Mutations Beget More Mutations—Rapid Evolution of Mutation Rate in Response to the Risk of Runaway Accumulation

5 authors, including:

Yongsen Ruan
Sun Yat-Sen University
2 PUBLICATIONS   **2** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

mutation rate evolution View project

Mutations Beget More Mutations—Rapid Evolution of Mutation Rate in Response to the Risk of Runaway Accumulation

Yongsen Ruan [ORCID],[1] Haiyu Wang,[1] Bingjie Chen [ORCID],[1] Haijun Wen,*,[1] and Chung-I Wu*,[1,2,3]

[1]State Key Laboratory of Biocontrol, School of Life Sciences, Sun Yat-Sen University, Guangzhou, China

[2]CAS Key Laboratory of Genomic and Precision Medicine, Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing, China

[3]Department of Ecology and Evolution, University of Chicago, Chicago, IL

*Corresponding authors: E-mails: wenhj5@mail.sysu.edu.cn; ciwu@uchicago.edu.

Associate editor: Yuseob Kim

Abstract

The rapidity with which the mutation rate evolves could greatly impact evolutionary patterns. Nevertheless, most studies simply assume a constant rate in the time scale of interest (Kimura 1983; Drake 1991; Kumar 2005; Li 2007; Lynch 2010). In contrast, recent studies of somatic mutations suggest that the mutation rate may vary by several orders of magnitude within a lifetime (Kandoth et al. 2013; Lawrence et al. 2013). To resolve the discrepancy, we now propose a runaway model, applicable to both the germline and soma, whereby mutator mutations form a positive-feedback loop. In this loop, any mutator mutation would increase the rate of acquiring the next mutator, thus triggering a runaway escalation in mutation rate. The process can be initiated more readily if there are many weak mutators than a few strong ones. Interestingly, even a small increase in the mutation rate at birth could trigger the runaway process, resulting in unfit progeny. In slowly reproducing species, the need to minimize the risk of this uncontrolled accumulation would thus favor setting the mutation rate low. In comparison, species that starts and ends reproduction sooner do not face the risk and may set the baseline mutation rate higher. The mutation rate would evolve in response to the risk of runaway mutation, in particular, when the generation time changes. A rapidly evolving mutation rate may shed new lights on many evolutionary phenomena (Elango et al. 2006; Thomas et al. 2010, 2018; Langergraber et al. 2012; Besenbacher et al. 2019).

Key words: mutation rate, mutation load, mutator, runaway model, cancer risk.

Introduction

Mutational process is fundamental to understanding evolution itself and the accuracy in estimating the mutation rate is crucial for molecular evolutionary studies. A recent example is the divergence between human and chimpanzee, which shows a 2-fold discrepancy in the parent–offspring data vis-a-vis the molecular evolutionary analyses (Scally and Durbin 2012; Besenbacher et al. 2019). It appears that the mutation rate itself has evolved substantially between the two closely related species with only 1.2% difference in their DNA sequences (Waterson et al. 2005). The molecular clock is far more erratic than postulated. The literature on phylogeny, adaptive evolution and population differentiation usually makes the assumption that the mutation rate is constant in the timeframe of interest (Kimura 1983; Drake 1991; Kumar 2005; Li 2007; Lynch 2010). How valid will these conclusions be if the mutation rate itself is rapidly evolving?

Because the mutation rate is too small to measure, the conventional approach is to substitute the evolutionary rate over a long period for the per generation estimate. Recently, for direct measurement, the parent–offspring comparisons have begun to appear in publications (Venn et al. 2014; Besenbacher et al. 2019). The bulk of such data, however, exist in the form of somatic mutation accumulation in an individual's lifetime. Such data have been collected for studying cancers (Kandoth et al. 2013; Lawrence et al. 2014) as exemplified by the Cancer Genome Atlas (TCGA; Cancer Genome Atlas 2012; Cancer Genome Atlas Research et al. 2013).

Germline mutations in many ways resemble somatic mutations that accrued in defined tissues although the fitness consequences of germline mutations are much greater. Nevertheless, many of the physiological and molecular mechanisms should be common across tissues and some general principles likely exist. In the germline, for example, mutation occurrence is usually treated as a time-homogeneous Poisson process. This oversimplified assumption seems an inadequate foundation of theories for determining the mutation rate (Kumar 2005; Lynch 2010; Scally and Durbin 2012; Sung et al. 2012; Wielgoss et al. 2013; Segurel et al. 2014; Lynch et al. 2016), or the evolution of sex (Charlesworth et al. 2005). Fortunately, the assumption of the simple Poisson process can be easily verified in somatic tissues (see below).

Mutation accumulation over a lifetime is likely a time-inhomogeneous process (Alexandrov et al. 2015; Milholland et al. 2015; Podolskiy et al. 2016; Zhang et al. 2019).

In particular, since some mutations are themselves mutators that increase the mutation rate, (mutator) mutations should beget more mutations (mutators or otherwise), thus leading to the runaway accumulation. Although genes responsible for DNA repair are likely mutators, evidence suggests that a much larger number of mutations may increase the mutation rate in a subtle manner (Hung et al. 2019). For example, any mutation that can influence the chromatin structure or the ion concentration can be a mutator if it can impair DNA polymerase and/or repair enzymes (Schuster-Bockler and Lehner 2012; El Meouche and Dunlop 2018). Unless held back by selection, the first mutator would initiate the runaway process and eventually greatly elevate the mutation rate. It will be shown that many weak mutators will trigger the runaway more quickly than a few strong ones, with the same aggregate mutation effect (see , supplementary material 2, Supplementary Material online).

The accumulation of mutations in somatic tissues has been extensively recorded (Kandoth et al. 2013; Lawrence et al. 2013; Martincorena et al. 2015; Blokzijl et al. 2016; Ju et al. 2017; Bae et al. 2018; Lodato et al. 2018; Chen et al. 2019; Yizhak et al. 2019). Since these mutations affect only local patches of tissues, the fitness consequences are much smaller than those of germline mutations. Indeed, somatic mutations experience such weak selection that the process is characterized as "quasi-neutral" (Chen et al. 2019). Importantly, as the runaway process is operative only when it is unchecked by natural selection, somatic tissues are the likely targets. Furthermore, if the runaway mutational process is triggered, it may lead to tumorigenesis. Therefore, the end-products of runaway accumulation are different in the germline than in the soma. In the former, they are eliminated by natural selection but, in the latter, they prominently present themselves in tumors. In TCGA data, the mutation load varies by >1,000-fold, even among cases of the same cancer type (Cancer Genome Atlas Research et al. 2013; Kandoth et al. 2013; Lawrence et al. 2013). Hence, the distribution of the mutation load typically exhibits a very long tail (supplementary fig. S1, Supplementary Material online). This highly skewed distribution suggests something akin to a runaway process.

For germline mutations, the mutation rate would be limited by natural selection due to their fitness consequences. Hence, how this limit is set has been a frequent topic in the literature (Sturtevant 1937; Kimura and Maruyama 1966; Leigh 1970; Drake 1991; Lynch 2008, 2010, 2011; Thomas et al. 2010; Wielgoss et al. 2013; Segurel et al. 2014). Since the mutation rate may not stay constant in a lifetime, it might entail a low initial rate in order to reduce the likelihood of runaway accumulation (see Theoretical Background below). There are many evolutionary implications for avoiding the runaway process (see Discussion).

We should note that mutations begetting more mutations refer specifically to "mutator mutations." The novelty of the present hypothesis is that mutators are mechanistically connected to further mutations. It is distinct from other hypotheses; for example, deleterious mutations have been suggested to render subsequent mutations less fit (Agrawal 2002; Shaw and Baer 2011; see supplementary material 2, Supplementary Material online). The key here is positive feedbacks among mutator mutations.

Results

The Runaway Mutation Accumulation in Somatic Tissues

Although a runaway accumulation of mutations is plausible, it has not been observed presumably because it is too deleterious to the organisms. It is hence expected that natural selection should have stopped the process in the germline. In this regard, somatic mutations would be more revealing as cells that experience runaway accumulation may not die but, instead, transform into tumors.

Figure 1a shows the mutation load in the coding regions of normal and cancerous colorectal tissues. In the normal tissues, the mutation load is tightly clustered ~40 per exome, whereas the distribution is skewed to the right with a median and mean load of 148 and 503, respectively, in cancerous tissues. The long tail with the load >5,000 is shown in the inset. Note that the mutation numbers in the least loaded 3% of tumor cases fall in the range of the normal tissue. As a result, the highest and lowest numbers among cancer cases differ by 300-fold (~15,000/50). Such a long tail is observed in most cancer types (Kandoth et al. 2013; Lawrence et al. 2013). We should note that the "tail" is in fact much thinner than the visual impression conveyed in figure 1a because cancer patients account for only a small fraction of the population. For example, colorectal cancer patients account for only 2% of the older segment (age > 50) of the general population (https://seer.cancer.gov/). Hence, the load distribution of the entire population would require the red bars be lowered to 2% of the height portrayed in figure 1a.

