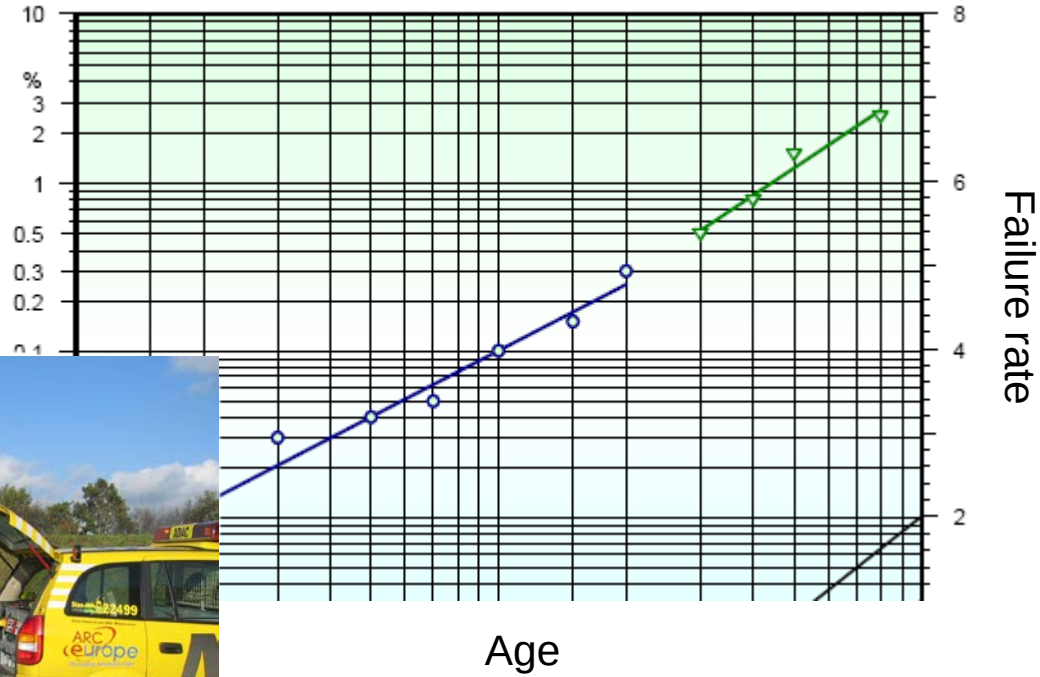
ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

