

On the patterns of genetic intra-tumour heterogeneity before and after treatment

Alexander Stein and Benjamin Werner

Evolutionary Dynamics Group, Centre for Cancer Genomics and Computational Biology, Barts Cancer Centre, Queen Mary University of London, Charterhouse Square, London, United Kingdom EC1M 6BQ.

Correspondence should be addressed to a.stein@qmul.ac.uk and b.werner@qmul.ac.uk

October 10, 2024

Abstract

Genetic intra-tumour heterogeneity (gITH) is a universal property of all cancers. It emerges from the interplay of cell division, mutation accumulation and selection with important implications for the evolution of treatment resistance. Theoretical and data-driven approaches extensively studied gITH in ageing somatic tissues or cancers at detection. Yet, the expected patterns of gITH during and after treatment are less well understood. Here, we use stochastic birth-death processes to investigate the expected patterns of gITH across different treatment scenarios. We consider homogeneous treatment response with shrinking, growing and stable disease, and follow up investigating heterogeneous treatment response with sensitive and resistant cell types. We derive analytic expressions for the site frequency spectrum, the total mutational burden and the single-cell mutational burden distribution that we validate with computer simulations. We find that the SFS after homogeneous treatment response retains its characteristic power-law tail, while emergent resistant clones cause peaks corresponding to their sizes. The frequency of the largest resistant clone is subdominant and independent of the population size at detection, whereas the relative total number of resistant cells increases with detection size. Furthermore, the growth dynamics under treatment determine whether the total mutational burden is dominated by preexisting or newly acquired mutations, suggesting different possible treatment strategies.

Introduction

Every cell in a human body accumulates thousands of unique mutations with age [1, 2]. From this genetic heterogeneity, pre-malignant cells may arise and eventually evolve into an expanding cancerous cell population. As the disease progresses, cancer cells continue to accumulate mutations, subject to genetic drift and selection, leading to a complex pattern of genetic intra-tumor heterogeneity (gITH) [3, 4].

Mathematical and computational modelling applied to sequencing data of normal and cancerous tissues has led to a better understanding of the emerging patterns of clonal diversity [5, 6, 7, 8, 9]. It is clear that gITH informs on past evolutionary paths and at least partially determines future evolution, e.g. the

risk of resistance or relapse in treated tumours [10, 11]. While patterns of gITH in both normal and cancerous cell populations prior to treatment are well studied, less is known about the dynamics of gITH in response to treatment.

In this study, we use birth-death processes with random accumulation of mutations to investigate gITH in cancers before and after treatment. We investigate three different measures of gITH: the site frequency spectrum (SFS), the total mutational burden (tMB) and the single cell mutational burden (scMB). We first derive analytic expressions for populations at detection and homogeneous treatment responses. We then investigate effects of emerging resistance and the corresponding heterogeneous treatment response on possible measures of gITH and their dependence on the detailed dynamics and timing of treatment.

Theoretical framework

We model the cancer cell population using a birth-death process with constant birth and death rates. Starting from a single ancestor cell, cells either divide or die giving rise to a phylogeny (fig. 1A). Initially, the birth rate is larger than the death rate ($b_1 > d_1$) leading to a growing population (phase 1). Once the cancer is detected at time t_d with size N_d , treatment is administered leading to new birth and death rates b_2 and d_2 (phase 2). Treatment is applied for time t_f after which we have the final population size N_f . Assuming that all cells have the same birth and death rates, there are three qualitatively different scenarios for the growth during treatment: The cancer decreases ($b_2 < d_2$ and $N_d > N_f$), remains approximately constant ($b_2 = d_2$ and $N_d \approx N_f$) or continues growing ($b_2 > d_2$ and $N_d < N_f$) as illustrated in fig. 1B.

We investigate the emergence of resistance by dividing the population into treatment-sensitive and treatment-resistant cell types. We start with a single sensitive progenitor cell. When a sensitive cell divides, each daughter cells acquires resistance mutations with rate ν . Once a cell obtains a resistance mutations, it will be inherited by its offspring and the resistant clone is given new birth and death rates (fig. 1C). In total, we have 8 growth parameters: the birth and death rate of sensitive cells before treatment $b_{s,1}, d_{s,1}$ and during treatment $b_{s,2}, d_{s,2}$, and the same set of parameters again for resistant cells $b_{r,1}, d_{r,1}$ and $b_{r,2}, d_{r,2}$. We assume that resistant cells behave neutral in the absence of the drug ($b_{s,1} = b_{r,1}$ and $d_{s,1} = d_{r,1}$), sensitive cells decrease in number in the presence of treatment ($b_{s,2} < d_{s,2}$) while resistant cells continue to increase ($b_{r,2} > d_{r,2}$).

To model the accumulation of neutral mutations, we assume that in each division, each cell obtains μ mutations, where μ is a random number drawn from a Poisson distribution with mean m . Since each division comes with 2 daughter cells, there are on average $2m$ mutations generated per division. We assume infinite-sites such that each mutation is treated unique and mutation reversion is neglected.

We denote the stochastic total number of cells by $Z_0(t)$, its expectation by $N(t) = E[Z_0(t)]$ and add a tilde when conditioned on survival $\tilde{N}(t) = E[Z_0(t)|Z_0(t) > 0]$. Next to the total population, subpopulations will follow stochastic growth too. We write $p(a \rightarrow n, t)$ for the probability that a (sub-)population has size n after time t when starting from size a . In its most general form, $p(a \rightarrow n, t)$ is a sum over $\min(a, n)$ terms that yield little analytic insight and is computationally expensive [12, 13]. Therefore, we synthesized exact and approximate results to balance computational efficiency and accuracy (Appendix A and *Methods*).

Some cancers, even if they have positive fitness, go extinct. If we observe a cancer it means that it survived genetic drift, which is mathematically incorporated by conditioning on survival. Given birth

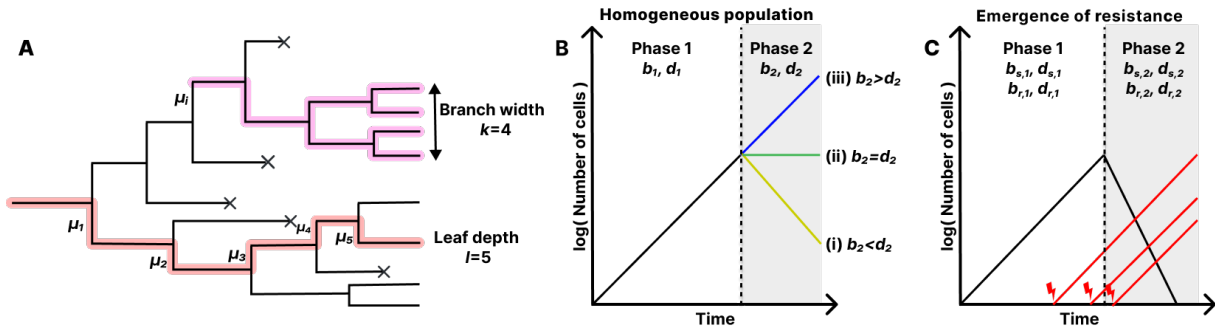


Figure 1: Model illustration. **A.** Phylogeny with highlighted branch (purple) and highlighted path from leaf to root (red). **B.** Growth of homogeneous population: In phase 1, the population grows exponentially. In phase 2, the population (i) decreases, (ii) remains constant or (iii) continues increasing. **C.** Growth with emergence of resistance: In phase 1, the cancer grows exponentially and resistance mutations lead to a treatment-resistant subpopulation. In phase 2, the sensitive population decreases while the resistant population continues increasing.

rate b and death rate d , we denote the probability for a population of size 1 to go extinct by time t with $\alpha(t)$. The extinction probability for a population of size a is then

$$p(a \rightarrow 0, t) = \alpha^a(t) \quad \text{with} \quad \begin{cases} \alpha(t) = \frac{de^{(b-d)t} - d}{be^{(b-d)t} - d} & \text{if } b \neq d \\ \alpha(t) = \frac{bt}{1-bt} & \text{if } b = d \end{cases} \quad (1)$$

We can then write the average growth conditioned on survival by

$$\tilde{N}(t) = \frac{a}{1 - \alpha^a(t)} e^{(b-d)t} \quad (2)$$

whereas the average growth including extinct trajectories is $N(t) = ae^{(b-d)t}$ [14, 15].

Observations are made either at a fixed size or a fixed time. If we stop the stochastic process once the population reaches a fixed time, we have variable population size. The random size is fully characterized by $p(a \rightarrow n, t)$. Vice versa, if we stop at a fixed size, time will be a random variable. Conditioned on survival, the random time T_N to grow to size N for the first time when starting from a single cell is known to approximately follow a Gumbel distribution (ref. [15] and Appendix B) with expectation

$$E[T_N] = \frac{1}{b-d} \left(\ln \left(\frac{b-d}{b} N \right) + \gamma \right), \quad (3)$$

where $\gamma \approx 0.5772$ is the Euler-Mascheroni constant. Both conditions have relevance in biological systems: In clinic, the starting time of the tumor is unknown and we observe the disease once it reaches detectable size. In contrast, in laboratory set-ups, experiments start and end at predefined times. It is possible to switch the conditioning of predictions between fixed-time and fixed-size (Appendix C). Using appropriate mapping between time and size, we find that the difference in the predictions between the two conditions is often small, in particular for large populations that behave nearly deterministically.

We investigate gITH of the cell population in terms of three summary statistics. First, we look at the site frequency spectrum (SFS). The SFS distributes mutations into classes \mathcal{S}_k according to the number of cells k they occur in. The sizes of the classes $S_k = |\mathcal{S}_k|$ build a distribution $\{S_k\}_{k=1}^N$ that we call the

SFS. A site frequency corresponds to the abundance k of a mutation that resides at a particular site. Second, we study the total mutational burden (tMB) that we define as the number of unique mutations in the entire cancer and denote by B . The tMB counts mutations that are either clonal or subclonal and is related to the SFS by $B = \sum_{k=1}^N S_k$. Third and last, we compute the single cell mutational burden (scMB) distribution. The scMB is the number of mutations in a single cell. The scMB distribution, $\{M_j\}_{j=1}^{\infty}$, is defined by the number of cells, M_j , with j mutations. In this sense, the scMB distribution distributes cells into classes according to the amount of mutations they carry.

Summary statistics can be computed from a single realization, and generally differ between different realizations. From this perspective, the SFS $\{S_k\}_{k=1}^N$ consists of N random variables, the total mutational burden B consists of 1 random variable, and the scMB distribution $\{M_j\}_{j=1}^{\infty}$ consists of infinitely many random variables. Mathematically speaking, the SFS and the scMB distribution are random distributions, i.e. distributions in which each value is a random variable. In the following, we are mainly interested in their expectations but also present some results of the variance between realisations. The latter corresponds to inter-tumour heterogeneity instead of intra-tumour heterogeneity.

The site frequency spectrum

We link the SFS to the phylogenetic tree and derive an analytic formula for the expected SFS in homogeneous populations. We then apply this formula to obtain the expected SFS at the time of detection and after treatment accompanied with discussions on the properties and interpretations.

Construction of the site frequency spectrum

Inspecting the phylogenetic tree in fig. 1A, we see that the number of mutations S_k that occur in k cells is obtained by summing over all mutations with branch width k . If we write the number of branches with width k as W_k , label the branches with width k by j and the number of mutations on branch j by $\mu_{k,j}$, we have

$$S_k = \sum_{j=1}^{W_k} \mu_{k,j}. \quad (4)$$

Since we assume infinite sites, the number of mutations $\mu_{k,j}$ are independent and by assumption follow a Poisson distribution, $\mu_{k,j} \sim \text{Poiss}(m)$. We call the collection $\{W_k\}_{k=1}^N$ the branch width distribution (BWD), which is a summary statistic of the tree topology that emerged from the stochastic growth of the underlying cell population. Since eqn. (4) is a random sum, we can write the expectation and variance as

$$\begin{aligned} E[S_k] &= E[W_k]E[\mu_{k,1}] = E[W_k]m \\ \text{Var}[S_k] &= \text{Var}[\mu_{k,1}]E[W_k] + E^2[\mu_{k,1}]\text{Var}[W_k] = mE[W_k] + m^2\text{Var}[W_k], \end{aligned} \quad (5)$$

which disentangles the effects of stochastic growth and mutation accumulation. As we shall see in further proceedings, it is useful to separate between mutations that are preexisting to time $t' = 0$ or newly emerging in a time interval $[0, t]$. We denote the number of newly emerging mutations found in k cells by $S_k^{(new)}$, and the number of preexisting mutations by $S_k^{(pre)}$. The total number of mutations S_k found in k cells is then $S_k^{(tot)} = S_k^{(new)} + S_k^{(pre)}$. The following theorem characterizes the expected SFS of both contributions.

Theorem 1. Suppose the expected total population size over the time interval $[0, t]$ is $N(t')$ with initial size N_0 , and the probability to grow from size x to y in time t is $p(x \rightarrow y, t)$. Then the expected SFS of newly emerging mutations conditioned on survival and fixed time t is

$$E[S_k^{(new)}] = \frac{1}{1 - p(N_0 \rightarrow 0, t)} \int_0^t 2mbN(t')p(1 \rightarrow k, t - t')dt'. \quad (6)$$

Suppose the expected initial SFS at time $t' = 0$ is $E[S_{k'}^{(init)}]$, then the expected SFS of pre-existing mutations conditioned on survival and fixed time t is

$$E[S_k^{(pre)}] = \frac{1}{1 - p(N_0 \rightarrow 0, t)} \sum_{k'=1}^{N_0} E[S_{k'}^{(init)}]p(k' \rightarrow k, t). \quad (7)$$

Proof. Appendix D. □

For the proof of eqn. (6), we follow the derivation of ref. [16] who derived the result for $N_0 = 1$. In both expressions, the first term is due to condition on survival of the entire population and can be dropped should one be interested in the expectation including extinction. Intuitively, eqn. (6) counts the number of mutations generated at time t' , conditions them to grow to size k and then sums over all mutations generated in the considered time interval. Eqn. (7) has similarities to a sampling formula. There are $E[S_{k'}^{(init)}]$ mutations of size k' that are "sampled" to size k with probability $p(k' \rightarrow k, t)$. We do not strictly talk about sampling since we allow $k > k'$.

The site frequency spectrum at detection

The expected SFS at detection, $E[S_k^{(det)}]$, is computed by applying theorem 1 over the time interval $[0, t_d]$ starting from a single cell with no mutations (Appendix E). Following ref. [17] and [16], the resulting integral can be simplified to

$$\begin{aligned} E[S_k^{(det)}] &= 2mN_d \int_0^{1-1/N_d} \left(1 - \frac{d_1}{b_1}y\right)^{-1} (1-y)y^{k-1} dy \\ &\rightarrow 2mN_d \sum_{l=0}^{\infty} \frac{\left(\frac{d_1}{b_1}\right)^l}{(k+l)(k+l+1)} \quad \text{for } N_d \rightarrow \infty \\ &\rightarrow 2mN_d \frac{b_1}{b_1 - d_1} \frac{1}{k(k+1)} \quad \text{for } k \rightarrow \infty. \end{aligned} \quad (8)$$

Here, $N_d = \tilde{N}(t_d)$ is the expected population size at time t_d conditioned on survival. Although not the focus of this study, it is straightforward to include mutations accumulated before expansion that will be clonal in the entire cancer (Appendix E).

The prediction in eqn. (8) perfectly aligns with the average SFS obtained through computer simulations (fig. 2A). For large k , we have $E[S_k^{(det)}] \sim k^{-2}$, which is a well-known characteristic of the SFS in exponentially growing populations observable in bulk sequencing data of sufficient coverage [18, 5, 6]. Noticeably, single realisations of the stochastic process deviate significantly from the expected scaling for

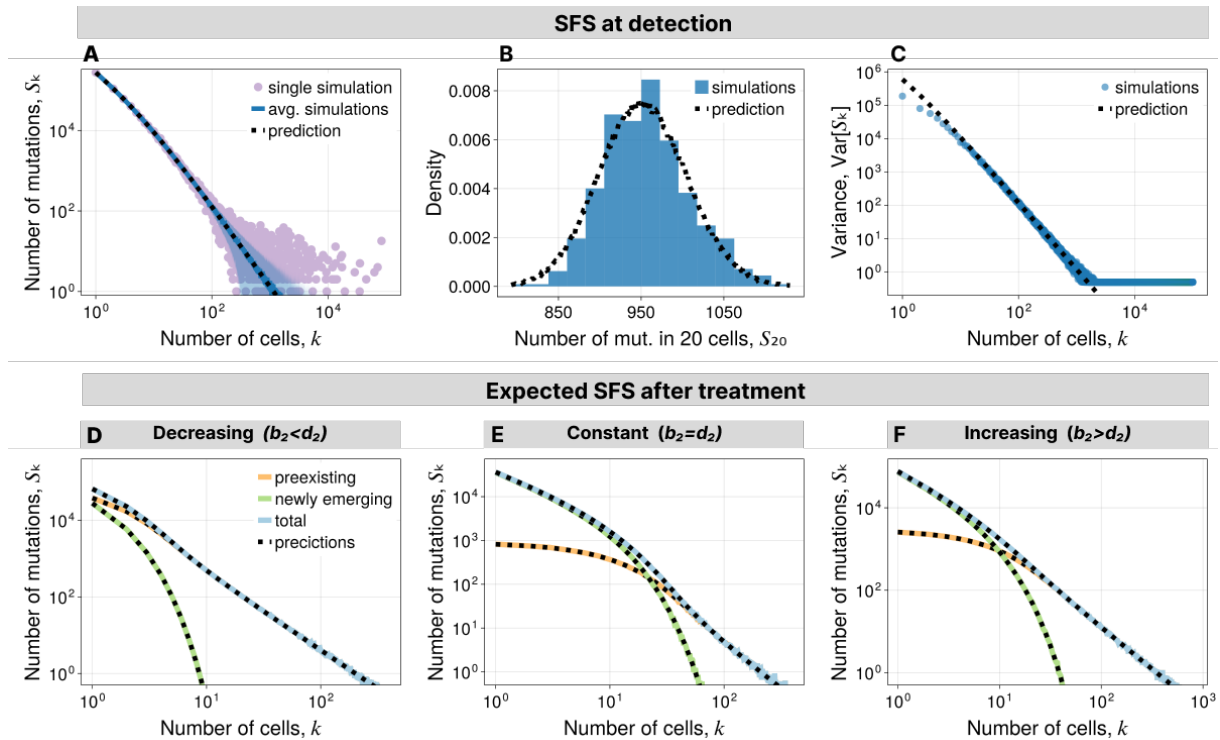


Figure 2: SFS in homogeneous populations. **A.-C.** SFS at the time of detection. **A.** Single simulation, average over many simulations with 1-sigma interval, prediction from eqn. (8). **B.** Distribution of S_{20} from simulations and prediction using a compound Poisson distribution. **C.** Variance in S_k from simulations and prediction from eqn. (10). **D.-F.** SFS after treatment. Average over many simulations and prediction from theorem 1 with explicit formulas in Appendix F. **D.** For decreasing population. **E.** For constant population. **F.** For increasing population. Parameters are in table S1.

large site frequencies (2A, dots). These are mutations occurring in the first few generations of growth and have a high degree of stochasticity.

Knowing the probability density $f_T(t|N'_d)$ for the time t to reach fixed size N'_d for the first time (Appendix B), we can change the conditioning from fixed time to fixed size (Appendix C). To avoid convergence issues, we consider the normalized SFS that we denote with a tilde $E[\tilde{S}_k|N'_d]$. We change the conditioning by computing

$$E[\tilde{S}_k|N'_d] = \int_0^\infty E[\tilde{S}_k|t] f_T(t|N'_d) dt, \quad (9)$$

Using the large population size solution $N_d \rightarrow \infty$ in eqn. (8), we show that the fixed-size expectation coincides with the fixed-time solution, $E[\tilde{S}_k|N'_d] = E[\tilde{S}_k|T = t]$ (Appendix E). This is in agreement with ref. [16] in which it is shown that the fixed-size expectation can be approximated by the fixed-time expectation for large N_d .

The expected SFS $E[S_k]$ after exponential growth is well studied and routinely observed in cancer genetic data. However, the variance $Var[S_k]$ of the SFS under exponential growth remains to the best of our knowledge unknown. In the pure-birth process, and within the fixed-size limit, we observe in simulations that W_k are approximately Poisson distributed such that S_k is described by a compound Poisson process

(fig. 2B). This suggests a heuristic expression for the variance

$$\text{Var}[S_k^{(det)}] = (1 + m)E[S_k^{(det)}], \quad (10)$$

that shows good agreement with simulations of the pure-birth process on fixed size (fig. 2C). There are deviations for very small site frequencies $k = 1, 2, 3, \dots$. To better understand this, we point out that when conditioning on growth to size N_d , there are exactly N_d branches, which are simultaneously leaves such that $\text{Var}[W_1^{(det)}] = 0$ and thus $\text{Var}[S_1^{(det)}] = mE[W_1^{(det)}]$.

Interestingly, eqn. (10) suggests that the variance in the SFS carries information on the mutation rate m . This can be leveraged to infer the mutation rate m from sequencing data, possibly at different stages of tumour growth.

The site frequency spectrum after treatment

We compute the SFS after homogeneous treatment response, and distinguish between decreasing, constant and continued increasing cell populations (fig. 1B). To apply theorem 1, we assume that the cancer has initially fixed size N'_d and expected SFS at detection, $E[S_k^{(init)}] = E[S_k^{(det)}]$. For analytical purposes, we approximate the fixed-size expected SFS with the fixed-time expected SFS in eqn. (8). Together with the expected growth $N(t') = N'_d e^{(b_2 - d_2)t'}$, the treatment time interval $[0, t_f]$ and probabilities $p(a \rightarrow n, t)$ (Appendix A), we can readily compute the expected SFS of newly emerging and preexisting mutations. We validate the general analytic predictions computed in Appendix F with simulations showing perfect alignment (fig. 2D-F). We proceed with highlighting some approximations and observations.

First, we observe that newly emerging mutations are restricted to small site frequencies, and thus the tail is determined by preexisting mutations. In shrinking populations, newly emerging mutations are prone to extinction and only survive short amounts of time and thus have almost negligible impact on the SFS. In contrast, for constant or increasing tumour populations, newly emerging mutations dominate the SFS at low frequencies whereas preexisting mutations continue to dominate the SFS at high frequencies (fig. 2D-F).