Source of Mutators in Cancer

If the long tail of the mutation load is caused by mutator mutations, we ask if the mutations in the 177 known DNA repair genes (DRGs) might be the cause (see Materials and Methods; Wood et al. 2001, 2005). In the TCGA data, 4,044 of the 9,979 cancer samples do not carry mutations in these genes. Figure 1b separates cases that do or do not have such mutations in red or blue, respectively. Cases with DRG mutations do have a higher mutation load. But the converse is also true: cases with a higher mutation load should have more DRG mutations. We hence randomly choose 207 genes with the same exon length as the 177 DNA repair genes. Indeed, the distribution of nonsynonymous mutations in the randomly chosen genes is the same as the distribution in the DRGs (fig. 1b vs. fig. 1d). Hence, the highly loaded cases have exactly the expected number of DRGs mutations. Moreover, if mutations in DRGs really act as mutators, the mutator effect should increase the mutational load exponentially, rather than linearly (fig. 1c). This will be modeled in below section "The Model." Since the right-skewed long tail is not caused by mutations in the known DRGs, it is plausible that a large number of weak mutators may drive the high mutation rate (Schuster-Bockler and Lehner 2012; El Meouche and Dunlop 2018; Hung et al. 2019).
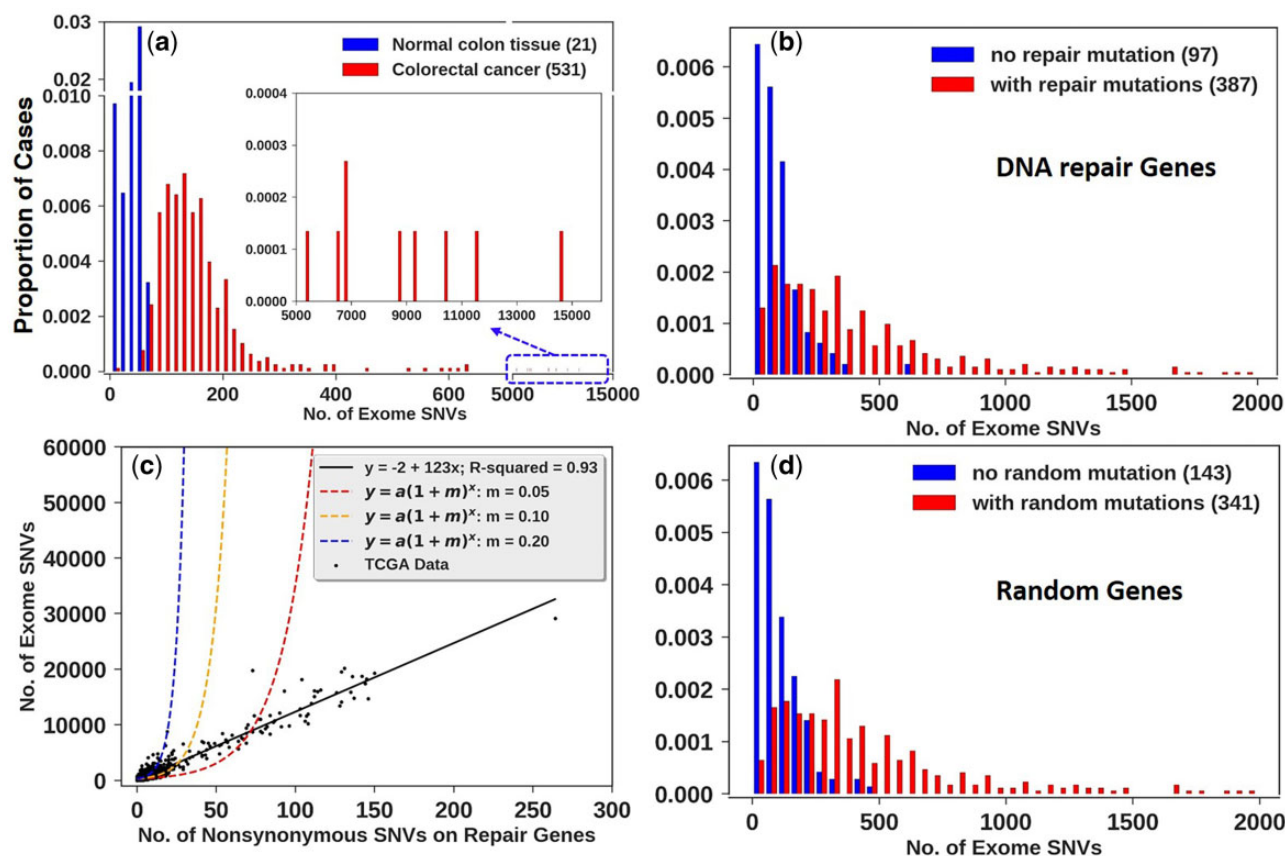
FIG. 1. Mutation load distribution. (a) Distributions of mutational load in the normal colon tissue (blue; Blokzijl et al. 2016) and colorectal cancer (red; Cancer Genome Atlas 2012; Cancer Genome Atlas Research et al. 2013). SNVs, single nucleotide variants. The number of cases is shown in the parentheses. The inset displays the tail end of the distribution. (b) Distribution of SNVS in LUAD cancer from TCGA data with and without nonsynonymous mutations in DNA repair genes. Red bars represent cases with at least one nonsynonymous mutation. The number of cases is shown in the parentheses. (c) Mutation load as a function of the number of nonsynonymous mutations in DNA repair genes. Each point represents a patient (9,979 patients in total) with the linear regression line shown. The three dashed lines are the models in which each DNA repair mutation increases the mutation rate according to the equation given (see The Model). The parameter $a = 267$ is the mean number of exome SNVs from TCGA data and m is the factor by which a nonsynonymous mutation in DNA repair genes can increase the mutation rate. There is no hint of any increase in mutation rate due to the mutations in DNA repair genes. (d) is the same as (b) but the genes are randomly chosen from the genome with their total length equal to the DNA repair genes. Note that (b) and (d) are nearly identical. They, together with (c), show that the cases with very high mutation load cannot be attributed to the mutations in DNA repair genes.

Previous analyses of the mutation load distribution often treat the accumulation as a composite of multiple mutational profiles, referred to as "mutation signatures." For example, one study identified 33 signatures and suggested five of them to be linearly correlated with the age (Alexandrov et al. 2015). These clock-like mutations account for ∼20% of the total, suggesting that 80% of them accumulate in a time-inhomogeneous manner. Others suggest an exponential increase in the mutation load as a function of age (Milholland et al. 2015; Podolskiy et al. 2016; Zhang et al. 2019). Furthermore, the proportion of clock-like mutations is larger in normal tissues, including early embryos, than in tumors (Blokzijl et al. 2016; Ju et al. 2017). These analyses point to the evolution of the mutation rate itself that increases with time. The model in the next section frames such time-dependent mutation rates in mechanistic terms.

The Model

Theoretical Background

In modeling the mutation rate, we track the evolution of one single cell, instead of a population of cells because the cancer genomic data of TCGA are effectively about the genome of the single progenitor cell that proliferates to become a full-blown tumor (Armitage and Doll 1954; Fearon and Vogelstein 1990; Wu et al. 2016). In doing so, the simulation would appear to ignore the effect of selection, which is addressed below.

First, we shall use the Ka/Ks ratio to represent the strength of selection, where Ka is the number of nonsynonymous substitutions per site and Ks is synonymous substitutions per site (Li et al. 1985). If selection on nonsynonymous mutations is largely absent, it is generally accepted that Ka/Ks would be ∼1 (Sharp et al. 1995; Keightley et al. 2005, but see Lu and Wu 2005). For germline mutations, the Ka/Ks ratio ranges between 0.05 and 0.3 from a wide range of taxa

(vertebrates, insects, nematodes, and plants), suggesting 70–95% of nonsynonymous mutations are removed by natural selection. In contrast, the Ka/Ks ratio ranges between 0.9 and 1.15 for normal and cancer tissues (Chen et al. 2019), thus suggesting somatic evolution to be nearly neutral. This dynamics has recently been characterized as quasi-neutral (Chen et al. 2019), thanks to the population structure of somatic cells (see below).

Second, the unit of somatic cell population is a single stem-cell niche (Wu et al. 2016). On the epithelia of human colons and small intestines, the population size (N) in a stem-cell niche is $N < 50$. In this range of population sizes, the efficacy of selection, both positive and negative, is low. A salient consequence is the canceling out of positive and negative selection that gives rise to quasi-neutrality. With quasi-neutrality, it would be feasible to simulate the evolution of a population by tracking a single cell lineage (see Supplementary Material online for detail). This practice is generally not justifiable when the net selection deviates from the quasi-neutrality.

Third, a direct approach to the effect (or lack of) of selection is presented in figure 2, whereby mutators are evaluated for their fitness consequences. We use Ka/Ks of each cancer sample to represent the influence of selection (the Y axis) and the mutation load to represent the mutator effect (the X axis). In the pan-cancer data (fig. 2a) of nearly 10,000 cases, R^2 is 0.0026. Because the Ka/Ks value is highly variable when the mutation load is small, we also remove cases with <100 coding mutations (fig. 2b) or < 300 coding mutations (fig. 2c). When the data are thus parsed, the correlation in fact becomes even lower, with $R^2 = 0.0011$ and 0.0034, respectively. In short, the mutator activity appears to be uncorrelated with the strength of selection.

Fourth, there is an additional reason to track one single cell lineage in the germline, where the fitness effect of the mutation impacts the progeny of the next generation rather than the survival of the germ cell bearing it (in Drosophila, it is known that sperm without chromosomes are functional; Muller 1928). Thus, the mutations would have little fitness consequences in the germ cells themselves.

The Basic Features
In the runaway model, we posit an initial mutation rate at birth, μ_0, referred to as the baseline mutation rate. With the increasing age, the mutation rate, $\mu(t)$, is time-dependent. A proportion (p) of mutations is assumed to be mutators that can increase $\mu(t)$ by a fraction λ. The mutation accumulation is therefore an inhomogeneous Poisson process. In parallel, we also assume a proportion (r) of mutations to be fitness-altering. In the germline, fitness-altering mutations are usually deleterious but, in the soma, mutations may often be advantageous in the form of cancer driver mutations. Fitness mutations can thus be of both kinds. The remaining fraction $[(1-p)(1-r)]$ of mutations has no phenotypic effect. The process of mutation accumulation lasts for an individual's lifetime, T, during which no recombination happens. Unless otherwise specified, T follows a truncated normal distribution ($\mu = 70$, $\sigma = 10$, lower $= 0$, upper $= 120$) according to the World Bank Data (https://data.worldbank.org/).

