Number of cells without mutations in compartments 1–5, arising from a single cell in compartment 1. Lines are equation (2.8), symbols are averages with corresponding standard deviations over 10^3 independent runs of stochastic individual-based computer simulations, and squares are equation (2.13). Parameters are $n_0 = 1$, $\varepsilon = 0.85$, $\gamma = 1.26$, $u = 10^{-6}$ and $r_0 = 1/400$. (b) Reproductive capacity of a single founder cell in compartment 1. Shown is the number of cells with 0–4 mutations in the first 31 compartments, acquired from a single cell in compartment 1. Symbols are numerical solutions of (2.8) in the limit of infinite time and lines are equation (2.13). The reproductive capacity increases exponentially for increasing compartment number. Cells carrying multiple mutations are strongly suppressed within a hierarchical tissue structure (see equation (2.14)). (Online version in colour.)

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

2.4. Number of distinct neutral mutations

So far, we have discussed the reproductive capacity of cells. However, we did not distinguish between different mutations, but grouped together cells with an equal number of mutations. Often, experimental studies focus on mutation landscapes, investigating the variation in clonal loads in healthy and sick individuals. Our approach allows us to estimate the expected number of distinct mutations that arise from a single founder cell, corresponding to the count of distinct symbols with the same colour in figure 2.

Let us assume that every mutation event is unique. Thus, we neglect the possibility that the same mutation is derived twice independently. The diversity in compartment i increases either by additional mutations of cells in compartment i or by differentiation of clones from compartment $i-1$ into compartment i . Assuming a differentiation probability $\varepsilon > 0$ the expected diversity n_i^{k+1} of cells with $k+1$ mutations in compartment i is

$$n_i^{k+1} = u \sum_{j=1}^i m_j^k = \alpha u^{k+1} \sum_{j=1}^i \frac{(2\varepsilon)^{j-1}}{(2\alpha-1)^{j+k}} \binom{j+k-1}{k}. \quad (2.16)$$

As an example, if we use the parametrization of the haematopoietic system, $\varepsilon = 0.85$ and $u = 10^{-6}$, we find approximately $n_{31}^1 \approx 30$ distinct single mutations in compartment 31 that were derived from the clonal progeny of a single founder cell in compartment 1. But the founder cell by chance could carry a mutation that changes their cell proliferation parameters. For example, the differentiation probability of the founder cell could change to $\varepsilon = 0.75$. In this case, equation (2.16) estimates 28 000 distinct mutations derived from this single founder cell. We note that the above change in the differentiation probability from 0.85 to 0.75 is sufficient to explain the manifestation of chronic myeloid leukaemia in otherwise healthy adults [38,52]. However, equation (2.16) represents an average, and in individual cases fluctuations, caused by stochasticity, are expected. But also small changes in ε or in u influence the expected diversity significantly. A linear

error analysis [53] reveals the dependency of (2.16) on the uncertainties of u and ε , which is given by

$$\Delta n_i^{k+1} = \sum_{j=1}^i \frac{(2\varepsilon)^{j-1}}{(2\alpha-1)^{j+k+1}} (c_j^\varepsilon + d_j^u) \binom{j+k-1}{k}, \quad (2.17)$$

with $c_j^\varepsilon = u^{k+1} |2\alpha(1-j-k) - 1| \Delta\varepsilon$ and $d_j^u = u^k |(k+1)\varepsilon(2\varepsilon-1) + u(2(k-j+3)\varepsilon - k - 2) + 2u^2(2-j)| \Delta u$. If we assume that $\Delta u = 10^{-7}$ and $\Delta\varepsilon = 0.01$ the uncertainty given by (2.17) becomes $\Delta n_{31}^1 \approx 35$, where the individual error contributions in u and ε are 7 and 28, respectively. Especially, note the strong dependency on $\Delta\varepsilon$. If we choose $\Delta\varepsilon = 0.05$, a deviation that might be difficult to detect *in vivo*, one gets $\Delta n_{31}^1 \approx 150$. Thus, small variations in ε lead to significant differences in the expected diversity of clonal populations, one aspect that might contribute to the explanation of the observation of very diverse mutation landscapes. Note also that an increasing mutation rate $\Delta u = 10^{-6}$ gives $\Delta n_{31}^1 \approx 65$. Of course, higher mutation rates increase the expected diversity of clonal populations. However, a higher mutation rate (or genomic instability) is neither the exclusive nor necessarily the dominant underlying cause of the diversity in the mutation landscape that is observed.

