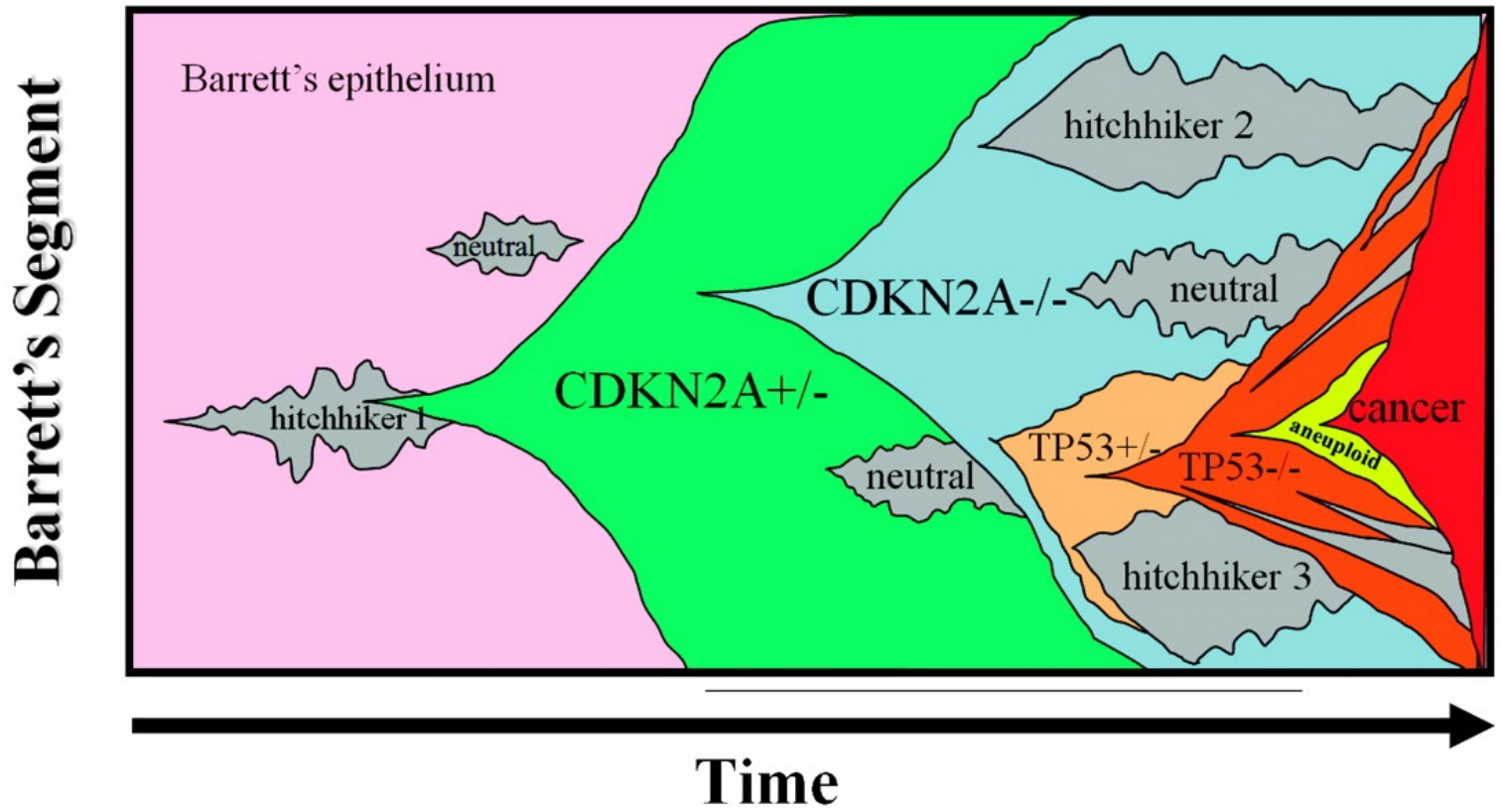
Genealogies in Growing Solid Tumors

Rick Durrett

Over the past two decades, the theory of tumor evolution has been based on the notion that the tumor evolves by a succession of selective sweeps. Sottoriva et al [1] sampled 349 individual tumor glands from the opposite sides of 15 colorectal tumors and large adenomas. Based on this they suggested that all of the driver mutations are present at the time of the initial expansion. In this talk we will describe a simple mathematical model that reproduces the observed phenomena and makes quantitative predictions.

[1] *Nature Genetics.* 47 (2015), 209-216

Clonal Expansion in Barrett's Esophagus



Observed Clonal Structure in AML

Clonal fractions at diagnosis: 13%, 29%, 5%, and 53%. Small clone was dominant at relapse. Had 78 new mutations compared to first sampling.

