Elucidating the correlations between cancer initiation times and lifetime cancer risks

Hamid Teimouri1,2,5, Maria P. Kochugaeva3,5 & Anatoly B. Kolomeisky1,2,4*

Cancer is a genetic disease that results from accumulation of unfavorable mutations. As soon as genetic and epigenetic modifications associated with these mutations become strong enough, the uncontrolled tumor cell growth is initiated, eventually spreading through healthy tissues. Clarifying the dynamics of cancer initiation is thus critically important for understanding the molecular mechanisms of tumorigenesis. Here we present a new theoretical method to evaluate the dynamic processes associated with the cancer initiation. It is based on a discrete-state stochastic description of the formation of tumors as a fixation of cancerous mutations in tissues. Using a first-passage analysis the probabilities for the cancer to appear and the times before it happens, which are viewed as fixation probabilities and fixation times, respectively, are explicitly calculated. It is predicted that the slowest cancer initiation dynamics is observed for neutral mutations, while it is fast for both advantageous and, surprisingly, disadvantageous mutations. The method is applied for estimating the cancer initiation times from experimentally available lifetime cancer risks for different types of cancer. It is found that the higher probability of the cancer to occur does not necessarily lead to the faster times of starting the cancer. Our theoretical analysis helps to clarify microscopic aspects of cancer initiation processes.

It is well known that tumor cells are characterized by abnormal cell division rates, which is a result of mutations in cancer-susceptible genes (known as oncogenes)1–4. Specifically, these mutations affect the regulation of cell proliferation and differentiation via activation of oncogenes or inactivation of tumor suppressor genes (TSGs)2–5. Mutations are taking place randomly, and after several cellular replications some of them might occasionally lead to significant genetic and epigenetic alterations such that the normal cells behavior changes to the uncontrolled proliferation, eventually starting a cancer3,6,7. After these cancer initiation events, rapid changes are taking place with a newly formed tumor being able to escape cellular control mechanisms, and the cancer progresses into more invasive forms3,4,6,8. But this happens only after the initial stage of cancer succeeds, and thus it is critically important to understand the dynamics of cancer initiation6.

Human tissues and organs are composed of heterogeneous mixtures of cells: not all cells are equal in their potential to proliferate. An important role in tissue maintenance and repair is played by a population of so-called stem cells10. These cells are characterized by their ability to self-renew and make more stem cells or ability to produce differentiated progenitor cells11. Epithelial tissues are also known for subdivision into compartments where homeostatic mechanism, a balance between self-renewal and differentiation, maintains the constant cell number. Cancer appears in such compartments, breaking the homeostatic tissue equilibrium. However, having only a single mutated cell in the compartment does not lead to cancer. The cancer initiation event generally is associated with a fixation of one or several mutations, i.e., when all cells in the compartment become mutated, or when a significant fraction of them is mutated, producing noticeable genetic and epigenetic changes6,9.

One of the most important quantities that determines if the person gets a cancer is a cancer lifetime risk. It refers to a probability of being diagnosed with or dying from cancer during the person’s lifespan. Lifetime risks strongly depend on the type of cancer. For example, a person’s risk of getting a lung cancer is more than 11 times higher than of developing of a brain cancer, and 8 times greater than that of a stomach cancer12,13. Various studies have attributed the differences in cancer rates to environmental risk factors, such as smoking, bad dietary habits.

1Department of Chemistry, Rice University, Houston, TX, United States. 2Center for Theoretical Biological Physics, Rice University, Houston, TX, United States. 3Department of Biomedical Engineering and Systems Biology Institute, Yale University, West Haven, CT, United States. 4Department of Chemical and Biomolecular Engineering, Rice University, Houston, TX, United States. 5These authors contributed equally: Hamid Teimouri and Maria P. Kochugaeva. *email: tolya@rice.edu
or exposure to UV light, as well as to heritable mutations. However, the environmental factors and the hereditary factors cannot fully explain the substantial differences in the cancer rates across tissues. Moreover, the total numbers of cells that make up these tissues also cannot explain varying cancer risks. Recent statistical analysis of 31 cancers by Tomasetti and Vogelstein suggested that there is a strong correlation between random mutations acquired during stem cell divisions and lifetime cancer risk.