2.5. Example: clonal diversity in acute lymphoblastic leukaemia

Let us now consider a specific example using data from childhood acute lymphoblastic leukaemia (ALL). The most common chromosomal abnormality in this disease is the t(12;21) translocation that results in the fusion gene ETV6–RUNX1 (also known as TEL/AML1). There is evidence that this mutation often arises *in utero*. This has been confirmed to be the case in at least one pair of monozygotic twins [54]. This mutation is a founder mutation and is considered to be critical for the disease. Cells that express this fusion gene appear to have a higher self-renewal rate and enhanced survival when compared with normal cells [54,55]. In our model, enhanced self-renewal implies a reduced differentiation probability for the cells carrying the mutation ($\varepsilon < 0.85$). Recently, Ma *et al.* [56] performed whole genome sequencing on leukaemic cells isolated from two pairs of monozygotic twins. In one pair of twins, the initial event occurred *in utero* since the

ETV6–RUNX1 fusion was shared by both siblings. They found that the incidence of non-synonymous single nucleotide mutations between the samples ranged from 708 to 1237 [56]. These mutations must have occurred after the founder mutation and independent of each other. The time from the putative appearance of the shared ETV6–RUNX1 mutation and disease was 48–55 months. The second pair of monozygotic twins shared a mutation in NF-1, and the time to diagnosis of ALL was 72–77 months after birth. The tumours in these two children had 949–975 unique non-synonymous single nucleotide mutations. Using these constraints and an estimate of 10–75% cancer cells at diagnosis, unchanged proliferation rates as well as $N_0 = 100$ [57], we use equation (2.4) to estimate the differentiation probability ε for the mutant cells, which is then in the range of 0.78–0.81, i.e. only slightly lower than that of normal cells. Based on equation (2.4), it takes approximately 50–80 months to reach this load. We further predict a range of 350–2600 distinct mutations by using equation (2.16), assuming a mutation rate of $u = 10^{-6}$ and the above range of differentiation probabilities. Thus, we expect slightly more distinct mutations than found in patients. But some of those theoretically predicted clonal populations are small and may escape detection. For example, if we neglect mutations that occur during the last differentiation step of cells in compartments 30–31 (expected to be small cell populations), we predict only 150–950 distinct mutations. Clearly, the model as presented can explain the large number of passenger mutations that can be expected in a typical patient with ALL and probably other types of leukaemia.

3. Discussion

The accumulation of multiple mutations in cells is considered to be critical for cancer initiation. However, mutations unavoidably accompany cell proliferation and thus cancer can potentially occur in any multicellular organism. Hierarchical tissue structures contribute to the protection against such mutations. So far, the suppression of single mutations in hierarchical tissue structures has been the focus. Here, we have shown that, in addition, the risk of the accumulation of multiple mutations is dramatically reduced by a hierarchical tissue organization. The cell population is divided into a few slowly proliferating stem cells and many faster proliferating progenitor cells. Stem cells have an almost infinite cell reproductive capacity, but the manifestation of a critical mutational load often exceeds an organism's expected natural lifetime. Progenitor cells proliferate faster, but their reproductive capacity usually is limited, and thus they give rise to clonal waves travelling through the hierarchy. Still, both cases can be observed. There are cancers that presumably originate from stem cell mutations, for example chronic myeloid leukaemia, and there are cancers that originate in the later stages of haematopoiesis, for example acute promyelocytic leukaemia and various other subtypes of acute myeloid leukaemia [17]. Understanding the clonal dynamics for both cases is of importance. Here, we have focused on the accumulation of non-stem cell-driven mutations in hierarchically organized tissues. We arrived at closed solutions for the clonal waves travelling through the hierarchy. From this, the reproductive capacity of cells can be deduced. This allows us to predict the expected risk of acquiring any number of mutations from one single cell, given the proliferation properties of this cell. We derived equations that allow the quantitative classification of multiple mutations in

hierarchically organized structures, highlighting the strong suppression of clones carrying multiple mutations by this architecture. Although we neglected clonal competition over limited resources, such as cytokines or nutrients, we expect the model to capture general patterns of clonal expansions in hierarchically organized tissues and to serve as a null model. Moreover, those cancer cells that divide independently of proliferation signals (for example, owing to mutated tyrosine kinases) escape such a competition.