For decreasing populations, we obtain a simple formula in the case of the pure-death process. To show this, we adapt the sampling process from theorem 3 in ref. [15] (Appendix G). Taking $b_2 = 0$ and sufficiently large k , we find

$$E[S_k^{(tot)}] = E[S_k^{(pre)}] \approx 2mN_f \frac{b_1}{b_1 - d_1} \frac{1}{k^2 + k}, \quad (11)$$

where $N_f = \tilde{N}(t_f)$ is the expected population size at time t_f conditioned on survival. There are no newly emerging mutations during treatment since $b_2 = 0$, and the k^{-2} tail of neutral mutations from the SFS at detection is maintained.

As $b_2 \rightarrow d_2$, the importance of genetic drift increases, and if $b_2 = d_2$, the SFS of newly emerging mutations takes the form

$$E[S_k^{(new)}] = 2mN_f \left(\frac{b_2 t_f}{b_2 t_f + 1} \right)^k \frac{1}{k}. \quad (12)$$

As $t_f \rightarrow \infty$, all preexisting mutations will be either extinct or fixed in the population such that only newly emerging mutations contribute to gITH. From eqn. (12), we can see that the SFS scales with k^{-1} in agreement with previously reported results for the SFS in constant populations at equilibrium [19, 20, 16].

Testing homogeneity

Theorem 1 provides us predictions for the SFS at detection and after treatment assuming that all cancer cells have the same growth parameters. In the spirit of neutral theory [21, 22], we can take this prediction as null hypothesis of homogeneity that can be compared to sequencing data [18]. If the population is homogeneous, we expect a good fit between the data and the prediction. If the population is heterogeneous, we expect to see deviations between data and prediction.

Sequencing data are noisy and mutations that occur in less than 1% of the cell population cannot reliably be observed. This restricts our test on the observable mutations occurring in at least 1% of the population ($\frac{k}{N_f} > 0.01$). However, considering large site frequencies dramatically simplifies the homogeneity test in the treatment scenario. From simulated data, we observe that the SFS in mutations above 1% of the population follow the power law $S_k \sim k^{-2}$ known from exponentially growing tumors (fig. S2). We support this observations by arguing that the observable SFS for biologically feasible parameters consists of large site frequencies that behave nearly deterministic (Appendix G). Remarkably, this holds true for all three considered cases of homogeneous growth, although deviations may occur for small population that are kept constant for long times [9]. We conclude that testing homogeneity can be achieved by evaluating whether the SFS follows the expected scaling $E[S_k^{(tot)}] \sim k^{-2}$ as demonstrated in prior analyses of the SFS at detection [18, 5].

The total mutational burden

Analogous to the SFS, we link the tMB to the topology of the phylogenetic tree, and provide a general formula to compute the expected tMB, which we then apply to homogeneous cancer cell populations before and after treatment. We connect the tMB to the risk of resistance and discuss its relevance for therapeutic interventions.

Construction of the total mutational burden

Looking at the phylogenetic tree, we denote the number of nodes including leaves with at least one living descendent by R , and label them with $i = 1, 2, 3, \dots, R$. Each node comes with μ_i unique mutations such that the tMB is

$$B = \sum_{i=1}^R \mu_i. \quad (13)$$

Here, R and μ_i are random numbers and $\mu_i \sim \text{Pois}(m)$ are independent such that

$$\begin{aligned} E[B] &= E[R]E[\mu_1] = E[R]m \\ \text{Var}[B] &= \text{Var}[\mu_1]E[R] + E^2[\mu_1]\text{Var}[R] = mE[R] + m^2\text{Var}[R]. \end{aligned} \quad (14)$$

We separate again between newly emerging and preexisting mutations such that the total tMB is the sum of the two contributions, $B^{(tot)} = B^{(new)} + B^{(pre)}$ that are characterized with the following theorem.

Theorem 2. *Suppose the expected total population size over the time interval $[0, t]$ is $N(t')$ with initial size N_0 , and the probability to grow from size x to y in time t is $p(x \rightarrow y, t)$. Then the expected tMB of*

newly emerging mutations conditioned on survival and time t is

$$E[B^{(new)}] = \frac{1}{1 - p(N_0 \rightarrow 0, t)} \int_0^t 2mbN(t')(1 - p(1 \rightarrow 0, t - t'))dt'. \quad (15)$$

Suppose the expected initial SFS at time $t' = 0$ is $E[S_{k'}^{(init)}]$, then the expected tMB of preexisting mutations conditioned on survival and time t is

$$E[B^{(pre)}] = \frac{1}{1 - p(N_0 \rightarrow 0, t)} \sum_{k'=1}^{N_0} E[S_{k'}^{(init)}] (1 - p(k' \rightarrow 0, t)). \quad (16)$$

Proof. Appendix H. □

The expected tMB can be obtained from theorem 1 by noting that $E[B] = \sum_k^\infty E[S_k]$. However, the expressions for the expected SFS after treatment are complicated and theorem 2 then provides more straightforward calculations.

The total mutational burden at detection

Following theorem 2, the expected tMB at detection is

$$\begin{aligned} E[B^{(det)}] &= 2mN_d \frac{b_1}{d_1} \ln \left(\frac{b_1}{b_1 - d_1} - \frac{1}{(1 - \alpha(t_d))} \frac{1}{N_d} \frac{d_1}{b_1 - d_1} \right) \\ &\sim 2mN_d \frac{b_1}{d_1} \ln \left(\frac{b_1}{b_1 - d_1} \right) \quad \text{for } N_d \rightarrow \infty, \end{aligned} \quad (17)$$

where $N_d = \tilde{N}(t_d)$ is the expected size at fixed time t_d . The expression coincides with the expression for the tMB in ref. [16], in which the expected tMB was computed by summing over the SFS. We focus on mutations that occur during growth and treatment. Should one be interested in mutations preexisting in the ancestor cell, those will be shared by the entire population and can be added as a constant. Detailed calculations are in Appendix I.

From eqn. (17), it is apparent that increasing the death rate increases the tMB if measured at the same population size. This is expected, as an increased death rate implies more cell divisions to reach the same population size. However, a considerable fraction of mutations are at low frequencies and will go extinct even if the cell population continues increasing.

In the pure-birth process, the tMB is simply $E[B^{(det)}] \approx 2mN_d$. To grow from 1 cell to N_d cells, the population undergoes $N_d - 1$ divisions, where each daughter cells obtains a random number of mutations. Thus, the tMB is the sum of $2 \times (N_d - 1)$ independent random variables that are Poisson distributed with mean m . Approximating $N_d - 1 \approx N_d$, we have $B^{(det)} \sim \text{Poiss}(2mN_d)$, which agrees with stochastic simulations conditioned on fixed size (fig. 3A). For observations made on a fixed time, the varying number of cells yields an additional layer of stochasticity. Also, if $d_1 > 0$, there are different growth trajectories to reach size N_d . From eqn. (14), it is clear that in either of the two cases $\text{Var}[B^{(det)}] > E[B^{(det)}]$ because the variance in division counts is positive, $\text{Var}[R^{(det)}] > 0$, implying an increased genetic heterogeneity between tumours (fig. 3B-C).

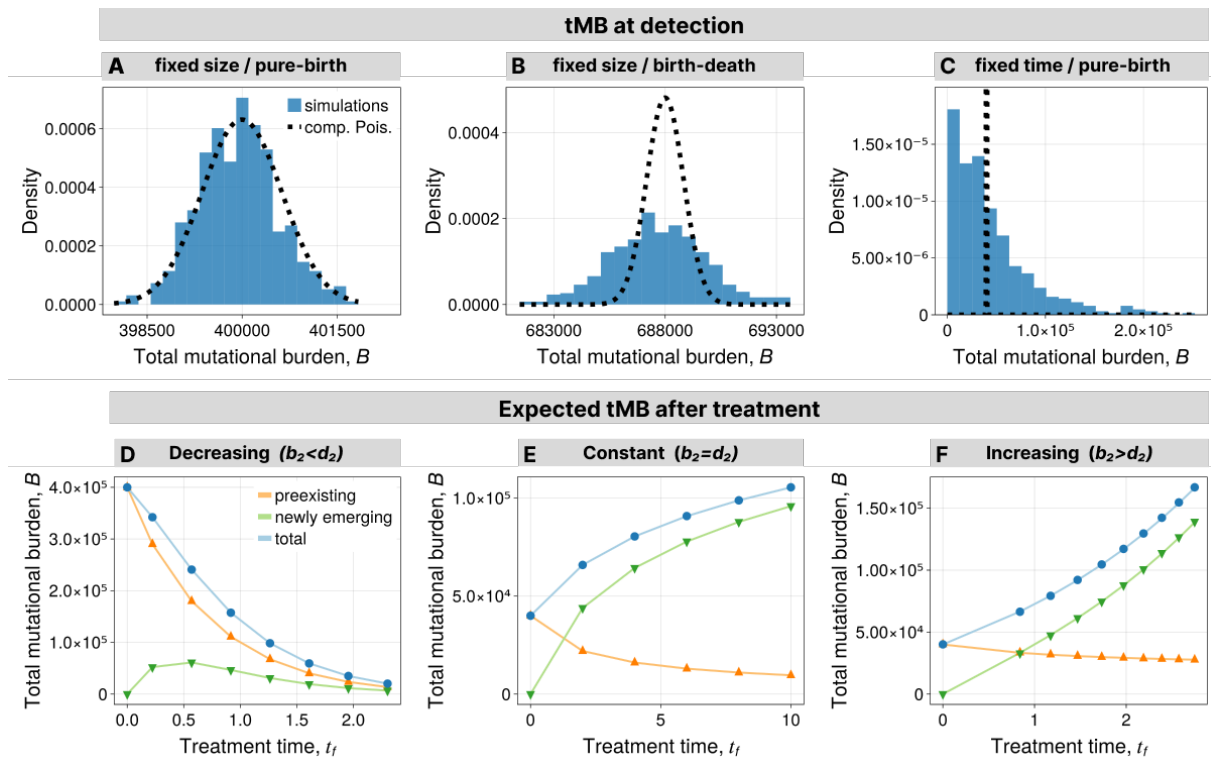


Figure 3: tMB in homogeneous populations. **A.-C.** Distribution of tMB at detection from simulations compared to compound Poisson distribution with mean given by eqn. (17). **A.** Pure-birth process conditioned on fixed size. **B.** Birth-death process conditioned on fixed size. **C.** Pure-birth process conditioned on fixed time. **D.-F.** Average tMB after treatment from simulations (points and triangles) and predictions (solid lines) using theorem 2 with explicit formulas in Appendix J. **D.** For decreasing population. **E.** For constant population. **F.** For increasing population. Parameters are in table S2.

The total mutational burden after treatment

We compute the expected tMB after homogeneous treatment response using theorem 2. We assume initial size N'_d , expected SFS $E[S_k^{(init)}] = E[S_k^{(det)}]$ approximated by eqn. (8), and treatment interval $[0, t_f]$. The tMB of mutations emerging during treatment is solved exactly. For decreasing or increasing population (i.e. $b_2 \neq d_2$), it is

$$E[B^{(new)}] = 2mN_f \frac{b_2}{d_2} \ln \left(\frac{b_2}{b_2 - d_2} - \frac{1}{(1 - \alpha(t_f))} \frac{N'_d}{N_f} \frac{d_2}{b_2 - d_2} \right) \quad (18)$$

that simplifies to eqn. (17) for $N'_d = 1$. For the constant population ($b_2 = d_2$), it is

$$E[B^{(new)}] = 2mN_f \ln(1 + b_2 t_f). \quad (19)$$

The expression for mutations preexisting to treatment is more involved due to summation over the initial SFS. However, we find a simple approximation for $d_1 = 0$, which we heuristically extend for general

$d_1 > 0$ that reads

$$E[B^{(pre)}] \approx E[B^{(det)}] \frac{1 - \alpha(t_f)}{\alpha(t_f)} \log \left(\frac{1}{1 - \alpha(t_f)} \right). \quad (20)$$

The separation between $b_2 = d_2$ and $b_2 \neq d_2$ is made through $\alpha(t_f)$ defined in eqn. (1). The approximation performs very well in all three cases with small deviations for increasing d_1 , in which case the approximation yields an useful upper bound (fig. S3).

In fig. 3D-F, we show the alignment of the exact theoretical predictions with stochastic simulations. It is worth noting that most mutations contributing to tMB are at low frequencies. After exponential growth, it is estimated that 50% of unique mutations occur in only 1 cell [16]. After treatment, in particular if the population remains approximately constant, this asymmetry is less extreme [9]. Still, the large majority of mutations remain at low frequencies.

Relationship to genetic resistance against therapy

The tMB is linked to the risk of resistance. An increasing number of mutations comes with an increased chance of genetic resistance mechanisms, which is confirmed in clinical data for various cancer types and treatments [4, 11]. Mathematically, the tMB is connected to the probability of having a resistant subpopulation through a probability p_r that a specific mutation leads to resistance. Given the tMB is $B(t)$ at time t the probability that none of those has led to resistance is $\text{Pr}(\text{sensitive}) = (1 - p_r)^{B(t)}$. Although this picture is oversimplified, it makes clear that reducing the tMB generally decreases the risk of resistance. Notable exceptions are immunotherapies, which rely on the accumulation of neoantigens to be effective. In this case, an increased tMB is a double-edged sword: Higher tMB comes with an increased treatment efficiency due to a higher load of neoantigens necessary for the immune system to detect the cancer cells as invaders but also increases the chance of resistance mechanisms such as immune escape [23, 24, 25].

In fig. 3D-F, we show that mutations arising during treatment contribute significantly and often dominate the tMB shortly after treatment initiation. Especially in the constant population scenario, over time, preexisting mutations are almost fully depleted and most mutations will have arisen during treatment. This is an important observation, especially for the planning of possible adaptive clinical trials that are already ongoing or in preparation [26, 27]. A constant population during treatment approximates adaptive therapy cycles in the limit of small cycle sizes. Our calculations suggest that drugs suppressing the accumulation of new mutations - if feasible - would significantly reduce the risk of novel acquired genetic resistance in such patients, while at the same time weeding out preexisting genetic diversity.

There also has been a long debate if treatment resistance is mostly caused by selection on preexisting genetic diversity or resistance is acquired during periods of treatment [28]. These questions are not only related to cancer therapies but similarly can be asked for antibiotic or antiviral therapies [29, 30]. As we show in fig. 3D-F, there is not a single correct answer, but the dominant mechanism does both depend on the timing and duration of therapy and the population dynamics of the treated population. In case of a shrinking population, diversity is for the most part dominated by preexisting mutations. In contrast, in constant or growing populations, genetic diversity is very fast dominated by newly acquired mutations. These observations should be taken into consideration in future planning of multi-drug combination or adaptive cancer therapies and possible underlying mechanisms of failed clinical trials.

The single cell mutational burden distribution

The single cell mutational burden (scMB) distribution has started recently to attract attention. Whole genome single-cell mutational burden information across different normal human tissues and ages has become available recently [1] and theoretical models of single cell mutational burdens during homeostasis have shown to complement information encoded in the SFS [9]. We extend on these considerations and discuss the scMB distribution and its properties under scenarios of cancer growth and during treatment focusing on homogeneous populations.

Construction of the single cell mutational burden distribution

The scMB distribution M_j is linked to the phylogenetic tree through the divisional distributions, which counts the number of cells D_l with l divisions. In terms of the tree topology, l is the leaf depth that is the distance, i.e. number of nodes, from the leaf to the root (fig. 1A). The distribution $\{D_l\}_{l=1}^{\infty}$ partitions the cell population into classes of cells according the number of divisions. Picking a cell with label m that has undergone l divisions, labelling its divisions by $n = 1, 2, \dots, l$ and noting that on each division the cell accumulated $\mu_{m,n}$ mutations, we can write the scMB distribution in term of the divisional distribution. Namely, we have

$$M_j = \sum_{l=1}^{\infty} \sum_{m=1}^{D_l} 1_{\{\sum_{n=1}^l \mu_{m,n}=j\}}, \quad (21)$$

where $1_{\{x=y\}}$ is an indicator function that is 1 if $x = y$ and 0 otherwise. The following theorem characterizes the expected scMB distribution.

Theorem 3. *Given population size N , the expected scMB distribution is expressed in terms of the expected divisional distribution $E[D_l]$ by*

$$E[M_j] = \sum_{l=1}^{\infty} E[D_l] \frac{(lm)^j e^{-lm}}{j!}. \quad (22)$$

Starting with N_0 cells that have undergone zero divisions, the mean-field solution for the expected divisional distribution for a population of size N at time t is

$$E[D_l] \stackrel{mf}{=} N \frac{(2bt)^l}{l!} e^{-2bt}, \quad (23)$$

such that the probability mass $P_l = \frac{E[D_l]}{N}$ follows a Poisson distribution with mean $2bt$.

Proof. Appendix K. □

For exponentially growing populations, eqn. (22) was also recently reported in ref. [31]. Eqn. (23) was derived in ref. [32] and [5]. Furthermore, it has been shown that the divisional distribution is normal distributed as time goes to infinity [33, 34, 35] and some exact expressions with higher complexity were obtained [36, 37]. Since the mean-field solution shows good agreement with simulations up to a small correction for small population size, we proceed with the much simpler mean-field solution.

The average number of divisions for a single cell underlies an inspection bias [36]. Consider a population that has lived for time t with birth rate b . Picking a leaf of the phylogenetic tree at random and tracing it back to the root, one may naively estimate that the leaf has undergone bt divisions on average. Yet,

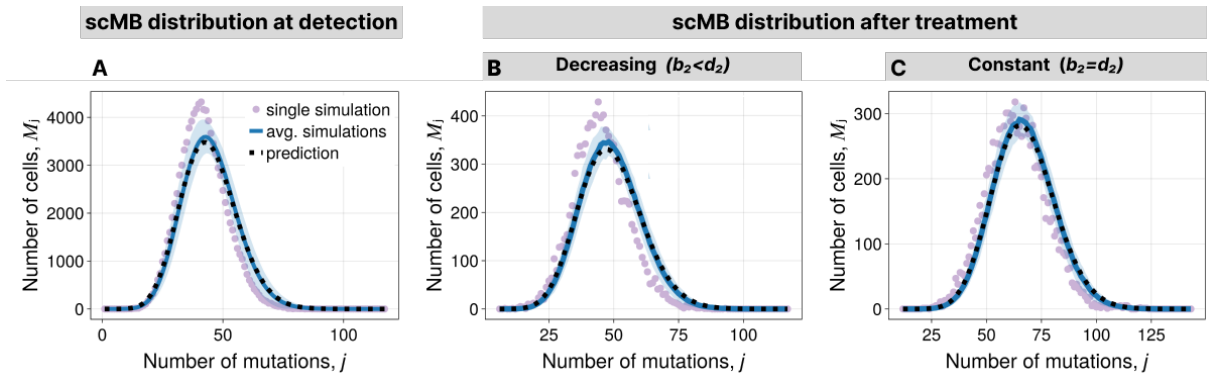


Figure 4: **A.** scMB distribution at detection for the pure-birth process. Average over many simulations with the 1-sigma interval, single realization, and the prediction from theorem 3 with divisional distribution as in eqn. (24). **B.-C.** scMB distribution after treatment. Average over many simulations with the 1-sigma interval, single simulation, and the prediction from theorem 3 with divisional distribution as in eqn. (25). **B.** For decreasing population. **C.** For constant population. Parameters are in table S3.

eqn. (23) predicts an average division number of $2bt$. To understand the increased number of divisions, note that some cells proliferate more than other by chance. Cells with more divisions have a larger number of offspring and contribute more to the average number of divisions in the observed population.

The single cell mutational burden distribution at detection

Following theorem 3, we first obtain the divisional distribution from eqn. (23) and then the scMB distribution from eqn. (22) for a cell populations starting with 1 cell growing to N_d cells. Comparison between the analytic prediction and simulations revealed a small constant offset of the mean. To resolve this discrepancy, we looked at the average number of divisions per cell defined by $\bar{l} = \frac{1}{N_d} \sum D_l$. Rigor treatment of the pure-birth process shows that the expected average number of divisions is given by $E[\bar{l}^{(exact)}] = 2(H_{N_d} - 1)$. From this it follows that the mean-field solution obtained through eqn. (23) differs from the exact solution by a constant 2 when conditioned on fixed size or $\ln(4)$ when conditioned on fixed time (Appendix L). Focusing on fixed-size observations, we integrate the correction constant into the divisional distribution setting

$$P_l \sim \text{Poiss}(2bt_d - 2), \quad (24)$$

where t_d is defined as the mean time to reach size N_d for the first time (see Appendix B). The adjusted scMB distribution shows excellent agreement with simulations of the pure-birth process (fig. 4A), and good but not perfect agreement when deaths are included (fig. S5). The scMB distribution of single realisation has the same shape as the expected distribution, although it comes with a small random shift of the mean. These shifts are caused by stochastic effects at small population sizes leading to mutations shared by a large fraction of the population.

The single cell mutational burden distribution after treatment

For homogeneous treatment response, the analytic prediction for decreasing, constant and increasing populations under treatment is the same. As long as the final population size is large, i.e. $N_f \gg 1$, small

population effects can be neglected such that the average number of divisions obtained during treatment time t_f is $E[\bar{l}] = 2bt_f$. Those divisions are again Poisson distributed and together with divisions before treatment, we have

$$P_l \sim \text{Pois}((2bt_d - 2) + 2bt_f), \quad (25)$$

where $t_f \approx \frac{\ln(N_f/N_d)}{b_2 - d_2}$. We put $E[D_l] = N_f P_l$ into eqn. (22) to obtain the scMB distribution showing good alignment with simulations (fig. 4B-C, fig. S6).

The average scMB increases linear in time (fig S4), in concordance with experimental observations of healthy human tissues [2]. Whereas this is also true for some limiting cases for the tMB normalized over the population size, the tMB shows in general more complex non-linear behaviour (fig. 3D-F). Similarly to the observable SFS, the scMB distribution does not change its shape after homogeneous treatment response yielding a null model to test homogeneity. In contrast to the SFS, the scMB distribution differs from the null hypothesis only for heterogeneity in birth rates, i.e. resistant cells have different birth rates than sensitive cells. Furthermore, the scMB distributions allows decoupling mutation rate and birth rate by studying its variance [9].