At first glance, the tumors might appear faster for higher lifetime cancer risks. But the correlation between cancer initiation time and the lifetime cancer risk has not been methodically investigated. There are certain types of cancers with low lifetime risks that occur at early ages, while there are other types with high lifetime risks that happen at older ages. For example, several pediatric cancers such as Wilms tumors, carcinomas and neuroendocrine tumors appear very early but they also have low lifetime risks. Therefore it is crucially important to estimate the initiation times for different cancer types and its relations with the corresponding lifetime risks. In recent years, several mathematical models have been developed for analysis of cancer initiation and progression dynamics. However, some important microscopic aspects of the evolutionary processes leading to cancer remain unexplained. For example, the state of the system when the mutated cells take over the whole tissue compartment, which is known as a fixation, is frequently associated with the formation of the tumor. While the probability to reach the fixation, called a fixation probability, has been explicitly evaluated, the time to reach the fixation (fixation time) has been estimated only approximately using numerical and computer simulations methods for limited range of parameters.

Here we develop a new theoretical framework of explicitly evaluating the cancer initiation dynamics. In our theoretical approach, the mutation fixation in the tissue compartment is assumed as the point of cancer initiation. Applying a discrete-state stochastic approach with a first-passage analysis, the fixation probabilities and fixation times are calculated explicitly. Utilizing our theoretical predictions, we extract relevant parameters from experimental data on lifetime risks for different types of cancer, which are used then to estimate the specific cancer initiation times. Our theoretical analysis suggests that generally there is no correlations between the probability and mean time of getting cancer, suggesting that both properties, as a minimal requirement, should be utilized in evaluation of cancer risks.

Methods
Theoretical model. Let us consider a tissue compartment that has at the beginning $N$ normal stem cells as shown in Fig. 1 (Top). At some specific time, which we set as a time zero (which, however, does not correspond to the person’s birth time), one stem cell undergoes a mutation with a probability $\mu$. Here we consider only driver mutations, i.e. those that promote the cancer development. Both normal and mutated stem cells in the tissue can replicate, but with different rates. To reflect the effect of somatic mutations, the mutated cells are characterized by a fitness parameter $r$, which is defined as the ratio of the division rate of the mutated cells to the division rate of the normal cells. If $r > 1$ the mutation is advantageous; if $r < 1$ the mutation is disadvantageous; and $r = 1$ describes a neutral mutation. It is expected that most mutations in oncogenes that lead to cancer are advantageous. The important characteristic of normal tissues is the homeostatic equilibrium, i.e., the total number of cells in the compartment remains constant. To incorporate this property into our theoretical model we assume that the system follows a birth-death process known as a Moran process. This means that after division of any randomly chosen cell the number of cells in the compartment increases by one, and then one of the randomly chosen cells should be instantly removed to keep the number of cell constant and equal to $N$. We identify the fixation state as a starting point

![Figure 1](https://www.nature.com/scientificreports/)
of the cancer because there are no normal stem cells left in the compartment. Thus, the cancer initiation dynamics in our model is viewed as a process of transitioning from the state 1 to the state \(N\). This is a stochastic process which is governed by various transition rates. Following the description of the Moran processes and considering the two-stage process for replication via cell division and immediate cell removal to fix the total number of cells (see the details of the derivation in the SI), the forward transition rate from the state \(n\) to the state \(n+1\) is given by

\[
\frac{r}{n(N-n)} \frac{n}{N+1},
\]

and \(b\) is a division rate of the normal stem cell. The factor \(r\) comes from the fact that this transition is taking place due to replication of the mutated cell and the corresponding instantaneous removal (to keep the homeostatic equilibrium) of the normal stem cell. The backward transition (from the state \(n\) to the state \(n-1\)) is equal just to \(an\) because it describes the replication of the normal cell and the removal of the mutated cell.

In our theoretical framework, the cancer starts when the system reaches the state \(N\) for the first time starting initially in the state 1. This suggests that the cancer initiation dynamics can be conveniently described as a first-passage process. One can define a first-passage probability density function \(F_t(n)\) that describes the probability to reach the state \(N\) for the first time at time \(t\) if at \(t=0\) the system started in the state \(n\). The temporal evolution of these functions can be described by a set of so-called backward master equations.

\[
\frac{dF_t(n)}{dt} = r a_n F_{n+1}(t) + a_n F_{n-1}(t) - a_n(1 + r)F_n(t),
\]

for \(1 < n < N\); and

\[
\frac{dF_t(1)}{dt} = r a_1 F_2(t) - a_1(1 + r)F_1(t).
\]

In addition, we have the boundary condition \(F_t(N) = \delta(t)\), the physical meaning of which is that the fixation process is immediately accomplished if the system starts from the state \(N\).

