Selective sweep probabilities in spatially expanding populations

Alexander Stein
Centre for Cancer Genomics and Computational Biology, Barts Cancer Institute, Queen Mary University of London, London, UK and Department of Physics, ETH Zurich, Zürich, Switzerland

Ramanarayanan Kizhuttil
Department of Physics, Indian Institute of Science Education and Research, Kolkata, India

Maciej Bak and Robert Noble*
Department of Mathematics, City, University of London, London, UK
(Dated: November 27, 2023)

Evolution during range expansions is an important feature of many biological systems including tumours, microbial communities, and invasive species. A selective sweep is a fundamental process, in which an advantageous mutation evades clonal interference and spreads through the population to fixation. However, most theoretical investigations of selective sweeps have assumed constant population size or have ignored spatial structure. Here we use mathematical modelling and analysis to investigate selective sweep probabilities in populations that grow with constant radial expansion speed. We derive probability distributions for the arrival time and location of the first surviving mutant and hence find simple approximate and exact expressions for selective sweep probabilities in one, two and three dimensions, which are independent of mutation rate. Namely, the selective sweep probability is approximately $(1 - c_{\mathrm{wt}}/c_{\mathrm{m}})^d$, where c_{wt} and c_m are the wildtype and mutant radial expansion speeds, and d is the spatial dimension. Using agent-based simulations, we show that our analytical results accurately predict selective sweep frequencies in the two-dimensional spatial Moran process. We further compare our results with those obtained for alternative growth laws. Parameterizing our model for human tumours, we find that selective sweeps are predicted to be rare except during very early solid tumour growth, thus providing a general, pan-cancer explanation for findings from recent sequencing studies.

I. INTRODUCTION

Range expansions abound across biological scales. Evolution during cell population range expansions determines the development and spatial heterogeneity of biofilms [1], tumours [2], mosaicism [3], and normal tissue [4]. At the species level, range expansions influenced human evolution [5] and are of growing importance in the current era of climate change as organisms are forced into new habitats [6, 7]. Range expansions alter the course of evolution in distinct and often profound ways [8] that are not fully understood. Recent theoretical investigations (for example, [9–12]) have typically assumed that proliferation is restricted to the population boundary, precluding full selective sweeps. Alternative modes of dispersal remain largely unexplored.

Somatic evolution provides especially strong motivation for investigating selective sweeps during range expansions. Cancer results from an accumulation of mutations that drive cells to proliferate uncontrollably and to invade surrounding tissue [13]. Some of these are clonal mutations that occur during tumourigenesis or early tumour development [14, 15]. However, recent advances in multi-region and single-cell sequencing have revealed pervasive genetic intratumour heterogeneity [16] and diverse modes of evolution [17, 18]. Although fitter mutant clones arise throughout cancer progression they seldom if ever achieve selective sweeps, except perhaps in very small tumours [15, 17].

With notable exceptions [19–21], mathematical studies of cancer initiation and early evolution have either ignored spatial structure [22–24] or have relied on intractable agent-based models [25–28]. The typical observed pattern of very early fixation events followed by branching or effectively neutral intratumour evolution – which has important clinical implications [29] – has eluded a general explanation.

Here we use mathematical analysis to explain why beneficial mutations typically fixate only – if at all – in the very early stages of range expansions. We focus on a simple model with short-range dispersal throughout the population. This model is relatively permissive to selective sweeps and therefore yields useful upper bounds on their prevalence. By solving our model we derive exact and simple approximate expressions for sweep probabilities in one, two and three dimensions. We confirm the accuracy and robustness of our analytical results using extensive agent-based simulations of a spatial Moran process, and we compare outcomes for alternative growth laws. We discuss how these findings shed light on the nature of evolution in range expansions in general, and cancer development in particular.

* robert.noble@city.ac.uk

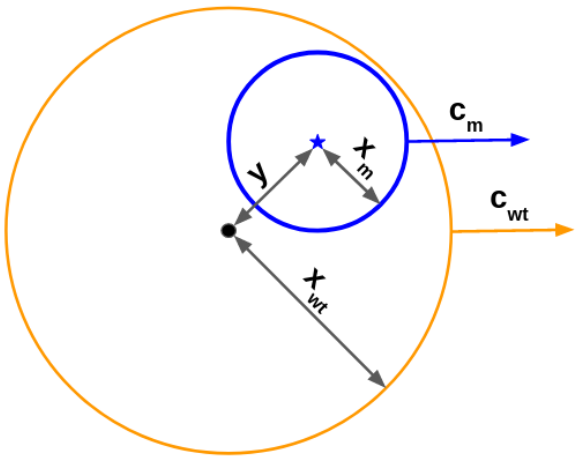

Figure 1. Illustration of the macroscopic model configuration. A mutant population (blue) expands within the wildtype (orange), represented by balls with radii x_{m} and x_{wt} and radial expansion speeds c_{wt} and c_{m}, respectively. The mutant arose at distance y from the centre of the wildtype.

II. MACROSCOPIC MODEL

We consider a wildtype population that begins at a point at time $t = 0$ and expands spherically, such that its radius x_{wt} grows at constant speed c_{wt}. Since we are interested in selective sweeps, we consider only advantageous mutations that spread within the wildtype at constant speed $c_{\mathrm{m}} > c_{\mathrm{wt}}$. These mutations occur at per-capita rate $\tilde{\mu}$ (independent of location) and each mutant survives with probability ρ. In our results, as the mutation rate is always accompanied by the survival probability, it is enough to consider the compound parameter $\mu = \rho\tilde{\mu}$. For brevity, we will refer to this μ as the mutation rate and imply that we consider only successful mutations. Model variables are illustrated in Figure 1.

Various models link propagation speeds to fitness values and migration rates. The most prominent formula is the solution of a reaction-diffusion equation associated with Fisher [30] and Kolmogorov [31]: $c = 2\sqrt{D\Delta r}$, where D is the diffusion coefficient and Δr is the difference in the local growth rate between wildtype and mutant. More recent studies have sought to refine and generalise this result [32–35]. Our methods are compatible with any model that generates approximately constant speeds.

The first surviving mutation achieves a complete (or "classic") selective sweep only if it reaches every part of the wildtype expansion front before a second mutant of equal or greater fitness arises within the wildtype (Figure 2). Otherwise the outcome is clonal interference (or possibly a soft selective sweep if the two mutations confer the same trait). For simplicity, we neglect mutants with fitness values between those of the wildtype and the first mutant, which would slow the expansion of the first mutant and so reduce rather than negate the selective sweep probability. Neither do we investigate the case of a yet-fitter mutant achieving a selective sweep after arising from the first.

The unconditional sweep probability is derived in four steps. First, we introduce random variable X, the radius of the wildtype population when the first mutant arises, and we compute its probability density $f_X(x)$. Second, we introduce the random variable Y to denote the distance between the wildtype and mutant origins, and we calculate its probability density function conditioned on X, namely $f_Y(y|X = x)$. Third, we derive an expression for the conditional sweep probability $\Pr(\text{sweep}|X = x, Y = y)$. Finally, we marginalize out X and Y to obtain the unconditional sweep probability $\Pr(\text{sweep})$. For brevity, we focus on the three-dimensional case; analogous results in one and two dimensions are derived in Appendices C and D.

III. SWEEP PROBABILITY

A. Probability that no mutations occur

We start by deriving the probability that no mutations occur before a given time. This result will be useful in various parts of subsequent derivations.

Claim 1. *Let $N(t)$ denote the population size as a function of time and μ denote the per capita mutation rate, conditional on mutant survival. The probability that no successful mutants arise during the time interval $[0, t]$ is then given by*

$$F(t) = e^{-\mu \int_0^t N(t')\,dt'} \tag{1}$$

Proof. The number of successful mutations in the infinitesimal small time interval $[t, t + dt]$ is given by $\mu N(t)dt$. In other words, $\mu N(t)$ is the rate at which successful mutations arise. The probability $F(t)$ that the population acquires zero mutants at time t decreases at this rate, such that

$$\frac{dF(t)}{dt} = -\mu N(t)F(t).$$

This is a simple first order differential equation with solution

$$F(t) = Ce^{-\mu \int_0^t N(t')\,dt'}.$$

We require $C = 1$ so that the probability is $F = 1$ for vanishing population size $N(t) = 0$. □

B. Arrival time of the first mutant

In the absence of mutants, the wildtype population in three dimensions grows as $N_{\mathrm{wt}} = \frac{4}{3}\pi x_{\mathrm{wt}}^3 = \frac{4}{3}\pi(c_{\mathrm{wt}}t)^3$.

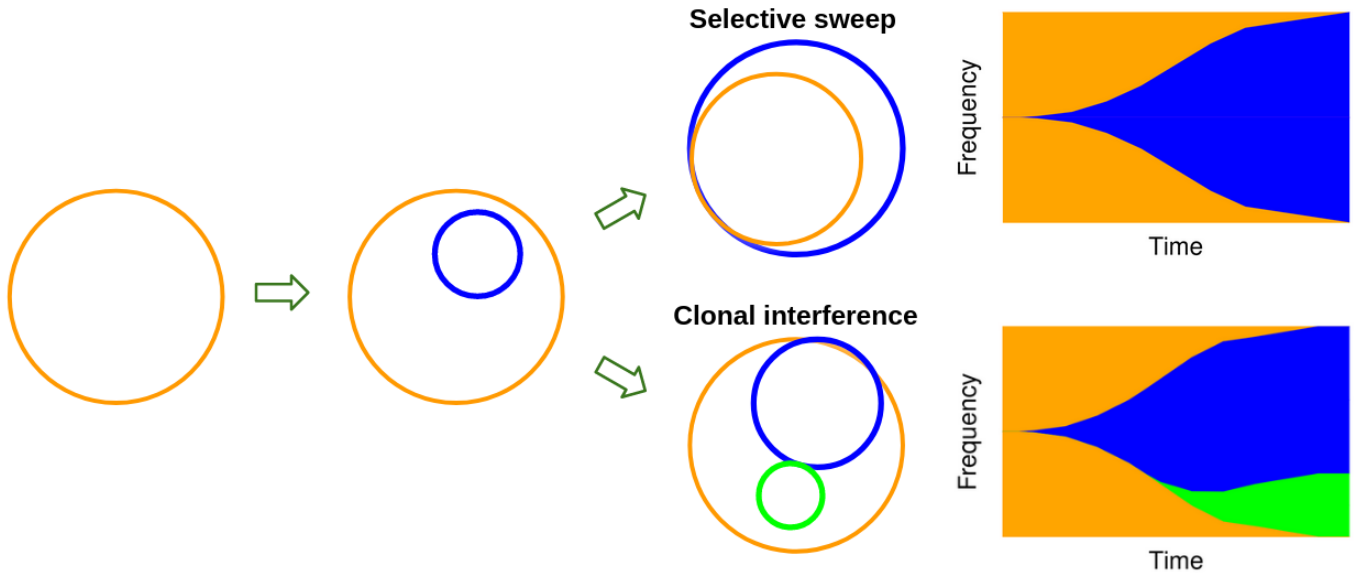

Figure 2. The two possible fates of the first mutant expansion (blue) within the wildtype population (orange). In the first case, the mutant reaches every part of the wildtype boundary before further mutations arise in the wildtype, and so achieves a selective sweep. Otherwise, a second (green) mutant arises in the wildtype population and generates clonal interference.

Applying Claim 1, we obtain the probability that no mutations arise in the time interval $[0, t]$:

$$F(t) = e^{-\mu \int_0^t \frac{4}{3}\pi (c_{\mathrm{wt}} t')^3 \, dt'} = e^{-(t/\kappa)^4} \tag{2}$$

in terms of characteristic length of time $\kappa = \sqrt[4]{\frac{3}{\mu \pi c_{\mathrm{wt}}^3}}$.