Genetic intra-tumour heterogeneity upon emerging resistance

So far, we discussed properties of gITH in cell populations with homogeneous treatment response. For certain treatment regimes such as chemo or radio therapy in early stage disease, this seems a reasonable null scenario. However, in other cases, emergence of resistant subpopulations is a major problem. Genetic resistance is well documented in many targeted cancer therapies, where specific point mutations induce resistance [38, 39]. Resistance mutations give rise to resistant clones leading to a cell population, in which some cells are less or not affected by treatment (fig. 1C). Building up on the results for homogeneous populations, we first study the size distribution for resistant clones, followed by investigations of the SFS and scMB distribution.

The resistant clone size distribution

We distinguish between mutations without effect accumulated at rate m and mutations that cause resistance accumulated at rate ν per cell and division. We denote the number of resistant clones of size κ , by Θ_κ and call $\{\Theta_\kappa\}_{\kappa=1}^N$ the clone size distribution (CSD). The CSD and the SFS have a conceptual difference: the CSD is not nested, i.e. clones are not overlapping, whereas the SFS allows nested lineages. However, since ν is very small, resistance mutations have negligible effect on the growth dynamics of sensitive cells from which the resistant clones arise. As a consequence, we can approximate the CSD by adapting theorem 1 (Appendix M).

At detection, the expected CSD is

$$E[\Theta_\kappa^{(det)}] \approx \frac{1}{1 - p(1 \rightarrow 0, t_d)} \int_0^{t_d} 2\nu e^{(b_{s,1} - d_{s,1})t} p(1 \rightarrow \kappa, t_d - t) dt, \quad (26)$$

where the term for conditioning on survival $\frac{1}{1 - p(1 \rightarrow 0, t_d)}$ is appropriately parameterized with growth parameters for sensitive cells $b_{s,1}$, $d_{s,1}$ while the term $p(1 \rightarrow \kappa, t_d - t)$ must be parameterized with growth parameters for resistant cells $b_{r,1}$, $d_{r,1}$. Assuming that the growth parameters of sensitive and resistant cells are the same in the absence of treatment, the CSD takes the same form as the SFS in homogeneous

populations in eqn. (8). Focusing on the limit $N_d \rightarrow \infty$ and writing $b_1 := b_{s,1} = b_{r,1}$ and $d_1 := d_{s,1} = d_{r,1}$, we have

$$E[\Theta_\kappa^{(det)}] \approx 2\nu N_d \sum_{l=0}^{\infty} \frac{\left(\frac{d_1}{b_1}\right)^l}{(\kappa+l)(\kappa+l+1)} \quad (27)$$

This is not surprising, given above assumptions, resistance inferring mutations are neutral before treatment.

Next, we consider the CSD after treatment time t_f . As initial condition, we assume fixed size N'_d , and initial CSD $E[\Theta_\kappa^{(init)}] = E[\Theta_\kappa^{(det)}]$. We adapt theorem 1 and obtain the expected CSD of clones that emerged before and after detection giving us

$$E[\Theta_\kappa^{(new)}] \approx \int_0^{t_f} 2\nu b_{s,2} N'_d e^{(b_{s,2}-d_{s,2})t'} p(1 \rightarrow \kappa, t_f - t') dt' \quad (28)$$

and

$$E[\Theta_\kappa^{(pre)}] \approx \sum_{\kappa'=1}^{N'_d} E[\Theta_{\kappa'}^{(det)}] p(\kappa' \rightarrow \kappa, t_f). \quad (29)$$

In eqn. (28), the relevant mutation generation is caused by sensitive cells ($2\nu b_{s,2} N'_d e^{(b_{s,2}-d_{s,2})t'}$). In both eqn.(28) and (29), the growth of clone sizes $p(a \rightarrow n, t)$ is parameterized for resistance cells $b_{r,2}$ and $d_{r,2}$.

Subdominance of the largest resistant clone

In an exponentially growing population, the first mutant clone has a large advantage as all later arriving clones need to catch up to become dominant. We thus asked whether the first resistant clone always dominates all other later arising resistant clones. Here, we show that this is not the case. We focus on detection since a dominating resistant clone at detection also dominates after treatment.

First, we find that the first arriving clone is not always the largest. We derived the probability densities for the arrival time T_1 of the first clone and the second clone T_2 given that $T_1 = t_1$ and survival of drift (Appendix N). The long-term growth of the clones can be written as $V_1 e^{(b_1-d_1)(t_d-T_1)}$ and $V_2 e^{(b_1-d_1)(t_d-T_2)}$ where V_1, V_2 are exponentially distributed with mean $\frac{b_1-d_1}{b_1}$ (Theorem 1 in ref. [15]). The difference in their size at time t_d is then

$$V_1 e^{(b_1-d_1)(t-T_1)} - V_2 e^{(b_1-d_1)(t-T_2)}, \quad (30)$$

which we sample and illustrate in fig. 5A. One can see that a considerable fraction is in the negative range (illustrated in red). This means that the second clone can outgrow the first clone purely by drift. These effects would become more pronounced for possible fitness distributions of resistant clones.

Next, we considered the largest clone at time t_d instead of the first arriving clone. Interpreting the CSD as a probability distribution for clone sizes, we can use methods from order statistics [40] and find an approximate probability density for the size of the largest resistant clone $\tilde{\kappa}$ that reads

$$f(\tilde{\kappa}) = \frac{A}{\tilde{\kappa}^2} \left(1 - \frac{1}{\tilde{\kappa}}\right)^{A-1} \quad \text{with } A = \frac{b_1}{b_1 - d_1} 2\nu N_d \quad (31)$$

and is valid for $\nu N_d \gg 1$. The derivation is presented in Appendix N and we validated our approximation with simulations (fig. 5B).

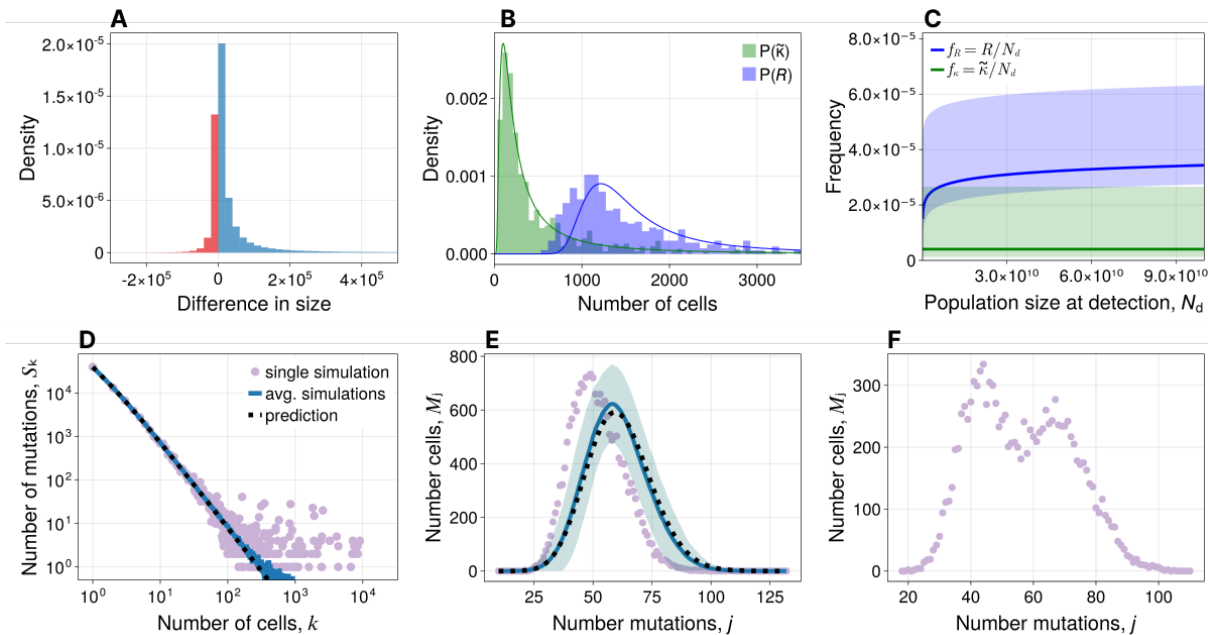


Figure 5: **A.** Difference in size between the first and second resistant clone at detection given by eqn. (30). Negative differences are shown in red, and positive differences in blue. **B.** Size distributions for the largest resistant clone and the resistant subpopulation at the time of detection. Simulations as histograms, and predictions as solid lines given by eqn. (31) and (32). **C.** Distribution of frequencies for the largest resistant clone and the resistant subpopulation at the time of detection. The median frequency is shown as a solid line with shadow ranging from the 10% to the 90% quantiles. **D.** SFS after treatment upon emergence of resistance from simulations and prediction of the compound tail in eqn. (34). Parameters are set such that all sensitive cells are extinct. **E.** scMB distribution upon emergence of resistance in which birth rates remain unaffected. For the prediction, we used treatment time t_f measured from one simulation. **F.** scMB distribution for a single population upon emergence of resistance with change of birth rates. Parameters are in table S4.

We compare the size of the largest resistant clone to the total number of resistant cells. Given small mutation rate ν but large population sizes N_d and $\nu N_d \gg 1$, the probability of having R resistant cells at the time of detection is described by a Landau distribution [41, 42]. We have

$$P(R) = \frac{b_1 - d_1}{b_1} \frac{1}{2\nu N_d} P_{\text{Landau}} \left(\frac{b_1 - d_1}{b_1} \frac{R}{2\nu N_d} - \ln(2\nu N_d) \right) \quad (32)$$

that also shows good agreement with simulations but with a small offset (fig. 5B).

Both distributions have a fat tail that decreases with exponent -2 . The mean and the median of the largest clones $\tilde{\kappa}$ are always below the mean and median of R . Interestingly, the median fraction of all resistant cells is weakly increasing with detection size N_d and is predicted to be around 3.0×10^{-5} for biologically feasible parameter values. In contrast, the median size of the largest resistant clone is approximately constant for different detection sizes and is around 4.0×10^{-6} (fig. 5C). This is reasonable as newly resistant cells continue to emerge from sensitive cells during growth and thus increase the fraction of resistant cells as the tumour grows. We conclude that the largest clone may make up a significant fraction but probably not a dominating fraction of the resistant subpopulation.

The site frequency spectrum

We assume resistant cells to be neutral in the absence of treatment such that the expected SFS at detection is unchanged given by eqn. (8). In the following, we focus on the SFS after treatment. We separate between neutral mutations that emerged in sensitive cells S_k^S and mutations that emerged in resistant cells S_k^R . We label K resistant clones with $i = 1, 2, \dots, K$ such that

$$S_k = S_k^S + S_k^R \quad \text{with} \quad S_k^R = \sum_i^K S_k^{R,i}. \quad (33)$$

By definition, S_k^S combines mutations in sensitive cells and mutations in resistant cells acquired in sensitive ancestors.

If resistance mutations lead to full resistance, i.e. $b_r := b_{r,1} = b_{r,2}$ and $d_r := d_{r,1} = d_{r,2}$, then the compound SFS of all resistant clones takes again the known form

$$E[S_k^R] = \sum_{\kappa=1}^{N_f} E[\Theta_\kappa] E[S_k(\kappa)] \approx 2mN_r \sum_{l=0}^{\infty} \frac{\left(\frac{d_r}{b_r}\right)^l}{(k+l)(k+l+1)}, \quad (34)$$

where N_r is the number of resistant cells, $E[S_k(\kappa)]$ is the SFS of an exponentially growing cell population at size κ . The calculations are presented in Appendix O and alignment with simulations is shown in fig. 5D.

To make progress on S_k^S , we assume sufficiently long treatment times such that all sensitive cells go extinct. Then, S_k^S consists of mutations in resistant cells that are clonal in at least one resistant clone. If we denote $\vec{\kappa} = (\kappa_1, \kappa_2, \dots, \kappa_K)$ as a vector of clone sizes κ_i with indices $i \in X$, we have

$$S_k^{S,R} = \sum_{Y \in 2^X} C_Y \delta \left(\sum_{j \in Y} \kappa_j = k \right). \quad (35)$$

Here, the first sum goes over all possible combinations Y in the set of indices X that is the power set 2^X , and C_Y is the number of mutations shared between the chosen clones. Those mutations will arise as peaks in the SFS (fig. 5D and S7). The heights and the locations of the peaks carry information on the timing and selection of resistant clones. However, a more in-depth understanding of the peaks requires analysing the relatedness between clones and goes beyond the scope of this study.

The single cell mutational burden

From our previous investigations, we know that the scMB is not affected by the death rate but only the birth rate. Thus, if the birth rate is unaffected in the selection process then the scMB distribution remains the shape described in theorem 3, which we confirmed with simulations (fig. 5E). However, if the birth rate of the resistant cell lines differs from the sensitive cells then the scMB distribution will differ too (fig. 5F).

We write the scMB distribution of the resistant population as $M_j^R = \sum_{i=1}^K M_j^{R,i}$, where $M_j^{R,i}$ is the scMB distribution of an individual clone with label i . Together with the sensitive cells, the total expected scMB

distribution is

$$\begin{aligned} E[M_j] &= E[M_j^S] + E[M_j^R] \\ &= E[M_j^S] + \sum_{\kappa=1}^{N_f} E[\Theta_\kappa] E[M_j(\kappa)]. \end{aligned} \tag{36}$$

The total scMB distribution consists of overlapping scMB distribution of sensitive cells and resistant clones. If the birth rate changes in resistant cells during treatment but is the same among all resistant cells, the majority of resistant cells consist of clones that arose before treatment with the same expected scMB leading to a combined single peak of resistant cells next to the peak of sensitive cells (fig. 5F). If the birth rate of resistant clones changes already before treatment, the clones will have different expected scMB leading to many overlying distributions.

Discussion

We used stochastic processes to investigate expected patterns of genetic heterogeneity in cancer before and after treatment. We discussed three summary statistics of gITH (SFS, tMB, scMB) for different treatment scenarios. For homogeneous treatment response, we found that the SFS keeps its k^{-2} power law tail regardless whether the cancer is shrinking, approximately constant or growing, and the scMB distribution is described by two entangled Poisson processes. Those predictions build the basis for possible homogeneity tests, and can provide insights into the mechanisms underlying resistance evolution. Similar usage of a null hypothesis of the SFS has led to biological insights into selection processes of healthy tissues [43] and cancer before treatment [18]. We further characterised the non-linear behaviour of the tMB. Notably, the tMB relates to the probability of having resistant cells and the scMB may be interpreted as scMB of antigens important for the response to immunotherapies.

Next, we investigated the effect of treatment resistance on gITH assuming a simple model of genetic resistance reasonable for treatment of late-stage cancers with targeted therapies that often show resistance due to specific point mutations [38, 39]. We found that the resistant subpopulation will consist of many resistant clones that arose from sensitive cells. The largest resistant clone likely makes up a considerable but not dominating fraction of the total resistant subpopulation. We showed that resistant clones leave traces in the SFS, and change the scMB distribution only if birth rates are affected.

For generality and tractability, here we used a minimal set of assumptions when defining our theoretical framework. Data generation often comes with additional complexities that must be incorporated before our results can be readily compared to the data. First and most important, sequencing comes with sampling effects as well as sequencing noise. However, these can be taken into account in both stochastic simulations and data analysis [9, 5]. Furthermore, there are possible biological complexities including (i) the impact of spatial structure [44] (ii) complex selection processes, e.g. varying levels of treatment response among resistant clones [45] (iii) the impact of the tumour-microenvironment and other non-genetic determinants [46]. These effects may contribute to evolutionary paths of individual tumours. As such, our results provide testable null hypotheses of which additional complexities can be tested. We hope that our results will set the stage for further theoretical and data-driven investigations.

Methods

Computer simulations were implemented in the Julia programming language and are based on the Gillespie algorithm [47, 48]. We adapted the simulation framework developed for ref. [49]. The cell population is saved in a tree structure. The root of the tree represents the ancestor cell and the leaves represent living cells. In case of a death event the chosen leaf is removed from the tree. In case of a birth event two new leaves are created and each carries the number of mutations drawn from a Poisson distribution. Then, each node that is not a leaf represents an ancestor of a currently living cell.

To obtain the SFS, we iterate over all nodes and count the number of mutations unique to this node and its ancestors. If the node with label i has μ_i mutations and leads to k leaves, we increase S_k by μ_i . After iterating over all nodes, we have the SFS.

For the scMB distribution, we iterate over all leaves. On top of the mutation of the leaf itself, we add its parent's mutations, then its parent's parent's mutations and go on until we reach the ancestor cell. The final sum is the scMB of this leaf. After iteration over all leaves, we have the scMB distribution.

Numerical solutions of integrals were computed using the HCubature module in the Julia programming language [50] that is based on an adaptive algorithm [51]. Integrals were computed up to a relative tolerance of 10^{-3} .

Mathematical analysis was supported using *Wolfram Mathematica* [52].

We balance efficiency and accuracy for the computation of the probability mass function of the birth-death process $p(a \rightarrow n, t)$. We always use exact solutions for the pure-birth process, pure-death process or when the population start from a single cell [53]. Otherwise, we use exact solutions written as computationally stable sum [12] for $\min(a, n) \geq 50$ and use the (unnormalized) saddlepoint-approximation method solution otherwise [13]. Details of the approximations are described in Appendix A.

Random samples for random variables T_1 and T_2 were generated using inverse transformation sampling [54]. Therefore, a random number u is drawn from a uniform distribution in $(0, 1)$ and is then transformed into a random number x with cumulative density function $F_X(x)$ by plugging the uniform random number into the inverse distribution function $x = F_X^{-1}(u)$.

We fit the exponent γ of the power-law $S_k \sim k^\gamma$ to analyse the SFS obtained from simulated data. Therefore, we use a maximum likelihood estimator for discrete power law distributions from ref. [55]. Adapted for our purposes, we have

$$\hat{\gamma} = 1 + B_{k_{\min}} \left(\sum_{k=k_{\min}}^N S_k \ln \frac{k}{k_{\min} - \frac{1}{2}} \right), \quad (37)$$

where k_{\min} defines the smallest site frequency used for the fitting, $B_{k_{\min}} = \sum_{k=k_{\min}}^N S_k$ is the number of mutations that occur in at least k_{\min} mutations and N is the population size.

Acknowledgement

AS was supported by the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie EvoGamesPlus grant agreement No. 955708. BW was supported by a

Barts Charity Lectureship (grant no. MGU045) and a UKRI Future Leaders Fellowship (grant no. MR/V02342X/1). We thank Nathaniel Mon Père and Christo Morison for helpful discussions at different stages of the manuscript.

Code availability

For the simulations, we adapted a package that was written for ref. [49] and is available at <https://github.com/jessierenton/SomaticEvolution.jl>.

The adapted package is available at <https://github.com/alexsteininfo/TreeStatistics.jl> and code to run and analyse the simulations is available at <https://github.com/alexsteininfo/GITH-Treatment-Patterns>.

Appendix

A Probability mass functions of the birth-death process

Given constant rates for birth b and death d , and starting with a cells, what is the probability $p(a \rightarrow n, t)$ that we have n cells after time t ? – A comprehensive answer to this question is given in ref. [53]. Here, we summarize the results with the addition of more recent results that circumvent issues with computational stability and efficiency.

A.1 Exact solutions

For general birth and death rates, $b \geq 0$ and $d \geq 0$, and general initial population size $a \geq 1$, the probability to grow to size n in time t is given by

$$p(a \rightarrow n, t) = \sum_{j=0}^{\min(a,n)} \binom{a}{j} \binom{a+n-j-1}{a-1} \alpha^{a-j} \beta^{n-j} (1-\alpha-\beta)^j \quad \text{for } n \geq 1 \quad (\text{A1})$$

$$p(a \rightarrow 0, t) = \alpha^a.$$

In the super- and subcritical case $b \neq d$, we have

$$\alpha = \frac{de^{rt} - d}{be^{rt} - d} \quad \text{and} \quad \beta = \frac{be^{rt} - b}{be^{rt} - d}, \quad (\text{A2})$$

and in the critical case $b = d$, we have

$$\alpha = \beta = \frac{bt}{1 + bt}. \quad (\text{A3})$$

The mean and variance is given by

$$m(t) = ae^{(b-d)t}$$

$$\sigma^2(t) = a \frac{b+d}{b-d} e^{(b-d)t} (e^{(b-d)t} - 1) \quad (\text{A4})$$

Whereas it is remarkable that an exact solution can be obtained, the solution is computationally expensive and unstable. If the term $(1 - \alpha - \beta)$ is negative and the binomial coefficients are large, the limits of machine precision is exceeded and one yields wrong values for $p(a \rightarrow n, t)$. An alternative form has been derived that sums over positive terms only [12] that reads

$$p(a \rightarrow n, t) = \sum_{j=0}^{\min(a,n)} \binom{a}{j} (1-\alpha)^j \alpha^{a-j} p_{jm}^* \quad (\text{A5})$$

with

$$p_{jm}^* = \binom{n-1}{n-j} (1-\beta)^j \beta^{(n-j)} \quad (\text{A6})$$

for $j \geq 0$ and $p_{00}^* = 1$ and $p_{0m}^* = 0$ for $j = 0$. Whereas the sum in eqn. (A5) is computationally stable, it is still computationally expensive for large values of a or n .

If the initial population size is $a = 1$ and $b \neq d$, the probability to grow to size n after time t simplifies to

$$\begin{aligned} p(1 \rightarrow n, t) &= (1 - \alpha)(1 - \beta)\beta^{n-1} \quad \text{for } n \geq 1 \\ p(1 \rightarrow 0, t) &= \alpha \end{aligned} \tag{A7}$$

with α and β given by eqn. (A2). The distribution is recognized as generalized geometric distribution.

For the critical case ($b = d$) with initial population size $a = 1$, the probability mass becomes

$$\begin{aligned} p(1 \rightarrow n, t) &= \frac{(bt)^{n-1}}{(1 + bt)^{n+1}} \quad \text{for } n \geq 1 \\ p(1 \rightarrow 0, t) &= \frac{bt}{1 + bt}. \end{aligned} \tag{A8}$$

For the pure-birth process ($d = 0$) and general initial population size $a \geq 1$, we have

$$p(a \rightarrow n, t) = \binom{n-1}{a-1} e^{-abt} (1 - e^{-bt})^{n-a} \tag{A9}$$

which is a negative binomial distribution, i.e. the probability of having n trials given a successes with success probability $p = e^{-bt}$.