First-passage probability functions contain a comprehensive dynamic description of the fixation process. In this work, we are interested in fixation probabilities \(\pi_n \equiv \int_0^T F_n(t)dt\) and fixation times \(\bar{T}_n \equiv \int_0^\infty tF_n(t)dt\), which are analytically calculated in the SI. For example, for the fixation probability from the state \(n\) we obtain

\[
\pi_n = \frac{1 - 1/r^n}{1 - 1/r^N}
\]

which is a well known result. For \(r \to 1\) we get \(\pi_n = n/N\). In Fig. 2a, the fixation probability \(\pi_1\) is presented for different values of the parameters \(r\) and \(N\). One can see that for large values of \(N\) the fixation probability depends only on the fitness parameter \(r\).

A critically important feature of the cancer initiation process is how long does it take to reach the cancer starting point, which can be defined as a cancer initiation time. In our language, it corresponds to the fixation time for the mutation that activates the oncogene. More specifically, it is given by \(\bar{T}_n\), which as a conditional mean first-passage time to reach the fixation state \((n = N)\) from the state with initially \(n = 1\) mutated cells before the mutation can be eliminated from the system \((n = 0)\). From biological point of view, the cancer initiation time can be interpreted as the average time interval between the occurrence of the initial mutation and the state when all cells in the tissue compartment become mutated. Our explicit calculations (presented in the SI) provide the following expression.
As shown in the SI, for $r \to 1$ and $N \to \infty$ we obtain

$$T_i \approx N/b.$$  

(6)

It can be also shown (see the SI for details) that for large $N$ the expression for the fixation time can be simplified into

$$T_i \approx \frac{1}{b} \frac{1}{r(1 - \frac{1}{r^N})} \left[ \frac{\text{Ei}(-\ln r)}{\ln r} \left(1 - \frac{1}{r}\right) + \frac{2}{\ln r} (\gamma + \ln [N \ln r]) \right].$$  

(7)

where $\text{Ei}(x)$ is the exponential integral defined as $\text{Ei}(x) \equiv -\int_{-\infty}^{x} \frac{e^{-z}}{z} dz$, and $\gamma$ is the Euler–Mascheroni constant.

The results for fixation times as functions of $N$ and $r$ are presented in Fig. 2b. The slowest cancer initiation dynamics is expected for the neutral mutations ($r = 1$). This can be easily understood because in this case the system performs the unbiased random walk between the discrete states (see Fig. 1 Bottom), redundantly visiting the same states many times. As expected, for the advantageous mutations ($r > 1$) the cancer initiation times are lower since the dynamics is biased in the direction of increasing the number of mutated cells in the tissue compartment (Fig. 1). This will drive the system faster in the direction of the fixation. Surprisingly, the fixation times are also fast for the disadvantageous mutations ($r < 1$). This unexpected result can be explained using the following arguments. The fixation time is a conditional time only for those events that reach the fixation. The system is biased in the direction of eliminating of mutations and this leads to very low fixation probabilities (see Fig. 2a). However, for those rare cases when the system goes to the fixation they must occur quickly in order not to be influenced by the bias in the opposite direction. Those trajectories consist mostly of the forward transitions (in the direction of fixation). But because at each intermediate step the probability to go forward is low ($r < 1$), the probability to observe such forward trajectories is also low and it exponentially decreases with the size of the system. Thus, these fixation events for $r < 1$ are fast, but the probability of observing them is very low. From this point of view, it seems that these situations physiologically probably are not very relevant.

**Results**

**Estimation of fitness parameters for different types of cancer.** To calculate explicitly the initiation times for different types of cancer, we need to estimate the fitness parameter $r$ and the number of stem cells $N$ in the specific tissues. The latter has been well evaluated before. However, the estimation of the fitness parameter $r$ is much more difficult, and it requires several approximations.