We identify $1 - F(t) = 1 - \Pr(T \geq t) = \Pr(T \leq t)$ as the cumulative distribution function for the arrival time T of the first surviving mutant. The probability density of T is then

$$f_T(t) = \frac{d(1 - F(t))}{dt} = \frac{4t^3}{\kappa^4} e^{-\left(\frac{t}{\kappa}\right)^4}. \tag{3}$$

With a change of variable, $t = \frac{x}{c_{\mathrm{wt}}}$ and $dt = \frac{1}{c_{\mathrm{wt}}} dx$, we obtain the probability density of the radius of the wildtype ball X at the time the first surviving mutant arises, which reads

$$f_X(x) = \frac{4x^3 e^{-(x/\theta)^4}}{\theta^4} \tag{4}$$

and is expressed in terms of the characteristic length $\theta = \sqrt[4]{\frac{3c_{wt}}{\pi \mu}}$. Hence f_X follows a Weibull distribution with shape parameter 4 and scale parameter θ. The expectation of the radius is

$$\mathbb{E}[X] = \Gamma\left(\frac{5}{4}\right)\theta \approx 0.91\,\theta,$$

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

$$\mathrm{Var}[X] = \left(\frac{\sqrt{\pi}}{2} - \Gamma\left(\frac{5}{4}\right)^2\right)\theta^2 \approx 0.065\,\theta^2.$$

Hence the standard deviation of the wildtype radius X is approximately 28% of its mean. Analogous calculations in one and two dimensions also yield Weibull distributions with shape parameter $d + 1$, where d is the spatial dimension (see Appendices C and D). Results for all three cases are illustrated in Figure 3A-C.

Next, we calculate the probability distribution for the distance Y between the wildtype's origin and the location of the first surviving mutant, conditional on $X = x$. The probability that a mutant arises at location y is proportional to the size $D(y)$ and the age $a(y)$ of the cell population at that location. We have

$$f_Y(y|X = x)\, dy \propto D(y)a(y), \tag{5}$$

where $D(y) = 4\pi y^2 \, dy$ is the number of cells in the infinitesimally thin shell at radius y and $a(y) = \frac{x-y}{c_{\mathrm{wt}}}$ is the age of the population at distance y at the time the first surviving mutant arises. After adding an indicator function $1\{y \leq x\}$ – to condition on the mutant arising inside the wildtype ball – and then normalizing we obtain the final expression

$$f_Y(y|X = x) = \frac{12y^2(x - y)}{x^4} 1\{y \leq x\}. \tag{6}$$

We calculate the unconditional probability density of Y by marginalizing out X,

$$f_Y(y) = \int_0^\infty f_Y(y|x) f_X(x) \, dx, \tag{7}$$

giving us

$$f_Y(y) = \frac{12y^2}{\theta^4}\left(\theta\Gamma\left(\frac{1}{4}, \frac{y^4}{\theta^4}\right) - y\Gamma\left(0, \frac{y^4}{\theta^4}\right)\right). \tag{8}$$

Here, $\Gamma(a,z) = \int_z^\infty t^{a-1}e^{-t}\,dt$ is the incomplete gamma function. Again, we calculate the expectation

$$\mathbb{E}[Y] = \left(3\Gamma\left(\frac{5}{4}\right) - \frac{3}{5}\Gamma\left(\frac{1}{4}\right)\right)\theta \approx 0.54\,\theta$$

and the variance

$$\mathrm{Var}[Y] = \left(\frac{\sqrt{\pi}}{5} - \left(3\Gamma\left(\frac{5}{4}\right) - \frac{3}{5}\Gamma\left(\frac{1}{4}\right)\right)^2\right)\theta^2 \approx 0.059\,\theta^2.$$

Compared to f_X, the distribution f_Y is shifted towards zero and is more positively skewed. Again, similar results are obtained in one and two dimensions (see Appendices C and D), which are shown together in Figure 3A-C.

C. Conditional sweep probability

Here, we derive an expression for the conditional sweep probability $\Pr(\text{sweep}|X = x, Y = y)$. As illustrated in Figure 2, we assume that the probability of a selective sweep equals the probability that no additional mutants arise in the remaining wildtype population after the first successful mutant starts expanding. We can therefore apply Claim 1 to the remaining wildtype population.

We introduce a new variable τ that measures time in the same unit as t but has its origin $\tau = 0$ when the first mutant arises. Recall that $t = 0$ relates to the origin of the wildtype. Thus, we have $t = \tau + x/c_{\text{wt}}$. Let $\Delta(\tau)$ denote the remaining wildtype population. Since both populations expand spherically, we have

$$\Delta(\tau) = N_{\text{wt}} - N_{\text{int}}, \tag{9}$$

where $N_{\text{wt}} = \frac{4}{3}\pi x_{\text{wt}}^3$ is the volume of the wildtype ball, and N_{int} is the volume of the intersection between the wildtype and mutant balls. While the mutant is completely surrounded by the wildtype, we have simply $N_{\text{int}} = \frac{4}{3}\pi x_{\text{m}}^3$. After the mutant expands beyond the wildtype ball, we must consider the intersection of the two spheres. We write

$$\Delta(\tau) = \Delta^{(1)}(\tau)1\{[0,\tau_1]\} + \Delta^{(2)}(\tau)1\{[\tau_1,\tau_2]\} \tag{10}$$

with

$$\begin{aligned}\Delta^{(1)}(\tau) &= \frac{4}{3}\pi x_{\text{wt}}^3 - \frac{4}{3}\pi x_{\text{m}}^3 \\ &= \frac{4}{3}\pi(x + c_{\text{wt}}\tau)^3 - \frac{4}{3}\pi(c_{\text{m}}\tau)^3, \end{aligned}\tag{11}$$

$$\begin{aligned}\Delta^{(2)}(\tau) &= \frac{4}{3}\pi x_{\text{wt}}^3 - N_{\text{s-s}}(x_{\text{wt}}, x_{\text{m}}) \\ &= \frac{4}{3}\pi(x + c_{\text{wt}}\tau)^3 - N_{\text{s-s}}(x + c_{\text{wt}}\tau, c_{\text{m}}\tau), \end{aligned}\tag{12}$$

where $N_{\text{s-s}}$ describes the intersection of two spheres with radii x_{wt} and x_{m} (see Appendix B or Ref. [36] for the formula). We substitute $x_{\text{wt}} = x + c_{\text{wt}}\tau$ and $x_{\text{m}} = c_{\text{m}}\tau$ because we want to integrate over τ, x and y. The time τ_1 at which the mutant reaches the boundary of the wildtype is determined by

$$\tau_1 c_{\text{m}} = (x - y) + \tau_1 c_{\text{wt}} \Rightarrow \tau_1 = \frac{x - y}{c_{\text{m}} - c_{\text{wt}}}. \tag{13}$$

Furthermore, the time τ_2 at which the mutant has entirely replaced the wildtype is given by

$$\tau_2 c_{\text{m}} = (x + y) + \tau_1 c_{\text{wt}} \Rightarrow \tau_2 = \frac{x + y}{c_{\text{m}} - c_{\text{wt}}}. \tag{14}$$

Using claim 1, the conditional sweep probability is then

$$\Pr(\text{sweep}|X = x, Y = y) = e^{-\mu\int_0^\infty \Delta(\tau)\,d\tau}. \tag{15}$$

Although the integral of $\Delta(\tau)$ fails to yield a simple solution when the intersection formula $N_{\text{s-s}}$ is inserted, we can obtain a useful upper bound by using a simplifying approximation. We note that, due to geometrical symmetry, the time needed for a mutant to sweep the entire wildtype (τ_2 in Equation 14) is minimal when $y = 0$. Therefore, instead of locating the emergence of the first mutant according to the probability distribution $f_Y(y)$ in eqn. 8, we can obtain an upper bound on the sweep probability by assuming that the first mutant arises at the centre of the wildtype population. We thus approximate the conditional probability density by a delta function at $y = 0$:

$$f_Y(y|X = x) = \delta(y). \tag{16}$$

It follows directly that the unconditional probability distribution of Y is $f_Y(y) = \delta(y)$. With this simplification, the time at which the mutant leaves the wildtype and the time at which the mutant overtakes the wildtype are the same ($\tau_1 = \tau_2 = \frac{x}{c_{\text{m}} - c_{\text{wt}}}$) and the formula for the remaining wildtype population is given by $\Delta^{(1)}(\tau)$ for the relevant time interval $[0, \tau_1]$. Solving the integral in eqn. 15 we obtain

$$\Pr(\text{sweep}|X = x, Y = 0) = e^{-(x/\alpha)^4}, \tag{17}$$

where $\alpha = \sqrt[4]{\frac{3(c_{\text{m}} - c_{\text{wt}})^3}{\pi\mu\left(c_{\text{wt}}^2 - 3c_{\text{wt}}c_{\text{m}} + 3c_{\text{m}}^2\right)}}$ is a characteristic length. As one might expect, this sweep probability decreases monotonically with the wildtype radius, meaning that late-arriving mutants are less likely to sweep (Figure 3D). In the analogous one-dimensional model, we can derive the exact as well as the approximate conditional sweep probabilities (Appendix C), while in two and three dimensions we instead numerically evaluated the integral in eqn. 15. We then marginalized out Y to obtain exact results for $\Pr(\text{sweep}|X = x)$ that we compared to the approximation $\Pr(\text{sweep}|X = x, Y = 0)$. We find that, for biologically relevant parameter values, the exact solutions are remarkably close to the upper bounds obtained with the approximation of Equation 17 (Figure 3D).

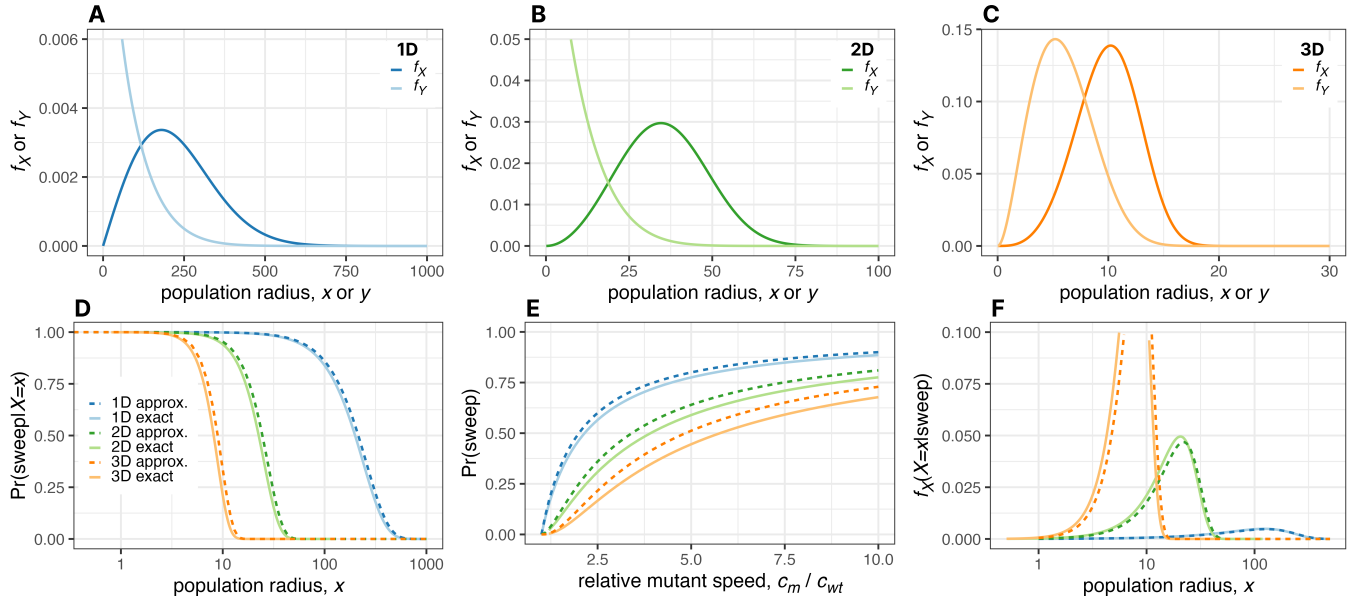

Figure 3. Analytical and numerical solutions of the macroscopic model. **A.-C.** Probability density of the wildtype population radius at the time the first surviving mutant arises (dark curves) and the distance between the origins of the mutation and wildtype expansions (light curves) in 1D **(A)**, 2D **(B)** and 3D **(C)**. **D.** Approximate and exact solutions for the sweep probability conditioned on the wildtype population radius. **E.** Approximate and exact solutions for the unconditional sweep probability. **F.** Approximate and exact solutions for the probability density of the wildtype population radius at the time the first surviving mutant arises, conditioned on this mutant achieving a selective sweep. In all plots we use parameter values relevant to cancer (see Section V): $c_{\mathrm{wt}} = 0.15$ cells per generation, $\tilde{\mu} = 10^{-5}$ mutations per division, $\rho = 0.23$ and $c_{\mathrm{m}} = 0.31$ cells per generation (except where varied). Formulas can be found in Table II.

