<https://www.overleaf.com/project/5bc76c610c744c6a4ef95381>

Temporal order of mutations influences cancer initiation dynamics

Hamid Teimouri ^{1,2}, Anatoly B. Kolomeisky ^{1,2,3,4*}

1 Department of Chemistry, Rice University, Houston, Texas, United States

2 Center for Theoretical Biological Physics, Rice University, Houston, Texas, United States

3 Department of Chemical and Biomolecular Engineering, Rice University, Houston, Texas, United States

4 Department of Physics and Astronomy, Rice University, Houston, Texas, USA

* tolya@rice.edu

Abstract

Cancer is a set of genetic diseases that are driven by mutations. It was recently discovered that the temporal order of genetic mutations affects the cancer evolution and even the nature of the disease itself. The mechanistic origin of these observations, however, remain not well understood. Here we present a theoretical model for cancer initiation dynamics that allows us to quantify the impact of the temporal order of mutations. In our approach, the cancer initiation process is viewed as a set of stochastic transitions between discrete states defined by the different numbers of mutated cells. Using a first-passage analysis, probabilities and times before the cancer initiation are explicitly evaluated for two alternative sequences of two mutations. It is found that the probability of cancer initiation is determined only by the first mutation, while the dynamics is specified by both mutations. In addition, it is shown that the acquisition of a mutation with higher fitness before mutation with lower fitness increases the probability of the tumor formation but delays the cancer initiation. Theoretical results are explained using effective free-energy landscapes.

Introduction

It is widely accepted that cancer tumors are caused by several genetic alterations, such as point mutations, unsuccessful DNA repairs and DNA methylations, in normal tissue cells [6, 7, 24, 25]. These changes lead to abnormal growth of those cells that can no longer support the required functions of the tissue. The mechanistic origin of the increased competitiveness of cancer cells have been the focus of significant research efforts that identified the key features of tumors [6, 11, 24, 25]. Traditionally, the cancer phenotype has been considered as a sum of complementary properties that each mutation drives separately [6]. These views, however, have been recently challenged by experimental observations suggesting that the order of mutations is also an important factor in cancer development [9, 13, 22]. But the microscopic picture behind the effect of the order of mutations in cancer remains largely unexplored.

There are now several examples of how different order of mutations lead to dramatically different outcomes in the cancer formation and development [3,8–10,14,18,22]. In the mouse model of adrenocortical tumors, it was observed that when the mutation in the oncogene Ras preceded the mutation in p53 gene this led to highly malignant tumors with strong metastatic properties. But the reverse order of mutations (p53 first and then Ras) produced only benign tumors [8]. Another striking example comes from the studies on chronic myeloproliferative neoplasms (MPNs), which are myeloid tumors that contain on average between 5 to 10 somatic mutations [22]. Two driver mutations, *TET2* and *JAK2*, were identified as the most common in these cancers. Blood cells analysis showed that patients with *TET2*-first mutations were a decade older than the patients with *JAK2*-first mutations. Moreover, it was observed that the order of mutation acquisition leads to different cancer types and probabilities of survival [22]. It was also shown in colorectal cancers using genome-sequencing studies that the type of the first mutation and the order in which mutations occur strongly affect the nature of diseases [13].

These observations stimulated discussions on how to explain why the mutation order alters cancer initiation and evolution [9,13]. Three possible mechanisms (not mutually exclusive) have been proposed [9]. One of them suggests that the first mutation might change the accessibility of specific genomic regions and this prevents the second mutation from activating or repressing this region. If the order of mutations is reversed the blocking does not happen. The second idea is that initially one mutation can induce rapid cell growth and differentiation, while starting from another mutation lead to much slower growth and less differentiation. The third mechanism argues that different starting mutations might lead to different cellular microenvironment, influencing this way the distinct disease evolution. These mechanisms, however, are qualitative, but the increasing amount of experimental data and the need to have new medical strategies to fight against the cancer require more quantitative understanding.

Theoretical models play an important role in uncovering the mechanisms of cancer initiation and progression [4,19,25]. They provide a critical quantitative link between the underlying microscopic processes and the appearance of malignant tumors. Due to these theoretical studies and a large number of other important experimental studies, we currently know that the cancer starts when cells accumulate certain number and types of genetic alterations. Significant efforts have been made for modeling the dynamics of mutation acquisition and how it is governed by relevant genetic parameters such as the rate of mutations, the size of the population of cells and the rate of mutations proliferation [1,5,12,16,19,20]. Recently we developed a new theoretical framework to evaluate the cancer initiation dynamics [23]. It is based on a discrete-state stochastic description of cancer initiation as a fixation of driver mutations in the tissues. It allows to explicitly evaluate the probabilities for the cancer to appear and the times before it happens, which are viewed as the fixation probabilities and the fixation times, respectively. This framework was applied for estimating the cancer initiation times from experimentally available data for different types of cancer.

In this paper, we present a theoretical model of cancer initiation that allows us to quantitatively explain the effect of different order of mutations. In our theoretical approach, the cancer appearance is viewed as a fixation of two different mutations. This so-called two-hit model assumes that the first type of mutations occupies the whole tissue before the second type of mutations starts and eventually fills the system. Using the method of first-passage processes, we explicitly evaluate the probability and the dynamics of cancer initiation. It is found that the order of mutations strongly influences both fixation probabilities and fixation times. We show that if the first mutation has a higher fitness than the second mutation this increases the overall probability of the cancer initiation, but, surprisingly, delays the formation of the tumor. These

observations are quantified using effective free-energy landscapes for the underlying stochastic processes of cancer initiation.

Figure 1. (a) A hypothetical scenario for the tissue to sequentially acquire two mutations, labeled A and B . The final outcome of two mutations is a cancer, but it might be different diseases depending on AB or BA sequences of mutations. (b) A schematic view of a two-hit mutation fixation process in the tissue compartment with N cells. Normal stem cells are green, while cells harboring single and double mutations are shown in yellow and red, respectively.