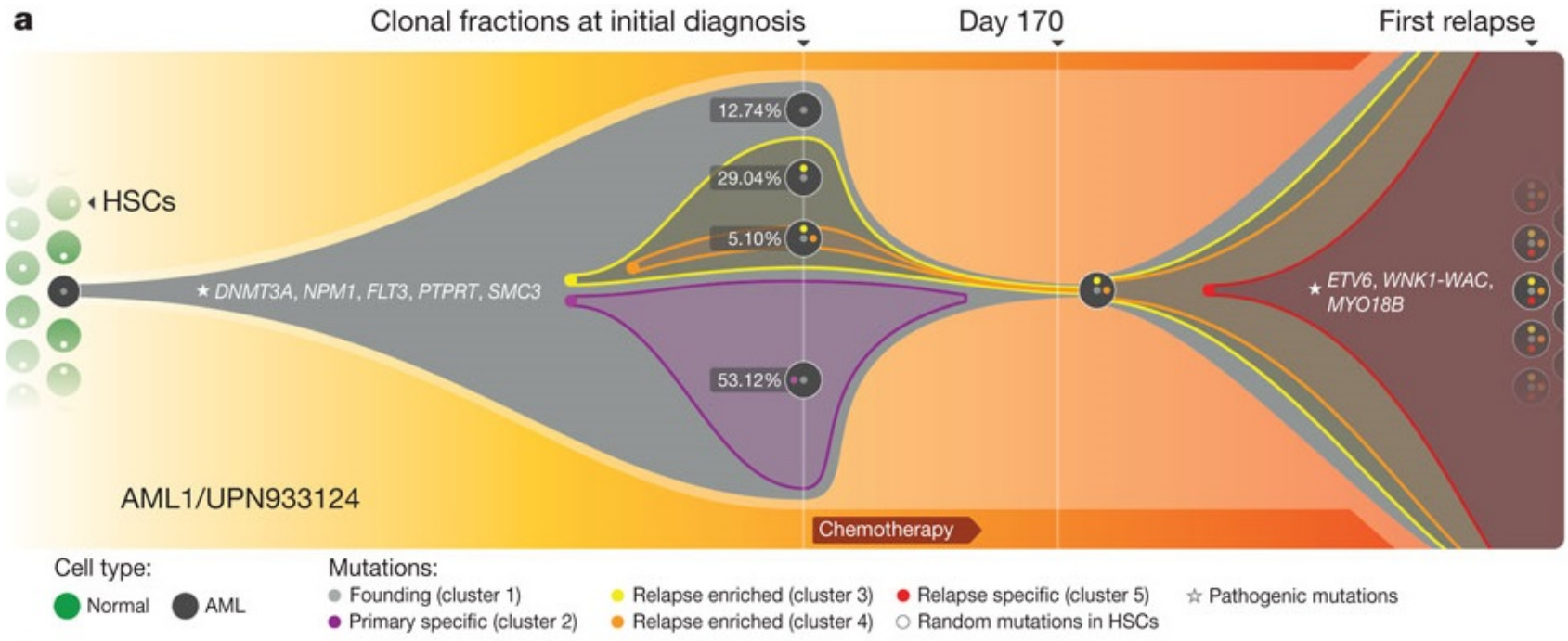


Figure: Ding et al. (2012) *Nature.* 481, 506–510

Sottoriva et al: sampling

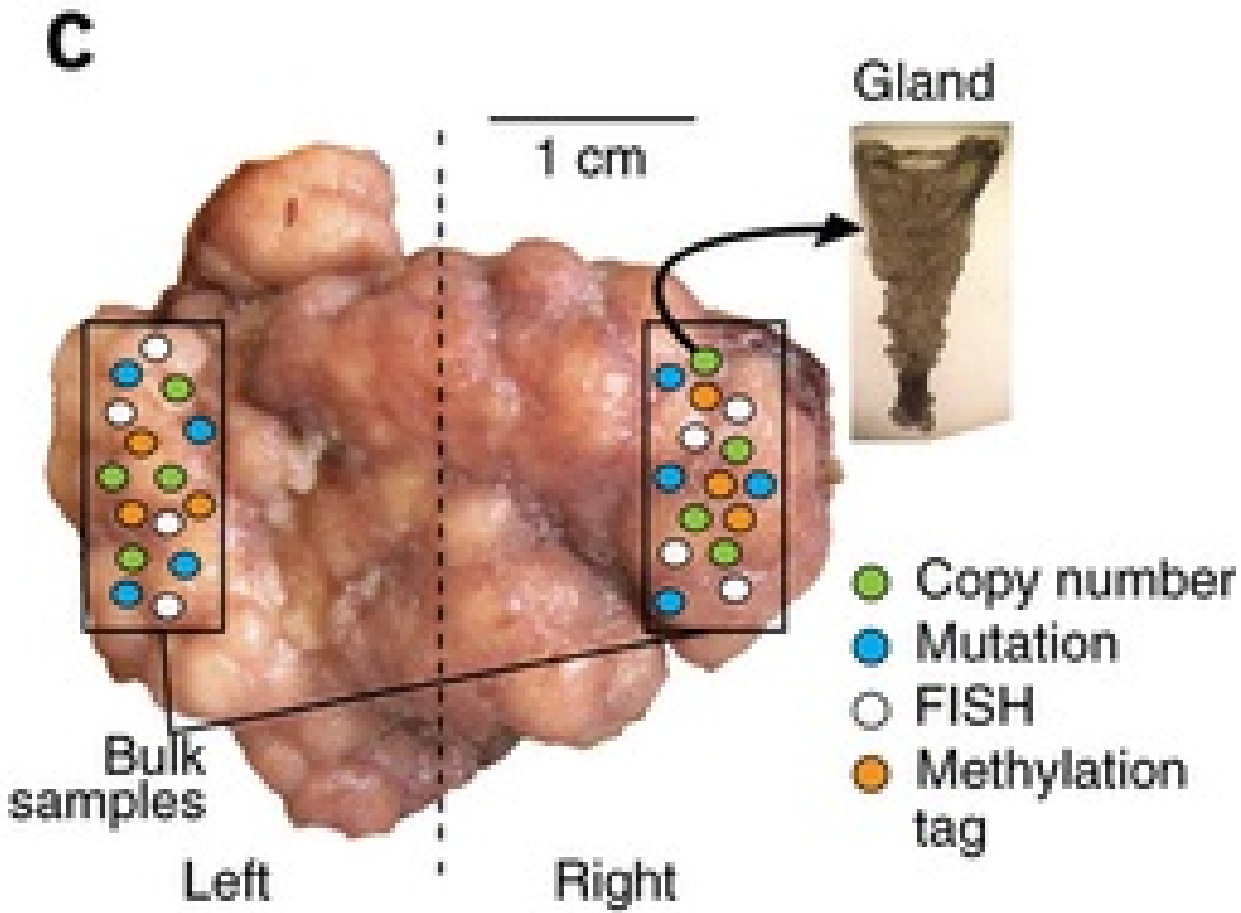


Figure: 349 individual tumor glands were sampled from the opposite sides of 15 colorectal tumors and large adenomas.

Big Bang Picture

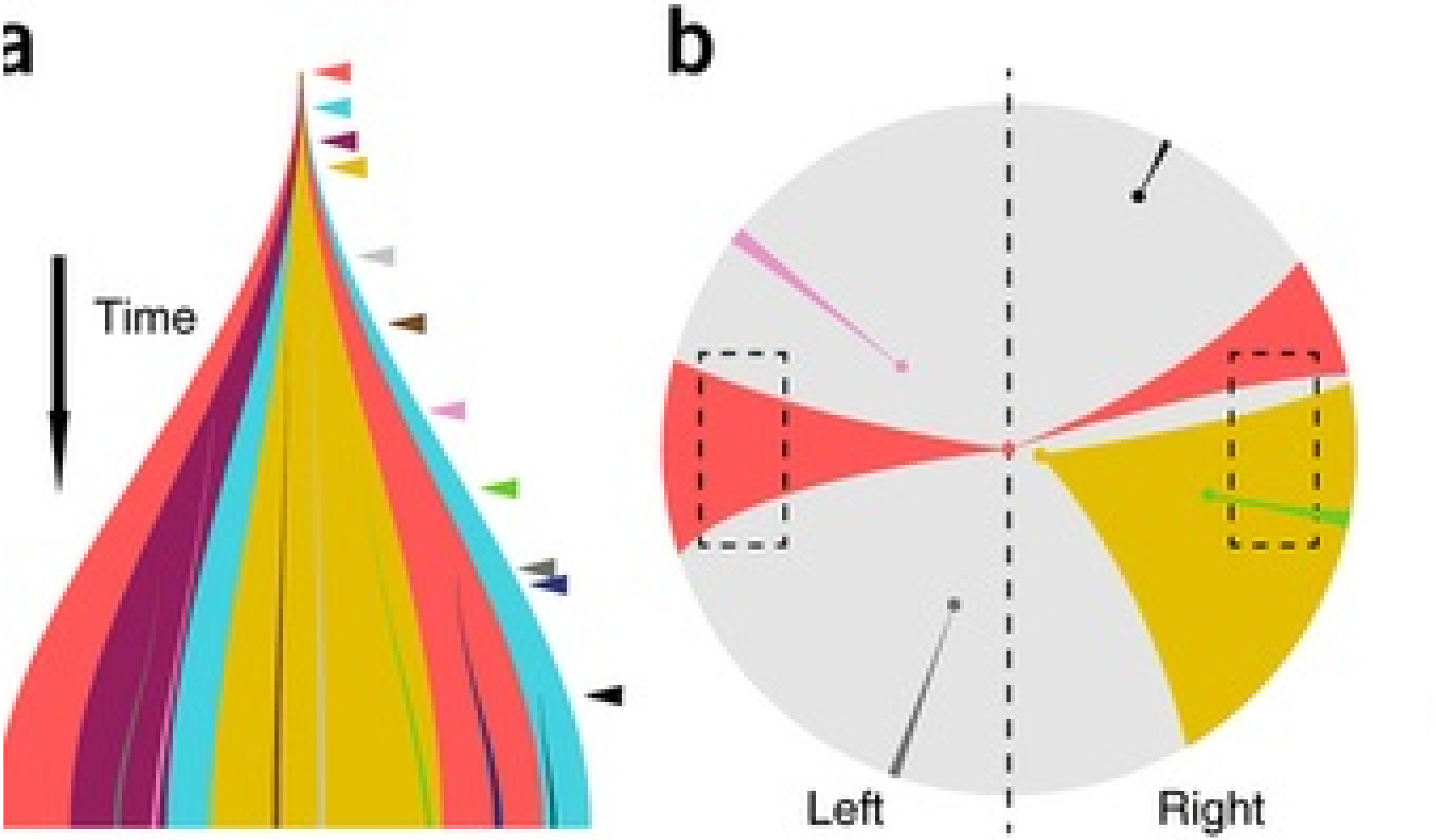
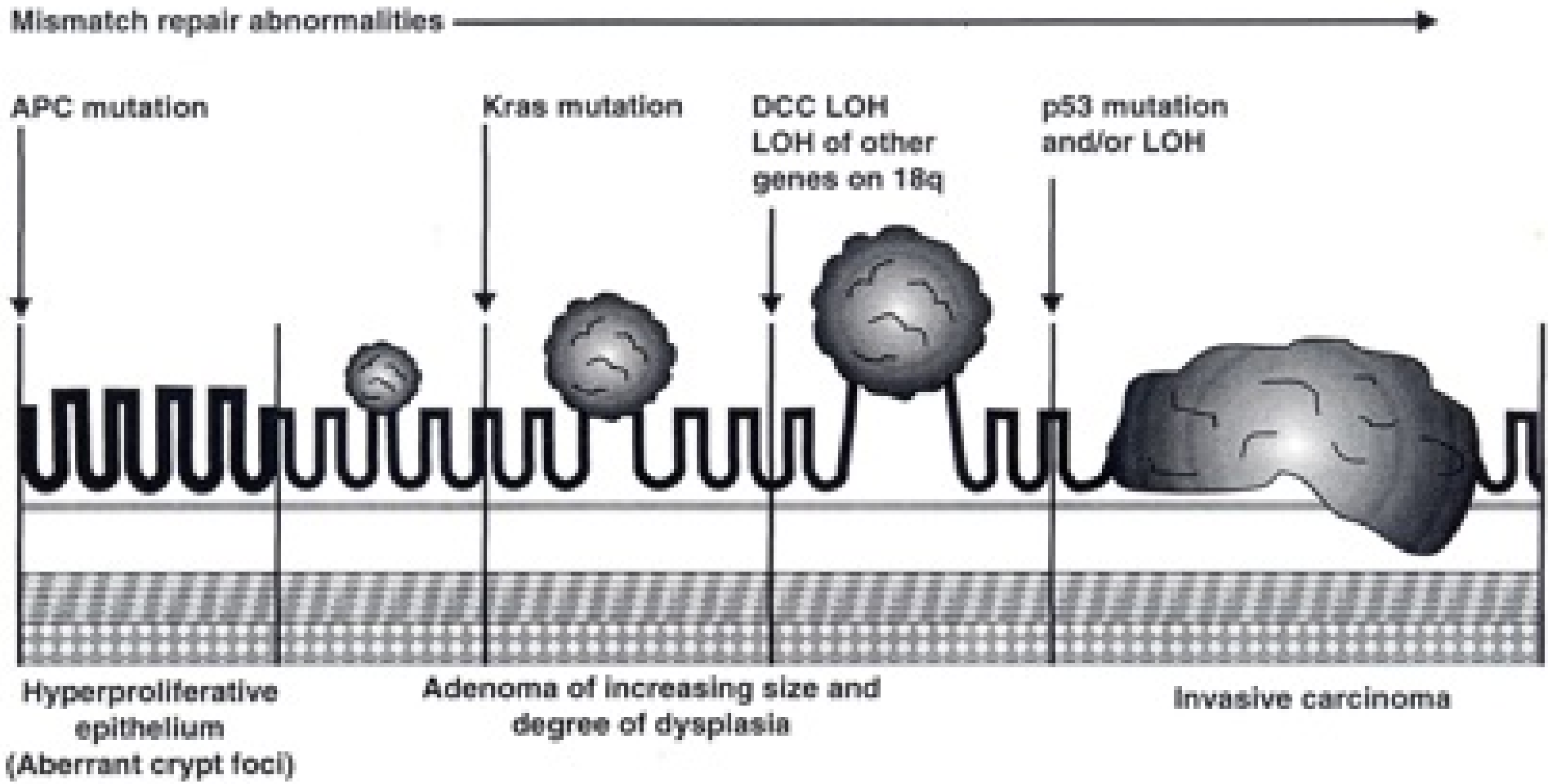


