Measuring single cell divisions in human tissues from multi-region sequencing data

Benjamin Werner[1,2,*], Jack Case[1,3], Marc J. Williams[4,5,6], Kate Chkhaidze[1], Daniel Temko[4], Javier Fernandez-Mateos[1], George D. Cresswell[1], Daniel Nichol[1], William Cross[4], Inmaculada Spiteri[1], Weini Huang[7,8], Ian Tomlinson[9], Chris P. Barnes[5,10], Trevor A. Graham[4,*] & Andrea Sottoriva[1,*]

[1]Evolutionary Genomics & Modelling Lab, Centre for Evolution and Cancer, The Institute of Cancer Research, Sutton, London SM2 5NG, UK

[2]Evolutionary Dynamics Group, Centre for Cancer Genomics & Computational Biology, Barts Cancer Institute, Queen Mary University of London, Charterhouse Square, London, UK EC1M 6BQ.

[3]University of Cambridge

[4]Evolution and Cancer Laboratory, Centre for Cancer Genomics & Computational Biology, Barts Cancer Institute, Queen Mary University London, London, Charterhouse Square, London, UK EC1M 6BQ.

[5]Department of Cell and Developmental Biology, University College London, London, UK

[6]Centre for Mathematics and Physics in the Life Sciences and Experimental Biology (CoMPLEX), University College London, London, UK

[7]Group of Theoretical Biology, The State Key Laboratory of Biocontrol, School of Life Science, Sun Yat-sen University, Guangzhou, 510060 China

[8]School of Mathematical Sciences, Queen Mary University London, London, UK.

[9]Institute of Cancer and Genomic Sciences, University of Birmingham, Birmingham, UK

[10]UCL Genetics Institute, University College London, London, UK

*Correspondence should be addressed to: b.werner@qmul.ac.uk, t.graham@qmul.ac.uk, andrea.sottoriva@icr.ac.uk

Abstract

Both normal tissue development and cancer growth are driven by a branching process of cell division and mutation accumulation that leads to intra-tissue genetic heterogeneity. However, quantifying somatic evolution in humans remains challenging. Here, we show that multi-sample genomic data from a single time point of normal and cancer tissues contains information on single-cell divisions. We present a new theoretical framework that, applied to whole-genome sequencing data of healthy tissue and cancer, allows inferring the mutation rate and the cell survival/death rate per division. On average, we found that cells accumulate 1.14 mutations per cell division in healthy haematopoiesis and 1.37 mutations per division in brain development. In both tissues, cell survival was maximal during early development. Analysis of 131 biopsies from 16 tumours showed 4 to 100 times increased mutation rates compared to healthy development and substantial inter-patient variation of cell survival/death rates.

Introduction

Most cells in human tissues have a limited life span and need to be replenished for tissues to remain functional[1-3]. This cell turnover leads to somatic evolution, with cells accumulating mutations upon which selection may act[4,5]. Inter- and intra-tumour genetic heterogeneity[6,7] as well as treatment resistance[8,9] are now understood to be consequences of somatic evolutionary processes. Recent studies demonstrate somatic evolution in healthy non-cancerous tissues

2

53   throughout live[10-14]. Normal brain cells carry hundreds of mutations weeks after

54   conception[12] and normal skin or esophagus cells accumulate hundreds of cancer

55   driver mutations during adulthood[10,11].

56   These observations call for a better quantitative understanding of the somatic

57   evolutionary forces in both cancerous and healthy tissues[15]. However, unlike

58   species evolution for which a timed fossil record exists[16,17], the lack of sequential

59   human data over time due to ethical and technical limitations is a major obstacle.

60   Furthermore, some evolutionary forces are difficult to measure even having the

61   data. For example, the mutational burden in a tissue is the combined effect of

62   per-cell mutation and per-cell survival rates, which remain hidden in sequencing

63   data[18,19] (Figure 1). Currently, we cannot independently infer these two for

64   somatic evolution fundamental quantities from single time point sequencing

65   data.

66   Here, we show that multiple bulk or single cell sequencing from the same patient

67   contain recoverable information on these important quantities that can be

68   recovered with evolutionary theory. This allows inferring *in vivo* cell mutation

69   and cell survival rates in tissues of individual humans from single time point

70   sequencing data.

71

72   We draw our inferences by defining and quantifying the distribution of

73   mutational distances amongst multiple samples. We first discuss the required

74   theoretical considerations and derive an analytical expression for the expected

75   distribution of mutational distances from multi-sample sequencing data. We

76   introduce a Bayesian sampling framework based on the mutational distance

3

77 distribution, allowing us to disentangle mutation rates per cell division and cell

78 survival/death rates. We apply this framework to whole genome single cell

79 sequencing data of haematopoiesis and brain tissue and measure both

80 evolutionary parameters during early development. Finally, we utilize multi-

81 sample sequencing data on 16 tumours to infer patient specific evolutionary

82 parameters in human cancers.

83 ## Results

84 **The distribution of mutational distances**

85 All cells in a human tissue must have descended from a most recent common

86 ancestor cell (MRCA) that existed briefly during early development. Similarly, all

87 cells in a sample of a tissue must have descended from a (different) MRCA that

88 was present in that tissue at an earlier time (Figure 1a). Mutations found in all

89 cells of the sample (clonal mutations) were present in this MRCA. If we take

90 multiple samples of the same tissue, we can reconstruct the mutational profile

91 (all mutations carried by a single cell) of multiple ancestral cells (Figure 1a).

92 Typically, these ancestral cells differ in their exact mutational profile between

93 one another, because mutations inevitably accumulate differently in distinct

94 lineages (Figure 1b). We use the differences of the mutational profiles between

95 ancestral cells to construct the distribution of mutational distances. We define a

96 mutational distance as the number of mutations different between any two

97 ancestral cells (Figure 1c). In the language of set theory, if ancestral cell 1 carries

98 a set of mutations A and ancestral cell 2 carries a set of mutations B, then by

99 definition, both cells must have coalesced from an earlier ancestral cell (Figure

100   1a). The mutational profile of this cell is given by the intersection $A \cap B$. This

101   allows us to construct two mutational distances given by

102

103        $$y_1 = |A \setminus (A \cap B)| \qquad \text{and} \qquad y_2 = |B \setminus (A \cap B)|. \qquad (1)$$

104

105   This process can be iterated for increasing combinations of samples per tumour.

106   We now turn to quantitative expressions for the expected distribution of

107   mutational distances $P(y)$. In a single division, the probability of a cell to acquire

108   X novel mutations follows a *Poisson* distribution

109

110        $$P(X) = \frac{(\mu L)^X}{X!} e^{-\mu L}. \qquad (2)$$

111

112   Here, μ is the mutation rate (in units of base pairs per cell division) and L the size

113   of the sequenced genome. Throughout the manuscript, we assume a constant

114   mutation rate and do not consider more punctuated catastrophic events or

115   mutational bursts. Distances between cells of a lineage may arise from more than

116   a single cell division. Instead, double, triple and higher modes of cell division

117   contribute to the distribution of mutational distances of multi-sample samples.

118   In general, a cell accumulates $X_1 + X_2 + \cdots + X_n$ number of novel mutations after

119   n divisions, which is again *Poisson* distributed.

120

121   In addition, we must account for cell death or differentiation, leading to lineage

122   loss. We therefore introduce a probability β of having two surviving lineages

123   after a cell division and a probability $1 - \beta$ of a single surviving lineage (cell

124   death). We can split the total of n cell divisions into r divisions that result in two

5

surviving lineages (branching divisions) and m divisions with only a single surviving lineage (non-branching divisions). The number of non-branching events m is again a random variable, which follows a *Negative Binomial* distribution

$$P(m|r) = \binom{r + m - 1}{r - 1} \beta^r (1 - \beta)^m. \tag{3}$$

The number of mutations acquired between two branching divisions depends jointly on the *Poisson* distributed number of mutations and the *Negative binomial* distributed number of non-branching divisions m. Formally, we can write for the total number of mutations between two branching divisions

$$Y = \sum_{i=1}^{m} X_i. \tag{4}$$

Equation (4) is a random sum of random variables and different combinations of X and m imply the same mutational burden Y within a single cell lineage. Intuitively, a measured mutational burden in a single lineage can result from either many non-branching divisions with a low mutation rate or, alternatively a few non-branching divisions with high mutation rate. The mutational burden of a single sample is insufficient to disentangle per-cell mutation and per-cell survival/death rates.

