Variation of mutational burden in healthy human tissues suggests non-random strand segregation and allows measuring somatic mutation rates

Benjamin Werner[1,*] & Andrea Sottoriva[1,*]

[1]Evolutionary Genomics & Modelling Lab, Centre for Evolution and Cancer, The Institute of Cancer Research, London, UK

[*]Correspondence should be addressed to: benjamin.werner@icr.ac.uk and andrea.sottoriva@icr.ac.uk

Abstract

The immortal strand hypothesis poses that stem cells could produce differentiated progeny while conserving the original template strand, thus avoiding accumulating somatic mutations. However, quantitating the extent of non-random DNA strand segregation in human stem cells remains difficult *in vivo*. Here we show that the change of the mean and variance of the mutational burden with age in healthy human tissues allows estimating strand segregation probabilities and somatic mutation rates. We analysed deep sequencing data from healthy human colon, small intestine, liver, skin and brain. We found highly effective non-random DNA strand segregation in all adult tissues (mean strand segregation probability: 0.98, standard error bounds (0.97,0.99)). In contrast, non-random strand segregation efficiency is reduced to 0.87 (0.78,0.88) in neural tissue during early development, suggesting stem cell pool expansions due to symmetric self-renewal. Healthy somatic mutation rates differed across

tissue types, ranging from 3.5×10^{-9}/bp/division in small intestine to 1.6×10^{-7}/bp/division in skin.

Author Summary

Cairn proposed in 1975 that upon proliferation, cells might not segregate DNA strands randomly into daughter cells, but preferentially keep the ancestral (blue print) template strand in stem cells. This mechanism would allow to drastically reduce the rate of mutation accumulation in human tissues. Testing the hypothesis in human stem cells within their natural tissue environment remains challenging. Here we show that the patterns of mutation accumulation in human tissues with age support highly effective non-random DNA strand segregation after adolescence. In contrast, during early development in infants, DNA strand segregation is less effective, likely because stem cell populations are continuing to grow.

Introduction

The immortal DNA strand hypothesis, originally proposed by Cairns in 1975, poses that adult mammalian stem cells do not segregate DNA strands randomly after proliferation [1]. Instead, stem cells might preferentially retain the parental ancestral strand, whereas the duplicated strand is passed onto differentiated cells with limited life span (Figure 1). In principle, such hierarchical tissues could produce differentiated progeny indefinitely without accumulating any proliferation-induced mutations in the stem cell compartment [2,3].

Experimental evidence supporting this hypothesis comes from BrdU stain tracing experiments both *in vitro* and *in vivo* [4-7]. Evidence from spindle orientation bias in mouse models of normal and precancerous intestinal tissue corroborated these findings, suggesting that strand segregation is then lost during tumourigenesis [8]. However, many of the experiments suffer from uncertainties in stem cell identity and a definite mechanism of strand recognition remains unknown [9]. Hence why Cairns hypothesis remains controversial [10].

Orthogonal studies based on the expected accumulation of somatic mutations in healthy human tissues have argued against the immortal strand hypothesis [11,12]. However, the mere accumulation of somatic mutations in healthy tissue neither supports nor negates the immortal strand hypothesis *in vivo*. Here, we show that measuring the change of the mutational burden and, most crucially, the change of the *variance* of the mutational burden with age allows determining the probability of DNA strand segregation and the per cell mutation rate in healthy human tissues. First, we outline the approach and then apply it to genomic data from healthy human colon, small intestine, liver, skin and brain tissue. The data comes from four recent independent studies on mutational burden in healthy tissues [13-16], which contain information on in total 39 individuals at different ages and analysed genomes of 341 single cells. We find evidence for non-random strand segregation in all adult tissues and significant differences in somatic mutation rates between tissues, but less prominent strand-segregation in brain tissue during early development.

Results

The expected change of mean and variance of mutational burden with age

We describe the accumulation of mutations with time in hierarchically organised human tissues by a stochastic mathematical and computational model, Figure 1. A detailed description and derivation of all equations is provided below (Materials and Methods). Briefly, our model considers a constant number of N stem cells that contribute to tissue homeostasis. Stem cells divide with a certain constant rate λ, e.g. once every week or month. During each division, the parental DNA strand is copied and χ novel mutations might occur on the daughter strand. Here χ is a random number that follows a Poisson distribution with mutation rate μ per bp/division and genome size L. Cell fate is also probabilistic in our model. Cells with the parental strand will keep a stem-cell fate with probability p, e.g. for $p = 1$ they will always remain stem cell, or differentiate otherwise, e.g. for $p = 1/2$ cell fate decisions are purely random (coin flip). We can understand the probability p as the probability of non-random strand segregation, e.g. $p \approx 1$ suggest highly non-random strand segregation, whereas $p = 1/2$ corresponds to random strand segregation.

With this model, we can describe the accumulation of mutations over time explicitly (see Materials and Methods for more details). Assuming the mutation rate μ as well as the cell proliferation rate λ to be constant, we find that both the mutational burden $\tilde{\mu}$ as well as the variance of the mutational burden σ^2 are predicted to increase linearly with time t:

$$\frac{1}{\lambda}\frac{\Delta\tilde{\mu}}{\Delta t} = (1-p)\mu L \quad (1)$$

$$\frac{1}{\lambda}\frac{\Delta\sigma^2}{\Delta t} = (1-p)\mu L + (\mu L)^2 p(1-p), \quad (2)$$

see Materials and Methods for a detailed derivation and Figure 2 for a verification by individual based computer simulations. However, the rates by which the mutational burden and the variance of the mutational burden increase over time depend differently on the mutation rate μ and the non-random strand segregation probability p. This allows us to independently measure the mutation rate μ and the non-random strand segregation probability p via:

$$p = \frac{\frac{\Delta\sigma^2}{\Delta\tilde{\mu}} - 1}{\frac{\Delta\sigma^2}{\Delta\tilde{\mu}} - 1 + \frac{1}{\lambda}\frac{\Delta\tilde{\mu}}{\Delta t}} \quad (3)$$

$$\mu L = \frac{\Delta\sigma^2}{\Delta\tilde{\mu}} - 1 + \frac{1}{\lambda}\frac{\Delta\tilde{\mu}}{\Delta t}. \quad (4)$$

Importantly, measuring the change in mutational burden $\frac{\Delta\tilde{\mu}}{\Delta t}$ and variance $\frac{\Delta\sigma^2}{\Delta t}$ over time in combination with equations (3) and (4) determines the mutation rate μ (per cell division) and the non-random strand segregation probability p for healthy tissues.