Figure: Temporal and spatial patterns of mutations

Progression of Colon Cancer



Mutation patterns: adenomas

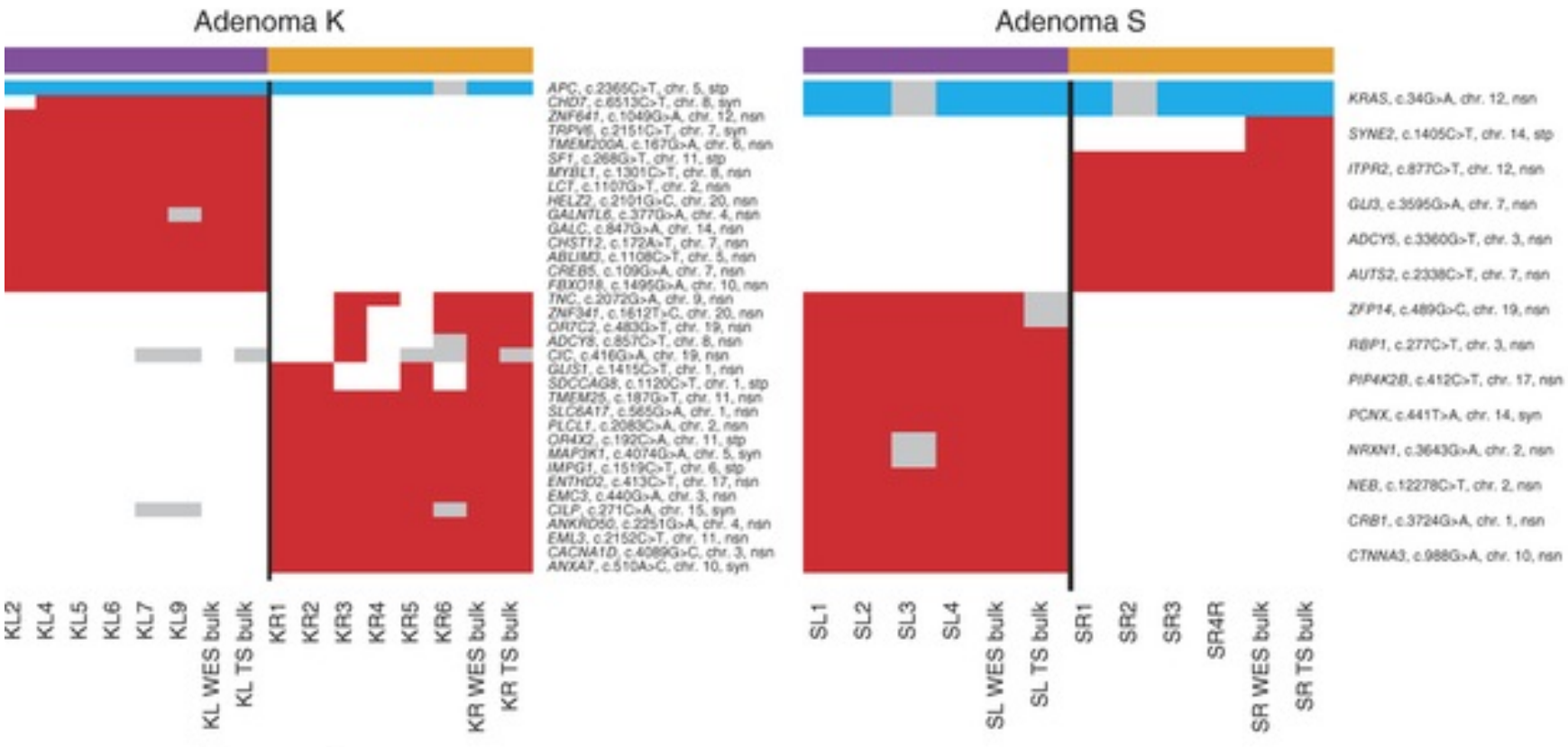


Figure: Adenomas were characterized by mutations and copy number aberrations (CNA) that segregated between tumor sides.