We therefore turn to the number of mutations different between ancestral cells. Suppose two ancestral cells are separated by r branching divisions. Following

149 from equation (4), we can calculate the probability distribution of the number of

150 acquired mutations $P(y|r)$ after r branching divisions

151

$$P(y|r) = \sum_{i=r}^{\infty} \binom{i-1}{r-1} \beta^r (1-\beta)^{i-r} e^{-i\mu L} \frac{(i\mu L)^y}{y!}. \tag{5}$$

153

154 Here the sum starts at r, as we need to have at least r branching divisions and

155 runs to infinity as in principal infinitely many non-branching divisions can occur

156 (with vanishingly low probability). Finally, we need the expected distribution of

157 branching divisions $P(r)$ in a growing population of cells, which follow from

158 coalescence theory[20-22]. For a growing population, e.g. human tissues during

159 early development or cancer growth, we find

160

$$P(r) = \frac{\exp\left(-\frac{e^{-\beta(r+1)}}{\beta}\right) - \exp\left(-\frac{e^{-\beta r}}{\beta}\right)}{1 - \exp\left(-\frac{e^{-\beta}}{\beta}\right)}. \tag{6}$$

162

163 We provide a more detailed derivation in the Methods. Combining equations (5)

164 and (6) we arrive at the final expression for the expected distribution of

165 mutational distances in an exponentially growing population

166

$$P(y) = \sum_{r=1}^{\infty} \sum_{i=r}^{\infty} P(r) \binom{i-1}{r-1} \beta^r (1-\beta)^{i-r} e^{-i\mu L} \frac{(i\mu L)^y}{y!}. \tag{7}$$

168

169 The two evolutionary parameters of interest here, the mutation rate per cell

170 division μ and the cell survival rate β, disentangle in equation (7). There are

171 approximately four possible regimes for the distribution of mutational distances,

172   discriminated by uni- or multimodality determined by combinations of small or

173   large μ and β. In Figure 2a we show four representative realisations of equation

174   (7). The distribution of mutational distance is unimodal for sufficiently small

175   mutation rate μ (bottom panels in Figure 2a) with a single peak at the mean

176   mutational distance μL. The per-cell survival probability β determines the

177   weight of the distribution towards larger distances. For $\beta = 1$ the distribution is

178   sharply located around the mean mutation rate. However, for smaller β more

179   weight is given to larger distances and the distribution gets a fat tail. The same is

180   true for the case of high mutation rate μ, except the distribution is multi-modal

181   with peaks separated by multiples of the mean mutational distance μL (Figure

182   2a). Again, β determines the weight to higher mutational distances with lower β

183   causing a distribution with a long oscillating tail (top right panel in Figure 2a).

184   Note, the y-axes in Figure 2a correspond to the probabilities of observing certain

185   mutational distances. Lower probabilities require a higher resolution and

186   therefore more sampling to resolve the exact shape of the distribution. In

187   practice, the distribution of mutational distances is easiest to recover from data

188   with low μ and high β (fewest number of tissue samples required), whereas most

189   samples are required for high μ and low β (top right panel in Figure 2a).

190

191   **Computational validation and MCMC inference framework**

192   We implemented stochastic spatial simulations of mutation accumulation in

193   growing tissues using previously published code[23]. Briefly, cell birth and death

194   on a 2- or 3-dimensional grid was simulated using a Gillespie algorithm[24]. During

195   division, cells accumulate a number of new mutations drawn from a *Poisson*

196   distribution. Simulations were stopped when the tissue reached ~1 million cells.

8

197 This allowed us to take samples (either single cells or bulks) and construct all

198 pairwise mutational distances of all ancestral cell lineages detectable in the

199 samples. In Figure 2b we show an example of the mutational distance

200 distribution derived from 200 samples of a stochastic simulation (dots)

201 compared to the theoretical prediction (dashed line).

202 We want to infer the microscopic evolutionary parameters μ and β given a

203 measured distribution of mutational distances. This can be done by Markov

204 chain Monte Carlo methods (MCMC). We implemented a standard Metropolis-

205 Hastings algorithm. In brief, a random pair of parameters μ and β is drawn from

206 uninformed uniform distributions and the likelihood of the model parameters

207 given the data is calculated. The new set of parameters is accepted with a

208 probability proportional to the likelihood ratio of the new and old parameter set

209 (see Methods for more details). This framework recovers the true underlying

210 parameters from stochastic simulations (Figure 2c & Supplementary Figures 17

211 to 21).

212

213 ***In vivo* mutation and cell survival rate inference in healthy**

214 **haematopoiesis during early development**

215 We discuss the *in vivo* mutation accumulation in healthy haematopoiesis during

216 early development as a first application. The cell population is growing and we

217 expect a low mutation rate and a high per-cell survival rate during the

218 development of early haematopoiesis[13,25]. In a recent study, Lee-Six and

219 colleagues[13] sequenced the genome of 89 healthy haematopoietic stem cells of a

220 single 59-year-old male and subsequently constructed the phylogeny of healthy

221 haematopoiesis. They estimated the per-cell mutation rate to be 1.2 mutations

222 per genome per division during early development assuming perfect cell

223 doublings. Using the same data we construct the pairwise mutational distances

224 of all ancestral cells limited to the 20 earliest branching events. The resulting

225 distribution of mutational distances is shown in Figure 3a. We then use the same

226 MCMC framework discussed above to jointly infer the mutation and cell survival

227 rate. The MCMC algorithm rapidly converges to a fixed set of parameters (

228 Supplementary Figure 17). In Figure 3a and 3b we show the posterior parameter

229 distributions after an initial burn in phase of 200 MCMC steps. In agreement with

230 Lee-Six and colleagues, we find a mutation rate of $\mu = 1.14^{+0.12}_{-0.24}$ mutations per

231 genome per division (shown is the medium mutation rate per bp/cell-division

232 and 95% credibility intervals inferred from the MCMC posterior parameter

233 distribution), which corresponds to a mutation rate of $\mu = 3.9 \times 10^{-10}$ base

234 pairs/division (assuming 3×10^9 bp in the human genome). Furthermore, we

235 infer a per-cell survival rate of $\beta = 0.96^{+0.038}_{-0.102}$, independently confirming the

236 original assumption of almost perfect cell doubling during early development[13].

237

238 ***In vivo* mutation and cell survival rate inference in single neurons**

239 **during development**

240 In a recent publication, Bae and colleagues[12] collected single neurons from 3

241 fetuses 15 to 21 weeks post conception. Cells were expanded in culture and the

242 whole genome was sequenced. Here we focus on the case where 14 whole

243 genome sequenced single neurons were available (1 fetus 17w4d after

244 conception). Again, we inferred all pairwise mutational differences, constructed

245 the corresponding distribution of mutational distances (Figure 4a) and used our

246 MCMC framework for joint parameter estimates. The MCMC converges rapidly

247 and we find sharply localised posterior distributions for the mutation and cell

248 survival rate. We infer a median mutation rate of $\mu = 1.37^{+0.1}_{-0.1}$ mutations per

249 genome per division (corresponding to a mutation rate of $\mu = 4.6 \times 10^{-10}$ base

250 pair/division) and a per-cell survival rate of $\beta = 0.998^{+0.002}_{-0.01}$. This inference

251 agrees with Bae and colleagues original estimate of 1.3 mutations per genome

252 per division based on a weighted average of all 3 fetuses, again assuming no cell

253 death during early development. It also agrees with estimates of 1.2 mutations

254 per division from de novo SNVs in familial trios[26]. The almost identical mutation

255 rates in haematopoietic and brain tissue during early development may not be

256 surprising. We would expect the DNA duplication and repair machinery to be

257 stable across tissues during early development. It may even remain stable

258 throughout live, as suggested by the linear rate of mutation accumulation with

259 age across individuals[27-29].