D-BSSE
Department of Biosystems
Science and Engineering

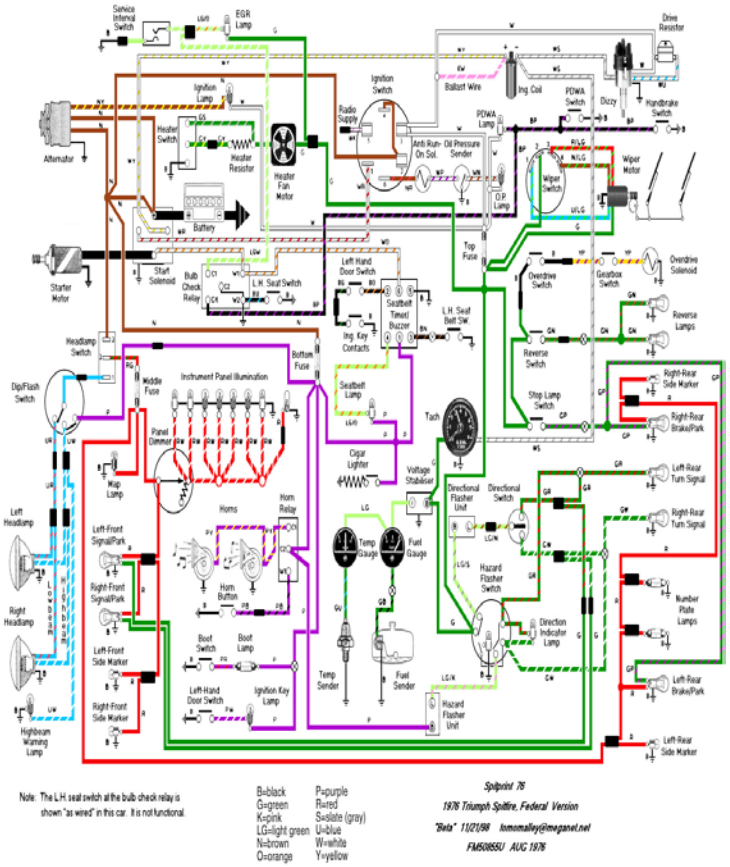
Cancer progression

Niko Beerenwinkel

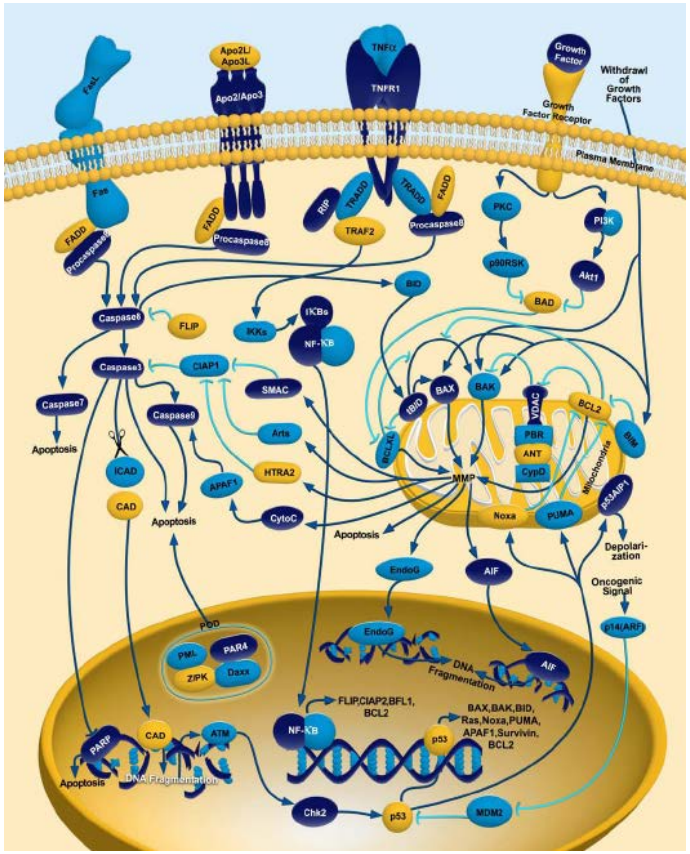
Car breakdown



Breakdown of complex systems

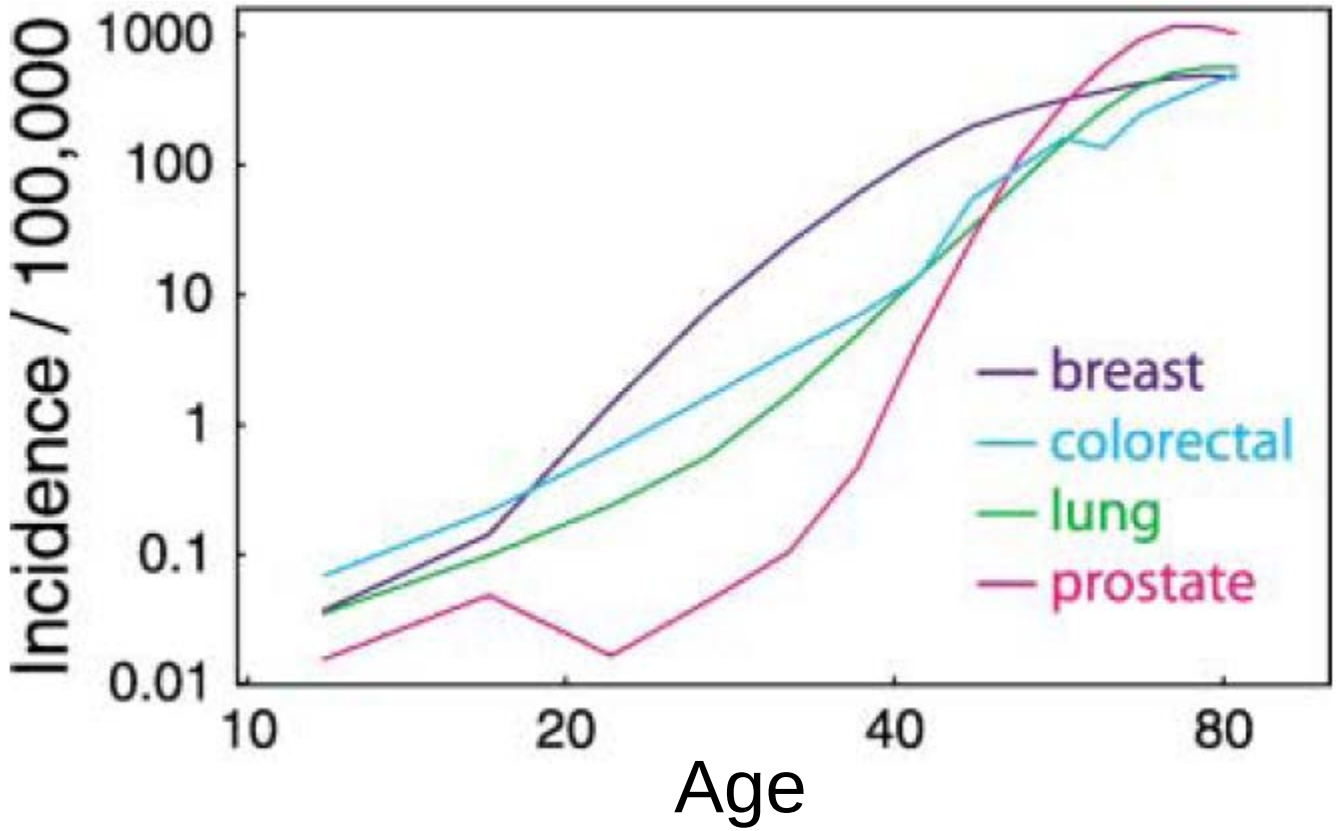


Electrical circuit

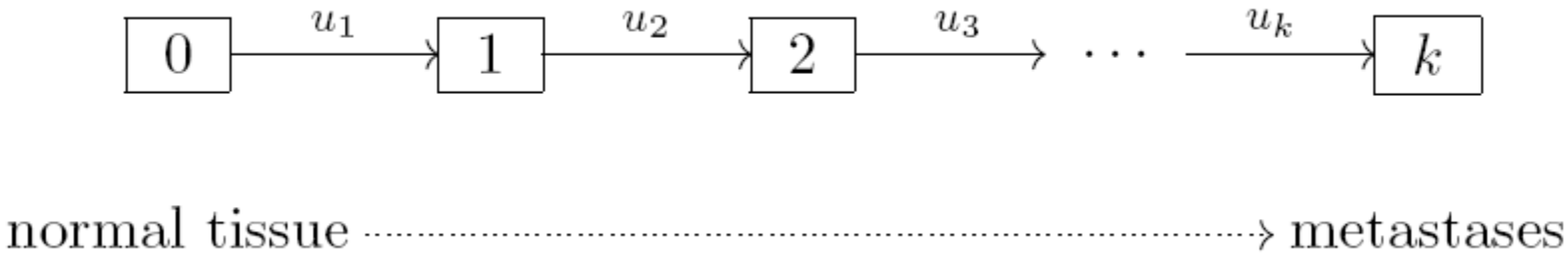


Biological signaling pathway

Cancer incidence data



Multistage theory



- u_j = u small $\Rightarrow$ Prob(one step by time t) = $1 - e^{-ut} \approx ut$.
- Then the incidence is $I_k(t) = (ut)^{k-1}u = u^k t^{k-1}$ and