For the pure death process $b = 0$, we have

$$p(a \rightarrow n, t) = \binom{a}{n} e^{-ndt} (1 - e^{-dt})^{a-n}, \tag{A10}$$

which is a binomial distribution, i.e. the probability of having n successes in a Bernoulli trials with success probability $p = e^{-dt}$.

A.2 Saddlepoint approximation method

The saddlepoint approximation was introduced to approximate the probability density function or probability mass function given that we know the moment generating function [56] and has found many applications [57].

Consider a discrete random variable X that can take values of integers, $k = 1, 2, 3, \dots$. Let us denote the moment generating function by $M(\theta)$, and the cumulant generating function $K(\theta) = \log(M(\theta))$. Then the saddlepoint mass function is computed by

$$\hat{p}(k) = \frac{1}{\sqrt{2\pi K''(\hat{\theta})}} e^{K(\hat{\theta}) - \hat{\theta}k} \tag{A11}$$

where $\hat{\theta}$ is the solution to $K'(\hat{\theta}) = k$. Noteworthy, by design the saddlepoint approximation is not normalized. Once normalized, one talks about the normalized saddlepoint mass function.

Davison et al. applied the saddlepoint approximation on the known moment generating function given

by Bailey [13]. For the super- and subcritical case ($b \neq d$) the saddlepoint mass function is given by

$$\hat{p}(a \rightarrow n, t) = \frac{1}{\sqrt{2\pi a}} \frac{1}{\tilde{s}^k} \left(\frac{d - b\tilde{s} + d(\tilde{s} - 1)e^{(b-d)t}}{d - b\tilde{s} + b(\tilde{s} - 1)e^{(b-d)t}} \right) \left(\frac{(e^{(b-d)t} - 1)e^{(b-d)t}\tilde{s}(b-d)^2(-b^2\tilde{s}^2 + be^{(b-d)t}d(\tilde{s}^2 - 1) + d^2)}{(b(e^{(b-d)t}(\tilde{s} - 1) - \tilde{s}) + d)^2 (b\tilde{s} + d(-e^{(b-d)t}\tilde{s} + e^{(b-d)t} - 1))^2} \right)^{-1/2} \quad (\text{A12})$$

where

$$\tilde{s} = \frac{1}{2A}(-B + \sqrt{B^2 - 4AC}) \quad \text{with} \quad \begin{cases} A = b(e^{(b-d)t} - 1)(b - de^{(b-d)t}) \\ B = 2bd(1 + e^{2(b-d)t} - e^{(b-d)t} - \frac{a}{n}e^{(b-d)t}) + e^{(b-d)t}(b^2 + d^2)(\frac{a}{n} - 1) \\ C = d(e^{(b-d)t} - 1)(d - be^{(b-d)t}) \end{cases} \quad (\text{A13})$$

For the critical case ($b = d$), it is

$$\hat{p}(a \rightarrow n, t) = \frac{1}{\sqrt{2\pi a}} \frac{1}{\tilde{s}^k} \left(\frac{bt(1 - \tilde{s}) + \tilde{s}}{1 - bt(\tilde{s} - 1)} \right)^a \times \left(\frac{bt\tilde{s}(-bt\tilde{s}^2 + bt + \tilde{s}^2 + 1)}{(bt(\tilde{s} - 1) - 1)^2(-bt\tilde{s} + bt + \tilde{s})^2} \right)^{-1/2} \quad (\text{A14})$$

where

$$\tilde{s} = \frac{1}{2A}(-B + \sqrt{B^2 - 4AC}) \quad \text{with} \quad \begin{cases} A = bt - (bt)^2 \\ B = 2(bt)^2 + \frac{a}{n} - 1 \\ C = -bt - (bt)^2 \end{cases} \quad (\text{A15})$$

Davison et al. have validated the approximation over a range of parameters. They have found that the approximation yields good results for sufficiently large n and a . Furthermore, their analysis shows that normalization of the saddlepoint mass function does not lead to significant improvement of the approximation such that we continue working with the unnormalized probabilities.

A.3 Heuristic approximation using over-dispersed Poisson distributions

We start by observing that for the pure-birth process, the probability mass follows a binomial distribution. In the pure-death process, it follows a negative binomial distribution which is commonly used to describe overdispersed count data. Furthermore, plotting eqn. (A5) for general b and d and sufficiently large n shows the typical bell shape similar to a Poisson distribution but with increased variance. Those observation motivated us to approximate the probability mass with an overdispersed Poisson distribution that we parameterise with the expressions for the mean and variance.

Consul Poisson distribution: We chose the Consul Poisson distribution [58] as one of the simplest formulas of an overdispersed Poisson distribution that reads

$$P(k) = \frac{\mu e^{-k\lambda - \mu} (k\lambda + \mu)^{k-1}}{k!} \quad (\text{A16})$$

and has mean $\frac{\mu}{1-\lambda}$ and variance $\frac{\mu}{(1-\lambda)^3}$. In the case $\lambda = 0$, we have the standard Poisson distribution.

We parameterize the Poisson distribution with the mean and the variance of the super- or subcritical birth-death process given in eqn. (A4) that leads to

$$\begin{aligned}\frac{\mu}{1-\lambda} &= m(t) = ae^{rt} \\ \frac{\mu}{(1-\lambda)^3} &= \sigma^2(t) = a \frac{b+d}{b-d} e^{(b-d)t} (e^{(b-d)t} - 1)\end{aligned}\tag{A17}$$

Solving this equation for μ and λ has two solutions. We chose the solution leading to positive parameters μ and λ in case of overdispersion (i.e. $\sigma^2(t) > m(t)$) that is

$$\begin{aligned}\mu &= \frac{m^{3/2}(t)}{\sigma(t)} = \frac{(ae^{(b-d)t})^{3/2}}{\sqrt{\frac{a(b+d)e^{(b-d)t}(e^{(b-d)t}-1)}{b-d}}} \\ \lambda &= \frac{\sigma(t) - m^{1/2}(t)}{\sigma(t)} = \frac{\sqrt{\frac{a(b+d)e^{(b-d)t}(e^{(b-d)t}-1)}{b-d}} - \sqrt{ae^{(b-d)t}}}{\sqrt{\frac{a(b+d)e^{(b-d)t}(e^{(b-d)t}-1)}{b-d}}}\end{aligned}\tag{A18}$$

Negative binomial distribution: The negative binomial distribution is a standard choice to describe an overdispersed Poisson distribution. It reads

$$P(k) = \binom{k+r-1}{k} (1-p)^k p^r,\tag{A19}$$

has mean $\frac{r(1-p)}{p}$ and variance $\frac{r(1-p)}{p^2}$. It converges towards a Poisson distribution in the limit $p \rightarrow 1$. The parameters p and r can be parameterized and expressed in terms of the mean and variance,

$$\begin{aligned}p &= \frac{m(t)}{\sigma^2(t)} \\ r &= \frac{m^2(t)}{\sigma^2(t) - m(t)},\end{aligned}\tag{A20}$$

where the mean $m(t)$ and variance $\sigma^2(t)$ are again taken from eqn. (A4).

B Probability for the time to reach size N

The probability for the time to reach size N starting from one cell ($a = 1$) was derived in ref. [15]. The probability density is approximately

$$\text{approximately } f_{T_N}(t) = e^{-\frac{b-d}{b}Ne^{-(b-d)t}} \frac{(b-d)^2 N}{b} e^{-(b-d)t}.\tag{B1}$$

This is a Gumbel distribution, which is commonly written in terms of a location parameter μ and shape parameter β . We can write the density as

$$f_{T_N}(t) = \frac{1}{\beta} e^{-(z+e^{-z})} \quad \text{with} \quad z = \frac{x-\mu}{\beta},\tag{B2}$$

such that $\beta = \frac{1}{b-d}$ and $\mu = \frac{1}{b-d} \ln\left(\frac{b-d}{b}N\right)$.

We can then make use of the known properties of the Gumbel distribution. In particular, the expectation is

$$E[T_N] = \mu + \beta\gamma = \frac{1}{b-d} \left(\ln\left(\frac{b-d}{b}N\right) + \gamma \right) \quad (\text{B3})$$

where $\gamma \approx 0.577$ is the Euler constant. Further, the moment-generating function defined by

$$\phi(u) = E[e^{uT_N}] = \int_{-\infty}^{\infty} e^{ut} f_{T_N}(t) dt \quad (\text{B4})$$

takes the form

$$\phi(u) = \Gamma(1 - \beta u) e^{\mu u} = \Gamma\left(1 - \frac{u}{b-d}\right) \left(\frac{b-d}{b}N\right)^{\frac{u}{b-d}}, \quad (\text{B5})$$

where $\Gamma(z) = \int_0^{\infty} t^{z-1} e^{-t} dt$ is the gamma function.

The Gumbel distribution is defined over $(-\infty, \infty)$ whereas we are only interested in positive times $T_N > 0$. However, in applications the integral over the density for negative t is typically negligible and we proceed with approximating $\int_{-\infty}^0 f_{T_N}(t) dt = 0$.

C Switching between fixed-time and fixed-size conditioning

The random process is stopped at fixed time $T = t$ or fixed size $N = n$ and an observable is measured. We know that the probability to grow to size n in time t (Appendix A) that we write as $p(n|t)$ to be explicit on the conditioning. Similarly, starting from a single cell, we know the probability for the time t to grow to size n for the first time (Appendix B) that we denote by $f_T(t|N = n)$. We investigate the expectation of an observable O .

Given we know the expected observation for fixed time $T = t$, we can compute the expected observation for fixed size $N = n$ by

$$\begin{aligned} E[O|N = n] &= \int_0^{\infty} E[E[O|N = n]|T = t] f_T(t|N = n) dt \\ &= \int_0^{\infty} E[O|T = t] f_T(t|N = n) dt. \end{aligned} \quad (\text{C1})$$

In the first equality, we decompose the expectation of $E[O|N = n]$ into expectations for all possible times t . In the second equality, we use the law of total expectation which states $E[E[O|N = n]|T = t] = E[O|T = t]$.

Similarly, if we know the expected observation for fixed size $N = n$, we can compute the expected observation for fixed time $T = t$ by

$$E[O|T = t] = \sum_{n=1}^{\infty} E[O|N = n] p(n|t), \quad (\text{C2})$$

where we condition on survival and thus chose $n = 1$ rather than $n = 0$.

D Proof of theorem 1

Part 1: We prove eqn. (6) for general N_0 following Appendix B of ref. [16] who proved the result for $N_0 = 1$. Restricted (but not conditioned) on survival of the entire population $\{Z_0(t) > 0\}$, the expected number of mutations that emerged in the time interval $[t', t' + dt']$ and grow to size k in the remaining time $t - t'$ was shown to be

$$E \left[\left(S_k^{(new)}(t' + dt') - S_k^{(new)}(t') \right) 1_{\{Z_0(t) > 0\}} \right] = bE[\Omega]E[Z_0(t')]p(1 \rightarrow k, t - t') dt'. \quad (D1)$$

Here, Ω is the number of mutations generated in a division. Given our *Theoretical framework*, we have $E[\Omega] = 2m$. The expected population size in the birth-death process is given by $E[Z_0(t')] = N(t') = N_0 e^{(b-d)t'}$. Conditioning the total population to survive time t leads to

$$E \left[\left(S_k^{(new)}(t' + dt') - S_k^{(new)}(t') \right) | Z_0(t) > 0 \right] = \frac{1}{1 - p(N_0 \rightarrow 0, t)} 2mbN(t')p(1 \rightarrow k, t - t') dt'. \quad (D2)$$

It remains to integrate over the time interval of entire time interval $t' \in [0, t]$ such that

$$E[S_k^{(new)}] = \frac{1}{1 - p(N_0 \rightarrow 0, t)} \int_0^t 2mbN(t')p(1 \rightarrow k, t - t') dt', \quad (D3)$$

which makes the proof complete.

Noteworthy, the expected BWD is given by

$$E[W_k^{(new)}] = E[S_k^{(new)}]/m = \frac{1}{1 - p(N_0 \rightarrow 0, t)} \int_0^t 2bN(t')p(1 \rightarrow k, t - t') dt', \quad (D4)$$

Part 2: Next, we prove eqn. (7). We start by writing the initial BWD as $W_k^{(init)}$, label branches of the phylogenetic tree with $i = 1, 2, \dots, L$ and their branch widths with k_i . During time t , branches grow stochastically and we filter for those that grow to size k by writing

$$W_k^{(pre)} = \sum_{i=1}^L 1_{\{k_i \rightarrow k\}} = \sum_{k'=1}^{N_0} \sum_{j=1}^{W_{k'}^{(init)}} 1_{\{k_j \rightarrow k\}} \quad (D5)$$

where $1_{\{k_i \rightarrow k\}} = 1$ if branch i of width k_i grows to size k and $1_{\{k_i \rightarrow k\}} = 0$ otherwise. In the second equality, we used that the L branches are partitioned into N_0 classes $\{W_{k'}^{(init)}\}_{k'=1}^{N_0}$ according to their initial width k' .

Taking the expectation, we have

$$\begin{aligned} E[W_k^{(pre)}] &= E \left[\sum_{k'=1}^{N_0} \sum_{j=1}^{W_{k'}^{(init)}} 1_{\{k_j \rightarrow k\}} \right] \\ &= \sum_{k'=1}^{N_0} E \left[\sum_{j=1}^{W_{k'}^{(init)}} 1_{\{k_j \rightarrow k\}} \right] \\ &= \sum_{k'=1}^{N_0} E[W_{k'}^{(init)}] p(k' \rightarrow k, t). \end{aligned} \quad (D6)$$

In the second equality, we used linearity of the expectation. In the third equality, we used Wald's identity together with $E[1_{\{k_j \rightarrow k\}}] = p(k' \rightarrow k, t)$ when k_j has width k' . We substitute $E[W_k^{(pre)}] = E[S_k^{(pre)}]/m$ and $E[W_k^{(init)}] = E[S_k^{(init)}]/m$ and condition the entire population on survival such that

$$E[S_k^{(pre)}] = \frac{1}{1 - p(N_0 \rightarrow 0, t)} \sum_{k'=1}^{N_0} E[S_{k'}^{(init)}] p(k' \rightarrow k, t), \quad (\text{D7})$$

which is the desired result.

E Site frequency spectra at detection

E.1 Fixed-time expectation

Starting from a single cell with u_0 mutations, the population grows time t_d to detection. The average growth in the time interval $[0, t_d]$ is given by $N(t') = e^{(b_1 - d_1)t'}$ and the expected population size at detection conditioned on survival is $N_d = \tilde{N}(t_d) = \frac{1}{1 - \alpha(t_d)} e^{(b_1 - d_1)t_d}$.

Newly emerging mutations: Following theorem 1, we have

$$E[S_k^{(new)}] = \frac{1}{1 - \alpha(t_d)} \int_0^{t_d} 2mb_1 e^{(b_1 - d_1)t'} p(1 \rightarrow k, t_d - t') dt'. \quad (\text{E1})$$

In ref. [17] and [16], the integral is rewritten and simplified for the limits $N_d \rightarrow \infty$ and $k \rightarrow \infty$ giving us

$$\begin{aligned} E[S_k^{(new)}] &= 2mN_d \int_0^{1 - 1/N_d} \left(1 - \frac{d_1}{b_1} y\right)^{-1} (1 - y)y^{k-1} dy \\ &\rightarrow 2mN_d \sum_{l=0}^{\infty} \frac{\left(\frac{d_1}{b_1}\right)^l}{(k+l)(k+l+1)} \quad \text{for } N_d \rightarrow \infty \\ &\rightarrow 2mN_d \frac{b_1}{b_1 - d_1} \frac{1}{k(k+1)} \quad \text{for } k \rightarrow \infty. \end{aligned} \quad (\text{E2})$$

Considering only mutations with infinite lineage, one can use the probability mass functions of the pure-birth process and the SFS was shown to be [16]

$$E[S_k^{(new)}] = 2m \frac{b_1}{b_1 - d_1} N_d \left(1 - \frac{1}{N_d}\right)^k \left(\frac{1}{k(k+1)} + \frac{1}{N_d} \frac{1}{k+1}\right). \quad (\text{E3})$$

Since all mutations have an infinite lineage in the pure-birth process, this formula also provides the SFS for the pure-birth process when setting $d_1 = 0$.

Preexisting mutations: Following theorem 1, we have

$$E[S_k^{(pre)}] = \frac{1}{1 - \alpha(t_d)} u_0 p(1 \rightarrow k, t_d) = u_0 (1 - \beta(t_d)) \beta(t_d)^{k-1}, \quad (\text{E4})$$

which is just the number of initial mutations multiplied by the probability of growing from size 1 to size k in time t_d conditioned on survival. $\beta(t_d)$ is defined in eqn. (A2).

Our main interest is in mutation accumulation during tumour progression. Therefore, we work with $u_0 = 0$ in the main text leading to $E[S_k^{(pre)}] = 0$ and $E[S_k^{(det)}] = E[S_k^{(new)}]$.

E.2 Fixed-size expectation

Following eqn. (C1), the expected SFS at fixed size $N = N'_d$ can be expressed in terms of the expected SFS at fixed time $T = t$ by writing

$$\begin{aligned} E[S_k|N = N'_d] &= \int_0^\infty E[S_k|T = t] f_T(t|N = N'_d) dt \\ &\approx \int_0^\infty C \tilde{N}(t) f_T(t|N = N'_d) dt. \end{aligned} \quad (\text{E5})$$

In the second line, we used the solution for large population size ($\tilde{N}(t) \rightarrow \infty$ in eqn. (8)) such that $E[S_k|T = t]$ is proportional to the expected size at time t . Here, C is independent of time and reads

$$C = 2m \sum_{l=0}^{\infty} \frac{\left(\frac{d_1}{b_1}\right)^l}{(k+l)(k+l+1)}. \quad (\text{E6})$$

The expected size at fixed time t_d can be written as

$$N_d = \tilde{N}(t_d) = \frac{1}{\sigma} e^{(b_1-d_1)t} - \frac{\rho}{\sigma} \quad (\text{E7})$$

with $\rho = \frac{d_1}{b_1}$ and $\sigma = \frac{b_1-d_1}{b_1}$ (eqn. 17 in ref. [16]). The integral can be expressed in terms of the moment generating function (eqn. (B5)) such that

$$E[S_k|N = n] = C \left(\frac{1}{\sigma} \phi(b_1 - d_1) - \frac{\rho}{\sigma} \right). \quad (\text{E8})$$

However, we have

$$\phi(b_1 - d_1) = \Gamma \left(1 - \frac{b_1 - d_1}{b_1} \right) \left(\frac{b_1 - d_1}{b_1} N'_d \right)^{\frac{b_1 - d_1}{b_1 - d_1}} \quad (\text{E9})$$

and $\Gamma(0)$ is not defined.

To avoid the convergence problem, we consider the normalized SFS that we define by $E[\tilde{S}_k|T = t] = \frac{E[S_k|T=t]}{E[Z_0(t)|Z_0(t)>0]} = C$. Again, we switch from the fixed-time expectation to the fixed-size expectation, which yields

$$E[\tilde{S}_k|N = N'_d] = \int_0^\infty C f_T(t|N = N'_d) dt = C = E[\tilde{S}_k|T = t]. \quad (\text{E10})$$

We find that the fixed-time expectation and fixed-size solution expectation. The perfect one-to-one correspondence vanishes if we take the exact expression for the fixed-time expectation instead of the large-size approximation in eqn. (8).

F Site frequency spectrum after treatment with homogeneous response

We compute the expected SFS after treatment with homogeneous treatment response. As initial condition, we consider fixed size N'_d and the expected SFS at detection. For the initial SFS, we approximate the fixed-size expectation of the SFS with the fixed-time expectation of the SFS by replacing the expected size $N_d = \tilde{N}(t_d)$ with the fixed-size N'_d in eqn. (8). We denote the treatment time by t_f after which the population conditioned on survival has expected size $N_f = \frac{1}{1-p(N'_d \rightarrow 0)} N'_d e^{(b_2-d_2)t_f}$.

F.1 Decreasing cell populations

We consider decreasing cell populations characterized by $b_2 < d_2$ and $N_f < N'_d$.

Newly emerging mutations: Following theorem 1, we have

$$E[S_k^{(new)}] = \frac{1}{1-p(N'_d \rightarrow 0, t_f)} \int_0^{t_f} 2mb_2 N'_d e^{(b_2-d_2)t'} p(1 \rightarrow k, t-t') dt'. \quad (\text{F1})$$

Here, $p(1 \rightarrow k, t-t')$ is given by eqn. (A7). Using the same transformation as for the exponential growth starting from a single cell (ref. [16] eqn. (B.5) to (B.6)), the integral can be rewritten as

$$E[S_k^{(new)}] = 2mN_f \int_0^{\frac{e^{(b_2-d_2)t_f}-1}{e^{(b_2-d_2)t_f}-\frac{d_2}{b_2}}} \left(1 - \frac{d_2}{b_2}y\right)^{-1} (1-y)y^{k-1} dy. \quad (\text{F2})$$

In general, the upper integral bound does not go to 1 such that we cannot transform the integral into a sum as done in ref. [16]. Instead, we solved the integral numerically.

The situation changes in the limit $b_2 \rightarrow d_2$. Then, the upper integral bound becomes 1 but at the same time the integrand becomes y^{k-1} resulting in $E[S_k^{(new)}] = 2mN_f \frac{1}{k}$. This result coincides with the equilibrium solution for $b_2 = d_2$ that will be further refined in the next subsection (see eqn. (F5)).

In the limit $\frac{d_2}{b_2} \rightarrow \infty$, the upper integral bound converges to 0 whereas the integrand converges to $\frac{1}{1-\infty y}(1-y)y^{k-1}$ resulting in $E[S_k^{(new)}] = 0$. The result is intuitive. If there are no birth events, then there are no new mutations.

Preexisting mutations: Following theorem 1, we have

$$E[S_k^{(pre)}] = \frac{1}{1-p(N'_d \rightarrow 0, t_f)} \sum_{k'=1}^{N'_d} E[S_{k'}^{(det)}] p(k' \rightarrow k, t_f). \quad (\text{F3})$$

Here, $p(k' \rightarrow k, t_f)$ is given by eqn. (A1) parameterized with b_2 and d_2 . By construction, the initial SFS is the SFS at detection whose expectation is given by eqn. (8).