1 Model

In our theoretical approach, the cancer initiation is viewed as an event that starts when two different mutations take over the healthy tissue as illustrated in Fig. 1. More specifically, we assume that originally in the well-mixed tissue compartment there are N normal stem cells and the total number of cells is always fixed. At some time, which we view as $t = 0$, one stem cell gets a first mutation with a probability u_1 . It is realistic to assume that the mutation rate is very low ($Nu_1 \ll 1$) [15, 19], and the further transformations in the system are taking place only via stem cells divisions and removals to keep the total number of cells fixed. Normal cells divide with a speed b , while the cells with the mutation divide with a rate $r_1 b$. The parameter r_1 is known as a fitness parameter, and it is equal to the ratio of the division rates for the mutated cell over the normal cell. It reflects the overall physiological influence of the mutation on cellular metabolism: if $r_1 < 1$ the mutation is disadvantageous, $r_1 = 1$ corresponds to a neutral effect, while for $r_1 > 1$ the mutation is advantageous. The overall number of mutated cells in the fixed population of stem cells will be fluctuating, and there is a time when all cells would become mutated. This is known as a mutation fixation, and the system cannot have normal cells after that. At this moment, we assume that the second mutation will appear with a probability u_2 , which is also very small ($Nu_2 \ll 1$).

The cells with one mutation divide with the rate $r_1 b$, while the cells with two mutations divide with the rate $r_2 b$. The fitness parameter r_2 reflects the accumulative effects of both mutations present in the cell.

Introducing two mutations in the tissue compartment generates three types of cells (see Fig. 1b). The normal wild-type cells are labeled as type 0 (green circles), the cells with one mutation are labeled as type 1 (yellow circles), and the cells harboring two mutations are labeled as type 2 (red circles). At any time, the tissue might only have the cells of type 0 and 1, or only the cells of type 1 and 2 (Fig. 1b). The dynamic changes in the system can be viewed as stochastic transitions between $2N$ discrete states, as shown in Fig. 2, and these states are specified by the number of cells with one or two mutations. We define a state n ($1 \leq n \leq N$) as the state that has n cells with one mutation and $N - n$ wild-type cells without mutations. If $N < n \leq 2N$ then the state n describes a situation with $n - N$ cells with two mutations and $2N - n$ cells with one mutations: see Fig. 2. The state $n = N$ corresponds to the fixation of the first mutation (all cells have one mutation), while the state $n = 2N$ corresponds to the fixation of both mutations (all cells have two mutations). It can be shown that for the states on the first branch ($1 \leq n \leq N$), the forward transition rate from the state n to the state $n + 1$ is given by $r_1 a_n$ and the backward transition rate from the state n to the state $n - 1$ is equal to a_n where, assuming that divisions and removals follow the Moran process [17], one obtains [23]

$$a_n = b \frac{n(N - n)}{N + 1}. \quad (1)$$

For the states on the second branch ($N < n \leq 2N$) it can be shown that the forward transition rate from the state n to the state $n + 1$ is given by $r_2 a_{n-N}$ and the reverse transition from the state n to the state $n - 1$ is given by $r_1 a_{n-N}$: see Fig. 2. The transition rate from the state N to the state $N + 1$, which is given by the appearance of the second mutation, is equal to Nu_2 . Transitions between all neighboring states are reversible except for two situations. When the system goes from the state $N - 1$ to N , this corresponds to the elimination of all wild-type cells and it cannot be reversed. Similarly, the transition from the state $2N - 1$ to $2N$ corresponds to the elimination of all cells with only one mutation. In addition, the single mutated cell can be eliminated from the system with the rate a_1 from the state 1 (Fig. 2).

Figure 2. A discrete-state stochastic model for two sequential mutations in the tissue compartment. The state n corresponds to n cells with one mutations for $1 \leq n \leq N$, and $n - N$ cells with two mutations for $N < n \leq 2N$.

The dynamics of cancer initiation with two sequential mutations can now be explicitly analyzed using the method of first-passage processes [23]. This is because in our discrete-state stochastic model (Fig. 2) the cancer initiation corresponds to the events that start in the state 1 and reach the state $2N$ for the first time. We define a function $F_n(t)$ as the corresponding first-passage probability density function to start from any site n at $t = 0$ and to reach the final fixation state $2N$ at time t . The

Figure 3. a) The ratio of fixation probabilities Π_{AB}/Π_{BA} for two alternative sequences of mutations. b) The ratio of fixation times T_{AB}/T_{BA} for two alternative sequences of mutations. For calculations, $N = 1000$ is used.

temporal evolution of these functions are governed by backward master equations,

$$\frac{dF_n(t)}{dt} = r_1 a_n F_{n+1}(t) + a_n F_{n-1}(t) - a_n(1 + r_1)F_n(t), \quad (2)$$

for $1 \leq n < N$, and

$$\begin{aligned} \frac{dF_n(t)}{dt} = & r_2 a_{n-N} F_{n+1}(t) + r_1 a_{n-N} F_{n-1}(t) \\ & - a_{n-N}(r_2 + r_1)F_n(t), \end{aligned} \quad (3)$$

for $N < n < 2N$, while

$$\frac{dF_N(t)}{dt} = Nu_2 F_{N+1}(t) - Nu_2 F_N(t) \quad (4)$$

for $n = N$. In addition we have the boundary condition, $F_{2N}(t) = \delta(t)$, which physically means that if the system starts in this state the cancer initiation process is immediately accomplished.

These equations can be solved explicitly (see the SI), and this allows us to obtain a comprehensive description of the cancer initiation dynamics. One can define a

probability to reach the fixation state starting from the state n , $\Pi_n = \int_0^\infty F_n(t)dt$, and the calculations presented in the SI show that

$$\Pi_n = \frac{1 - 1/r_1^n}{1 - 1/r_1^N}, \quad (5)$$

for $1 \leq n < N$, while $\Pi_n = 1$ for $N \leq n < 2N$. This result can be easily understood by considering the properties of the discrete-state stochastic model in Fig. 2. If the system is found in one of the states $1 \leq n < N$ then there is always a non-zero chance that the mutated cells will be eliminated from the system (exiting eventually to the left from the state 1 with the rate a_1). However, because the transition from the state $N - 1$ to the state N is irreversible, for the states $N \leq n < 2N$ the system can never eliminate the mutated cells and with a probability one it will reach the fixation state. But there is another surprising observation from Eq. (5). It suggests that the probability of cancer initiation by two mutations (starting from the state $n = 1$) depends only on the properties of the first mutation. This is again the consequence of the irreversibility in the transition from the first branch and transient fixation of the first mutation before the system can proceed further to the final fixation of both mutations. Since the probability of cancer initiation is directly related to the cancer lifetime risk [19, 21], this clearly indicates that the order of mutations does matter. The results for the fixation probability also simplify in several limiting cases. When $r_1 \rightarrow 1$, we obtain $\Pi_n = n/N$, while for the tissues with very large number of cells and advantageous first mutation ($N \gg 1$ and $r_1 > 1$) we have $\Pi_n = 1 - 1/r_1^n$.