In figure 3, examples of mutation accumulation are displayed. All three types of mutations (neutral, fitness-altering, and mutator) are shown. Case 1 does not have a mutator mutation; hence, mutations accrue steadily and slowly. In case 2, a large effect mutator leads to a large jump in mutation rate and a quick succession of fitness mutations. In case 3, several small-effect mutators lead to the gradual acceleration of mutation accumulation. In cases 2 and 3, the waiting time between new mutations becomes shorter and shorter, resulting in the accumulation of many fitness-altering mutations.

Mutator Mutations
Because mutators in the cancer genomic data are not associated with the "strong-effect" DNA repair genes, it seems plausible that mutations affecting chromatin structure, DNA polymerase fidelity, ion concentration in the nucleus, nuclear membrane permeability and numerous other factors may all have a mutator effect, albeit a weak one. In addition, the mutators could be SNVs (single nucleotide variants), CNVs (copy number variations), epigenetic changes, or even the expression level of genes (Schuster-Bockler and Lehner 2012; El Meouche and Dunlop 2018; Hung et al. 2019). Given that the mutation rate in the human germline at $\sim 10^{-9}$/bp/year is very low, a 50% increase in the error rate should be considered weak (in Escherichia coli, strong effect mutators can alter the mutation rate by three orders of magnitude; Loh et al. 2010). Nevertheless, if the error rate is multiplicative, ten weak mutators, each having a 50% effect, would collectively enhance the error rate by 58-fold.

In the model, the baseline mutation rate (i.e., the initial mutation rate at birth) is μ_0, which is set at 0.33 mutations per exome per year in the soma according to the mutational load of normal tissues (Blokzijl et al. 2016). The mutation rate changes when a new mutator mutation occurs. (For convenience, we use μ' and μ and drop the "t"). Thus,

$$\mu' = \lambda\mu \, , \tag{1}$$

where λ is the effect of a new mutator. We assume $ln(\lambda) = x$ follows the exponential distribution, $1/me^{-x/m}$. The mean of the mutator effect is hence $\bar{\lambda} = e^m$ (see Materials and Methods). If $m = 0$, the mutator effect λ is always equal to 1 and $\mu(t) = \mu_0$ for the lifetime.

Fitness-Altering Mutations
We now denote the number of fitness-altering mutations by K. The accumulation of these mutations will lead to a phenotypic state, for example, cancer in the soma or lethality in the progeny (via germline mutations). In the germline, mutations affecting fitness are dominated by deleterious variants (Kimura 1983; Fay et al. 2001; Eyre-Walker and Keightley 2007; Li 2007), whereas nonneutral mutations are often advantageous in the soma, resulting in the overproliferation of cells (Fearon and Vogelstein 1990; Hanahan and Weinberg 2011). Due to the relatively weak influence of negative selection in the soma as presented in "Theoretical Background," the runaway process
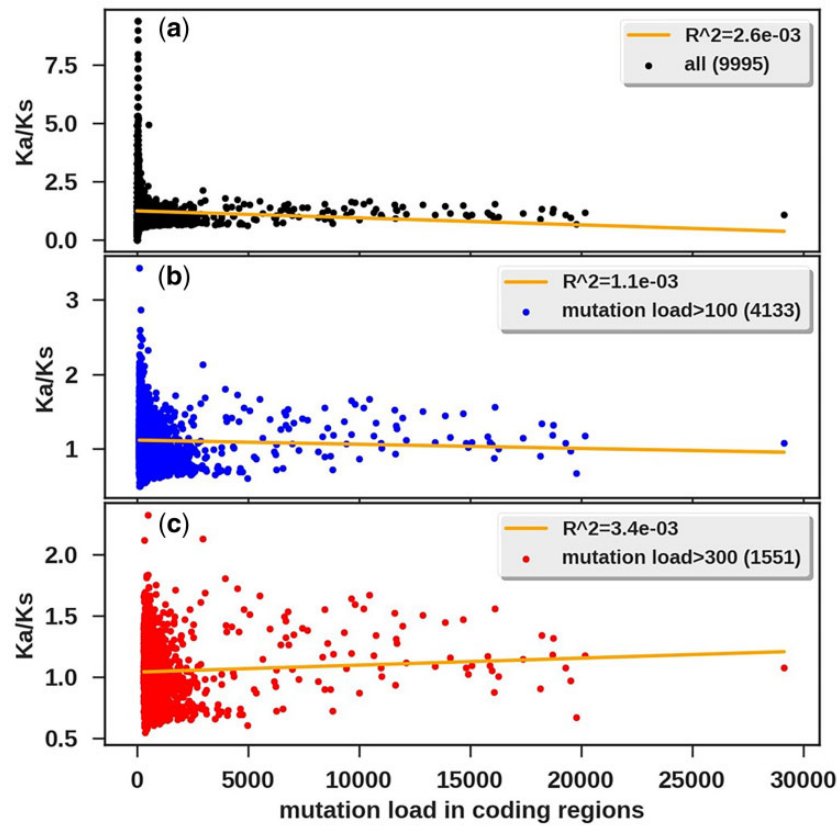
FIG. 2. The low correlation between the strength of selection (Ka/Ks) and the mutator activity (mutation load). (*a*) *X* axis is the mutation load in the coding regions of each cancer case from the TCGA data. *Y* axis is the Ka/Ks ratio, representing the strength of selection. The orange line is the linear regression line and R^2, as shown, is very low. In short, cases with a higher mutation load do not experience stronger or weaker selective pressure. The number of cases is shown in the parentheses. In (*b*) and (*c*), only cases with >100 and >300 mutations are used.
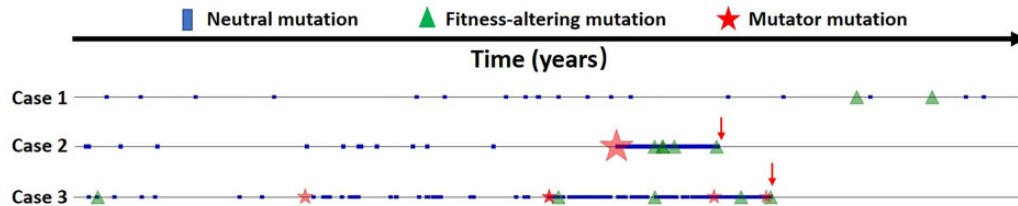

FIG. 3. Three examples of mutation accumulation. Three types of mutations are considered: neutral (blue box), fitness-altering (green triangle), and mutators (red star, the size proportional to effect size). Fitness-altering mutations in somatic tissues can lead to cancer when the number reaches a threshold (usually 5). The process lasts for an individual's lifetime or until reaching the cancerous state.

is prominent in cancer samples. We shall present the analysis of somatic mutations first.

(1) Fitness-altering mutations in the soma leading to tumorigenesis—a mutation is assumed to affect fitness with probability *r*. Based on the analysis of 299 cancer driver genes (Bailey et al. 2018), we set *r* at 0.04 (see Materials and Methods; Wu et al. 2016). The number of genetic events needed to start tumorigenesis has been estimated to be five to ten (Fearon and Vogelstein 1990; Hanahan and Weinberg 2011). Mutations underlying these events are often referred to as "cancer drivers" which, by definition, are advantageous mutations. The cancer risk (R_0) is the probability of developing cancer in one's lifetime. Hence, in the deterministic runaway model,

$$P(\text{Cancer}|K) = \begin{cases} 0, & K < 5 \\ 1, & K \geq 5 \end{cases}.$$

In the probabilistic runaway model, the potential function is:

$$P(\text{Cancer}|K) = \begin{cases} 0, & K < 5 \\ 1, & K \geq 10 \\ \dfrac{K-4}{12}, & 5 \leq K < 10 \end{cases}.$$

Beyond driver and neutral (passenger) mutations, the existence of deleterious mutations in tumors has been controversial (Wu et al. 2016; Martincorena et al. 2017;

5

Zapata et al. 2018). Such mutations presumably could slow down cancer evolution and should be relevant only when data on noncancerous clonal expansions are included (Chen B, Ruan Y, Wu CI, unpublished results).

(2) Fitness-altering mutations in the germline—the description will be given in the section Application of the Model to Germline Evolution: Lower μ_0 Retarding the Runaway.

Approximate Analytical Solutions

Although we rely on extensive simulations, approximate analytical solutions have also been obtained as summarized below (see Supplementary Material online for details):

The expected number of mutations at age t is given by:

$$\lambda(t) = -\frac{\ln(1 - mp\mu_0 t)}{mp} \ . \tag{2}$$

The probability of an individual with x mutations at age t is as follows:

$$P\left(n_k^*(t) = x\right) = e^{-\lambda(t)} \frac{\lambda(t)^x}{x!} = e^{\frac{\ln(1-mp\mu_0 t)}{mp}} \frac{\left(-\frac{\ln(1-mp\mu_0 t)}{mp}\right)^x}{x!}. \tag{3}$$

Similarly, we can obtain the probability of having x mutator mutations as follows:

$$P(n_k(t) = x) = e^{\frac{\ln(1-mp\mu_0 t)}{m}} \frac{\left(-\frac{\ln(1-mp\mu_0 t)}{m}\right)^x}{x!}. \tag{4}$$

The probability of having x fitness-altering mutations is as follows:

$$P(K(t) = x) = e^{r\frac{\ln(1-mp\mu_0 t)}{mp}} \frac{\left(-r\frac{\ln(1-mp\mu_0 t)}{mp}\right)^x}{x!}. \tag{5}$$

In the deterministic runaway model, the risk at age t is given by:

$$R_0(t) = 1 - \sum_{x=0}^{K_m-1} e^{r\frac{\ln(1-mp\mu_0 t)}{mp}} \frac{\left(-r\frac{\ln(1-mp\mu_0 t)}{mp}\right)^x}{x!}. \tag{6}$$

The comparisons between simulations and the analytical solutions are given in supplementary figures S3 and S4, Supplementary Material online. The disparity increases with the mutator strength due to the approximation based on the mean mutator effect (see Supplementary Material online).

