

# 1 Variation of mutational burden in 2 healthy human tissues suggests non- 3 random strand segregation and allows 4 measuring somatic mutation rates

---

5 Benjamin Werner<sup>1,\*</sup> & Andrea Sottoriva<sup>1,\*</sup>

6

7 <sup>1</sup>Evolutionary Genomics & Modelling Lab, Centre for Evolution and Cancer, The Institute of Cancer  
8 Research, London, UK

9 \*Correspondence should be addressed to: [benjamin.werner@icr.ac.uk](mailto:benjamin.werner@icr.ac.uk) and [andrea.sottoriva@icr.ac.uk](mailto:andrea.sottoriva@icr.ac.uk)

10

## 11 **Abstract**

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

25 tissue types, ranging from  $3.5 \times 10^{-9}$  /bp/division in small intestine to  $1.6 \times$   
26  $10^{-7}$  /bp/division in skin.

27

## 28 **Author Summary**

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

## 39 **Introduction**

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

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

54

55 Orthogonal studies based on the expected accumulation of somatic mutations in  
56 healthy human tissues have argued against the immortal strand hypothesis  
57 [11,12]. However, the mere accumulation of somatic mutations in healthy tissue  
58 neither supports nor negates the immortal strand hypothesis *in vivo*. Here, we  
59 show that measuring the change of the mutational burden and, most crucially,  
60 the change of the *variance* of the mutational burden with age allows determining  
61 the probability of DNA strand segregation and the per cell mutation rate in  
62 healthy human tissues. First, we outline the approach and then apply it to  
63 genomic data from healthy human colon, small intestine, liver, skin and brain  
64 tissue. The data comes from four recent independent studies on mutational  
65 burden in healthy tissues [13-16], which contain information on in total 39  
66 individuals at different ages and analysed genomes of 341 single cells. We find  
67 evidence for non-random strand segregation in all adult tissues and significant  
68 differences in somatic mutation rates between tissues, but less prominent  
69 strand-segregation in brain tissue during early development.

## 70 Results

### 71 The expected change of mean and variance of mutational burden with 72 age

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

88

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

94

$$95 \quad \frac{1}{\lambda} \frac{\Delta \tilde{\mu}}{\Delta t} = (1 - p)\mu L \quad (1)$$

$$96 \quad \frac{1}{\lambda} \frac{\Delta \sigma^2}{\Delta t} = (1 - p)\mu L + (\mu L)^2 p(1 - p), \quad (2)$$

97

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

104

$$105 \quad p = \frac{\frac{\Delta \sigma^2}{\Delta \tilde{\mu}} - 1}{\frac{\Delta \sigma^2}{\Delta \tilde{\mu}} - 1 + \frac{1}{\lambda} \frac{\Delta \tilde{\mu}}{\Delta t}} \quad (3)$$

$$106 \quad \mu L = \frac{\Delta \sigma^2}{\Delta \tilde{\mu}} - 1 + \frac{1}{\lambda} \frac{\Delta \tilde{\mu}}{\Delta t}. \quad (4)$$

107

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

112

### 113 **Measured mean and variance of mutational burden with age from** 114 **sequencing data**

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

116 accumulation in healthy colon, small intestine and liver tissue by whole genome  
117 sequencing multiple single stem cell derived organoids of healthy donors of  
118 different ages. In addition, Martincorena and colleagues [14] measured  
119 mutational burden in multiple skin samples of four individuals with ages  
120 between 58 and 73 years. Furthermore, two recent publications [15,16]  
121 performed large-scale single cell whole genome sequencing of neurons at  
122 different ages. In the experiments by Blokzijl and colleagues [1,13], they isolated  
123 single cells and expanded those into organoids. These cells can therefore be  
124 thought of as tissue specific stem cells. In contrast, the other experiments  
125 [2,3,14-16] do not directly measure mutational burden in stem but more  
126 differentiated progenitor cells. However, compared to the total number of cell  
127 divisions in the tissue, the number of divisions separating stem and progenitor  
128 cells is neglectable.