D. Unconditional sweep probability

We now have all the necessary ingredients for the unconditional sweep probability, which is obtained by marginalizing out X and Y from the conditional sweep probability,

$$\begin{aligned}\Pr(\text{sweep}) &= \\ &\int_0^\infty \int_0^\infty \Pr(\text{sweep}|X=x, Y=y) f_Y(y|X=x) f_X(x)\, dy\, dx \\ &= \int_0^\infty \int_0^x e^{-\mu \int_0^\infty \Delta(\tau)\, d\tau} \frac{12y^2(x-y)}{x^4} \frac{4x^3 e^{-x^4/\theta^4}}{\theta^4}\, dy\, dx. \end{aligned} \tag{18}$$

Evaluating the integral using the simplified model of eqn. 16 and 17, we find

$$\begin{aligned}\Pr(\text{sweep}) &\leq \int_0^\infty \int_0^\infty e^{-\left(\frac{x}{\alpha}\right)^4} \delta(y) \frac{4x^3 e^{-x^4/\theta^4}}{\theta^4}\, dy\, dx \\ &= \left(\frac{c_{\mathrm{m}} - c_{\mathrm{wt}}}{c_{\mathrm{m}}}\right)^3. \end{aligned} \tag{19}$$

We thus arrive at a surprisingly simple formula that is independent of the mutation rate. This result generalizes to analogous one- and two-dimensional models, with the exponent 3 replaced by the respective spatial dimension (Appendices C and D). In each case, the sweep probability increases monotonically with the speed ratio $c_{\mathrm{m}}/c_{\mathrm{wt}}$ and converges slowly to 1 (Figure 3E). The approximate expressions are close to the exact solution in one dimension, and to numerical evaluations of the integral in two and three dimensions.

The sweep probability is independent of the mutation rate not only in the case where $y = 0$ but also in the general case with $f_Y(y|X = x)$ in eqn. 18. This follows from a more general result:

Claim 2. *Let $a > 0$. Let $H(x,y)$, $P(x,y)$ and $Q(x,y)$ be homogeneous functions in x and y with degrees $h, p, q > 0$ respectively and $Q(x,y) \neq 0$ for all $x > 0$ and $y > 0$. Then, the integral*

$$I = \int_0^\infty \int_0^x e^{-aH(x,y)} \frac{P(x,y)}{Q(x,y)}\, dy\, dx \tag{20}$$

is proportional to $a^{\frac{q-p-2}{h}}$.

Proof. See Appendix A. □

Now we consider the particular functions

$$H(x,y) = \int_0^\infty \Delta(\tau)\, d\tau + x^4 \frac{3c_{\mathrm{wt}}}{\pi}, \tag{21}$$

$$P(x,y) = 12y^2(x-y) \times 4x^3, \tag{22}$$

$$Q(x,y) = x^4 \frac{3c_{\mathrm{wt}}}{\pi}, \tag{23}$$

which are all independent of the mutation rate μ. Clearly $P(x,y)$ and $Q(x,y)$ are homogeneous functions of degree 6 and 4, respectively, and we also have

Claim 3. *$H(x,y)$ is homogeneous of degree 4.*

Proof. See appendix B. □

Combining these functions, we obtain the sweep probability

$$\Pr(\text{sweep}) = \mu \int_0^\infty \int_0^x e^{-\mu H(x,y)} \frac{P(x,y)}{Q(x,y)} \, dy \, dx. \quad (24)$$

Applying claim 2 with $a = \mu$ then yields

$$\Pr(\text{sweep}) \propto \mu^1 \times \mu^{\frac{4-6-2}{4}} = \mu^0 = 1. \quad (25)$$

Applying claim 2 in the one- and the two-dimensional cases also leads to results that are independent of the mutation rate. This agrees with the exact analytical solution for the unconditional sweep probability in one dimension (Appendix C).

To gain intuition as to why the sweep probability is independent of mutation rate, we note that increasing the mutation rate will have two opposing effects. The first surviving mutant is likely to arrive earlier and so has less distance to travel to achieve a sweep. On the other hand, a second mutant is more likely to arise in a given volume of wildtype cells. In our model, these two effects cancel each other out, whether we use the probability distribution of eqn. 6, eqn. 16, or indeed any other probability distribution $f_Y(y|X = x)$ that is homogeneous with degree 1. Nevertheless, independence does not necessarily hold true when our assumptions are varied. For instance, a time-dependent propagation speed will generally lead to non-homogeneous functions for $N(t)$ and $\int_0^\infty \Delta(\tau) \, d\tau$, which in turn will lead to a more complicated relationship between the sweep probability and the mutation rate.

E. Conditional arrival time of the first mutant

In biological systems where it is infeasible to track evolutionary dynamics, selective sweeps must be inferred from subsequent genetic data. For example, we might observe a public mutation in a tumour and ask when this mutation occurred. We can use our model to find a probability distribution of the radius X at the time the first mutant arose, given that we observe a selective sweep, simply by applying Bayes' theorem:

$$f_X(x|\text{sweep}) = \frac{\Pr(\text{sweep}|X = x) f_X(x)}{\Pr(\text{sweep})}. \quad (26)$$

Using the results from the simplified model, we find that the conditional probability distribution is the Weibull distribution

$$f_X(x|\text{sweep}) \, dx = \frac{4x^3}{\beta^3 \theta^4} e^{-\frac{x^4}{\beta^3 \theta^4}} \, dx \quad (27)$$

where $\beta = \frac{c_\text{m} - c_\text{wt}}{c_\text{m}}$ rescales the characteristic length scale from eqn. 4. Similar results are obtained in one and two dimensions and are together presented in Figure 3F.

IV. SIMULATIONS

A. An agent-based realisation of the two-dimensional spatial Moran process

To gauge the robustness and generality of our predictions, we measured the frequency of selective sweeps in a two-dimensional agent-based model. We suppose that the wildtype population invades a habitat initially occupied by a resident competitor, which is a plausible biological scenario for both invasive species [37] and invasive tumours [27, 29]. Our agent-based model thus has three types of individuals: resident, wildtype invader, and mutant invader. Localised competition between wildype and resident slows the wildtype expansion and creates potential for selective sweeps. We implemented this model using the *demon* agent-based modelling framework [38] within the *warlock* computational workflow [39], which facilitates running large numbers of simulations on a high-performance computing cluster. We have previously applied the same framework to studying cancer evolution [27, 39]. Further model details are given in Appendix G.

The agent-based simulations provide a useful test of our results because, although the general setup is the same, they differ in several ways from our macroscopic model. Space in the simulations is divided into discrete patches; the times between birth and dispersal events are exponentially distributed random variables (constituting another source of stochasticity); population boundaries are rough, not smooth; and the expansion wave front is typically not sharp, and changes shape as the wave progresses. Hence we would not expect perfect agreement between the results of the two models.

B. Linking the microscopic and macroscopic models

Our agent-based model approximately resembles a spatial death-birth Moran process (also known as the stepping stone model) [40, 41]. Expansion speeds in the spatial Moran process can in turn be approximated using the Fisher-Kolmogorov-Petrovsky-Piscounov (FKPP) equation [34, 35], which predicts that the mutant will expand within the wildtype with constant radial expansion speed c_m dependent on the difference in their proliferation rates $\Delta r_\text{m} = r_\text{m} - r_\text{wt}$ [30, 31, 33].

To compare the results of our discrete-space simulations to our continuous-space macroscopic model, we measured the propagation speeds of the wildtype

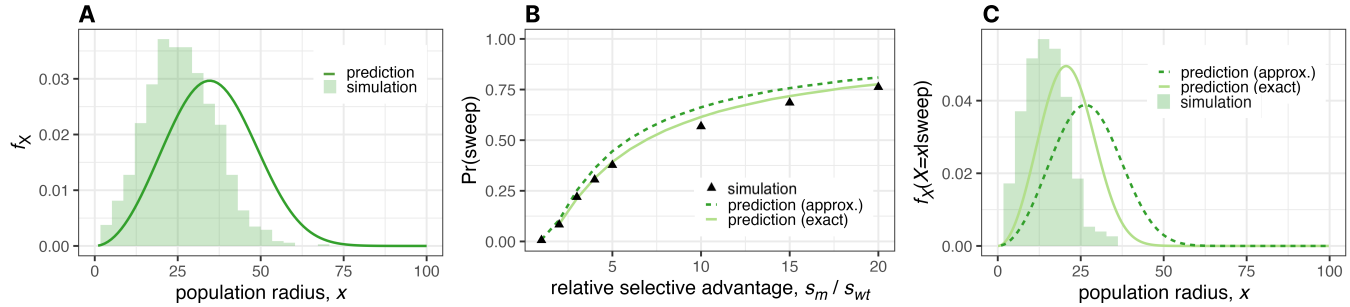

Figure 4. Simulation results versus predictions from the macroscopic model. **A.** Probability density of the wildtype population radius at the time the first surviving mutant arises. The histogram represents 1,000 simulations. **B.** Unconditional selective sweep probability versus ratio of selective advantages. Each data point is computed from 1,000 simulations. **C.** Probability density of the wildtype population radius at the time the first surviving mutant arises, conditioned on the mutant achieving a selective sweep. The histogram consists of radii measured in simulations that resulted in a selective sweep, among a total of 1,000 simulations. In all plots we use parameter values relevant to cancer (see Section V). For panels **A** and **C**, we use $r_{\mathrm{n}} = 0.91$, $r_{\mathrm{wt}} = 1$ such that $s_{\mathrm{wt}} = \frac{r_{\mathrm{wt}} - r_{\mathrm{n}}}{r_{\mathrm{n}}} = 0.1$ leading to speed $c_{\mathrm{wt}} = 0.15$ (see Appendix H). Also $r_{\mathrm{m}} = 1.3$, which implies $s_{\mathrm{m}} = \frac{r_{\mathrm{m}} - r_{\mathrm{wt}}}{r_{\mathrm{wt}}} = 0.3$ and hence $c_{\mathrm{m}} = 0.31$ (see Appendix H). The mutation rate is set to $\mu = 10^{-5}$ per division and the deme carrying capacity is $K = 16$, leading to local survival probability $\rho = \frac{1-(r_{\mathrm{wt}}+s_{\mathrm{m}})^{-1}}{1-(r_{\mathrm{wt}}+s_{\mathrm{m}})^{-N}} \approx 0.23$. In **B**, varying r_{m} from 1.1 to 3 corresponds to varying s_{m} from 0.1 to 2. The correspondence between s and c is given in Table I.

within the resident, and of the mutant within the wildtype (Table I). We did this by applying linear regression to the linear phase of the effective radius growth curves (the effective radius is the square root of the population size over π). Further investigations of propagation speeds in this model will be the subject of a further study.

C. Simulation results

Given the considerable differences between the models, the probability density functions resulting from the macroscopic and microscopic models are reassuringly consistent. The radius at the time the first surviving mutant arises is slightly lower in the simulations than in our analytical model (Figure 4A shows a typical case). This difference persists when conditioning on the mutant achieving a selective sweep (Figure 4C). These offsets can likely be explained by discretization effects and the fact that, in the simulations, the propagation front needs to be established before the expansion can proceed. The sweep probabilities in simulations are nevertheless very close to our analytical predictions (Figure 4B).

V. PARAMETER ESTIMATION

Our model has only three parameters: the mutation rate conditional on survival, μ, and the wildtype and mutant propagation speeds, c_{wt} and c_{m}. Here, we estimate these parameters for cancers such that we obtain numerical estimates for the expectations that can be compared to experimental observations.