Measured mean and variance of mutational burden with age from sequencing data

In a recent publication Blokzijl and colleagues [13] measured mutation

accumulation in healthy colon, small intestine and liver tissue by whole genome sequencing multiple single stem cell derived organoids of healthy donors of different ages. In addition, Martincorena and colleagues [14] measured mutational burden in multiple skin samples of four individuals with ages between 58 and 73 years. Furthermore, two recent publications [15,16] performed large-scale single cell whole genome sequencing of neurons at different ages. In the experiments by Blokzijl and colleagues [1,13], they isolated single cells and expanded those into organoids. These cells can therefore be thought of as tissue specific stem cells. In contrast, the other experiments [2,3,14-16] do not directly measure mutational burden in stem but more differentiated progenitor cells. However, compared to the total number of cell divisions in the tissue, the number of divisions separating stem and progenitor cells is neglectable.

These datasets enable measurements for the change in mutational burden $\frac{\Delta\tilde{\mu}}{\Delta t}$ and the variance $\frac{\Delta\sigma^2}{\Delta t}$ of the mutational burden with age in those healthy human tissues, see Figure 3 & 4. Equations (3) and (4) have a single undetermined parameter, the stem cell proliferation rate λ. Strictly speaking, they therefore only provide possible ranges for the mutation rate and the strand segregation probability. However, the possible ranges are narrow for any biologically meaningful stem cell proliferation rate, see Figure 5.

Estimations of non-random strand segregation probability in healthy human tissues

For all tissues, the experimental observations confirm our expectation of a linearly increasing mean and variance of the mutational burden. Using linear regression on the data in [4-7,13-16], we find for colon that the change in mutational burden over time was: $\frac{\Delta\tilde{\mu}}{\Delta t} = 37.2 \pm 3.1$, for small intestine: $\frac{\Delta\tilde{\mu}}{\Delta t} = 34.6 \pm 6.9$, for liver: $\frac{\Delta\tilde{\mu}}{\Delta t} = 30.5 \pm 2.1$, for prefrontal cortex: $\frac{\Delta\tilde{\mu}}{\Delta t} = 16.2 \pm 1.1$ and for hippocampal dentate gyrus: $\frac{\Delta\tilde{\mu}}{\Delta t} = 21.8 \pm 7.9$ mutations per whole genome per year. We found for skin: $\frac{\Delta\tilde{\mu}}{\Delta t} = 1.66 \pm 0.15$ mutations per 0.69 Mb per year. We found for neurons during early development: $\frac{\Delta\tilde{\mu}}{\Delta t} = 4.2 \pm 1.3$ mutations per whole genome per day. Uncertainties here are standard errors. Similarly, for the change of variance we found for colon: $\frac{\Delta\sigma^2}{\Delta t} = 985.5 \pm 103$, for small intestine: $\frac{\Delta\sigma^2}{\Delta t} = 747.3 \pm 304$, for liver: $\frac{\Delta\sigma^2}{\Delta t} = 1564 \pm 56$, for prefrontal cortex: $\frac{\Delta\sigma^2}{\Delta t} = 7500 \pm 965$, for hippocampal dentate gyrus: $\frac{\Delta\sigma^2}{\Delta t} = 15016 \pm 6234$ mutations per whole genome per year, for skin: $\frac{\Delta\sigma^2}{\Delta t} = 5.23 \pm 0.37$ mutations per 0.69 Mb per year and for neurons during early development: $\frac{\Delta\sigma^2}{\Delta t} = 252.2 \pm 191$ mutations per whole genome per day (Figure 3 & 4).

If stem cells divide once per week this implies (Equation (3)) for the probability of DNA strand segregation in colon: $p = 0.973$ (0.971; 0.974), small intestine: $p = 0.969$ (0.966; 0.97), liver: $p = 0.988$ (0.987; 0.989), prefrontal cortex: $p = 0.999$ (0.998; 0.9993), hippocampal dentale gyrus: $p = 0.999$ (0.998; 0.9998), skin: $p = 0.985$ (0.983; 0.987). In contrast for neurons during early development we find: $p = 0.876$ (0.78; 0.88) if cells divide

every 48h. Numbers in brackets correspond to the range of the DNA strand segregation probabilities given the upper and lower bound of the error estimates of the linear regressions. Dependencies of the estimates on the proliferation rate can be found in the caption of Figure 5.

This suggests highly effective non-random DNA strand segregation in human adult stem cells and is in line with previous observations of predominantly asymmetric stem cell divisions [6-8,17,18]. It would require extreme stem cell proliferation rates of approximately one division per stem cell per year for the data to be consistent with solely random strand segregation ($p = 0.5$), Figure 5. This is an unlikely scenario as all tissues analysed here are thought to have high stem cell proliferation rates [9,19,20]. Interestingly, during development non-random DNA strand segregation is less prominent. One explanation is an expanding stem cell population due to symmetric stem cell self-renewals during early development[1,10], which also would explain the increased accumulation of mutations early, as well as typical increased telomere shortening early in life [11,12,21].

Measurements of somatic mutation rates in healthy human tissues

Based on Equation (4) we find for the *in vivo* mutation rate per base pair per cell division in colon: $\mu = 4.37\ (4.26;\ 4.46)\times10^{-9}$, small intestine: $\mu = 3.54\ (2.61;\ 4.17)\times10^{-9}$, liver: $\mu = 8.48\ (8.22;\ 8.77)\times10^{-9}$, prefrontal cortex: $\mu = 7.68\ (7.18;\ 8.12)\times10^{-8}$, hippocampal dentale gyrus: $\mu = 1.14\ (1.04;\ 1.68)\times10^{-7}$, neurons during early development: $\mu = 1.23\ (0.43;\ 1.52)\times10^{-8}$ and skin: $\mu = 1.57\ (1.54;\ 1.63)\times10^{-7}$. The ranges of

these values agree with a recent estimate of the somatic mutation rate in human fibroblasts [13-16,22] and are one to two orders of magnitude larger than germline mutation rates [13,23,24]. However, our method does not require precise estimates of the total number of cell divisions since conception (Figure 5). We find surprising differences in the somatic mutation rates across tissue types that cannot be explained by for example different stem cell proliferation rates alone. The mutation rate estimate in skin is particularly high. This might be due to the nature of the samples used by Martincorena and colleagues [14], as the mutational burden was measured in eye lids of individuals that were exposed to high levels of UV radiation for decades. It is plausible that this contributed to the very high mutation rate estimate. It remains to be seen, if these differences across tissues prevail for denser sampling in more individuals.