260

261 ***In vivo* mutation and cell survival rates in human tumours**

262

263 We then investigated the per-cell mutation and survival rates in individual

264 tumours. We analysed whole genome or exome sequencing of 131 biopsies from

265 16 tumours comprised of 1 colon adenoma, 7 colon carcinomas, 5 clear cell renal

266 carcinomas and 2 lung squamous cell carcinomas (Table 1). When whole genome

267 sequencing was available, the mutational load was sufficient to apply the

268 inference framework to each chromosome separately (Figure 5 and

269 Supplementary Figures 1-9). The analysis was restricted to regions of

11

270   chromosomes with same copy number profile in all samples of a tumour and

271   inferences were normalised by copy-number and genome content. The

272   resolution to infer the distribution of mutational distances from tumours was

273   lower compared to healthy haematopoiesis or brain during development.

274   Nevertheless, in most cases, the reconstructed distributions recover important

275   features of the theoretical distribution (Supplementary Figures 1-9 and 14). We

276   found that mutation rates per cell division were 4 to 100 times higher in tumours

277   compared to healthy tissue, ranging from 2.91×10^{-9} (bp/division) in the colon

278   adenoma to 53×10^{-9} (bp/division) in one lung squamous cell carcinoma (Table

279   1). Mutation rates differ significantly between patients but not across

280   chromosomes of the same patient (Supplementary Figures 11 and 12). Overall

281   this suggests important differences in mutation accumulation at the single cell

282   level between tumours and is in agreement with recent experimental *in vitro*

283   single cell mutation rate inferences[29,30].

284

285   To further unravel the underlying differences in mutation accumulation during

286   tumour growth, we decomposed somatic mutations into the most prevalent

287   trinucleotide mutational signatures[31] for three whole-genome sequenced

288   colorectal carcinomas and inferred per-cell mutation and per-cell survival rates

289   per signature in each chromosome (Figure 5). Again, we find significant

290   differences between patients (Supplementary Figure 13), further supporting

291   inter-tumour differences of mutation accumulation at the single cell level.

292

293   The inter-patient variation of the cell survival rate was evident. Whereas in

294   healthy tissue almost all cells survive during development, in tumours cell

295   survival rates vary between 0.34 in one MSI+ colon carcinoma up to 0.86 in one

296   renal cell carcinoma (Table 1). Again, per-cell survival rates were overall

297   consistent if inferred from chromosomes of individuals, but varied significantly

298   between patients (Figure 6 and Supplementary Figure 12). The underlying

299   reasons for this inter-patient variation may be cell intrinsic and/or extrinsic, e.g.

300   high cell death due to genomic instability, high mutational burden or immune

301   surveillance. It will be of high interest to further unravel these differences on a

302   patient specific basis in future studies. It should be noted that the inferred cell

303   survival rates are high compared to previous estimates[32,33]. This is a direct

304   consequence of the joint inference of mutation and cell survival rates that was

305   not possible in earlier work.

306

307   ## Discussion

308   Here we presented a framework that allows disentangling the microscopic

309   evolutionary forces of mutation and survival rates per cell division in humans

310   from single time point measurements. Leveraging data on mutations in healthy

311   haematopoiesis[13] and brain tissue[12], we found, in agreement with previous

312   estimates, mutation rates of 1.14 and 1.37 mutations per whole genome per cell

313   division. Mutation rates were 4 to 100 times higher in cancers and showed

314   considerable inter-patient variation.

315

316   The inference framework presented here relies on some assumptions. Mutation

317   and cell survival rates are kept constant trough time and spatial location. We do

318   not consider significant changes in cell fitness during growth and/or spatial

13

319　resource constraints. These limitations are more important for tumour specific

320　inferences and less relevant for healthy tissue. The exact temporal and spatial

321　change of the underlying microscopic evolutionary parameters over the lifetime

322　of an individual tumour remains an open question. In some cases, there is

323　evidence for singular catastrophic events[34] and mutational signatures may

324　change between resection and relapse[35]. However, it will also be important to

325　disentangle mutation and cell population dynamic processes in these cases. A

326　more fine-grained sampling over space and time is needed to better access if and

327　how evolutionary parameters change within tumours. Given the technological

328　advances in single cell genomics[36,37], sequencing of potentially thousands of

329　single cells would lead to significant information gain. This will allow probing

330　potential changes of these evolutionary parameters over time.

331

332　Furthermore, we expect the inter-patient variation of per-cell mutation and

333　survival rates to directly influence clinically important variables, such as the

334　likelihood of pre-existing treatment resistance[38], tumour age and

335　aggressiveness[39]. Measuring microscopic evolutionary forces in humans allows

336　for a mechanistic foundation for precision medicine.

337

338
339
340
341
342

Methods

343

344

Branching distribution in exponentially growing populations

346   To calculate the expected distribution of branching events in an exponentially

347   growing population, we can make use of coalescence theory[20,21]. Note that in

348   coalescence theory one usually uses a backward time convention. If a population

349   grows exponentially with $N(\tau) = e^{\beta\tau}$, coalescence considers backward time

350     $t = -\tau$ such that populations effectively shrink exponentially. The probability of

351   coalescence $P_\zeta(t)$ at time t in an exponentially growing population is given by

352

$$P_\zeta(t) = \frac{1}{N(t)}\prod_{s=0}^{t-1}\left[1 - \frac{1}{N(s)}\right] \approx \frac{e^{\beta t}}{N_0}\exp\left(\frac{1-e^{\beta t}}{\beta N_0}\right), \tag{8}$$

354

355   where $N(t)$ is the size of the growing population at time t. In our case, we are

356   concerned with mutational distances and thus we ask for the distribution of

357   times between coalescence events Δt rather than the distribution of coalescence

358   times t. However, we can directly infer this distribution from equation (8), by

359   rewriting $\Delta t = t_0 - t$ as the time of the initiating cell population at some point in

360   the past. By substituting $t_0 = \log(N_0)/(\beta)$, we have $\Delta t = \frac{\log(N_0)}{\beta} - t$ and we find

361   for the distribution of times between coalescence events

362

$$P(\Delta t) = P\left(\frac{\log(N_0)}{\beta} - t\right) = e^{-\beta\Delta t}\exp\left(\frac{1-N_0 e^{-\beta\Delta t}}{\beta N_0}\right). \tag{9}$$

364

365   This is for large N_0 well approximated by

366

$$P(\Delta t) = e^{-\beta \Delta t} \exp\left(-\frac{e^{-\beta \Delta t}}{\beta}\right). \tag{10}$$

368

369 We show the validity of this approximation in Supplementary Figure 16. The

370 normalized expression holds for all $N_0 \geq 1$. We can discretise this probability

371 density function to derive at the probability for the number of branching

372 divisions r via

373

$$P(r) = \int\limits_{r}^{r+1} d(\Delta t)\, P(\Delta t) = \int\limits_{r}^{r+1} d(\Delta t)\, e^{-\beta \Delta t} \exp\left(-\frac{e^{-\beta \Delta t}}{\beta}\right)$$

374
$$= \exp\left(-\frac{e^{-\beta(r+1)}}{\beta}\right) - \exp\left(-\frac{e^{-\beta r}}{\beta}\right). \tag{11}$$

375

376 As we are interested in positive branch length only, we need to normalise the

377 distribution for non-negative integers such that $1 = \frac{1}{C}\sum_{i=1}^{\infty} P(r = i)$. The

378 normalising factor is $C = 1 - \exp\left(-\frac{e^{-\beta}}{\beta}\right)$, and the distribution of branching

379 divisions r in an exponentially expanding cell population becomes

380

$$P(r) = \frac{\exp\left(-\frac{e^{-\beta(r+1)}}{\beta}\right) - \exp\left(-\frac{e^{-\beta r}}{\beta}\right)}{1 - \exp\left(-\frac{e^{-\beta}}{\beta}\right)}. \tag{12}$$

382

383 Equation (12) together with equation (6) in the main text allows a complete

384 description of the expected distribution of mutational distances in exponentially

385 growing populations. It has to be noted that the coalescence approximation used

16

386 here is based on a deterministic exponential growth function. It is known that

387 such approaches do not always fully capture the full stochasticity especially at

388 small population sizes and birth-death processes often perform better[22]. The

389 individual based computer simulations used here are implementations of the

390 Gillespie algorithm and are exact numerical representations of the underlying

391 stochastic process. However, a further analysis on the stochasticity of the

392 process for small population sizes is warranted.