129

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

137

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

140 For all tissues, the experimental observations confirm our expectation of a  
141 linearly increasing mean and variance of the mutational burden. Using linear  
142 regression on the data in [4-7,13-16], we find for colon that the change in  
143 mutational burden over time was:  $\frac{\Delta\tilde{\mu}}{\Delta t} = 37.2 \pm 3.1$ , for small intestine:  $\frac{\Delta\tilde{\mu}}{\Delta t} =$   
144  $34.6 \pm 6.9$ , 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  
145 for hippocampal dentate gyrus:  $\frac{\Delta\tilde{\mu}}{\Delta t} = 21.8 \pm 7.9$  mutations per whole genome per  
146 year. We found for skin:  $\frac{\Delta\tilde{\mu}}{\Delta t} = 1.66 \pm 0.15$  mutations per 0.69 Mb per year. We  
147 found for neurons during early development:  $\frac{\Delta\tilde{\mu}}{\Delta t} = 4.2 \pm 1.3$  mutations per  
148 whole genome per day. Uncertainties here are standard errors. Similarly, for the  
149 change of variance we found for colon:  $\frac{\Delta\sigma^2}{\Delta t} = 985.5 \pm 103$ , for small intestine:  
150  $\frac{\Delta\sigma^2}{\Delta t} = 747.3 \pm 304$ , for liver:  $\frac{\Delta\sigma^2}{\Delta t} = 1564 \pm 56$ , for prefrontal cortex:  $\frac{\Delta\sigma^2}{\Delta t} =$   
151  $7500 \pm 965$ , for hippocampal dentate gyrus:  $\frac{\Delta\sigma^2}{\Delta t} = 15016 \pm 6234$  mutations per  
152 whole genome per year, for skin:  $\frac{\Delta\sigma^2}{\Delta t} = 5.23 \pm 0.37$  mutations per 0.69 Mb per  
153 year and for neurons during early development:  $\frac{\Delta\sigma^2}{\Delta t} = 252.2 \pm 191$  mutations  
154 per whole genome per day (Figure 3 & 4).

155

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

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

166

167 This suggests highly effective non-random DNA strand segregation in human  
168 adult stem cells and is in line with previous observations of predominantly  
169 asymmetric stem cell divisions [6-8,17,18]. It would require extreme stem cell  
170 proliferation rates of approximately one division per stem cell per year for the  
171 data to be consistent with solely random strand segregation ( $p = 0.5$ ), Figure 5.  
172 This is an unlikely scenario as all tissues analysed here are thought to have high  
173 stem cell proliferation rates [9,19,20]. Interestingly, during development non-  
174 random DNA strand segregation is less prominent. One explanation is an  
175 expanding stem cell population due to symmetric stem cell self-renewals during  
176 early development[1,10], which also would explain the increased accumulation  
177 of mutations early, as well as typical increased telomere shortening early in life  
178 [11,12,21].

179

## 180 **Measurements of somatic mutation rates in healthy human tissues**

181 Based on Equation (4) we find for the *in vivo* mutation rate per base pair per cell  
182 division in colon:  $\mu = 4.37 (4.26; 4.46) \times 10^{-9}$  , small intestine:  
183  $\mu = 3.54 (2.61; 4.17) \times 10^{-9}$  , liver:  $\mu = 8.48 (8.22; 8.77) \times 10^{-9}$  , prefrontal  
184 cortex:  $\mu = 7.68 (7.18; 8.12) \times 10^{-8}$ , hippocampal dentate gyrus:  
185  $\mu = 1.14 (1.04; 1.68) \times 10^{-7}$ , neurons during early development:  
186  $\mu = 1.23 (0.43; 1.52) \times 10^{-8}$  and skin:  $\mu = 1.57 (1.54; 1.63) \times 10^{-7}$ . The ranges of



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

199

## 200 **Explaining strand segregation in terms of symmetric stem cell divisions**

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

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

## 227 Discussion

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

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

243

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

252

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

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

266

267

268

269

270

271

272

273

274

275

276

277

278

279

280

281

282

283

284

## 285 **Materials and Methods**

286

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

305

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

309

$$310 \quad P(k, n, p) = \binom{n}{k} p^k (1-p)^{n-k}. \quad (S1)$$

311

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

316

$$317 \quad P(\chi) = \frac{(\mu L)^\chi}{\chi!} e^{-\mu L}. \quad (S2)$$

318

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

321

$$322 \quad \tilde{\chi} = \sum_{i=1}^{n-k} \chi_i. \quad (S3)$$

323

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

327

$$328 \quad \tilde{\mu} = E[\tilde{\chi}] = E[n-k]E[\chi] = n(1-p)\mu L, \quad (S4)$$

329

330 and the variance of the mutational burden  $\sigma^2$  is given by

331

$$332 \quad \sigma^2 = E[n-k]Var[\chi] + (E[\chi])^2Var[n-k] = n(1-p)\mu L + n(\mu L)^2 p(1-p). \quad (S5)$$

333

334 These expressions allow quantifying the change of the mutational burden as well  
335 as the change of the variance of the mutational burden after a number of  $\Delta n$   
336 divisions per stem cell