Another important characteristics of the cancer initiation process is the cancer initiation time. It is defined as a time for the tissue to reach the two-mutations fixation for the first time. In our theoretical framework, it also corresponds to the mean first-passage time to reach the state $2N$ starting from the state n . It is generally defined as $T_n = \frac{\int_0^\infty t F_n(t) dt}{\Pi_n}$. As shown explicitly in the SI, the cancer initiation time starting from the state $n = 1$ can be written as

$$T_1 = \frac{N+1}{b} \sum_{n=1}^{N-1} \frac{1}{n(N-n)} \left(\frac{r_1^n - 1}{r_1 - 1} \right) \left(\frac{r_1^{N-n} - 1}{r_1^N - 1} \right) + \frac{1}{Nu_2} \left[\frac{1 - (r_1/r_2)^N}{1 - r_1/r_2} \right] + \frac{N+1}{br_2} \sum_{n=1}^{N-1} \frac{1}{n(N-n)} \left(\frac{1 - (r_1/r_2)^{N-n}}{1 - r_1/r_2} \right) \quad (6)$$

There are three contributions to the average fixation time of two mutations in Eq. (6), i.e., $T_1 = T_{11} + T_{trans} + T_{12}$. The first term T_{11} describes the time for the system to reach the state N , which corresponds to the fixation of the first mutation. The second term T_{trans} describes the effective rate of acquiring the second mutation in cells that are fully fixed by the first mutation. It can be shown that

$$T_{trans} = \frac{1}{Nu_2} \frac{1}{\Pi_{12}}, \quad \Pi_{12} = \frac{1 - r_1/r_2}{1 - (r_1/r_2)^N}, \quad (7)$$

where Π_{12} is the fixation probability of the second mutation starting from the state $N + 1$. Thus, this contribution to the overall fixation time reflects the possibility of reverse transitions from the state $N + 1$ back to the state N . The third term, T_{12} describes the time to reach the final fixation starting from the state $N + 1$.

Again, it is interesting to consider limiting cases. For $r_1 = r_2 = 1$, i.e., for two successive neutral mutations and $N \rightarrow \infty$, we obtain,

$$T_1 \simeq \frac{N}{b} + \frac{1}{bu_2} + \frac{N}{b} \ln N \simeq \frac{N}{b} \left[1 + \frac{1}{Nu_2} + \ln N \right]. \quad (8)$$

In this limit, the system spends most of the time on the second branch of discrete states. Also, for large N Eq. (6) can be simplified into (see the SI):

$$T_1 \simeq \frac{2(1+r_1^N) \ln N + \text{Ei}(\ln r_1) + 2r_1^N \text{Ei}(-\ln r_1)}{b(r_1-1)(r_1^N-1)} + \frac{1}{Nu_2} \left[\frac{1-(r_1/r_2)^N}{1-r_1/r_2} \right] + \frac{2 \ln N + \left(\frac{r_1}{r_2}\right)^N \text{Ei}(-\ln \frac{r_1}{r_2}) + \text{Ei}(\ln \frac{r_1}{r_2})}{br_2(1-r_1/r_2)} \quad (9)$$

where $\text{Ei}(x)$ is the exponential integral defined as $\text{Ei}(x) \equiv -\int_{-x}^{\infty} \frac{e^{-z}}{z} dz$, and γ is the Euler Mascheroni constant. For $r_1 < r_2$, terms multiplied by $(r_1/r_2)^N$ vanish and consequently the fixation time is further simplified,

$$T_1 \simeq \frac{2(1+r_1^N) \ln N + \text{Ei}(\ln r_1) + 2r_1^N \text{Ei}(-\ln r_1)}{b(r_1-1)(r_1^N-1)} + \frac{1}{Nu_2} \left[\frac{1}{1-r_1/r_2} \right] + \frac{1}{br_2(1-r_1/r_2)} \left[2 \ln N + \text{Ei}(\ln \frac{r_1}{r_2}) \right]. \quad (10)$$

Now we can analyze the effect of the mutations order by considering two specific mutations A and B with the fitness parameters given by r_A and r_B , respectively. There are two different mutation sequences in this system, labeled as AB and BA . For the events of type AB , the mutation A is the first one, while the mutation B is the second one. Then we have $r_1 = r_A$ while $r_2 = \varepsilon r_A r_B$. Here, the combined fitness parameter r_2 reflects the presence of both mutations and the parameter ε describes the cooperativity between the mutations. When $\varepsilon > 1$ the presence of both mutations is more advantageous and this is a positive cooperativity, while for $\varepsilon < 1$ we have a negative cooperativity when the mutations counterbalance each other. To simplify our calculations, we assume $\varepsilon = 1$, which corresponds to a neutral cooperation, i.e., when the effect of both mutations is independent of each other. But the cooperativity effects can be easily taken into account in our theoretical approach. The situation is opposite for events of type BA . Here we have $r_1 = r_B$ and $r_2 = r_A r_B$. These two mutational sequences have different probabilities of cancer initiation defined as Π_{AB} and Π_{BA} , respectively. To quantify the difference we define a ratio,

$$\Delta_p = \frac{\Pi_{AB}}{\Pi_{BA}} = \left(\frac{1-1/r_A}{1-1/r_B} \right) \left(\frac{1-1/r_B^N}{1-1/r_A^N} \right) \quad (11)$$

Fig. 3a shows the difference in the cancer initiation probabilities for different mutations sequences as a function of the relative fitness parameters. One can see that, when the first mutation is more advantageous ($r_A/r_B > 1$), the AB fixation probability is higher than the BA fixation probability. The effect is stronger when the fitness parameters are closer to being equal to one. This observation can be easily explained by noting again that the overall fixation probability depends only on the nature of the first mutation. Thus, if the first mutation has a larger fitness parameter for one sequence it will lead to a higher probability of the cancer initiation. Similarly, we can explicitly calculate the fixation times T_{AB} and T_{BA} for AB and BA mutation sequences, respectively. To quantify the difference, the following function is defined,

$$\Delta_T = \frac{T_{AB}}{T_{BA}}, \quad (12)$$

and the results are presented in Fig. 3b. It shows that there is a different cancer initiation dynamics depending on the mutations order, and the effect is stronger when fitness parameters are close to unity. However, the unexpected result is that if the first mutation is more advantageous ($r_A/r_B > 1$) it takes longer to initiate the cancer. This contrasts with the higher probability of the overall fixation for this situation, i.e., something that is more probable takes longer to achieve, which is opposite to naive expectations.