Mutation patterns: carcinomas

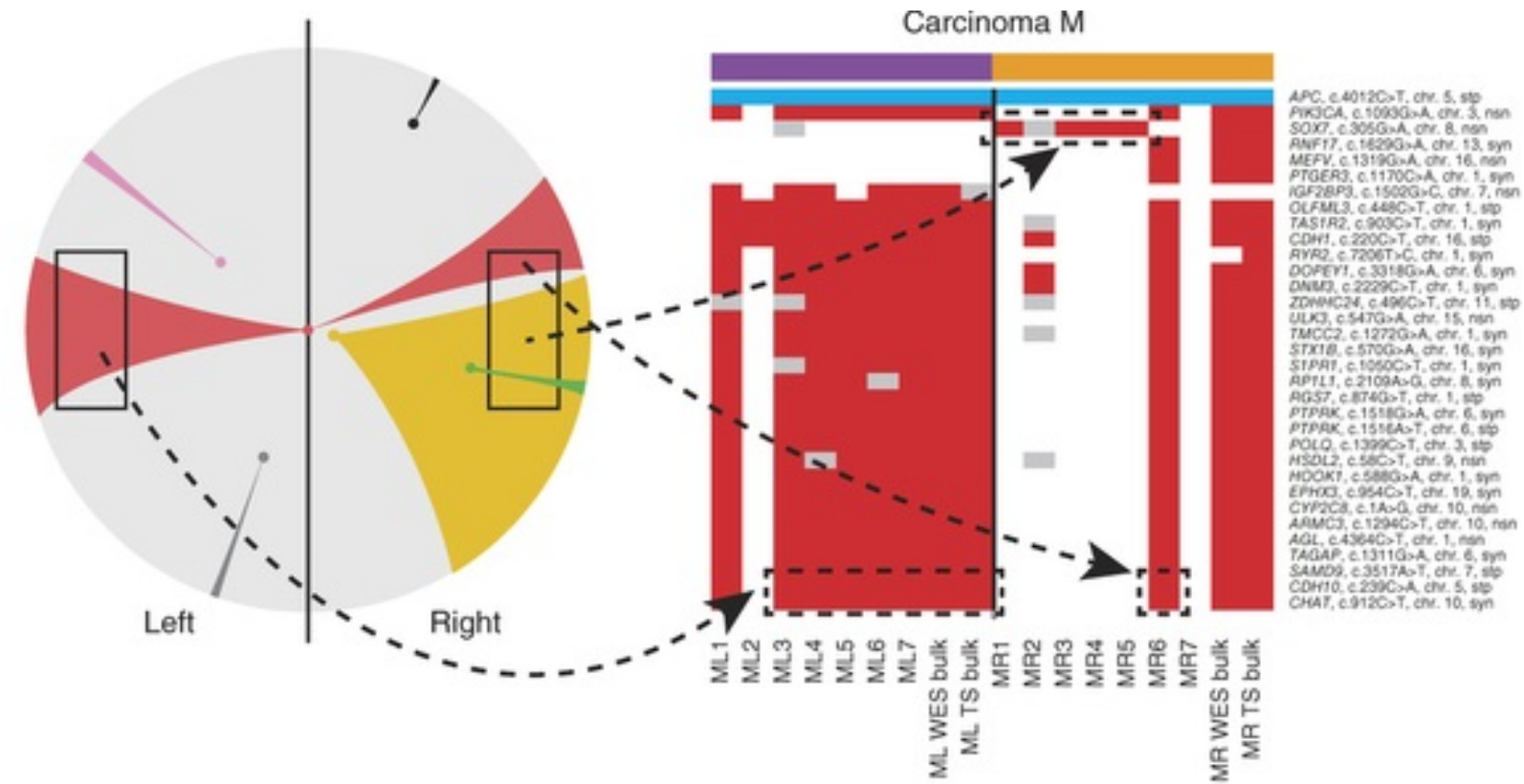


Figure: In contrast the majority of carcinomas exhibited the same private CNA in individual glands from different sides of the tumor.

The difference is early mixing

Kang et al (list of authors contains Sottoriva, Cristinda Curtis, and Darryl Shibata.) J. Pathology 237 (2015), 355–362.

In order to be detectable (at a frequency of 10%) a private mutation must occur in the first few cell divisions ($1/8 = 12.5\%$). Our work will show this is wrong.

Ryser, Min, Siegmund and Shibata, manuscript in preparation
Found evidence of early abnormal cell movement in 8 of 15 invasive colorectal carcinomas ("born to be bad") but not in four benign adenomas