In general, the exact probabilities $p(k' \rightarrow k, t_f)$ take a complicated form. Consequently, the evaluation of $E[S_k^{(pre)}]$ for specific k consists of computing a sum ranging from 0 to N'_d over sums ranging from

0 to $\min(k, k')$. The computational cost is high but can be decreased with the approximation methods discussed in Appendix A.

In the pure-death process with $b_1 = 0$, we can use (A10) that resolves the second sum without the use of approximation methods. Resolving the first sum remains complicated but is possible when we condition on fixed size instead of fixed time (see Appendix G).

F.2 Constant cell populations

Next, we consider cell populations that remain approximately constant characterized by $b_2 = d_2$ and $N'_d \approx N_f$.

Newly emerging mutations: Following theorem 1, we have

$$E[S_k^{(new)}] = \frac{1}{1 - p(N'_d \rightarrow 0, t_f)} \int_0^{t_f} b_2 2m N'_d p(1 \rightarrow k, t_f) dt', \quad (\text{F4})$$

where $p(1 \rightarrow k, t_f)$ are given by eqn. (A8). We solve the integral and obtain

$$E[S_k^{(new)}] = \frac{1}{1 - p(N'_d \rightarrow 0, t_f)} 2mb_2 N'_d \int_0^{t_f} \frac{(b_2 t_f)^{k-1}}{(1 + b_2 t_f)^{k+1}} dt' = 2m N_f \left(\frac{b_2 t_f}{1 + b_2 t_f} \right)^k \frac{1}{k}. \quad (\text{F5})$$

As $t \rightarrow \infty$, the SFS is proportional to $\frac{1}{k}$, which is known as the equilibrium solution from other constant population size models, namely the Moran process or the Wright-Fisher process [20].

Preexisting mutations: Following theorem 1, we have

$$E[S_k^{(pre)}] = \frac{1}{1 - p(N'_d \rightarrow 0, t_f)} \sum_{k'=1}^{N'_d} E[S_k^{(det)}] p(k' \rightarrow k, t_f). \quad (\text{F6})$$

Here, $p(k' \rightarrow k, t_f)$ is given by eqn. (A1) parameterized with b_2 . Again, $E[S_k^{(det)}]$ is given by eqn. (8).

Similar to the decreasing population, the computation of $E[S_k^{(pre)}]$ for one specific k consists of evaluating the sum of sums where the second sum can be resolved using the approximations described in Appendix A.

F.3 Increasing cell populations

Eventually, we consider increasing cell populations characterized by $b_2 > d_2$ and $N_f > N'_d$.

Newly emerging mutations: The exact expressions for increasing populations coincide with the ones for decreasing populations such that

$$\begin{aligned} E[S_k^{(new)}] &= \frac{1}{1 - p(N'_d \rightarrow 0, t_f)} \int_0^{t_f} b_2 2m N'_d e^{(b_2 - d_2)t'} p(1 \rightarrow k, t - t') dt', \\ &= 2m N_f \int_0^{\frac{e^{(b_2 - d_2)t_f} - 1}{e^{(b_2 - d_2)t_f} - \frac{d_2}{b_2}}} \left(1 - \frac{d_2}{b_2} y \right)^{-1} (1 - y) y^{k-1} dy. \end{aligned} \quad (\text{F7})$$

We solved the integral numerically.

In the limit $t_f \rightarrow \infty$, the upper integral bound converges to 1 and following ref. [16] (computations after eqn. (B.6)), we can express the integral as infinite sum

$$E[S_k^{(new)}] = 2mN_f \sum_{l=0}^{\infty} \frac{\left(\frac{d_2}{b_2}\right)^l}{(k+l)(k+l+1)}. \quad (\text{F8})$$

This limit has restricted application. After detection, the cancer is unlikely to continue growing over a long time since the cancer likely becomes lethal by then. However, this scenario is applicable in case birth and death rates are changed before detection, e.g. through activation of the immune system, angiogenesis or other environmental factors.

In the limit $b_2 \rightarrow d_2$, we recover the $E[S_k^{(pre)}] = 2mN_f \frac{1}{k}$ like we did for decreasing populations.

In the limit $d_2 \rightarrow 0$, we have

$$E[S_k^{(new)}] = 2mN_f (1 - e^{-b_2 t_f})^k \left(\frac{1}{k(k+1)} + e^{-b_2 t_f} \frac{1}{k+1} \right). \quad (\text{F9})$$

Preexisting mutations: Again, the expressions for increasing populations coincide with the ones for decreasing populations such that

$$E[S_k^{(pre)}] = \frac{1}{1 - p(N'_d \rightarrow 0, t_f)} \sum_{k=1}^{N'_d} E[S_k^{(det)}] p(k' \rightarrow k, t_f), \quad (\text{F10})$$

where $p(k' \rightarrow k, t_f)$ is given by eqn. (A1) parameterized with b_2 and d_2 , and $E[S_k^{(det)}]$ is given by eqn. (8). Obtaining a single value comes with evaluation of two sums where the second sum can be resolved using approximation methods for $p(k' \rightarrow k, t_f)$ in Appendix A.

F.4 Approximating the fixed-size expectation with the fixed-time expectation

For increasing or decreasing populations, we approximate the expected SFS at fixed size N'_f with the expected SFS at fixed time t_f . Therefore, we transform the treatment time into size after treatment using the expression for the expected size after treatment for time t_f giving us

$$N'_f = \frac{1}{1 - p(N'_d \rightarrow 0, t_f)} N'_d e^{(b_2 - d_2)t_f}. \quad (\text{F11})$$

Inverting the exact expression analytically is complicated. However, in applications we have $N'_d > 10^9$ and given that we still have an observable cancer $N_f > 10^3$. For population sizes of this magnitude, we have $p(N'_d \rightarrow 0, t_f) \approx 0$ such that $N'_f \approx N'_d e^{(b_2 - d_2)t_f}$ and

$$t_f \approx \frac{\log\left(\frac{N_f}{N'_d}\right)}{b_2 - d_2}. \quad (\text{F12})$$

We inserted this expression into the fixed-time expectation to approximate the fixed-size expectation.

G The site frequency spectrum retains its k^{-2} tail after treatment with homogeneous response

After treatment of cancers with homogeneous treatment response, the expected SFS retains its power law tail $E[S_k] \sim k^{-2}$. We assume that mutations that occur in less than 1% of the population are not observable due to sequencing limitations. Thus, we define the tail distribution by S_k with $k \geq k_{\min} = 1\%N$. First, we show that the k^{-2} tail is retained in simulated data that have population sizes in the magnitudes $10^4 - 10^5$. Second, we make a parameter estimation for realistically large tumors and argue why the k^{-2} tail is retained. We show that newly emerging mutations remain at low unobservable frequencies, and observable preexisting mutations are usually large and behave nearly deterministic.

G.1 Observation of the k^{-2} tail in simulated data

We observe a power law of the form $E[S_k] \sim k^{-\gamma}$ from fig. 2D-F by noting of the plots build a straight line in the loglog-plot for sufficiently large k . Next, we infer the exponent γ from the simulated data using the maximum likelihood estimator from ref. [55] (see *Methods*). We first infer exponents of simulated data at detection (fig. S1A-D). As expected, we obtain exponents $\gamma \approx 2$. Depending on the cutoff k_{\min} , we observe a small bias towards larger exponents (fig. S1B and D). This bias may be explained by the high variability in the largest site frequencies. We also note that strictly speaking, we did not show that the SFS of a single realization follows a power law distribution but the expected SFS does.

Next, we inferred the exponents in simulated data after treatment with homogeneous response. Again, we find exponents $\gamma \approx 2$ with small biases towards larger exponents consistent with the inference artefact that we also obtained for the SFS at detection. We conclude the observation of the $S_k \sim k^{-2}$ tail after treatment with homogeneous response.

G.2 Estimating biologically reasonable parameters

We assume that a cancer is detected at population sizes in the range $N_d = 10^9 - 10^{12}$. Assuming a typical division time $\tau_2 = 5$ days, the birth rate can be estimated to be $b_1 = \frac{\ln(2)}{\tau_2} \approx 0.14 \text{ days}^{-1}$. For cancers growing to detection size in around 5-15 years, the growth rate is in the range from $0.005 - 0.015 \text{ days}^{-1}$ translating into death rates close to the birth rate in the range $d_2 = 0.125 - 0.135 \text{ days}^{-1}$. In the following, we work with an intermediate death rate $d_1 = 0.13 \text{ days}^{-1}$.

We set treatment time to $t_f = 180$ days. We assume that treatment affects the death rate only such that we keep the birth rate $b_2 = 0.14 \text{ days}^{-1}$. The death rate is adapted according to the three treatment scenarios. In an effective treatment, the cancer may decrease to 1% of its original size translating into death rate $d_2 \approx 0.166 \text{ days}^{-1}$. For approximately constant disease, we have $d_2 = 0.14 \text{ days}^{-1}$. For ineffective treatment, we consider the case in which treatment has no effect at all such that $d_2 = 0.13 \text{ days}^{-1}$.

G.3 Decreasing cell populations

In the pure-death process, the expected SFS has a k^{-2} tail. If $b_2 = 0.0$, then there are no newly emerging mutations such that $E[S_k^{(tot)}] = E[S_k^{(pre)}]$. Instead of conditioning the expectation on fixed time t_f , let us condition on fixed size N'_f starting from fixed detection size N'_d . In the pure-death process, growing from size k' to size k during treatment is the same as sampling k cells from k' cells without replacement. Thus, the probability to grow to size k is given by the hypergeometric distribution,

$$p(k' \rightarrow k, N'_d \rightarrow N'_f) = \frac{\binom{k'}{k} \binom{N'_f - k'}{N'_d - k}}{\binom{N'_f}{N'_d}}. \quad (\text{G1})$$

If $N'_d \gg N'_f$, the hypergeometric distribution can be approximated by the binomial distribution (sampling with replacement), which is

$$p(k' \rightarrow k, N'_d \rightarrow N'_f) \approx \binom{N'_f}{k} \left(\frac{k'}{N'_d}\right)^k \left(1 - \frac{k'}{N'_d}\right)^{N'_f - k}. \quad (\text{G2})$$

Interestingly, we arrive at a binomial distribution like the probability mass function for fixed-time in eqn. (A10), although the binomial distribution is different parameterized.

From here, we can follow the derivation of Durrett who studied the form of the sampled SFS to describe frequencies in a sample (theorem 3 in ref. [15]). We transform the sum for the SFS into an integral and focus on the tail of the SFS at detection such that

$$\begin{aligned} E[S_k^{(tot)}] &= \sum_{k'=1}^{N'_d} E[S_{k'}^{(det)}] p(k' \rightarrow k, N'_d \rightarrow N'_f) \\ &\approx \int_1^{N'_d} \frac{2mN'_d}{k'^2} \binom{N'_f}{k} \left(\frac{k'}{N'_d}\right)^k \left(1 - \frac{k'}{N'_d}\right)^{N'_f - k} dk'. \end{aligned} \quad (\text{G3})$$

We substitute the initial mutant size with its frequency, $x = \frac{k'}{N'_d}$ and $dx = \frac{1}{N'_d} dk'$. Furthermore, we assume $N'_d \rightarrow \infty$ such that the lower integral boundary goes to 0. We have

$$E[S_k^{(tot)}] = \int_0^1 \frac{2m}{x^2 N'_d} \binom{N'_f}{k} x^k (1-x)^{N'_f - k} (N'_d dx). \quad (\text{G4})$$

Using that $\int_0^1 x^a (1-x)^b dx = \frac{a!b!}{(a+b+1)!}$ (see Lemma 2 in ref. [15]), we obtain

$$E[S_k^{(tot)}] = 2mN'_f \frac{1}{k(k-1)} \quad \text{for } 2 \geq k \geq N'_f, \quad (\text{G5})$$

which scales as k^{-2} for large k . Noteworthy, the continuity approximation works only for sufficiently large k .

The SFS of preexisting mutations retains its k^{-2} tail. Consider a cancer cell population with detection size $N_d = 10^{12}$ that after treatment for time $t_f = 180$ days shrinks to 1% of its original size, $N_f = 10^{10}$. We compute the mean and the standard deviation for a single site frequency that is initially

in 1% of the cells, i.e. $k' = 10^{10}$. Inserting the birth and death rates from our parameter estimation $b_2 = 0.14 \text{ days}^{-1}$ and $d_2 = 0.166 \text{ days}^{-1}$ into eqn. (A4), we have

$$\begin{aligned} m(t_f) &= k' e^{(b_2 - d_2)t_f} = 10^8 \\ \sigma(t_f) &= \sqrt{k' \frac{b_2 + d_2}{b_2 - d_2} e^{(b_2 - d_2)t_f} (e^{(b_2 - d_2)t_f} - 1)} \approx 3.4 \times 10^4 \end{aligned} \quad (\text{G6})$$

Starting from a large populations, we know that the probability mass function takes a bell-shaped curve as long as we are not too close to the extinction boundary [13]. Observing that the standard deviation is multiple magnitudes lower than the mean, we conclude a sharply peaked probability mass function that we interpret as approximately deterministic behaviour.

If the growth dynamics of site frequencies is deterministic, it is straightforward to show that the initial shape of the SFS is retained. Focusing on the tail of the SFS at detection and assuming deterministic growth during treatment, we have

$$E[S_k^{(pre)}] = \sum_{k'=1}^{N'_d} 2m N_d \frac{d_1}{b_1 - d_1} \frac{1}{k'^2} \delta(k - k' e^{(b_2 - d_2)t_f}) = 2m \frac{d_1}{b_1 - d_2} N_f \frac{1}{k^2}. \quad (\text{G7})$$

Here, we use deterministic growth for the total population $N_f = N_d e^{(b_2 - d_2)t_f}$ and approximate the probability mass function by $\delta(k - k' e^{(b_2 - d_2)t_f})$ that is 1 if $k - k' e^{(b_2 - d_2)t_f} = 0$ and 0 otherwise.

Newly emerging mutations are restricted to small frequencies. Since $b_2 < d_2$, newly emerging mutations survive only due to drift. Given $b_2 = 0.14$ and $d_2 \approx 0.166$, the probability to survive time t_f is only $1 - \alpha(t_f) \approx 0.000243$. Further, the probability to reach larger sizes $k > 1$ is decreasing with time.

G.4 Constant cell populations

The SFS of preexisting mutations retains its k^{-2} tail. Consider a cancer cell population with detection size $N_d = 10^{12}$ that remains approximately constant over treatment time $t_f = 180$ days. We compute the mean and the standard deviation for a single site frequency that is initially in 1% of the cells, i.e. $k' = 10^{10}$. Inserting the birth and death rates from our parameter estimation $b_2 = d_2 = 0.14 \text{ days}^{-1}$ into eqn. (A4), we have

$$\begin{aligned} m(t) &= k' = 10^{10} \\ \sigma(t) &= \sqrt{2kb_2 t_f} \approx 7.1 \times 10^5 \end{aligned} \quad (\text{G8})$$

With analogous argumentation to the decreasing population, we conclude that the growth is approximately deterministic and the SFS of preexisting mutations retains its k^{-2} tail.

Newly emerging mutations are restricted to small frequencies. Using the same parameters as before, we compute the probability that a newly emerging mutation grows to abundance $k > 100$ during treatment time $t_f = 180$ days

$$1 - \sum_{k'=0}^{100} p(1 \rightarrow k', t_f) = 1 - \frac{bt}{1 + bt} - \sum_{k'=1}^{100} \frac{(b_2 t_f)^{k'-1}}{(1 + b_2 t_f)^{k'+1}} \approx 1 - 0.9618 - 0.0374 = 0.0008. \quad (\text{G9})$$

The chance is very small, and we conclude that genetic drift alone is not enough to grow to observable site frequencies in the order $k = 10^{10}$.

This is also reflected in the expression $E[S_k^{(new)}] = 2mN_f \left(\frac{b_2 t_f}{b_2 t_f + 1}\right)^k \frac{1}{k}$, which has an exponential cutoff with k . Inserting the parameters above, we note that there is a dominating exponential cutoff with k that reads

$$\left(\frac{b_2 t_f}{b_2 t_f + 1}\right)^k = 0.96^{10^{10}}, \quad (\text{G10})$$

which is extremely small such that $E[S_k^{(new)}]$ is virtually zero for large k .

G.5 Increasing cell populations

The SFS of preexisting mutations retains its k^{-2} tail. Consider a cancer cell population with detection size $N_d = 10^{11}$ that after ineffective treatment for time $t_f = 180$ days continues increasing to approximately $N_f \approx 6 \times 10^{11}$. We compute the mean and the standard deviation for a single site frequency that is initially in 1% of the cells, i.e. $k' = 10^9$. Inserting the birth and death rates from our parameter estimation $b_2 = 0.14 \text{ days}^{-1}$ and $d_2 = 0.13 \text{ days}^{-1}$ into eqn. (A4), we have

$$\begin{aligned} m(t_f) &= k' e^{(b_2 - d_2)t_f} \approx 6.0 \times 10^9 \\ \sigma(t_f) &= \sqrt{k' \frac{b_2 + d_2}{b_2 - d_2} e^{(b_2 - d_2)t_f} (e^{(b_2 - d_2)t_f} - 1)} \approx 9.1 \times 10^5 \end{aligned} \quad (\text{G11})$$

With analogous argumentation to the decreasing population, we conclude that the growth is approximately deterministic and the SFS of preexisting mutations retains its k^{-2} tail.

Newly emerging mutations are restricted to small frequencies. In contrast to decreasing and the constant populations, newly emerging mutations grow in their abundance not only to due to genetic drift. However, since the cancer is usually detected at very large sizes, and preexisting site frequencies grow at the same rate, it is highly unlikely that newly emerging mutations grow to 1% of the population. With the estimated birth and death rate, we have mean and standard deviation for the abundance of a newly emerging mutation growing for time $t_f = 180$ days,

$$\begin{aligned} m(t_f) &= e^{(b_2 - d_2)t_f} \approx 6.0 \\ \sigma(t_f) &= \sqrt{\frac{b_2 + d_2}{b_2 - d_2} e^{(b_2 - d_2)t_f} (e^{(b_2 - d_2)t_f} - 1)} \approx 28.7. \end{aligned} \quad (\text{G12})$$

Although the standard deviation is very large compared to the mean, the likelihood to reach observable size $k_{\min} \approx 6.0 \times 10^9$ is again virtually zero.

H Proof of theorem 2

Part 1: We prove eqn. (15) building up on the proof of theorem 1. Restricted (but not conditioned) on survival of the entire population $\{Z_0(t) > 0\}$, the expected number of mutations that emerged in the

time interval $[t', t' + dt']$ and survive in the remaining time $t - t'$ is

$$\begin{aligned} E \left[\left(B^{(new)}(t' + dt') - B^{(new)}(t) \right) 1_{\{Z_0(t) > 0\}} \right] &= \sum_{k \geq 1} E \left[\left(S_k^{(new)}(t' + dt') - S_k^{(new)}(t) \right) 1_{\{Z_0(t) > 0\}} \right] \\ &= bE[\Omega] E[Z_0(t')] \sum_{k \geq 1} p(1 \rightarrow k, t - t') \\ &= bE[\Omega] E[Z_0(t')] (1 - p(1 \rightarrow 0, t - t')). \end{aligned} \quad (\text{H1})$$

In the first equality, we used that $B^{(new)} = \sum_{k \geq 1} S_k^{(new)}$ together with linearity of the expectation. The second equality is equivalent to the first equation in the proof of theorem 1 taken from ref. [16]. Again, Ω is the number of mutations generated in a division such that $E[\Omega] = 2m$ and the expected population size is $E[Z_0(t')] = N(t') = N_0 e^{(b-d)t'}$. After conditioning the total population to survive time t , we have

$$E \left[\left(S^{(new)}(t' + dt') - S^{(new)}(t) \right) | Z_0(t) > 0 \right] = \frac{1}{1 - p(N_0 \rightarrow 0, t)} 2mbN(t')(1 - p(1 \rightarrow 0, t - t')). \quad (\text{H2})$$

We integrate over the time interval of interest $t' \in [0, t]$ giving us

$$E[B^{(new)}] = \frac{1}{1 - p(N_0 \rightarrow 0, t)} \int_0^t 2mbN(t')(1 - p(1 \rightarrow 0, t - t')) dt', \quad (\text{H3})$$

which makes the proof complete.

The expected number of divisions is

$$E[R^{(new)}] = E[B^{(new)}]/m = \frac{1}{1 - p(N_0 \rightarrow 0, t)} \int_0^t 2bN(t')(1 - p(1 \rightarrow 0, t - t')) dt'. \quad (\text{H4})$$

Part 2: Next, we prove eqn. (16). Analogously to the proof of the second part of theorem 1, we start by writing the initial BWD as $W_k^{(init)}$, label its L branches with index $i = 1, 2, \dots, L$ and their branch widths with k_i . During time t , branches may go extinct and we count branches leading to at least one living cell that is

$$R^{(pre)} = \sum_{i=1}^L 1_{\{k_i \rightarrow k > 0\}} = \sum_{k'=1}^{N_0} \sum_{j=1}^{W_{k'}^{(init)}} 1_{\{k_j \rightarrow k > 0\}}. \quad (\text{H5})$$

Here, $1_{\{k_i \rightarrow k > 0\}} = 1$ if branch i of initial width k_i grows to size $k > 0$ and $1_{\{k_i \rightarrow k\}} = 0$ otherwise. Again, we used that the L branches are partitioned into N_0 classes $\{W_{k'}^{(init)}\}_{k'=1}^{N_0}$ according to their initial width k' .

Taking the expectation and using the same steps as in the proof of theorem 1, we have

$$\begin{aligned} E[B^{(pre)}] &= E \left[\sum_{k'=1}^{N_0} \sum_{j=1}^{W_{k'}^{(init)}} 1_{\{k_j \rightarrow k > 0\}} \right] \\ &= \sum_{k'=1}^{N_0} E \left[\sum_{j=1}^{W_{k'}^{(init)}} 1_{\{k_j \rightarrow k > 0\}} \right] \\ &= \sum_{k'=1}^{N_0} E[W_{k'}^{(init)}] (1 - p(k' \rightarrow 0, t)). \end{aligned} \quad (\text{H6})$$

We substitute $E[R^{(pre)}] = E[B^{(pre)}]/m$ and $E[W_k^{(init)}] = E[S_k^{(init)}]/m$ and condition the entire population on survival such that

$$E[B^{(pre)}] = \frac{1}{1 - p(N_0 \rightarrow 0, t)} \sum_{K=1}^{N_0} E[S_{k'}^{(init)}] (1 - p(k' \rightarrow 0, t)), \quad (\text{H7})$$

which is the desired result.