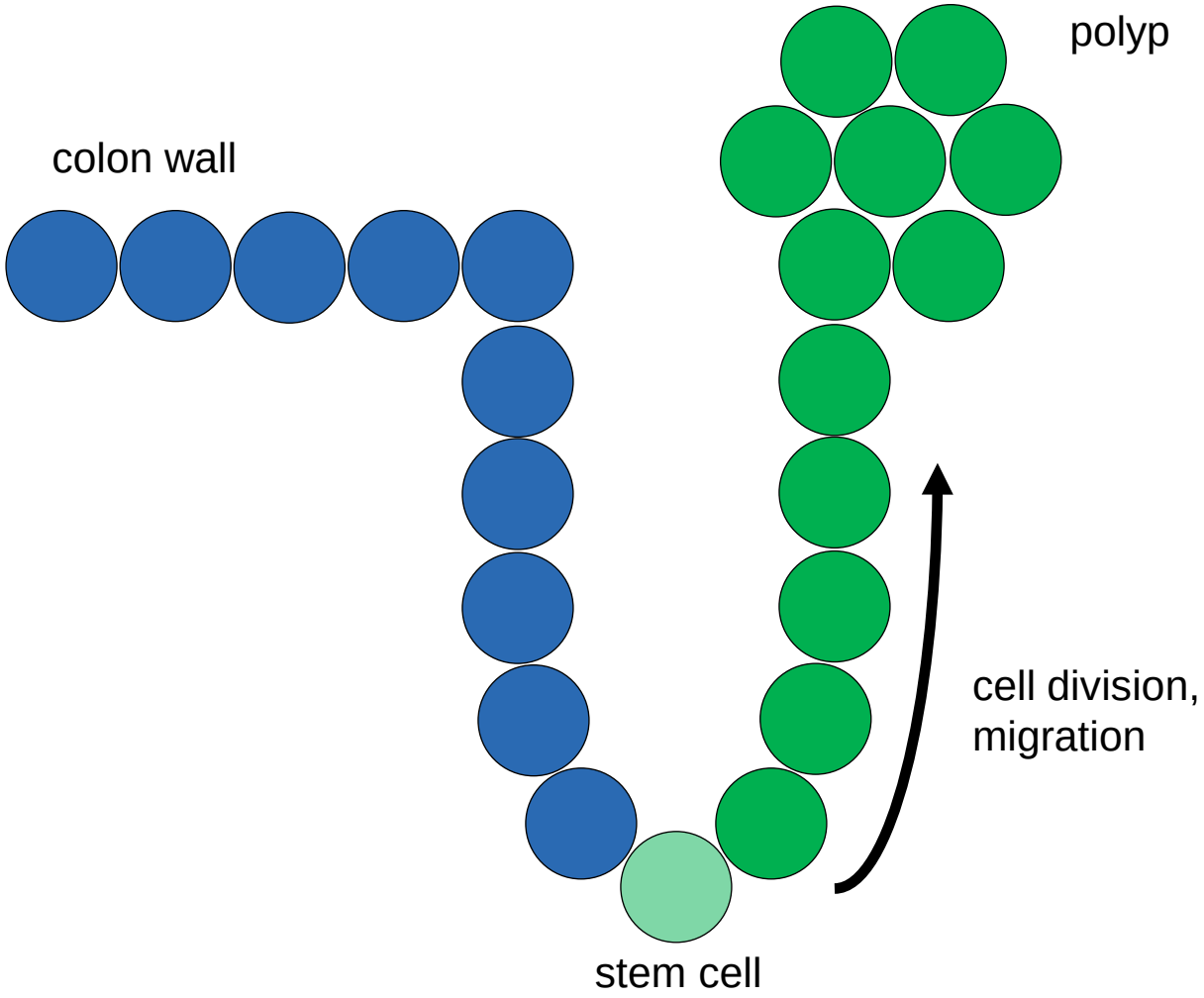
$$\log I_k = k \log u + (k-1) \log t$$

Armitage & Doll 1954, Nordling 1953

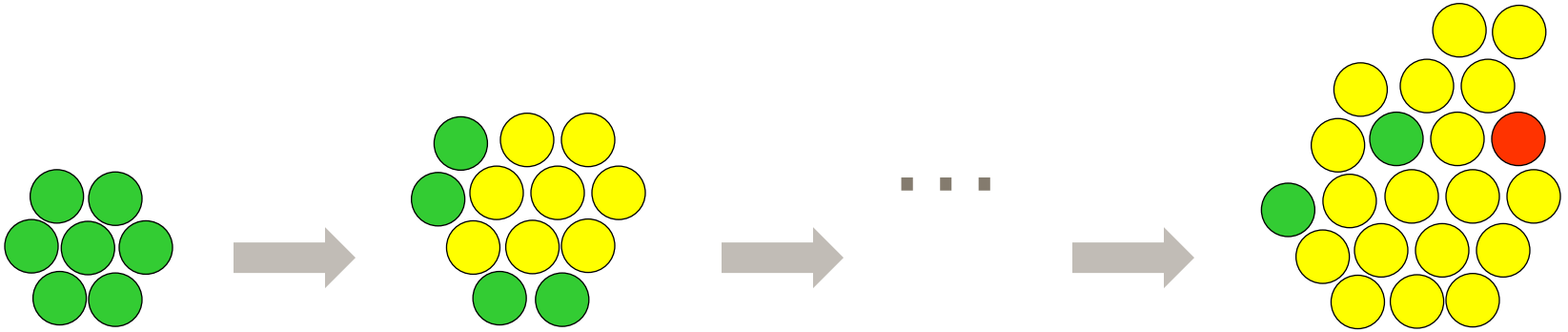
Outline

- Colon cancer
- Multistage theory, Waiting times
- The Wright-Fisher process
- The speed of adaptation

