

# Genealogies in Growing Solid Tumors

Rick Durrett, Duke U.

January 5, 2018

## Abstract

Over the past two decades, the theory of tumor evolution has largely focused on the selective sweeps model. According to this theory, tumors evolve by a succession of clonal expansions that are initiated by driver mutations that have a fitness advantage over the resident types. A 2015 study of colon cancer [44] has suggested an alternative theory of tumor evolution, the so-called Big Bang model, in which all of the necessary driver mutations are acquired before expansion began, and the evolutionary dynamics within the expanding population are predominantly neutral. In this paper, we will describe a simple mathematical model inspired by work of Hallatschek and Nelson [25] that makes quantitative predictions about spatial patterns of genetic variability. While this model has some success in matching observed patterns in two dimensions, it fails miserably in three dimensions. Despite this failure, we think the model analyzed here will be a useful first step in building an accurate model.

## 1 Introduction

In the 1950s Fisher and Holloman [20] and Nordling [41] found that within the age range 25–74, the logarithm of the cancer death rate increased in direct proportion to the logarithm of the age, with a slope of about 6 on a log-log plot. Nordling grouped all types of cancer together and considered only men, but the pattern persisted when Armitage and Doll [1] separated cancers by their type and considered men and women separately. Nordling [41] suggested that the slope of six on a log-log plot would be explained if a cancer cell was the end result of seven successive mutations.

In the studies cited above, the stages were unspecified events. That changed in 1971 with Knudson’s study of retinoblastoma [31]. Based on observations of 48 cases of retinoblastoma and published reports, he hypothesized that the disease is a cancer caused by two mutational events. He based this on the observation that the incidence time was exponential for the patients with multiple bilateral tumors, but looked like a  $\text{gamma}(2, \lambda)$  distribution for the less serious unilateral cases. He hypothesized that the more serious patients had a germline mutation in the underlying gene while the less serious patients needed to have mutations occur in both copies of the underlying gene. In current terminology the gene, later identified as *RB1* is a tumor suppressor gene. Trouble begins when both copies are knocked out. For more on the development of these ideas see [32], which was written to mark the 30th anniversary of the original paper.

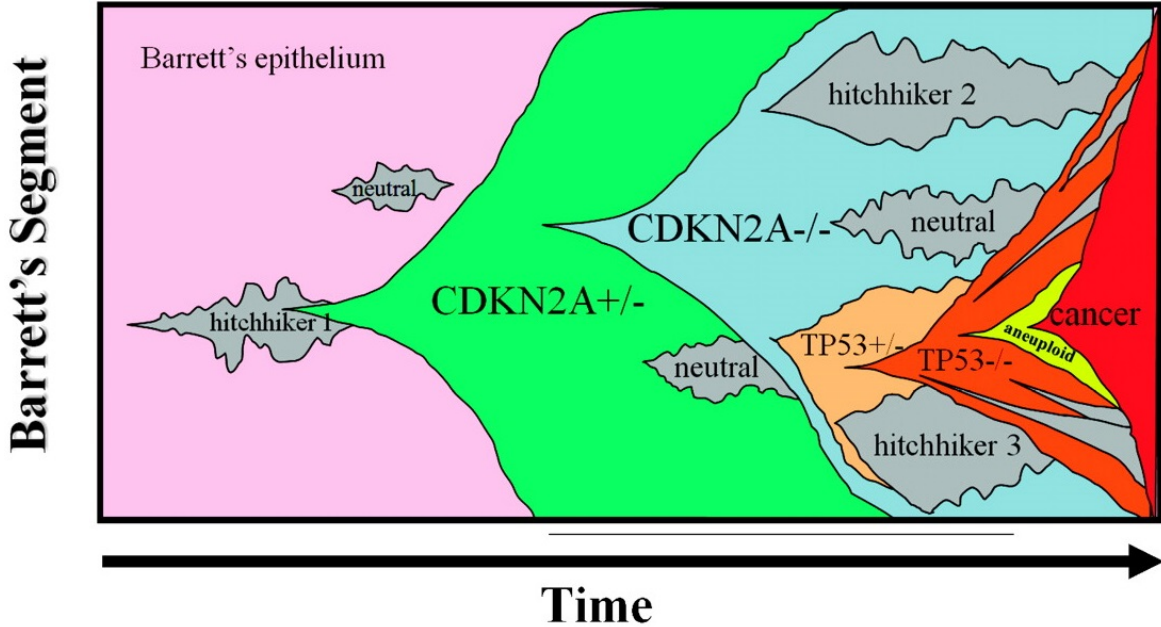


Figure 1: Clonal expansion in Barrett's Esophagus. Figure is from [36].

In colon cancer the initiating event is thought to involve the inactivation of the tumor suppressor gene APC (adenomatous polyposis coli). In 1990, Fearon and Vogelstein [18] found a second piece of the puzzle when they noted that approximately 50% of colorectal carcinomas, and a similar percentage of adenomas greater than 1 cm have mutations in the RAS gene family, while only 10% of adenomas smaller than 1 cm have these mutations. In the modern terminology, the members of the RAS family are *oncogenes*. A mutation in a single copy is sufficient to allow progression. The analysis in [18] also suggested a role for TP53 (which produces the tumor protein p53) in the progression to cancer. The protein p53 has been described as "the guardian of the genome" because of its role in conserving stability by preventing genome mutations. Mutations in the gene TP53 have since been implicated in many cancers, see [21] and [48]. Combining these ideas leads to a four (or five) stage description for colon cancer that is described for example in the books of Vogelstein and Kinzel [47], and Frank [19].

In the multistage theory of carcinogenesis, it is thought that the sequence of "driver" mutations produces a series of selective sweeps. This theory has been confirmed by whole genome sequencing of cancer cells. Ding et al [10] have identified the clonal structure of eight relapsed acute myeloid leukemia (AML) patients. In one patient the founding clone 1 accounted for 12.74% of the tumor at the time of diagnosis. The additional mutations in clones 2 and 3 may have resulted in growth or survival advantages because they were 53.12% and 29.04% of the tumor respectively. Only 5.10% of the cells were in clone 4 indicating that it may have arisen last. However, the relapse evolved from clone 4 with the resultant clone 5 having 78 new somatic mutations compared to the sampling at day 170.

A mathematical theory has been developed for the clonal expansion model to make predictions about the level of intratumor heterogeneity [4, 15], the number of passenger

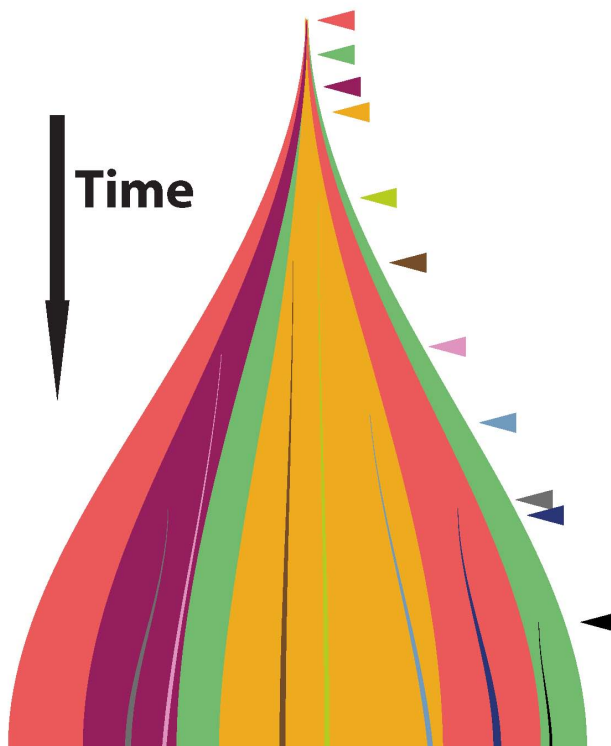


Figure 2: Cartoon picture of the Big Bang theory. Arrows mark the times of mutations. Colors indicate their spatial extent.

(neutral) mutations in cancer cells [2, 46], and more practical questions such the evolution of resistance to treatment [26, 35, 45], the effectiveness of combination therapy [3, 29, 38], and the potential effectiveness of screening to reduce ovarian cancer [9]. See [13] for an introduction to the mathematics underlying many of these applications.