Explaining strand segregation in terms of symmetric stem cell divisions

Our analysis suggests in general highly effective non-random DNA strand segregation in human colon, small intestine, liver, skin and brain. However, approximately 1% to 5% of divisions in adults do not seem to segregate strands properly and stem cells accumulate additional mutations over time. The reason for this improper segregation could be either wrongly segregated strands during an asymmetric stem cell division or the loss of a stem cell by either a symmetric stem cell differentiation or cell death followed by a symmetric stem cell self-renewal. Arguments are made for both symmetric and asymmetric stem cell divisions in human tissues [15,16,25-28]. We wondered if our approach provides a mean to distinguish both possibilities. We therefore implemented stochastic simulations of mutation accumulation in either asymmetric dividing stem cell

populations with imperfect strand segregation or a stem cell population with a mix of symmetric and asymmetric divisions (SI Figure 1). Both scenarios lead to linearly increasing mean and variance of the mutational burden, with small differences in the actual rates. However, as predicted, the ratio of the variance and the mean $\sigma^2/\tilde{\mu}$ are in both scenarios independent of time and on average the same (see also Equation (S10)). Interestingly, the distribution of $\sigma^2/\tilde{\mu}$ differs. Whereas the variance of the distribution of $\sigma^2/\tilde{\mu}$ increases with time for symmetric stem cell divisions, it approximately remains constant for asymmetric stem cell divisions. However, measuring this effect reliably would require measuring the mean and variance of the mutational burden in many more independent samples of many more healthy humans of different ages than the currently available datasets. Hence, lack of resolution in currently available data precludes us to determine the cause of imperfect strand segregations. However, this effect might provide a future mean to quantitate the amount of symmetric self-renewal in human stem cell populations.

Discussion

Stem cells in fast proliferating healthy adult tissues such as colon have been reported to accumulate approximately 40 new mutations per year [13] (Figure 3). However, if mutation rates are in the order of 10^{-9} per base pair per cell division, which seems to be the current consensus and agrees with our measurements here, and the human genome consists of 6×10^9 base pairs, this would on average only allow for 6 to 7 divisions per stem cell per year. This is in contradiction to current measures on stem cell turnover rates in for example healthy colonic crypts [19,29]. This discrepancy is resolved by non-random

strand segregation, where many stem cell proliferations would not induce novel mutations on the stem cell level and the effective observed mutation accumulation on a population level can remain low despite high stem cell turnover rates. A clear molecular mechanism of strand recognition remains unknown. However, direct and indirect evidence to which our observations may contribute increasingly hint on the importance of strand segregation to maintain genomic integrity within healthy human tissues.

Our joined inference of mutation rate and strand segregation probability also reveals that mutation rates per cell division are likely higher than was assumed in previous studies [11]. We therefore find stronger signals of strand segregation in human sequencing data than was thought previously [11]. Our inference neglects the effects of cell-division independent mutations that may contribute to mutational burden in tissues at a low rate. This can lead to an underestimation of the true strand-segregation probability as well as the per-cell mutation rate in human tissues, see SI Figure 2.

A loss of strand segregation in stem cells implies a 50 to 100 times increased effective mutation rate on the cell population level without any other changes to the intrinsic DNA repair machinery. In a non-homeostatic setting, such as a growing tumour, in which the number of self-renewing cells (whether they are all or only a subset of cells) increases, the rate of random strand segregation events is much higher. This effect may contribute to the usually high mutational burden in cancers [30-33]. However, we note that our model has been developed for normal tissue and does not account for chromosomal rearrangements in

malignancies, which likely impact the estimation of mutation rates. It is an intriguing thought that early organ growth during development constitutes a very similar situation in which strand segregation is less effective within expanding stem cell populations and the increased rate of mutation accumulation early in life emerges as a natural consequence [34].

Materials and Methods

We assume that homeostasis in a healthy adult human tissue is maintained by a constant pool of N stem cells. Each of these stem cells undergoes n cell divisions during a time interval Δt. With each division, a stem cell non-randomly segregates DNA strands with a probability p. If $p = 1$ the ancestral strand will remain in the stem cell and the duplicated strand will be passed onto a daughter cell that becomes a non-stem cell, whereas $p = 0.5$ implies random strand segregation (i.e. no strand segregation), see Figure 1. We assume the probability p to be the same for all stem cells and don't account for possible variation by for example specific mutations that would change strand segregation probabilities for individual stem cells. The non-ancestral duplicated strand inherits on average μL novel mutations, where μ is the mutation rate per base pair per cell division and L the length of the copied genome (e.g. $L \approx 6{\times}10^9$ base pairs in humans). Throughout the manuscript we assume a constant mutation rate μ. In principal the mutation rate could depend on time explicitly, e.g. $\mu \to \mu(t)$. However, this would lead to non-linear dependencies, which is not supported by the currently available data, e.g. Figure 3 & 4. Thus assuming a constant mutation rate is retrospectively justified by the actual change of the mutational burden in human tissues.

It follows that for n cell divisions, the probability to segregate parental DNA strands k times is binomially distributed (k successes in n draws given a success probability of p)

$$P(k,n,p) = \binom{n}{k} p^k (1-p)^{n-k}. \quad \text{(S1)}$$

This implies that on average $E[k,n,p] = np$ cell divisions do not induce additional mutations in stem cells. However, $n(1-p)$ cell divisions will increase mutational burden within a single stem cell lineage, each division by a random number χ, given by a Poisson distribution:

$$P(\chi) = \frac{(\mu L)^{\chi}}{\chi!} e^{-\mu L}. \quad \text{(S2)}$$

The mutational burden $\tilde{\chi}$ within a single stem cell lineage consequently increases by

$$\tilde{\chi} = \sum_{i=1}^{n-k} \chi_i. \quad \text{(S3)}$$

Exact expressions for the mutational burden $\tilde{\mu}$ and variance σ^2 for such distributions are known[35]. The mutational burden $\tilde{\mu}$ after n stem cell divisions is given by

$$\tilde{\mu} = E[\tilde{\chi}] = E[n-k]E[\chi] = n(1-p)\mu L, \quad \text{(S4)}$$

and the variance of the mutational burden σ^2 is given by

$$\sigma^2 = E[n-k]Var[\chi] + (E[\chi])^2 Var[n-k] = n(1-p)\mu L + n(\mu L)^2 p(1-p). \quad \text{(S5)}$$