Polyp formation in a colonic crypt



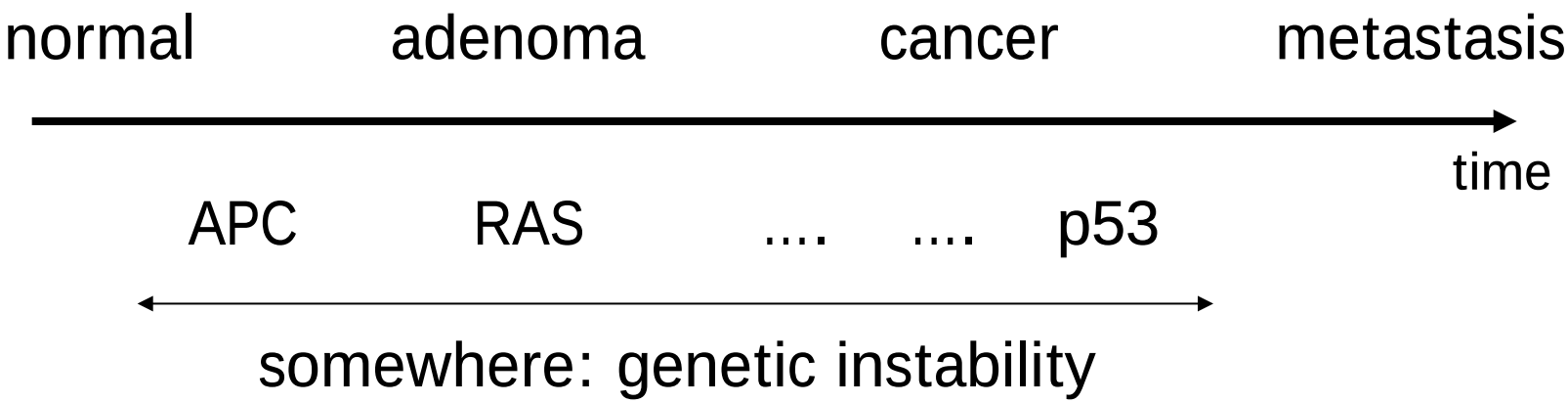
Tumor progression



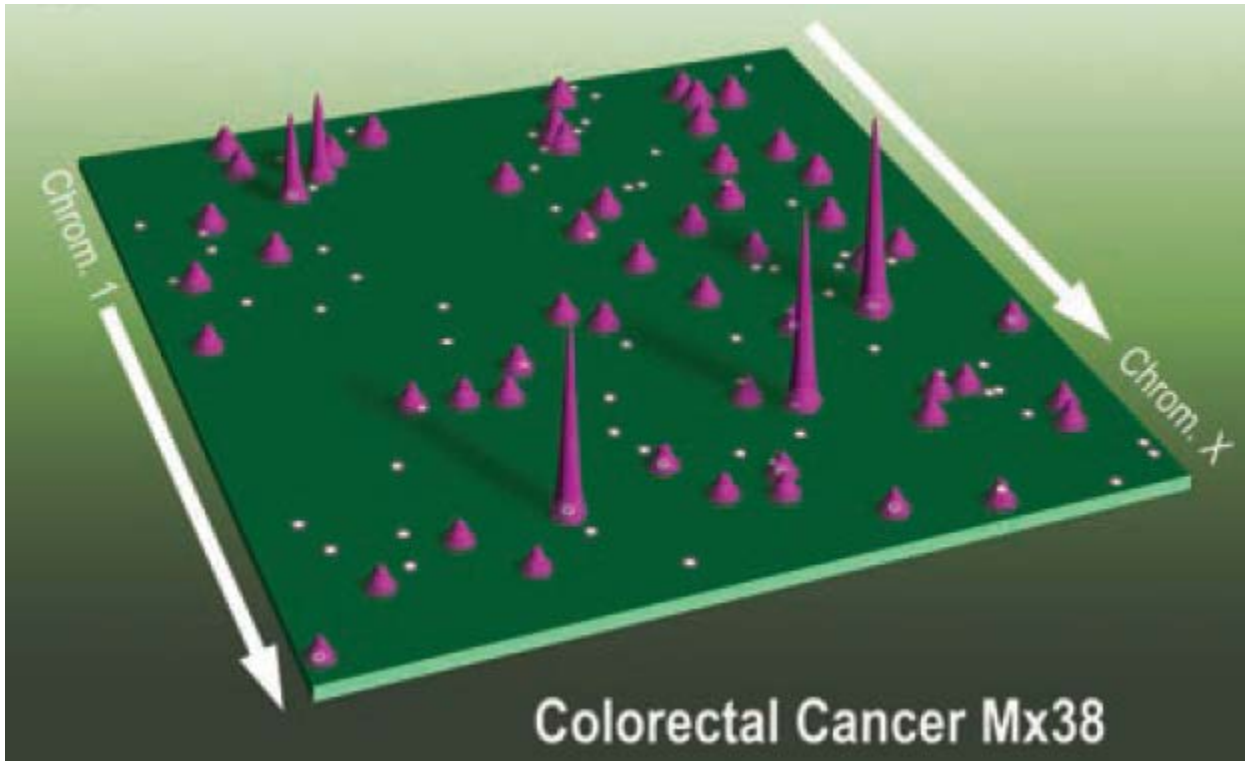
adenoma (polyp)