Application of the Model to Tumorigenesis: Many Weak Mutators Causing the Runaway

To trigger the runaway process, the mutator strength ($\bar{\lambda}$) and the initial baseline mutation rate (μ_0) are both crucial. In this section, we analyze the mutator strength (strong but infrequent vs. weak but common mutators) by varying p and $\bar{\lambda}$ but keeping $p \times \bar{\lambda}$ constant (at 0.12). The results are shown in figure 4 for the deterministic model. The

probabilistic model yields qualitatively similar results as shown in supplementary figure S2, Supplementary Material online.

In figure 4, the lifetime cancer risk (R_0) and mutation load both increase as the mutator strength increases (fig. 4a–c). However, above a certain level, the risk starts to decrease (fig. 4c–e). Because $p \times \bar{\lambda}$ is kept constant, the total mutator activity is expected to be the same, except in figure 4a, where the mean mutator effect is 0 (i.e., $\bar{\lambda}=1$). Nevertheless, if $\bar{\lambda}$ is small, the mutation load does not build up in the limited lifetime to yield a high R_0 as shown in figure 4b. When $\bar{\lambda}$ is very large (=500, fig. 4e), the total mutation load is the same as in simulations with lower $\bar{\lambda}$ (fig. 4c and d). However, with such strong mutators, the mutation distribution is bimodal with cases of either very low or very high mutation load (fig. 4e). Since only cases with very high mutation load are at risk of cancer, R_0 is low with strong mutators. Overall, many weak mutators are much more compatible with the observed distributions of mutation load than a few strong ones (fig. 4a–e vs. fig. 1a).

Although the model predicts a long tail of high mutation load, the maximal number of mutations is still smaller than 500 (fig. 4b–e), whereas the observations of figure 1a show a much longer and thinner tail. In the framework of the runaway model, a very long tail is possible, provided that there is time for mutations to accrue. For example, there may exist a time gap (G, also known as incubation period or latency time; Nadler and Zurbenko 2014) between the emergence of the single progenitor cancer cell and its subsequent expansion. Even with a small time gap ($G = 1$ year), the maximum number of mutations in tumors can reach nearly 4,000 (fig. 4f). The distribution with $G = 1$ indeed agrees well with the observations on colorectal cancer (fig. 4f vs. fig. 1a).

Application of the Model to Germline Evolution: Lower μ_0 Retarding the Runaway

We now transition from analyzing somatic mutations to predicting the germline mutation rate. Such a transition requires some modifications of the model of the section Model and Simulation. Unlike in the soma, all germline mutations in the gamete are present in every cell of the progeny. For this reason, deleterious mutations in the germline would have a much greater fitness consequence than somatic mutations.

Modifications of the Basic Model for Germline Evolution

Organisms should be able to regulate, at least to some extent, their germline mutation rate, which is expected to be lower than that in the soma. The question is "how much lower." It is natural to ask how the baseline mutation rate, μ_0, would affect the runaway process in the germline. After all, a higher μ_0 is equivalent to the presence of some mutators. Although the mutators may also evolve in terms of both p and λ, the two quantities are not empirically measurable. As μ_0 remains the only measurable quantity at present, we keep p and λ constant so we can compare different μ_0 values, when given the same mutator activity (i.e., p and λ).
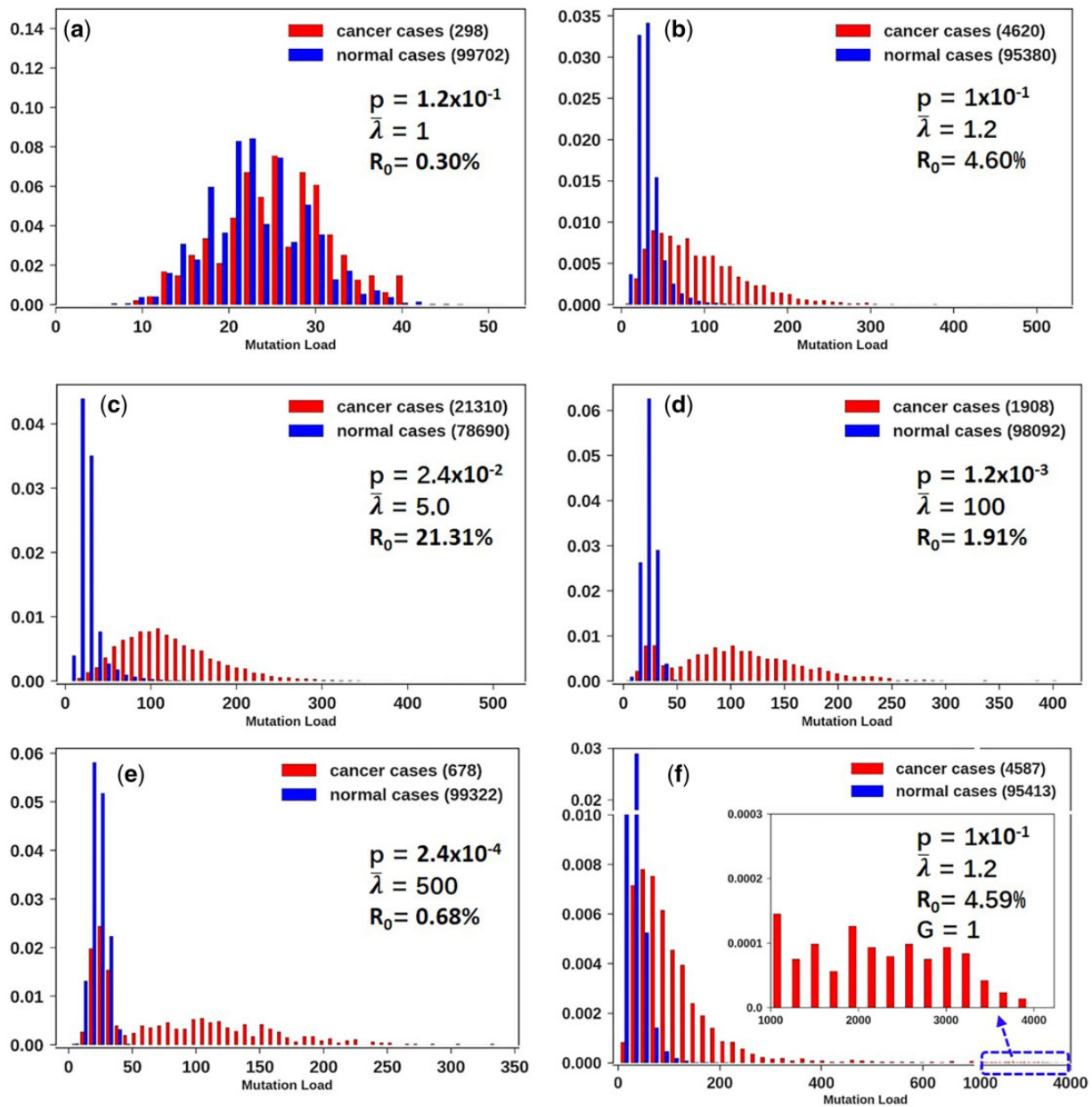
FIG. 4. Mutation load distribution across mutator effect sizes. (a–e) Normalized histogram of mutation load with $p \times \bar{\lambda} = 0.12$, where p is the probability of mutator mutations and $\bar{\lambda}$ is the mean mutator strength. (a–e): the mutator strength continues to increase but the mutators become less prevalent (see the main text). R_0 represents the lifetime cancer risk, that is, the probability of becoming cancerous by accumulating at least five cancer drivers. For each parameter set, we simulate 100,000 cases. (f) The same as panel (b) except for an additional parameter of $G = 1$ (year), which is the latency time between the acquisition of the five driver mutations and the exponential expansion of the cancer cell population. Recent evidence has increasingly suggested the existence of this latency time due to intense local competitions among cell clones (Martincorena et al. 2018; Yizhak et al. 2019; Zhu et al. 2019). During this time gap, mutation accumulation may also accelerate, thus giving rise to the strong right-skewed distribution.

(1) Mutation rate—this rate of germline mutation in humans, referred to as μ^*, is estimated to be ~1 de novo mutations per year per gamete, corresponding to about 0.4×10^{-9}/bp/year (Scally and Durbin 2012; Jonsson et al. 2017). In this section, we are interested in the coding regions and set $\mu^* = 0.02$ per year. We then ask about the consequence of mutation accumulation if μ_0 is larger than μ^*, with $\mu_0 = 2\mu^*$, $5\mu^*$, or $10\mu^*$, in comparison with $\mu_0 = \mu^*$. The average

strength of the mutator, $\bar{\lambda}$, is assumed to be either 2 or 5. Given that $\mu^* = 0.4 \times 10^{-9}$/bp/year is a very small number to begin with, a 2- or 5-fold increase would still yield a very small mutation rate.