These expressions allow quantifying the change of the mutational burden as well as the change of the variance of the mutational burden after a number of Δn divisions per stem cell

$$\frac{\Delta\widetilde{\mu}}{\Delta n} = \frac{\widetilde{\mu_2} - \widetilde{\mu_1}}{n_2 - n_1} = (1-p)\mu L, \tag{S6}$$

$$\frac{\Delta\sigma^2}{\Delta n} = \frac{\sigma_2^2 - \sigma_1^2}{n_2 - n_1} = (1-p)\mu L + (\mu L)^2 p(1-p). \tag{S7}$$

However, in actual data the number of stem cell divisions is unknown and change would be measured in time t. Assuming a constant rate of stem cell proliferations λ we can write $\Delta n = \lambda \Delta t$. This allows us to rewrite above equations for the change of the mean and the variance of the mutational burden over real time t via

$$\frac{1}{\lambda}\frac{\Delta\widetilde{\mu}}{\Delta t} = (1-p)\mu L \tag{S8}$$

$$\frac{1}{\lambda}\frac{\Delta\sigma^2}{\Delta t} = (1-p)\mu L + (\mu L)^2 p(1-p). \tag{S9}$$

Importantly, both the change of the mutational burden $\frac{\Delta\widetilde{\mu}}{\Delta t}$ as well as the change of the variance of the mutational burden $\frac{\Delta\sigma^2}{\Delta t}$ can be measured from human somatic mutation data, see Figure 3 & 4. Furthermore, equations (S8) and (S9) imply that the mutational burden as well as the variance of the mutational burden are expected to increase linearly with age in adult tissues. Even if strand segregation is highly effective, mutations still accumulate linearly with age.

However, the rate of mutation accumulation is decreased by a factor of $(1-p)$. As we neither know the somatic mutation rate μ nor the stem cell proliferation rate λ with certainty, a linear increase in mutational burden with age at most suggests imperfect strand segregation (e.g. $0 \leq p < 1$). Importantly, the linear increase of both the mean and the variance in time is a result of the sum of Poisson distributed random variables and does by itself not imply the presence or absence of non-random strand segregation.

However, the ratio of variance and mean is independent of time t and the stem cell proliferation rate λ

$$\frac{\Delta\sigma^2}{\Delta\tilde{\mu}} = 1 + \mu L p, \tag{S10}$$

and therefore provides natural bounds for possible mutation rates per cell division μ and strand segregation probabilities p in human tissues, see SI Figure 1. Furthermore, rearranging equation (S8) and substituting $\mu L p = \mu L - \frac{1}{\lambda}\frac{\Delta\tilde{\mu}}{\Delta t}$ into equation (S10), the strand segregation probability p and the mutation rate μ disentangle, allowing us independent estimates via

$$p = \frac{\frac{\Delta\sigma^2}{\Delta\tilde{\mu}} - 1}{\frac{\Delta\sigma^2}{\Delta\tilde{\mu}} - 1 + \frac{1}{\lambda}\frac{\Delta\tilde{\mu}}{\Delta t}} \tag{S11}$$

$$\mu L = \frac{\Delta\sigma^2}{\Delta\tilde{\mu}} - 1 + \frac{1}{\lambda}\frac{\Delta\tilde{\mu}}{\Delta t}. \tag{S12}$$

The relative change of the mutational burden and the variance variance of the

mutational burden allow estimates of the mutation rate μ (per cell division) and the non-random strand segregation probability p. Estimating the mutation rate as well as the strand segregation probability, we need to measure the change of the mutational burden as well as the change of the variance of the mutational burden. This requires multiple measurements of the mutational burden within single cells of a single individual that ideally would be followed over time. This is unpractical and such data currently does not exist. We therefore measure the mutational burden and variance in multiple cells of multiple individuals of different ages. To calculate the variance and the mean of the mutational burden, we require at least 3 samples per individual, see Figure 3 & 4. For completeness we also show expressions for the mutation rate μ and p in dependence of stem cell proliferations n. They are given by

$$p = \frac{\frac{\Delta\sigma^2}{\Delta\tilde{\mu}} - 1}{\frac{\Delta\sigma^2}{\Delta\tilde{\mu}} - 1 + \frac{\Delta\tilde{\mu}}{\Delta n}} \tag{S13}$$

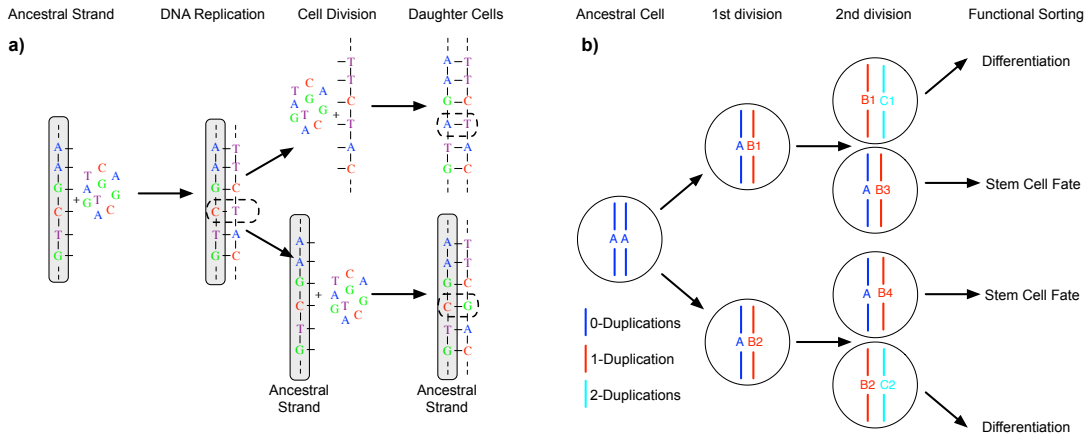
and

$$\mu L = \frac{\Delta\sigma^2}{\Delta\tilde{\mu}} - 1 + \frac{\Delta\tilde{\mu}}{\Delta n}. \tag{S14}$$

We recognize that our model is based on some assumptions and approximations. For example, telomeres, the protective ends of chromosomes, shorten with each cell division. Upon reaching a critically short telomere length, cells enter senescent. Senescence is not modelled in our model, however we argue that since this is likely to occur at very old ages [21,36], this process is unlikely to influence our results significantly.