carcinoma

Genetic progression of colon cancer

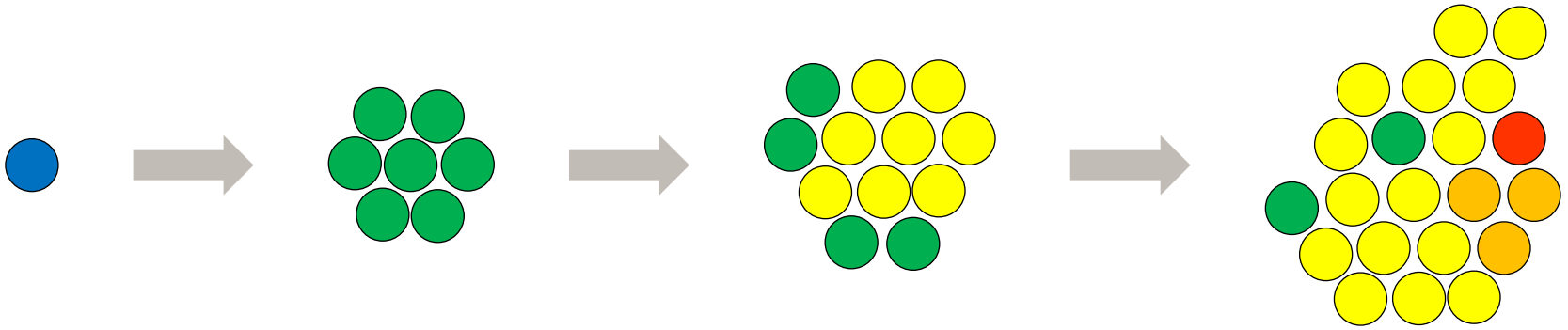


The mutational landscape of colon cancer: few mountains, many hills



- In each patient, 15 to 20 mutated genes seem to drive progression.
- This set of genes differs considerably among patients.

Genetic progression of cancer



adenoma

carcinoma

1 cell

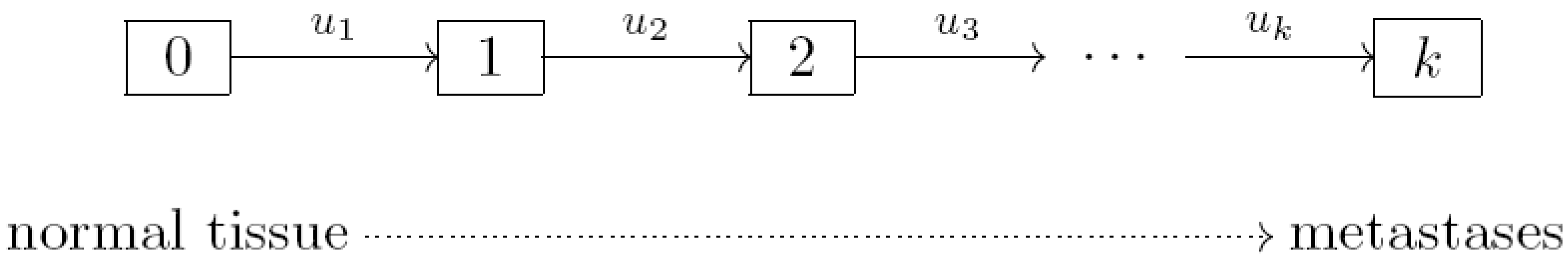
10^6 cells

10^9 cells

few specific mutations

many unspecific mutations

Multistage theory



- Tumorigenesis is assumed to be a linear, multi-step process.
- Transitions from state j – 1 to state j are rare, i.e., u_j is small.
- The waiting time for each step is exponential, Exp(u_j)
- Let τ_k be the waiting time until stage k is reached. Then

$$\tau_k \sim \mathrm{Exp}(u_1) + \cdots + \mathrm{Exp}(u_k)$$

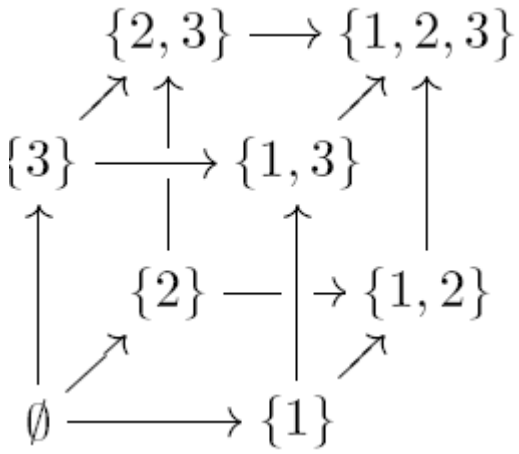
The expected waiting time

$$
\begin{aligned}
E[\tau_k] &= E\left[\sum_{j=1}^{k} \mathrm{Exp}(u_j)\right] \\
&= \sum_{j=1}^{k} E[\mathrm{Exp}(u_j)] \\
&= \sum_{j=1}^{k} \frac{1}{u_j}
\end{aligned}
$$

Independent mutations

- Each mutation occurs independently at time

$$T_j \sim \text{Exp}(\lambda_j)$$



Waiting time to any *k* independent mutations

- Let τ_k be the waiting time until any *k* out of *d* mutations have occurred,

$$\tau_k = \min_{\{j_1,\ldots,j_k\}\subset\{1,\ldots,d\}} \max\{T_{j_1},\ldots,T_{j_k}\}$$