337

$$338 \quad \frac{\Delta \tilde{\mu}}{\Delta n} = \frac{\tilde{\mu}_2 - \tilde{\mu}_1}{n_2 - n_1} = (1 - p)\mu L, \quad (S6)$$

$$339 \quad \frac{\Delta \sigma^2}{\Delta n} = \frac{\sigma_2^2 - \sigma_1^2}{n_2 - n_1} = (1 - p)\mu L + (\mu L)^2 p(1 - p). \quad (S7)$$

340

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

346

$$347 \quad \frac{1}{\lambda} \frac{\Delta \tilde{\mu}}{\Delta t} = (1 - p)\mu L \quad (S8)$$

$$348 \quad \frac{1}{\lambda} \frac{\Delta \sigma^2}{\Delta t} = (1 - p)\mu L + (\mu L)^2 p(1 - p). \quad (S9)$$

349

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

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

363

364 However, the ratio of variance and mean is independent of time  $t$  and the stem  
 365 cell proliferation rate  $\lambda$

366

367 
$$\frac{\Delta\sigma^2}{\Delta\tilde{\mu}} = 1 + \mu L p, \quad (\text{S10})$$

368

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

374

375 
$$p = \frac{\frac{\Delta\sigma^2}{\Delta\tilde{\mu}} - 1}{\frac{\Delta\sigma^2}{\Delta\tilde{\mu}} - 1 + \frac{1}{\lambda} \frac{\Delta\tilde{\mu}}{\Delta t}} \quad (\text{S11})$$

376 
$$\mu L = \frac{\Delta\sigma^2}{\Delta\tilde{\mu}} - 1 + \frac{1}{\lambda} \frac{\Delta\tilde{\mu}}{\Delta t}. \quad (\text{S12})$$

377

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



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

$$391 \quad p = \frac{\frac{\Delta\sigma^2}{\Delta\tilde{\mu}} - 1}{\frac{\Delta\sigma^2}{\Delta\tilde{\mu}} - 1 + \frac{\Delta\tilde{\mu}}{\Delta n}} \quad (S13)$$

392 and

$$393 \quad \mu L = \frac{\Delta\sigma^2}{\Delta\tilde{\mu}} - 1 + \frac{\Delta\tilde{\mu}}{\Delta n}. \quad (S14)$$

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

## 400 **Acknowledgments**

401 B.W. is supported by the Geoffrey W. Lewis Post-Doctoral Training fellowship. A.S. is  
402 supported by The Chris Rokos Fellowship in Evolution and Cancer and by Cancer  
403 Research UK (A22909). A.S. is also supported by the Wellcome Trust  
404 (202778/B/16/Z). This work was also supported by Wellcome Trust funding to the  
405 Centre for Evolution and Cancer (105104/Z/14/Z).