To estimate the propagation speed, we follow a similar procedure to Ref. [20] and [42]. Consider a tumor that grows to volume V between 1 and 10 cm^3 in time T between 5 and 20 years. The propagation speed can then be estimated by

$$\tilde{c} = \frac{r}{T} = \frac{\sqrt[3]{\frac{3}{4\pi}V}}{T},$$

which equates to between 1 and 40 μm per day. Given that the diameter of a colon cancer cell is $l \approx 20\,\mu$m [43] and the generation time (cell cycle time) is $\tau_G \approx 4$ days [44], we can switch units to obtain $c = \tilde{c} \times \tau_G / l$, which is between 0.15 and 7 cell diameters per generation.

The consensus in the literature is that the rate of acquiring advantageous or "driver" mutations is of the order of $\tilde{\mu} = 10^{-5}$ per cell per generation [27, 42]. For the survival probability, assuming a spatial Moran process [35, 41] with a local population size of at least 10 cells and mutant relative fitness between 1.1 and 2, we obtain a survival probability ρ between 0.09 and 0.5, in agreement with a recently inferred value in colorectal tumors [45].

These parameter values imply a typical length scale

$$\theta = \sqrt[4]{\frac{3c_{wt}}{\pi\mu}} \approx 10 \text{ to } 50 \text{ cells}.$$

The expected values of X and Y are then $\mathbb{E}[X] = 10$ to 50 cell diameters and $\mathbb{E}[Y] = 7$ to 30 cell diameters.

Assuming c_{m} is at most ten times c_{wt}, we find that the expected tumour radius when a selective sweep occurred, given that a sweep did occur, is no more than 50 cells, corresponding to a population size N of no more

than 400,000 cells. The time for the sweep to be completed is then $\tau_2 \approx 40$ generations, at a population size of approximately $800,000$ cells.

To examine the robustness of these results, we note that the peak of the conditional sweep probability distribution in d dimensions (e.g. Equation 27 for $d = 3$) is at $C\theta\gamma^{\frac{d}{d+1}}$, where $C = \left(\frac{d}{d+1}\right)^{\frac{1}{d+1}}$. It follows that, in three dimensions, the radius that maximizes the sweep probability is highly robust to varying μ or c_{wt}. For example, if we vary μ or c_{wt} by a factor of 100 then the peak radius will change by a factor of only $\sqrt[4]{100} \approx 3$.

We conclude that sweeps are likely to occur only during early tumour growth.

VI. DISCUSSION

Here we have used mathematical modelling and analysis to determine the expected frequency of "classic" selective sweeps versus incomplete or soft sweeps during range expansions. We find that this frequency is generally expected to be low, even for mutations with very high selection coefficients. Moreover, when the wildtype and mutant radial expansion speeds are constant, the sweep probability is independent of the mutation rate and can be expressed solely in terms of those speeds (which can in turn be related, through the FKPP equation or other standard models, to the selection coefficient, dispersal rates, and other basic parameters). An intuitive explanation for this finding is that if the mutation rate is higher then, on the one hand, the first advantageous mutation is likely to arise in a smaller population – meaning that it has less distance to travel to achieve a sweep – but, on the other hand, competing mutations will also tend to arise sooner than in the case of a lower mutation rate. These two effects exactly cancel out under the assumption of constant radial growth speeds.

We make three arguments to justify our focus on this particular model of range expansions. First, our model corresponds to the continuum approximation of standard mathematical models of range expansions and spatial population genetics: the spatial Moran process (or stepping stone model) and the biassed voter model (which is equivalent to a spatial Moran process with only one individual per deme, or to an Eden growth model extended to allow local dispersal and competition throughout the population). These models are well understood, intuitive, and easy to parameterise. Second, because our model is relatively permissive to selective sweeps, it provides useful upper bounds for selective sweep probabilities in more complex scenarios. Selective sweeps will be even less frequent when dispersal rates are higher near the boundary, or when the wildtype radial expansion speed increases over time. Third, in much the same way as the Moran and Wright-Fisher processes are the most useful, tractable models of evolution in constant-sized, non-spatial populations, so the constant-radial-speed model yields the clearest rule of thumb results for range expansions. For instance, if the radial expansion speed at which a mutant spreads within the wildtype population is twice the speed at which the wildtype expands then the probability of this mutant achieving a selective sweep can be approximated simply as $(1 - 1/2)^2 = 1/4$ in two dimensions and $1/8$ in three dimensions.

Alternative models have been considered previously. Antal and colleagues [20] used a macroscopic model similar to ours to investigate the case in which mutations arise only at the boundary of a range expansion. Given that selective sweeps are then impossible, the interesting outcome is when the mutant envelops the wildtype. Ralph & Coop [46] and Martens and colleagues [19, 47] instead considered constant-sized populations. They found that selective sweeps are likely only if the population width is much smaller than a characteristic length scale, which depends on the mutation rate, the dispersal rate, the effective local population density, and the strength of selection. In Appendix E we compare our results to those of these prior studies and consider other alternative growth laws.

In our model, a selective sweep can occur only if the rate of spread of an advantageous mutation exceeds the expansion speed of the wildtype population. This scenario is plausible, for example, in the case of an invasive species that is still adapting to its new conditions and whose range expansion is slowed by the need to modify its environment (niche construction) or by interspecific interactions [37]. When the invader must displace a resident competitor (as in our agent-based simulations), the sweep probability can be approximated via the FKPP solution as $(1 - \sqrt{\frac{r_{\mathrm{m}} - r_{\mathrm{wt}}}{r_{\mathrm{wt}} - r_{\mathrm{n}}}})^d$, where r_{m}, r_{wt} and r_{n} are the growth rates of mutant invader, wildtype invader and resident populations respectively, and d is the spatial dimension. This expression makes clear that selective sweeps are most likely to occur in species invading essentially linear habitats, such as coastlines.

Our three-dimensional results, though less applicable to invasive species, are particularly relevant to understanding the general nature of solid tumour evolution. There are two ways in which tumours might acquire the multiple clonal drivers (advantageous mutations) observed in sequencing studies. The first route is via sequential population bottlenecks. Suppose that early tumour development comprises not one but several successive range expansions. Growth repeatedly stalls due to constraints such as hypoxia, immune control, and physical barriers. Driver mutations enable subclones to escape these constraints and invade new territory, each time purging genetic diversity so that the final, prolonged expansion originates from a single highly transformed cell. This episodic model is conventional and has been particularly well characterised in recent studies of colorectal cancer [48, 49] and breast

cancer [50]. The second way to acquire clonal drivers is via selective sweeps during the final range expansion. Our results suggest that the first process is the more important. Selective sweeps of even extremely strong drivers are highly unlikely to occur in tumours containing more than a few hundred thousand cells, equivalent to less than a cubic millimeter in volume. Pervasive genetic heterogeneity and parallel evolution are therefore revealed to be straightforward consequences of spatial structure, expected to be ubiquitous across solid tumours, irrespective of mutation rate and degree of genomic instability.

CONTRIBUTIONS

RN conceived the research question and supervised the project. RN and AS designed the research. AS and RK carried out the mathematical analysis. MB performed agent-based simulations. All authors wrote and approved the manuscript.

ACKNOWLEDGMENTS

AS was supported by the European Union's Horizon 2020 research and innovation programme under the Marie Składowska-Curie EvoGamesPlus grant agreement no. 955708. MB was supported by an award from the City, University of London Research Pump-priming Fund. RN was supported by the National Cancer Institute of the National Institutes of Health under Award Number U54CA217376. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

[1] Oskar Hallatschek, Pascal Hersen, Sharad Ramanathan, and David R Nelson. Genetic drift at expanding frontiers promotes gene segregation. *Proceedings of the National Academy of Sciences*, 104(50):19926–19930, 2007.

[2] Zaira Seferbekova, Artem Lomakin, Lucy R Yates, and Moritz Gerstung. Spatial biology of cancer evolution. *Nature Reviews Genetics*, 24(5):295–313, 2023.

[3] Leslie G Biesecker and Nancy B Spinner. A genomic view of mosaicism and human disease. *Nature Reviews Genetics*, 14(5):307–320, 2013.

[4] Iñigo Martincorena, Amit Roshan, Moritz Gerstung, Peter Ellis, Peter Van Loo, Stuart McLaren, David C Wedge, Anthony Fullam, Ludmil B Alexandrov, Jose M Tubio, et al. High burden and pervasive positive selection of somatic mutations in normal human skin. *Science*, 348(6237):880–886, 2015.

[5] Luigi L Cavalli-Sforza, Paolo Menozzi, and Alberto Piazza. Demic expansions and human evolution. *Science*, 259(5095):639–646, 1993.

[6] Gretta T Pecl, Miguel B Araújo, Johann D Bell, Julia Blanchard, Timothy C Bonebrake, I-Ching Chen, Timothy D Clark, Robert K Colwell, Finn Danielsen, Birgitta Evengård, et al. Biodiversity redistribution under climate change: Impacts on ecosystems and human well-being. *Science*, 355(6332):eaai9214, 2017.

[7] Emily V Moran and Jake M Alexander. Evolutionary responses to global change: lessons from invasive species. *Ecology Letters*, 17(5):637–649, 2014.

[8] Laurent Excoffier, Matthieu Foll, and Rémy J Petit. Genetic consequences of range expansions. *Annual Review of Ecology, Evolution, and Systematics*, 40:481–501, 2009.

[9] Oskar Hallatschek and David R Nelson. Life at the front of an expanding population. *Evolution*, 64(1):193–206, 2010.

[10] Kirill S Korolev, Melanie JI Müller, Nilay Karahan, Andrew W Murray, Oskar Hallatschek, and David R Nelson. Selective sweeps in growing microbial colonies. *Physical biology*, 9(2):026008, 2012.

[11] Diana Fusco, Matti Gralka, Jona Kayser, Alex Anderson, and Oskar Hallatschek. Excess of mutational jackpot events in expanding populations revealed by spatial luria–delbrück experiments. *Nature communications*, 7(1):12760, 2016.

[12] Serhii Aif, Nico Appold, Lucas Kampman, Oskar Hallatschek, and Jona Kayser. Evolutionary rescue of resistant mutants is governed by a balance between radial expansion and selection in compact populations. *Nature Communications*, 13(1):7916, 2022.

[13] Douglas Hanahan and Robert A Weinberg. The hallmarks of cancer. *cell*, 100(1):57–70, 2000.

[14] Eric R Fearon and Bert Vogelstein. A genetic model for colorectal tumorigenesis. *cell*, 61(5):759–767, 1990.

[15] Moritz Gerstung, Clemency Jolly, Ignaty Leshchiner, Stefan C Dentro, Santiago Gonzalez, Daniel Rosebrock, Thomas J Mitchell, Yulia Rubanova, Pavana Anur, Kaixian Yu, et al. The evolutionary history of 2,658 cancers. *Nature*, 578(7793):122–128, 2020.

[16] Marco Gerlinger, Andrew J Rowan, Stuart Horswell, James Larkin, David Endesfelder, Eva Gronroos, Pierre Martinez, Nicholas Matthews, Aengus Stewart, Patrick Tarpey, et al. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *New England journal of medicine*, 366(10):883–892, 2012.

[17] Alexander Davis, Ruli Gao, and Nicholas Navin. Tumor evolution: Linear, branching, neutral or punctuated? *Biochimica et Biophysica Acta (BBA)-Reviews on Cancer*, 1867(2):151–161, 2017.

[18] Samra Turajlic, Andrea Sottoriva, Trevor Graham, and Charles Swanton. Resolving genetic heterogeneity in cancer. *Nature Reviews Genetics*, 20(7):404–416, 2019.

[19] Erik A Martens, Rumen Kostadinov, Carlo C Maley, and Oskar Hallatschek. Spatial structure increases the waiting time for cancer. *New journal of physics*, 13(11):115014, 2011.

[20] Tibor Antal, PL Krapivsky, and MA Nowak. Spatial evolution of tumors with successive driver mutations. *Physical Review E*, 92(2):022705, 2015.

[21] Chay Paterson, Martin A Nowak, and Bartlomiej Waclaw. An exactly solvable, spatial model of mutation accumulation in cancer. *Scientific reports*, 6(1):39511, 2016.

[22] Peter Armitage and Richard Doll. The age distribution of cancer and a multi-stage theory of carcinogenesis. *British journal of cancer*, 91(12):1983–1989, 2004.

[23] Chay Paterson, Hans Clevers, and Ivana Bozic. Mathematical model of colorectal cancer initiation. *Proceedings of the National Academy of Sciences*, 117(34):20681–20688, 2020.