- For $k = 1$, we have

$$\tau_1 = \min\{T_1,\ldots,T_d\} \sim \mathrm{Exp}(\lambda_1 + \cdots + \lambda_d)$$

Independent mutations, identical rates

- If $\lambda_1 = \ldots = \lambda_k = \lambda$, then

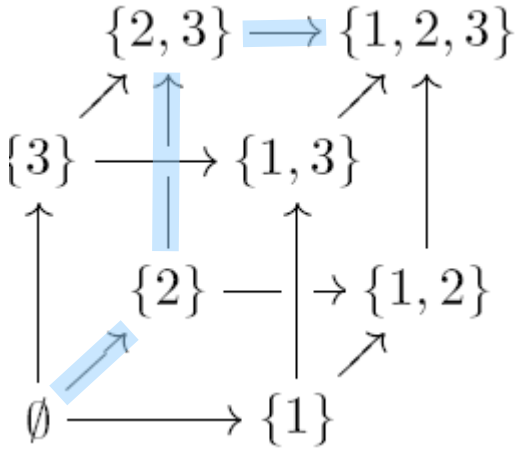
$$\begin{aligned} \tau_1 &\sim \text{Exp}(d\lambda) \\ \tau_2 &\sim \tau_1 + \text{Exp}((d-1)\lambda) \\ &\vdots \\ \tau_j &\sim \tau_{j-1} + \text{Exp}((d-j+1)\lambda) \end{aligned}$$

- Hence,

$$E[\tau_k] = \frac{1}{\lambda} \sum_{j=1}^{k} \frac{1}{d-j+1}$$

Mutational pathways

- Each total order of mutations
  $j_1 < \ldots < j_k$
  defines a mutational pathway in the genotype lattice.
- For a fixed path, let Exit_i denote the set of all possible mutations in step i.
- For example,
  - $\text{Exit}_1 = \{1, 2, 3\}$
  - $\text{Exit}_2 = \{1, 3\}$
  - $\text{Exit}_3 = \{1\}$

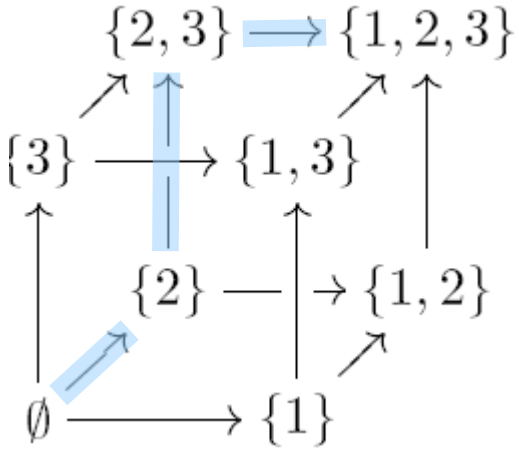


Probability and waiting time of pathways

- The probability of a pathway $P = j_1 \to \ldots \to j_k$ is

$$\text{Prob}(P) = \prod_{i=1}^{k} \frac{\lambda_{j_i}}{\sum_{j \in \text{Exit}_i} \lambda_j}$$

(competing exponentials)



- The expected waiting time of P is

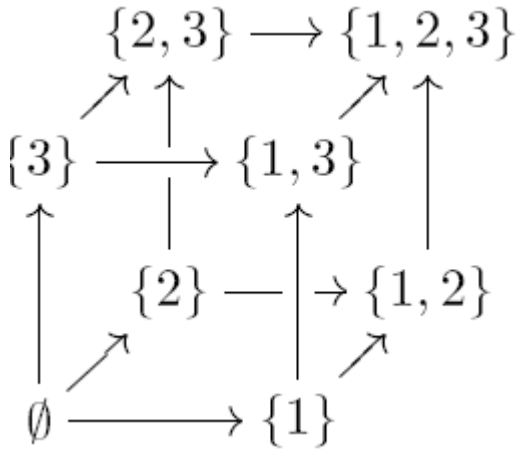
$$E(\tau_P) = \sum_{i=1}^{k} \frac{1}{\sum_{j \in \text{Exit}_i} \lambda_j}$$

Independent mutations, arbitrary rates

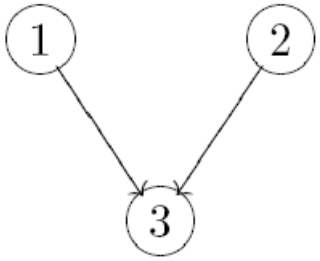
- The expected waiting time of k out of d independent mutations is

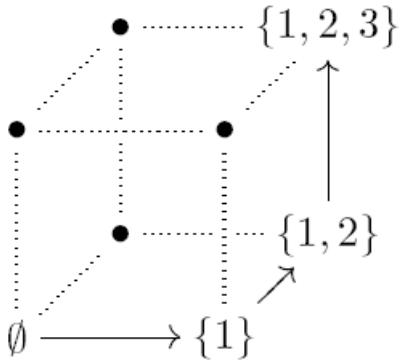
$$E(\tau_k) = \sum_{P=j_1\to\cdots\to j_k} \mathrm{Prob}(P)\, E(\tau_P)$$

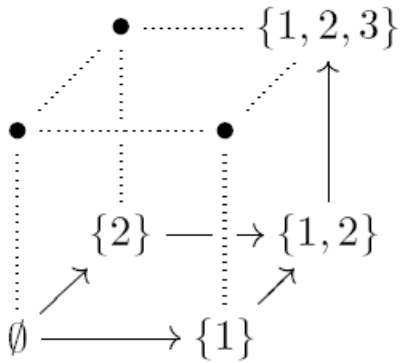
where P runs over all pathways of length k starting from the wild type.