393

394 **Interpretation of effective survival rate**

395

396 Throughout the manuscript we use the concept of the effective cell survival rate

397 β. One can also formulate cell death with a microscopic perspective given a

398 probability α for a daughter cell to die (or differentiate) after division. Such a

399 probability allows for three outcomes after a cell division: with probability

400 $(1 - \alpha)^2$ both daughter cells survive, with probability $2\alpha(1 - \alpha)$ one daughter

401 cell survives and with probability α^2 both daughter cells die. However, as we are

402 bound to find surviving cell lineages in every possible measure of tumours, none

403 of the observed cell lineages can have gone extinct. Mathematically, this implies

404 that measurement conditions cell division on non-extinction of both daughter

405 cells and we can write

406

407 $\beta \equiv P(successful\ division|\ non\ extinction) = \frac{P(successful\ division\ \&\ non\ extinction)}{P(non\ extinction)}.$

408

409 With the corresponding probabilities α we get

17

410

$$\beta = \frac{(1-\alpha)^2}{1-\alpha^2} = \frac{1-\alpha}{1+\alpha}. \tag{13}$$

411

412

413   We also can rearrange equation (13) to solve for α,

414

$$\alpha = \frac{1-\beta}{1+\beta}. \tag{14}$$

415

416

417   If we interpret α as the probability of random cell death after a division, α must

418   be smaller than $1/2$. If α were larger than $1/2$, tumour populations extinct

419   almost surely after sufficiently many cell divisions. This implies $\beta > 1/3$ for

420   growing populations.

421

422   **Simulations of mutation accumulation in growing tissues**

423   We simulated cell populations of $\sim$1 million cells on a grid with varying birth

424   death and mutation rates using an implementation of the Gillespie algorithm

425   based on code published in[23]. The code is available at

426   https://github.com/sottorivalab/CHESS.cpp. A cell division produces two

427   surviving cells with probability β or one surviving cell with probability $1-\beta$.

428   During each division, each daughter cell inherits the mutations of its parent and

429   in addition accumulates novel mutations. The number of novel mutations is

430   drawn from a *Poisson* distribution with mean μ. During simulations, the

431   mutations for each cell as well as the division history of each cell are recorded.

432

433   We took samples (between 1 and 10k cells per sample) from each simulated

434   tumour. For most inferences, we used maximal distance sampling. Sequencing

435   errors were simulated for each bulk by binomial sampling assuming sequencing

436   depths of 100x, by generating dispersed coverage values for input mutations. We

437   do that by sampling a coverage from a Poisson distribution: $\text{Poisson}(\lambda = Z)$ with

438   mean λ equal to a desired sequencing depth Z. Once we have sampled a depth

439   value k for a mutation, we sample its frequency (number of reads with the

440   variant allele frequency) with a Binomial trail. We use $f \sim \text{Binomial}(n, k)$, where

441   n is the proportion of cells carrying this mutation given all cells sampled in the

442   simulated biopsy. This generates realistic mutation distributions comparable to

443   available genomic sequencing data.

444

445   **Bayesian parameter inference**

446   We use a Markov chain Monte Carlo method (MCMC) to recover the mutational

447   distance μL and the cell survival rate β given a measured distribution of

448   mutational distances. More precisely we implemented a standard Metropolis-

449   Hastings-algorithm following below steps:

450   (i)   Create a new random set of model parameters $\boldsymbol{w}$ given the current set

451   of parameters $\boldsymbol{v}$ from a defined probability density Q, such that

452           $Q(x|y) = Q(y|x)$.

453   (ii)   Calculate the likelihood $L(P(\boldsymbol{w}))$ of the model distribution $P(\boldsymbol{w})$ given

454   the data.

455   (iii)   Calculate the ratio of the new and old likelihood $\rho = L(P(\boldsymbol{w}))/$

456           $L(P(\boldsymbol{v}))$. Accept the new parameter set with probability ρ otherwise

457   reject.

458   (iv)   Repeat

459   In our case the model distribution is given by equation (7) in the main text. To

460   calculate the likelihood of equation (7) given the data, we have to choose a cut off

461   for the infinite sums. However, real data always has a maximum mutational

462   distance. Higher terms of the infinite sums contribute to higher mutational

463   distances. The distribution of interest does not change for a sufficiently high cut

464   off and each observed data set only requires finite many terms. Here we used

465    $r = i = 30$ as upper cut-off, which is a conservative choice. We used uninformed

466   uniform prior distributions for mutational distance μL and the per-cell survival

467   rate β in all cases. Point estimates were extracted as sample medians from the

468   MCMC inferences. The ranges of the uniform priors were adjusted to optimise

469   acceptance rates and computational time. In our implementation, a new set of

470   parameters is relative to the previously accepted parameter set $\boldsymbol{w}_{\text{New}} = \boldsymbol{w}_{Old} +$

471    $\Phi(\boldsymbol{w})$, where Φ is the prior parameter distribution. A typical range used in our

472   inference   scheme   is $\Phi_{\text{uniform}}(\beta) = [-0.06, +0.06]$ and $\Phi_{\text{uniform}}(\mu) =$

473    $[-0.15, +0.15]$. We also tested *Gamma* prior distributions and did not see

474   differences in convergence. One numerical realisation of the Log-Likelihood

475   function is shown in Supplementary Figure 18 and example traces of the MCMC

476   algorithm are shown in Supplementary Figure 17. We also tested the influence of

477   sequencing depth and spatial sampling strategies on the performance of the

478   MCMC inference framework (Supplementary Figures 19 & 20). The code for the

479   MCMC   inference   is   available   at   https://github.com/sottorivalab/MCMC-

480   MutationalDistances-.

481

482   **Mutational signature analysis**

483   For each sample we found the set of signatures (among those signatures

484   reported in CRC) that best explained the totality of mutations in the sample. We

485   did a non-negative regression of the sample's mutations against all the CRC

486   signatures[44] and found those signatures with non-zero coefficients. We took

487   these as the candidate signatures for each sample.

488

489   For each mutation in each sample, we determined the likelihood of the mutation

490   under each of the candidate signatures. We assigned a mutation to a candidate

491   signature where the likelihood under that signature was at least twice that under

492   any other. If there was no such signature, we assigned the mutation to "Other".

493   The method was originally developed in[44] and is based on the R-package

494   "SomaticSignatures"[45]. We did not adjust for differences in nucleotide

495   composition when calculating differences between coding and non-coding

496   regions as we wanted to infer the overall point mutation rate in these regions.

497   Nucleotide dependent mutation rate estimates are shown in Supplementary

498   Figures 10 and 15. Nucleotide composition was adjusted for to calculate the

499   mutation rates of mutational signatures using standard tools[45].

500

501   **Data availability**

502   Sequencing data from healthy haematopoiesis is available from Lee-Six et al.[13],

503   brain data during early development from Bae et al.[12], colorectal cancer data

504   from Cross et al.[40] and Roerink et al.[41], renal cell carcinoma data from Gerlinger

505   et al.[42] and lung carcinoma data from Jamal-Hanjani et al.[43].

506

Acknowledgments

Author Contributions

B.W. and A.S. conceived the study. B.W. and J.C. performed mathematical analysis. B.W., M.J.W., K.C., D.T., J.F.M., G.D.C., D.N. W.C., I. S., W.H. & I.T. contributed to data analysis and simulations. A.S. and T.A.G. supervised the study. B.W., C.P.B., T.A.G. & A.S. wrote the manuscript. All authors read and approved the manuscript.

Competing Interests

The authors declare no competing interests.

Code availability

The code for stochastic simulations of tumour growth is available at https://github.com/sottorivalab/CHESS.cpp. The code for the MCMC inference is available at https://github.com/sottorivalab/MCMC-MutationalDistances-.