(2) The fitness consequences of germline mutations—Fay et al. (2001) estimated that strongly selected mutations (f_2, in their notation) account for 1–5% of all coding mutations on the basis of human polymorphism data (Fay et al. 2001; Eyre-Walker et al. 2006). We thus

assume that nonneutral mutations occur among 1% ($r = 0.01$) or 5% ($r = 0.05$) of the coding mutations and these mutations would cause lethality in a combination of two or more. With the assumed lethality, the fate of the mutator is determined in one generation. Importantly, the results presented below do not depend strongly on the assumptions about fitness interactions. Other parameter sets yield qualitatively similar results (see Supplementary Material online). Nevertheless, if the mutations are not assumed to cause lethality, the fitness effect would be manifested in several generations, depending on how recombination dissociates the mutators from the fitness mutations. The tracking would require much more efforts but the qualitative conclusion is not altered.

We wish to know the fitness reduction associated with a germline mutation, M, that increases μ_0 to $2\mu^*$, $5\mu^*$, or $10\mu^*$. As defined above, a gamete has a probability of acquiring lethal mutations when the runaway process gets started. We shall refer to this probability as the "fitness risk" (see fig. 5). Since the gamete is otherwise normal, P would be the fitness reduction associated with M. Given the effective population size of humans in the range of $\sim 5 \times 10^3$ (Charlesworth 2009; Prado-Martinez et al. 2013), $s \sim (1/2N_e) \sim 10^{-4}$ should be effectively neutral. We thus consider a fitness reduction of 0.005–0.01 caused by M to be medium to high.

Predictions of the Runaway Model on Germline Mutation Rate

Results from two schemes of mutator effect are presented. In scheme A with $\bar{\lambda} = 2$ and $r = 0.05$, fitness-altering mutations are common but mutator strength is moderate (fig. 5a and supplementary table S1, Supplementary Material online). In this scheme and in the absence of mutators, the risk stays low until age 40 if $\mu_0 = 2\mu^*$. Nevertheless, because the fitness mutations are common, the risk would be moderate at age 25 if $\mu_0 = 5\mu^*$. With mutators, if $\mu_0 = 2\mu^*$, the risk would be moderate at age 30 and high at age 40. If $\mu_0 = 5\mu^*$, then the risk is already high at age 18.

In scheme B, with $\bar{\lambda} = 5$ and $r = 0.01$, the fitness-altering mutations are less common but the mutators are strong (fig. 5b and supplementary table S2, Supplementary Material online). We shall now consider the fitness risk if there are no mutators. First, according to simulations, the fitness risk would be imperceptible for all parameter values (fig. 5b and supplementary table S2, Supplementary Material online). Second, without mutators, the calculated fitness risk would remain $< 10^{-3}$ even if $\mu_0 = 10\mu^*$ (fig. 5c, see Materials and Methods). In contrast, if we consider mutators with the associated risk of runaway process, even $\mu_0 = \mu^*$ would incur a moderate risk of 0.005 at age 25. If $\mu_0 = 2\mu^*$, the risk is moderate at age 13 (probably before sexual maturity) and high at age 19. With strong mutators and $\mu_0 = 5\mu^*$, the fitness risk would be exceedingly high at age 10.

The trend is most discernible if we focus on the risk at age 25, assumed to be the mean reproductive age of humans

(fig. 5c). With mutators (the top two lines), a 2-fold increase in the baseline mutation rate would make the risk rather high at age 25. A 5-fold increase to $\mu_0 = 0.1$ would make the risk of fitness reduction unacceptable. On the other hand, without the runaway process, the baseline mutation could be 2-, or even 5-fold, higher than its current level, regardless of the proportion of lethal mutations. The lowest line shows that, given $r = 0.01$, μ_0 can easily be 10-fold higher than it is now if there is no threat of runaway mutation. In conclusion, mutators could set a limit on the baseline mutation rate, μ_0, that has to be sufficiently low to avoid the runaway.

The threat of runaway accumulation is much reduced if reproduction happens much earlier. For example, at age 5, the cost of letting μ_0 increase by even 10-fold would be modest (fig. 5d). This age-dependency could explain why the mutation rate per year in mice is ten times higher than in humans (Lindsay et al. 2018). Furthermore, there is a negative correlation between generation time and per year mutation rate. Comparing figure 5d with figure 5c, one may make a case that the reduced risk of runaway mutation accumulation may permit the baseline mutation rate to rise high in taxa with shorter generation time. In contrast, long-living species like humans may have set μ_0 low to avoid mutation rates increasing at an accelerating pace. Indeed, in >1,500 trios in Iceland, where the parental age rarely exceeds 50, no runaway accumulation was observed (Jonsson et al. 2017).

Discussion

Mutation is unique among evolutionary forces as mutations themselves form a positive feedback loop. With mutator mutations begetting more mutations, the process can evolve very quickly, even within a lifetime. Indeed, somatic tissues with a high mutation load often evolve into tumors (Loeb 1991; Hanahan and Weinberg 2011; Cancer Genome Atlas 2012; Cancer Genome Atlas Research et al. 2013). The source of mutators is crucial for the runaway process. An interesting example is a panel of 66 DNA polymerase I mutants in *E. coli* that yield replication fidelities spanning three orders of magnitude (Loh et al. 2010). Such strong mutators, however, are not found to drive high mutation rates as shown in figure 1, thus suggesting the source of mutator mutations to be rather diverse.

Until now, mutators driving a runaway process have not yet been documented for germline evolution. We thus reevaluate the assumption of an unvarying mutation rate across time. In the evolutionary time span, the postulate of a constant mutation rate driving a molecular clock has been rejected (Wu and Li 1985; Kumar and Subramanian 2002; Kumar 2005; Elango et al. 2006; Moorjani et al. 2016). Furthermore, by tracking the change of the ratio through time (being the mutation rate in males vs. that in females), Makova and Li (2002) report the ratio in the extant hominids to be much smaller than in the earlier period of primate evolution. More recently, the mutation rate in extant humans, observed by the parents–offspring trio sequencing, is found to be only 2/3 as large as those of other apes—chimpanzee, gorilla, and orangutan
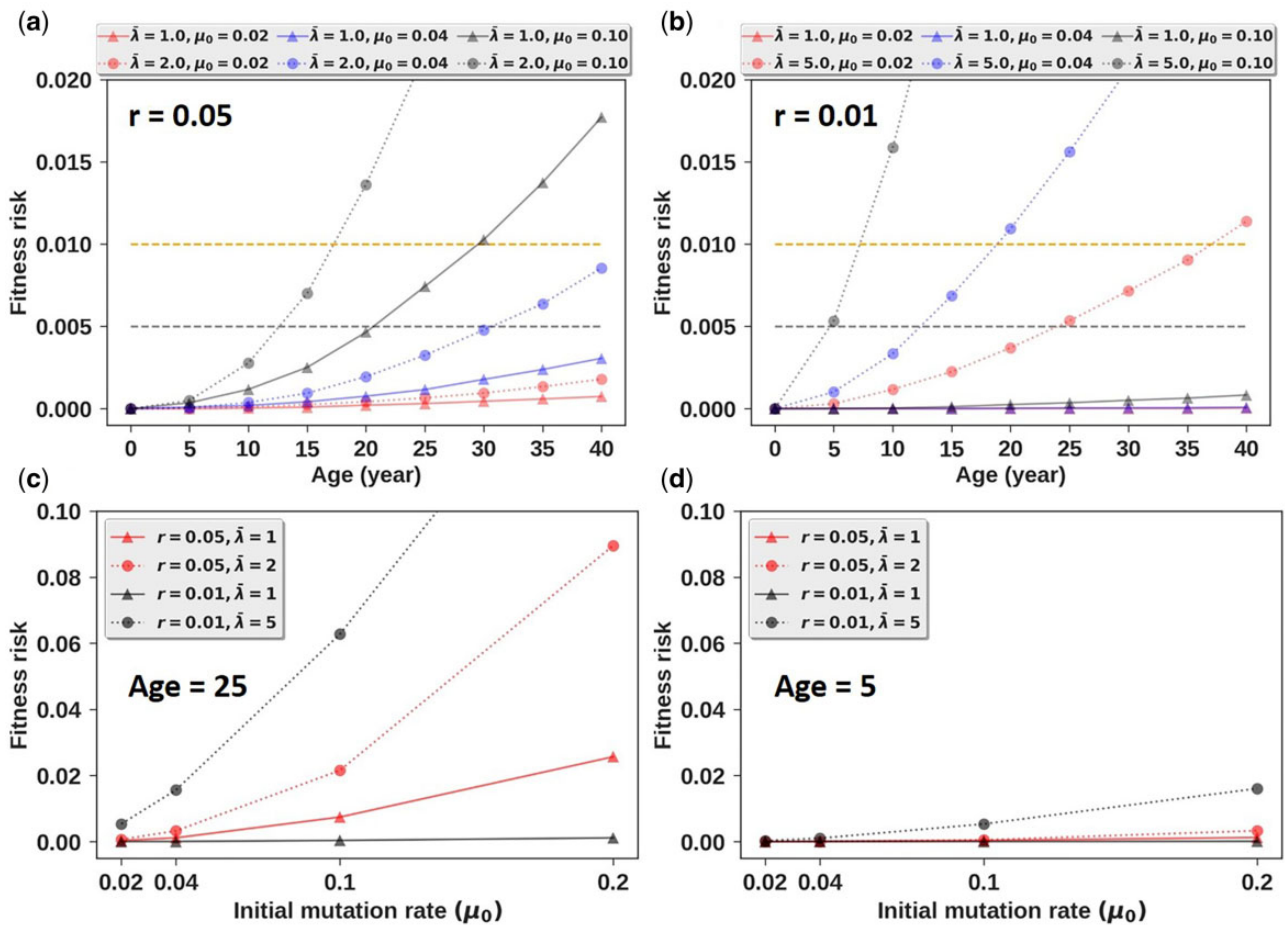
FIG. 5. Probability of fitness reduction (or risk) as a function of the germline mutation rate at age 0, μ_0. This value in extant humans is estimated to be 0.02. We use $\mu_0 = 0.02, 0.04, 0.10, 0.20$ in the simulations. $p = 0.1$ for all panels. (a) Fitness reduction due to runaway mutation accumulation under frequent fitness-altering mutations ($r = 0.05$) and weak mutators ($\bar{\lambda}=2$). (b) As in (a), but with less frequent fitness-altering variants ($r = 0.01$) and stronger mutators ($\bar{\lambda}=5$). (c and d) Fitness reduction at age 25 or 5 as a function of the baseline mutation rate. The low fitness values in these panels are given numerically in supplementary tables S1 and S2, Supplementary Material online.