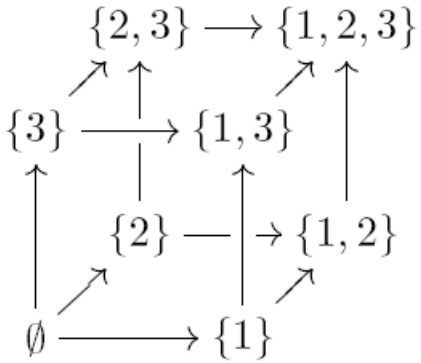


Posets define the geometry of genotype space







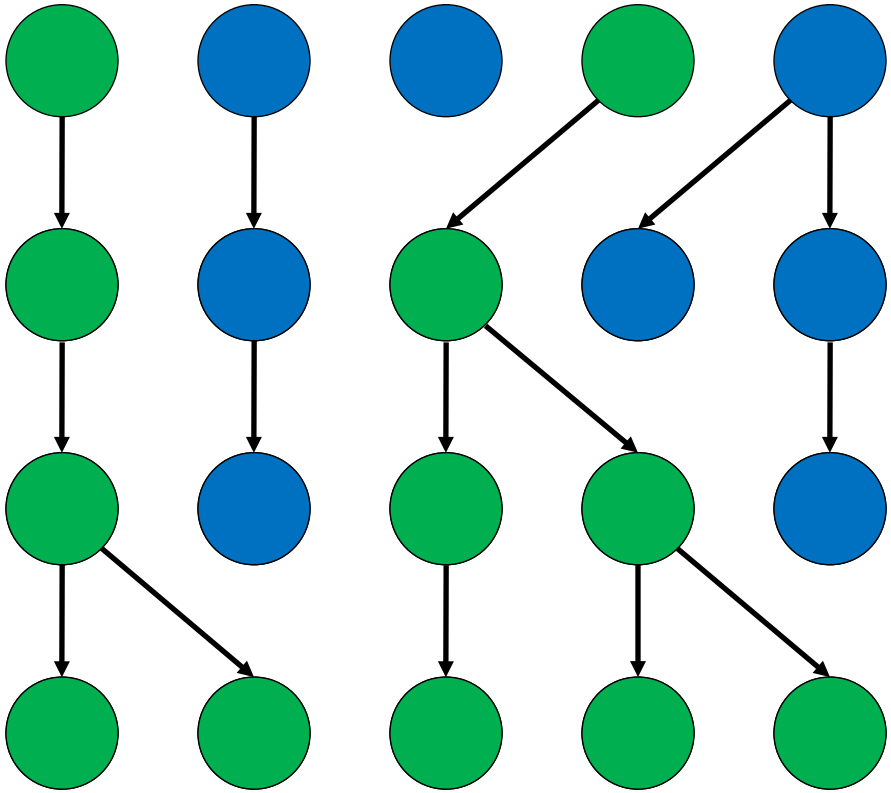


Linear order
(multistage theory)

Partial order

Independence

The Wright-Fisher process



generation 1

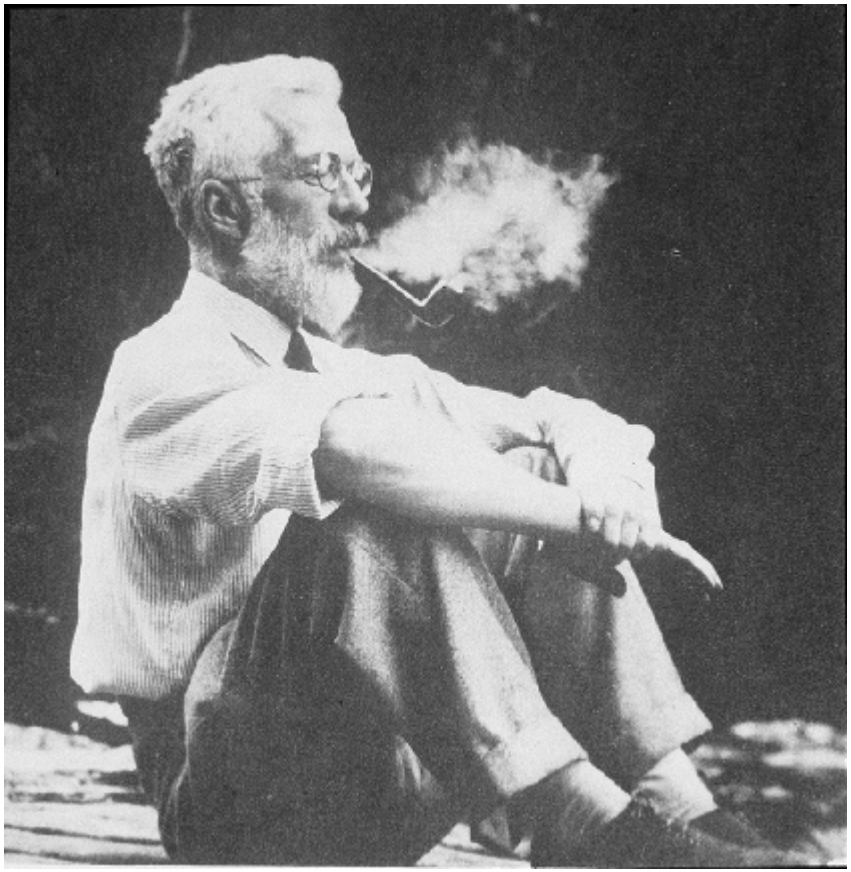
generation 2

generation 3

generation 4

Wright-Fisher model

Sewall Wright (1889-1988)



Ronald A. Fisher (1890-1962)

The Wright-Fisher process defines a Markov chain

- Consider a haploid population of constant size N.
- There are two different types, A and B.
- Reproduction occurs in discrete, non-overlapping generations, i.e., individuals are synchronized.
- Let X(t) be the number of type A individuals in generation t = 0, 1, 2, 3, …
- X(t) has state space {0, 1, …, N}.