[24] Michael D Nicholson, David Cheek, and Tibor Antal. Sequential mutations in exponentially growing populations. *PLOS Computational Biology*, 19(7):e1011289, 2023.

[25] Ruping Sun, Zheng Hu, Andrea Sottoriva, Trevor A Graham, Arbel Harpak, Zhicheng Ma, Jared M Fischer, Darryl Shibata, and Christina Curtis. Between-region genetic divergence reflects the mode and tempo of tumor evolution. *Nature genetics*, 49(7):1015–1024, 2017.

[26] Jeffrey West, Ryan O Schenck, Chandler Gatenbee, Mark Robertson-Tessi, and Alexander RA Anderson. Normal tissue architecture determines the evolutionary course of cancer. *Nature communications*, 12(1):2060, 2021.

[27] 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.

[28] Xiao Fu, Yue Zhao, Jose I Lopez, Andrew Rowan, Lewis Au, Annika Fendler, Steve Hazell, Hang Xu, Stuart Horswell, Scott TC Shepherd, et al. Spatial patterns of tumour growth impact clonal diversification in a computational model and the tracerx renal study. *Nature ecology & evolution*, 6(1):88–102, 2022.

[29] Robert Noble, John T Burley, Cécile Le Sueur, and Michael E Hochberg. When, why and how tumour clonal diversity predicts survival. *Evolutionary applications*, 13(7):1558–1568, 2020.

[30] Ronald Aylmer Fisher. The wave of advance of advantageous genes. *Annals of eugenics*, 7(4):355–369, 1937.

[31] Andrey Nikolaevich Kolmogorov, Ivan Georgievich Petrovsky, and Nikolay Semenovich Piskunov. A study of the diffusion equation with increase in the amount of substance, and its application to a biological problem. *Bullentin of Moscow State University*, 1937.

[32] Eric Brunet and Bernard Derrida. Shift in the velocity of a front due to a cutoff. *Physical Review E*, 56(3):2597, 1997.

[33] James Dickson Murray. *Mathematical Biology: I: An introduction*, volume 17. Springer, 2002.

[34] Oskar Hallatschek and Kirill S Korolev. Fisher waves in the strong noise limit. *Physical review letters*, 103(10):108103, 2009.

[35] Bahram Houchmandzadeh and Marcel Vallade. Fisher waves: An individual-based stochastic model. *Physical Review E*, 96(1):012414, 2017.

[36] Eric W Weisstein. Sphere-sphere intersection. from mathworld–a wolfram web resource. https://mathworld.wolfram.com/Sphere-SphereIntersection.html. Accessed: 22/11/2023.

[37] Jens-Christian Svenning, Dominique Gravel, Robert D Holt, Frank M Schurr, Wilfried Thuiller, Tamara Münkemüller, Katja H Schiffers, Stefan Dullinger, Thomas C Edwards Jr, Thomas Hickler, et al. The influence of interspecific interactions on species range expansion rates. *Ecography*, 37(12):1198–1209, 2014.

[38] Robert Noble. demon: deme-based oncology model. https://github.com/robjohnnoble/demonmodel, 2023.

[39] Maciej Bak, Blair Colyer, Veselin Manojlović, and Robert Noble. Warlock: an automated computational workflow for simulating spatially structured tumour evolution. *arXiv preprint arXiv:2301.07808*, 2023.

[40] Richard Durrett, Jasmine Foo, and Kevin Leder. Spatial moran models, ii: cancer initiation in spatially structured tissue. *Journal of mathematical biology*, 72:1369–1400, 2016.

[41] Motoo Kimura and George H Weiss. The stepping stone model of population structure and the decrease of genetic correlation with distance. *Genetics*, 49(4):561, 1964.

[42] Ivana Bozic, Tibor Antal, Hisashi Ohtsuki, Hannah Carter, Dewey Kim, Sining Chen, Rachel Karchin, Kenneth W Kinzler, Bert Vogelstein, and Martin A Nowak. Accumulation of driver and passenger mutations during tumor progression. *Proceedings of the National Academy of Sciences*, 107(43):18545–18550, 2010.

[43] Babita Shashni, Shinya Ariyasu, Reisa Takeda, Toshihiro Suzuki, Shota Shiina, Kazunori Akimoto, Takuto Maeda, Naoyuki Aikawa, Ryo Abe, Tomohiro Osaki, et al. Size-based differentiation of cancer and normal cells by a particle size analyzer assisted by a cell-recognition pc software. *Biological and Pharmaceutical Bulletin*, 41(4):487–503, 2018.

[44] Siân Jones, Wei-dong Chen, Giovanni Parmigiani, Frank Diehl, Niko Beerenwinkel, Tibor Antal, Arne Traulsen, Martin A Nowak, Christopher Siegel, Victor E Velculescu, et al. Comparative lesion sequencing provides insights into tumor evolution. *Proceedings of the National Academy of Sciences*, 105(11):4283–4288, 2008.

[45] Benjamin Werner, Jack Case, Marc J Williams, Ketevan Chkhaidze, Daniel Temko, Javier Fernández-Mateos, George D Cresswell, Daniel Nichol, William Cross, Inmaculada Spiteri, et al. Measuring single cell divisions in human tissues from multi-region sequencing data. *Nature communications*, 11(1):1035, 2020.

[46] Peter Ralph and Graham Coop. Parallel adaptation: one or many waves of advance of an advantageous allele? *Genetics*, 186(2):647–668, 2010.

[47] Erik A Martens and Oskar Hallatschek. Interfering waves of adaptation promote spatial mixing. *Genetics*, 189(3):1045–1060, 2011.

[48] Marc D Ryser, Diego Mallo, Allison Hall, Timothy Hardman, Lorraine M King, Sergei Tatishchev, Inmaculada C Sorribes, Carlo C Maley, Jeffrey R Marks, E Shelley Hwang, et al. Minimal barriers to invasion during human colorectal tumor growth. *Nature communications*, 11(1):1280, 2020.

[49] William Cross, Michal Kovac, Ville Mustonen, Daniel Temko, Hayley Davis, Ann-Marie Baker, Sujata Biswas, Roland Arnold, Laura Chegwidden, Chandler Gatenbee, et al. The evolutionary landscape of colorectal tumorigenesis. *Nature ecology & evolution*, 2(10):1661–1672, 2018.

[50] Artem Lomakin, Jessica Svedlund, Carina Strell, Milana Gataric, Artem Shmatko, Gleb Rukhovich, Jun Sung Park, Young Seok Ju, Stefan Dentro, Vitalii Kleshchevnikov, et al. Spatial genomics maps the structure, nature and evolution of cancer clones. *Nature*, 611(7936):594–602, 2022.

[51] Eric W Weisstein. Circle-circle intersection. from mathworld–a wolfram web resource. https://mathworld.wolfram.com/Circle-CircleIntersection.html. Accessed: 22/11/2023.

[52] Sébastien Benzekry, Clare Lamont, Afshin Beheshti, Amanda Tracz, John ML Ebos, Lynn Hlatky, and Philip Hahnfeldt. Classical mathematical models for description and prediction of experimental tumor growth. *PLoS computational biology*, 10(8):e1003800, 2014.

[53] Richard Durrett. *Branching process models of cancer*. Mathematical Biosciences Institute Lecture Series. Springer, 2015.

[54] Ivana Bozic, Chay Paterson, and Bartlomiej Waclaw. On measuring selection in cancer from subclonal mutation frequencies. *PLoS computational biology*, 15(9):e1007368, 2019.

APPENDICES

Appendix A: Proof of claim 2

We show the relation by making the substitutions

$$\hat{x} = a^{1/h} x \quad \text{and} \quad \hat{y} = a^{1/h} y$$

such that

$$\begin{aligned} H(\hat{x}, \hat{y}) &= a H(x, y) \\ P(\hat{x}, \hat{y}) &= a^{p/h} P(x, y) \\ Q(\hat{x}, \hat{y}) &= a^{q/h} Q(x, y) \end{aligned}$$

and

$$dx = a^{-1/h} d\hat{x} \quad \text{and} \quad dy = a^{-1/h} d\hat{y}.$$

Inserting the substitutions in eqn. 20, we obtain

$$\begin{aligned} I &= \int_0^\infty \int_0^x e^{-aH(x,y)} \frac{P(x,y)}{Q(x,y)} \, dy \, dx \\ &= \int_0^\infty \int_0^{\hat{x}} e^{-H(\hat{x},\hat{y})} \frac{a^{-p/h} P(\hat{x},\hat{y})}{a^{-q/h} Q(\hat{x},\hat{y})} a^{-1/h} a^{-1/h} \, d\hat{y} \, d\hat{x} \\ &= a^{-2/h} a^{(q-p)/h} \int_0^\infty \int_0^{\hat{x}} e^{-H(\hat{x},\hat{y})} \frac{P(\hat{x},\hat{y})}{Q(\hat{x},\hat{y})} \, d\hat{y} \, d\hat{x} \end{aligned}$$

The remaining integral is no longer dependent on parameter a and shows the desired scaling with $a^{\frac{q-p-2}{h}}$.

Appendix B: Proof of claim 3

We show that $H(x, y)$ as given in eqn. 21 is homogeneous with degree $h = 4$. We first note that the second term is a polynomial of degree 4 such that homogeneity is trivial. The sum of two homogeneous functions is again homogeneous such that we need to show homogeneity only for the first term, which we can write out as the sum of two integrals,

$$\int_0^\infty \Delta(\tau) \, d\tau = \int_0^{\tau_1} \Delta^{(1)}(\tau) \, d\tau + \int_{\tau_1}^{\tau_2} \Delta^{(2)}(\tau) \, d\tau, \quad \text{(B1)}$$

which is specified by eqn. 11 - 14.

The first term is homogeneous in x and y with degree $h = 4$. We take the integral as function of x and y, and insert μx and μy,

$$\begin{aligned} &\left[\int_0^{\tau_1} \Delta^{(1)}(\tau) d\tau \right] (\mu x, \mu y) \\ &= \int_0^{\tau_1} \frac{4}{3} \pi ((c_{\mathrm{wt}} \tau + \mu x)^3 - (c_{\mathrm{m}} \tau)^3) d\tau, \end{aligned} \quad \text{(B2)}$$

where we note that also x and y in the integral boundary τ_1 are replaced such that $\tau_1 = \frac{\mu x - \mu y}{c_{\mathrm{m}} - c_{\mathrm{wt}}}$. We now substitute $\hat{\tau} = \mu^a \tau$ and obtain

$$\int_0^{\hat{\tau}_1} \frac{4}{3} \pi ((c_{\mathrm{wt}} \mu^{-a} \hat{\tau} + \mu x)^3 - (c_{\mathrm{m}} \mu^{-a} \hat{\tau})^3) (\mu^{-a} \, d\hat{\tau}), \quad \text{(B3)}$$

where $\hat{\tau}_1 = \frac{\mu x - \mu y}{c_{\mathrm{m}} - c_{\mathrm{wt}}} \times \mu^a$. We choose $a = -1$ such that $\hat{\tau}_1 = \frac{x-y}{c_{\mathrm{m}} - c_{\mathrm{wt}}}$ and

$$\mu^4 \int_0^{\hat{\tau}_1} \frac{4}{3}\pi((c_{\mathrm{wt}}\hat{\tau} + x)^3 - (c_{\mathrm{m}}\hat{\tau})^3) d\hat{\tau}, \tag{B4}$$

which proves the homogeneity of degree $h = 4$ for the first term in eqn. B1.