To understand the microscopic origin of the surprising observations of anti-correlations between the probability and the time before the cancer initiation, we plot in Fig. 4 the ratio of the first and third terms in the fixation times (T_{11}/T_{12}) for different fitness parameters. These times correspond to the system being found on the first branch when only the wild-type cells and the cells with one mutation are present (T_{11} , $1 \leq n < N$); and for being on the second branch when only the cells with one or two mutations are present (T_{12} , $N + 1 \leq n < 2N$). For convenience, let us consider the AB mutation sequence. When the first mutation has a higher fitness parameter than the second mutation (Fig. 4, upper left corner), it is found that $T_{11}/T_{12} < 1$, i.e., the system spends most of the time on the second branch of discrete states. But if the second mutation is more advantageous (Fig. 4, lower right corner) the situation is reversed, and the system spends most of the time on the first branch of the discrete states. Based on these observations, we can construct an effective "free-energy" landscape for the cancer initiation dynamics driven by stochastic transitions. One can associate longer times with higher barriers and shorter times with smaller barriers. This leads to a schematic picture shown in Fig. 5 for alternative sequences of mutations. The first barrier corresponds to the states on the first branch, and the second barrier corresponds to the states on the second branch. The deep region between two barriers reflects the irreversible transition between two branches of the states. Now we can understand better the difference in cancer initiation for two sequences of mutations. The probability of cancer initiation is determined only by the first barrier, and then it is clear that if the first mutation is more advantageous ($r_A > r_B$, left panel on Fig. 5) the cancer lifetime risk is higher. But in the opposite case ($r_B > r_A$, right panel in Fig. 5), the cancer lifetime risk is lower. At the same, the cancer initiation dynamics depends on both barriers, although the highest barrier dominates the behavior. It is reasonable to assume that the highest barrier will lead to the slowest dynamics, and this is described by the situation on the left panel of Fig. 5 (AB sequence). This is because the second barrier (counting from the deep intermediate state) is higher. For the case on the right panel in Fig. 5 (sequence BA), while the first transition is relatively slow it is not as slow as the second transition for the AB sequence. These arguments explain the effect of temporal order of mutations in cancer initiation.

In conclusion, we developed a theoretical model to quantitatively explain the effect of temporal order of mutations in cancer initiation. The appearance of tumor is associated with fixation of several mutations in the tissue compartment that originally had a fixed number of normal stem cells. Our idea is that the cancer initiation process can be viewed as a sequence of stochastic transitions between discrete states that are defined by different numbers of cells with one or two mutations. This allows us to evaluate properties of cancer initiation and compare different mutations sequences. Specifically, we analyzed in detail the cancer initiation dynamics after the fixation of two different sequential mutations. It is shown that the probability of cancer initiation is fully determined by the fixation of the first mutation. This also suggests that the probability of the tumor formation is higher if the first mutation is more advantageous than the second one. These theoretical results explain recent experimental observations that emphasize the special role of initial truncating mutations in the human cancers [13]. We also found that cancer initiation times depend on both mutations, and the fastest

Figure 4. Heat maps for the ratio of the contributions to the overall fixation time T_{11}/T_{12} as a function of different fitness parameters. In calculations, the AB mutation sequence and $N = 1000$ are chosen.

dynamics is observed if the second mutation is more advantageous. These theoretical predictions are able to explain striking observations that the reverse order of mutations delays the formation of tumor by many years or might even not lead to cancer at all [9]. One of the most surprising observations of our theoretical analysis is the apparent anti-correlation between the probability of cancer initiation and the speed of tumor formation, which contrasts with naive expectations. To better understand this result, we developed effective free-energy landscapes for cancer initiation dynamics that allowed us to clarify the mechanistic origin of this phenomenon. It should be noted that such anti-correlations are also frequently observed in other stochastic processes such as chemical reactions.

While the presented theoretical model provides a quantitative description of the effect of mutations order in the cancer initiation, it is crucial to discuss its limitations. It assumes that the second mutation does not start until the first one is fully fixed. This is known as a linear evolution in tumor formation [2]. However, there is a limited amount of experimental data that support the linear evolution in cancers [2, 20]. Most studies favor a branching evolution picture in which different cellular clones with different mutations evolve in parallel in the growing tumors. At the same time, the linear evolution is a limiting case of the branched evolution when the mutation rate are very low, and this is frequently the case in many biological systems. In addition, there are arguments suggesting that early stages of cancer initiation can be well described by the linear evolution models [2]. Another simplification in our approach is the assumption that two mutations would lead to the formation of the tumor, while current data suggest that as many as 10 mutations are needed to drive some human cancers. Our method can be extended to multiple hits that follow the linear evolution. However, despite these limitations, the model still provides a clear physical picture for complex processes that are taking place during the formation of tumors.

Figure 5. The effective "free-energy" landscapes for cancer initiation with two alternative sequences of mutations: left panel describes AB mutations ($r_A > r_B$), while right panel describes BA mutations ($r_B > r_A$).

Acknowledgments

The work was supported by the Welch Foundation (C-1559), by the NSF (CHE-1664218), and by the Center for Theoretical Biological Physics sponsored by the NSF (PHY-1427654).

Appendix

I. Calculation of fixation probability and fixation times

One can define the corresponding first-passage probability $F_n(t)$ density functions to start from any state n and reach the site $2N$ at time t . The temporal evolution of these functions are governed by the following backward master equations,

$$\frac{dF_n(t)}{dt} = r_1 a_n F_{n+1}(t) + a_n F_{n-1}(t) - a_n(1 + r_1)F_n(t), \quad (13)$$

for $1 \leq n < N$, and

$$\frac{dF_n(t)}{dt} = r_2 a_{n-N} F_{n+1}(t) + r_1 a_{n-N} F_{n-1}(t) - a_{n-N}(r_2 + r_1)F_n(t), \quad (14)$$

for $N < n < 2N$, while

$$\frac{dF_N(t)}{dt} = Nu_2 F_{N+1}(t) - Nu_2 F_N(t) \quad (15)$$

for $n = N$.

$$\frac{dF_{N+1}(t)}{dt} = r_2 a_1 F_{N+2}(t) + r_1 a_1 F_N(t) - a_1(r_2 + r_1)F_{N+1}(t), \quad (16)$$

for $n = N + 1$. In addition we have the boundary condition $F_{2N} = \delta(t)$. It is convenient to solve this problem using the Laplace transformation, which changes the backward master equations:

$$\frac{s\widetilde{F}_n}{r_1 a_n} + (1 + 1/r_1)\widetilde{F}_n = \widetilde{F}_{n+1} + \frac{1}{r_1}\widetilde{F}_{n-1} \quad (17)$$

for $1 < n < N$, and

$$\frac{s\widetilde{F}_1}{r_1 a_1} + (1 + 1/r_1)\widetilde{F}_1 = \widetilde{F}_2 \quad (18)$$

for $n = 1$.