406

407

408

409

410

411

412

413

414

415

416

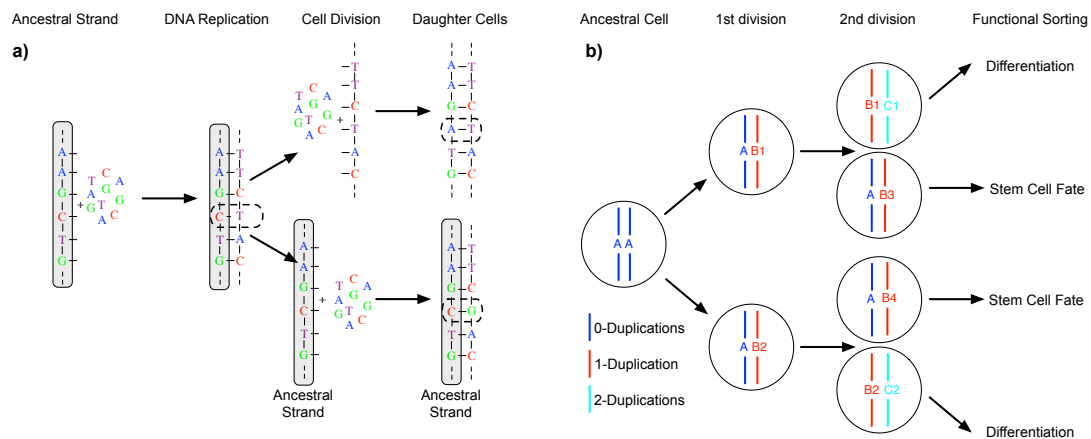
417

418

419

420

421



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

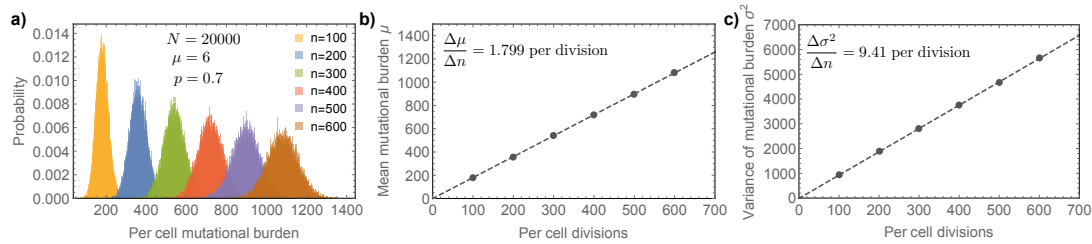
433

434

435

436

437



438

439 **Figure 2: Predicted mutational burden in individual stem cells with age. a)**

440 We show simulated stochastic mutation accumulation in a stem cell population

441 of constant size. Here  $N = 20,000$  stem cells segregating DNA strands with

442 probability  $p = 0.7$  and a mutation rate of  $\mu = 6$  per cell division (corresponding

443 to a mutation rate of  $\mu = 10^{-9}$  per bp per cell division). **b)** Mutational burden

444 and **c)** variance of the mutational burden increase linear. Linear regression

445 (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

446 on above parameters and equation (1) and (2) are  $\frac{\Delta\tilde{\mu}}{\Delta t} = 1.8$  and  $\frac{\Delta\sigma^2}{\Delta t} = 9.36$ .

447 Equation (3) and (4) yield for the strand segregation probability  $p = 0.702$  and

448 for the mutation rate  $\mu = 6.03$ , (exact values imposed on the simulation were

449  $p = 0.7$  and  $\mu = 6$ ).

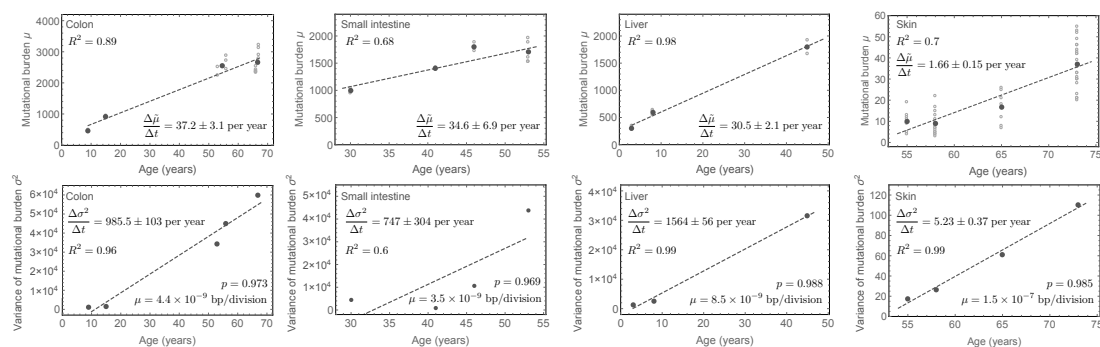
450

451

452

453

454



455

456 **Figure 3: Mutational burden and variance in healthy human tissues.**

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

470

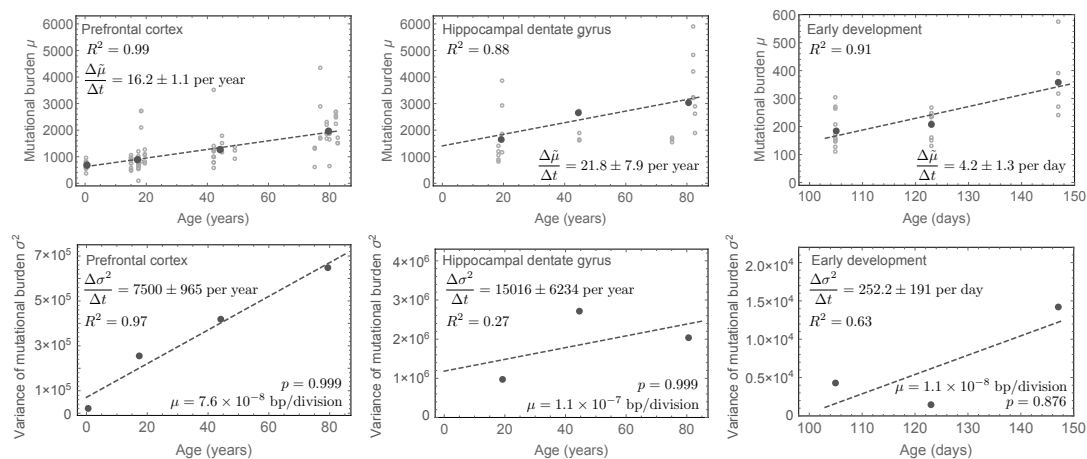
471

472

473

474

475



476

477 **Figure 4: Mutational burden and variance in human neurons during early**

478 **development and adulthood.** Mutational and variance measured from single

479 whole genome sequencing of neurons in the prefrontal cortex and the

480 hippocampus dental gyrus [15] as well as in single neurons during early

481 childhood development after birth [16] (uncertainties represent standard

482 errors). Mutation accumulation in early childhood is highly increased compared