I The total mutational burden at detection

Pure-birth process If we consider the fixed-size limit, we know that there are exactly $N_d - 1$ divisions that a population must undergo to grow from 1 cell to N_d cells. Thus, the expected tMB is

$$E[B^{(det)}] = (N_d - 1)2m \approx 2N_d m \quad (\text{I1})$$

Furthermore, since each division comes with $\mu_i \sim \text{Poiss}(m)$ mutations, we know that the tMB is Poisson distributed

$$B^{(det)} \sim \text{Poiss}(2N_d m). \quad (\text{I2})$$

If there were u_0 mutations in the ancestor cell, one simply needs to add them as constant outside the Poisson distribution, $B^{(det)} \sim \text{Poiss}(2N_d m) + u_0$.

In the fixed-time condition, we consider a population growing time t_d at which we expect population size $N_d = E[Z_0(t_d)]$. This translates into an expected number of $N_d(t_d) - 1$ divisions similar to the fixed-size expectation. However, the distribution of $B^{(det)}$ is not Poisson distributed anymore. The number of divisions differs between realizations thus leading to additional stochastic effect.

Birth-death process Following theorem 2, starting from a single cell with no mutations that grows for time t_d and is conditioned to survive, we have

$$\begin{aligned} E[B^{(new)}] &= \frac{1}{1 - p(k \rightarrow 0, t_d)} \int_0^{t_d} b_1 2m N(t) \times (1 - p(1 \rightarrow 0, t_d - t)) dt \\ &= -2m \frac{e^{(b_1 - d_1)t_d}}{1 - p(k \rightarrow 0, t_d)} \frac{b_1}{d_1} \log \left(\frac{b_1 - d_1}{b_1 - d_1 e^{-(b_1 - d_1)t_d}} \right) \\ &= 2m N_d \frac{b_1}{d_1} \log \left(\frac{b_1}{b_1 - d_1} \frac{1}{1 - p(1 \rightarrow 0, t_d - t)} \right). \end{aligned} \quad (\text{I3})$$

We can further manipulate the expression to show agreement with ref. [16]. We have

$$\begin{aligned} E[B^{(new)}] &= -2m N_d \frac{b_1}{d_1} \log \left(\frac{b_1 - d_1}{b_1 - d_1 (N_d \sigma + \rho)^{-1}} \right) \\ &= -2m N_d \frac{b_1}{d_1} \log \left(\sigma + \frac{\rho}{N_d} \right) \\ &\sim -2m N_d \frac{\log(\sigma)}{\rho} \quad \text{for } N_d \rightarrow \infty, \end{aligned} \quad (\text{I4})$$

where we used that $N_d = \frac{1}{1 - p(1 \rightarrow 0, t_d)} e^{(b_1 - d_1)t_d} = \frac{e^{(b_1 - d_1)t_d + \rho}}{\sigma}$.

If there be u_0 mutations preexisting in the first cell, they simply need to be added as a constant such that total tMB is $E[B^{(tot)}] = E[B^{(new)}] + u_0$.

J The total mutational burden after treatment with homogeneous response

We compute the expected tMB after treatment with homogeneous treatment response. As initial condition, we consider fixed size N'_d and the expected SFS at detection. For the initial SFS, we approximate the fixed-size expectation of the SFS with the fixed-time expectation of the SFS by replacing the expected size $N_d = \tilde{N}(t_d)$ with the fixed-size N'_d in eqn. (8). We denote the treatment time by t_f after which the population conditioned on survival has expected size $N_f = \frac{1}{1-p(N'_d \rightarrow 0)} N'_d e^{(b_2-d_2)t_f}$. Since the extinction probability appears in the final expressions, it is convenient to write $\alpha(t)$ for $p(1 \rightarrow 0, t)$, and $\alpha^a(t)$ for $p(a \rightarrow 0, t)$.

J.1 Decreasing cell populations

We consider decreasing cell populations characterized by $b_2 < d_2$ and $N_f < N'_d$.

Newly emerging mutations: Following theorem 2, we have

$$\begin{aligned} E[B^{(new)}] &= \frac{1}{1 - \alpha^{N'_d}(t_f)} \int_0^{t_f} 2mb_2 N'_d e^{(b_2-d_2)t} (1 - \alpha(t_f - t)) dt \\ &= -2mN_f \frac{b_2}{d_2} \log \left(\frac{b_2 - d_2}{b_2 - d_2 e^{-(b_2-d_2)t_f}} \right) \\ &= 2mN_f \frac{b_2}{d_2} \log \left(\frac{b_2}{b_2 - d_2} - e^{-(b_2-d_2)t_f} \frac{d_2}{b_2 - d_2} \right) \\ &= 2mN_f \frac{b_2}{d_2} \log \left(\frac{b_2}{b_2 - d_2} - \frac{1}{1 - \alpha(t_f)} \frac{N'_d}{N_f} \frac{d_2}{b_2 - d_2} \right), \end{aligned} \tag{J1}$$

For $t_f \rightarrow \infty$, the expression coincides with the the tMB at detection. In particular, it is independent of the initial population size N'_d . This makes sense. Newly emerging mutations are only short-lived and the major contributions to $B^{(new)}$ are very recent mutations.

Preexisting mutations: Following theorem 2, we have

$$E[B^{(pre)}] = \frac{1}{1 - \alpha^{N'_d}(t_f)} \sum_{k=1}^{N'_d} E[S_k^{(det)}] (1 - \alpha^k(t_f)). \tag{J2}$$

Here, $\alpha(t)$ is given by eqn. (1) parameterized with b_2 and d_2 . By construction, the initial SFS is the SFS at detection whose expectation is given by eqn. (8).

J.2 Constant cell populations

Next, we consider cell populations that remain approximately constant characterized by $b_2 = d_2$ and $N'_d \approx N_f$.

Newly emerging mutations: Following theorem 2, we have

$$\begin{aligned} E[B^{(new)}] &= \frac{1}{1 - \alpha^{N'_d}(t_f)} \int_0^{t_f} (2mb_2N'_d) (1 - \alpha(t_f - t)) dt \\ &= 2mN_f \log(1 + b_2t_f). \end{aligned} \quad (\text{J3})$$

Preexisting mutations: Following theorem 2, we have

$$E[B^{(pre)}] = \frac{1}{1 - \alpha^{N'_d}(t_f)} \sum_{k=1}^{N'_d} E[S_k^{(det)}] (1 - \alpha^k(t_f)), \quad (\text{J4})$$

Again, $\alpha(t)$ is given by eqn. (1) parameterized with b_2 and d_2 . The initial SFS is given by eqn. (8).

J.3 Increasing cell populations

Eventually, we consider increasing cell populations characterized by $b_2 > d_2$ and $N_f > N'_d$.

Newly emerging mutations: The expressions for increasing populations coincide with the ones for decreasing populations such that

$$\begin{aligned} E[B^{(new)}] &= \frac{1}{1 - \alpha^{N'_d}(t_f)} \int_0^{t_f} 2mb_2N'_d e^{(b_2-d_2)t} (1 - \alpha(t_f - t)) dt \\ &= 2mN_f \frac{b_2}{d_2} \log\left(\frac{b_2}{b_2 - d_2} - \frac{1}{1 - \alpha(t_f)} \frac{N'_d}{N_f} \frac{d_2}{b_2 - d_2}\right). \end{aligned} \quad (\text{J5})$$

In the limit $\frac{N'_d}{N_f} \rightarrow 0$, the expression is independent

Preexisting mutations: Following theorem 2, we have

$$E[B^{(pre)}] = \frac{1}{1 - \alpha^{N'_d}(t_f)} \sum_{k=1}^{N'_d} E[S_k^{(det)}] (1 - \alpha^k(t_f)). \quad (\text{J6})$$

Once more, $\alpha(t)$ is given by eqn. (1) parameterized with b_2 and d_2 . The initial SFS is given by eqn. (8).

J.4 Approximation for preexisting mutations

We resolve the sum for $E[B^{(pre)}]$ using diverse approximation. In this section, the extinction probability will always be the extinction probability after time t_f such that we write $\alpha = \alpha(t_f)$ that are parameterized with b_2 and d_2 .

Consider $E[S_k^{(det)}] \sim \frac{1}{k^2}$ as expected SFS at detection. For large site frequencies k , the SFS at detection is approximately $E[S_k^{(det)}] \approx \frac{b_1}{b_1-d_1} 2mN'_d \frac{1}{k^2}$. Using this expression for $E[B^{pre}]$ and with help of *Wolfram Mathematica*, we obtain

$$\begin{aligned} E[B^{(pre)}] &= \frac{1}{1 - \alpha^{N'_d}(t_f)} \sum_{k=1}^{N'_d} \frac{b_1}{b_1 - d_1} 2mN'_d \frac{1}{k^2} (1 - \alpha^k) \\ &= \frac{1}{1 - \alpha^{N'_d}(t_f)} \frac{b_1}{b_1 - d_1} 2mN'_d \left(\alpha^{N'_d+1} \Phi(\alpha, 2, N'_d + 1) + H_{N'_d}^{(2)} - \text{Li}_2(\alpha) \right), \end{aligned} \quad (\text{J7})$$

where $\Phi(z, s, a) = \sum_{j=0}^{\infty} \frac{z^j}{(j+a)^s}$ is the Lerch phi, $\text{Li}_n(z) = \sum_{j=0}^{\infty} \frac{z^j}{j^n}$ is the polylogarithm and $H_n^{(2)} = \sum_{j=1}^n \frac{1}{j^2}$ are generalized harmonic numbers. For sufficiently large N'_d , the first term in the parentheses becomes very small and we are left with

$$E[B^{(pre)}] \approx \frac{b_1}{b_1 - d_1} 2mN'_d \left(H_{N'_d}^{(2)} - \text{Li}_2(\alpha) \right). \quad (\text{J8})$$

Notably, the $H_{N'_d}^{(2)}$ is time-independent and positive whereas $-\text{Li}_2(\alpha)$ is time-dependent and negative. We can interpret the first term as the number of mutations at detection which we can also obtain directly by summing over $\frac{1}{k^2}$. Then, the second term describes the number of mutations that are lost over time.

Consider $E[S_k^{(det)}] \sim \frac{1}{k^2+k}$ as expected SFS at detection. Most mutations occur at low abundance where $E[S_k^{(det)}] \sim \frac{1}{k}$ is not valid. To get one step closer to the exact expected SFS at detection, we next consider $E[S_k^{(det)}] \approx 2mN'_d \frac{b_1}{b_1-d_1} \frac{1}{k^2+k}$. Using *Wolfram Mathematica*, we obtain

$$\sum_k^{N'_d} \frac{1}{k^2+k} (1-\alpha)^k = \alpha^{N'_d+1} \Phi(\alpha, 1, N'_d+1) - \alpha^{N'+1} \Phi(\alpha, 1, N'_d+2) + \frac{(\alpha-1) \log(1-\alpha)}{\alpha} - \frac{1}{N'_d+1} \quad (\text{J9})$$

For large N'_d , the first two terms and the last term become very small leaving us with

$$\sum_k^{N'_d} \frac{1}{k^2+k} (1-\alpha)^k \approx \frac{(\alpha-1) \log(1-\alpha)}{\alpha}. \quad (\text{J10})$$

Using this expression to compute the tMB of preexisting mutations, we obtain

$$E[B^{(pre)}] \approx \frac{b_1}{b_1-d_1} 2mN'_d \frac{(\alpha-1) \log(1-\alpha)}{\alpha}. \quad (\text{J11})$$

Heuristic adjustment at $t_f = 0$. At $t_f = 0$, the extinction probability is $\alpha = 0$, and

$$\lim_{\alpha \rightarrow 0} \frac{(\alpha-1) \log(1-\alpha)}{\alpha} = 1 \quad (\text{J12})$$

such that $E[B^{(pre)}] \approx \frac{b_1}{b_1-d_1} 2mN'_d$ at time $t_f = 0$. This is not accurate for the general birth-death process as we know from the section on the tMB at detection. Requiring a correct result at time $t_f = 0$, we make a heuristic correction and write

$$E[B^{(pre)}] \approx E[B^{(det)}] \frac{(\alpha-1) \log(1-\alpha)}{\alpha}, \quad (\text{J13})$$

where $E[B^{(det)}]$ is given by eqn. (17).

K Proof of theorem 3

Part 1: We prove eqn. (22). Taking the expectation of the scMB described by eqn. (21), we have

$$\begin{aligned}
 E[M_j] &= E \left[\sum_{l=1}^{\infty} \sum_{m=1}^{D_l} 1_{\{\sum_{n=1}^l \mu_n = j\}} \right] \\
 &= \sum_{l=1}^{\infty} E \left[\sum_{m=1}^{D_l} 1_{\{\sum_{n=1}^l \mu_n = j\}} \right] \\
 &= \sum_{l=1}^{\infty} E[D_l] E \left[1_{\{\sum_{n=1}^l \mu_n = j\}} \right] \\
 &= \sum_{l=1}^{\infty} E[D_l] \Pr \left(\sum_{n=1}^l \mu_n = j \right) \\
 &= \sum_{l=1}^{\infty} E[D_l] \frac{(lm)^k e^{-lm}}{k!}.
 \end{aligned} \tag{K1}$$

Part 2: Next, we prove eqn. (23). We denote the number of cells with l divisions at time t by $D_l(t)$. We follow the derivation of ref. [5]. We start from a population with N_0 cells at time $t = 0$ such that $D_0(t = 0) = N_0$. Considering only the mean behaviour, we write the following differential equations for D_l

$$\begin{aligned}
 \frac{D_0}{dt} &= -(b+d)D_0 \\
 \frac{D_l}{dt} &= -(b+d)D_l + 2bD_{l-1},
 \end{aligned} \tag{K2}$$

which has solutions

$$\begin{aligned}
 D_0(t) &= N_0 e^{-(b+d)t} \\
 D_l(t) &= \frac{(2bt)^l}{l!} D_0(t)
 \end{aligned} \tag{K3}$$

The expected population size after time t is $N = N_0 e^{(b-d)t}$, which leads to the desired result.

L Average number of divisions in the population

L.1 Increase of the average generation per division

The following derivation was motivated by ref. [9]. Consider a population of N cells, where the number of divisions that each cell has undergone are saved in a vector $\vec{x} = [x_1, \dots, x_N]$ with mean $\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i$. After one division, the population has increased to size $N+1$ and we have division vector $\vec{y} = [y_1, \dots, y_{N+1}]$ with mean $\bar{y} = \frac{1}{N+1} \sum_{i=1}^{N+1} y_i$. Let us denote the index of the dividing cell by j such that $x_i = y_i$ for

$i \neq j$ and $y_j = y_{N+1} = x_j + 1$. Then, we have

$$\begin{aligned}\bar{y} &= \frac{1}{N+1} \sum_{i=1}^{N+1} y_i \\ &= \frac{1}{N+1} (N\bar{x} + x_j + 2)\end{aligned}\tag{L1}$$

If we now average over many realizations, then

$$\begin{aligned}E[\bar{y}] &= E\left[\frac{N\bar{x} + x_j + 2}{N+1}\right] \\ &= \frac{NE[\bar{x}] + E[x_j] + 2}{N+1}\end{aligned}\tag{L2}$$

Now, note that

$$E[\bar{x}] = E\left[\frac{1}{N} \sum_{i=1}^N x_i\right] = \frac{1}{N} NE[x_i] = E[x_i],\tag{L3}$$

such that we have

$$E[\bar{y}] = E[\bar{x}] + \frac{2}{N+1}.\tag{L4}$$

Thus, the average increase for going from size N to size $N+1$ is given by $\frac{2}{N+1}$.

L.2 Fixed-size solution in the pure-birth process

For a pure-birth process starting from 1 cell with 0 divisions, we have $N-1$ divisions to grow to size N where the i -th division increases the generation average by $\frac{1}{i+1}$ such that the expected average number of division reads

$$E[\bar{l}] = \frac{2}{1+1} + \frac{2}{2+1} + \frac{2}{3+1} \dots = \sum_{i=1}^{N-1} \frac{2}{i+1} = 2(H_N - 1).\tag{L5}$$

More generally starting from N_0 cells with 0 divisions growing to size N , we have

$$E[\bar{l}] = \frac{2}{N_0+1} + \frac{2}{N_0+2} + \dots = \sum_{i=1}^{N-1} \frac{2}{N_0+i} = 2(H_N - H_{N_0}).\tag{L6}$$

L.3 Fixed-time solution in the pure-birth process

In the spirit of the proof of theorem 1 and 2 but without strict proof, we write the increase of the average number of divisions in time interval $[t', t'+dt']$ as

$$E[\bar{l}(t'+dt') - \bar{l}(t')] = bN(t') \frac{2}{N(t') + 1}.\tag{L7}$$

We restrict ourselves to the pure-birth process as conditioning on survival complicates the derivation. After summing over the entire time interval $[0, t]$, we have

$$E[\bar{l}] = 2b \int_0^t \frac{N(t')}{N(t') + 1} dt'.\tag{L8}$$

Taking $N(t) = N_0 e^{bt}$, this reads

$$E[\bar{l}] = 2bt_{\text{corr}} \quad \text{with} \quad t_{\text{corr}} = \frac{\log(bN_0 e^{bt} + b) - \log(b(N_0 + 1))}{b}. \quad (\text{L9})$$

L.4 Comparison between mean-field solution and exact solution in the pure-birth process

Fixed-time solutions: Interestingly, even as time goes to infinity, there is a constant off-set between eqn. L9 and the mean field solution $E[\bar{l}^{(mf)}] = 2bt$ from theorem 3. We have

$$\lim_{t \rightarrow \infty} E[\bar{l}^{(mf)}] - E[\bar{l}^{(exact)}] = 2 \ln \left(\frac{N_0 + 1}{N_0} \right) \quad (\text{L10})$$

For $N_0 = 1$, this yields a constant off-set of $\ln(4) \approx 1.386$, and for $N_0 \rightarrow \infty$, the difference is vanishing and we have $E[\bar{l}^{(mf)}] = E[\bar{l}^{(exact)}]$.

Fixed-size solutions: We obtain a similar result in the fixed size condition. We transform the mean-field fixed-time solution into a fixed-size solution using according to Appendix C such that

$$E[\bar{l}^{(mf)}|N] = \int_0^\infty (2bt) f_T(t) dt = 2(\gamma + \ln(N)). \quad (\text{L11})$$

We note that $f_T(t)$ assumes that $N_0 = 1$. To compare this to the exact fixed-size solution, we approximate the harmonic number with its asymptotic limit $H_N = \gamma + \ln(N) + O\left(\frac{1}{N}\right)$ giving us

$$E[\bar{l}^{(exact)}|N] = 2(\gamma + \ln(N) - 1) + O\left(\frac{1}{N}\right) \quad (\text{L12})$$

such that

$$E[\bar{l}^{(mf)}|N] - E[\bar{l}^{(exact)}|N] = 2 + O\left(\frac{1}{N}\right). \quad (\text{L13})$$

The asymptotic solution differs from the exact solution by 2, which agrees with the correction term computed for the critical branching process in ref. [36]. Further, computer simulations show that the correction term 2 is in good agreement with the birth-death process starting from a single cell.

M Size distribution of resistant clones

We consider sensitive cells growing with rates $b_{s,1}$ and $d_{s,1}$ in the absence of treatment and $b_{s,2}$ and $d_{s,2}$ during treatment. During division of sensitive cells, each daughter cells acquires resistance mutations at rate $\nu \ll 1$. Resistant cells grow with rates $b_{r,1}$ and $d_{r,1}$ in the absence of treatment and $b_{r,2}$ and $d_{r,2}$ during treatment. Since the generation of resistance mutations is really small and our model does not include any other interactions between sensitive and resistant cells, we can approximate the growth of sensitive cells using growth rates $b_{s,1} - d_{s,1}$ and $b_{s,2} - d_{s,2}$ rather than $b_{s,1} - d_{s,1} - \nu$ and $b_{s,2} - d_{s,2} - \nu$. We further neglect the chance that more than one resistance mutation emerge in one daughter during one division. Such events come with probability of the order ν^2 . Using these approximations, the CSD can be computed in the same way as the SFS.

Clone size distribution at detection. Before treatment, resistant clones are generated at rate $2\nu e^{(b_{s,1}-d_{s,1})t}$ and grow to size κ with probability $p(1 \rightarrow \kappa, t_d - t)$. Following theorem 1, we have

$$E[\Theta_{\kappa}^{(det)}] = \frac{1}{1 - p(1 \rightarrow 0, t_d)} \int_0^{t_d} 2\nu e^{(b_{s,1}-d_{s,1})t} p(1 \rightarrow \kappa, t_d - t) dt. \quad (M1)$$

Whereas the term $p(1 \rightarrow \kappa, t_d - t)$ described the growth of resistant clones and thus must be parameterized with $b_{r,1}$ and $d_{r,1}$, the term $\frac{1}{1 - p(1 \rightarrow 0, t_d)}$ describes the conditioning on survival of the entire population that is appropriately parameterized with $b_{s,1}$ and $d_{s,1}$.

If sensitive and resistant cells share the same growth parameters in the absence of treatment, i.e. $b_{s,1} = b_{r,1} := b_1$ and $d_{s,1} = d_{r,1} := d_1$, then the CSD takes the same form as the SFS from eqn. (8), namely

$$\begin{aligned} E[\Theta_{\kappa}^{(det)}] &= 2\nu N_d \int_0^{1-1/N_d} \left(1 - \frac{d_1}{b_1} y\right)^{-1} (1-y) y^{\kappa-1} dy \\ &\rightarrow 2\nu N_d \sum_{l=0}^{\infty} \frac{\left(\frac{d_1}{b_1}\right)^l}{(\kappa+l)(\kappa+l+1)} \quad \text{for } N_d \rightarrow \infty \\ &\rightarrow 2\nu N_d \frac{b_1}{b_1 - d_1} \frac{1}{\kappa(\kappa+1)} \quad \text{for } \kappa \rightarrow \infty. \end{aligned} \quad (M2)$$

Clone size distribution after treatment. The CSD after treatment for time t_f is also obtained by adapting theorem 1. For newly emerging mutations, the relevant mutation generation comes from sensitive cells. Starting from N'_d cells, the mutation generation is described by $2\nu b_{s,2} N'_d e^{(b_{s,2}-d_{s,2})t}$. We drop the term $\frac{1}{1 - p(N'_d \rightarrow 0, t_f)}$ for conditioning on survival such that

$$E[\Theta_{\kappa}^{(new)}] = \int_0^{t_f} b_{s,2} 2\nu N'_d e^{(b_{s,2}-d_{s,2})t} p(1 \rightarrow \kappa, t_f - t) dt. \quad (M3)$$

The term $p(1 \rightarrow \kappa, t_f - t)$ described the growth of resistant clones and thus must be parameterized with $b_{r,2}$ and $d_{r,2}$. For the CSD of clones preexisting to treatment, we again drop the conditioning on survival and obtain

$$E[\Theta_{\kappa}^{(pre)}] = \sum_{i=1}^{N'_d} E[\Theta_{\kappa'}^{(det)}] p(\kappa' \rightarrow \kappa, t_f), \quad (M4)$$

with $p(\kappa' \rightarrow \kappa, t_f)$ parameterized with $b_{r,2}$ and $d_{r,2}$.