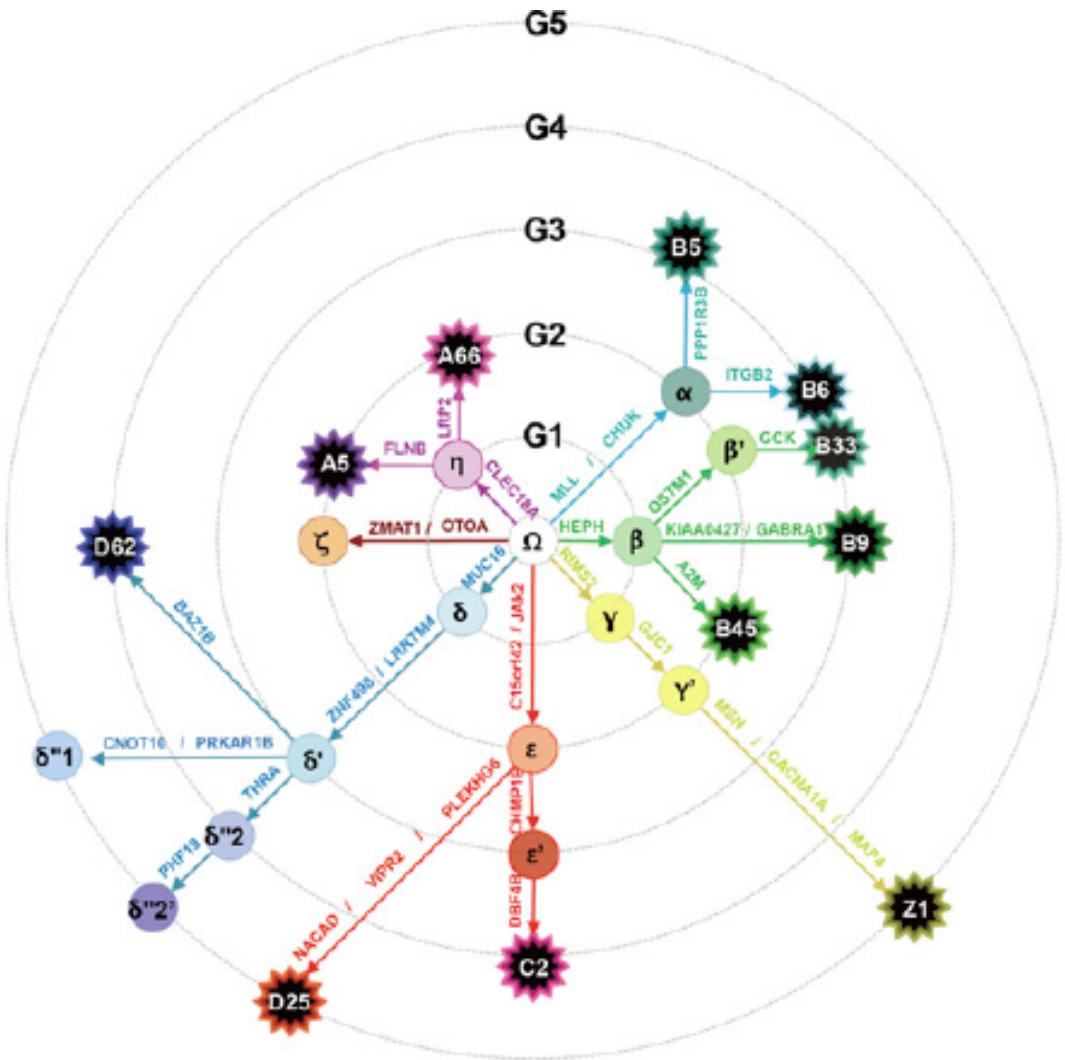


Figure: Genealogy in a hepatocellular carcinoma

Hallatschek et al (2007) PNAS 104, 19926–19930

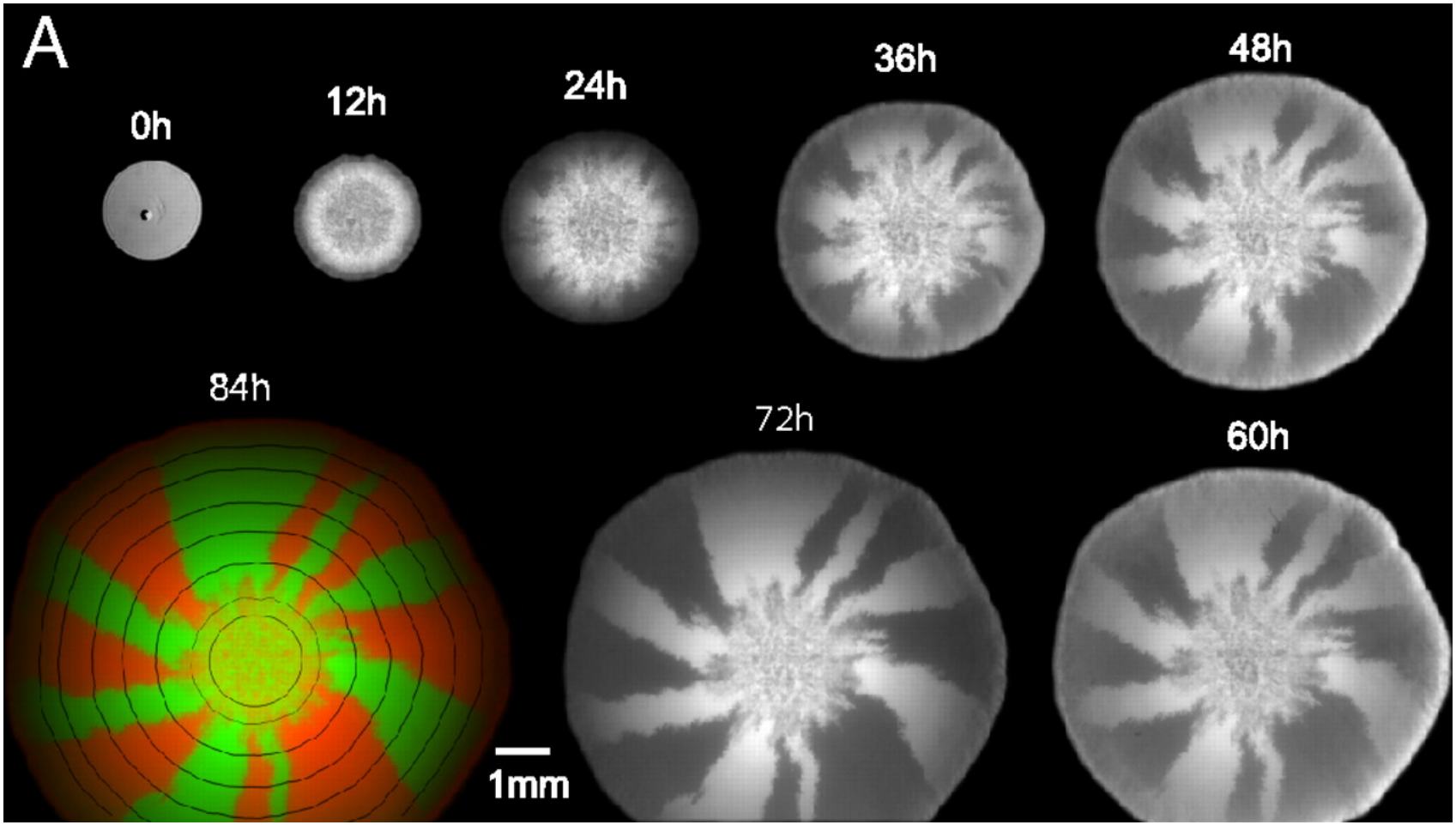


Figure: Competition of flourescently labelled yeast cells

Model 1: Biased voter model

In the biased voter model, ξ_t, 0's (normal cells) give birth at rate 1, and 1's (cancer cells) at rate λ. In either case the new individual is sent to a randomly chosen nearest neighbor on $\mathbb{Z}^d$.

Williams, T., and Bjerknes, R. (1972) Stochastic model for abnormal clone spread through epithelial basal layer. *Nature.* 235, 19–21

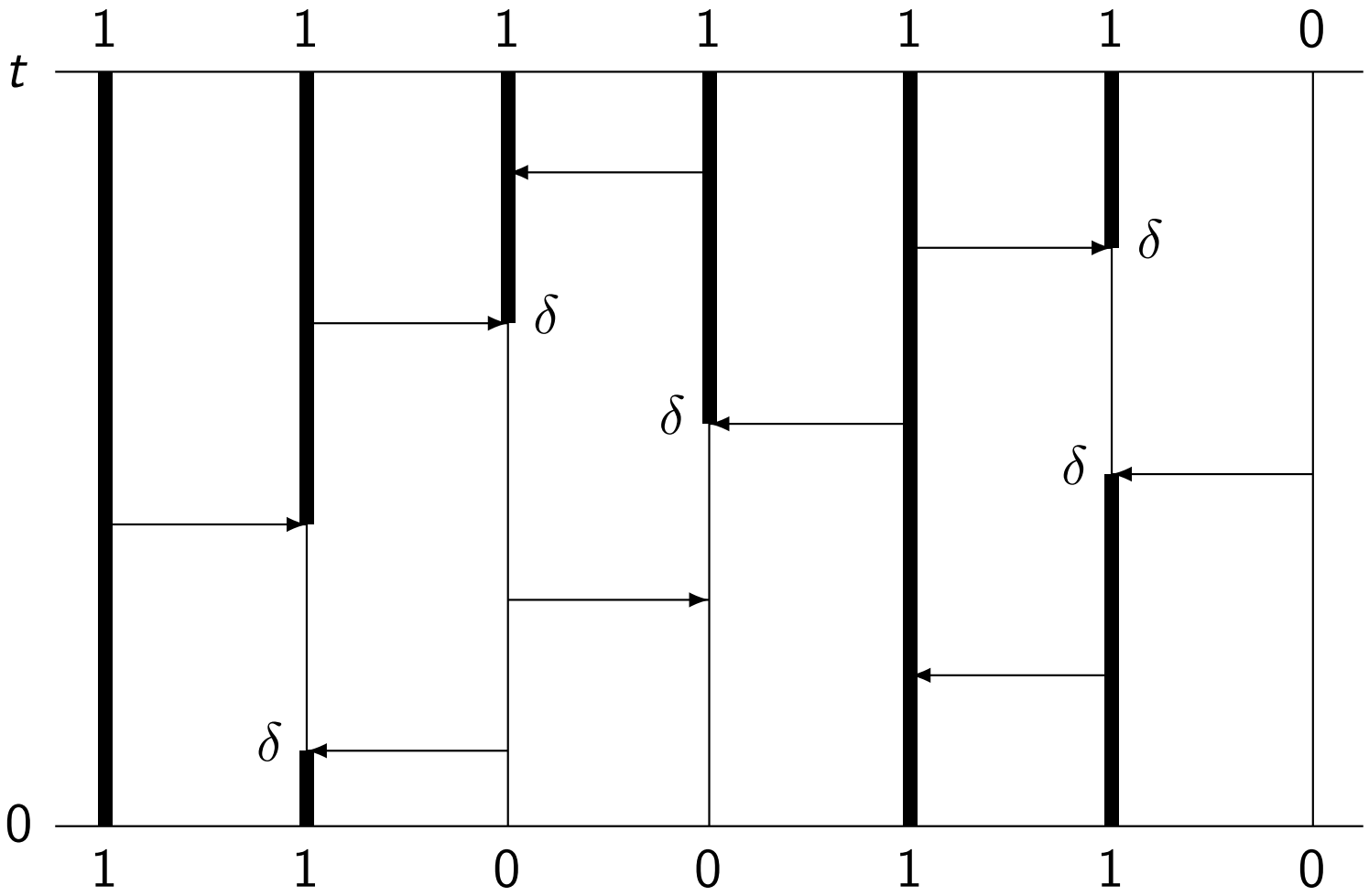
Bramson, M., and Griffeath, D. On the Williams-Bjerknes tumour growth model. II. *Math. Proc. Cambridge Philos. Soc.* 88 (1980), 339–357. I. *Ann. Probab.* 9 (1981), 173–185.

Durrett, R., Foo, J., Leder, K. (2016) Spatial Moran Models. II. Cancer inititaion in spatially structured tissue. *J. Math. Biol.* 72, 1369–1400

J. Foo, K, Leder, M.D. Ryser. (2014) Multifocality and recurrence risk: a quantitative model of field cancerization. *J. Theor. Biol.* 355, 170–184

Construction of the Biased Voter Model

$\to \delta$'s (Poisson rate 1) spread either type, $\to$'s (rate $\lambda - 1$) only spread 1's