In our analysis, we follow a recently proposed simple mathematical approach. According to this method, the cancer initiation probability, i.e., the probability that a mutation is fixed in the compartment of $N$ cells during a human lifetime, is given by:

$$P_{in} \approx NbT_{life} \mu \pi_i,$$  

(8)

where $T_{life}$ is the typical human lifetime (we assume here $T_{life} = 80$ years), and $\mu$ is the probability of mutation (activation of the oncogene) multiplied by the number of possible oncogenes, which varies for the different tissues. This result can be physically explained using the following arguments. The system can move in the direction of the fixation state, which is associated with the start of the cancer, only after cell divisions are taking place. There are $Nb$ such divisions per unit time in the tissue with $N$ cells, and over the human lifetime the total number of such divisions will be $NbT_{life}$. The cancer will not start until, at least, one of the oncogenes is activated, which has the probability $\mu$. Finally, $\pi_i$ describes the fixation probability that this mutation will not disappear but will occupy the whole tissue compartment.

However, the cancer initiation probability $P_{in}$ is not exactly the cancer lifetime risk $R_{tr}$, that has been evaluated from various clinical data. At the same time, it can be argued that both quantities are related as:

$$R_{tr} = P_{in}Q_{pp},$$  

(9)

where $Q_{pp}$ is a probability of cancer progression, i.e., the probability that after the cancer initiation the tumor will grow and the homeostatic equilibrium will be broken. From Eqs. (8) and (9), we obtain

$$R_{tr} = NbT_{life} \mu \pi_i Q_{pp}.$$  

(10)

Another factor that helps in estimating the fitness parameter $r$ is that, typically, the number of stem cells $N$ is very large. This leads to

$$\pi_i \approx \begin{cases} 0; & r < 1 \\ 1; & r = 1 \\ 1 - \frac{1}{r}; & r > 1 \end{cases}$$  

(11)
Finally, Eq. (12) is used to estimate the fitness parameters for using the data for breast cancers. One can reasonably assume that several types of cancers, and the results are presented in Table 1. Because the number of stem cells is very large, it can be shown from Eqs. (7) and (12) that

$$r \simeq \frac{1}{1 - \frac{R_{tr}}{b R_{tr} t_{life} Q_{pr} N}} \simeq 1 + \left[ \frac{R_{tr}}{b R_{tr} t_{life} Q_{pr} N} \right] \frac{1}{N}. \tag{12}$$

This is an important result because it relates the fitness parameter to the number of stem cells. It is also consistent with ideas presented earlier, where it was argued that cancer initiation corresponds to gaining a fitness value greater than the threshold value $r^* \simeq 1 + 1/N$. Finally, Eq. (12) is used to estimate the fitness parameters for several types of cancers, and the results are presented in Table 1.

### Table 1. Cancer development properties for 28 different cancer types. Data are adapted from 13. In calculations of fixation times and $t_{life}$, $\mu = 3 \times 10^{-8}$ and $Q_{pr} = 0.001$ were utilized. Note that 3 cancers from 13 are not analyzed because no division rate $b$ was reported for them.