The picture of successive selective sweeps has been confirmed by comparing primary tumors and their metastases [50], and by regional sequencing of breast cancer [39], glioblastoma [43], and renal carcinoma [22]. Thus it was surprising when Sottoriva et al introduced and validated a ‘big Bang’ model in which all driver mutations were present at the time of tumor initiation. They collected genetic data of various types from 349 individual tumor glands were sampled from the opposite sides of 15 colorectal tumors and large adenomas. Data presented in Figure 3 of their paper shows that adenomas were characterized by mutations and copy number aberrations (CNA) that segregated between tumor sides. In contrast the majority of carcinomas exhibited the same private CNA in individual glands from different sides of the tumor. Follow up work of Ryser et al [42] found evidence of early abnormal cell movement in 8 of 15 invasive colorectal carcinomas (“born to be bad”) but not in four benign adenomas

The mutation patterns found by Sottoriva et al. [44] are similar to those found by Halatschek et al [23] in a remarkable experimental paper. Two fluorescently labeled strains of *E. coli* were mixed and placed at the center of an agar plate containing a rich growth medium. The central region of the plate exhibits a dense speckled pattern reminiscent of

the initial mixed population. From this ring toward the boundary of the colony, the population segregates into single colored sectors with boundaries that fluctuate. [25] and [30] have developed and analyzed models for the development of sectors in the system.

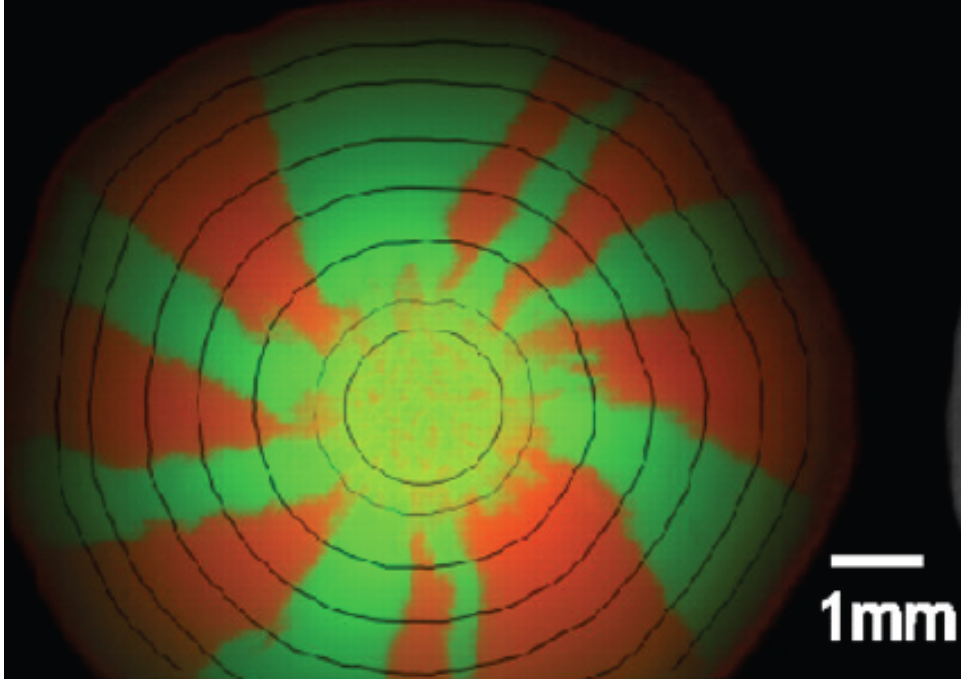


Figure 3: Yeast growth patterns.

## 2 Biased Voter Model

A natural first step is to investigate genealogies in the biased voter model, which was introduced by Williams and Bjerknes [49] as a cancer model in 1972. In contrast to Durrett, Foo and Leder [16] there will be only two types of cells: 0 = wild type, 1 = cancer cells, and there are only neutral mutations, i.e., no type 2's will be created, and there are no new mutations from type 0 to type 1.

In the biased voter model, 0's give birth at rate 1, and 1's at rate  $\lambda$ . In either case the new individual is sent to a randomly chosen nearest neighbor on  $\mathbb{Z}^d$ . Since we are concerned with solid tumors we will be primarily concerned with  $d = 2, 3$ . The key to analyzing the biased voter model is its duality with a branching coalescing random walk  $\zeta_t$ , see [6, 5]. To explain the duality we have to construct the process. For each  $x \in \mathbb{Z}^d$  and nearest neighbor  $y$  we have a Poisson process  $T_n^{x,y}$  with rate  $1/2d$ , and a Poisson process  $S_n^{x,y}$  with rate  $(\lambda - 1)/2d$ . At times  $T_n^{x,y}$  we put a  $\delta$  at  $x$  and draw an arrow from  $y$  to  $x$ . The arrow reaches the point  $x$  just above the  $\delta$ . At times  $S_n^{x,y}$  we draw an arrow from  $y$  to  $x$ . We think of  $\delta$ 's as dams that stop the flow of fluid, and arrows that spread the fluid in the direction of the arrow. If we inject fluid at the bottom at the sites of  $A$  and allow the fluid to only flow up, being blocked by  $\delta$ 's and flowing across arrows in the direction of their orientation the sites reached by fluid at time  $t$  are  $\xi_t^A$ . See Figure 4 for an example.

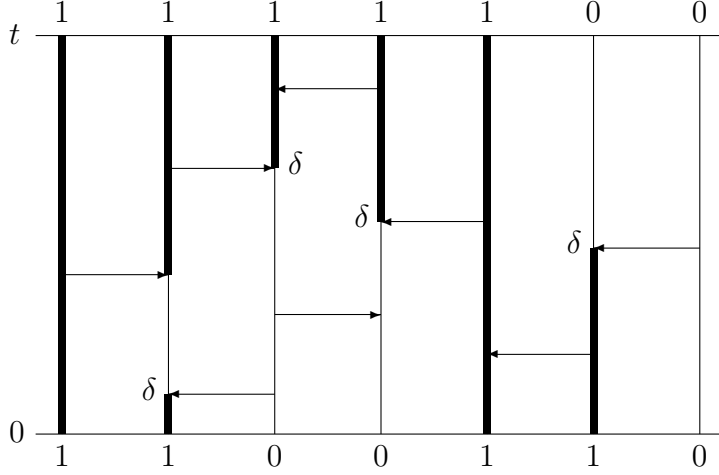


Figure 4: Construction of the biased voter model. Thicklines show the flow of fluid and indicate the  $(x, t)$  with  $\xi_t(x) = 1$ .

To define the dual  $\zeta_s^{x,t}$ ,  $0 \leq s \leq t$  we start at  $x$  at time  $t$ . Fluid flows down, is blocked by  $\delta$ 's and moves across the arrows in the opposite direction to their orientation. From the definition it should be clear that  $x \in \xi_t$  if and only if some site in  $\zeta_t^{x,t}$  is occupied at time 0. That is,

$$\{x \in \xi_t^A\} = \{\zeta_t^{x,t} \cap A \neq \emptyset\}$$

To define the dual starting from a set of sites  $B$  let  $\zeta_s^{B,t} = \cup_{x \in B} \zeta_s^{x,t}$ . It is immediate that

$$\{\xi_t^A \cap B \neq \emptyset\} = \{\zeta_t^{B,t} \cap A \neq \emptyset\}$$

From the construction it is easy to see in the dual, particles jump to each nearest neighbor at rate  $1/2d$ , give birth onto each nearest neighbor at rate  $(\lambda - 1)/2d$ . This system is called a coalescing branching random walk

$\zeta_t^{x,t}$  gives the set of potential ancestors of the individual at  $x$  at time  $t$ . To find the actual ancestor, we could follow the approach of the *ancestral selection graph* and starting from the values of  $\xi_0$  on  $\zeta_t^{x,t}$  and work our way back up the graphical representations to see which arrows produced births. Here we will follow the approach in Neuhauser's work on competing contact processes [40] and inductively define an ordering of the points in the dual  $\zeta_s^{x,t}$  as we work backwards. The actual ancestor will be the first occupied site in the list. Looking at the example drawn in Figure 5 and working backwards from time  $t$ , the arrow from 0 to  $-1$  will be an actual birth only if at that point  $-1$  is vacant and 0 is occupied. To encode this we write the dual as  $-1, 0$ . The next two arrows as we work down is a voter arrow so the dual changes to  $-2, 0$  and then to  $-2, 1$ . The arrow from  $-3$  to  $-2$  will be an actual birth only if at that point  $-2$  is vacant and  $-3$  is occupied, but if this is the case  $-3$  will be the actual ancestor so the dual is now  $-2, -3, 1$ . Taking into account the last two arrows we see that the actual ancestor is the first occupied site on the list  $-1, -3, 1, 2$ .