483 to adulthood. However, the per-cell mutation rate per division appears higher in

484 adulthood. The non-random strand segregation in contrast is with

485  $p = 0.999$  (0.998; 0.9993) extremely high in adults, whereas with

486  $p = 0.876$  (0.78; 0.88) it is lower in early childhood. This can be understood as a

487 consequence of cell population expansions due to symmetric self-renewals in

488 early childhood. For details of mutation rate estimates, see the main text.

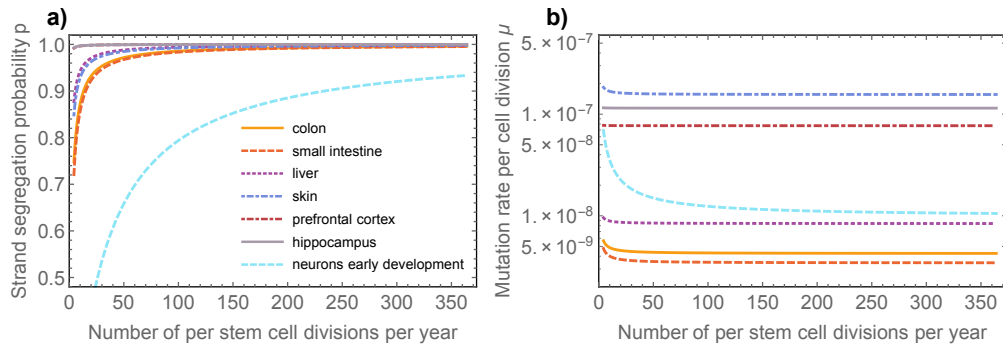
489

490

491

492

493



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  $\lambda$ . 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