| Cancer type | $R_{tr}$ | $b$ (per year) | $t_{life}$ (years) | $t_{0}$ (years) |
|-------------|---------|----------------|-------------------|---------------|
| Acute myeloid leukemia | 0.0041 | $1.35 \times 10^4$ | 12.00 | 1.0011 | 1964.0 | 0.021 |
| Basal cell carcinoma | 0.3000 | $5.82 \times 10^4$ | 7.60 | 1.0028 | 1595.8 | 0.001 |
| Chronic lymphocytic leukemia | 0.0052 | $1.35 \times 10^4$ | 12.00 | 1.0013 | 1577.6 | 0.021 |
| Colorectal adenocarcinoma | 0.0480 | $2.00 \times 10^4$ | 73.00 | 1.0014 | 261.5 | 0.002 |
| Colorectal adenocarcinoma with FAP | 1.0000 | $2.00 \times 10^4$ | 73.00 | 1.0285 | 15.0 | 0.002 |
| Colorectal adenocarcinoma with Lynch syndrome | 0.5000 | $2.00 \times 10^4$ | 73.00 | 1.0143 | 29.2 | 0.002 |
| Duodenum adenocarcinoma | 0.0003 | $4.00 \times 10^4$ | 24.00 | 1.0013 | 583.6 | 0.35 |
| Duodenum adenocarcinoma with FAP | 0.0350 | $4.00 \times 10^4$ | 24.00 | 1.1519 | 6.6 | 0.35 |
| Esophageal squamous cell carcinoma | 0.0019 | $8.64 \times 10^4$ | 17.40 | 1.0537 | 22.9 | 2.2 |
| Gallbladder non papillary adenocarcinoma | 0.0028 | $1.60 \times 10^4$ | 0.58 | 2.2486 | 18.3 | 35.7 |
| Head & neck squamous cell carcinoma | 0.0138 | $1.85 \times 10^4$ | 21.50 | 1.0145 | 82.8 | 0.08 |
| Head & neck squamous cell carcinoma with HPV-16 | 0.0794 | $1.85 \times 10^4$ | 21.50 | 1.0831 | 15.2 | 0.08 |
| Hepatocellular carcinoma | 0.0071 | $3.01 \times 10^4$ | 0.91 | 1.0011 | 3164.3 | 0.012 |
| Hepatocellular carcinoma with HCV | 0.0170 | $3.01 \times 10^4$ | 0.91 | 1.0108 | 3593.6 | 0.012 |
| Lung adenocarcinoma (nonsmokers) | 0.0045 | $1.22 \times 10^4$ | 0.07 | 1.0220 | 22468.1 | 0.39 |
| Lung adenocarcinoma (smokers) | 0.0810 | $1.22 \times 10^4$ | 0.07 | 1.3952 | 1058.7 | 0.39 |
| Melanoma | 0.0203 | $3.80 \times 10^4$ | 2.48 | 1.0009 | 14019.2 | 0.004 |
| Osteosarcoma | 0.0004 | $4.18 \times 10^4$ | 0.07 | 1.5207 | 566.6 | 119.02 |
| Osteosarcoma of the arms | 0.0004 | $6.50 \times 10^4$ | 0.07 | 1.3827 | 725.5 | 765.4 |
| Osteosarcoma of the head | 0.0003 | $8.60 \times 10^4$ | 0.07 | 1.2169 | 1423.3 | 578.5 |
| Osteosarcoma of the legs | 0.00022 | $1.59 \times 10^4$ | 0.07 | 1.8605 | 272.0 | 312.9 |
| Osteosarcoma of the pelvis | 0.0003 | $4.50 \times 10^4$ | 0.07 | 1.4146 | 639.9 | 1105.6 |
| Pancreatic ductal adenocarcinoma | 0.0136 | $4.18 \times 10^4$ | 1.00 | 1.0014 | 23772.1 | 0.008 |
| Pancreatic endocrine carcinoma | 0.0002 | $7.40 \times 10^4$ | 1.00 | 1.0011 | 21716.2 | 0.45 |
| Small intestine adenocarcinoma | 0.0007 | $1.00 \times 10^4$ | 36.00 | 1.0001 | 6566.2 | 0.009 |
| Testicular germ cell cancer | 0.0037 | $7.20 \times 10^4$ | 5.80 | 1.0369 | 117.3 | 0.798 |
| Thyroid papillary/follicular carcinoma | 0.0103 | $6.50 \times 10^4$ | 0.09 | 1.7560 | 314.5 | 5.9 |
| Thyroid medullary carcinoma | 0.0003 | $6.50 \times 10^4$ | 0.09 | 1.2387 | 1143.0 | 58.95 |

The estimation of the fixation times and times before the first mutation appears. After determining the fitness parameter $r$, we can now estimate the cancer initiation times, which in our theoretical framework are the same as the fixation times. Because the number of stem cells is very large, it can be shown from Eqs. (7) and (12) that

$$T_{i} = \frac{1}{r(1 - r^{-\gamma})} \left[ \frac{1}{b} - \frac{R_{tr}}{b R_{tr} t_{life} Q_{pr} N} \right] + \frac{2 \mu Q_{pr} N}{r(1 - r^{-\gamma}) R_{tr}} + \ln \left( \frac{R_{tr}}{b R_{tr} t_{life} Q_{pr}} \right) T_{life} \tag{13}$$