Duality with coalescing branching random walk

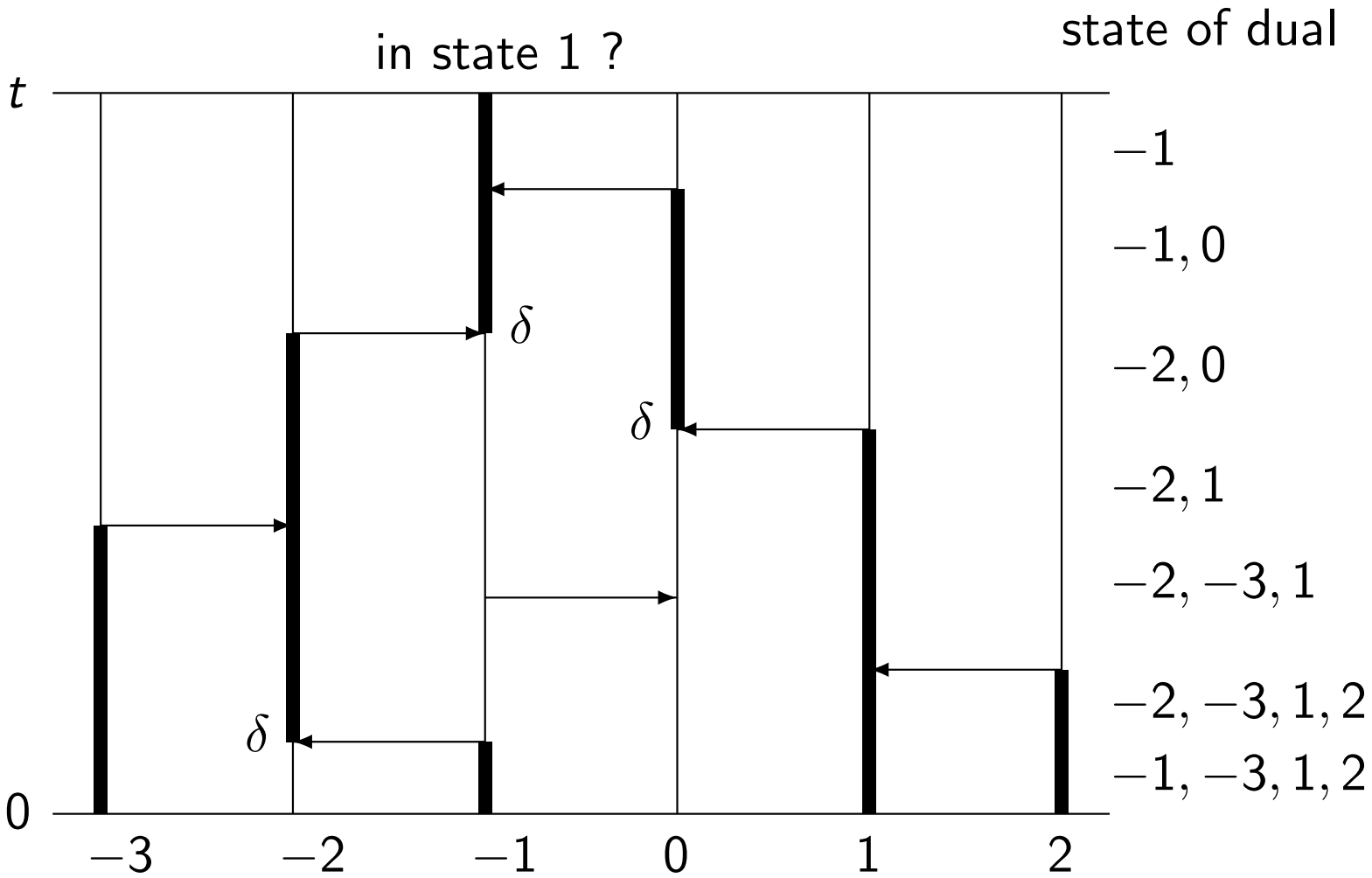
Particles in the dual $\zeta_t^{B,t}$ jump to each nearest neighbor at rate $1/2d$, give birth onto each nearest neighbor at rate $(\lambda - 1)/2d$. When two particles occupy the same site they immediately coalesce into one particle.

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

Harris (1976), Griffeath (1978).

$\zeta_t^{x,t}$ gives the set of potential ancestors of the individual at x at time t. To determine the actual ancestor we introduce an ordering on $\zeta_s^{x,t}$. Ancestor is first occupied site in the order. In population genetics, this construction is called the **ancestral selection graph.**

Krone, S.M., and Neuhauser, C. (1997) *Genetics.* 145, 519–534


in state 1 ?
state of dual
t
0
−3
−2
−1
0
1
2
δ
δ
δ
−1
−1, 0
−2, 0
−2, 1
−2, −3, 1
−2, −3, 1, 2
−1, −3, 1, 2

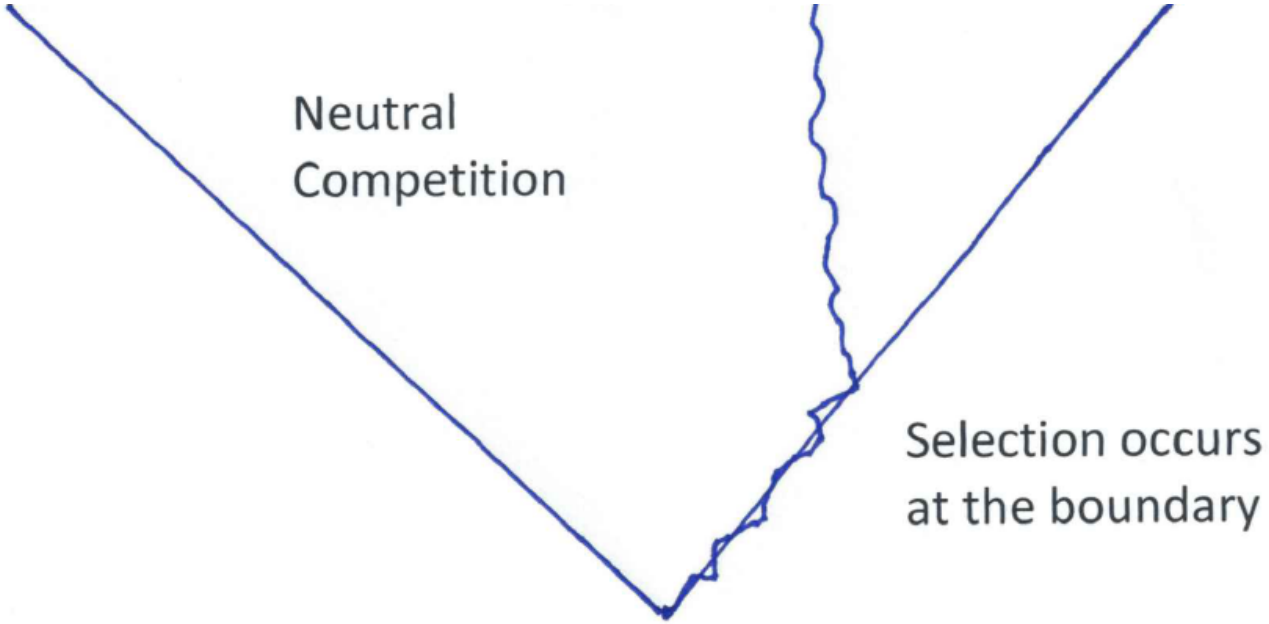
Shape theorem

Bramson and Griffeath (1980,1981) showed that if we start with a single type 1 at the origin then when ξ_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_ϵ (which depends on the outcome ω) so that on $\{T_0 = \infty\}$ we have