To justify dropping the conditioning on survival, we note that we start from an initially large number of sensitive and resistant cells. First, it is highly unlikely that resistant cells go extinct. Second, the relevant mutational input for the CSD comes from the sensitive cell population. For small times t , the average growth of sensitive cells with conditioning on survival is $\tilde{N}_s(t) = \frac{1}{1-\alpha(t)} N'_d e^{(b_{s,2}-d_{s,2})t}$ that is well approximated by the average growth of sensitive cells without conditioning $N_s(t) = N'_d e^{(b_{s,2}-d_{s,2})t}$. For larger t , when the sensitive cell population approaches the extinction threshold, there will be so few sensitive cells that the generation of new clones will be negligible, $b_{s,2} 2\nu N'_d e^{(b_{s,2}-d_{s,2})t} \rightarrow 0$. Together, this allows us to drop the term $\frac{1}{1 - p(N'_d \rightarrow 0, t_f)}$ in theorem 1 while using $N_s(t)$ rather than $\tilde{N}_s(t)$ for the generation of new mutations.

N Resistant subpopulation versus largest resistant clone

In growing populations, it is apparent that the random timing of mutations plays an important role in determining the mutant population size. Here, we ask whether the largest resistant clone dominates the total resistant subpopulation. First, we compute the arrival time of the first and second mutant and note that the second arriving mutant may outgrow the first due to genetic drift. Secondly, we derive the probability distributions for the size of the largest clone and compare this to the probability distribution for the total number of mutant cells.

We focus on clone sizes at detection since a resistant clone dominating at detection will also dominate after treatment. Therefore, we consider a birth-death process with birth rate b_1 and death rate d_1 growing from a single cell up to size N_d , while each daughter cell acquires resistance mutation at rate $\nu \ll 1$ during a division. The mutation rate per division is then $\tilde{\nu} = 2\nu$.

N.1 Arrival time of resistant clones

In order to condition on survival of genetic drift, we can adapt the mutation rate $\tilde{\nu} \rightarrow \tilde{\nu}\sigma$, where $\sigma = \frac{b_1 - d_1}{b_1}$ is the probability to survive drift in the linear birth-death process [15].

Considering deterministic growth of the entire population with random mutations, the number of acquired resistance mutations follows an in-homogeneous Poisson process such that the resistance mutations acquired in the time interval $[\tau_1, \tau_2]$ is given by

$$P_k = \frac{e^{-\lambda} \lambda^k}{k!} \quad \text{with } \lambda = \int_{\tau_1}^{\tau_2} b_1 \tilde{\nu} e^{(b_1 - d_1)t} dt. \quad (\text{N1})$$

To compute the arrival time of the first mutant, we set $\tau_1 = 0$ and $\tau_2 = t_1$. The probability that no mutations were acquired up to time t_1 is P_0 . We identify $1 - P_0$ as the cumulative density function for the arrival time of the first surviving mutant. Taking the derivative of the cumulative density function, we obtain the probability density function for T_1 that reads

$$f_{T_1}(t_1) = \frac{d(1 - P_0)}{dt_1} = b_1 \tilde{\nu} e^{(b_1 - d_1)t_1 - \frac{b_1 \tilde{\nu}}{b_1 - d_1} e^{(b_1 - d_1)t_1}}. \quad (\text{N2})$$

Similar results were previously reported by making approximations on the fully stochastic model (e.g. chapter 5 in ref. [15]).

Given that the first resistant clone arose at $T_1 = t_1$, the probability that no second mutations occurs until time t_2 is P_0 but with $\tau_1 = t_1$ and $\tau_2 = t_2$. Analogously to T_1 , we compute the probability density function for T_2 ,

$$f_{T_2}(t_2|T_1 = t_1) = \frac{d(1 - P_0)}{dt_2} = b_1 \tilde{\nu} e^{(b_1 - d_1)t_2 - \frac{b_1 \tilde{\nu}}{b_1 - d_1} (e^{(b_1 - d_1)t_2} - e^{(b_1 - d_1)t_1})} \mathbf{1}_{[t_1, \infty]}(t_2). \quad (\text{N3})$$

The unconditional density of T_2 is obtained by marginalizing out T_1 ,

$$\begin{aligned} f_{T_2}(t_2) &= \int_0^\infty f_{T_2}(t_2|T_1 = t_1) f_{T_1}(t_1) dt_1 \\ &= \frac{b_1^2 \tilde{\nu}^2}{b_1 - d_1} \left(e^{(b_1 - d_1)t_2} - 1 \right) e^{(b_1 - d_1)t_2 - \frac{b_1 \tilde{\nu}}{b_1 - d_1} e^{(b_1 - d_1)t_2}} \\ &\approx \frac{b_1^2 \tilde{\nu}^2}{b_1 - d_1} e^{2(b_1 - d_1)t_2 - \frac{b_1 \tilde{\nu}}{b_1 - d_1} e^{(b_1 - d_1)t_2}}. \end{aligned} \quad (\text{N4})$$

Inverse cumulative density function: For fig. 5A, we generate random numbers for T_1 and T_2 . Therefore, we use inverse transformation sampling, which requires the inverse distribution function. For T_1 , we have

$$t_1 = F_{T_1}^{-1}(u) = \frac{1}{b_1 - d_1} \log \left(\frac{(b_1 - d_1) \log \left(\frac{1}{1-u} \right)}{b_1 \tilde{\nu}} + 1 \right). \quad (\text{N5})$$

For T_2 conditioned on $T_1 = t_1$, we have

$$t_2 = F_{T_2|T_1}^{-1}(u) = \frac{1}{b_1 - d_1} \log \left(e^{(b_1 - d_1)t_1} - \frac{(b_1 - d_1) \log(1-u)}{b_1 \tilde{\nu}} \right). \quad (\text{N6})$$

N.2 Size of the resistant subpopulation

In the limit $\tilde{\nu}N \gg 1$, the probability to have R resistant cells is described by a Landau distribution [41, 42] of the form

$$P(R) = \frac{b-d}{b} \frac{1}{\tilde{\nu}N} P_{\text{Landau}} \left(\frac{b-d}{b} \frac{R}{\tilde{\nu}N} - \ln(\tilde{\nu}N) \right). \quad (\text{N7})$$

For $\tilde{\nu}N \sim 1$, the solution is slightly more complicated but also given in ref. [42]. The fixed-time solution was computed in ref. [59] and some asymptotic solutions including selection are derived in ref. [60]. However, we focus on the case $\tilde{\nu}N \gg 1$.

N.3 Size of the largest clone

We denote the number of clones with size κ by Θ_κ such that $\{\Theta_\kappa\}_{\kappa=1}^N$ is the clone size distribution (CSD) similar to the definition of the SFS in the main text. As explained in the main text, the CSD is described by the same expression as the SFS. Since we are interested in the largest clone only that is unlikely to go extinct (should the population continue growing indefinitely), it is sufficient to consider skeleton spectrum (i.e. clones that will never go extinct) in which case the CSD takes the form described in Proposition 1 of ref. [16], namely

$$\begin{aligned} E[\Theta_\kappa^{(skel)}] &\approx \frac{b_1}{b_1 - d_1} \tilde{\nu} N_d \left(1 - \frac{1}{N_d} \right)^\kappa \left(\frac{1}{\kappa(\kappa+1)} + \frac{1}{N} \frac{1}{\kappa} \right) \\ &\sim \frac{b_1}{b_1 - d_1} \tilde{\nu} N_d \frac{1}{\kappa(\kappa+1)} \quad \text{for } N_d \rightarrow \infty \end{aligned} \quad (\text{N8})$$

For simplicity, we take $\kappa \in [1, \infty)$ on a continuous scale and approximate $\frac{1}{\kappa^2 + \kappa} \approx \frac{1}{\kappa^2}$. Then, the skeleton CSD can be written as product of an amplitude A and probability density $f(\kappa)$,

$$E[\Theta_\kappa^{(skel)}] = Af(\kappa) \quad \text{with } A = \frac{b_1}{b_1 - d_1} \tilde{\nu} N_d \text{ and } f(\kappa) = \frac{1}{\kappa^2}. \quad (\text{N9})$$

The amplitude is interpreted as the number of mutations with infinite lineage. If we take this number as fixed, then the largest clone size can be understood as order statistic. In other words, the largest clone size $\tilde{\kappa}$ is the largest number of A clone sizes that are sampled from $f(\kappa)$.

Assuming independence between clone sizes, the cumulative density function for $\tilde{\kappa}$ is simply the A -th power of the cumulative density function for κ [40]. The cumulative density function for κ reads $F(\kappa) = 1 - \frac{1}{\kappa}$ such that

$$F(\tilde{\kappa}) = \left(1 - \frac{1}{\tilde{\kappa}}\right)^A \quad (\text{N10})$$

and the probability density is

$$f(\tilde{\kappa}) = \frac{dF}{d\tilde{\kappa}} = \frac{A}{\tilde{\kappa}^2} \left(1 - \frac{1}{\tilde{\kappa}}\right)^{A-1}. \quad (\text{N11})$$

O Site frequency spectrum upon emergence of resistance

Following the main text, we separate between mutations that occur in sensitive cells and mutations that occur in resistant cells. The sum of both contributions will make the total SFS,

$$S_k = S_k^S + S_k^R \quad \text{with} \quad S_k^R = \sum_{i=1}^K S_k^{R,i}. \quad (\text{O1})$$

O.1 Mutations emerging in resistant cells

We start by analysing S_k^R , which is the sum over all mutations arising in resistant clones $i = 1, \dots, K$. Given expected CSD $E[\Theta_\kappa]$, we have

$$E[S_k^R] = E\left[\sum_{i=1}^K S_k^{R,i}\right] = \sum_{\kappa=1}^{N_d} E[\Theta_\kappa] E[S_k(\kappa)], \quad (\text{O2})$$

where $E[S_k(\kappa)]$ denotes the expected SFS of a resistant clone with size κ excluding mutations that were acquired prior to obtaining the resistance mutation.

Before treatment, clones grow with constant birth and death rate analogously to a homogeneous popu-

lation at detection. Thus, the fixed-time SFS of a single clone with size κ is described by

$$\begin{aligned} E[S_k(\kappa)] &= 2m\kappa \int_0^{1-1/\kappa} \left(1 - \frac{d_{r,1}}{b_{r,1}}y\right)^{-1} (1-y)y^{k-1} dy \\ &\rightarrow 2m\kappa \sum_{l=0}^{\infty} \frac{\left(\frac{d_{r,1}}{b_{r,1}}\right)^l}{(k+l)(k+l+1)} \quad \text{for } \kappa \rightarrow \infty \\ &\rightarrow 2m\kappa \frac{b_{r,1}}{b_{r,1} - d_{r,1}} \frac{1}{k(k+1)} \quad \text{for } k \rightarrow \infty. \end{aligned} \quad (\text{O3})$$

During treatment, the growth rates of the resistant clones will change and $E[S_k(\kappa)]$ is more appropriately described by homogeneous population with continued increase given by eqn. (F7) and (F8). In the case of full resistance such that $b_r := b_{r,1} = b_{r,2}$ and $d_r := d_{r,1} = d_{r,2}$, we can keep the expression in eqn. (O3) for the treatment phase such that

$$E[S_k(\kappa)] \rightarrow 2m\kappa \sum_{l=0}^{\infty} \frac{\left(\frac{d_r}{b_r}\right)^l}{(k+l)(k+l+1)} \quad \text{for } \kappa \rightarrow \infty, \quad (\text{O4})$$

which once inserted into eqn. (O2) gives us the SFS of mutations that emerged in resistant cells. In particular, we then have

$$\begin{aligned} E[S_k^R] &= \sum_{i=1}^K E[S_k^{(R,i)}] \\ &\approx \sum_{i=1}^K 2m\kappa_i \sum_{l=0}^{\infty} \frac{\left(\frac{d_r}{b_r}\right)^l}{(k+l)(k+l+1)} \\ &= 2mN_r \sum_{l=0}^{\infty} \frac{\left(\frac{d_r}{b_r}\right)^l}{(k+l)(k+l+1)}, \end{aligned} \quad (\text{O5})$$

where N_r is the number of resistant cells. If treatment is applied long enough, there will be no more sensitive cells and we have the total population given by the resistant cells $N_f = N_r$ giving us the neutral tail known from homogeneous populations.

O.2 Mutations emerging in sensitive cells

Obtaining the SFS of mutations that emerged in sensitive cells is more complicated. Neglecting the effect of resistance mutations on the growth of sensitive cells, the SFS of the sensitive cell population can be computed analogous to the homogeneous decreasing cell population whose expected SFS is described by eqn. (F1) and (F3). However, there are also mutations that emerged in sensitive cells, and are inherited to resistant cells.

For simplicity, we restrict ourselves on case that all sensitive cells went extinct. Then, mutations that emerged in sensitive cells must be clonal in at least one resistant clone, and we can write

$$S_k^S = \sum_{Y \in 2^X} C_Y 1_{\{\sum_{j \in Y} \kappa_j = k\}}. \quad (\text{O6})$$

Here, the first sum goes over all possible combinations Y in the set of clone indices $X = 1, 2, \dots, K$ that is the power set 2^X , and C_Y is the number of mutations shared between the selected clones.

Using theorem 3, we can connect the number of clonal mutations in a resistant clone i' to its arrival time $t_{\text{arrival}} \in [0, t_d + t_f]$. We have

$$E \left[\sum_{i' \in Y} C_Y \right] \approx \begin{cases} 2mb_{r,2}(t_d + t_f - t_{\text{arrival}}) & \text{if } t_{\text{arrival}} \in [t_d, t_f] \\ 2mb_{r,1}(t_d - t_{\text{arrival}}) + 2mb_{r,2}t_f & \text{if } t_{\text{arrival}} \in [0, t_d] \end{cases} \quad (\text{O7})$$

To split this predictions into individual peaks C_Y in the SFS, one must study the relatedness between clone i' to other clones.

References

- [1] Emily Mitchell, Michael Spencer Chapman, Nicholas Williams, Kevin J Dawson, Nicole Mende, Emily F Calderbank, Hyunchul Jung, Thomas Mitchell, Tim HH Coorens, David H Spencer, et al. Clonal dynamics of haematopoiesis across the human lifespan. *Nature*, 606(7913):343–350, 2022.
- [2] Federico Abascal, Luke MR Harvey, Emily Mitchell, Andrew RJ Lawson, Stefanie V Lensing, Peter Ellis, Andrew JC Russell, Raul E Alcantara, Adrian Baez-Ortega, Yichen Wang, et al. Somatic mutation landscapes at single-molecule resolution. *Nature*, 593(7859):405–410, 2021.
- [3] Mel Greaves and Carlo C Maley. Clonal evolution in cancer. *Nature*, 481(7381):306–313, 2012.
- [4] Samra Turajlic, Andrea Sottoriva, Trevor Graham, and Charles Swanton. Resolving genetic heterogeneity in cancer. *Nature Reviews Genetics*, 20(7):404–416, 2019.
- [5] Marc J Williams, Benjamin Werner, Timon Heide, Christina Curtis, Chris P Barnes, Andrea Sottoriva, and Trevor A Graham. Quantification of subclonal selection in cancer from bulk sequencing data. *Nature genetics*, 50(6):895–903, 2018.
- [6] Giulio Caravagna, Timon Heide, Marc J Williams, Luis Zapata, Daniel Nichol, Ketevan Chkhaidze, William Cross, George D Cresswell, Benjamin Werner, Ahmet Acar, et al. Subclonal reconstruction of tumors by using machine learning and population genetics. *Nature genetics*, 52(9):898–907, 2020.
- [7] Benjamin D Simons. Deep sequencing as a probe of normal stem cell fate and preneoplasia in human epidermis. *Proceedings of the National Academy of Sciences*, 113(1):128–133, 2016.
- [8] Caroline J Watson, AL Papula, Gladys YP Poon, Wing H Wong, Andrew L Young, Todd E Druley, Daniel S Fisher, and Jamie R Blundell. The evolutionary dynamics and fitness landscape of clonal hematopoiesis. *Science*, 367(6485):1449–1454, 2020.
- [9] Marius E Moeller, Nathaniel V Mon Père, Benjamin Werner, and Weini Huang. Measures of genetic diversification in somatic tissues at bulk and single-cell resolution. *eLife*, 12:RP89780, 2024.
- [10] Javier Fernandez-Mateos, George D Cresswell, Nicholas Trahearn, Katharine Webb, Chirine Sakr, Andrea Lampis, Christine Stuttle, Catherine M Corbishley, Vasilis Stavrinides, Luis Zapata, et al. Tumor evolution metrics predict recurrence beyond 10 years in locally advanced prostate cancer. *Nature Cancer*, pages 1–18, 2024.
- [11] Noemi Andor, Trevor A Graham, Marnix Jansen, Li C Xia, C Athena Aktipis, Claudia Petritsch, Hanlee P Ji, and Carlo C Maley. Pan-cancer analysis of the extent and consequences of intratumor heterogeneity. *Nature medicine*, 22(1):105–113, 2016.
- [12] Simon Tavaré. The linear birth–death process: an inferential retrospective. *Advances in Applied Probability*, 50(A):253–269, 2018.
- [13] Anthony C Davison, Sophie Hautphenne, and Andrea Kraus. Parameter estimation for discretely observed linear birth-and-death processes. *Biometrics*, 77(1):186–196, 2021.
- [14] David G Kendall. On the generalized” birth-and-death” process. *The annals of mathematical statistics*, 19(1):1–15, 1948.
- [15] Richard Durrett. *Branching process models of cancer*. Springer, 2015.
- [16] Einar Bjarki Gunnarsson, Kevin Leder, and Jasmine Foo. Exact site frequency spectra of neutrally evolving tumors: A transition between power laws reveals a signature of cell viability. *Theoretical Population Biology*, 142:67–90, 2021.

- [17] Hisashi Ohtsuki and Hideki Innan. Forward and backward evolutionary processes and allele frequency spectrum in a cancer cell population. *Theoretical Population Biology*, 117:43–50, 2017.
- [18] Marc J Williams, Benjamin Werner, Chris P Barnes, Trevor A Graham, and Andrea Sottoriva. Identification of neutral tumor evolution across cancer types. *Nature genetics*, 48(3):238–244, 2016.
- [19] Robert C Griffiths and Simon Tavaré. The age of a mutation in a general coalescent tree. *Stochastic Models*, 14(1-2):273–295, 1998.
- [20] Richard Durrett and Richard Durrett. *Probability models for DNA sequence evolution*, volume 2. Springer, 2008.
- [21] Motoo Kimura et al. Evolutionary rate at the molecular level. *Nature*, 217(5129):624–626, 1968.
- [22] Motoo Kimura. The neutral theory of molecular evolution: a review of recent evidence. *The Japanese Journal of Genetics*, 66(4):367–386, 1991.
- [23] Naiyer A Rizvi, Matthew D Hellmann, Alexandra Snyder, Pia Kvistborg, Vladimir Makarov, Jonathan J Havel, William Lee, Jianda Yuan, Phillip Wong, Teresa S Ho, et al. Mutational landscape determines sensitivity to pd-1 blockade in non–small cell lung cancer. *Science*, 348(6230):124–128, 2015.
- [24] Nicholas McGranahan, Andrew JS Furness, Rachel Rosenthal, Sofie Ramskov, Rikke Lyngaa, Sunil Kumar Saini, Mariam Jamal-Hanjani, Gareth A Wilson, Nicolai J Birkbak, Crispin T Hiley, et al. Clonal neoantigens elicit t cell immunoreactivity and sensitivity to immune checkpoint blockade. *Science*, 351(6280):1463–1469, 2016.
- [25] Eszter Lakatos, Marc J Williams, Ryan O Schenck, William CH Cross, Jacob Househam, Luis Zapata, Benjamin Werner, Chandler Gatenbee, Mark Robertson-Tessi, Chris P Barnes, et al. Evolutionary dynamics of neoantigens in growing tumors. *Nature genetics*, 52(10):1057–1066, 2020.
- [26] Jingsong Zhang, Jessica J Cunningham, Joel S Brown, and Robert A Gatenby. Integrating evolutionary dynamics into treatment of metastatic castrate-resistant prostate cancer. *Nature communications*, 8(1):1816, 2017.
- [27] Adaptive chemotherapy for ovarian cancer in patients with replaced platinum-sensitive high grade serous or high grade endometrioid ovarian cancer (actov). <https://clinicaltrials.gov/ct2/show/NCT05080556>. Accessed: 2024-08-08.
- [28] Xuan Wang, Haiyun Zhang, and Xiaozhuo Chen. Drug resistance and combating drug resistance in cancer. *Cancer drug resistance*, 2(2):141, 2019.
- [29] Jessica MA Blair, Mark A Webber, Alison J Baylay, David O Ogbolu, and Laura JV Piddock. Molecular mechanisms of antibiotic resistance. *Nature reviews microbiology*, 13(1):42–51, 2015.
- [30] Kristen K Irwin, Nicholas Renzette, Timothy F Kowalik, and Jeffrey D Jensen. Antiviral drug resistance as an adaptive process. *Virus evolution*, 2(1):vew014, 2016.
- [31] Christo Morison, Dudley Stark, and Weini Huang. Single-cell mutational burden distributions in birth-death processes. *arXiv preprint arXiv:2309.06355*, 2023.
- [32] BP Kharlamov. On the generation numbers of particles in a branching process with overlapping generations. *Theory of Probability & Its Applications*, 14(1):44–50, 1969.
- [33] ML Samuels. Distribution of the branching-process population among generations. *Journal of Applied Probability*, 8(4):655–667, 1971.