$$\frac{s\widetilde{F}_n}{r_2 a_{n-N}} + (1 + r_1/r_2)\widetilde{F}_n = \widetilde{F}_{n+1} + \frac{r_1}{r_2}\widetilde{F}_{n-1} \quad (19)$$

for $N < n < 2N$, and

$$(s + Nu_2)\widetilde{F}_N = Nu_2\widetilde{F}_{N+1} \quad (20)$$

for $n = N$.

$$\frac{s\widetilde{F}_{N+1}}{r_2 a_1} + (1 + r_1/r_2)\widetilde{F}_{N+1} = \widetilde{F}_{N+2} + \frac{r_1}{r_2} \left(\frac{Nu_2}{s + Nu_2} \right) \widetilde{F}_{N+1} \quad (21)$$

for $n = N + 1$. Because we are interested only in the fixation probabilities and fixation times, there is no need to obtain full analytical expressions for F_n , but it is needed to determine the expansion of this function up to the linear term in s . Thus, we can write

$$\widetilde{F}_n(s) \simeq \begin{cases} \Pi_n^{(1)} + B_n^{(1)} \frac{s}{r_1}; & 1 \leq n \leq N \\ \Pi_n^{(2)} + B_n^{(2)} \frac{s}{r_2}; & N + 1 \leq n \leq 2N \end{cases} \quad (22)$$

where $\Pi_n = \widetilde{F}_n(s=0)$, is the fixation probability starting from n single mutations, and the unknown parameters B_n are related to the fixation times (viewed as conditional mean first-passage times) as

$$T_n = \frac{-\partial \widetilde{F}_n / \partial s|_{s=0}}{\widetilde{F}_n(s=0)} = \frac{B_n^{(1)}}{r_1 \Pi_n^{(1)}} \quad (23)$$

Note that because of boundary condition at $n = 2N$ we have $\Pi_{2N}^{(2)} = 1$ and $B_{2N}^{(2)} = 0$. Substituting Eq. 22 into Eqs. 17 and 18 we obtain for the fixation probabilities

$$\frac{s}{r_1 a_n} \left[\Pi_n^{(1)} + B_n^{(1)} \frac{s}{r_1} \right] + (1 + 1/r_1) \left[\Pi_n^{(1)} + B_n^{(1)} \frac{s}{r_1} \right] = \left[\Pi_{n+1}^{(1)} + B_{n+1}^{(1)} \frac{s}{r_1} \right] + \frac{1}{r_1} \left[\Pi_{n-1}^{(1)} + B_{n-1}^{(1)} \frac{s}{r_1} \right] \quad (24)$$

$$\frac{s}{r_1 a_1} \left[\Pi_1^{(1)} + B_1^{(1)} \frac{s}{r_1} \right] + (1 + 1/r_1) \left[\Pi_1^{(1)} + B_1^{(1)} \frac{s}{r_1} \right] = \left[\Pi_2^{(1)} + B_2^{(1)} \frac{s}{r_1} \right] \quad (25)$$

These expressions yield following equation for fixation probabilities:

$$(1 + 1/r_1)\Pi_n^{(1)} = \Pi_{n+1}^{(1)} + \left(\frac{1}{r_1} \right) \Pi_{n-1}^{(1)} \quad (26)$$

$$(1 + 1/r)\Pi_1^{(1)} = \Pi_2^{(1)} \quad (27)$$

As shown in our previous work [23], these equations can be easily solved, leading to the following explicit expressions for the fixation probability,

$$\Pi_n^{(1)} = \frac{1 - 1/r_1^n}{1 - 1/r_1^N}. \quad (28)$$

To proceed further we substitute Eq. 22 into Eqs. 19, 20, and 21

$$\frac{s}{r_2 a_n} \left[\Pi_n^{(2)} + B_n^{(2)} \frac{s}{r_2} \right] + (1 + r_1/r_2) \left[\Pi_n^{(2)} + B_n^{(2)} \frac{s}{r_2} \right] = \left[\Pi_{n+1}^{(2)} + B_{n+1}^{(2)} \frac{s}{r_2} \right] + \frac{r_1}{r_2} \left[\Pi_{n-1}^{(2)} + B_{n-1}^{(2)} \frac{s}{r_2} \right] \quad (29)$$

$$(s + Nu_2) \left[\Pi_N^{(1)} + B_N^{(1)} \frac{s}{r_1} \right] = Nu_2 \left[\Pi_{N+1}^{(2)} + B_{N+1}^{(2)} \frac{s}{r_2} \right] \quad (30)$$

$$\begin{aligned} \frac{s}{r_2 a_1} \left[\Pi_{N+1}^{(2)} + B_{N+1}^{(2)} \frac{s}{r_2} \right] + (1 + r_1/r_2) \left[\Pi_{N+1}^{(2)} + B_{N+1}^{(2)} \frac{s}{r_2} \right] = \\ \left[\Pi_{N+2}^{(2)} + B_{N+2}^{(2)} \frac{s}{r_2} \right] + \frac{r_1}{r_2} \left(\frac{Nu_2}{s + Nu_2} \right) \left[\Pi_{N+1}^{(2)} + B_{N+1}^{(2)} \frac{s}{r_2} \right] \end{aligned} \quad (31)$$

From these equations we obtain:

$$\left(1 + \frac{r_1}{r_2} \right) \Pi_n^{(2)} = \Pi_{n+1}^{(2)} + \frac{r_1}{r_2} \Pi_{n-1}^{(2)} \quad (32)$$

$$\Pi_N^{(1)} = \Pi_{N+1}^{(2)} \quad (33)$$

$$\Pi_{N+1}^{(2)} = \Pi_{N+2}^{(2)} \quad (34)$$

From eqn 28 and from boundary condition $\Pi_{2N}^{(2)} = 1$ we obtain:

$$\Pi_{N+1}^{(2)} = \dots = \Pi_n^{(2)} = \Pi_{n+1}^{(2)} = \dots = \Pi_{2N}^{(2)} = 1 = \Pi_N^{(1)} \quad (35)$$

Thus fixation probabilities are given by:

$$\Pi_1^{(1)} = \frac{1 - 1/r_1}{1 - 1/r_1^N}. \quad (36)$$

$$\Pi_1^{(2)} = 1. \quad (37)$$

From Eqs. 24, the corresponding equations for parameters $B_n^{(1)}$ can be written as,

$$\frac{\Pi_n^{(1)}}{a_n} + \left(1 + \frac{1}{r_1} \right) B_n^{(1)} = B_{n+1}^{(1)} + \frac{B_{n-1}^{(1)}}{r_1}, \quad (38)$$

$$\frac{\Pi_1^{(1)}}{a_1} + \left(1 + \frac{1}{r} \right) B_1^{(1)} = B_2^{(1)}. \quad (39)$$