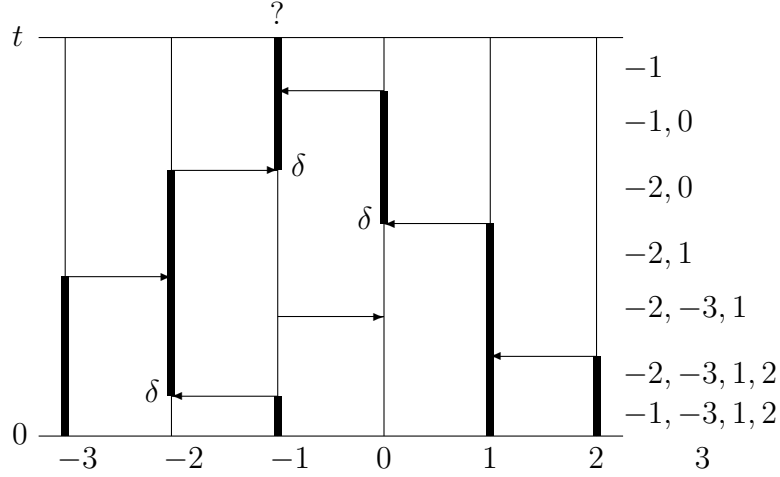


Figure 5: The dual  $\zeta_t^{-1,t} = \{-3, -1, 1, 2\}$  so  $-1$  will be occupied at time  $t$  if one of these sites is.

Bramson and Griffeath [6, 5] showed that if we start with a single type 1 at the origin then when  $\xi_t^0$  does not die out, it grows linearly and has an asymptotic shape  $D$ . That is, for any  $\epsilon > 0$ , there is a  $t_\epsilon$  (which depends on the outcome  $\omega$ ) so that on  $\{T_0 = \infty\}$  we have

$$(1 - \epsilon)tD \cap \mathbb{Z}^d \subset \xi_t \subset (1 + \epsilon)tD \quad \text{for } t \geq t_\epsilon(\omega). \quad (1)$$

$D$  is convex and has the same symmetries as those of  $\mathbb{Z}^d$  that leave the origin fixed, e.g., rotation by 90 degrees around an axis, or reflection through a hyperplane through the origin perpendicular to an axis.

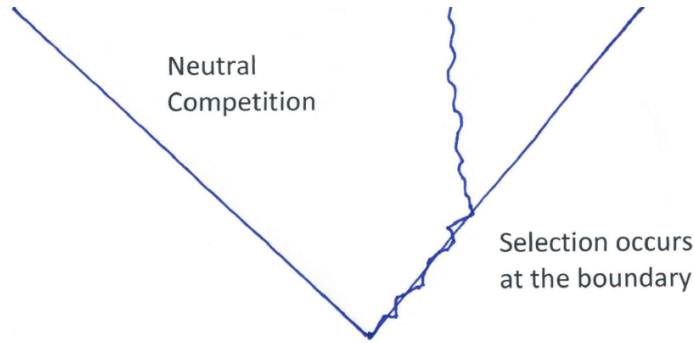


Figure 6: Behavior of the ancestral lineage of a cell in the biased voter model.

If we select an  $x$  well away from the boundary at time  $t$  then initially all branching arrows connect two occupied sites. In this case they will not be part of the actual ancestral lineage, so it consists entirely of voter arrows. Individual trajectories will move like a mean zero random walk, and hence travel  $O(\sqrt{s})$  in time  $s$ . As Figure 6 shows, this means that

the lineage will go almost straight down until it comes to the edge of the space time cone  $\Gamma = \{(x, t) : x \in tD\}$ . At this point the lineage will begin to contain branching events that will take it back toward the origin. It is far from obvious what the path is doing when it is near the boundary but it must get to the origin since  $x \in \xi_t$  implies  $0 \in \zeta_t^{x,t}$ .

### 3 A simplified model

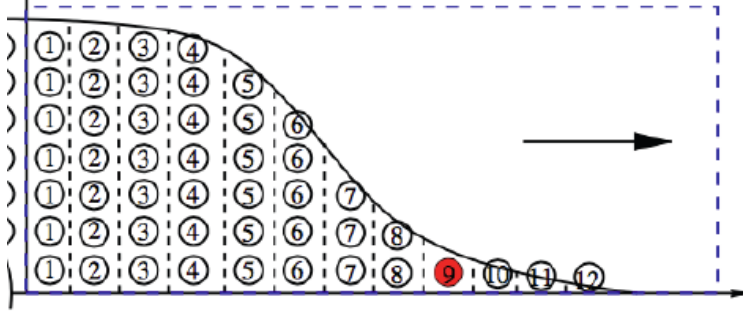


Figure 7: A one dimensional model with many individuals per site (or deme).

Hallatschek and Nelson [24] studied genealogies in a one dimensional system in which sites (demes) can carry up to  $N$  individuals. Here we will follow the formulation of Durrett and Fan [14]. There is one cell at each point of  $(L_n^{-1}\mathbb{Z}) \times \{1, \dots, M_n\}$ , whose cell-type is either 1 or 0. The cells in deme  $w \in L_n^{-1}\mathbb{Z}$  only interact with those in demes  $w - L_n^{-1}$  and  $w + L_n^{-1}$  and interact equally with all of those points. Hence each cell  $x = (w, i)$  has  $2M_n$  neighbors. Type-0 cells reproduce at rate  $2M_n r_n$ , type-1 cells at rate  $2M_n(r_n + \theta R_n^{-1})$ . When reproduction occurs the offspring replaces a neighbor chosen uniformly at random. We define the approximate density by

$$u_t^n(w) := \frac{1}{M_n} \sum_{i=1}^{M_n} \xi_t(w, i)$$

and linearly interpolate between demes to define  $u_t^n(w)$  for all  $w \in \mathbb{R}$ .

let  $C_b(\mathbb{R})$  be the set of bounded continuous functions equipped with the metric

$$\|f\| = \sum_{k=1}^{\infty} \sup_{|x| \leq k} |f(x)|$$

which induces the topology of uniform convergence on compact sets.

**Theorem 1.** *Suppose that as  $n \rightarrow \infty$ , the initial condition  $u_0^n$  converges in  $C_b(\mathbb{R})$  to  $f_0$  and that:*

- (a)  $r_n M_n / L_n^2 \rightarrow \alpha \in (0, \infty)$
- (b)  $r_n / L_n \rightarrow \gamma \in [0, \infty)$
- (c)  $M_n / R_n \rightarrow \beta \in [0, \infty)$

(d)  $L_n \rightarrow \infty$  and  $L_n R_n \rightarrow \infty$

Then the approximate density process  $(u_t^n)_{t \geq 0}$  converges in distribution in  $D([0, \infty), C_b(\mathbb{R}))$ , as  $n \rightarrow \infty$ , to a continuous  $C_b(\mathbb{R})$  valued process  $(u_t)_{t \geq 0}$  which is the weak solution to the stochastic partial differential equation (SPDE)

$$\partial_t u = \alpha \Delta u + 2\theta \beta u(1 - u) + |4\gamma u(1 - u)|^{1/2} \dot{W} \quad (2)$$

with initial condition  $u_0 = f_0$ . Here  $\dot{W}$  is the space-time white noise on  $[0, \infty) \times \mathbb{R}$ .