Acknowledgments

B.W. is supported by the Geoffrey W. Lewis Post-Doctoral Training fellowship. A.S. is supported by The Chris Rokos Fellowship in Evolution and Cancer and by Cancer Research UK (A22909). A.S. is also supported by the Wellcome Trust (202778/B/16/Z). This work was also supported by Wellcome Trust funding to the Centre for Evolution and Cancer (105104/Z/14/Z).



422

423 **Figure 1: The Immortal DNA strand hypothesis. a)** During replication of the
424 ancestral DNA strand, errors (dashed line) might occur. If these errors are not
425 corrected by intrinsic DNA repair mechanisms, they become permanently fixed
426 in daughter cells after the next cell division. However, the original ancestral
427 strand is still present and can provide the blue print for additional non-mutated
428 copies of DNA. **b)** In principle, a stem cell driven tissue allows for non-random
429 DNA strand segregation. Preferentially segregating ancestral DNA strands into
430 stem cells and duplicated strands into differentiated cells with limited life span
431 can drastically reduce the accumulation of somatic mutations in tissues.

432

433

434

435

436

437

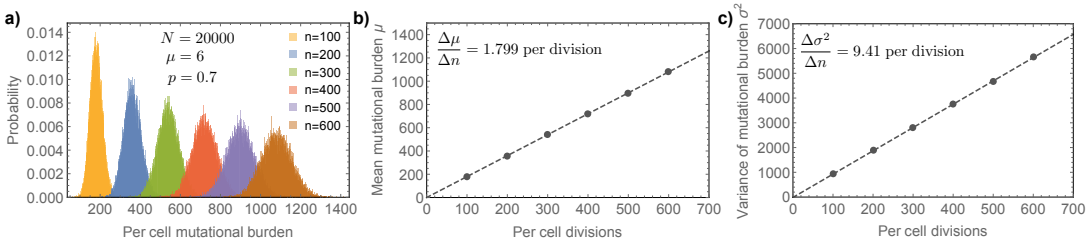


Figure 2: Predicted mutational burden in individual stem cells with age. a) We show simulated stochastic mutation accumulation in a stem cell population of constant size. Here $N = 20{,}000$ stem cells segregating DNA strands with probability $p = 0.7$ and a mutation rate of $\mu = 6$ per cell division (corresponding to a mutation rate of $\mu = 10^{-9}$ per bp per cell division). **b)** Mutational burden and **c)** variance of the mutational burden increase linear. Linear regression (dashed lines) gives $\frac{\Delta\tilde{\mu}}{\Delta t} = 1.799$ and $\frac{\Delta\sigma^2}{\Delta t} = 9.41$. The expected exact values based on above parameters and equation (1) and (2) are $\frac{\Delta\tilde{\mu}}{\Delta t} = 1.8$ and $\frac{\Delta\sigma^2}{\Delta t} = 9.36$. Equation (3) and (4) yield for the strand segregation probability $p = 0.702$ and for the mutation rate $\mu = 6.03$, (exact values imposed on the simulation were $p = 0.7$ and $\mu = 6$).

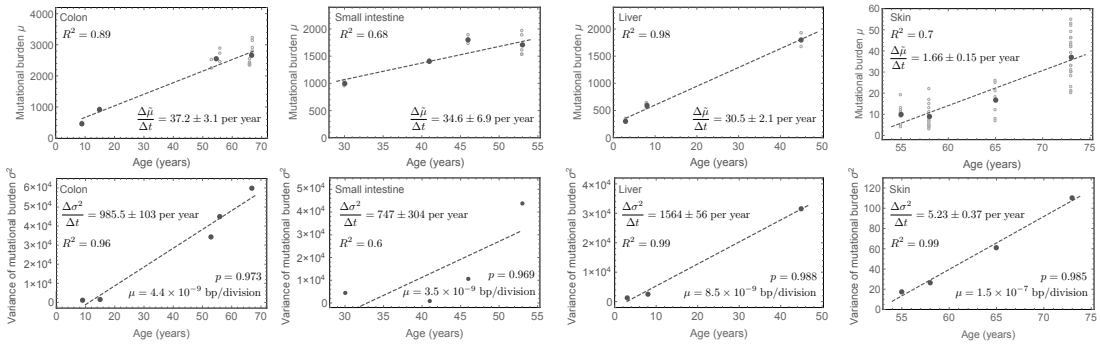


Figure 3: Mutational burden and variance in healthy human tissues. Mutational burden and variance of the mutational burden in colon, small intestine liver and skin tissue in healthy adult humans of different ages, data taken from [13,14]. Open circle represent mutational burden of single cells, whereas dark grey dots represent the mean mutational burden or variance respectively. In all cases, the data well supports our expectation of a linearly increasing mean and variance with age. Linear regressions (dashed lines) give estimates for the change of the mutational burden and the variance with age, see main text (uncertainties represent standard errors). Equations (3) and (4) then allow to estimate the non-random strand segregation probability as well as the per-cell mutation rate per cell division. In all cases, the probability of non-random strand segregation is high (median: $p = 0.979$ (0.97,0.99)), whereas the mutation rate per cell division varies between tissues and is highest in skin, see insets and main text for tissue specific estimates.