As was shown before, [23] $B_n^{(1)}$ is given by:

$$B_n^{(1)} = B_1^{(1)} \left(1 + \frac{1}{r_1} + \dots + \frac{1}{r_1^{n-1}} \right) + \sum_{j=1}^{n-1} \sum_{l=0}^{j-1} \frac{1}{r_1^l} \frac{\Pi_{j-l}^{(1)}}{a_{j-l}}. \quad (40)$$

This expression is valid for any $1 \leq n \leq N$. The constant $B_1^{(1)}$ can be found from the boundary condition at $n = N$.

$$B_N^{(1)} = \frac{r_1}{r_2} B_{N+1}^{(2)} - \frac{r_1}{u_1} \quad (41)$$

$$B_1^{(1)} = - \frac{\sum_{j=1}^{N-1} \sum_{l=0}^{j-1} \frac{1}{r_1^l} \frac{\Pi_{j-l}^{(1)}}{a_{j-l}}}{\left(1 + \frac{1}{r_1} + \dots + \frac{1}{r_1^{N-1}} \right)} + \frac{\frac{r_1}{r_2} B_{N+1}^{(2)} - \frac{r_1}{Nu_2}}{\left(1 + \frac{1}{r_1} + \dots + \frac{1}{r_1^{N-1}} \right)} \quad (42)$$

To proceed further we must calculate $B_{N+1}^{(2)}$. Similarly for $B_n^{(2)}$ we obtain following recursion relation:

$$\frac{1}{a_n} + \left(1 + \frac{r_1}{r_2}\right) B_n^{(2)} = B_{n+1}^{(2)} + \frac{r_1}{r_2} B_{n-1}^{(2)} \quad (43)$$

Now we define :

$$K_n = B_{n+1}^{(2)} - B_n^{(2)} \quad (44)$$

Thus Eqn. 43 takes the form:

$$\frac{1}{a_n} = K_n - xK_{n-1} \quad (45)$$

where $x = \frac{r_1}{r_2}$. Using Eqn. 41, K_N can be written as:

$$K_N = \left(1 - \frac{r_1}{r_2}\right) B_{N+1}^{(2)} + \frac{r_1}{Nu_2} \quad (46)$$

It can be shown that solution of recurrence relation 45 is given by:

$$K_n = \sum_{l=0}^{n-N-2} \frac{x^l}{a_{n-l-N}} + x^{n-N-1} K_{N+1} \quad (47)$$

Thus $B_n^{(2)}$ is given by:

$$B_n^{(2)} = B_{N+1}^{(2)} + \sum_{j=N+1}^{n-1} K_j \quad (48)$$

$$B_n^{(2)} = B_{N+1}^{(2)} + \sum_{j=N+1}^{n-1} \sum_{l=0}^{j-N-2} \frac{x^l}{a_{j-l-N}} + \sum_{j=N+1}^{n-1} x^{j-N-1} K_{N+1} \quad (49)$$

Now we can calculate K_{N+1} :

$$K_{N+1} = B_{N+2}^{(2)} - B_{N+1}^{(2)} = \frac{1}{a_1} + \frac{r_1}{Nu_2} \quad (50)$$

Combining Eqns. 50 and 49 yields:

$$B_n^{(2)} = B_{N+1}^{(2)} + \sum_{j=N+1}^{n-1} \sum_{l=0}^{j-N-2} \frac{x^l}{a_{j-l-N}} + \sum_{j=N+1}^{n-1} x^{j-N-1} \left(\frac{1}{a_1} + \frac{r_1}{Nu_2} \right) \quad (51)$$

We can obtain $B_{N+1}^{(2)}$ using boundary condition $B_{2N}^{(2)} = 0$:

$$B_{N+1}^{(2)} = - \sum_{j=N+1}^{2N-1} \sum_{l=0}^{j-N-2} \frac{x^l}{a_{j-l-N}} - \sum_{j=N+1}^{2N-1} x^{j-N-1} \left(\frac{1}{a_1} + \frac{r_1}{Nu_2} \right) \quad (52)$$

Substituting Eqn. 52 into Eqn. 42 yields:

$$B_1^{(1)} = - \frac{\sum_{j=1}^{N-1} \sum_{l=0}^{j-1} \frac{1}{r_1^l} \frac{\Pi_{j-l}^{(1)}}{a_{j-l}}}{\left(1 + \frac{1}{r_1} + \dots + \frac{1}{r_1^{N-1}}\right)} - \frac{\frac{r_1}{Nu_2} \left[1 + \frac{r_1}{r_2} \sum_{j=N+1}^{2N-1} x^{j-N-1}\right]}{\left(1 + \frac{1}{r_1} + \dots + \frac{1}{r_1^{N-1}}\right)} - \frac{\left[\frac{r_1}{r_2 a_1} \sum_{j=N+1}^{2N-1} x^{j-N-1} + \frac{r_1}{r_2} \sum_{j=N+1}^{2N-1} \sum_{l=0}^{j-N-2} \frac{x^l}{a_{j-l-N}}\right]}{\left(1 + \frac{1}{r_1} + \dots + \frac{1}{r_1^{N-1}}\right)} \quad (53)$$

It can be shown that:

$$1 + \frac{1}{r_1} + \dots + \frac{1}{r_1^{N-1}} = \sum_{i=0}^{N-1} r_1^{-i} = \frac{1 - 1/r_1^N}{1 - 1/r_1} = \frac{1}{\Pi_1^{(1)}} \quad (54)$$

Now we can calculate fixation times using eqn. 23,

$$T_1 = \frac{N+1}{br_1(1-r_1^{-N})} \sum_{j=1}^{N-1} \sum_{l=0}^{j-1} \frac{(r_1^{-l} - r_1^{-j})}{(j-l)(N-j+l)} + \frac{1}{Nu_2} \left[\frac{1 - (r_1/r_2)^N}{1 - r_1/r_2} \right] + \frac{N+1}{br_2} \sum_{j=1}^{N-1} \sum_{l=0}^{j-1} \frac{r_1^l r_2^{-l}}{(j-l)(N-j+l)} \quad (55)$$

This expression can be further simplified by defining a new index $n = j - l$.