531

Figure Legends

533

Figure 1: Multi-region bulk sequencing encodes information on single cell lineages and single cell divisions. **a)** Each of the seven spatially separated tissue samples (in grey) consists of thousands to millions of cells that descended from a single most recent common ancestor (MRCA) cell. The genomic make-up of the single ancestral cell is described by the mutations clonal to the bulk sample. Those appear at high variant allele frequency in the sample (bottom-left panel, in purple). The intersection of mutations in any two bulk MRCA cells corresponds to the genomic profile of another more ancestral cell. This process continues back in time until the MRCA cell of all the sampled cells is reached. **b)** The level of genomic variation within a growing tissue (e.g. development or cancer) is the direct consequence of mutation accumulation during cell divisions, leading to a branching structure. Importantly, the most fundamental parameters, the mutation rate μ and survival rate β of cells per division that drive this process are not directly observable. **c)** Mutation rate per division μ and cell survival rate β leave identifiable fingerprints in the observable patterns of genetic heterogeneity within a tissue. Cell divisions occur in increments of natural numbers and thus the mutational distance between any two ancestral cells is a multiple of the mutation rate μ.

549

Figure 2: Distribution of mutational distances and computational validation. a**)** The quantized nature of cell divisions leads to a characteristic predicted distribution of mutational distances across cell lineages. The shape of the distribution depends on the exact values of μ and β. Roughly four different scenarios of combinations of small and large μ and β are possible. They influence the shape of the distribution differently and thus constructing the distribution of mutational distances allows disentangling the mutation rate μ and cell survival rate β. **b)** Spatial stochastic simulations confirm the ability of mutational distance distributions to disentangle mutation and lineage expansion rates (red area shows the spatial spread of a subclonal mutation). Dots show mutational distances inferred from 200 samples of a single stochastic computer simulation ($\mu = 20, \beta = 0.95$), the dashed line is the predicted distribution based on our equation 7. **c)** A Monte Carlo Markov Chain inference framework based on mutational distance distributions reliably identifies mutation and lineage expansion rates in simulations of spatial and stochastically growing tissues (2 dimensional spatial stochastic simulations, μ:

23

563   Spearman Rho $= 0.98, p = 4 \times 10^{-23}$; β: Spearman Rho $= 0.93, p = 8 \times 10^{-16}$, Relative error:

564   $\eta_\mu = 0.056, \eta_\beta = 0.045$).

565

566   **Figure 3: Per-cell mutation and per-cell survival rate inferences in healthy haematopoiesis**

567   **during development. a)** Mutational distance distribution inferred from 89 whole genome

568   sequenced healthy haematopoietic stem cells from ref[13] (black dots), and best theoretical fit

569   (grey line). Posterior parameter distribution of the MCMC inference for **b)** the mutation rate per

570   cell division ($\mu L = 1.14^{+0.12}_{-0.24}$ mutations per genome per cell division) and **c)** the cell survival rate

571   ($\beta = 0.96^{+0.038}_{-0.102}$). Median point estimates and 95% credibility intervals were taken from the

572   posterior parameter distributions. The inferred mutation rate per cell division agrees with the

573   original estimation of 1.2 mutations per cell division. Furthermore, our joined inference of

574   mutation and cell survival rate confirms the original assumption of no cell death during early

575   development of haematopoiesis.

576

577   **Figure 4: Per-cell mutation and per-cell survival rate inferences in single neurons during**

578   **development a)** Mutational distance distribution inferred from 14 whole genome sequenced

579   single neurons from ref[12] derived from one fetus (17w4d past conception) (black dots), and best

580   theoretical fit (grey line). MCMC inference for **b)** the mutation rate per cell division ($\mu L =$

581   $1.37^{+0.1}_{-0.1}$ mutations per genome per cell division) and **c)** the per-cell survival rate ($\beta =$

582   $0.998^{+0.002}_{-0.01}$). Median point estimates and 95% credibility intervals were taken from the posterior

583   parameter distributions. The inferred mutation rate per cell division agrees with the original

584   estimation of 1.3 mutations per cell division. Furthermore, our joined inference of mutation and

585   cell survival rate confirms the original assumption of no cell death during early brain

586   development.

587

588   **Figure 5: Mutational distance for three colorectal tumours. a-c)** Examples of the mutational

589   distance distribution on single chromosomes for three different colorectal carcinomas for which

590   6, 7 and 9 multi-region bulk samples were sequenced at whole-genome resolution (dots=data,

591   dashed line=theoretical prediction based on MCMC parameter estimates – see insets). The

592   distribution of mutational distances differs between patients, with Patient 04 (MSI –

593   Microsatellite Instability) showing one order of magnitude larger mutational distances. **d-f)** Per-

594   cell mutation rate per chromosome separated by trinucleotide mutational signature. Results are

595   consistent across chromosomes, as expected (Methods). **g-i)** The mean overall mutation rates are

596   ($\mu_{02} = (1.0^{+0.46}_{-0.07}) \times 10^{-8}, \mu_{03} = (2.4^{+0.41}_{-0.19}) \times 10^{-8}$ and $\mu_{04} = (3.1^{+0.35}_{-0.12}) \times 10^{-8}$ bp/division   ,

597   dashed lines), 20 to 60 times higher compared to healthy somatic cells. Patient 04 is MSI+

598   highlighted by signature 6. **j-l)** Estimates of per-cell survival rates per chromosome are

599   consistent   across   chromosomes   of   the   same   patient   (Median:

600   $\beta_{02} = 0.51^{+0.05}_{-0.05}, \beta_{03} = 0.65^{+0.02}_{-0.02}, \beta_{04} = 0.34^{+0.01}_{-0.01}$), but vary considerably between patients

601   (Supplementary Figure 12).

602

Figure 6: Map of per-cell mutation and per-cell survival rates across cancer types. For each of the 16 tumours analysed we plot the per-cell mutation rate versus the per-cell survival rate. Median estimates and 95% credibility intervals for the mutation and cell survival rate are derived from the MCMC inferences as described in the main text. Dashed lines correspond to values of healthy tissue ($\mu_h = 1 \times 10^{-9}, \beta_h = 1/3$). White background corresponds to β values that allow for growing cell populations as $\beta = 1/3$ corresponds to stable (homeostatic) populations. Shaded area describes values of β that would lead to population extinction. Most cancers scatter across the map, indicating extensive inter-patient heterogeneity.

611
612
613
614
615

Tissue type	Sequencing	# Samples	μ $\times 10^{-9}$	β	Source
HSC (development)	Whole Genome	89	0.39	0.96	Lee-Six
Neuron (development)	Whole Genome	14	0.46	0.99	Bae
CRA	Exome	6	2.91	0.46	Cross
CRC (MSS)	Exome	13	30.1	0.84	Cross
CRC (MSS)	Exome	8	12.5	0.43	Cross
CRC (MSS)	Whole Genome	6	24.0	0.65	Cross
CRC (MSS)	Whole Genome	7	10	0.51	Cross
CRC (MSS)	Whole Genome	9	8.9	0.45	Roerink
CRC (MSS)	Whole Genome	9	9.9	0.50	Roerink
CRC (MSI)	Whole Genome	9	30.9	0.34	Cross
CRC (MSI)	Whole Genome	7	17.9	0.47	Roerink
CCRCC	Exome	8	21.7	0.66	Gerlinger
CCRCC	Exome	11	31.2	0.86	Gerlinger
CCRCC	Exome	8	15.8	0.47	Gerlinger
CCRCC	Exome	8	2.3	0.80	Gerlinger
CCRCC	Exome	8	2.1	0.72	Gerlinger
NSCLC	Exome	7	53	0.36	Jamal-Hanjani
NACLC	Exome	7	14	0.59	Jamal-Hanjani

616

Table 1: Data summary and evolutionary parameter inferences. The data of healthy tissue during development was taken from Lee-Six et al[13] and Bae et al[12]. Data on colorectal cancers is from Cross et al[40] and Roerink et al[41], data on renal cell carcinoma from Gerlinger et al[42] and data

620   on lung carcinomas from Jamal-Hanjani et al[43]. Estimates for mutation and cell survival rates are

621   from best MCMC fits based on the distribution of mutational distances.