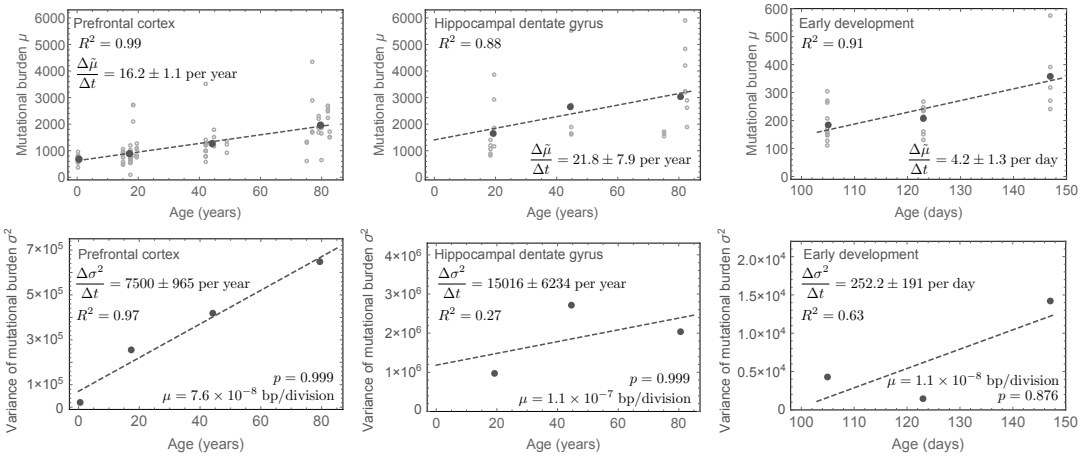
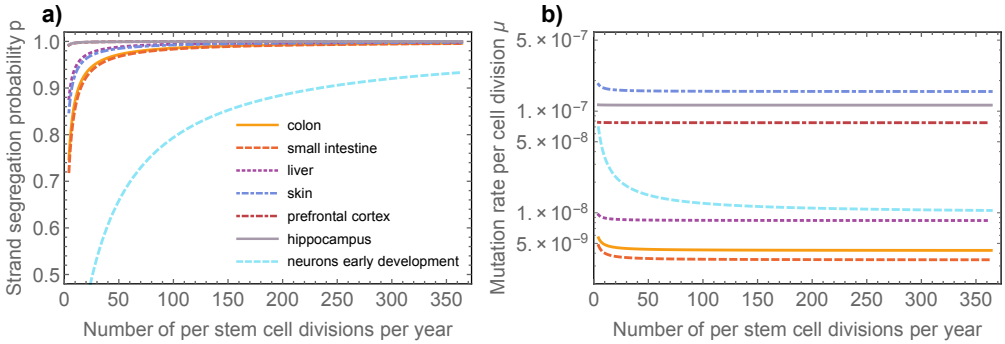


Figure 4: Mutational burden and variance in human neurons during early development and adulthood. Mutational and variance measured from single whole genome sequencing of neurons in the prefrontal cortex and the hippocampus dental gyrus [15] as well as in single neurons during early childhood development after birth [16] (uncertainties represent standard errors). Mutation accumulation in early childhood is highly increased compared to adulthood. However, the per-cell mutation rate per division appears higher in adulthood. The non-random strand segregation in contrast is with $p = 0.999\ (0.998; 0.9993)$ extremely high in adults, whereas with $p = 0.876\ (0.78; 0.88)$ it is lower in early childhood. This can be understood as a consequence of cell population expansions due to symmetric self-renewals in early childhood. For details of mutation rate estimates, see the main text.



494

495 **Figure 5**: **Dependence of parameter inferences on stem cell proliferation**
496 **rate.** Inferences of **a)** the DNA strand segregation probability and **b)** mutation
497 rate per cell division are robust against wide ranges of the stem cell proliferation
498 rate λ. If stem cells divide once per week this implies (Equation (3)) for the
499 probability of DNA strand segregation in colon: $p = 0.973\ (0.892;\ 0.996)$, small
500 intestine: $p = 0.969\ (0.877;\ 0.995)$, liver: $p = 0.988\ (0.952; 0.998)$, prefrontal
501 cortex: $p = 0.999\ (0.997; 0.9999)$, hippocampal dentale gyrus:
502 $p = 0.999\ (0.997; 0.9999)$, skin: $p = 0.985\ (0.94;\ 0.998)$. Numbers in brackets
503 correspond to the range of the DNA strand segregation probabilities for stem cell
504 replication rates between once per month and every day respectively. In contrast
505 for neurons during early development we find: $p = 0.876\ (0.67; 0.96)$ if cells
506 divide every 48h (number in brackets correspond to cell divisions once per week
507 and twice a day respectively). Based on Equation (4) we find for the *in vivo*
508 mutation rate per base pair per cell division in colon: $\mu = 4.37\ (4.27;\ 4.77)\times$
509 10^{-9}, small intestine: $\mu = 3.54\ (3.45;\ 3.91)\times 10^{-9}$, liver: $\mu = 8.48\ (8.39;\ 8.8)\times$
510 10^{-9}, prefrontal cortex: $\mu = 7.68\ (7.67;\ 7.7)\times 10^{-8}$, hippocampal dentale gyrus:
511 $\mu = 1.14\ (1.14;\ 1.15)\times 10^{-7}$, neurons during early development:
512 $\mu = 1.23\ (1.02;\ 1.47)\times 10^{-8}$ and skin: $\mu = 1.57\ (1.56;\ 1.65)\times 10^{-7}$.

513

References:

1. Cairns J. Mutation selection and the natural history of cancer. Nature. 1975 Oct 28;255:197–200.

2. Werner B, Dingli D, Lenaerts T, Pacheco JM, Traulsen A. Dynamics of Mutant Cells in Hierarchical Organized Tissues. PLoS Comput Biol. 2011 Dec 1;7(12):e1002290.

3. Rodriguez-Brenes IA, Wodarz D, Komarova NL. Minimizing the risk of cancer: tissue architecture and cellular replication limits. Journal of The Royal Society Interface. 2013 Jun 26;10(86):20130410–0.

4. Shinin V, Gayraud-Morel B, Gomès D, Tajbakhsh S. Asymmetric division and cosegregation of template DNA strands in adult muscle satellite cells. Nat Cell Biol. 2006 Jun 25;8(7):677–82.

5. Conboy MJ, Karasov AO, Rando TA. High Incidence of Non-Random Template Strand Segregation and Asymmetric Fate Determination In Dividing Stem Cells and their Progeny. Goodell MA, editor. PLoS biology. 2007 Apr 17;5(5):e102.

6. Booth C, Potten CS. Gut instincts: thoughts on intestinal epithelial stem cells. Journal of Clinical Investigation. 2000 Jun 1;105(11):1493–9.

7. Falconer E, Chavez EA, Henderson A, Poon SSS, McKinney S, Brown L, et al. Identification of sister chromatids by DNA template strand sequences. Nature. Nature Publishing Group; 2010 Jul 1;463(7277):93–7.