Another important question emerged more recently with the accessibility of whole genome sequencing technology. These techniques revealed very complex and diverse mutation landscapes for many different cancers, with the classification of driver and passenger mutations as the final goal. Knowing the exact driver mutations might help our understanding of the properties of specific cancer cells, allowing the development of effective treatment strategies. A promising example is the design of molecularly targeted agents such as the various tyrosine kinase inhibitors (imatinib, nilotinib, etc.) to treat patients with chronic myeloid leukaemia. These molecules specifically bind to kinase domains encoded by the BCR-ABL oncogene and strongly suppress the proliferative capacity of these cells [58–60]. Our work can contribute to this question by predicting the average number of distinct neutral (passenger) mutations acquired from a single cell at any stage of the hierarchy. This approach directly corresponds to the method of a neutral marker. The genetically marked cell represents the founder cell of the clonal population and one can follow the offspring of this founder cell throughout the hierarchy. This enables the prediction of the size and the variability of the clonal population. For example, in normal haematopoiesis, we expect cells to have a differentiation probability of $\varepsilon = 0.85$, leading to approximately 30 distinct single mutations (sub-clones) in adult cells acquired from this single cell. If this founder cell acquired a mutation that changed the differentiation probability to $\varepsilon = 0.75$ by chance, the expected number of distinct single mutations increases to approximately 28 000.

We have shown that even a slight change in the self-renewal probability of progenitor cells can lead to substantial differences in the number of passenger mutations observed in ALL. This probably holds true for other malignancies. In acute promyelocytic leukaemia, apart from the t(15;17) that is a critical event in the origin of this disease (akin to the ETV6–RUNX1 discussed earlier for ALL), approximately 440 non-synonymous single nucleotide mutations were found, which were unique to the tumour clone [21]. Interestingly, it is highly likely that the cell of origin of acute promyelocytic leukaemia is downstream (a progenitor cell) of the cell of origin of ETV6–RUNX1-driven ALL and this may, in part, explain the less diverse mutational landscape reported in acute promyelocytic leukaemia compared with ALL and would fit well with our model. Of course, any genomic instability will further increase the repertoire of passenger mutations that is observed in any given tumour.

We also note that the effect of a specific mutation on a cell needs not be large for the effect to spread throughout the tumour. Tissue architecture and dynamics, such as in haematopoiesis, serve as a deterrent against the accumulation of mutations, in particular multiple mutations occurring in one cell. Once a driver mutation appears, if this changes either the self-renewal of the cell or the mutation rate, then the appearance of many passenger mutations becomes inevitable in such an architecture because of the amplification of cells that occurs. Thus, a minor change in the differentiation probabilities that

might be difficult to detect *in vivo* drastically changes the expected number of passenger mutations that would be observed. Concomitantly, this will increase the risk of acquiring mutations in 'driver genes' and so lead to malignancy. However, if the initiation of disease requires several co-occurring

mutations, a hierarchical tissue structure is a powerful mechanism of tumour suppression.

B.W. and A.T. thank the Max Planck Society and the Emmy-Noether programme of the German Research Foundation for generous funding.