The absolute value is to make the coefficient well defined when  $u < 0$ . When  $\theta > 0$  the solutions will stay nonnegative. For related results see the earlier work of Muller and Tribe [37] and Doering, Mueller, and Smereka [11] on limits of long range voter models

To explain the terms in the limit

- $\alpha \partial_x^2 u(x, t)$  = diffusion generated by the voter model component of the dynamics
- $2\theta \beta u(1 - u)$  = increase in 1's due to births from 1's onto 0's at rate  $2M_n \theta / R_n$
- $2\sqrt{\gamma u(1 - u)} \dot{W}$  fluctuation due to random reproduction

Hallatschek and Nelson [24] viewed the system in a reference frame moving at rate  $v$  and made different parameter choices, so they ended up with the solution of the stochastic PDE

$$\partial_t u(x, t) = D \partial_x^2 u(x, t) + v \partial_x u(x, t) + su(1 - u) + \epsilon \sqrt{u(1 - u)} \eta \quad (3)$$

where  $\eta$  is space-time white noise, see their equation (2). The term  $v \partial_x u(x, t)$  arises due to the moving frame of reference.

Using arguments about the behavior of tracer particles placed into an expanding fluid the authors of [24] argue that the probability density  $G(y, t|x, T)$  that an individual at  $x$  at time  $T$  was descended from an ancestor that lived at time  $y$  at time  $t$  satisfies

$$\partial_\tau G(y, t|x, T) = D \partial_y^2 G - \partial_y [(v + 2D \partial_y \ln(u(y, t))) G] \quad (4)$$

Recalling that a diffusion process with generator  $L = a(x)f''(x) + b(x)f'(x)$  has a transition probability that satisfies

$$\partial_t p_t(x, y) = \partial_y^2 (a(y)p_t(x, y)) - \partial_y (b(y)p_t(x, y))$$

and comparing with (4), we see that in a fixed reference frame the coordinates of the ancestor will be a diffusion process with generator

$$Lf(y) = \frac{1}{2} f''(y) + \frac{d}{dy} \ln(u(y, \tau)) \cdot f'(y) \quad (5)$$

where  $\tau = T - t$  since we are working backwards in time starting from  $x$  at time  $T$ . To see this is reasonable note that if the interface was as it is drawn in Figure 7 then in the flat part of the curve there would be no drift and the genealogy would be Brownian motion, in

agreement with the heuristic picture in Figure 6. In the right side of the picture where  $u$  is decreasing, there will be a drift to the left, which will keep the lineage close to the boundary.

If  $u$  is the solution to (3), then it is not smooth enough for the drift  $(d/dy) \ln[u(y, t)]$  in (4) to make sense mathematically. Even worse, in two dimensions there is no analogue of Theorem 3, since in dimensions  $d > 1$  SPDE do not have function-valued solutions. To avoid these problems we could let  $\bar{P}$  denote the probability measure for the biased voter model conditioned not to die out and switching to fixed reference frame replace the solution of (3) by

$$u(x, t) = \bar{P}(\xi_t(x) = 1).$$

Since we do not know much about the right-hand side we will go one step further and let

$$u(x, t) = \exp[\varphi(|x| - vt)/t^\alpha] \quad (6)$$

where  $\varphi(z)$  converges to 0 as  $z \rightarrow -\infty$  and  $\rightarrow -\infty$  as  $z \rightarrow \infty$ . Here we have replaced the limiting shape in (1) by a ball and introduced  $t^\alpha$  as a measure of the fluctuations of the boundary. Thinking of the central limit theorem a natural choice is  $\alpha = 1/2$ . However, physicists' arguments [28] and simulations of the Eden model (which has births and no deaths) [51, 52, 34] suggest that in two dimensions the fluctuations of the boundary will be of order  $t^{1/3}$  rather than the usual  $t^{1/2}$ . It is not known what the fluctuations should be in  $d > 2$ , so to accommodate a range of scalings we will allow  $\alpha$  to vary and see how the behavior changes.

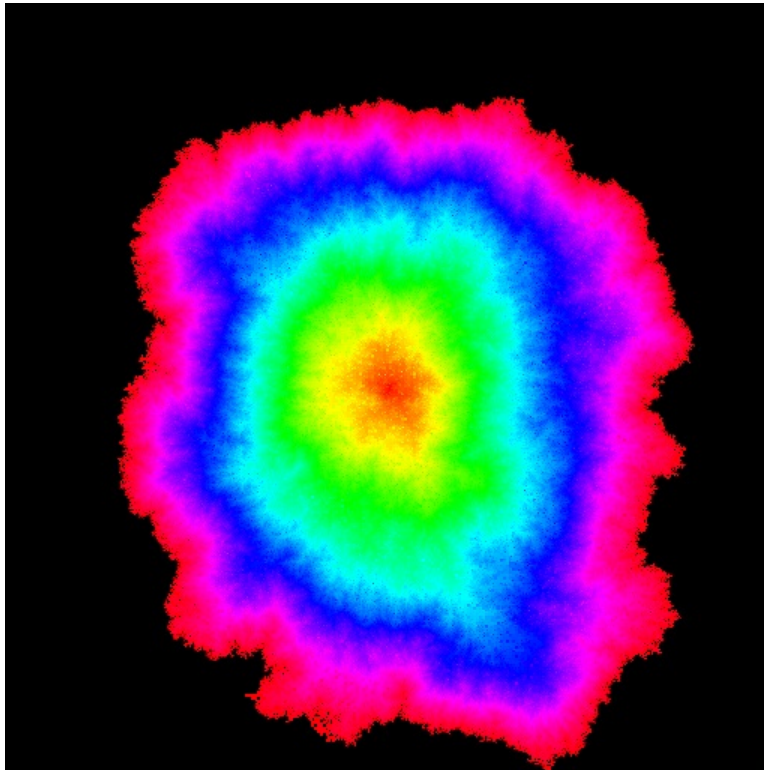


Figure 8: The Eden model has a very rough boundary.

Using the reasoning that led to (5), the process in  $d > 1$  is:

$$Lf(y) = \frac{1}{2}\Delta f(y) + \nabla \ln(u(y, \tau)) \cdot f'(y) \quad (7)$$

where again  $\tau = T - t$ . Suppose, for the moment, that we are in two dimensions. Our next step is to change to polar coordinates. If we do this to Brownian motion (and add a superscript 0 to indicate that we are transforming Brownian motion) the result has

$$dR_t^0 = dB_t^1 + \frac{1}{R_t^0} ds \quad d\theta_t^0 = dB_t^2/R_t^0$$

where the  $B_s^i$  are independent Brownian motions. See [12] for this and other stochastic calculus facts that we use.

Since the drift in (7) is radial, in polar coordinates its angular part has

$$d\theta_t = dB_t^2/R_t.$$

The radial component

$$dR_t = dB_t^1 + \left( \frac{1}{R_t} + \tau^{-\alpha} \phi'((R_t - v\tau)/\tau^\alpha) \right) dt,$$

where  $\psi = \varphi'/\varphi$ . Writing  $U_t = R_t - v(T - t)$  to return to the moving frame of reference, and dropping the first term which is small

$$dU_t = dB_t^1 + (v + \tau^{-\alpha} \phi'(U_t/\tau^\alpha)) dt.$$

If we let  $\psi(u) = vu + \phi(u/\tau^\alpha)$  then we can write the last equation as

$$dU_t = dB_t^1 + \psi'(U_t) ds. \quad (8)$$

To understand the behavior of this process we note that its generator can be written as

$$L_U f(u) = \frac{1}{2} e^{-2\psi(u)} \partial_u (e^{2\psi(u)} \partial_u f),$$

so if we let  $\langle f, g \rangle = \int f(u)g(u)e^{2\psi(u)} du$  then

$$\langle g, L_U f \rangle = \langle L_U g, f \rangle.$$

That is,

**Lemma 1.**  *$L_U$  is self-adjoint with respect to  $e^{2\psi}$ .*

To have a concrete example we let

$$\phi(u) = \begin{cases} 0 & u \leq 0 \\ -u^2/2 & u \geq 0 \end{cases}$$

so that

$$2\psi(u) = \begin{cases} 2vu & u \leq 0 \\ 2vu - u^2/\tau^{2\alpha} & u \geq 0 \end{cases}$$

and the drift

$$b(u) = \begin{cases} v & u \leq 0 \\ v - x/\tau^{2\alpha} & u \geq 0 \end{cases}$$