In calculating the fixation times, cancer lifetime risks $R_{tr}$, number of stem cells $N$ and cell division rates $b$ were taken from the recently assembled data. The probability for activating a single oncogene was estimated to be $3 \times 10^{-8}$ and multiplying it by the average number of oncogenes $\sim 10$ we obtain $\mu \approx 3 \times 10^{-8}$ as was estimated before. Much less information is known about the probability of cancer progression ($Q_{pr}$). It has been argued theoretically and supported by some medical data that not all lesions progress to full cancers, and we estimate $Q_{pr} \approx 0.001$ using the data for breast cancers. One can reasonably assume that $Q_{pr}$ for all cancers are similar or of the same order of magnitude. Of course, this is a first attempt approximation, and we employ it only because of...
the lack of available information. Using our theoretical framework, one can also estimate the time before the first mutation appears, \( t_0 \). It can be shown that it is given by

\[
t_0 = \frac{T_{\text{lifetime}}}{NbT_{\text{mut}}/\mu} = \frac{1}{N\mu}.
\]  

(14)

This formula can be understood by noting that \( NbT_{\text{lifetime}} \) gives the total number of driver mutations during the lifetime, and dividing the lifetime by this number gives the average time between mutations.

The results of our calculations for the fixation times and for the times before the first mutation appears are presented in Table 1. Our theoretical method suggests that the fixation times vary strongly for different types of cancer. Slightly smaller variations are predicted for \( t_0 \). However, one should be cautioned to take these numbers literally because they are sensitive to absolute values of \( \mu \) and \( Q_{pr} \), and we used the same values of \( \mu \) and \( Q_{pr} \) for all cancers just to illustrate our method, which is also not realistic. Nevertheless, we believe that this quantitative analysis provides a reasonable description of trends in the cancer initiation dynamics because the same calculations for different sets of the parameters \( \mu \) and \( Q_{pr} \) produced no correlations: see the SI. In addition, for all different tumors the cancer initiation is associated with the mutation fixation, i.e., with the same biological process. For this reason, most cancers probably might be viewed similar to a some degree with respect to the initiation mechanism, and these arguments strongly support our theoretical analysis.

**Correlation between cancer lifetime risks and cancer initiation times.** The cancer lifetime risks are widely utilized for predicting the chances of getting the cancer. It is also frequently implicitly assumed that the higher the risk, the faster cancer will develop. However, the relations between cancer lifetime risks and cancer initiation times have not been thoroughly investigated. Our theoretical method allows us to measure the correlations between these quantities because both properties can be explicitly evaluated.

Figure 3 shows the fixation times, estimated using our method, and experimental data on cancer lifetime risks for 28 different types of cancer. Statistical analysis of the results gives a Spearman’s correlation coefficient \( -0.2 \) between the cancer lifetime risks and the fixation times, magnitude of which is significantly smaller than the value \( -1 \) expected for the perfect correlation. To test the validity of the null hypothesis that there is no correlation, we also performed a p-value analysis of these data. A large p-value of \( p = 0.31 \) suggests that the null hypothesis of no correlations cannot be rejected for the given set of data. Similar results are obtained for other sets of parameters \( \mu \) and \( Q_{pr} \), as shown in the SI. Of course, the conclusion from the p-value analysis might change if more data will be available. But at this moment, our analysis suggests that for cancer lifetime risks and fixation times there is no correlation between these quantities, or it is too weak to be observed in existing data. This is a very important result because it indicates that cancer lifetime risks alone cannot be utilized to evaluate the danger of getting cancer. Cancer initiation times should also be utilized, and we provide the quantitative framework how to estimate them.

**Discussion**

We developed a simple mathematical approach to evaluate the cancer initiation dynamics. The appearance of tumor is associated with fixation of some random mutation in the originally healthy tissue with fixed number of stem cells. The initial stage of cancer development is viewed as a stochastic process of transitions between discrete states with different numbers of mutated cells, and the first-passage analysis is utilized for calculating exactly the fixation probabilities and the fixation times.

It is shown that the cancer initiation dynamics depends strongly on the fitness parameter \( r \) that describes how faster the mutated cell divides in comparison with the normal cell. The effect of the number of cells \( N \) in the tissue is much weaker. It is found that for large fitness parameters the probability of fixation is high, as expected. However, the dependence of fixation times is non-monotonic with the maximum for neutral mutations \( (r = 1) \). What is surprising that even for disadvantageous mutations \( (r < 1) \) the fixation might start quite quickly. We are
able to explain these observations by utilizing arguments that view the fixation process as a random walk on the sequence of states with different degrees of mutations. Neutral mutations correspond to the unbiased random walk, which is slow because many states are repeatedly visited during the process. For disadvantageous mutations, which can be viewed as a biased random walk in the direction opposite to the fixation, the probability of fixation is small. Then only those rare events reach the fixation that are fast enough so that the opposing bias does not have time to act. But we also argued that this situation is probably not relevant for cancer dynamics due to low probability of this to occur.