622

623

References

625

626   1.   Sun, J. *et al.* Clonal dynamics of native haematopoiesis. *Nature* **514,**
627   322–327 (2014).
628   2.   Busch, K. *et al.* Fundamental properties of unperturbed
629   haematopoiesis from stem cells in vivo. *Nature* **518,** 542–546 (2015).
630   3.   Frank, S. A. Somatic evolutionary genomics: Mutations during
631   development cause highly variable genetic mosaicism with risk of
632   cancer and neurodegeneration. *Proceedings of the National Academy*
633   *of Sciences* **107,** 1725–1730 (2010).
634   4.   Nowell, P. The Clonal Evolution of Tumor Cell Populations. *Science*
635   **194,** 23–28 (1976).
636   5.   Greaves, M. & Maley, C. C. Clonal evolution in cancer. *Nature* **481,**
637   306–313 (2012).
638   6.   Swanton, C. Intratumor Heterogeneity: Evolution through Space and
639   Time. *Cancer Research* **72,** 4875–4882 (2012).
640   7.   McGranahan, N. & Swanton, C. Clonal Heterogeneity and Tumor
641   Evolution: Past, Present, and the Future. *Cell* **168,** 613–628 (2017).
642   8.   McGranahan, N. & Swanton, C. Biological and Therapeutic Impact of
643   Intratumor Heterogeneity in Cancer Evolution. *Cancer Cell* **27,** 15–26
644   (2015).
645   9.   Khan, K. H. *et al.* Longitudinal Liquid Biopsy and Mathematical
646   Modeling of Clonal Evolution Forecast Time to Treatment Failure in
647   the PROSPECT-C Phase II Colorectal Cancer Clinical Trial. *Cancer*
648   *Discov* (2018).
649   10.   Martincorena, I. *et al.* High burden and pervasive positive selection of
650   somatic mutations in normal human skin. *Science* **348,** 880–886
651   (2015).
652   11.   Martincorena, I. *et al.* Somatic mutant clones colonize the human
653   esophagus with age. *Science* **57,** eaau3879 (2018).
654   12.   Bae, T. *et al.* Different mutational rates and mechanisms in human
655   cells at pregastrulation and neurogenesis. *Science* **543,** eaan8690
656   (2017).
657   13.   Lee-Six, H. *et al.* Population dynamics of normal human blood
658   inferred from somatic mutations. *Nature* **561,** 473–478 (2018).
659   14.   Werner, B. & Sottoriva, A. Variation of mutational burden in healthy
660   human tissues suggests non-random strand segregation and allows
661   measuring somatic mutation rates. *PLoS Comput Biol* **14,** e1006233

662   (2018).

663   15.   Burrell, R. A., McGranahan, N., Bartek, J. & Swanton, C. The causes and
664   consequences of genetic heterogeneity in cancer evolution. *Nature*
665   **501,** 338–345 (2013).

666   16.   Lynch, M. Evolution of the mutation rate. *Trends in Genetics* **26,** 345–
667   352 (2010).

668   17.   Lynch, M. *et al.* Genetic drift, selection and the evolution of the
669   mutation rate. *Nature Reviews Genetics* **17,** 704–714 (2016).

670   18.   Williams, M. J., Werner, B., Barnes, C. P., Graham, T. A. & Sottoriva, A.
671   Identification of neutral tumor evolution across cancer types. *Nature*
672   *Genetics* **48,** 238–244 (2016).

673   19.   Williams, M. J. *et al.* Quantification of subclonal selection in cancer
674   from bulk sequencing data. *Nature Genetics* **50,** 895–903 (2018).

675   20.   Slatkin, M. & Hudson, R. R. Pairwise Comparisons of Mitochondrial
676   DNA Sequences in Stable and Exponentially Growing Population.
677   *Genetics* **129,** 555–562 (1991).

678   21.   Donelly, P. & Tavare, S. Coalescence and Genealogical Structure under
679   Neutrality. *Annual Reviews of Genetics* **29,** 401–421 (1995).

680   22.   Stadler, T. *et al.* How well can the exponential-growth coalescent
681   approximate constant-rate birth–death population dynamics?
682   *Proceedings of the Royal Society B: Biological Sciences* **282,** 20150420
683   (2015).

684   23.   Chkhaidze, K. *et al.* Spatially constrained tumour growth affects the
685   patterns of clonal selection and neutral drift in cancer genomic data.
686   *PLoS Comput Biol* **15,** e1007243 (2019).

687   24.   Gillespie, D. T. A general method for numerically simulating the
688   stochastic time evolution of coupled chemical reactions. *Journal of*
689   *Computational Physics* **22,** 403–434 (1976).

690   25.   Werner, B. *et al.* Reconstructing the *in vivo* dynamics of
691   hematopoietic stem cells from telomere length distributions. *eLife*
692   **10.7554,** e08687v2 (2015).

693   26.   Rahbari, R. *et al.* Timing, rates and spectra of human germline
694   mutation. *Nature Genetics* **48,** 126–133 (2015).

695   27.   Blokzijl, F. *et al.* Tissue-specific mutation accumulation in human
696   adult stem cells during life. *Nature* **538,** 260–264 (2016).

697   28.   Lodato, M. A. *et al.* Aging and neurodegeneration are associated with
698   increased mutations in single human neurons. *Science* **359,** 555–559
699   (2018).

700   29.   Milholland, B. *et al.* Differences between germline and somatic
701   mutation rates in humans and mice. *Nature Communications* **8,** 1–8
702   (2017).

703   30.   Brody, Y. *et al.* Quantification of somatic mutation flow across
704   individual cell division events by lineage sequencing. *Genome Res.* **28,**
705   1901–1918 (2018).

706   31.   Alexandrov, L. B. *et al.* Clock-like mutational processes in human
707   somatic cells. *Nature Genetics* **47,** 1402–1407 (2015).

708   32.   Bozic, I. *et al.* Accumulation of driver and passenger mutations during
709   tumor progression. *Proceedings of the National Academy of Science*
710   **107,** 18545–18550 (2010).

711   33.   Bozic, I., Gerold, J. M. & Nowak, M. A. Quantifying Clonal and
712   Subclonal Passenger Mutations in Cancer Evolution. *PLoS Comput*
713   *Biol* **12,** e1004731 (2016).
714   34.   Gerstung, M. *et al.* The evolutionary history of 2,658 cancers. *BioRxiv*
715   (2017). doi:10.1101/312041
716   35.   Rubanova, Y. *et al.* TrackSig: reconstructing evolutionary trajectories
717   of mutations in cancer. *BioRxiv* **49,** 1015 (2018).
718   doi:10.1101/260471
719   36.   Chen, X. *et al.* Single-cell analysis at the threshold. *Nature*
720   *Biotechnology* **34,** 1111–1118 (2016).
721   37.   Davis, A. & Navin, N. E. Computing tumor trees from single cells.
722   *Genome Biology* 1–4 (2016). doi:10.1186/s13059-016-0987-z
723   38.   Diaz, L. A. *et al.* The molecular evolution of acquired resistance to
724   targeted EGFR blockade in colorectal cancers. *Nature* **486,** 537–540
725   (2012).
726   39.   Lipinski, K. A. *et al.* Cancer Evolution and the Limits of Predictability
727   in Precision Cancer Medicine. *TRENDS in CANCER* **2,** 49–63 (2016).
728   40.   Cross, W. *et al.* The evolutionary landscape of colorectal
729   tumorigenesis. *Nature Ecology & Evolution* 1–14 (2018).
730   41.   Roerink, S. F. *et al.* Intra-tumour diversification in colorectal cancer at
731   the single-cell level. *Nature* **556**, 457-462 (2018).
732   42.   Gerlinger, M. *et al.* Genomic architecture and evolution of clear cell
733   renal cell carcinomas defined by multiregion sequencing. *Nature*
734   *Genetics* **46,** 225–233 (2014).
735   43.   Jamal-Hanjani, M. *et al.* Tracking the Evolution of Non–Small-Cell
736   Lung Cancer. *New England Journal of Medicine* **376,** 2109–2121
737   (2017).
738   44.   Temko, D., Tomlinson, I. P. M., Severini, S., Schuster-Böckler, B. &
739   Graham, T. A. The effects of mutational processes and selection on
740   driver mutations across cancer types. *Nature Communications* 1–10
741   (2018).
742   45.   Gehring, J. S., Fischer, B., Lawrence, M. & Huber, W.
743   SomaticSignatures: inferring mutational signatures from single-
744   nucleotide variants. *Bioinformatics* **31,** 3673–3675 (2015).
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759