514 **References:**

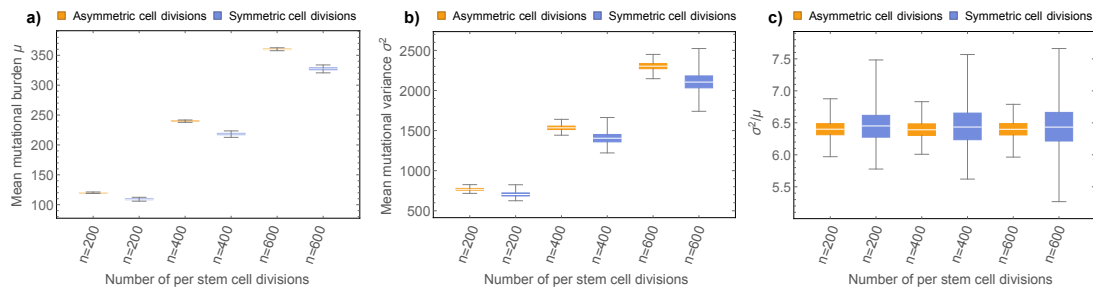
515

- 516 1. Cairns J. Mutation selection and the natural history of cancer. *Nature*. 1975  
517 Oct 28;255:197–200.
- 518 2. Werner B, Dingli D, Lenaerts T, Pacheco JM, Traulsen A. Dynamics of  
519 Mutant Cells in Hierarchical Organized Tissues. *PLoS Comput Biol*. 2011  
520 Dec 1;7(12):e1002290.
- 521 3. Rodriguez-Brenes IA, Wodarz D, Komarova NL. Minimizing the risk of  
522 cancer: tissue architecture and cellular replication limits. *Journal of The  
523 Royal Society Interface*. 2013 Jun 26;10(86):20130410–0.
- 524 4. Shinin V, Gayraud-Morel B, Gomès D, Tajbakhsh S. Asymmetric division  
525 and cosegregation of template DNA strands in adult muscle satellite cells.  
526 *Nat Cell Biol*. 2006 Jun 25;8(7):677–82.
- 527 5. Conboy MJ, Karasov AO, Rando TA. High Incidence of Non-Random  
528 Template Strand Segregation and Asymmetric Fate Determination In  
529 Dividing Stem Cells and their Progeny. Goodell MA, editor. *PLoS biology*.  
530 2007 Apr 17;5(5):e102.
- 531 6. Booth C, Potten CS. Gut instincts: thoughts on intestinal epithelial stem  
532 cells. *Journal of Clinical Investigation*. 2000 Jun 1;105(11):1493–9.
- 533 7. Falconer E, Chavez EA, Henderson A, Poon SSS, McKinney S, Brown L, et al.  
534 Identification of sister chromatids by DNA template strand sequences.  
535 *Nature*. Nature Publishing Group; 2010 Jul 1;463(7277):93–7.
- 536 8. Quyn AJ, Appleton PL, Carey FA, Steele RJC, Barker N, Clevers H, et al. Short  
537 Article. *Cell Stem Cell*. Elsevier Ltd; 2010 Feb 5;6(2):175–81.
- 538 9. Lansdorp PM. Immortal Strands? Give Me a Break. *Cell*. 2007  
539 Jun;129(7):1244–7.
- 540 10. Legrauerend C, Escobar M, Jay P. “The Immortal DNA Strand”: Difficult to  
541 Digest? *Cell Stem Cell*. Elsevier Inc; 2010 Apr 2;6(4):298–9.
- 542 11. Tomasetti C, Bozic I. The (not so) immortal strand hypothesis. *Stem Cell  
543 Research*. Elsevier B.V; 2015 Mar 1;14(2):238–41.
- 544 12. Kiel MJ, He S, Ashkenazi R, Gentry SN, Teta M, Kushner JA, et al.  
545 Haematopoietic stem cells do not asymmetrically segregate chromosomes  
546 or retain BrdU. *Nature*. 2007 Aug 29;449(7159):238–42.
- 547 13. Blokzijl F, de Ligt J, Jager M, Sasselli V, Roerink S, Sasaki N, et al. Tissue-  
548 specific mutation accumulation in human adult stem cells during life.  
549 *Nature*. Nature Publishing Group; 2016 Oct 13;538(7624):260–4.



- 550 14. Martincorena I, Roshan A, Gerstung M, Ellis P, Peter Van Loo, McLaren S, et  
551 al. High burden and pervasive positive selection of somatic mutations in  
552 normal human skin. *Science*. 2015 May 18;348:880–6.
- 553 15. Lodato MA, Rodin RE, Bohrson CL, Coulter ME, Barton AR, Kwon M, et al.  
554 Aging and neurodegeneration are associated with increased mutations in  
555 single human neurons. *Science*. 2018 Jan 29;359:555–9.
- 556 16. Bae T, Tomasini L, Mariani J, Zhou B, Roychowdhury T, Franjic D, et al.  
557 Different mutational rates and mechanisms in human cells at  
558 pregastrulation and neurogenesis. *Science*. 2017 Dec 7;543:eaan8690.
- 559 17. Yatabe Y, Tavaré S, Shibata D. Investigating stem cells in human colon by  
560 using methylation patterns. *Proceedings of the National Academy of  
561 Science*. 2001 Aug 30;98:1–6.
- 562 18. Knoblich JA. Asymmetric cell division: recent developments and  
563 their implications for tumour biology. *Nature Publishing Group. Nature  
564 Publishing Group*; 2010 Dec 1;473(7342):31–8.
- 565 19. Clevers H. The Intestinal Crypt, A Prototype Stem Cell Compartment. *Cell.  
566 Elsevier Inc*; 2013 Jul 18;154(2):274–84.
- 567 20. Tomasetti C, Vogelstein B. Variation in cancer risk among tissues can be  
568 explained by the number of stem cell divisions. *Science*. 2015 Jan  
569 1;347(6217):78–81.
- 570 21. Werner B, Beier F, Hummel S, Balabanov S, Lassay L, Orlikowsky T, et al.  
571 Reconstructing the *in vivo* dynamics of hematopoietic stem cells from  
572 telomere length distributions. *eLife*. 2015 Oct 15;10.7554:e08687v2.
- 573 22. Milholland B, Dong X, Zhang L, Hao X, Suh Y, Vijg J. Differences between  
574 germline and somatic mutation rates in humans and mice. *Nature  
575 Communications. Nature Publishing Group*; 2017 Apr 28;8:1–8.
- 576 23. Lynch M. Evolution of the mutation rate. *Trends in Genetics. Elsevier Ltd*;  
577 2010 Aug 1;26(8):345–52.
- 578 24. Lynch M, Ackerman MS, Gout J-F, Long H, Sung W, Thomas WK, et al.  
579 Genetic drift, selection and the evolution of the mutation rate. *Nature  
580 Reviews Genetics. Nature Publishing Group*; 2016 Nov 1;17(11):704–14.
- 581 25. Sánchez-Danés A, Hannezo E, Larsimont J-C, Liagre M, Youssef KK, Simons  
582 BD, et al. Defining the clonal dynamics leading to mouse skin tumour  
583 initiation. *Nature. Nature Publishing Group*; 2016 Aug 18;536(7616):298–  
584 303.
- 585 26. Scheele CLGJ, Hannezo E, Muraro MJ, Zomer A, Langedijk NSM, van  
586 Oudenaarden A, et al. Identity and dynamics of mammary stem cells  
587 during branching morphogenesis. *Nature. Nature Publishing Group*; 2017  
588 Feb 16;542(7641):313–7.