References

- Hayflick L, Moorhead PS. 1961 The serial cultivation of human diploid cell strains. *Exp. Cell Res.* **25**, 585–621. (doi:10.1016/0014-4827(61)90192-6)
- Raff MC. 1992 Social controls on cell survival and cell death. *Nature* **356**, 397–400. (doi:10.1038/356397a0)
- Pardee AB. 1989 G1 events and regulation of cell proliferation. *Science* **246**, 603–608. (doi:10.1126/science.2683075)
- McCulloch JE, Till JE. 2005 Perspectives on the properties of stem cells. *Nat. Med.* **11**, 1026–1028. (doi:10.1038/nm1005-1026)
- Michor F, Nowak MA, Frank SA, Iwasa Y. 2003 Stochastic elimination of cancer cells. *Proc. R. Soc. Lond. B* **270**, 2017–2024. (doi:10.1098/rspb.2003.2483)
- Dingli D, Traulsen A, Pacheco JM. 2007 Compartmental architecture and dynamics of hematopoiesis. *PLoS ONE* **4**, e345. (doi:10.1371/journal.pone.0000345)
- Nowak MA. 2006 *Evolutionary dynamics*. Cambridge, MA: Harvard University Press.
- Michor F, Hughes TP, Iwasa Y, Branford S, Shah NP, Sawyers CL, Nowak MA. 2005 Dynamics of chronic myeloid leukaemia. *Nature* **435**, 1267–1270. (doi:10.1038/nature03669)
- Wodarz D, Komarova N. 2005 *Computational biology of cancer*. Lecture Notes and Mathematical Modeling. Singapore: World Scientific Publishing.
- Werner B, Dingli D, Lenaerts T, Pacheco JM, Traulsen A. 2011 Dynamics of mutant cells in hierarchical organized tissues. *PLoS Comput. Biol.* **7**, e1002290. (doi:10.1371/journal.pcbi.1002290)
- Loeffler M, Wichmann HE. 1980 A comprehensive mathematical model of stem cell proliferation which reproduces most of the published experimental results. *Cell Tissue Kinetics* **13**, 543–561. (doi:10.1111/j.1365-2184.1980.tb00494.x)
- Fuchs E. 2008 Skin stem cells: rising to the surface. *J. Cell Biol.* **180**, 273–284. (doi:10.1083/jcb.200708185)
- Tumbar T, Guasch G, Greco V, Blanpain C, Lowry WE, Rendl M, Fuchs E. 2004 Defining the epithelial stem cell niche in skin. *Science* **303**, 359–363. (doi:10.1126/science.1092436)
- Potten CS, Gandara R, Mahida YR, Loeffler M, Wright NA. 2009 The stem cells of small intestinal crypts: where are they? *Cell Prolif.* **42**, 731–750. (doi:10.1111/j.1365-2184.2009.00642.x)
- Araten DJ, Nafa K, Pakdeesuwan K, Luzzatto L. 1999 Clonal populations of hematopoietic cells with paroxysmal nocturnal hemoglobinuria genotype and phenotype are present in normal individuals. *Proc. Natl Acad. Sci. USA* **96**, 5209–5214. (doi:10.1073/pnas.96.9.5209)
- Daley GQ, Van Etten RA, Baltimore D. 1990 Induction of chronic myelogenous leukemia in mice by the P210bcr/abl gene of the Philadelphia chromosome. *Science* **247**, 824–830. (doi:10.1126/science.2406902)
- Guibal FC *et al.* 2009 Identification of a myeloid committed progenitor as the cancer-initiating cell in acute promyelocytic leukemia. *Blood* **114**, 5415–5425. (doi:10.1182/blood-2008-10-182071)
- Stratton MR, Campbell PJ, Futreal AP. 2009 The cancer genome. *Nature* **458**, 719–724. (doi:10.1038/nature07943)
- Beerenwinkel N, Antal T, Dingli D, Traulsen A, Kinzler KW, Velculescu VEE, Vogelstein B, Nowak MA. 2007 Genetic progression and the waiting time to cancer. *PLoS Comput. Biol.* **3**, e225. (doi:10.1371/journal.pcbi.0030225)
- Gerstung M, Beerenwinkel N. 2010 Waiting time models of cancer progression. *Math. Popul. Stud.* **17**, 115–135. (doi:10.1080/08898480.2010.490994)
- Welch JS *et al.* 2012 The origin and evolution of mutations in acute myeloid leukemia. *Cell* **150**, 264–278. (doi:10.1016/j.cell.2012.06.023)