760
761
762
763
764

a) Multi-region bulk sequencing

Single cells ● ●----● Single lineages MRCA → ● ◄ Bulk sample

Bulk mutations (sub-clonal)

Number of mutations

Bulk sample

Single-cell mutations (clonal)

Allelic frequency f

Time t_7 t_6 t_5 t_4 t_3 t_2 t_1 t_0
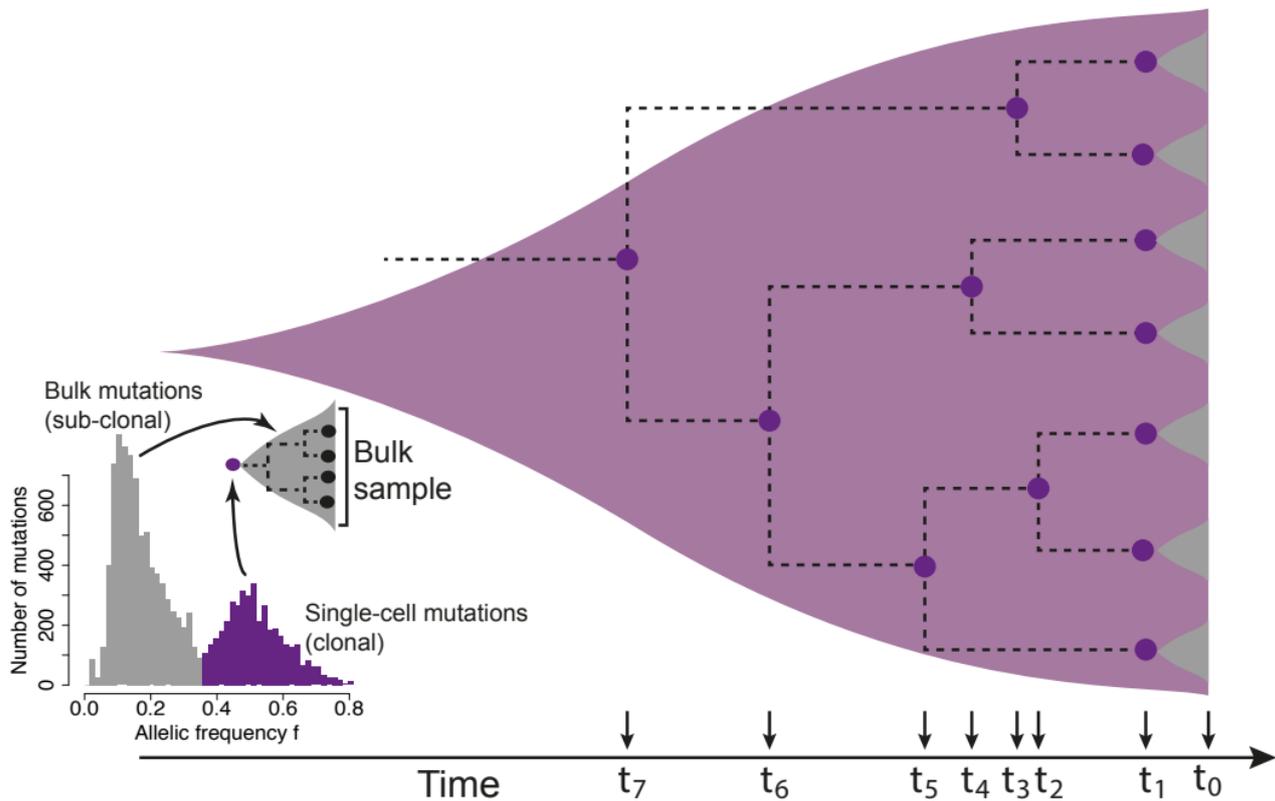
b) Unobservable cell dynamics
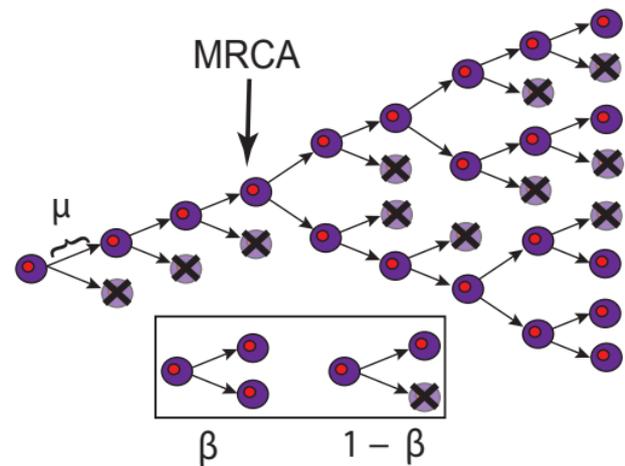
MRCA