(Besenbacher et al. 2019). These observations suggest that the mutation rate has been evolving rapidly.

Since the mutation rate evolves, the extant rate should reflect the past selective pressure (Lynch 2010, 2011; Scally and Durbin 2012; Sung et al. 2012; Wielgoss et al. 2013). The extant baseline mutation rate at $\mu_0 = 0.02$ may reflect the avoidance of the runaway process as a mere 2-fold increase (to $\mu_0 = 0.04$) is already at risk of runaway mutation accumulation and, hence, fitness reduction. As the potential for such accumulation should increase with age, one may seek evidence for this process in men of advanced age, say, >70. There are hints that the accumulation of mutations is closer to exponential than linear when very old paternal parents are included (Risch et al. 1987; Wyrobek et al. 2006; Kong et al. 2012; Khandwala et al. 2018). If this conclusion holds when more data from old age paternal parents become available, it would mean that humans' reproductive tissues may be at the risk of runaway mutation accumulation. The risk has been averted simply because humans did not live that long in much of the duration of our existence. In this context, advanced paternal age would have theoretical, medical, and social–cultural significance.

The avoidance of runaway mutation accumulation may also explain why the average age of reproduction has such an impact on the baseline mutation rate. In general, short-living animals have a much higher baseline rate (Welch et al. 2008; Thomas et al. 2010; Wilson Sayres et al. 2011; Wang et al. 2019). For example, the per year mutation rate in mice is ten times higher than in humans (Lindsay et al. 2018). Even among hominoids, the slight delay in the age of reproduction ($\sim$ age 29) may have contributed to the much lower baseline rate in humans than in other apes (age 19–25) (Langergraber et al. 2012). In figure 4f, it is shown that one extra year can extend the tail length many-fold. As pointed out (Besenbacher et al. 2019), this strong generation-time dependence may explain why the mutation rate obtained by trio sequencing in humans is only half as large as that inferred from the evolutionary analysis between human and other apes. Given their longer generation time, modern humans appear to have a lower baseline mutation rate than their ancestors and relatives.

Although our calculation shows that the extant rate of mutation in the human germline, at $\mu_0 = 0.02$, may be at the cusp of incurring runaway, there are alternative

hypotheses. For example, this rate could also be dictated by the drift barrier (Lynch 2008, 2010, 2011). In the drift-barrier hypothesis, the fitness reduction (Δw; w stands for fitness) associated with $\mu_0 = 0.02$ would be $\Delta w \sim 1/(2Ne)$, where Ne is the effective population size. In this view, $2Ne \, \Delta w \sim 1$ is expected. In the Supplementary Material online, applying the published approach of surveying human polymorphisms (Fay et al. 2001; Eyre-Walker et al. 2006) to the 1,000 human genomes (Genomes Project et al. 2015), we find the strength of selection against deleterious mutations, as is commonly assumed, only shows $2Ne \, \Delta w \ll 1$. If this calculation stands, the current mutation rate in humans may be lower than the drift barrier and the avoidance of runaway may fill in the gap.

The process of mutation is another example, in addition to genetic drift (Chen et al. 2017), natural selection (Chen et al. 2019), and sex chromosome evolution (Xu et al. 2017) that cancer evolution studies can enrich the general theories of biological evolution (Wu et al. 2016; Wen et al. 2018). This study shows that the evolution of the mutation rate could be a highly dynamic process due to the positive feedback of mutators on mutations, as revealed by cancer genomic studies (Cancer Genome Atlas 2012; Cancer Genome Atlas Research et al. 2013; Kandoth et al. 2013; Lawrence et al. 2014). Depending on how rapidly the mutation rate has been evolving, the conclusions in many molecular evolutionary studies will have to be re-evaluated if they rely heavily on the constancy of the mutation rate over an extended evolutionary time.

Materials and Methods

Data Handling

Estimating the Proportions of DNA Repair Genes and Cancer Driver Genes in the Genome

In the runaway model, p is the proportion of mutator mutations and r is the proportion of fitness-altering mutations. Both have to be estimated. We first use the proportion of DNA repair genes as a lower bound on p. For fitness mutations, we use the cancer driver mutations compiled from the literature.

The 177 DNA repair genes and 299 cancer driver genes were obtained from the public databases as well as published papers (Wood et al. 2001, 2005; Lawrence et al. 2014; Bailey et al. 2018). We use the software GTFtools (Li 2018) to calculate nonoverlapping exon length for these genes (see supplementary tables S5 and S6, Supplementary Material online). The total lengths of exons of the 177 DNA repair genes and the 299 cancer driver genes are 1,026,236 and 2,372,955 bp, respectively. The exome of the human genome consists of roughly 180,000 exons constituting ~1% of the total genome (Ng et al. 2009). Thus, the exome proportion of DNA repair genes is $1,026,326/(3.2 \times 10^7) = 0.032$ and $2,372,955/(3.2 \times 10^7) = 0.074$ for cancer driver genes. Considering that the ratio of nonsynonymous mutations over synonymous mutations is ~2:1, the proportion of function-changing sites in DNA repair genes is 0.02 and 0.04 in cancer driver genes.

Extraction of SNVs from Whole Exome (or Genome) Data from Cancerous and Normal Tissues

We obtained exome mutation data called by Mutect2 and the associated clinical information from TCGA. There are 10,429 patients across 33 cancer types. For each case, we only care about single nucleotide variants (SNVs) in exons and discard all other mutations. We divide the SNVs into synonymous ("5′-UTR," "3′-UTR," "Silent") and nonsynonymous mutations ("Missense Mutation," "Nonsense Mutation," "Nonstop Mutation," "Translation Start Site," "Splice Site," "Splice Region"). We then counted the numbers of nonsynonymous and synonymous mutations in the 177 DNA repair and 299 cancer driver genes in each patient. To compare with the contribution of DNA repair mutations to the mutation load, we randomly chose 207 genes with the same length as the 177 DNA repair genes and likewise counted the nonsynonymous and synonymous mutations.

We obtained filtered vcf files (whole-genome sequencing data) corresponding to 45 normal tissue samples (21 colons, 14 small intestines, and 10 livers) from Blokzijl et al. (2016). We used ANNOVAR (Yang and Wang 2015) for variant annotation. We processed these data similarly to the TCGA data to obtain nonsynonymous and synonymous mutation counts in DNA repair, cancer driver, and random genes.

Model and Simulation

The structure of the runaway mutation accumulation model is presented in the main text and the approximate analytical solutions are given in Supplementary Material online. We describe the simulation procedures below.

Simulation Steps

(1) Initialization. Determine the lifetime (T) of an individual. Set initial age $t = 0$, initial mutation rate $\mu = \mu_0$, the number of mutator mutations $n_k = 0$, the number of mutations $n_k^* = 0$, the number of fitness-altering mutations $K = 0$.

(2) Determine the waiting time Δt to a new mutation by sampling from the exponential distribution with mean $1/\mu$.

(3) The new mutations are mutators with probability p, and fitness-altering with probability r. p and r are mutually exclusive. If the new mutation is fitness-altering, go to step 4. If the new mutation is a mutator, go to step 5.

(4) Fitness mutation: Update fitness-altering mutation number $K = K + 1$.

(5) Mutator mutation: Sample a value k from the exponential distribution with mean m $\left(\frac{1}{m}e^{-\frac{x}{m}}\right)$ (see Mutator Strength below). Update mutation rate $\mu = \mu \times e^k$. Update mutator mutation number $n_k = n_k + 1$.

(6) Update $t = t + \Delta t$, $n_k^* = n_k^* + 1$. If we are simulating the mutation accumulation in somatic tissues (i.e., the tumorigenesis process), go to step 7. If in the germline, go to step 8.

(7) According to the tumor onset function (see main text), determine whether the individual forms a tumor. If a tumor is formed or $t > T$, stop simulation (Note: if there

is a latency time G, the cancer case can also accumulate mutations during latency). Otherwise, return to step 2.

(8) If $K > 1$ (synthetic lethal) or $t > T$, stop simulation. Otherwise, return to step 2.

For each parameter space, we simulate 100,000 cases unless otherwise specified.