8. Quyn AJ, Appleton PL, Carey FA, Steele RJC, Barker N, Clevers H, et al. Short Article. Cell Stem Cell. Elsevier Ltd; 2010 Feb 5;6(2):175–81.

9. Lansdorp PM. Immortal Strands? Give Me a Break. Cell. 2007 Jun;129(7):1244–7.

10. Legraverend C, Escobar M, Jay P. "'The Immortal DNA Strand'": Difficult to Digest? Cell Stem Cell. Elsevier Inc; 2010 Apr 2;6(4):298–9.

11. Tomasetti C, Bozic I. The (not so) immortal strand hypothesis. Stem Cell Research. Elsevier B.V; 2015 Mar 1;14(2):238–41.

12. Kiel MJ, He S, Ashkenazi R, Gentry SN, Teta M, Kushner JA, et al. Haematopoietic stem cells do not asymmetrically segregate chromosomes or retain BrdU. Nature. 2007 Aug 29;449(7159):238–42.

13. Blokzijl F, de Ligt J, Jager M, Sasselli V, Roerink S, Sasaki N, et al. Tissue-specific mutation accumulation in human adult stem cells during life. Nature. Nature Publishing Group; 2016 Oct 13;538(7624):260–4.

14. Martincorena I, Roshan A, Gerstung M, Ellis P, Peter Van Loo, McLaren S, et al. High burden and pervasive positive selection of somatic mutations in normal human skin. Science. 2015 May 18;348:880–6.

15. Lodato MA, Rodin RE, Bohrson CL, Coulter ME, Barton AR, Kwon M, et al. Aging and neurodegeneration are associated with increased mutations in single human neurons. Science. 2018 Jan 29;359:555–9.

16. Bae T, Tomasini L, Mariani J, Zhou B, Roychowdhury T, Franjic D, et al. Different mutational rates and mechanisms in human cells at pregastrulation and neurogenesis. Science. 2017 Dec 7;543:eaan8690.

17. Yatabe Y, Tavare S, Shibata D. Investigating stem cells in human colon by using methylation patterns. Proceedings of the National Academy of Science. 2001 Aug 30;:1–6.

18. Knoblich JA. Asymmetric cell division:recent developments and theirimplications for tumour biology. Nature Publishing Group. Nature Publishing Group; 2010 Dec 1;11(12):849–60.

19. Clevers H. The Intestinal Crypt,A Prototype Stem Cell Compartment. Cell. Elsevier Inc; 2013 Jul 18;154(2):274–84.

20. Tomasetti C, Vogelstein B. Variation in cancer risk among tissues can be explained by the number of stem cell divisions. Science. 2015 Jan 1;347(6217):78–81.

21. Werner B, Beier F, Hummel S, Balabanov S, Lassay L, Orlikowsky T, et al. Reconstructing the *in vivo* dynamics of hematopoietic stem cells from telomere length distributions. eLife. 2015 Oct 15;10.7554:e08687v2.

22. Milholland B, Dong X, Zhang L, Hao X, Suh Y, Vijg J. Differences between germline and somatic mutation rates in humans and mice. Nature Communications. Nature Publishing Group; 2017 Apr 28;8:1–8.

23. Lynch M. Evolution of the mutation rate. Trends in Genetics. Elsevier Ltd; 2010 Aug 1;26(8):345–52.

24. Lynch M, Ackerman MS, Gout J-F, Long H, Sung W, Thomas WK, et al. Genetic drift, selection and the evolution of the mutation rate. Nature Reviews Genetics. Nature Publishing Group; 2016 Nov 1;17(11):704–14.

25. Sánchez-Danés A, Hannezo E, Larsimont J-C, Liagre M, Youssef KK, Simons BD, et al. Defining the clonal dynamics leading to mouse skin tumour initiation. Nature. Nature Publishing Group; 2016 Aug 18;536(7616):298–303.

26. Scheele CLGJ, Hannezo E, Muraro MJ, Zomer A, Langedijk NSM, van Oudenaarden A, et al. Identity and dynamics of mammary stem cells during branching morphogenesis. Nature. Nature Publishing Group; 2017 Feb 16;542(7641):313–7.

27. Ju YS, Martincorena I, Gerstung M, Petljak M, Alexandrov LB, Rahbari R, et al. Somatic mutations reveal asymmetric cellular dynamics in the early human embryo. Nature. Nature Publishing Group; 2017 Mar 30;543(7647):714–8.

28. Dingli D, Traulsen A, Michor F. (A)Symmetric Stem Cell Replication and Cancer. PLoS Comput Biol. 2007;3(3):e53.

29. Basak O, Beumer J, Wiebrands K, Seno H, van Oudenaarden A, Clevers H. Induced Quiescence of Lgr5+ Stem Cells in Intestinal Organoids Enables Differentiation of Hormone- Producing Enteroendocrine Cells. Stem Cell. Elsevier; 2017 Feb 2;20(2):177–190.e4.

30. Greenman C, Stephens P, Smith R, Dalgliesh GL, Hunter C, Bignell G, et al. Patterns of somatic mutation in human cancer genomes. Nature. Nature Publishing Group; 2007 Mar 8;446(7132):153–8.

31. McGranahan N, Swanton C. Clonal Heterogeneity and Tumor Evolution: Past, Present, and the Future. Cell. Elsevier Inc; 2017 Feb 9;168(4):613–28.

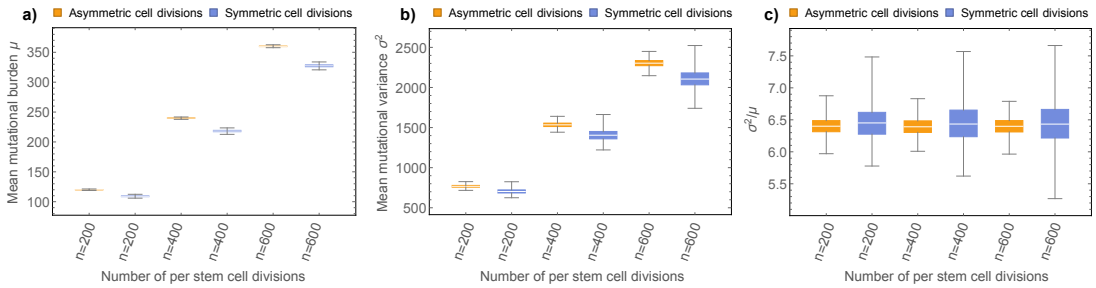
32. Gerlinger M, Rowan AJ, Horswell S, Larkin J, Endesfelder D, Gronroos E, et al. Intratumor Heterogeneity and Branched Evolution Revealed by Multiregion Sequencing. New England Journal of Medicine. 2012 Mar 8;366(10):883–92.