$$T_1 = \frac{N+1}{b} \sum_{n=1}^{N-1} \frac{1}{n(N-n)} \left(\frac{r_1^n - 1}{r_1 - 1} \right) \left(\frac{r_1^{N-n} - 1}{r_1^{N-1} - 1} \right) + \frac{1}{Nu_2} \left[\frac{1 - (r_1/r_2)^N}{1 - r_1/r_2} \right] + \frac{N+1}{br_2} \sum_{n=1}^{N-1} \frac{1}{n(N-n)} \left(\frac{1 - (r_1/r_2)^{N-n}}{1 - r_1/r_2} \right) \quad (56)$$

II. Explicit expressions for fixation times for $N \rightarrow \infty$

In general it is difficult to perform explicit summation for larger N . For $N \rightarrow \infty$, we can convert summation to integration:

$$T_1 \simeq \frac{1}{b(r_1 - 1)(r_1^N - 1)} [2(1 + r_1^N) \ln(N-1) + \text{Ei}(\ln r_1) + 2r_1^N \text{Ei}(-\ln r_1)] + \frac{1}{br_2(1 - r_1/r_2)} \left[2 \ln(N-1) + \left(\frac{r_1}{r_2} \right)^N \text{Ei}(-\ln \frac{r_1}{r_2}) + \text{Ei}(\ln \frac{r_1}{r_2}) \right] + \frac{1}{Nu_2} \left[\frac{1 - (r_1/r_2)^N}{1 - r_1/r_2} \right] \quad (57)$$

III. Full exact solution of F_n for $N = 3$

In this section we obtain a full exact solution of the first-passage probability density function F_n for $N = 3$. This analytical solution is extremely useful, because it allows us to verify the general exact solution, which was obtained based on an steady state ansatz. As shown in figure 1, for $N = 3$ there are 6 states. Correspondingly we define the first-passage probability $F_n(t)$ functions to start from any state n and reach the site $2N$ at time t . The temporal evolution of these functions is governed by a set of backward master equations:

$$\frac{dF_1(t)}{dt} = r_1 a_1 F_2(t) - a_1(1 + r_1) F_1(t) \quad (58)$$

$$\frac{dF_2(t)}{dt} = a_2 F_1(t) + r_1 a_2 F_3(t) - a_2(1 + r_1) F_2(t) \quad (59)$$

$$\frac{dF_3(t)}{dt} = 2u_2 F_4(t) - 2u_2 F_3(t) \quad (60)$$

$$\frac{dF_4(t)}{dt} = r_2 a_4 F_5(t) + r_1 a_4 F_3(t) - a_4(r_1 + r_2) F_4(t) \quad (61)$$

$$\frac{dF_5(t)}{dt} = r_1 a_5 F_4(t) + r_2 a_5 F_6(t) - a_5(r_1 + r_2) F_5(t) \quad (62)$$

In addition we have the boundary condition $F_6(t) = \delta(t)$. Performing Laplace transformation over these equations we get:

$$(s + a_1(1 + r_1))\widetilde{F}_1 = r_1 a_1 \widetilde{F}_2 \quad (63)$$

$$(s + a_2(1 + r_1))\widetilde{F}_2 = a_2 \widetilde{F}_1 + r_1 a_2 \widetilde{F}_3 \quad (64)$$

$$(s + 2u_2)\widetilde{F}_3 = 2u_2 \widetilde{F}_4 \quad (65)$$

$$(s + a_4(r_1 + r_2))\widetilde{F}_4 = r_2 a_4 \widetilde{F}_5 + r_1 a_4 \widetilde{F}_3 \quad (66)$$

$$(s + a_5(r_1 + r_2))\widetilde{F}_5 = r_1 a_5 \widetilde{F}_4 + r_2 a_5 \quad (67)$$

We note that because of symmetry we have $a_4 = a_1 = \frac{1}{2}$ and $a_5 = a_2 = \frac{1}{2}$. Solving this

Appendix II - Figure 1. Discrete-state stochastic model of 2-hit mutation for $N = 3$.

system of five equations and five unknowns yields $\widetilde{F}_1$, $\widetilde{F}_2$, $\widetilde{F}_3$, $\widetilde{F}_4$ and $\widetilde{F}_5$. Specifically we need $\widetilde{F}_1(s)$ to calculate fixation time and fixation probability. After some algebra we obtain:

$$\widetilde{F}_1(s) = \frac{r_1^2 r_2^2 2u_2}{[r_1(4s + 1) + r_1^2 + (2s + 1)^2][r_1 s(r_2 + 4s + 4u_2) + (r_2 + 2s)^2(s + 2u_2) + r_1^2 s]} \quad (68)$$

Expanding this function in terms of s yields the fixation probability,

$$\Pi_1 = \widetilde{F}_1(s = 0) = \frac{r_1^2}{r_1^2 + r_1 + 1} = \frac{1 - 1/r_1}{1 - 1/(r_1)^3} \quad (69)$$

and, the fixation time:

$$T_1 = \frac{-\frac{\partial \widetilde{F}_1}{\partial s}|_{s=0}}{\widetilde{F}_1(s=0)} = \frac{4(1 + r_1)}{r_1^2 + r_1 + 1} + \frac{1}{2u_2} \left[1 + \frac{r_1}{r_2} + \left(\frac{r_1}{r_2}\right)^2 \right] + \frac{2}{r_2} \left[2 + \frac{r_1}{r_2} \right] \quad (70)$$

Similarly, the Taylor expansion of $\widetilde{F}_3$, $\widetilde{F}_4$ and $\widetilde{F}_5$ in terms of s yields the other fixation probabilities,

$$\Pi_3 = \widetilde{F}_3(s = 0) = 1 \quad (71)$$

$$\Pi_4 = \widetilde{F}_4(s = 0) = 1 \quad (72)$$

$$\Pi_5 = \widetilde{F}_5(s = 0) = 1 \quad (73)$$

Now we expand 56 for $N = 3$ and after some algebra we obtain equation 70:

$$\begin{aligned} T_1 &= 4 \left[\frac{1}{2} \left(\frac{r_1 - 1}{r_1 - 1} \right) \left(\frac{r_1^2 - 1}{r_1^3 - 1} \right) + \frac{1}{2} \left(\frac{r_1^2 - 1}{r_1 - 1} \right) \left(\frac{r_1 - 1}{r_1^3 - 1} \right) \right] + \frac{1}{2u_2} \left[1 + \frac{r_1}{r_2} + \left(\frac{r_1}{r_2}\right)^2 \right] \\ &\quad + \frac{4}{r_2} \left[\frac{1}{2} \left(\frac{1 - (r_1/r_2)^2}{1 - (r_1/r_2)} \right) + \frac{1}{2} \left(\frac{1 - r_1/r_2}{1 - r_1/r_2} \right) \right] \\ &= \frac{4(1 + r_1)}{r_1^2 + r_1 + 1} + \frac{1}{2u_2} \left[1 + \frac{r_1}{r_2} + \left(\frac{r_1}{r_2}\right)^2 \right] + \frac{2}{r_2} \left[2 + \frac{r_1}{r_2} \right] \quad (74) \end{aligned}$$