Mutator Strength

In the main text, the effect of a new mutator on the mutation rate is as follows:

$$\mu' = \lambda\mu.$$

Most of mutator mutations only increase mutation rate a little, so we assume that $x = ln(\lambda)$ follows an exponential distribution with mean m. λ thus follows the log-exponential distribution. A good estimation for the mutator strength is the mean λ. Here, we just obtain the mean of the log-exponential variable (λ) to estimate the average effect of a new mutator.

$$P(\lambda \leq k) = P(e^x \leq k) = P(x \leq \ln k) = \int_0^{\ln k} \frac{1}{m} e^{-x/m} dx$$
$$= 1 - k^{-1/m} \quad (k \geq 1).$$

That is, the cumulative density function (CDF) of λ is as follows:

$$F(\lambda) = 1 - \lambda^{-1/m} \quad (\lambda \geq 1).$$

Taking the derivative gives the density of λ,

$$f(\lambda) = \frac{dF(\lambda)}{d\lambda} = \frac{1}{m} \lambda^{-1/m-1} \quad (\lambda \geq 1).$$

Then the mean of λ is given by:

$$\bar{\lambda} = \int_1^\infty \lambda f_\Lambda(\lambda) d\lambda = \int_1^\infty \frac{1}{m} \lambda^{-1/m} d\lambda.$$

When $0 < m < 1$,

$$\bar{\lambda} = \int_1^\infty \frac{1}{m} \lambda^{-1/m} d\lambda = \frac{\frac{1}{m} \lambda^{1-\frac{1}{m}}}{1 - \frac{1}{m}} \Big|_{\lambda=1}^{\infty} = \frac{1}{1-m}.$$

When $m = 1$,

$$\bar{\lambda} = \int_1^\infty \lambda^{-1} d\lambda = \ln(\lambda) \Big|_{\lambda=1}^{\infty} = \infty.$$

When $m > 1$,

$$\bar{\lambda} = \int_1^\infty \frac{1}{m} \lambda^{-1/m} d\lambda = \frac{\frac{1}{m} \lambda^{1-\frac{1}{m}}}{1 - \frac{1}{m}} \Big|_{\lambda=1}^{\infty} = \infty.$$

The mean of λ goes to infinity when $m \geq 1$. However, the mutator effect, λ, of a new mutator cannot be infinite in our case. In addition, when m is small,

$$\frac{1}{1-m} \approx e^m.$$

Thus, we just let the mean effect of a mutator approximately be:

$$\bar{\lambda} = e^m \tag{7}$$

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Acknowledgments

References

Agrawal AF. 2002. Genetic loads under fitness-dependent mutation rates. *J Evol Biol.* 15(6):1004–1010.

Alexandrov LB, Jones PH, Wedge DC, Sale JE, Campbell PJ, Nik-Zainal S, Stratton MR. 2015. Clock-like mutational processes in human somatic cells. *Nat Genet.* 47(12):1402–1407.

Armitage P, Doll R. 1954. The age distribution of cancer and a multi-stage theory of carcinogenesis. *Br J Cancer.* 8(1):1–12.

Bae T, Tomasini L, Mariani J, Zhou B, Roychowdhury T, Franjic D, Pletikos M, Pattni R, Chen BJ, Venturini E, et al. 2018. Different mutational rates and mechanisms in human cells at pregastrulation and neurogenesis. *Science* 359(6375):550–555.

Bailey MH, Tokheim C, Porta-Pardo E, Sengupta S, Bertrand D, Weerasinghe A, Colaprico A, Wendl MC, Kim J, Reardon B, et al. 2018. Comprehensive characterization of cancer driver genes and mutations. *Cell* 173(2):371–385 e318.

Besenbacher S, Hvilsom C, Marques-Bonet T, Mailund T, Schierup MH. 2019. Direct estimation of mutations in great apes reconciles phylogenetic dating. *Nat Ecol Evol.* 3(2):286–292.

Blokzijl F, de Ligt J, Jager M, Sasselli V, Roerink S, Sasaki N, Huch M, Boymans S, Kuijk E, Prins P, et al. 2016. Tissue-specific mutation accumulation in human adult stem cells during life. *Nature* 538(7624):260–264.

Cancer Genome Atlas N. 2012. Comprehensive molecular characterization of human colon and rectal cancer. *Nature* 487:330–337.

Cancer Genome Atlas Research N, Weinstein JN, Collisson EA, Mills GB, Shaw KR, Ozenberger BA, Ellrott K, Shmulevich I, Sander C, Stuart JM. 2013. The Cancer Genome Atlas Pan-cancer analysis project. *Nat Genet.* 45:1113–1120.

Charlesworth B. 2009. Fundamental concepts in genetics: effective population size and patterns of molecular evolution and variation. *Nat Rev Genet.* 10(3):195–205.

Charlesworth D, Charlesworth B, Marais G. 2005. Steps in the evolution of heteromorphic sex chromosomes. *Heredity* 95(2):118–128.

Chen B, Shi Z, Chen Q, Shen X, Shibata D, Wen H, Wu CI. 2019. Tumorigenesis as the paradigm of quasi-neutral molecular evolution. *Mol Biol Evol.* 36(7):1430–1441.

Chen Y, Tong D, Wu CI. 2017. A new formulation of random genetic drift and its application to the evolution of cell populations. *Mol Biol Evol.* 34(8):2057–2064.

Drake JW. 1991. A constant rate of spontaneous mutation in DNA-based microbes. *Proc Natl Acad Sci U S A.* 88(16):7160–7164.

El Meouche I, Dunlop MJ. 2018. Heterogeneity in efflux pump expression predisposes antibiotic-resistant cells to mutation. *Science* 362(6415):686–690.

Elango N, Thomas JW, Program NCS, Yi SV. 2006. Variable molecular clocks in hominoids. *Proc Natl Acad Sci U S A.* 103(5):1370–1375.

Eyre-Walker A, Keightley PD. 2007. The distribution of fitness effects of new mutations. *Nat Rev Genet.* 8(8):610–618.

Eyre-Walker A, Woolfit M, Phelps T. 2006. The distribution of fitness effects of new deleterious amino acid mutations in humans. *Genetics* 173(2):891–900.

Fay JC, Wyckoff GJ, Wu CI. 2001. Positive and negative selection on the human genome. *Genetics* 158(3):1227–1234.

Fearon ER, Vogelstein B. 1990. A genetic model for colorectal tumorigenesis. *Cell* 61(5):759–767.

Genomes Project C, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA, et al. 2015. A global reference for human genetic variation. *Nature* 526(7571):68–74.

Hanahan D, Weinberg RA. 2011. Hallmarks of cancer: the next generation. *Cell* 144(5):646–674.

Hung S, Saiakhova A, Faber ZJ, Bartels CF, Neu D, Bayles I, Ojo E, Hong ES, Pontius WD, Morton AR, et al. 2019. Mismatch repair-signature mutations activate gene enhancers across human colorectal cancer epigenomes. *Elife* 8: e40760.

Jonsson H, Sulem P, Kehr B, Kristmundsdottir S, Zink F, Hjartarson E, Hardarson MT, Hjorleifsson KE, Eggertsson HP, Gudjonsson SA, et al. 2017. Parental influence on human germline de novo mutations in 1,548 trios from Iceland. *Nature* 549:519–522.

Ju YS, Martincorena I, Gerstung M, Petljak M, Alexandrov LB, Rahbari R, Wedge DC, Davies HR, Ramakrishna M, Fullam A, et al. 2017. Somatic mutations reveal asymmetric cellular dynamics in the early human embryo. *Nature* 543(7647):714–718.

Kandoth C, McLellan MD, Vandin F, Ye K, Niu B, Lu C, Xie M, Zhang Q, McMichael JF, Wyczalkowski MA, et al. 2013. Mutational landscape and significance across 12 major cancer types. *Nature* 502(7471):333–339.

Keightley PD, Lercher MJ, Eyre-Walker A. 2005. Evidence for widespread degradation of gene control regions in hominid genomes. *PLoS Biol.* 3(2):e42.

Khandwala YS, Baker VL, Shaw GM, Stevenson DK, Lu Y, Eisenberg ML. 2018. Association of paternal age with perinatal outcomes between 2007 and 2016 in the United States: population based cohort study. *BMJ* 363:k4372.

Kimura M. 1983. The neutral theory of molecular evolution. Cambridge: Cambridge University Press.

Kimura M, Maruyama T. 1966. The mutational load with epistatic gene interactions in fitness. *Genetics* 54(6):1337–1351.

Kong A, Frigge ML, Masson G, Besenbacher S, Sulem P, Magnusson G, Gudjonsson SA, Sigurdsson A, Jonasdottir A, Jonasdottir A, et al. 2012. Rate of de novo mutations and the importance of father's age to disease risk. *Nature* 488(7412):471–475.

Kumar S. 2005. Molecular clocks: four decades of evolution. *Nat Rev Genet.* 6(8):654–662.

Kumar S, Subramanian S. 2002. Mutation rates in mammalian genomes. *Proc Natl Acad Sci U S A.* 99(2):803–808.

Langergraber KE, Prufer K, Rowney C, Boesch C, Crockford C, Fawcett K, Inoue E, Inoue-Muruyama M, Mitani JC, Muller MN, et al. 2012. Generation times in wild chimpanzees and gorillas suggest earlier divergence times in great ape and human evolution. *Proc Natl Acad Sci U S A.* 109(39):15716–15721.

Lawrence MS, Stojanov P, Mermel CH, Robinson JT, Garraway LA, Golub TR, Meyerson M, Gabriel SB, Lander ES, Getz G. 2014. Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature* 505(7484):495–501.

Lawrence MS, Stojanov P, Polak P, Kryukov GV, Cibulskis K, Sivachenko A, Carter SL, Stewart C, Mermel CH, Roberts SA, et al. 2013. Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* 499(7457):214–218.

Leigh EG. 1970. Natural selection and mutability. *Am Nat.* 104:301–305.

Li H. 2018. GTFtools: a Python package for analyzing various modes of gene models. bioRxiv:263517.

Li W-H. 2007. Molecular evolution. New York; Basingstoke: W.H. Freeman; Palgrave [distributor].

Li WH, Wu CI, Luo CC. 1985. A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. *Mol Biol Evol.* 2(2):150–174.

Lindsay SJ, Rahbari R, Kaplanis J, Keane T, Hurles ME. 2018. Striking differences in patterns of germline mutation between mice and humans. bioRxiv: 082297.