After time τ_1 and we need to use the intersection formula for two spheres as stated in eqn. 12. We use Ref. [36] to write the intersection volume of two spheres with radii x_{wt} and x_{m} that are apart by distance y as

$$\begin{aligned} N_{\text{s-s}} =& \frac{\pi}{12y}(x_{\mathrm{wt}} + x_{\mathrm{m}} + y)^2 \\ & (y^2 + 2yx_{\mathrm{m}} - 3x_{\mathrm{m}}^2 + 2yx_{\mathrm{wt}} + 6x_{\mathrm{wt}}x_{\mathrm{m}} - 3x_{\mathrm{wt}}^2). \end{aligned} \tag{B5}$$

We want to show that

$$\begin{aligned} & \left[\int_{\tau_1}^{\tau_2} \Delta^{(2)}(\tau) d\tau\right](\mu x, \mu y) \\ & = \mu^4 \left[\int_{\hat{\tau}_1}^{\hat{\tau}_2} \frac{4}{3}\pi x_{\mathrm{wt}}^3 - N_{\text{s-s}} d\hat{\tau}\right](x, y) \\ & = \mu^4 \left[\int_{\tau_1}^{\tau_2} \frac{4}{3}\pi x_{\mathrm{wt}}^3 - N_{\text{s-s}} d\tau\right](x, y). \end{aligned} \tag{B6}$$

This can be done analogously to the previous calculation. In brief, we first replace $x_{\mathrm{wt}} = x + c_{\mathrm{wt}}\tau$ and $x_{\mathrm{m}} = c_{\mathrm{m}}\tau$ in eqn. B5. Next, we write out the integral in the first line of eqn. B6 and insert μx and μy for x and y. Now, we use the same trick as before and substitute $\hat{\tau} = \mu^{-1}\tau$. This way, the integration boundaries transform back to $\hat{\tau}_1 = \frac{x-y}{c_{\mathrm{m}} - c_{\mathrm{wt}}}$ and $\hat{\tau}_2 = \frac{x+y}{c_{\mathrm{m}} - c_{\mathrm{wt}}}$. The factor μ^4 will drop out and we find the desired homogeneity from eqn. B6. Finally, the homogeneity of $H(x, y)$ follows from the homogeneity of the single terms.

Appendix C: Sweep probability in one dimension

We consider a population that is expanding in one dimension in two directions. If x_{wt} describes the length from the origin of the population to the leading edge, the total population size is $N = 2x_{\mathrm{wt}}$. We can follow the same derivations as in section III B, and obtain the probability density for the population radius X at the time the first surviving mutant arises

$$f_X(x) = \frac{2xe^{-x^2/\theta_{1\mathrm{D}}^2}}{\theta_{1\mathrm{D}}^2}, \tag{C1}$$

with $\theta_{1\mathrm{D}} = \sqrt{\frac{c_{\mathrm{wt}}}{\mu}}$ (incidentally note that, since $\pi \approx 3$, $\theta_{1\mathrm{D}}^2 \approx \theta_{2\mathrm{D}}^3 \approx \theta_{3\mathrm{D}}^4$). Furthermore, we obtain the conditional probability density for the location of the first surviving mutant Y,

$$f_Y(y|X = x) = \frac{2(x-y)}{x^2} 1\{y \le x\}, \tag{C2}$$

The unconditional probability density is

$$f_Y(y) = \frac{2}{\theta_{1\mathrm{D}}}\Gamma\left(\frac{1}{2}, \frac{y^2}{\theta_{1\mathrm{D}}^2}\right) - \frac{2y}{\theta_{1\mathrm{D}}^2}\Gamma\left(0, \frac{y^2}{\theta_{1\mathrm{D}}^2}\right) \tag{C3}$$

1. Results for the simplified model

We set $f_Y(y|X = x) = \delta(y)$. The conditional sweep is then

$$\Pr(\text{sweep}|X = x) = e^{-(x/\alpha_{1\mathrm{D}})^2}, \tag{C4}$$

where $\alpha_{1\mathrm{D}} = \sqrt{\frac{c_{\mathrm{m}} - c_{\mathrm{wt}}}{\mu}}$.

The unconditional sweep probability is

$$\Pr(\text{sweep}) = \frac{c_{\mathrm{m}} - c_{\mathrm{wt}}}{c_{\mathrm{m}}}. \tag{C5}$$

The probability distribution of X given that we observe sweep

$$f_X(X = x|\text{sweep}) = \frac{2x}{\theta_{1\mathrm{D}}^2\beta} e^{-\frac{x^2}{\theta_{1\mathrm{D}}^2\beta}} \tag{C6}$$

where $\beta = \frac{c_{\mathrm{m}} - c_{\mathrm{wt}}}{c_{\mathrm{m}}}$.

2. Results for the full model

The conditional sweep is found to be

$$\Pr(\text{sweep}|X = x, Y = y) = e^{(x^2+y^2)/\alpha_{1\mathrm{D}}^2} \tag{C7}$$

where $\alpha_{1\mathrm{D}} = \sqrt{\frac{c_{\mathrm{m}} - c_{\mathrm{wt}}}{\mu}}$ as in the simplified case. In the first step, we marginalize out Y, which gives

$$\Pr(\text{sweep}|X = x) = \frac{\sqrt{\pi}\frac{x}{\alpha_{1D}}\mathrm{Erf}\left(\frac{x}{\alpha_{1D}}\right) - 1 + e^{-2\frac{x^2}{\alpha_{1D}^2}}}{x^2/\alpha_{1D}^2}, \tag{C8}$$

where $\mathrm{Erf}(x) = \frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2}\,dt$ is the Gaussian error function.

The exact formula for the unconditional sweep probability is obtained by next marginalizing out X. The result is

$$\Pr(\text{sweep}) = \beta'\left[\frac{2\cot^{-1}(\sqrt{1+\beta'})}{\sqrt{1+\beta'}} + \ln\left(\frac{\beta'+1}{\beta'+2}\right)\right],$$

where $\beta' = \frac{c_{\mathrm{m}}-c_{\mathrm{wt}}}{c_{\mathrm{wt}}}$ is the speed difference relative to the wildtype speed. The two curves are very close to each other and we see that the approximation is indeed an upper bound. We also see that the full model yields an expression independent of the mutation rate. An alternative approximate sweep probability can be derived as the first-order Taylor expansion of the exact expression at zero. The result is $(\pi/2 - \log 2)\beta \approx 0.88\beta$, which is more accurate than β when $c_m/c_{wt} < 2.35$.

The probability distribution of X given that we have a sweep $f_X(x|\text{sweep})$ can be obtained using Bayes' theorem as in Section III E. We omit the inelegant result.

Appendix D: Sweep probability in two dimensions

1. Results for the simplified model

We consider a population that is expanding in two dimensions as disks. If x_{wt} is the radius of the population, the total population size is $N = \pi x_{\mathrm{wt}}^2$. We can follow the same derivations as in section III B, and obtain the probability density for the population radius X at the time the first surviving mutant arises

$$f_X(x) = \frac{3x^2 e^{-x^3/\theta_{2\mathrm{D}}^3}}{\theta_{2\mathrm{D}}^3}, \tag{D1}$$

with $\theta_{2\mathrm{D}} = \sqrt[3]{\frac{3c_{\mathrm{wt}}}{\pi\mu}}$. Furthermore, we obtain the conditional probability density for the location of the first surviving mutant Y,

$$f_Y(y|X=x) = \frac{6y(x-y)}{x^3} 1\{y \le x\}, \tag{D2}$$

The conditional sweep in the simplified model is

$$\Pr(\text{sweep}|X=x) = e^{-(x/\alpha_{2\mathrm{D}})^3}, \tag{D3}$$

where $\alpha_{2\mathrm{D}} = \sqrt[3]{\frac{3(c_{\mathrm{m}}-c_{\mathrm{wt}})^2}{\pi\mu(2c_{\mathrm{m}}-c_{\mathrm{wt}})}}$ and following section III C the unconditional sweep probability is then

$$\Pr(\text{sweep}) = \left(\frac{c_{\mathrm{m}} - c_{\mathrm{wt}}}{c_{\mathrm{m}}}\right)^2 \tag{D4}$$

The probability distribution of X given that we observe sweep is

$$f_X(X=x|\text{sweep}) = \frac{3x^2}{\theta_{2\mathrm{D}}^3\beta^2} e^{-\frac{x^3}{\theta_{2\mathrm{D}}^3\beta^2}}, \tag{D5}$$

where $\beta = \frac{c_{\mathrm{m}}-c_{\mathrm{wt}}}{c_{\mathrm{m}}}$.

2. Full model formulation

In order to calculate the probability of a selective sweep in a circular range expansion, we need to calculate the remaining wildtype population once a mutant has arisen. To do so, we need to calculate the intersection of two circles given that we know both radii x_{wt} and x_{m} and the distance of the centers of the circles y. In the simplified model, we had $y = 0$ such that the intersection is simply given by the area of the inner disk. However, we need a more general formula for the case $y \neq 0$.

The intersecting area of two circles with radii x_{wt}, x_{m} at distance y is given by [51]

$$\begin{aligned} N_{\text{c-c}} =& x_{\mathrm{wt}}^2 \cos^{-1}\left(\frac{y^2 + x_{\mathrm{wt}}^2 - x_{\mathrm{m}}^2}{2yx_{\mathrm{wt}}}\right) \\ &+ x_{\mathrm{m}}^2 \cos^{-1}\left(\frac{y^2 + x_{\mathrm{m}}^2 - x_{\mathrm{wt}}^2}{2yx_{\mathrm{m}}}\right) \\ &- \frac{1}{2}\big((-y + x_{\mathrm{wt}} + x_{\mathrm{m}})(y + x_{\mathrm{wt}} - x_{\mathrm{m}}) \\ &(y - x_{\mathrm{wt}} + x_{\mathrm{m}})(y + x_{\mathrm{wt}} + x_{\mathrm{m}})\big)^{1/2}. \end{aligned} \tag{D6}$$

As in the 3D case, we have the remaining wildtype population given by two formulas as written in eqn. 10 with

$$\begin{aligned} \Delta^{(1)}(\tau) &= \pi x_{\mathrm{wt}}^2 - \pi x_{\mathrm{m}}^2 \\ \Delta^{(2)}(\tau) &= \pi x_{\mathrm{wt}}^2 - N_{\text{c-c}}. \end{aligned} \tag{D7}$$

The formulas for the time τ_1 at which the mutant leaves the wildtype and the time τ_2 at which the mutant has overtaken the wildtype remain the same as given in eqn. 13 and 14.

We then need to replace $x_{\mathrm{wt}} = x + c_{\mathrm{wt}}\tau$ and $x_{\mathrm{m}} = c_{\mathrm{m}}\tau$.

The full integral that needs to be solved for the unconditional sweep probability reads

$$\begin{aligned} \Pr(\text{sweep}) &= \int_0^\infty \int_0^\infty \Pr(\text{sweep}|X=x, Y=y) \\ &\qquad f_Y(y|X=x) f_X(x)\, dy\, dx, \\ &= \int_0^\infty \int_0^x e^{-\mu \int_0^\infty \Delta(\tau) d\tau} \\ &\qquad \frac{6y(x-y)}{x^3} \frac{3x^2 e^{-x^2/\theta_{2\mathrm{D}}^3}}{\theta_{2\mathrm{D}}^3}\, dy\, dx. \end{aligned}$$

3. Independence of the mutation rate

We apply claim 2 on the integral form of Pr(sweep) by making the following definitions

$$\begin{aligned} H(x,y) &= \int_0^\infty \Delta(\tau)d\tau + x^3\frac{3c_{\text{wt}}}{\pi}, \\ P(x,y) &= 6y(x-y)\times 3x^2, \\ Q(x,y) &= x^3\frac{\pi}{3c\text{wt}}, \end{aligned} \tag{D8}$$

which are homogeneous and have degrees $h = 3$, $p = 4$ and $q = 3$. The application of claim 2 yields

$$\begin{aligned} \Pr(\text{sweep}) &= \mu \times \int_0^\infty \int_0^x e^{-aH(x,y)} \frac{P(x,y)}{Q(x,y)}\, dy\, dx \\ &\propto \mu \times \mu^{\frac{3-4-2}{3}} = \mu^0, \end{aligned} \tag{D9}$$

which proves independence in the two-dimensional case. Note that the homogeneity over $\int_0^\infty \Delta(\tau)d\tau$ can be shown in the same way with the same substitutions as in appendix B. The independence of the mutation rate is confirmed by simulations (Figure 5).