μ

β $1 - \beta$

c) Observable phylogeny

3μ 2μ 3μ

3μ 3μ

3μ 2μ

Mutational distance μ μ

μ

a) Distribution of mutational distances

$\boldsymbol{\mu} \uparrow$
$\boldsymbol{\beta} \uparrow$

$\mu = 100$
$\beta = 0.95$

$\boldsymbol{\mu} \uparrow$
$\boldsymbol{\beta} \downarrow$

$\mu = 100$
$\beta = 0.40$

$\boldsymbol{\mu} \downarrow$
$\boldsymbol{\beta} \uparrow$

$\mu = 10$
$\beta = 0.95$

$\boldsymbol{\mu} \downarrow$
$\boldsymbol{\beta} \downarrow$

$\mu = 10$
$\beta = 0.40$
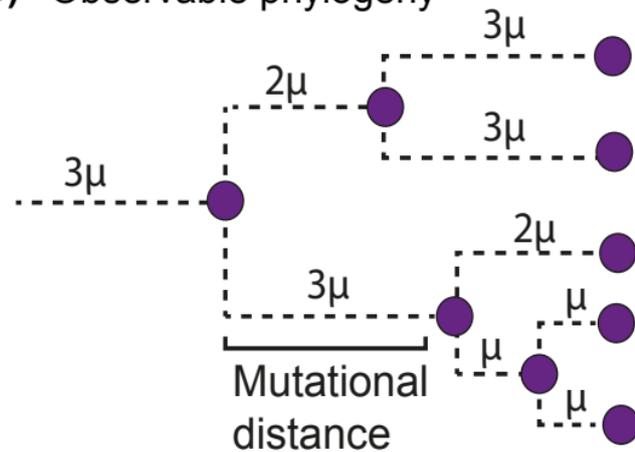
b) Computational validation
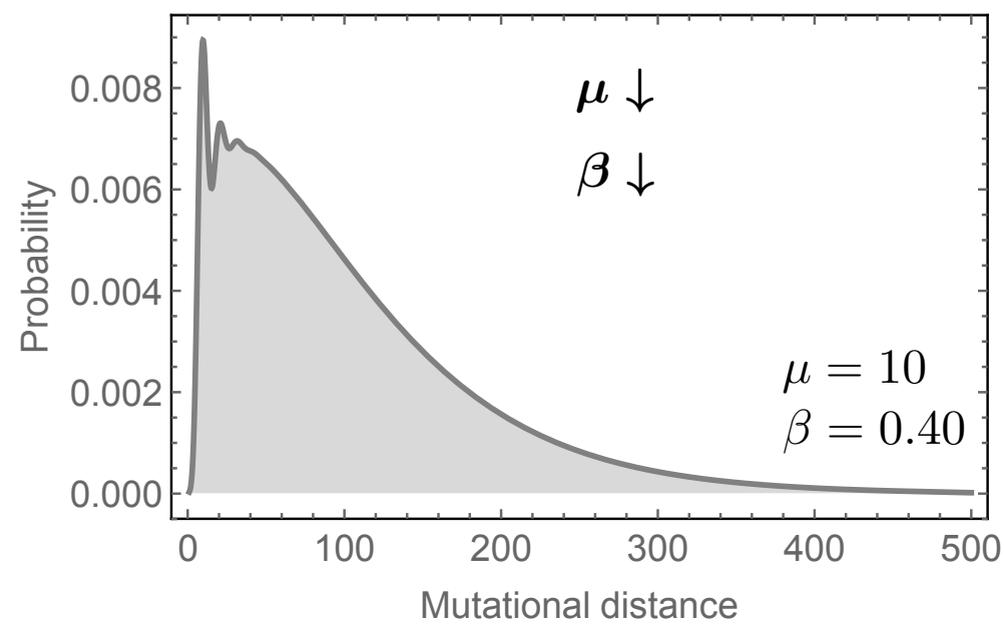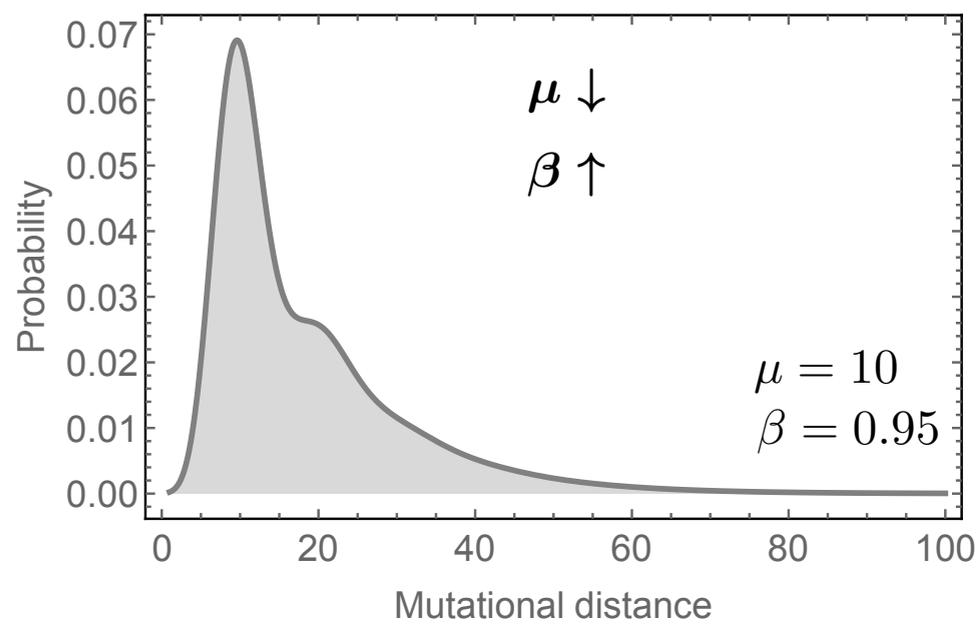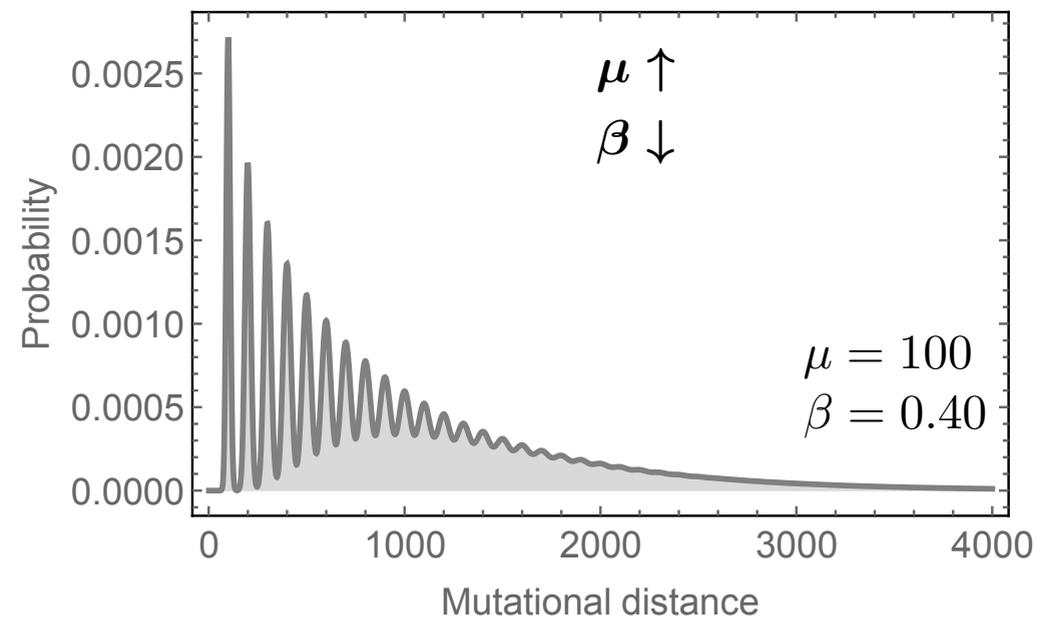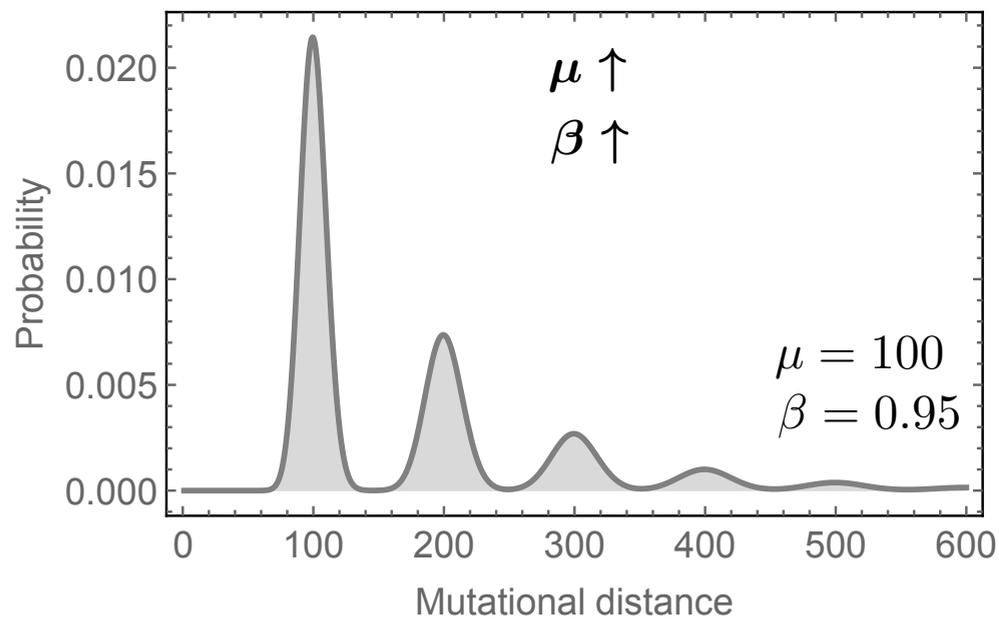
○ Simulated tumour
···· Theoretical prediction

c) MCMC parameter inference

-- True value
○ MCMC inference

$\rho_S = 0.98$
$\eta_\mu = 0.056$

-- True value
○ MCMC inference

$\rho_S = 0.93$
$\eta_\beta = 0.045$

a) Plot showing Probability vs Mutational distance y, with Theory (gray) and Data (open circles) curves. $R^2 = 0.96$, $\ln \overline{L} = -2.56$.

b) Histogram of Inferred mutation rate μL with Count on the y-axis. Inset shows Inferred mutational distance μ vs MCMC step. $\mu = 1.14^{+0.01}_{-0.01}$.

c) Histogram of Inferred per-cell survival rate β with Count on the y-axis. Inset shows Inferred per-cell survival rate β vs MCMC step. $\beta = 0.96^{+0.002}_{-0.004}$.

a)
$R^2 = 0.87$
$\ln \overline{L} = -1.45$

b)
$\mu = 1.37^{+0.1}_{-0.1}$

c)
$\beta = 0.998^{+0.002}_{-0.01}$