To make  $e^{2\psi(u)}$  a probability measure we have to normalize. To do this, we first suppose that the second half of the definition applies on the whole line and write

$$2\psi(u) = 2vu - \frac{u^2}{\tau^{2\alpha}} = -\left(\frac{u}{\tau^\alpha} - v\tau^\alpha\right)^2 + (v\tau^\alpha)^2$$

from this we see that

**Lemma 2.** *When  $e^{2\psi(u)}$  is normalized to be a probability measure it is a normal with mean  $v\tau^{2\alpha}$  and variance  $\tau^{2\alpha}/2$ .*

From this, we see that it is legitimate to ignore the contribution from  $u \leq 0$ . This result shows us that it is not sensible to take  $\alpha > 1/2$  since in that case the distance of the lineage from the tumor is much larger than the size of the tumor. The case  $\alpha = 1/2$  is also somewhat suspicious from this point of view since the displacement of the lineage from the boundary is of the same order as the diameter of the tumor. Later in Conclusion 2 we will see that when  $\alpha = 1/2$  all the coalescences occur near the beginning of tumor growth.

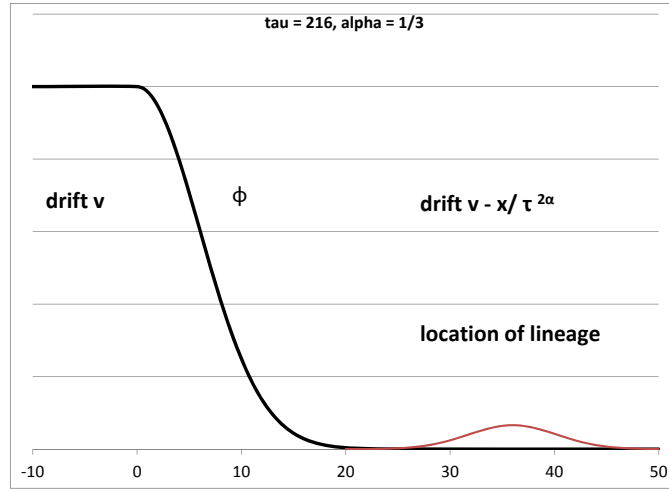


Figure 9: Picture of our concrete example indicating where the lineage is in equilibrium.

The reader should note that while (6) goes from 1 at  $v\tau$  to  $\approx 0$  over a window of size  $O(\tau^\alpha)$  the center of the stationary distribution is at  $v\tau^{2\alpha}$ , which is the point where the drift is 0. Because of this it is natural to look instead at  $\bar{U}_t = U_t - \tau^{2\alpha}v$ . When we do this the drift is just  $-\bar{u}/\tau^{2\alpha}$ , i.e.,  $\bar{U}_t$  is an Ornstein-Uhlenbeck process.

## 4 Discretization of Radial Process

To have the possibility for lineages to coalesce, we will discretize  $\bar{U}_s$  and  $\theta_s$  to take place on  $\epsilon\mathbb{Z}$ . This is biologically natural since the tumor is composed of cells. To fix the units, we will think of space as measured in  $cm$ , and time as measured in years.

- Since cells have a diameter of roughly  $10 \mu m = 10^{-5}m$ ,  $\epsilon \approx 10^{-3}$ .
- Since tumors are several  $cm$  in diameter and take years to form the speed  $v \approx 1$ . For simplicity we will set  $v = 1$

The Ornstein-Uhlenbeck has unbounded drift, so a naive approach to discretization can lead to negative jump rates. To avoid this, we will instead use a continuous time birth and death process with stationary measure

$$\mu(x) = \exp(-x^2/\tau^{2\alpha}) \quad (9)$$

in which the rate of jumps from  $x$  to  $y$ ,  $q(x, y)$ , satisfies the detailed balance condition:

$$\mu(x)q(x, x + \epsilon) = \mu(x + \epsilon)q(x + \epsilon, x). \quad (10)$$

To have (10) we need

$$\frac{q(x, x + \epsilon)}{q(x + \epsilon, x)} = \frac{\mu(x + \epsilon)}{\mu(x)} = \exp(-(2\epsilon x + \epsilon^2)/\tau^{2\alpha}).$$

To do this we can take

$$q(x, x + \epsilon) = \frac{\epsilon^{-2}}{2} \exp(-(2\epsilon x + \epsilon^2)/2\tau^{2\alpha}), \quad q(x + \epsilon, x) = \frac{\epsilon^{-2}}{2} \exp((2\epsilon x + \epsilon^2)/2\tau^{2\alpha}). \quad (11)$$

To see that discretized chain is reasonable note that replacing  $x$  by  $x - \epsilon$  in the second formula this chain has infinitesimal mean

$$\epsilon q(x, x + \epsilon) - \epsilon q(x, x - \epsilon) = \frac{\epsilon^{-1}}{2} \left[ \exp^{-(2\epsilon x + \epsilon^2)/2\tau^{2\alpha}} + \exp^{(2\epsilon x - \epsilon^2)/2\tau^{2\alpha}} \right] \approx -\frac{x}{\tau^\alpha},$$

and the infinitesimal variance

$$\epsilon^2 q(x, x + \epsilon) + \epsilon^2 q(x, x - \epsilon) = \frac{1}{2} \left[ \exp^{-(2\epsilon x + \epsilon^2)/2\tau^{2\alpha}} + \exp^{(2\epsilon x - \epsilon^2)/2\tau^{2\alpha}} \right] \approx 1.$$

## 5 Angular Process

To discretize the angular process  $\theta_t$  when  $R_0 = R$  we let  $\bar{\theta}_t$  be a process that jumps  $y \rightarrow y \pm \epsilon/R$  at rate  $(\epsilon^{-2}/2)(R^2/R_s^2)$  so that the infinitesimal variance is

$$2 \frac{\epsilon^2}{R^2} \cdot \frac{\epsilon^{-2}}{2} \cdot \frac{R^2}{R_s^2} = \frac{1}{R_s^2}.$$

The angular variable is easy to understand. If we condition on  $R_t$ ,  $t \geq 0$  then

$$\bar{\theta}_t - \bar{\theta}_0 \approx \text{normal}(0, \sigma_t^2) \quad \text{where} \quad \sigma_t^2 = \int_0^t 1/R_s^2 ds \quad (12)$$

Using the fact that in the limit as  $R \rightarrow \infty$   $R_s \sim T - s = R - s$  we have

$$\int_0^t \frac{1}{(R-s)^2} ds = \frac{1}{R-t} - \frac{1}{R} = \frac{t}{R(R-t)}$$

**Remark 1.** Note that for any  $\delta > 0$  this calculation is valid until  $t = R - R^\delta$  because in that case  $\sigma^2 = O(R^{-\delta})$ .

Consider now two points at distance  $R$  from the center whose angular parts differ by  $\theta_0 = aR^{-1/2}$ . For the lineages can hit by time  $t$  we need

$$\frac{t}{R(R-t)} = a^2 R^{-1} \quad \text{or} \quad t = a^2(R-t) \quad \text{or} \quad t = \frac{a^2 R}{1+a^2}$$

**Conclusion 1.** Tumor cells whose angles to the origin are separated by distance  $aR^{-1/2}$  will coalesce after a time with mean

$$\approx \frac{a^2 R}{1+a^2} \quad (13)$$

Due to the somewhat informal nature of our arguments, we will call them conclusions not theorems. To see what this says, note that when  $a = 3$ ,  $t = 0.9R$ . If  $R = 5000$  cells which is a tumor with of order  $10^8$  cells and a radius 5 cm,  $3R^{1/2} \approx 210$  cells. When  $a = 0.1$   $t \approx 0.01R$ . In the concrete example  $0.1R^{1/2} \approx 7$  cells

## 6 Coalescence of two lineages on the boundary

Consider now two independent genealogical processes  $(\bar{X}_1(t), \bar{\theta}_1(t))$  and  $(\bar{X}_2(t), \bar{\theta}_2(t))$ . To have both coordinates of order  $R$  we will let  $\bar{Y}_i(t) = R\bar{\theta}_i(t)$  to get a process that jumps  $y \rightarrow y \pm \epsilon$  at rate  $(\epsilon^{-2}/2)(R^2/R_s^2)$ . Let  $Z_i(t) = (\bar{X}_i(t), \bar{Y}_i(t))$ . Let  $T_\theta = \min\{t : \bar{Y}_1(t) = \bar{Y}_2(t)\}$ , and let  $T_c = \min\{t : Z_1(t) = Z_2(t)\}$  be the coalescence time. Conclusion 1 gives a lower bound on the coalescence time by using  $T_\theta \leq T_c$ . We will now identify a situation in which  $T_c$  occurs soon after  $T_\theta$ .