- Hanahan D, Weinberg RA. 2000 The hallmarks of cancer. *Cell* **100**, 57–70. (doi:10.1016/S0092-8674(00)81683-9)
- Vogelstein B, Kinzler KW. 2004 Cancer genes and the pathways they control. *Nat. Med.* **10**, 789–799. (doi:10.1038/nm1087)
- Sjöblom T *et al.* 2006 The consensus coding sequences of human breast and colorectal cancers. *Science* **314**, 268–274. (doi:10.1126/science.1133427)
- Jones S *et al.* 2008 Comparative lesion sequencing provides insights into tumor evolution. *Proc. Natl Acad. Sci. USA* **105**, 4283–4288. (doi:10.1073/pnas.0712345105)
- Pleasant ED *et al.* 2010 A comprehensive catalogue of somatic mutations from a human cancer genome. *Nature* **463**, 191–196. (doi:10.1038/nature08658)
- Shah SP *et al.* 2012 The clonal and mutational evolution spectrum of primary triple-negative breast cancers. *Nature* **486**, 395–399. (doi:10.1038/nature10933)
- Ding L *et al.* 2012 Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. *Nature* **481**, 506–510. (doi:10.1038/nature10738)
- Walter MJ *et al.* 2012 Clonal architecture of secondary acute myeloid leukemia. *N. Engl. J. Med.* **366**, 1090–1098. (doi:10.1056/NEJMoa1106968)
- Haber DA, Settleman J. 2007 Cancer: drivers and passengers. *Nature* **446**, 145–146. (doi:10.1038/446145a)
- Bozic I *et al.* 2010 Accumulation of driver and passenger mutations during tumor progression. *Proc. Natl Acad. Sci. USA* **107**, 18 545–18 550. (doi:10.1073/pnas.1010978107)
- Greenman C *et al.* 2007 Patterns of somatic mutation in human cancer genomes. *Nature* **446**, 153–158. (doi:10.1038/nature05610)
- Rodriguez-Brenes IA, Komarova NL, Wodarz D. 2011 Evolutionary dynamics of feedback escape and the development of stem-cell-driven cancers. *Proc. Natl Acad. Sci. USA* **108**, 18 983–18 988. (doi:10.1073/pnas.1107621108)
- Roeder I, Horn M, Glauche I, Hochhaus A, Mueller MC, Loeffler M. 2006 Dynamic modeling of imatinib-treated chronic myeloid leukemia: functional insights and clinical implications. *Nat. Med.* **12**, 1181–1184. (doi:10.1038/nm1487)
- Marciniak-Czochra A, Stiehl T, Ho AD, Jäger W, Wagner W. 2009 Modeling of asymmetric cell division in hematopoietic stem cells—regulation of self-renewal is essential for efficient repopulation. *Stem Cells Dev.* **18**, 377–385. (doi:10.1089/scd.2008.0143)
- Glauche I, Thielecke L, Roeder I. 2011 Cellular aging leads to functional heterogeneity of hematopoietic stem cells: a modeling perspective. *Aging Cell* **10**, 457–465. (doi:10.1111/j.1474-9726.2011.00692.x)
- Lenaerts T, Pacheco JM, Traulsen A, Dingli D. 2010 Tyrosine kinase inhibitor therapy can cure chronic myeloid leukemia without hitting leukemic stem cells. *Haematologica* **95**, 900–907. (doi:10.3324/haematol.2009.015271)
- Lenaerts T, Castagnetti F, Traulsen A, Pacheco JM, Rosti G, Dingli D. 2011 Explaining the *in vitro* and *in vivo* differences in leukemia therapy. *Cell Cycle* **10**, 1540–1544. (doi:10.4161/cc.10.10.15518)
- Traulsen A, Lenaerts T, Pacheco JM, Dingli D. 2013 On the dynamics of neutral mutations in a mathematical model for a homogeneous stem cell population. *J. R. Soc. Interface* **10**, 20120810. (doi:10.1098/rsif.2012.0810)
- Komarova NL, Wang L. 2004 Initiation of colorectal cancer: where do the two hits hit? *Cell Cycle* **3**, 1558–1565. (doi:10.4161/cc.3.12.1186)
- Basanta D, Scott JG, Rockne R, Swanson KR, Anderson AR. 2011 The role of IDH1 mutated tumour cells in secondary glioblastomas: an evolutionary game theoretical view. *Phys. Biol.* **8**, 015016. (doi:10.1088/1478-3975/8/1/015016)