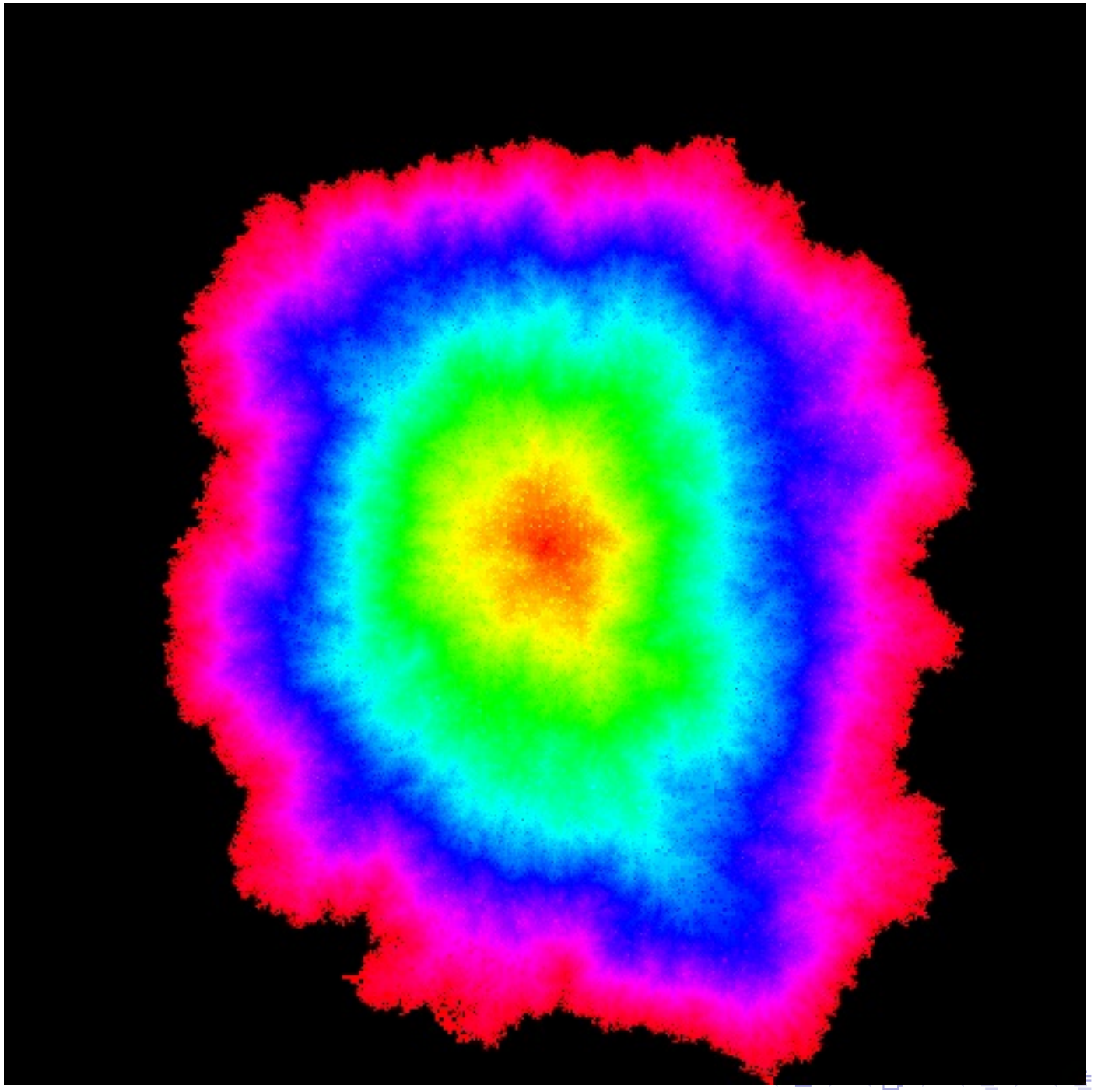
$$(1-\epsilon)tD \cap \mathbb{Z}^d \subset \xi_t \subset (1+\epsilon)tD \quad \text{for } t \geq t_\epsilon(\omega). \tag{1}$$

In the interior of the growing ball the branching arrows have no effect since they go from one site occupied by 1 to another. The genealogy becaomes a random walk, which moves by $O(\sqrt{t})$ in time t.

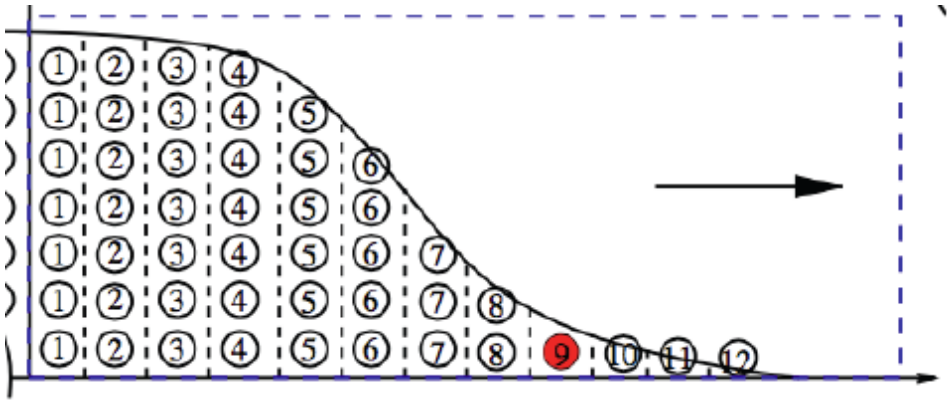
Mental Picture



By duality the genealogy does make its way back to the origin. However, it seems very difficult to study what happens at the boundary.



Motivation for Model 2: $d = 1$ scaling limit



Hallatschek and Nelson (2008) studied genealogies in a one dimensional system where sites (demes) can carry up to N individuals. and concluded that if space, time, and parameters arre rescaled appropriately then the system (observed from a reference frame moving at rate v) converged to the solution of the stochastic PDE

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

where η is space-time white noise.

Explanation of the Limit

$D\partial_x^2 u(x,t) =$ diffusion from the voter model component $\longrightarrow \delta$

$v\partial_x u(x,t) =$ drift due to moving reference frame

$su(1-u) =$ influence of selection

$\sigma\sqrt{u(1-u)}\,\eta$ fluctuation due to random reproduction

To get a random limit selection has to go to 0 at the right rate, otherwise we get a PDE

Genealogies in SPDE

Using arguments about the behavior of tracer particles placed into an expanding fluid they 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

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

If u is the solution to then SPDE, then it is not smooth enough for the drift $\partial_y \log[u(y, t)]$ to make sense mathematically. Even worse, in two dimensions SPDE on the previous slide does not have function-valued solutions.

Simplified model

Suppose the fraction of cancer cells near y at time t is

$$u(y,t) = \exp[\varphi((|y| - vt)/t^{\alpha})]$$

where $\varphi(z)$ converges to 0 as $z \to \infty$ and $\to -\infty$ as $z \to \infty$. $\alpha = 1/2$ would be central limit theorem fluctuations. Simulations of the Eden model suggest it has $\alpha = 1/3$.

By analogy with HN(2008) the coordinates of the ancestor in a fixed reference frame will be a diffusion process with generator

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

where $\tau = T - t$ since we are working backwards in time starting from x at time t.

Change to polar coordinates

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

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

If $\tau = T - t$ the radial component

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

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

$$dU_t \approx dB_t^1 + \left(v + \tau^{-\alpha}\varphi'(U_t/\tau^\alpha) \right) \, dt$$

Radial component

If we let $\psi(u) = vu + \varphi(u/\tau^\alpha)$ then we can write the radial SDE as

$$dU_t = dB_t^1 + \psi'(U_t)\, ds$$

Its generator $L_U f(u) = (1/2)f''(u) + \psi'(u)f'(u)$

$$= \frac{1}{2} e^{-2\psi(u)} \frac{d}{du}\left(e^{2\psi(u)} \frac{d}{du} f \right)$$

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, L_U is self-adjoint with respect to $e^{2\psi}$ (stationary distribution).