**Conclusion 2.** Suppose that

$$\bar{X}_i(0) \sim R, \quad |\bar{X}_1(t) - \bar{X}_2(t)| = O(R^\alpha) \quad |\bar{Y}_1(t) - \bar{Y}_2(t)| = R^\beta.$$

If  $\alpha < \beta < 1/2$ .

$$T_c = T_\theta + O(R^{2\alpha} \log^2 R)$$

Since  $T_\theta = O(R^{2\beta})$  and  $\alpha < \beta$  this says that  $T_c$  occurs soon after  $T_\theta$ .

*Ideas that go into the proof.* In Section 8 we will show that the radial process comes to equilibrium in time  $O(R^{2\alpha})$ , and that in equilibrium  $P(\bar{X}_1(t) = \bar{X}_2(t)) \approx R^{-\alpha}$ . Because of this we can hope for coalescence when the amount of time the angular parts have been equal  $|\{t : \bar{\theta}_1(t) = \bar{\theta}_2(t)\}| = O(R^\alpha)$ . Since the probability of equality is  $O(t^{-1/2})$  this should take time  $R^{2\alpha}$ . However, due to the fact that two dimensional random walk is recurrent, we need time  $O(R^{2\alpha} \log^2 R)$ . See Section 9 for details.  $\square$

Note that the  $Y$  coordinates of lineages separated by  $R^\beta$  with  $\beta < 1/2$  will by conclusion 1 coalesce with high probability at a time that is  $o(R)$  so there will be solid patches of size  $R^\alpha$  by  $R^\beta$  in polar coordinates in which most cells with the same ancestor.

## 7 Coalescence with interior lineages

**Conclusion 3.** *Consider two lineages with*

$$\bar{X}_1(0) \sim b_1 R \quad \bar{X}_2(0) \sim b_2 R \quad |\bar{Y}_1(t) - \bar{Y}_2(t)| = o(R^{1/2})$$

where  $b_1 < b_2$ . Then for any  $\delta > 0$  the probability the two lineages coalesce by time  $R - R^\delta$  tends to 0.

*Ideas that go into the proof.* The closer lineage will do a two dimensional random walk until time  $(1 - b_1)R$ . The first time that the two lineages are at the same distance from the center, the two angles will be separated by  $\sqrt{R}$ , so by the argument in Section 9 they will need time at least  $O(R \log^2 R)$  to coalesce. The stated result comes from the fact that Remark 1 implies our calculations are only valid until time  $R - R^\delta$ .  $\square$

Using results of Cox and Griffeath [8] we can study the case  $b_1 = b_2$ . The next result is a corollary of their formula (3.2). (The factor  $\theta(1 - \theta)$  comes from the fact that they consider the voter model starting from product measure with density  $\theta$ .)

**Conclusion 4.** *Consider two lineages with*

$$\bar{X}_1(0) \sim b_1 R \quad \log \|\bar{Z}_1(0) - \bar{Z}_2(0)\| \sim (\gamma/2) \log R$$

then the probability the two lineages coalesce by time  $t$  converges to  $1 - \gamma$ .

The theorem says that the density of the cluster of sites whose genealogies coalesce with that of  $Z_1(0)$  has a density that varies with the scale at which you examine it. Note that when  $\gamma = 1/2$  the limit is 0. Using the hitting time results from Section 9 one can show that if  $\|\bar{X}_1(0) - \bar{X}_2(0)\| \sim a \log R$  then the coalescence probability  $\sim f(a)/\log n$ .

## 8 Equilibration of the centered radial process

When the initial separation is  $O(R^{1/2})$  the coalescence may or may not occur depending on the variability of the radial component. In this section, we will show that the time for the radial process seen from the moving reference frame to reach equilibrium is  $O(\tau^{2\alpha})$ . As we have already shown in Lemma 2 the standard deviation in equilibrium is  $O(\tau^\alpha)$ . Consider

first times  $t \ll T$ . In this case  $\tau = T - t$  is constant and the radial process in the moving reference frame is

$$dX_t = dB_t - X_t/\tau^{2\alpha} dt.$$

If we let  $\beta = 1/\tau^{2\alpha}$  to simplify notation, this SDE can be solved explicitly by

$$X_t = e^{-\beta t} \left( X_0 + \int_0^t e^{\beta s} dB_s \right). \quad (14)$$

A proof can be found on page 180 of [12] but it is easily checked by formally differentiating the expression in (14)

$$dX_t = -\beta e^{-\beta t} \left( X_0 + \int_0^t e^{\beta s} dB_s \right) + dB_t.$$

From this we see that when  $X_0 = x$ ,  $X_t$  is normal with mean 0 and variance

$$\int_0^t e^{-2\beta(t-s)} ds = \frac{1}{2\beta}(1 - e^{-2\beta t}).$$

As  $t \rightarrow \infty$  this converges to normal(0, 1/2 $\beta$ ). In addition, we can see from the formula if  $t \gg 1/\beta = \tau^{2\alpha}$  we are close to equilibrium.

We need to derive a convergence to equilibrium for the discretized process. To lead up to this we will take another approach to the previous result. We start with the infinitesimal generator for the OU process

$$Lf(x) = \frac{1}{2}f''(x) - \frac{x}{\tau^{2\alpha}}f'(x).$$

To look for an eigenfunction, we guess  $f_1(x) = x$  which has

$$Lf_1(x) = -x/\tau^{2\alpha} = (-1/\tau^{2\alpha})f_1(x).$$

For the Ornstein-Uhlenbeck generator it is known that  $-1/\tau^{2\alpha}$  is the largest negative eigenvalue so the rate of convergence to equilibrium is  $\tau^{2\alpha}$ .

To mimic this argument in the discrete setting note that what  $\mu$  defined in (9) is not-malized to be a probability measure it is approximately

$$\pi(x) = \epsilon(\pi\tau^{2\alpha})^{1/2}e^{-x^2/\tau^{2\alpha}}, \quad (15)$$

since summing over  $x$  gives a Riemann sum approximating the integral of a normal density. From the definition of the rates  $q(x, y)$  it satisfies the detailed balance condition. Thus the flow of probability across the edge  $(x, x + \epsilon)$  in equilibrium is

$$\pi(x)q(x, x + \epsilon) = \pi(x + \epsilon)q(x + \epsilon, x) \equiv Q(x, x + \epsilon).$$

To bound the spectral gap we look at the Dirichlet form

$$\mathcal{E}(f, f) = \sum_x Q(x, x + \epsilon)[f(x + \epsilon) - f(x)]^2,$$

and note

$$1 - \lambda_1 = \inf\{\mathcal{E}(f, f) : \text{var}_\pi(f) = 1\}. \quad (16)$$

If we take  $f_1(x) = x$  then  $\text{var}_\pi(f) = \tau^{2/\alpha}/2$  while using (9) and (11)

$$\begin{aligned} \mathcal{E}(f_1, f_1) &= \sum_{x \in \epsilon\mathbb{Z}} \epsilon(\pi\tau^{2\alpha})^{1/2} e^{-x^2/\tau^{2\alpha}} \frac{\epsilon^{-2}}{2} \exp(-(2\epsilon x + \epsilon^2)/2\tau^{2\alpha}) \epsilon^2 \\ &= \frac{1}{2} \sum_{x \in \epsilon\mathbb{Z}} \epsilon(\pi\tau^{2\alpha})^{1/2} \exp\left(\frac{-x^2 - \epsilon x - \epsilon^2/2}{\tau^{2\alpha}}\right) \\ &= \frac{1}{2} \sum_{x \in \epsilon\mathbb{Z}} \epsilon(\pi\tau^{2\alpha})^{1/2} \exp\left(\frac{-(x + \epsilon/2)^2}{\tau^{2\alpha}}\right) e^{-\epsilon^2/4\tau^{2\alpha}} \approx 1/2 \end{aligned} \quad (17)$$

since when the last term is removed this is close to the integral of the density function for the normal with mean  $-\epsilon/2$  and variance  $\tau^{2\alpha}/2$ . Using (16) now we conclude that in the discretized version the spectral gap is close to  $1/\tau^{2\alpha}$ , i.e., the distribution of  $\bar{X}_t$  comes to equilibrium in time  $\tau^{2\alpha} \leq T$  since  $2\alpha < 1$ .