- [34] Wolfgang J Bühler. Generations and degree of relationship in supercritical markov branching processes. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 18(2):141–152, 1971.
- [35] W Bühler. The distribution of generations and other aspects of the family structure of branching processes. In *Proc. 6th Berkeley Symp. Math. Statist. Prob*, volume 3, pages 463–480, 1972.
- [36] David Cheek and Samuel GG Johnston. Ancestral reproductive bias in branching processes. *Journal of Mathematical Biology*, 86(5):70, 2023.
- [37] Mike Steel and Andy McKenzie. Properties of phylogenetic trees generated by yule-type speciation models. *Mathematical biosciences*, 170(1):91–112, 2001.
- [38] Luis A Diaz Jr, Richard T Williams, Jian Wu, Isaac Kinde, J Randolph Hecht, Jordan Berlin, Benjamin Allen, Ivana Bozic, Johannes G Reiter, Martin A Nowak, et al. The molecular evolution of acquired resistance to targeted egfr blockade in colorectal cancers. *Nature*, 486(7404):537–540, 2012.
- [39] Khurum H Khan, David Cunningham, Benjamin Werner, Georgios Vlachogiannis, Inmaculada Spiteri, Timon Heide, Javier Fernandez Mateos, Alexandra Vatsiou, Andrea Lampis, Mahnaz Darvish Damavandi, et al. Longitudinal liquid biopsy and mathematical modeling of clonal evolution forecast time to treatment failure in the prospect-c phase ii colorectal cancer clinical trial. *Cancer discovery*, 8(10):1270–1285, 2018.
- [40] Herbert A David and Haikady N Nagaraja. *Order statistics*. John Wiley & Sons, 2004.
- [41] David A Kessler and Herbert Levine. Large population solution of the stochastic luria–delbrück evolution model. *Proceedings of the National Academy of Sciences*, 110(29):11682–11687, 2013.
- [42] David A Kessler and Herbert Levine. Scaling solution in the large population limit of the general asymmetric stochastic luria–delbrück evolution process. *Journal of statistical physics*, 158(4):783–805, 2015.
- [43] Ryan O Schenck, Eunjung Kim, Rafael R Bravo, Jeffrey West, Simon Leedham, Darryl Shibata, and Alexander RA Anderson. Homeostasis limits keratinocyte evolution. *Proceedings of the National Academy of Sciences*, 119(35):e2006487119, 2022.
- [44] Robert Noble, Dominik Burri, Cécile Le Sueur, Jeanne Lemant, Yannick Viossat, Jakob Nikolas Kather, and Niko Beerenwinkel. Spatial structure governs the mode of tumour evolution. *Nature ecology & evolution*, 6(2):207–217, 2022.
- [45] Maise Al Bakir, Ariana Huebner, Carlos Martínez-Ruiz, Kristiana Grigoriadis, Thomas BK Watkins, Oriol Pich, David A Moore, Selvaraju Veeriah, Sophia Ward, Joanne Laycock, et al. The evolution of non-small cell lung cancer metastases in tracerx. *Nature*, 616(7957):534–542, 2023.
- [46] Anna S Nam, Ronan Chaligne, and Dan A Landau. Integrating genetic and non-genetic determinants of cancer evolution by single-cell multi-omics. *Nature Reviews Genetics*, 22(1):3–18, 2021.
- [47] Daniel T Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *Journal of computational physics*, 22(4):403–434, 1976.
- [48] Daniel T Gillespie. Exact stochastic simulation of coupled chemical reactions. *The journal of physical chemistry*, 81(25):2340–2361, 1977.
- [49] Lei Yu, Jessie Renton, Agata Burian, Marina Khachatryan, Till Bayer, Jonne Kotta, John J Stachowicz, Katherine DuBois, Iliana B Baums, Benjamin Werner, et al. A somatic genetic clock for clonal species. *Nature Ecology & Evolution*, pages 1–10, 2024.

- [50] Steven G. Johnson. The HCubature.jl package for multi-dimensional adaptive integration in Julia. <https://github.com/JuliaMath/HCubature.jl>, 2017.
- [51] A. C. Genz and A. A. Malik. Remarks on algorithm 006: An adaptive algorithm for numerical integration over an n -dimensional rectangular region. *Journal of Computational and Applied Mathematics*, 6:295–302, 1980.
- [52] Wolfram Research, Inc. Mathematica, Version 14.1. Champaign, IL, 2024.
- [53] Norman TJ Bailey. *The elements of stochastic processes with applications to the natural sciences*, volume 25. John Wiley & Sons, 1991.
- [54] Luc Devroye. Nonuniform random variate generation. *Handbooks in operations research and management science*, 13:83–121, 2006.
- [55] Aaron Clauset, Cosma Rohilla Shalizi, and Mark EJ Newman. Power-law distributions in empirical data. *SIAM review*, 51(4):661–703, 2009.
- [56] Henry E Daniels. Saddlepoint approximations in statistics. *The Annals of Mathematical Statistics*, pages 631–650, 1954.
- [57] Ronald W Butler. *Saddlepoint approximations with applications*, volume 22. Cambridge University Press, 2007.
- [58] Prem C Consul and Gaurav C Jain. A generalization of the poisson distribution. *Technometrics*, 15(4):791–799, 1973.
- [59] Tibor Antal and PL Krapivsky. Exact solution of a two-type branching process: models of tumor progression. *Journal of Statistical Mechanics: Theory and Experiment*, 2011(08):P08018, 2011.
- [60] Michael D Nicholson, David Cheek, and Tibor Antal. Sequential mutations in exponentially growing populations. *PLOS Computational Biology*, 19(7):e1011289, 2023.

Supplementary figures

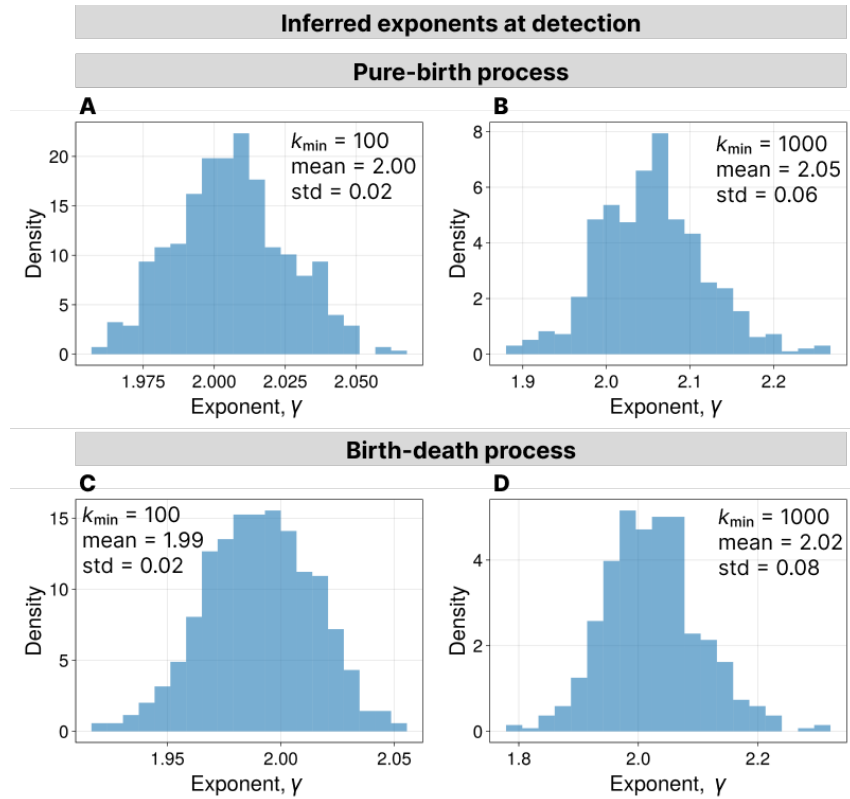


Figure S1: Fitted exponents to the SFS of simulated populations at detection using eqn. (37). **A.** Parameters: $b_1 = 1.0$, $d_1 = 1.0$, $m = 2.0$, $N_d = 10^5$ and $k_{\min} = 100$. **B.** Parameters: $b_1 = 1.0$, $d_1 = 1.0$, $m = 2.0$, $N_d = 10^5$ and $k_{\min} = 1000$. **C.** Parameters: $b_1 = 1.0$, $d_1 = 0.7$, $m = 2.0$, $N_d = 10^5$ and $k_{\min} = 100$. **D.** Parameters: $b_1 = 1.0$, $d_1 = 0.7$, $m = 2.0$, $N_d = 10^5$ and $k_{\min} = 1000$. Simulations were repeated 500 times.

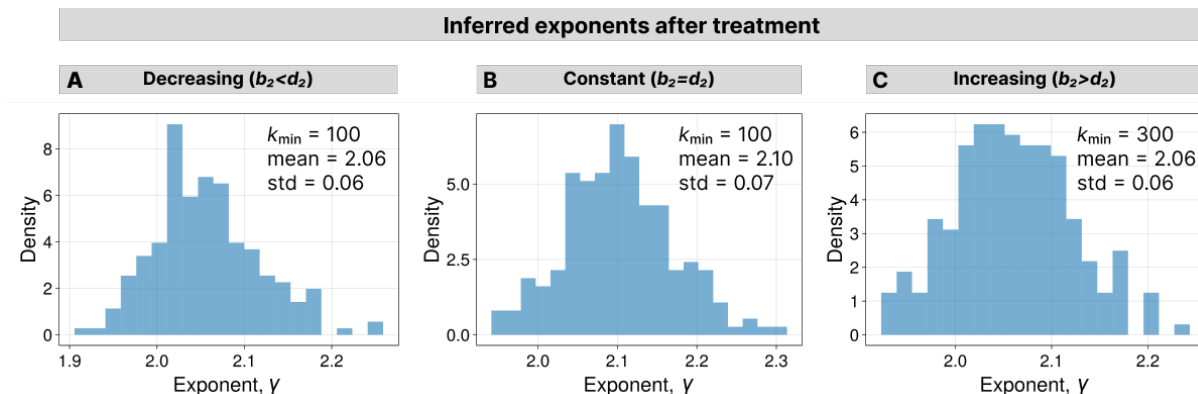


Figure S2: Fitted exponents to the SFS in of simulated homogeneous populations after treatment using eqn. (37). **A.** For decreasing population. Parameters before treatment: $b_1 = 1.0$, $d_1 = 0.0$, $m = 2.0$ and $N_d = 10^5$. Parameters during treatment: $b_2 = 1.0$, $d_2 = 3.0$, $m = 2.0$ and $N_f = 10^4$. Parameter for inference: $k_{\min} = 100$. **B.** For constant population. Parameters before treatment: $b_1 = 1.0$, $d_1 = 0.0$, $m = 2.0$ and $N_d = 10^5$. Parameters during treatment: $b_2 = d_2 = 1.0$, $m = 2.0$ and $t_f = 8.0$. Parameter for inference: $k_{\min} = 100$. **C.** For increasing population. Parameters before treatment: $b_1 = 1.0$, $d_1 = 0.0$, $m = 2.0$ and $N_d = 10^4$. Parameters during treatment: $b_2 = 1.0$, $d_2 = 0.6$, $m = 2.0$ and $N_f = 3.0 \times 10^4$. Simulations were repeated 200 times for each treatment scenario. Parameter for inference: $k_{\min} = 300$. Simulations were repeated 200 times for each treatment scenario.

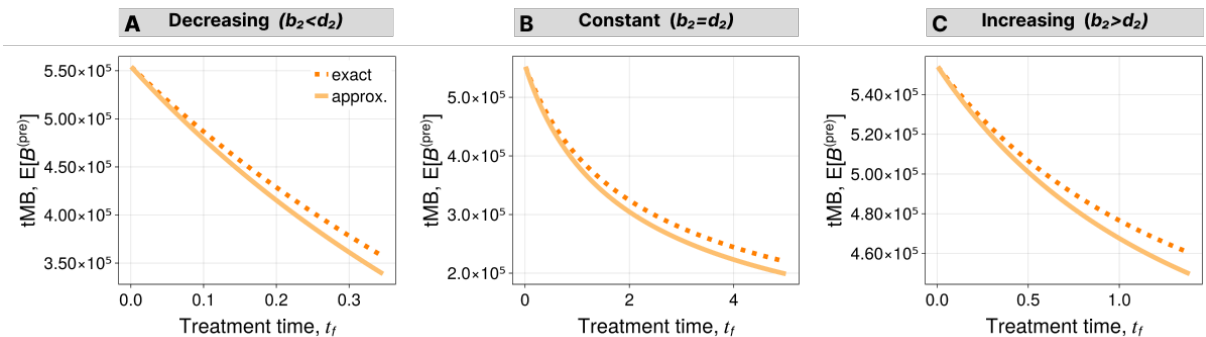


Figure S3: Comparison of approximate and exact expressions for the expected tMB of preexisting mutations after homogeneous treatment response. Exact tMB is given by eqn. (J2) and approximate tMB is given by eqn. (20). Parameters before treatment: $b_1 = 1.0$, $d_1 = 0.5$, $m = 2.0$ and $N_d = 10^5$. Parameters during treatment for decreasing population: $b_2 = 1.0$, $d_2 = 3.0$, $m = 2.0$ and t_f varied. Parameters during treatment for constant population: $b_2 = d_2 = 1.0$, $m = 2.0$ and t_f varied. Parameters during treatment for increasing population: $b_2 = 1.0$, $d_2 = 0.5$, $m = 2.0$ and t_f varied.

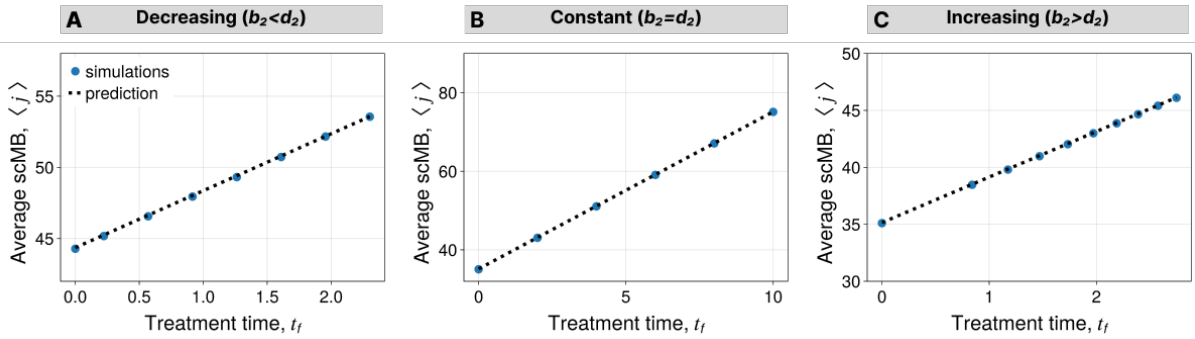


Figure S4: Average scMB after treatment for homogeneous population. **A.** For decreasing population. Parameters before treatment: $b_1 = 1.0$, $d_1 = 0.0$, $m = 2.0$ and $N_d = 10^5$. Parameters during treatment: $b_2 = 1.0$, $d_2 = 3.0$, $m = 2.0$ and N_f is varied. **B.** For constant population. Parameters before treatment: $b_1 = 1.0$, $d_1 = 0.0$, $m = 2.0$ and $N_d = 10^5$. Parameters during treatment: $b_2 = d_2 = 1.0$, $m = 2.0$ and t_f is varied. **C.** For increasing population. Parameters before treatment: $b_1 = 1.0$, $d_1 = 0.0$, $m = 2.0$ and $N_d = 10^4$. Parameters during treatment: $b_2 = 1.0$, $d_2 = 0.6$, $m = 2.0$ and N_f is varied. Simulations were repeated 200 times.

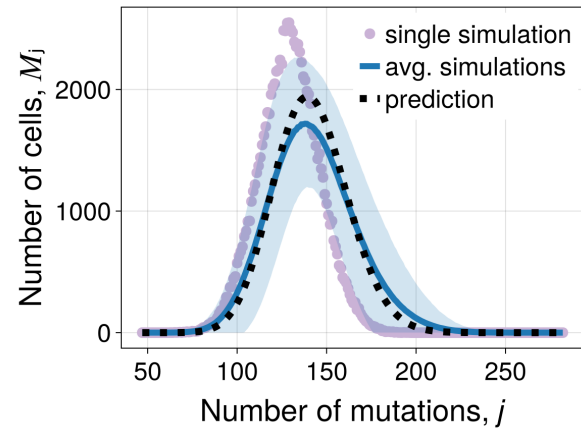


Figure S5: Single cell mutational burden distribution at detection for birth-death process. Parameters: $b_1 = 1.0$, $d_1 = 0.7$, $m = 2.0$ and $N_d = 10^5$. Simulations were repeated 500 times.

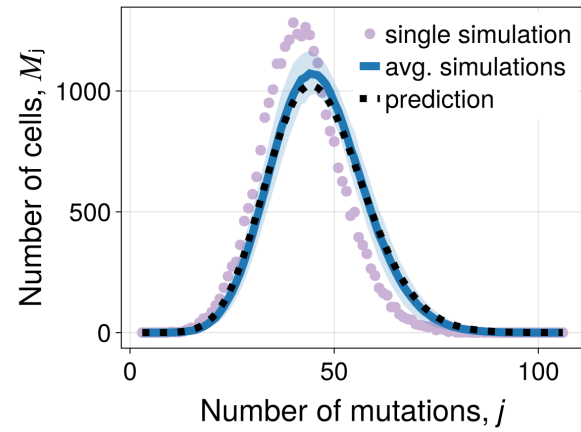


Figure S6: Single cell mutational burden distribution for continued increasing population. Parameters before treatment: $b_1 = 1.0$, $d_1 = 0.0$, $m = 2.0$ and $N_d = 10^4$. Parameters during treatment: $b_2 = 1.0$, $d_2 = 0.6$, $m = 2.0$ and $N_f = 3.0 \times 10^4$. Simulations were repeated 200 times.

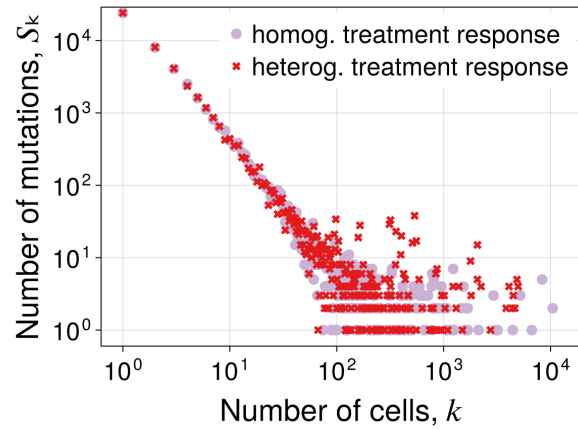


Figure S7: SFS of homogeneous vs heterogeneous population. For homogeneous population, we use parameters $b_1 = 1.0$, $d_1 = 0.0$, $m = 2.0$, $N_d = 1.2 \times 10^4$. For the heterogeneous population, we use parameters before treatment $b_{s_1} = b_{r_1} = 1.0$, $d_{s_1} = d_{r_1} = 0.0$, $m = 2.0$, $\nu = 10^{-3}$, $N_d = 10^4$ and during treatment $b_{s_2} = b_{r_2} = 1.0$, $d_{s_2} = 3.0$, $d_{r_2} = 0.0$, $m = 2.0$, $\nu = 10^{-3}$, $N_f = 1.2 \times 10^4$. The shown SFS is the result of a single realization at the end of each scenario.

Supplementary tables

Figure	b_1	d_1	N_d	b_2	d_2	N_f	t_f	m	realizations
2A	1.0	0.7	10^5	—	—	—	—	2.0	500
2B	1.0	0.0	10^5	—	—	—	—	2.0	500
2C	1.0	0.0	10^5	—	—	—	—	2.0	500
2D	1.0	0.0	10^5	1.0	3.0	10^4	—	2.0	200
2E	1.0	0.0	10^5	1.0	1.0	—	8.0	2.0	200
2F	1.0	0.0	10^4	1.0	0.6	3×10^4	—	2.0	200

Table S1: Parameters for fig. 2.

Figure	b_1	d_1	N_d	t_d	b_2	d_2	N_f	t_f	m	realizations
3A	1.0	0.0	10^5	—	—	—	—	—	2.0	500
3B	1.0	0.7	10^5	—	—	—	—	—	2.0	500
3C	1.0	0.0	—	9.21	—	—	—	—	2.0	500
3D	1.0	0.0	10^5	—	1.0	3.0	varied	—	2.0	200
3E	1.0	0.0	10^5	—	1.0	1.0	—	varied	2.0	200
3F	1.0	0.0	10^4	—	1.0	0.6	varied	—	2.0	200

Table S2: Parameters for fig. 3.

Figure	b_1	d_1	N_d	b_2	d_2	N_f	t_f	m	realizations
4A	1.0	0.0	10^5	—	—	—	—	2.0	500
4B	1.0	0.0	10^5	1.0	3.0	10^4	—	2.0	200
4C	1.0	0.0	10^5	1.0	1.0	—	8.0	2.0	200

Table S3: Parameters for fig. 4.

Figure	b_1	d_1	N_d	t_d	$b_{s,2}$	$d_{s,2}$	$b_{r,2}$	$d_{r,2}$	N_f	m	ν	realizations
5A	0.14	0.13	—	2555	—	—	—	—	—	—	10^{-7}	10^6
5B	1.0	0.5	10^5	—	—	—	—	—	—	—	10^{-3}	500
5C	0.14	0.13	varied	—	—	—	—	—	—	—	10^{-7}	500
5D	1.0	0.0	10^4	—	1.0	3.0	1.0	0.0	2.0×10^4	2.0	10^{-3}	200
5E	1.0	0.0	10^4	—	1.0	3.0	1.0	0.0	2.0×10^4	2.0	10^{-3}	200
5F	1.0	0.0	10^4	—	1.0	1.5	0.7	0.0	1.2×10^4	2.0	10^{-3}	1

Table S4: Parameters for fig. 5. We set $b_{s,1} = b_{r,1} = b_1$ and $d_{s,1} = d_{r,1} = d_1$.