- 589 27. Ju YS, Martincorena I, Gerstung M, Petljak M, Alexandrov LB, Rahbari R, et  
590 al. Somatic mutations reveal asymmetric cellular dynamics in the early  
591 human embryo. *Nature*. Nature Publishing Group; 2017 Mar  
592 30;543(7647):714–8.
- 593 28. Dingli D, Traulsen A, Michor F. (A)Symmetric Stem Cell Replication and  
594 Cancer. *PLoS Comput Biol*. 2007;3(3):e53.
- 595 29. Basak O, Beumer J, Wiebrands K, Seno H, van Oudenaarden A, Clevers H.  
596 Induced Quiescence of Lgr5+ Stem Cells in Intestinal Organoids Enables  
597 Differentiation of Hormone- Producing Enteroendocrine Cells. *Stem Cell*.  
598 Elsevier; 2017 Feb 2;20(2):177–190.e4.
- 599 30. Greenman C, Stephens P, Smith R, Dalgliesh GL, Hunter C, Bignell G, et al.  
600 Patterns of somatic mutation in human cancer genomes. *Nature*. Nature  
601 Publishing Group; 2007 Mar 8;446(7132):153–8.
- 602 31. McGranahan N, Swanton C. Clonal Heterogeneity and Tumor Evolution:  
603 Past, Present, and the Future. *Cell*. Elsevier Inc; 2017 Feb 9;168(4):613–28.
- 604 32. Gerlinger M, Rowan AJ, Horswell S, Larkin J, Endesfelder D, Gronroos E, et  
605 al. Intratumor Heterogeneity and Branched Evolution Revealed by  
606 Multiregion Sequencing. *New England Journal of Medicine*. 2012 Mar  
607 8;366(10):883–92.
- 608 33. Davies H, Staaf J, Ramakrishna M, Glodzik D, Zou X, Martincorena I, et al.  
609 Landscape of somatic mutations in 560 breast cancer whole-genome  
610 sequences. *Nature*. Nature Publishing Group; 2016 May 2;:1–20.
- 611 34. Frank SA. Somatic evolutionary genomics: Mutations during development  
612 cause highly variable genetic mosaicism with risk of cancer and  
613 neurodegeneration. *Proceedings of the National Academy of Sciences*.  
614 2010 Jan 26;107(suppl\_1):1725–30.
- 615 35. Ross SM. *Introduction to Probability Models*. Academic Press; 2014. 1 p.
- 616 36. Rodriguez-Brenes IA, Wodarz D, Komarova NL. Quantifying replicative  
617 senescence as a tumor suppressor pathway and a target for cancer therapy.  
618 *Scientific Reports*. Nature Publishing Group; 2015 Nov 27;:1–13.
- 619
- 620
- 621
- 622
- 623
- 624
- 625
- 626
- 627
- 628



629

630

631 **SI Figure 1: Strand segregation in terms of symmetric stem cell divisions. a)**

632 Mean mutational burden  $\mu$ , **b)** mutational variance  $\sigma^2$ , and **c)** the ratio of

633 mutational variance and mutational burden  $\sigma^2/\mu$  for purely asymmetrically or a

634 mix of symmetrically and asymmetrically dividing stem cells. Here we compare

635 stochastic simulations for  $N = 5000$  purely asymmetrically dividing stem cells

636 with a strand segregation probability of  $p = 0.9$  and stem cells with perfect

637 strand segregation  $p = 1$  but a fraction of 10% of stem cell divisions being

638 symmetric differentiations followed by symmetric self-renewals. Both scenarios

639 lead to a linear increase of mean and variance of mutational burden with

640 minimal rate differences. However, as predicted, the ratio of variance and mean

641 become time independent and are the same on average for both processes.