## 9 Estimating the Coalescence Time

We begin by explaining the method we will use, starting with the simpler problem of determining how long it takes for two independent continuous time random walks to hit on  $\mathbb{Z}^2$ . The methodology follows Section 3 of [53]. Let  $W_t$  be the difference in the positions of the two walks and let  $T_0 = \min\{t : W_t = 0\}$ . Breaking things down according to the value of  $T_0$  and writing the density function of  $T_0$  as  $P_x(T_0 = s)$

$$\begin{aligned} \int_0^t P_x(W_r = 0) dr &= \int_0^t P_x(T_0 = s) \int_0^{t-s} P_0(W_r = 0) dr ds \\ &\leq \int_0^t P_x(T_0 = s) \int_0^t P_0(W_r = 0) dr ds \end{aligned}$$

so we have

$$P_x(T_0 \leq t) \geq \frac{\int_0^t P_x(W_r = 0) dr}{\int_0^t P_0(W_r = 0) dr}. \quad (18)$$

Note that in using this result we want a lower bound on the numerator and an upper bound on the denominator.

For a bound in the other direction we use

$$\int_0^{t+t_0} P_x(W_r = 0) dr \geq \int_0^t P_x(T_0 = s) \int_0^{t_0} P_0(W_r = 0) dr ds$$

which gives

$$P_x(T_0 \leq t) \leq \frac{\int_0^{t+t_0} P_x(W_r = 0) dr}{\int_0^{t_0} P_0(W_r = 0) dr}. \quad (19)$$

Note that in using this result we want an upper bound on the numerator and a lower bound on the denominator.

*Proof of Conclusion 2.* The assumption  $\alpha < \beta$  implies that the centered radial component is in equilibrium at times  $t \geq R^{2\beta}$ . Suppose  $\bar{X}_1(t)$  and  $\bar{X}_2(t)$  are independent and have the equilibrium distribution (15). Using the fact that  $\pi(x) = \pi(-x)$  on  $\epsilon\mathbb{Z}$  and is an approximation to  $\text{normal}(0, \tau^{2\alpha}/2)$ , we see that if  $t = o(R)$  then

$$\begin{aligned} P(\bar{X}_1(t) = \bar{X}_2(t)) &= \sum_x \pi(x)^2 = \sum_x \pi(x)\pi(-x) \\ &= (\pi * \pi)(0) \approx \epsilon(2\pi\tau^{2\alpha})^{-1/2} \approx C_1\epsilon R^{-\alpha} \end{aligned} \quad (20)$$

where  $*$  indicates convolution and  $C_1 = (2\pi)^{-1/2}$  is a constant. Using this and the local central limit theorem we conclude that if  $K$  is large

$$\begin{aligned} E \int_K^t P(Z_1(T_\theta + s) = Z_2(T_\theta + s)) ds &\approx C_1\epsilon R^{-\alpha} \cdot \epsilon \int_K^t (2\pi s)^{-1/2} ds \\ &\sim C_2\epsilon^2 R^{-\alpha} t^{-1/2} \end{aligned} \quad (21)$$

where  $C_2 = C_1(2\pi)^{-1/2} \cdot 2 = \pi^{-1}$

To get a lower bound on  $P(T_\theta \leq t)$  using (18) we have to compute

$$\int_0^t P_e(Z_1(s) = Z_2(s)) ds \quad (22)$$

where  $e$  denotes the process starting from  $Z_1(0) = Z_2(0)$ . The integrand is

$$\approx \begin{cases} 1/(2\pi s) & s \ll R^{2\alpha} \\ R^{-\alpha} & s \gg R^{2\alpha} \end{cases}$$

since in the first case the drift in the OU process is not significant and in the second the radial part is in equilibrium. In the gap between the two cases the formulas are almost the same, so the expression in (22) is

$$\begin{aligned} &\approx \epsilon^2 \left( \int_1^{R^{2\alpha}} (2\pi s)^{-1} ds + C_1 R^{-\alpha} \int_{R^{2\alpha}}^t (2\pi t s)^{-1/2} ds \right) \\ &= \epsilon^2 \left( \frac{\alpha}{\pi} \log R + C_2 R^{-\alpha} t^{1/2} \right) \end{aligned} \quad (23)$$

Combining the last result with (21) and using (18) we see that if  $t = KR^{2\alpha} \log^2 R$  then

$$P(T_c \leq t) \geq \frac{KC_2}{\alpha/\pi + KC_2} \quad (24)$$

which is close to 1 if  $K$  is large. Since  $T_c \geq T_\theta$  we do not need a lower bound.  $\square$

Conclusion 3 can be proved in almost the same way using (19).

## 10 Discussion

At this point we have achieved our goal of analyzing the behavior of the lineages in the simplified model and found that near the boundary the coalescence patterns are similar to those found in the tumor data. Unfortunately, some other properties do not agree with observed patterns

- The patches of cells that share the same genealogy with an interior lineage have diameter  $O(R^{1/2})$  are not dense enough to be detected by biopsy.
- The genealogies we follow are  $\tau^{2\alpha}$  behind the front where the density of tumor cells is very small.

$$\exp(\varphi(\tau^\alpha)) = \exp(-\tau^{2\alpha}/2)$$

Taking  $\varphi(u) = u^\gamma/\gamma$  with  $\gamma > 2$  keeps genealogies closer but still the density at their location is low.

- In  $d = 3$  if two cells are separated by  $\log R$  at time  $t$  then, with high probability, they coalesce after time  $R - R^\delta$ , since three dimensional random walk is transient.

One of the problems with the current model is that there is considerable amount of turnover of cells in the interior of the tumor. This is not realistic since cell divisions in the interior are rare, and in some cases the cells will be hypoxic and die. We are currently investigating variants that only have growth at the frontier.

## Acknowledgements

This was partially supported by DMS 16164838 from the math biology program at NSF. The author would like to thank Jasmine Foo, Kevin Leder and Matthew Junge for useful comments. He would also like to thank David Basanta and Sandy Anderson for an invitation to visit the Integrated Mathematical Oncology at the Moffitt Cancer Center and an opportunity to give a talk about this work while it was still under development. Similar thanks go to Maury Bramson and the probability seminar at Minnesota.

## References

- [1] Armitage, P., and Doll, R. (1954) The age distribution of cancer and a multi-stage theory of carcinogenesis. *British J. Cancer*. 8, 1–12
- [2] Bozic, I. et al. (2010) Accumulation of driver and passenger mutations during tumor progression. *Proc. Natl. Acad. Sci.* 107(43), 18545–18550
- [3] Bozic, I. et al (2013) Evolutionary dynamics of cancer in response to targeted combination therapy. *eLife* 2, e00747.
- [4] Bozic, I., Gerold, J.M., and Nowak, M.A. (2016) Quantifying clonal and subclonal passenger mutations in cancer evolution. *PLOS Computational Biology* 12, e1004731.