IV. Effect of the number of stem cells

Another relevant parameter in the cancer initiation is the number of stem cells. Figure 2 shows how varying N influences the differences in the cancer initiation probabilities and times. The differences are relatively small for low number of cells, but they start to increase and eventually saturate at $N \gg 1$. The differences are stronger for the fitness parameters that are closer to unity. This can be easily explained for the fixation probabilities because $\Pi_1 = 1 - 1/r_1$ in the limit of $N \rightarrow \infty$. In this limit, from Eq. (57) the fixation time can be approximated as $T_1 \sim \ln N$ with the coefficient that depends only the fitness parameters r_1 and r_2 , and this leads to the constant ratio of the fixation times for different mutation sequences.

Appendix IV - Figure 2. (a) The ratio of fixation probabilities Π_{AB}/Π_{BA} as a function of the number of cells N . (b) The ratio of fixation times T_{AB}/T_{BA} as a function of the number of cells N . For calculations, $u_2 = 10^{-4}$ is used.

References

1. P. Ashcroft, F. Michor, and T. Galla. Stochastic tunneling and metastable states during the somatic evolution of cancer. *Genetics*, 199(4):1213–1228, 2015.
2. A. Davis, R. Gao, and N. Navin. Tumor evolution: Linear, branching, neutral or punctuated? *Biochimica et Biophysica Acta (BBA)-Reviews on Cancer*, 1867(2):151–161, 2017.

3. F. Delhommeau, S. Dupont, V. D. Valle, C. James, S. Trannoy, A. Masse, O. Kosmider, J.-P. Le Couedic, F. Robert, A. Alberdi, et al. Mutation in tet2 in myeloid cancers. *New England Journal of Medicine*, 360(22):2289–2301, 2009.
4. S. A. Frank. *Dynamics of cancer: incidence, inheritance, and evolution*, volume 1. Princeton University Press, 2007.
5. H. Haeno, Y. E. Maruvka, Y. Iwasa, and F. Michor. Stochastic tunneling of two mutations in a population of cancer cells. *PloS one*, 8(6):e65724, 2013.
6. D. Hanahan and R. A. Weinberg. Hallmarks of cancer: the next generation. *cell*, 144(5):646–674, 2011.
7. M. R. Hassler and G. Egger. Epigenomics of cancer—emerging new concepts. *Biochimie*, 94(11):2219–2230, 2012.
8. M. Herbet, A. Salomon, J.-J. Feige, and M. Thomas. Acquisition order of ras and p53 gene alterations defines distinct adrenocortical tumor phenotypes. *PLoS genetics*, 8(5):e1002700, 2012.
9. D. G. Kent and A. R. Green. Order matters: the order of somatic mutations influences cancer evolution. *Cold Spring Harbor perspectives in medicine*, 7(4):a027060, 2017.
10. T. Klampfl, H. Gisslinger, A. S. Harutyunyan, H. Nivarthi, E. Rumi, J. D. Milosevic, N. C. Them, T. Berg, B. Gisslinger, D. Pietra, et al. Somatic mutations of calreticulin in myeloproliferative neoplasms. *New England Journal of Medicine*, 369(25):2379–2390, 2013.
11. A. G. Knudson. Two genetic hits (more or less) to cancer. *Nature Reviews Cancer*, 1(2):157, 2001.
12. N. L. Komarova, A. Sengupta, and M. A. Nowak. Mutation–selection networks of cancer initiation: tumor suppressor genes and chromosomal instability. *Journal of theoretical biology*, 223(4):433–450, 2003.
13. A. J. Levine, N. A. Jenkins, and N. G. Copeland. The roles of initiating truncal mutations in human cancers: the order of mutations and tumor cell type matters. *Cancer cell*, 35(1):10–15, 2019.
14. R. L. Levine and D. G. Gilliland. Jak-2 mutations and their relevance to myeloproliferative disease. *Current opinion in hematology*, 14(1):43–47, 2007.
15. M. Lynch. Rate, molecular spectrum, and consequences of human mutation. *Proceedings of the National Academy of Sciences*, 107(3):961–968, 2010.
16. F. Michor, Y. Iwasa, and M. A. Nowak. Dynamics of cancer progression. *Nature reviews cancer*, 4(3):197, 2004.
17. P. MORAN et al. The statistical processes of evolutionary theory. *The statistical processes of evolutionary theory.*, 1962.
18. J. Nangalia, C. E. Massie, E. J. Baxter, F. L. Nice, G. Gundem, D. C. Wedge, E. Avezov, J. Li, K. Kollmann, D. G. Kent, et al. Somatic calr mutations in myeloproliferative neoplasms with nonmutated jak2. *New England Journal of Medicine*, 369(25):2391–2405, 2013.
19. M. A. Nowak. *Evolutionary dynamics: exploring the equations of life*. Harvard university press, 2006.

20. M. A. Nowak, F. Michor, and Y. Iwasa. The linear process of somatic evolution. *Proceedings of the national academy of sciences*, 100(25):14966–14969, 2003. 420
421
21. M. A. Nowak and B. Waclaw. Genes, environment, and “bad luck”. *Science*, 355(6331):1266–1267, 2017. 422
423
22. C. A. Ortmann, D. G. Kent, J. Nangalia, Y. Silber, D. C. Wedge, J. Grinfeld, E. J. Baxter, C. E. Massie, E. Papaemmanuil, S. Menon, et al. Effect of mutation order on myeloproliferative neoplasms. *New England Journal of Medicine*, 372(7):601–612, 2015. 424
425
426
427
23. H. Teimouri, M. P. Kochugaeva, and A. B. Kolomeisky. Elucidating the correlations between cancer initiation times and lifetime cancer risks. *Scientific reports*, 9(1):1–8, 2019. 428
429
430
24. B. Vogelstein, N. Papadopoulos, V. E. Velculescu, S. Zhou, L. A. Diaz, and K. W. Kinzler. Cancer genome landscapes. *science*, 339(6127):1546–1558, 2013. 431
432
25. R. A. Weinberg. *The biology of cancer*. Garland science, 2013. 433