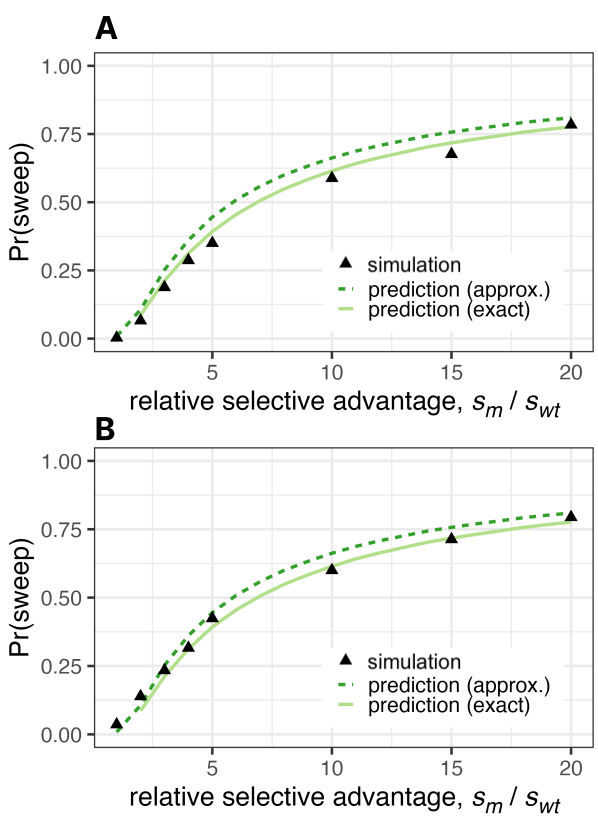

Figure 5. Simulation results of the unconditional sweep probability with mutation rate $\tilde{\mu} = 10^{-6}$ (**A**) and $\tilde{\mu} = 10^{-4}$ (**B**). Parameters other than the mutation rate are as in Figure 4.

Appendix E: Alternative growth models

In our model of expanding population, we considered polynomial growth $N(t) \propto t^d$ with $d = 1, 2, 3$ denoting the spatial dimension. However, our method to calculate the sweep probability works for arbitrary $N(t)$. Other relevant choices for the growth dynamics might be a constant population size, exponential growth or sigmoidal growth. For the case of cancer, several relevant growth models are discussed in [52]. The mathematics in the polynomial growth case are really neat and competition is well defined by Kimura's stepping stone model or the FKPP equation in the continuum limit. However, thinking of the underlying microscopic picture, interpretation of the results can become unclear.

1. Constant population size

For constant population sizes, we immediately know the wildtype radius at the moment the mutant arises and at any other time point. This fixed radius x_0 implies fixed population size $N_0 = \frac{4}{3}\pi x_0^3$. The probability distribution of X can be formally written as the delta function

$$f_X(x) = \delta(x - x_0). \tag{E1}$$

Following the main text, we have the location of the first mutant given by $f_Y(y|X = x)\, dy \propto D(y)a(y)$. Assuming the population is constant from the beginning of the observation, the cell ages are uniformly distributed. We write $a(y) = a_0$, where $a(y)$ is the age of the population at location y at the moment the first mutant arises, and a_0 is constant. The density remains the same $D(y) = 4\pi y^2\, dy$ assuming a spherical population. After considering the boundary and normalization, we have

$$f(y|X = x) = \frac{3y^2}{x^3} 1\{y < x\}. \tag{E2}$$

Since the population radius of the wildtype is fixed, we can drop the condition by replacing x with the constant x_0. This leads to

$$f_Y(y) = f_Y(y|X = x) = \frac{3y^2}{x_0^3} 1\{y < x_0\}. \tag{E3}$$

Next, we calculate the sweep probability conditioned on X and Y. We start by writing the remaining wildtype population once a mutant arose and analogously to the previous calculation introduce the time measure τ. We have

$$\Delta(\tau) = N_{\text{wt}} - N_{\text{int}} = N_0 - N_{\text{int}}, \tag{E4}$$

where the formula for the intersectional space N_{int} has to be split in a similar fashion as in section III C. We

keep the form of eqn. 10 and can adapt also eqn. 11 and 12 by replacing x_{wt} by x_0. The timings in the indicator functions of 12 become $\tau_1 = \frac{x_0 - y}{c_{\mathrm{m}}}$ and $\tau_2 = \frac{x_0 + y}{c_{\mathrm{m}}}$. Using claim 1, we obtain the conditional sweep probability by

$$\Pr(\text{sweep}|Y = y) = e^{\int_0^\infty \Delta(\tau)\, d\tau}. \tag{E5}$$

Then, the unconditional sweep probability can be obtained by the following integral

$$\Pr(\text{sweep}) = \int_0^\infty \Pr(\text{sweep}|Y = y) f_Y(y)\, dy. \tag{E6}$$

The integrals can be evaluated numerically.

For analytical insight, let us assume again that the mutant originates in the center of the wildtype population, $f_Y(y) = \delta(y)$. The integrands simplify and we obtain

$$\Pr(\text{sweep}) = \Pr(\text{sweep}|Y = 0) = e^{-\frac{\mu \pi x_0^4}{c_m}}. \tag{E7}$$

Because the timing of the first mutation plays no role in constant populations, our argumentation for the independence of the mutation rate no longer hold, and we obtain a sweep probability dependent on the mutation rate.

Performing the same analysis in one and two dimensions, we obtain

$$\begin{aligned} \Pr(\text{sweep}) &= e^{-\frac{\mu x_0^2}{2 c_{\mathrm{m}}}} \quad \text{in 1D} \\ \Pr(\text{sweep}) &= e^{-\frac{2\mu \pi x_0^3}{3 c_{\mathrm{m}}}} \quad \text{in 2D} \end{aligned} \tag{E8}$$

a. Comparison with Ralph & Coop: Motivated by parallel adaptation on the species level, Ralph & Coop investigated selective sweeps in constant populations using a very similar approach [46]. However, whereas we assume spherical wildtype populations, Ralph & Coop consider more general shapes. To derive a concrete expression, they ignore boundary effects and obtain "the expected number of other mutations to arise in an area of diameter a in the time it takes the wave to cover that area". Simplifying eqn. (4) in Ref. [46] and keeping their notation, we have

$$\frac{2a^3\lambda}{v} \quad \text{in 1D} \quad \text{and} \quad \frac{\pi a^3 \lambda}{v} \quad \text{in 2D}. \tag{E9}$$

The number of mutations arising in an area is a Poisson process. Thus, we can interpret this number as mean of a Poisson distribution. Then, the probability of a selective sweep given by

$$\begin{aligned} P(k = 0) &= e^{-\frac{2a^2\lambda}{v}} \quad \text{in 1D}, \\ P(k = 0) &= e^{-\frac{\pi a^3 \lambda}{v}} \quad \text{in 2D}. \end{aligned} \tag{E10}$$

We identify v as the speed of the mutant c_{m}, λ as the local mutation rate conditioned on survival $\mu = \tilde{\mu} * \rho$ and a as the maximal travelled distance by the first mutant, which is x_0 in our simplified case. Finally, we find the solution of Ralph & Coop (eqn. E10) to be identical to our solution (eqn. E8) up to multiplication by a constant. The main conclusion is that full selective sweeps are highly unlikely for sufficiently large population radius x_0.

b. Comparison with Martens et al.: Martens and colleagues investigated the likelihood of clonal interference and its impact on the speed of evolution in spatially structured constant-size populations [47], and then studied the implications for understanding cancer initiation [19]. Two modes of evolution are considered: (i) acquisition via subsequent selective sweeps and (ii) acquisition with parallel arising mutations which interfere with each other. To distinguish between these two modes, the authors compare the timescale for a surviving mutation to occur, t_{mut}, and the timescale for a mutation to sweep through the entire constant wildtype population, t_{fix}. Equating these two timescales leads to a critical length of the wildtype population:

$$\begin{aligned} L_c &= \left(\frac{c_0}{2 s_0 \mu}\right)^{\frac{1}{2}} \quad \text{in 1D} \\ L_c &= \left(\frac{c_0}{2 s_0 \mu}\right)^{\frac{1}{3}} \quad \text{in 2D}. \end{aligned} \tag{E11}$$

The authors argue that clonal interference is very likely in populations of size $L \gg L_c$, whereas we should expect selective sweeps when $L \ll L_c$. To compare this result with our model, we put eqn. E11 into eqn. E8. Therefore, we match using $c_0 \leftrightarrow c_{\mathrm{m}}$, $2s\mu \leftrightarrow \rho\tilde{\mu} = \mu$ and $x_0 \leftrightarrow L$ leading to

$$\begin{aligned} \Pr(\text{sweep}) &= e^{-\frac{x_0^2}{2L_c^2}} \quad \text{in 1D} \\ \Pr(\text{sweep}) &= e^{-\frac{\pi x_0^3}{3L_c^3}} \quad \text{in 2D}. \end{aligned} \tag{E12}$$

Indeed, we have $\Pr(\text{sweep}) \to 0$ for $x_0 \gg L_c$ and $\Pr(\text{sweep}) \to 1$ for $x_0 \ll L_c$. We conclude that the result of Martens et al. are in agreement with our result for constant population sizes.

2. Exponential growth

We write the population size as $N(t) = N_0 e^{rt}$, and set $N_0 = 1$. Then applying claim 1, the probability that no mutants occur until time t is given by

$$F(t) = e^{-\mu \int_0^t e^{rt'}\, dt'} = e^{-\frac{\mu}{r}(e^{rt} - 1)}. \tag{E13}$$

Dropping the -1 term, we obtain the same approximate expression as Durrett [53]. We proceed to calculate the probability distribution of T by

$$f(t) = \frac{d(1 - F(t))}{dt} = \mu e^{rt} e^{-\frac{\mu}{r} e^{rt}}, \tag{E14}$$

which agrees with eqn.(25) in Ref. [53]. Instead of asking for the radius, it is more intuitive to relate the population size of the wildtype at the arrival of the first mutant. This way, we not need to assume spherical growth. We write $N_x = e^{rt}$, such that $t = \frac{1}{r}\ln(N_x)$. After substitution, we have

$$f_{N_x}(N_x) = \frac{\mu}{r} e^{-\frac{\mu}{r}N_x}, \tag{E15}$$

which is an exponential distribution. The probability for the radius X can then be calculated by assuming spherical growth starting from one cell, $N_x = \frac{4}{3}\pi x + 1$. For mathematical convenience, we assume $N_x = \frac{4}{3}\pi x$ neglecting the first cell. The probability distribution of X reads

$$f_X(x) = \frac{3x^2}{\theta_{\text{exp}}^3} e^{-\frac{x^3}{\theta_{\text{exp}}^3}}, \tag{E16}$$

with $\theta_{\text{exp}} = \sqrt[3]{\frac{3r}{4\pi\mu}}$.

To derive $f_Y(y)$, we assume spherical growth of the wildtype. Thus, the number of cells at radius y is given by $D(y) = 4\pi y^2\, dy$. The age at radius y at the time the first mutant arises can be obtained by investigating the increase of the radius over time. Starting from radius y, the population grows to radius x in time a. Mathematically, we have $N_x/N_y = e^{ar}$ leading to

$$a = \frac{1}{r}\ln\left(\frac{N_x}{N_y}\right) = \frac{1}{r}\ln\left(\frac{x^3}{y^3}\right). \tag{E17}$$

The conditional probability distribution of Y is proportional to $f_Y(y|X=x)\,dy \propto D(y)a(y)$. After normalization, we have

$$f_Y(y|X=x) = \frac{9y^2}{x^3}\ln\left(\frac{x}{y}\right) 1\{y \le x\}. \tag{E18}$$

To obtain the unconditional probability for y, we can marginalize out X:

$$f_Y(y) = \int_0^\infty f_Y(y|X=x) f_X(x)\, dx. \tag{E19}$$

The integral can be evaluated numerically.

As the mutant grows independently of the wildtype, it can never reach fixation. Bozic and colleagues calculated the probability distribution of the mutant frequency α assuming a stochastic branching process [54]. Setting $\alpha = 1$ in eqn. (7) leads immediately to probability 0, meaning that the mutant can never reach frequency 1. Nevertheless, their formula allows us to consider cases in which the mutant frequency is sufficiently high as to be practically indistinguishable from clonality, given imperfect measurement.