- [5] Bramson, M., and Griffeath, D. (1980b) On the Williams-Bjerknes tumour growth model. II. *Math. Proc. Cambridge Philos. Soc.* 88, 339–357
- [6] Bramson, M., and Griffeath, D. (1981) On the Williams-Bjerknes tumour growth model. I. *Ann. Probab.* 9, 173–185
- [7] Bramson, M., and Griffeath, D. (1980) Asymptotics for interacting particle systems on  $\mathbb{Z}^d$ . *Z. Wahrsch. Verw. Gebiete.* 53, 183–196
- [8] Cox, J.T., and Griffeath, D. (1986) Diffusive clustering in the two-dimensional voter model. *Ann. Probab.* 14, 347–370.
- [9] Danesh, K., Durrett, R., Havriliesky, L., and Myers, E. (2013) A branching process model of ovarian cancer. *J. Theor. Biol.* 314, 10–15
- [10] Ding, Li, et al. (2012) Clonal evolution in relapsed acute myeloid leukemia revealed by whole-genome sequencing. *Nature.* 481, 506–510
- [11] Doering, C.R., Mueller, C., Semereka, P. (2003) Interacting particles, the stochastic Fisher-Kolmogorov-Petrovsky-Piscounov equation, and duality. *Physics A.* 325, 243–259
- [12] Durrett, R. (1996) *Stochastic Calculus*. CRC Press
- [13] Durrett, R. (2015) *Branching Process Models of Cancer*. Springer-Verlag, New York
- [14] Durrett, R., and Fan, Louis (2016) Genealogies in expanding populations. *Annals Applied Probability.* 16, 3456–3490
- [15] Durrett, R., Foo, J., Leder, K., Mayberry, J., Michor, F. (2011) Intratumor heterogeneity in evolutionary models of tumor progression. *Genetics.* 188, 461–477
- [16] Durrett, R., Foo, J. and Leder, K. (2016) Spatial Moran Models. II. Cancer initiation in spatially structured tissue. *Journal Mathematical Biology.* 72, 1369–1400
- [17] Edmonds, C.A., Lille, A.S., and Cavalli-Sforza, L.L. (2004) Mutations in the wave front of an expanding population. *Proc. Natl. Acad. Sci.* 101, 975–979
- [18] Fearon, E.R., and Vogelstein, B. (1990) A genetic model for colorectal tumorigenesis. *Cell.* 759–767
- [19] Frank, S.A. (2007) *Dynamics of Cancer: Incidence, Inheritance and Evolution*. Princeton U. Press
- [20] Fisher, J.C., and Hollomon, J.H. (1951) A hypothesis for the origin of cancer foci. *Cancer.* 4, 916–918
- [21] Garraway, L.A., and Lander, E.S. (2013) Lessons from the cancer genome. *Cell.* 153, 17–37

- [22] Gerlinger, M, et al (2015) Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *The New England Journal of Medicine*. 366, 883–892
- [23] Hallatschek, O., Hersen, P., Ramanathan, S., and Nelson, D.R. (2007) Genetic drift at expanding frontiers promotes gene segregation. *Proc. Natl. Acad. Sci.* 104, 19926–19930
- [24] Hallatschek, O., and Nelson, D.R. (2008) Gene surfing in expanding populations. *Theor. Pop. Biol.* 73, 158–170
- [25] Hallatschek, O., and Nelson, D.R. (2009) Life at the front of an expanding population. *Evolution*. 64, 193–206
- [26] Iwasa, Y., Nowak, M.A., and Michor, F. (2006) Evolution of resistance during clonal expansion. *Genetics*. 172, 2557–2566
- [27] Kang, H., et al. (2015) Many private mutations originate from the first few divisions of a human colorectal adenoma. *Journal of Pathology* 237, 355–362
- [28] Kardar, M., Parisi, G., and Zhang, Y.C. (1986) Dynamics of growing interfaces. *Phys. Rev. Letters*. 56, 889–892
- [29] Kaveh K, Takahashi Y, Farrar MA, Storme G, Guido M, Piepenburg J, Penning J, Foo J, Leder KZ, Hui SK. (2017) Combination therapeutics of Nilotinib and radiation in acute lymphoblastic leukemia as an effective method against drug-resistance. *PLoS Computational Biology*. 13(7):e1005482.
- [30] Korolev, K.S. et al. (2012) Selective sweeps in growing microbial colonies. *Physical Biology*. 9, paper 026008
- [31] Knudson, A.G. (1971) Mutation and cancer: Statistical study of retinoblastoma. *Proc. Natl. Acad. Sci.* 68, 820–823
- [32] Knudson, A.G. (2001) Two genetic hits (more or less) to cancer. *Nature Reviews Cancer*. 1, 157–162
- [33] Krone, S.M., and Neuhauser, C. (1997) The genealogy of samples in models with selection. *Genetics*. 145, 519–534
- [34] Krug, J., and Spohn, H. (1991) Kinetic roughening of growing surfaces. Pages 479–582 in *Solids Far from Equilibrium: Growth, Morphology, and Defects*. Cambridge U. Press
- [35] Leder, K., Foo, J., Skaggs, B., Gorre, M., Sawyers, C.L., and Michor, F. (2011) Fitness conferred by BCR-ABL kinase domain mutations determines the risk of pre-existing resistance in chronic myeloid leukemia. *PLoS One*. 6, paper e27682
- [36] Merlo, L.M.F., et al (2010) A Comprehensive Survey of Clonal Diversity Measures in Barrett’s Esophagus as Biomarkers of Progression to Esophageal Adenocarcinoma *Cancer Prevention Research*. 3, 1388–1397

- [37] Mueller, C., and Tribe, R. (1995) Stochastic p.d.e.s arising from the long range contact process and long range voter model. *Probability Theory and Related Fields*. 102, 519-545
- [38] Mumenthaler S, Foo J, Leder K, et al. (2011) Evolutionary modeling of combination strategies to overcome resistance to tyrosine kinase inhibitors in non-small cell lung cancer. *Mol. Pharm.* (2011) 8(6).
- [39] Navin, N., et al (2011) Tumour evolution inferred by single-cell sequencing. *Nature*. 472, 90–94
- [40] Neuhauser, C. (1992) Ergodic theorems for the multitype contact process. *Probab. Theory Related Fields*. 91, 467-506
- [41] Nordling, C.O. (1953) A new theory on cancer inducing mechanism. *Brit. J. Cancer*. 7, 68–72
- [42] Ryser, M.D., Min, B.H., Siegmund, K., and Shibata, D. (2017) Spatial mutation patterns as markers of early tumor cell mobility.
- [43] Sottoriva, A., et al (2013) Intratumor heterogeneity in human glioblastoma reflects cancer evolutionary dynamics. *Proc. Natl. Acad. Sci.* 110, 4009–4014
- [44] Sottoriva, A., et al. (2015) A Big Bang model of human colorectal cancer growth. *Nature genetics*. 47, 209–216
- [45] Tomasetti, C., and Levy, D. (2010) Roles of symmetric and asymmetric division of stem cells in the evolution of resistance. *Proc. Natl. Acad. Sci.* 107, 16766–16771
- [46] Tomasetti, C., Vogelstein, B., and Parmigiani, G. (2013) Half or more somatic mutations in cancers of self-renewing tissues originate prior to cancer initiation. *Proc. Natl. Acad. Sci.* 110, 1999–204
- [47] Vogelstein, B., and Kinzler, K.W. (1998) *The Genetic Basis of Human Cancer*. McGraw Hill
- [48] Vogelstein, B, et al. (2013) Cancer genome landscapes. *Science* 339, 1546–1558
- [49] Williams, T., and Bjerknes, R. (1972) Stochastic model for abnormal clone spread through epithelial basal layer. *Nature*. 235, 19–21
- [50] Yachida, S., et al (2010) Distant metastases occur late during the genetic evolution of pancreatic cancer. *nature*. 467, 1114–1117
- [51] Zabolitsky, J.G., and Stuaffer, D. (1986) Simulation of large Eden clusters. *Phys. Rev. A* 34, 1523–1530
- [52] Zabolitsky, J.G., and Stuaffer, D. (1986) Dynamics scaling of Eden cluster surfaces. *Phys. Rev. Letters*. 57, 1809 (only one page)
- [53] Zahle, I., Cox, J.T., and Durrett (2005) The stepping stone model, II. Genealogies and the infinite sites model. *Ann. Appl. Probab.* 15, 671–699