We applied our theoretical approach for evaluating explicitly the initiation times for different types of cancer. This is done by connecting theoretically calculated fixation probabilities with available clinical data on cancer lifetime risks, from which the fitness parameters are estimated. This allows us to calculate exactly the fixation times that are associated with the starting point of the cancer. The initiation times are determined for 28 different types of cancer. We performed then the analysis of correlations between cancer lifetime risks and the cancer starting times. In contrast to expectations, it is found that there is no correlations between these properties of cancer initiation dynamics, assuming that our theoretical method correctly predicts the starting times for cancer. This has an important consequence suggesting that both dynamic features, lifetime risks and initiation times, are required to comprehensively evaluate the risks of getting cancer.

While our theoretical method cannot provide molecular details to explain the observation on the lack of correlations, we can give the following microscopic arguments using the analogy with thermodynamics and kinetics of chemical processes. It is known that although thermodynamics gives the probability for the chemical reaction to occur, only kinetics determines if the reaction actually takes place on experimentally observable time scales. Thermodynamic probability is proportional to an equilibrium constant for the process, which is the ratio of forward over the backward reaction rates. At the same time, kinetics is determined by the slowest transition rates. In our theoretical language, the fixation probability depends on the product of ratios of forward to backward transition rates between all sequential discrete states (see Fig. 1), which is equal to the fitness parameter $r$. However, the fixation times depend more on the slowest forward transition rates, which are $q_{N1}$ (from the state 1 to the state 2) and $q_{N−1}$ (from the state $N − 1$ to the state $N$). The transition from the state 1 is slow because there is only 1 mutated cell in the tissue. The transition from the state $N − 1$ to the fixation is slow because only one normal cell left and the probability that it will be picked out for removal is very low. Thus, the slow speed of initial and final transitions during the mutation fixation process might be the reason for general lack of correlations between the fixation probabilities and fixation times.

It is also important to critically evaluate our theoretical method since it involves several approximations and assumptions. We assume that after a single mutation appears in the tissue no other mutations can occur in the system. This is probably a reasonable assumption because the probability of activating the oncogenes $μ$ is typically very low and normal cell division rates are typically fast. If, however, the second mutation appears before the fixation of the first one, it is expected that the overall fixation time will be longer because the normal cells will be eliminated faster. Multiple studies also suggest that more than one mutation in tumor-suppressing genes (“hits”) is required for cancer to start. In this case, it is expected that while the fixation times will be longer, other qualitative trends should remain the same as discussed here. Each individual mutation has the potential to drive the cancer process. In this case, it can be argued that the fixation times depend on the probability of mutation during the replication $μ$ and the probability of cancer progression $Q_p$. Because both of these parameters are not well determined and depend on the type of the cancer, we varied them by several orders of magnitude (see details in the SI). It is found that the magnitude of initiation times are quite sensitive to variations in $μ$ and $Q_p$.

In addition, it can be argued that the cancer might start when a large fraction of cells in the tissue, but not all of them, is mutated. Furthermore, the fitness parameter might increase with number of cell replications, and this is also not accounted in our model. Both these effects will shorten the cancer initiation times. Our model also does not take into account spatial effects. But we notice that our theoretical framework can be adapted to evaluate the cancer initiation dynamics for some of these more advanced cases. One more difficulty in applying our theoretical method to real medical data is the fact that tumors are frequently diagnosed at late stages of cancer, long after the initiations stages that we investigate. However, improving of diagnostics method in future should be able to identify better the cancer initiation steps.

All these critical comments suggest that it is probably unreasonable to view all the cancer initiation times reported in Table 1 as realistic. This might also explain too large values for some fixation times. However, the trends predicted by our theoretical method should be valid because all data are considered in the similar way, and the universal process of oncogene mutation fixation is behind the formation of tumors in all cancers. In addition, our theoretical approach gives a convenient, simple and versatile method to evaluate the cancer initiation dynamics, and it is expected that in the future better estimates of relevant parameters will make the evaluation of cancer initiation times more precise and reliable. The proposed theoretical method is also useful in designing more quantified approaches in cancer prevention. For example, it can be argued that $T_0 + T_1$ might be a better time estimate of the age at which the testing of different cancers should start. Furthermore, our theoretical framework is flexible enough to be extended and generalized to include more complex biochemical and biophysical processes that are taking place during the cancer development.