Lodato MA, Rodin RE, Bohrson CL, Coulter ME, Barton AR, Kwon M, Sherman MA, Vitzthum CM, Luquette LJ, Yandava CN, et al. 2018. Aging and neurodegeneration are associated with increased mutations in single human neurons. *Science* 359(6375):555–559.

Loeb LA. 1991. Mutator phenotype may be required for multistage carcinogenesis. *Cancer Res.* 51(12):3075–3079.

Loh E, Salk JJ, Loeb LA. 2010. Optimization of DNA polymerase mutation rates during bacterial evolution. *Proc Natl Acad Sci U S A.* 107(3):1154–1159.

Lu J, Wu CI. 2005. Weak selection revealed by the whole-genome comparison of the X chromosome and autosomes of human and chimpanzee. *Proc Natl Acad Sci U S A.* 102(11):4063–4067.

Lynch M. 2008. The cellular, developmental and population-genetic determinants of mutation-rate evolution. *Genetics* 180(2):933–943.

Lynch M. 2010. Evolution of the mutation rate. *Trends Genet.* 26(8):345–352.

Lynch M. 2011. The lower bound to the evolution of mutation rates. *Genome Biol Evol.* 3:1107–1118.

Lynch M, Ackerman MS, Gout JF, Long H, Sung W, Thomas WK, Foster PL. 2016. Genetic drift, selection and the evolution of the mutation rate. *Nat Rev Genet.* 17(11):704–714.

Makova KD, Li WH. 2002. Strong male-driven evolution of DNA sequences in humans and apes. *Nature* 416(6881):624–626.

Martincorena I, Fowler JC, Wabik A, Lawson ARJ, Abascal F, Hall M, Cagan A, Murai K, Mahbubani K, Stratton MR, et al. 2018. Somatic mutant clones colonize the human esophagus with age. *Science* 362(6417):911–917.

Martincorena I, Raine KM, Gerstung M, Dawson KJ, Haase K, Van Loo P, Davies H, Stratton MR, Campbell PJ. 2017. Universal patterns of selection in cancer and somatic tissues. *Cell* 171(5):1029–1041.e1021.

Martincorena I, Roshan A, Gerstung M, Ellis P, Van Loo P, McLaren S, Wedge DC, Fullam A, Alexandrov LB, Tubio JM, et al. 2015. Tumor evolution. High burden and pervasive positive selection of somatic mutations in normal human skin. *Science* 348(6237):880–886.

Milholland B, Auton A, Suh Y, Vijg J. 2015. Age-related somatic mutations in the cancer genome. *Oncotarget* 6(28):24627–24635.

Moorjani P, Amorim CE, Arndt PF, Przeworski M. 2016. Variation in the molecular clock of primates. *Proc Natl Acad Sci U S A.* 113(38):10607–10612.

Muller HJ. 1928. The measurement of gene mutation rate in Drosophila, its high variability, and its dependence upon temperature. *Genetics* 13(4):279–357.

Nadler DL, Zurbenko IG. 2014. Estimating cancer latency times using a Weibull model. *Adv Epidemiol Adv Epidemiol.* 2014:1–8.

Ng SB, Turner EH, Robertson PD, Flygare SD, Bigham AW, Lee C, Shaffer T, Wong M, Bhattacharjee A, Eichler EE, et al. 2009. Targeted capture and massively parallel sequencing of 12 human exomes. *Nature* 461(7261):272–276.

Podolskiy DI, Lobanov AV, Kryukov GV, Gladyshev VN. 2016. Analysis of cancer genomes reveals basic features of human aging and its role in cancer development. *Nat Commun.* 7(1):12157.

Prado-Martinez J, Sudmant PH, Kidd JM, Li H, Kelley JL, Lorente-Galdos B, Veeramah KR, Woerner AE, O'Connor TD, Santpere G, et al. 2013. Great ape genetic diversity and population history. *Nature* 499(7459):471–475.

Risch N, Reich EW, Wishnick MM, McCarthy JG. 1987. Spontaneous mutation and parental age in humans. *Am J Hum Genet.* 41(2):218–248.

Scally A, Durbin R. 2012. Revising the human mutation rate: implications for understanding human evolution. *Nat Rev Genet.* 13(10):745–753.

Schuster-Bockler B, Lehner B. 2012. Chromatin organization is a major influence on regional mutation rates in human cancer cells. *Nature* 488:504–507.

Segurel L, Wyman MJ, Przeworski M. 2014. Determinants of mutation rate variation in the human germline. *Annu Rev Genom Hum Genet.* 15:47–70.

Sharp PM, Averof M, Lloyd AT, Matassi G, Peden JF. 1995. DNA sequence evolution: the sounds of silence. *Philos Trans R Soc Lond B Biol Sci.* 349(1329):241–247.

Shaw FH, Baer CF. 2011. Fitness-dependent mutation rates in finite populations. *J Evol Biol.* 24(8):1677–1684.

Sturtevant AH. 1937. Essays on evolution. I. On the effects of selection on mutation rate. *Q Rev Biol.* 12(4):464–467.

Sung W, Ackerman MS, Miller SF, Doak TG, Lynch M. 2012. Drift-barrier hypothesis and mutation-rate evolution. *Proc Natl Acad Sci U S A.* 109(45):18488–18492.

Thomas GWC, Wang RJ, Puri A, Harris RA, Raveendran M, Hughes DST, Murali SC, Williams LE, Doddapaneni H, Muzny DM, et al. 2018. Reproductive longevity predicts mutation rates in primates. *Curr Biol.* 28(19):3193–3197 e3195.

Thomas JA, Welch JJ, Lanfear R, Bromham L. 2010. A generation time effect on the rate of molecular evolution in invertebrates. *Mol Biol Evol.* 27(5):1173–1180.

Venn O, Turner I, Mathieson I, de Groot N, Bontrop R, McVean G. 2014. Nonhuman genetics. Strong male bias drives germline mutation in chimpanzees. *Science* 344(6189):1272–1275.

Wang L, Ji Y, Hu Y, Hu H, Jia X, Jiang M, Zhang X, Zhao L, Zhang Y, Jia Y, et al. 2019. The architecture of intra-organism mutation rate variation in plants. *PLoS Biol.* 17(4):e3000191.

Waterson RH, Lander ES, Wilson RK, The Chimpanzee S, Analysis C. 2005. Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature* 437:69–87.

Welch JJ, Bininda-Emonds OR, Bromham L. 2008. Correlates of substitution rate variation in mammalian protein-coding sequences. *BMC Evol Biol.* 8(1):53.

Wen H, Wang HY, He X, Wu CI. 2018. On the low reproducibility of cancer studies. *Natl Sci Rev.* 5(5):619–624.

Wielgoss S, Barrick JE, Tenaillon O, Wiser MJ, Dittmar WJ, Cruveiller S, Chane-Woon-Ming B, Medigue C, Lenski RE, Schneider D. 2013. Mutation rate dynamics in a bacterial population reflect tension between adaptation and genetic load. *Proc Natl Acad Sci U S A.* 110(1):222–227.

Wilson Sayres MA, Venditti C, Pagel M, Makova KD. 2011. Do variations in substitution rates and male mutation bias correlate with life-history traits? A study of 32 mammalian genomes. *Evolution* 65(10):2800–2815.

Wood RD, Mitchell M, Lindahl T. 2005. Human DNA repair genes, 2005. *Mutat Res.* 577(1–2):275–283.

Wood RD, Mitchell M, Sgouros J, Lindahl T. 2001. Human DNA repair genes. *Science* 291(5507):1284–1289.

Wu CI, Li WH. 1985. Evidence for higher rates of nucleotide substitution in rodents than in man. *Proc Natl Acad Sci U S A.* 82(6):1741–1745.

Wu CI, Wang HY, Ling S, Lu X. 2016. The ecology and evolution of cancer: the ultra-microevolutionary process. *Annu Rev Genet.* 50(1):347–369.

Wyrobek AJ, Eskenazi B, Young S, Arnheim N, Tiemann-Boege I, Jabs EW, Glaser RL, Pearson FS, Evenson D. 2006. Advancing age has differential effects on DNA damage, chromatin integrity, gene mutations, and aneuploidies in sperm. *Proc Natl Acad Sci U S A.* 103(25):9601–9606.

Xu J, Peng X, Chen Y, Zhang Y, Ma Q, Liang L, Carter AC, Lu X, Wu CI. 2017. Free-living human cells reconfigure their chromosomes in the evolution back to uni-cellularity. *Elife* 6:e28070.

Yang H, Wang K. 2015. Genomic variant annotation and prioritization with ANNOVAR and wANNOVAR. *Nat Protoc.* 10(10):1556–1566.

Yizhak K, Aguet F, Kim J, Hess JM, Kubler K, Grimsby J, Frazer R, Zhang H, Haradhvala NJ, Rosebrock D, et al. 2019. RNA sequence analysis reveals macroscopic somatic clonal expansion across normal tissues. *Science* 364(6444):eaaw0726.

Zapata L, Pich O, Serrano L, Kondrashov FA, Ossowski S, Schaefer MH. 2018. Negative selection in tumor genome evolution acts on essential cellular functions and the immunopeptidome. *Genome Biol.* 19(1):67.

Zhang L, Dong X, Lee M, Maslov AY, Wang T, Vijg J. 2019. Single-cell whole-genome sequencing reveals the functional landscape of somatic mutations in B lymphocytes across the human lifespan. *Proc Natl Acad Sci U S A.* 116(18):9014–9019.

Zhu M, Lu T, Jia Y, Luo X, Gopal P, Li L, Odewole M, Renteria V, Singal AG, Jang Y, et al. 2019. Somatic mutations increase hepatic clonal fitness and regeneration in chronic liver disease. *Cell* 177(3):608–621.e612.