3. Sigmoidal growth

We calculate the probability distribution for the time the first mutant arises and outline the calculation for the position it occurs assuming spherical expansion. Among the various sigmoidal growth laws (e.g. logistic, Gompertzian, and von Bertalanffy growth [52]), we pick logistic growth as representative. Starting with a single cell, the population growth over time can be written as $N(t) = \frac{K}{1+Ke^{-rt}}$. We apply claim 1, and obtain the probability no mutation occur before time t,

$$F(t) = e^{-\frac{\mu K}{r}\ln\left(\frac{e^{rt}+K}{K+1}\right)}. \tag{E20}$$

We proceed to calculate the probability distribution of the time T the first mutant occurs,

$$f_T(t) = \frac{\mu K (K+1)^{\frac{\mu K}{r}} e^{rt}}{(e^{rt}+K)^{\frac{\mu K}{r}+1}}. \tag{E21}$$

The probability distribution for the population size $f_{N_x}(N_x)$ at the time the first mutant can be obtained by substituting $N_x = \frac{K}{1+Ke^{-rt}}$ yielding

$$f_{N_x}(N_x) = \frac{\mu\,(K+1)^{\frac{\mu K}{r}}\,(K-N_x)^{\frac{\mu K}{r}-1}}{rK^{\frac{2\mu K}{r}-1}}. \tag{E22}$$

Assuming the spherical growth the probability distribution $f_X(x)$ for the radius can be obtained by substituting $N_x = \frac{4}{3}\pi x^3$ where we again neglect the first cell. We obtain

$$f_X(x) = \frac{\mu\,(K+1)^{\frac{\mu K}{r}}\left(K - \frac{4\pi x^3}{3}\right)^{\frac{\mu K}{r}-1} 4\pi x^2}{rK^{\frac{2\mu K}{r}-1}}. \tag{E23}$$

To obtain the location at which the first mutant occurs, we need the age distribution dependent on the radius $a(y)$. In logistic growth, we compute this to be

$$a(y) = \frac{1}{r}\ln\left(\left(\frac{K-N_x}{N_x}\right)\left(\frac{N_y}{K-N_y}\right)\right).$$

Together with the density of cells at radius y of the sphere, $D(y) = 4\pi y^2 dy$, we can compute the location of y conditioned on X as

$$f_Y(y|X=x) = \frac{3y^2}{x_M^3}\frac{\ln\left[\left(\frac{x_M^3-x^3}{x^3}\right)\left(\frac{y^3}{x_M^3-y^3}\right)\right]}{\ln\left(\frac{x_M^3-x^3}{x_M^3}\right)} 1\{y \le x\}, \tag{E24}$$

where we define x_M to be the radius at which the population reaches the carrying capacity K determined by $K = \frac{4\pi x_M^3}{3}$.

To obtain the unconditional probability for y, we can marginalize out X,

$$f_Y(y) = \int_0^\infty f_Y(y|X=x) f_X(x)\, dx. \tag{E25}$$

The integral can be evaluated numerically.

Whereas sigmoidal growth functions include competition, the interpretation of spatial competition is still ambiguous and goes beyond the scope of this manuscript.

4. Boundary driven growth without proliferation in the interior

Antal and colleagues consider boundary-driven growth [20], which has been studied extensively [10]. Whereas we assume turnover of the entire population, Antal et al. consider turnover only at the boundary and focus on the three-dimensional case. A full selective sweep is impossible in this model since individuals located away from the boundary neither proliferate nor die. Instead, we can consider the probability that a mutant will envelop the wildtype and thus become the only proliferating population. Using this interpretation, eqn. (17) in Ref. [20] provides us the unconditional sweep probability,

$$P(N=1) = \frac{9+\beta^2}{\beta^2} \frac{2}{1+e^{3\pi/\beta}}, \tag{E26}$$

with $\beta = \sqrt{v^2-1}$ where we identify $v = \frac{c_\mathrm{m}}{c_\mathrm{wt}}$. The unconditional sweep probability is independent of the mutation rate, as in our case described in Section III D. Furthermore, Antal et al. find similar expressions for the arrival time of the first mutant $f_T(t)$ (eqn. 15 in [20]) and the conditional sweep probability $\Pr(\text{sweep}|X=x)$ (eqn. 16 in [20]) and the size of the wildtype population when the first mutant occurs (eqn. 25 in [20] provides the cumulative density function).

5. Mixed models

There are two ways to mix models. First, we can assume that the wildtype grows differently from the mutant population. We have done this in section E 1 assuming that the wildtype is constant and the mutant radius propagates with a constant radius. Second, the growth of the population can be described by two different formulas applied for different time intervals. For example, the exponential-linear model describes exponential growth that goes over into linear growth [52].

As the number of models increases rapidly, we focus on the discussion of the basic models. In general, our methods always lead to the formulation of the integrals for the sweep probability as long as we have the population growth as a function of time and other constant parameters of the wildtype N_wt and the mutant N_m. We therefore need to assume spherical growth of the wildtype and the mutant. Solving the integrals can become complicated, and might need to be performed numerically. Furthermore, the interpretation of the results might be questionable.

Appendix F: Numerical integration

We computed numerical integrals using the MATLAB function 'trapz', with interval widths and integration ranges tailored according to the nature of $f_X(x)$. For values of x close to 0 and beyond 2θ, $f_X(x)$ is negligibly small and makes virtually no contribution to the integral. Hence we set the lower and upper bound of integration to 0.01θ and 3θ, respectively. We set interval widths to 0.01θ. We used analagous values to calculate $\Pr(\text{sweep}|X=x)$.

Appendix G: Agent-based simulations

Individuals in our agent-based model are subdivided into well-mixed demes on a regular two-dimensional grid. The demes have identical carrying capacities and are initially filled with residents, except that a single wildtype invader is introduced at the centre of the grid. At each time step, an individual is chosen at random to be replaced by two offspring, with probabilities weighted by fitness. Each offspring then either migrates, with probability m, to a neighbouring deme in a randomly chosen direction, or remains in its parent deme. Local density dependence is implemented by imposing a relatively very high death rate whenever a deme is above carrying capacity. Mutation is coupled to wildtype reproduction. To improve computational efficiency, resident individuals are not permitted to disperse; we verified with additional simulations (to be published in a later study) that this asymmetry has only a minor effect on the wildtype expansion speed. Further model details have been published previously [27, 39].

Appendix H: Speed selection correspondence

We make a correspondence of propagation speed and selective advantage. We set the migration rate to $m = 0.05$ and the deme size to $K = 16$. Normal cells have birth rate $r_\mathrm{n} = 0.909090$, wildtype cells have birth rate $r_\mathrm{wt} = 1.0$ such that $s_\mathrm{wt} = \frac{r_\mathrm{wt}-r_\mathrm{n}}{r_\mathrm{n}} = 0.1$. For the mutant, we vary the birth rate r_m from 1.1 to 3.0 giving us $s_\mathrm{wt} = \frac{r_\mathrm{wt}-r_\mathrm{n}}{r_\mathrm{n}}$ ranging from 0.1 to 2.0. The speed is measured 10 times and is then averaged. The standard deviation was consistently below 2% of the mean and is neglected.

The wildtype speed is $c_{\mathrm{wt}} = 0.15$. The mutant speeds are presented in table I. The wildtype propagates faster than the mutant when the selective advantage is the same $s_{\mathrm{wt}} = s_{\mathrm{m}} = 0.100$. The reason for this is that normal cells cannot migrate whereas wildtype cells can, which in return inhibits the growth of the mutants.

s_{m}	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90	1.00
c_{m}	0.14	0.23	0.31	0.38	0.46	0.53	0.60	0.68	0.75	0.82
s_{m}	1.10	1.20	1.30	1.40	1.50	1.60	1.70	1.80	1.90	2.00
c_{m}	0.89	0.96	1.03	1.11	1.17	1.24	1.31	1.39	1.46	1.52

Table I. Correspondence of selective advantage s_{m} to the propagation speed c_{m} for the mutant population measured in our simulations. We fixed the migration rate $m = 0.05$, the deme size $K = 16$ and fitness of the wildtype $r_{\mathrm{wt}} = 1.0$.

	1D	2D	3D
$f_X(x)$	$\frac{2x}{\theta_{1D}^2} e^{-\frac{x^2}{\theta_{1D}^2}}$	$\frac{3x^2}{\theta_{2D}^3} e^{-\frac{x^3}{\theta_{2D}^3}}$	$\frac{4x^3}{\theta^4} e^{-\frac{x^4}{\theta^4}}$
$f_Y(y\|X=x)$	$\frac{2(x-y)}{x^2} 1\{y \leq x\}$	$\frac{6y(x-y)}{x^3} 1\{y \leq x\}$	$\frac{12y^2(x-y)}{x^4} 1\{y \leq x\}$
$f_Y(y)$	$\frac{2}{\theta_{1D}} \Gamma\left(\frac{1}{2}, \frac{y^2}{\theta_{1D}^2}\right) - \frac{2y}{\theta_{1D}^2} \Gamma\left(0, \frac{y^2}{\theta_{1D}^2}\right)$	$\frac{3}{2\theta_{2D}^3} \Bigg(3\theta_{2D}^2 \Gamma\left(\frac{5}{3}, \frac{y^3}{\theta_{2D}^3}\right) - 12\theta_{2D} y \Gamma\left(\frac{4}{3}, \frac{y^3}{\theta_{2D}^3}\right) + 9y^2 e^{-\frac{y^3}{\theta_{2D}^3}} + 2y^2 \Gamma\left(0, \frac{y^3}{\theta_{2D}^3}\right) \Bigg)$	$\frac{12y^2}{\theta^4} \left(\theta \Gamma\left(\frac{1}{4}, \frac{y^4}{\theta^4}\right) - y \Gamma\left(0, \frac{y^4}{\theta^4}\right) \right)$
$\Pr(\text{sweep}\|X=x, Y=y)$	$e^{\frac{x^2+y^2}{\alpha_{1D}^2}}$	numerical evaluation	numerical evaluation
$\Pr(\text{sweep}\|X=x)$	$\frac{\sqrt{\pi}\nu x \text{Erf}(\nu x) - 1 + e^{-2\nu^2 x^2}}{\nu^2 x^2}$	numerical evaluation	numerical evaluation
$\Pr(\text{sweep}\|X=x, Y=0)$	$e^{-\left(\frac{x}{\alpha_{1D}}\right)^2}$	$e^{-\left(\frac{x}{\alpha_{2D}}\right)^3}$	$e^{-\left(\frac{x}{\alpha}\right)^4}$
$\Pr(\text{sweep})$ [approx.]	$\frac{c_m - c_{wt}}{c_m}$	$\left(\frac{c_m - c_{wt}}{c_m}\right)^2$	$\left(\frac{c_m - c_{wt}}{c_m}\right)^3$
$\Pr(\text{sweep})$ [exact]	$\beta' \left[\frac{2\cot^{-1}\left(\sqrt{1+\beta'}\right)}{\sqrt{1+\beta'}} + \ln\left(\frac{\beta'+1}{\beta'+2}\right) \right]$	numerical evaluation	numerical evaluation
$f_X(X=x\|\text{sweep})$ [approx.]	$\frac{2x}{\theta_{1D}^2 \beta} e^{-\frac{x^2}{\theta_{1D}^2 \beta}}$	$\frac{3x^2}{\theta_{2D}^3 \beta^2} e^{-\frac{x^3}{\theta_{2D}^3 \beta^2}}$	$\frac{4x^3}{\theta^4 \beta^3} e^{-\frac{x^4}{\theta^4 \beta^3}}$
$f_X(X=x\|\text{sweep})$ [exact]	complicated expression	numerical evaluation	numerical evaluation

Table II. Summary of analytical results in 1D, 2D and 3D.

	1D	2D	3D
θ	$\sqrt{\frac{c_{wt}}{\mu}}$	$\sqrt[3]{\frac{3c_{wt}}{\pi\mu}}$	$\sqrt[4]{\frac{3c_{wt}}{\pi\mu}}$
α	$\sqrt{\frac{c_m - c_{wt}}{\mu}}$	$\sqrt[3]{\frac{3(c_m - c_{wt})^2}{\pi\mu(2c_m - c_{wt})}}$	$\sqrt[4]{\frac{3(c_m - c_{wt})^3}{\pi\mu\left(c_{wt}^2 - 3c_{wt}c_m + 3c_m^2\right)}}$
β	$\frac{c_m - c_{wt}}{c_m}$		
β'	$\frac{c_m - c_{wt}}{c_{wt}}$		

Table III. Summary of compound parameters in 1D, 2D and 3D.