Received: 17 June 2019; Accepted: 26 November 2019;
Published online: 12 December 2019

References

1. Hanahan, D. & Weinberg, R. A. The hallmarks of cancer. cell 100, 57–70 (2000).
2. Nowak, M. A. Evolutionary dynamics (Harvard University Press, 2006).
3. Frank, S. A. Dynamics of cancer: incidence, inheritance, and evolution (Princeton University Press, 2007).
23. Pon, J. R. & Marra, M. A. Driver and passenger mutations in cancer.

22. Kimura, M. Average time until fixation of a mutant allele in a finite population under continued mutation pressure: Studies by

20. Komarova, N. L. Spatial stochastic models for cancer initiation and progression.

19. Gentry, S. N. & Jackson, T. L. A mathematical model of cancer stem cell driven tumor initiation: implications of niche size and loss

18. Stransky, B. & Desouza, S. Modeling tumor evolutionary dynamics.

17. Beerenwinkel, N. "Cancer evolution: the mathematics of cancer progression and therapy resistance.”

16. Bleyer, A., Viny, A. & Barr, R. Cancer in 15-to 29-year-olds by primary site. "The oncologist" 11, 590–601 (2006).

15. Reed, K. B. et al. Increasing incidence of melanoma among young adults: an epidemiological study in olmsted county, minnesota. In Mayo Clinic Proceedings, vol. 87, 328–334 (Elsevier, 2012).

14. Tomasetti, C., Li, L. & Vogelstein, B. Stem cell divisions, somatic mutations, cancer etiology, and cancer prevention. "Science" 355, 1330–1334 (2017).

13. Tomasetti, C. & Vogelstein, B. Variation in cancer risk among tissues can be explained by the number of stem cell divisions. "Science" 347, 78–81 (2015).

12. Dubal, P. M. et al. Laryngeal chondrosarcoma: a population-based analysis. "The Laryngoscope" 124, 1877–1881 (2014).

11. Bota, S. et al. Follow-up of bronchial precancerous lesions and carcinoma in situ using fluorescence endoscopy. "Am. journal respiratory critical care medicine" 164, 1688–1693 (2001).

10. Tong, W. W. et al. Progression to and spontaneous regression of high-grade anal squamous intraepithelial lesions in hiv-infected and uninfectedmen. "Aids" 27, 2233–2243 (2013).

9. Lytle, N. K., Barber, A. G. & Reya, T. Stem cell fate in cancer growth, progression and therapy resistance. "Nat. Rev. Cancer" 1 (2001).

8. McFarland, C. D., Mirny, L. A. & Korolev, K. S. Tug-of-war between driver and passengermutations in cancer and other adaptive processes. Proc. Natl. Acad. Sci. 111, 15138–15143 (2014).

7. Altrock, P. M., Liu, L. L. & Michor, F. The mathematics of cancer: integrating quantitative models. "Nat. Rev. Cancer" 15, 730 (2015).

6. Redner, S. "A guide to first-passage processes" (Cambridge University Press, 2015).

5. Michor, F., Iwasa, Y. & Nowak, M. A. Dynamics of cancer progression. "Nat. reviews cancer" 4, 197 (2004).

4. Dominik, W. & Natalia, K. "Dynamics of cancer: mathematical foundations of oncology" (World Scientific, 2014).

3. Breyer, A., Viny, A. & Barr, R. Cancer in 15-to 29-year-olds by primary site. "The oncologist" 11, 590–601 (2006).

2. Kimura, M. Average time until fixation of a mutant allele in a finite population under continued mutation pressure: Studies by analytical, numerical, and semi‐sampling methods. "Proc. Natl. Acad. Sci." 77, 522–526 (1980).

1. Tomasetti, C. & Vogelstein, B. Variation in cancer risk among tissues can be explained by the number of stem cell divisions. "Science" 347, 78–81 (2015).

Acknowledgements

We thank O. Igoshin, K. Vanaja, D. Ostapenko, X. Li and F. Bocci for useful discussions. 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).

Author contributions

A.K. designed the research, H.T. and M.P.K. performed the research. H.T. and A.K. wrote the manuscript. All authors reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41598-019-55300-w.

Correspondence and requests for materials should be addressed to A.B.K.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2019