Concrete example

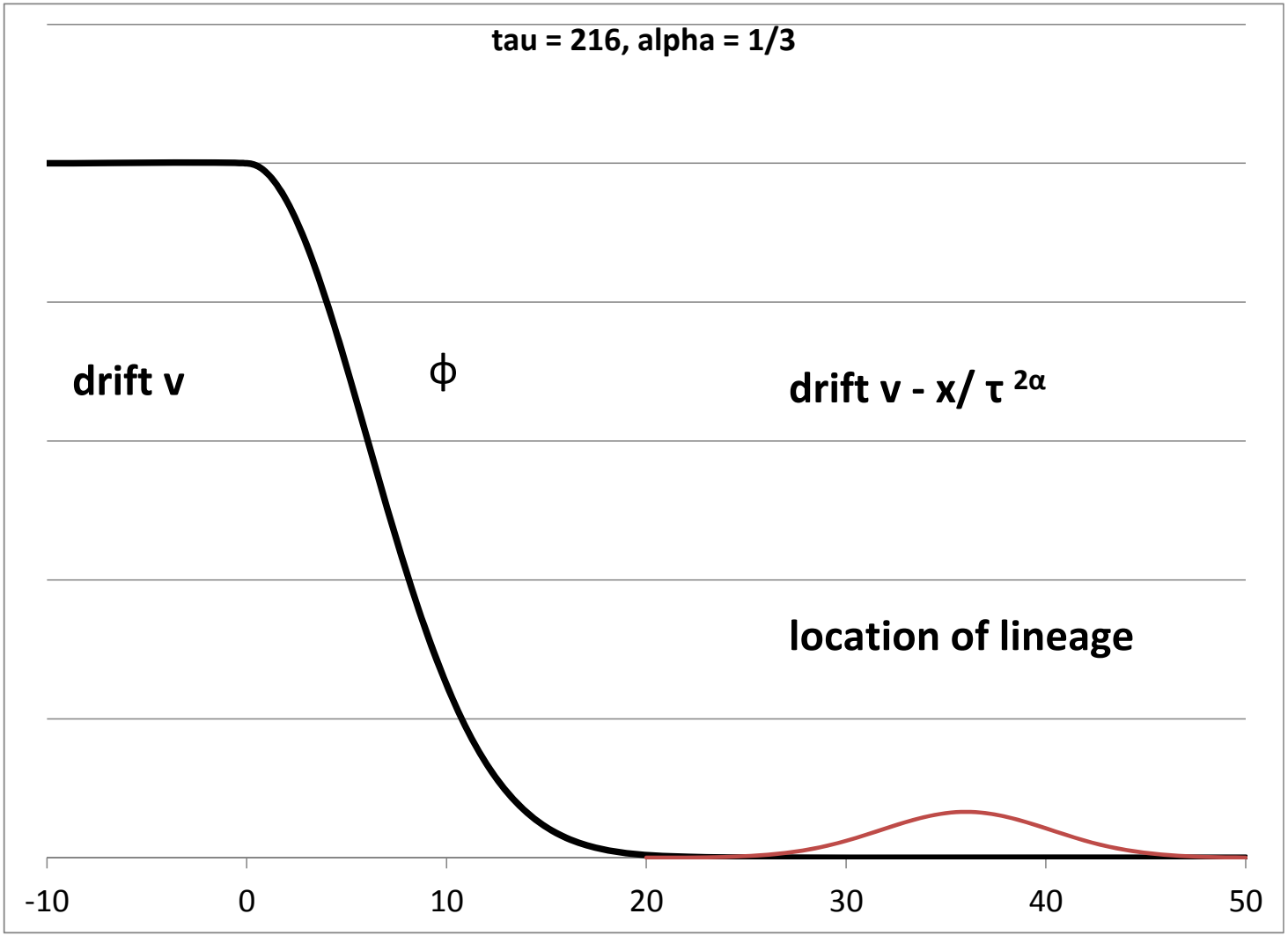
$$\varphi(u) = \begin{cases} 0 & u \le 0 \\ -u^2/2 & u \ge 0 \end{cases} \qquad 2\psi(u) = \begin{cases} 2vu & u \le 0 \\ 2vu - u^2/\tau^{2\alpha} & u \ge 0 \end{cases}$$

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

If $\alpha > 1/2$ the mean at time $s = 0$ is $T^{2\alpha} \gg T = R$ (recall $v = 1$) so the distance from the edge is larger than the tumor radius which makes no sense. $(\alpha = 1/2??)$

If $\alpha \le 1/2$ then $X_t = U_t - \tau^{2\alpha}$ is an Ornstein-Uhlenbeck process with a time dependent drift

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


tau = 216, alpha = 1/3
drift v
ϕ
drift v - x/ τ 2α
location of lineage
-10
0
10
20
30
40
50

Discretization

Two Brownian motions in $d = 2$ won't hit it each other so we have to consider the model on $\epsilon\mathbb{Z}^2$ to allow lineages to coalesce. To fix units we will think of

space is measured in cm

time is measured in years

cell have diameter $\approx 10\mu m = 10^{-3} cm$, so $\epsilon = 10^{-3}$

for simplicity we set $v = 1$

In this talk I'll ignore the discretization.

Coalescence: lower bound

$$\theta_t - \theta_0 \approx \text{normal}(0, \sigma_t^2) \quad \text{where}$$

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

Conclusion 1. *Consider two points on the boundary whose angles from the center differ by* $\theta_0 = aR^{-1/2}$. *The time for the lineages to hit has mean* $\geq a^2 R/(1+a^2)$.

$\theta_t - \sigma_t^2$ is martingale. $t/R(R-t) = a^2 R^{-1}$ when $t = a^2 R/(1+a^2)$.

$a = 3$, $t = 0.9R$. $R = 5000$ cells or 5 *cm*. $3R^{1/2} \approx 210$ cells.

Comparison with data

Kang et al. Two 0.5 cm^3 bulk samples from a 6 cm adenoma

mitotic ages were estimated as 250–1130 cell divisions

Big bang data. Adenomas (sizes in cm) 2.5, 3.5, 6, 6

Carcinomas: 1.8, 2, 2.3, 3.0, 3.4, 3.5, 3.5, 3.9, 4, 5, 5.6, 5.7, 6.1, 6.4, 9.5

Complication: tumor glands (crypts) have 10,000 cells and are typically clonal since the population of cells is produced by a small number of stem cells.

Coalescence: two lineages on the boundary

Let X_t be the radial part and scale the angular part to $Y_t = R\theta_t$.

Consider two starting points $X_1(t), Y_1(t)$ and $X_2(t), Y_2(t)$ on the edge of the tumor, in a ball of radius R^β

$$T_c = \min\{t : (X_1(t), Y_1(t)) = (X_2(t), Y_2(t))\}.$$
$$T_\theta = \min\{t : Y_1(t) = Y_2(t)\}$$

Conclusion 2. If $\alpha < \beta \le 1/2$ then

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

Since $T_\theta = O(R^{2\beta})$ the second term is smaller. Most of the lineages in the ball of radius R^β will coalesce at a time $o(R)$.

OU equilibrates in time $\tau^{2\alpha}$

Consider first times $t \ll T$ let $\beta = 1/\tau^{2\alpha}$ to simplify notation.

$$dX_t = dB_t - \beta X_t\, dt \quad \text{has solution}$$

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

From this we see that when $X_0 = x$, X_t is normal with mean $e^{-\beta t}x$ and variance

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

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

Proof of Upper Bound

Since $T_\theta \gg R^\alpha$

$$P(\bar{X}_1(T_\theta + s) = \bar{X}_2(T_\theta + s)) \approx C_1 \epsilon R^{-\alpha}$$

$$E \int_0^t P(Z_1(T_\theta + s) = Z_2(T_\theta + s))\, ds = C_1 \epsilon R^{-\alpha} \cdot \epsilon \int_0^t (2\pi s)^{-1/2}\, ds$$

$$= C_2 \epsilon^2 R^{-\alpha} t^{1/2}$$

The result now follows from

$$P_{x_1,x_2}(T_c \le t) \ge \frac{\int_0^t P_{x_1,x_2}(Z_1(T_\theta + s) = Z_2(T_\theta + s))\, ds}{\int_0^t P_{x_1=x_2}(Z_1(T_\theta + s) = Z_2(T_\theta + s))\, ds}$$

Denominator is $C_3 \epsilon^2 \log t$. Take $t = R^{2\alpha} \log^2 R$.

Coalescence: one boundary, one interior

If a lineage starts at bR then until time $(1-b)R$ it and its nearby lineages will do two dimensional a coalescing random walk. A result of Bramson and Griffeath implies that the density of particles decays to 0 like $c(\log t)/t$. Thus at time $(1-b)R$ a walker will contain $O(R/\log R)$ lineages but they will be scattered throughout the tumor and will not be detectable by a biopsy.

Even if the boundary lineage starts at the exact same angle the two will be separated by $\approx \sqrt{(1-b)R}$ when they are at the same radius. By earlier results they will take roughly time $R(1-b)/(2-b)$ to coalesce.

Problems with the model

In $d = 3$ if two cells are separated by $\log R$ at time t then with high probability they coalesce at a time $o(R)$ from time 0. (Three dimensional random walk is transient.)

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 $\varphi(u) = u^{\gamma}/\gamma$ keeps genealogies closer but still the density at their location is low.

A better model?

Thinking of the competing yeast species, we seed a small region of the grid with a number of cells of different colors. Given an occupied site x and a vacant neighbor y, x gives birth at rate λ to a cell of its color that is placed at y. We draw an arrow from y to x indicate where it came from.

This looks like a very simple model but the genealogies are geodesics in first passage percolation, which are not understood after 50 years of trying.

Determining the size of the fluctuations of the boundaries between sectors is an open mathematical question. Physicists (see e.g., Derrida and Dickman (1991) J. Phys A. 24, L191–L195) tell us that the fluctuation exponent $\chi = 2/3$ in $d = 2$.