42. Gerlee P, Basanta D, Anderson AR. 2011 Evolving homeostatic tissue using genetic algorithms. *Prog. Biophys. Mol. Biol.* **106**, 414–425. (doi:10.1016/j.pbiomolbio.2011.03.004)
43. Gillespie DT. 1976 A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J. Comput. Phys.* **22**, 403–434. (doi:10.1016/0021-9991(76)90041-3)
44. Gillespie DT. 2007 Stochastic simulation of chemical kinetics. *Annu. Rev. Phys. Chem.* **58**, 35–55. (doi:10.1146/annurev.physchem.58.032806.104637)
45. Dingli D, Pacheco JM. 2006 Allometric scaling of the active hematopoietic stem cell pool across mammals. *PLoS ONE* **1**, e2. (doi:10.1371/journal.pone.0000002)
46. Buescher ES, Alling DW, Gallin JI. 1985 Use of an X-linked human neutrophil marker to estimate timing of lyonization and size of the dividing stem cell pool. *J. Clin. Invest.* **76**, 1581–1584. (doi:10.1172/JCI112140)
47. Rufer N, Brümmendorf THH, Kolvraa S, Bischoff C, Christensen K, Wadsworth L, Schulzer M, Lansdorp PM. 1999 Telomere fluorescence measurements in granulocytes and T lymphocyte subsets point to a high turnover of hematopoietic stem cells and memory T cells in early childhood. *J. Exp. Med.* **190**, 157–167. (doi:10.1084/jem.190.2.157)
48. Dingli D, Traulsen A, Pacheco JM. 2007 Stochastic dynamics of hematopoietic tumor stem cells. *Cell Cycle* **6**, 461–466. (doi:10.4161/cc.6.4.3853)
49. Gerrits A, Dykstra B, Kalmykova OJ, Klauke K, Verovskaya E, Broekhuis MJ, de Haan G, Bystriykh LV. 2010 Cellular barcoding tool for clonal analysis in the hematopoietic system. *Blood* **115**, 2610–2618. (doi:10.1182/blood-2009-06-229757)
50. Raymond P, Warrell J, Hugues de T, Wang ZY, Degos L. 1993 Acute promyelocytic leukemia. *N. Engl. J. Med.* **329**, 177–189. (doi:10.1056/NEJM199307153290307)
51. Armanios M, Blackburn EH. 2012 The telomere syndromes. *Nat. Rev. Genet.* **13**, 693–704. (doi:10.1038/nrg3246)
52. Dingli D, Traulsen A, Pacheco JM. 2008 Chronic myeloid leukemia: origin, development, response to therapy, and relapse. *Clin. Leuk.* **2**, 133–139. (doi:10.3816/CLK.2008.n.017)
53. Taylor JR. 1997 *An introduction to error analysis: the study of uncertainties in physical measurements*, 2nd edn. Mill Valley, CA: University Science Books.
54. Hong D *et al.* 2008 Initiating and cancer-propagating cells in TEL-AML1-associated childhood leukemia. *Science* **319**, 336–339. (doi:10.1126/science.1150648)
55. Fischer M *et al.* 2005 Defining the oncogenic function of the TEL/AML1 (ETV6/RUNX1) fusion protein in a mouse model. *Oncogene* **24**, 7579–7591. (doi:10.1038/sj.onc.1208931)
56. Ma Y *et al.* 2013 Developmental timing of mutations revealed by whole-genome sequencing of twins with acute lymphoblastic leukemia. *Proc. Natl Acad. Sci. USA* **110**, 7429–7433. (doi:10.1073/pnas.1221099110)
57. Dingli D, Pacheco JM. 2007 Ontogenic growth of the haematopoietic stem cell pool in humans. *Proc. R. Soc. B* **274**, 2497–2501. (doi:10.1098/rspb.2007.0780)
58. Druker BJ *et al.* 2006 Five-year follow-up of patients receiving imatinib for chronic myeloid leukemia. *N. Engl. J. Med.* **355**, 2408–2417. (doi:10.1056/NEJMoa062867)
59. Saglio G *et al.* 2010 Nilotinib versus imatinib for newly diagnosed chronic myeloid leukemia. *N. Engl. J. Med.* **362**, 2251–2259. (doi:10.1056/NEJMoa0912614)
60. Werner B, Lutz D, Brümmendorf TH, Traulsen A, Balabanov S. 2011 Dynamics of resistance development to imatinib under increasing selection pressure: a combination of mathematical models and *in vitro* data. *PLoS ONE* **6**, e28955. (doi:10.1371/journal.pone.0028955)