642 However, the variance of the distribution of the ratio of the variance and mean

643 increases with time for symmetric stem cell divisions but is approximately

644 constant for asymmetric stem cell divisions. This effect might provide a future

645 method to distinguish and quantitate the amount of symmetric self-renewal in

646 human stem cell populations.

647

648

649

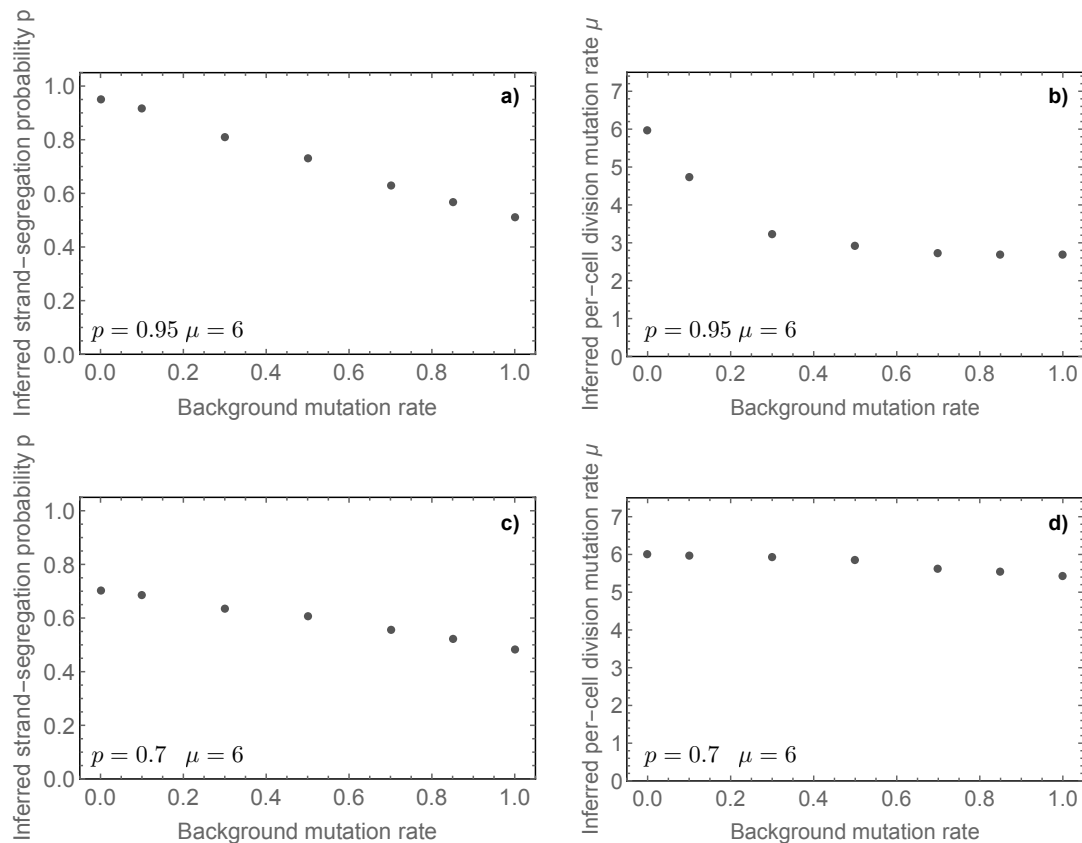
650

651

652

653

654



655

656

657

## SI Figure 2: Influence of cell division independent background mutation

658

rate on inference of non-random strand segregation probability and per-

659

cell mutation rate. Plots a) to d) show the non-random strand segregation

660

probability  $p$  and the per cell division mutation rate  $\mu$  based on equations (S11)

661

and (S12) inferred from stochastic simulations if we in addition allow for a

662

constant cell-division independent mutation rate that influences both the

663

ancestral and the duplicated DNA strand equally. In the upper panels a) and b)

664

the underlying true parameters per cell division are  $\mu = 6$  and  $p = 0.95$ , whereas

665

in the lower panels c) and d) we have  $\mu = 6$  and  $p = 0.7$ . If the background

666

mutation rate is 0, we recover the original parameters. Both the non-random

667

strand segregation probability  $p$  as well as the per cell division mutation rate  $\mu$

668

are slightly underestimated for an increasing background mutation rate.

669

Importantly, the non-random strand segregation probability is always

670 underestimated and inferences become biologically meaningless (e.g.  $p < 0.5$ )  
671 for large background mutation rates. The actual data suggests high non-random  
672 strand segregation probabilities (see main text) and therefore implies small  
673 background mutation rates compared to cell division induced mutations.  
674