33. Davies H, Staaf J, Ramakrishna M, Glodzik D, Zou X, Martincorena I, et al. Landscape of somatic mutations in 560 breast cancer whole-genome sequences. Nature. Nature Publishing Group; 2016 May 2;:1–20.

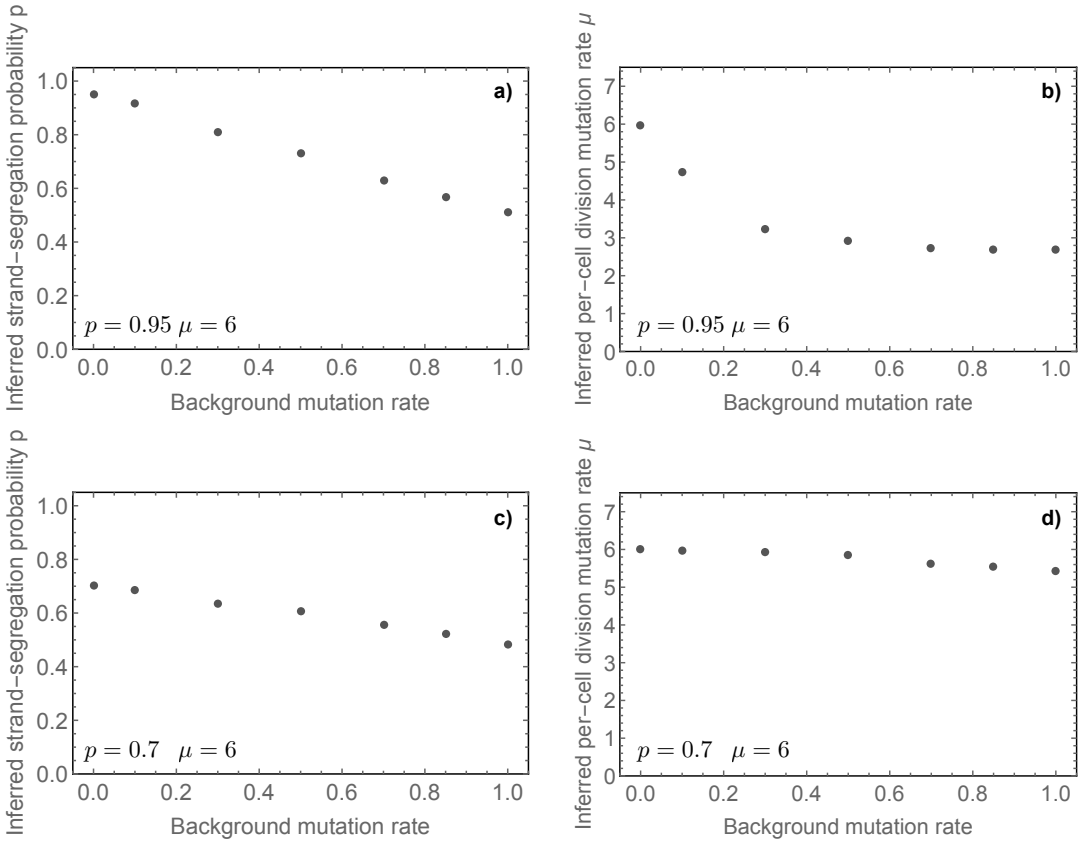
34. Frank SA. Somatic evolutionary genomics: Mutations during development cause highly variable genetic mosaicism with risk of cancer and neurodegeneration. Proceedings of the National Academy of Sciences. 2010 Jan 26;107(suppl_1):1725–30.

35. Ross SM. Introduction to Probability Models. Academic Press; 2014. 1 p.

36. Rodriguez-Brenes IA, Wodarz D, Komarova NL. Quantifying replicative senescenceas a tumor suppressor pathwayand a target for cancer therapy. Scientific Reports. Nature Publishing Group; 2015 Nov 27;:1–13.



SI Figure 1: Strand segregation in terms of symmetric stem cell divisions. a) Mean mutational burden μ, **b)** mutational variance σ^2, and **c)** the ratio of mutational variance and mutational burden σ^2/μ for purely asymmetrically or a mix of symmetrically and asymmetrically dividing stem cells. Here we compare stochastic simulations for $N = 5000$ purely asymmetrically dividing stem cells with a strand segregation probability of $p = 0.9$ and stem cells with perfect strand segregation $p = 1$ but a fraction of 10% of stem cell divisions being symmetric differentiations followed by symmetric self-renewals. Both scenarios lead to a linear increase of mean and variance of mutational burden with minimal rate differences. However, as predicted, the ratio of variance and mean become time independent and are the same on average for both processes. However, the variance of the distribution of the ratio of the variance and mean increases with time for symmetric stem cell divisions but is approximately constant for asymmetric stem cell divisions. This effect might provide a future method to distinguish and quantitate the amount of symmetric self-renewal in human stem cell populations.



SI Figure 2: Influence of cell division independent background mutation rate on inference of non-random strand segregation probability and per-cell mutation rate. Plots **a)** to **d)** show the non-random strand segregation probability p and the per cell division mutation rate μ based on equations (S11) and (S12) inferred from stochastic simulations if we in addition allow for a constant cell-division independent mutation rate that influences both the ancestral and the duplicated DNA strand equally. In the upper panels **a)** and **b)** the underlying true parameters per cell division are $\mu = 6$ and $p = 0.95$, whereas in the lower panels **c)** and **d)** we have $\mu = 6$ and $p = 0.7$. If the background mutation rate is 0, we recover the original parameters. Both the non-random strand segregation probability p as well as the per cell division mutation rate μ are slightly underestimated for an increasing background mutation rate. Importantly, the non-random strand segregation probability is always

underestimated and inferences become biologically meaningless (e.g. $p < 0.5$) for large background mutation rates. The actual data suggests high non-random strand segregation probabilities (see main text) and therefore implies small background mutation rates compared to cell division induced mutations.