Binomial sampling

- Each generation is sampled from the previous generation according to the binomial distribution,

$$(X(t+1) \mid X(t) = i) \sim \text{Binom}(N, i/N)$$

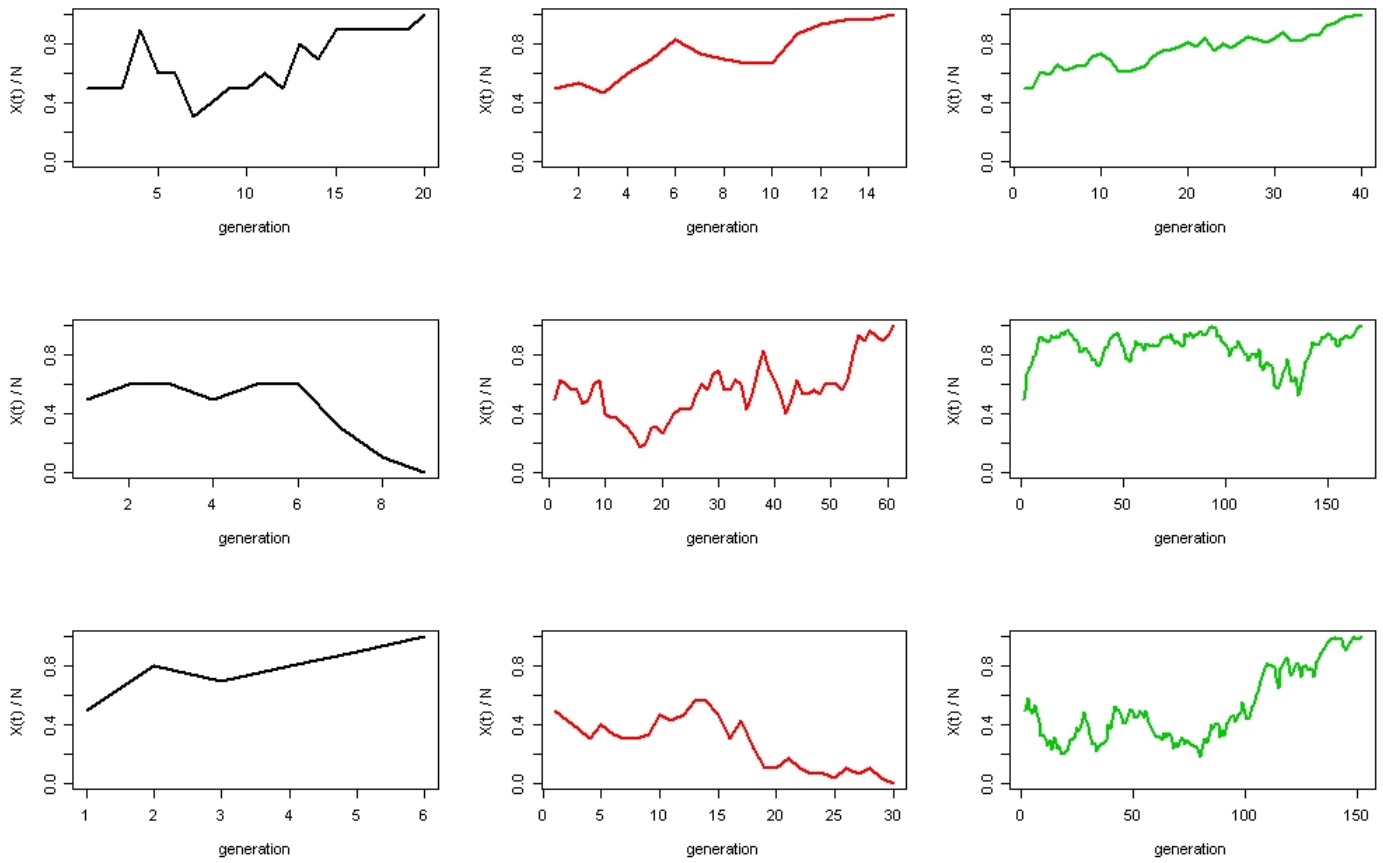
- The transition probabilities are given by

$$P_{ij} = P(X(t+1) = j \mid X(t) = i)$$

$$= \binom{N}{j} \left(\frac{i}{N}\right)^j \left(\frac{N-i}{N}\right)^{N-j}$$

Dynamics

N = 10 N = 30 N = 100



Properties of the Wright-Fisher process

- Expectation: $E[X(0)] = X(0) = i$
- $X(1) \sim \text{Binom}(N, p=i/N)$, hence $E[X(1)] = Np = i$.
- The Wright-Fisher process is unbiased: the average frequency of type A individuals does not change over time, $E[X(t)] = i$ for all $t = 0, 1, 2, \ldots$

- The variance of $X(t)$ is $Ni(1 – p)[1 – (1 – p)^t]$.
- The frequency of type A individuals is subject to random fluctuations.

- The Wright-Fisher process models random genetic drift.

Fixation probabilities

- The Wright-Fisher process has the two absorbing states, X(t) = 0 and X(t) = N.
- Let $x_i = \lim_{t \to \infty} P[X(t) = N \mid X(0) = i]$ be the probability of fixing type A when starting with i copies of it.
- We have

$$i = \lim_{t \to \infty} \mathrm{E}[X(t)] = 0 \cdot (1 - x_i) + N \cdot x_i$$

and therefore $x_i = i/N$.

Mean fixation times

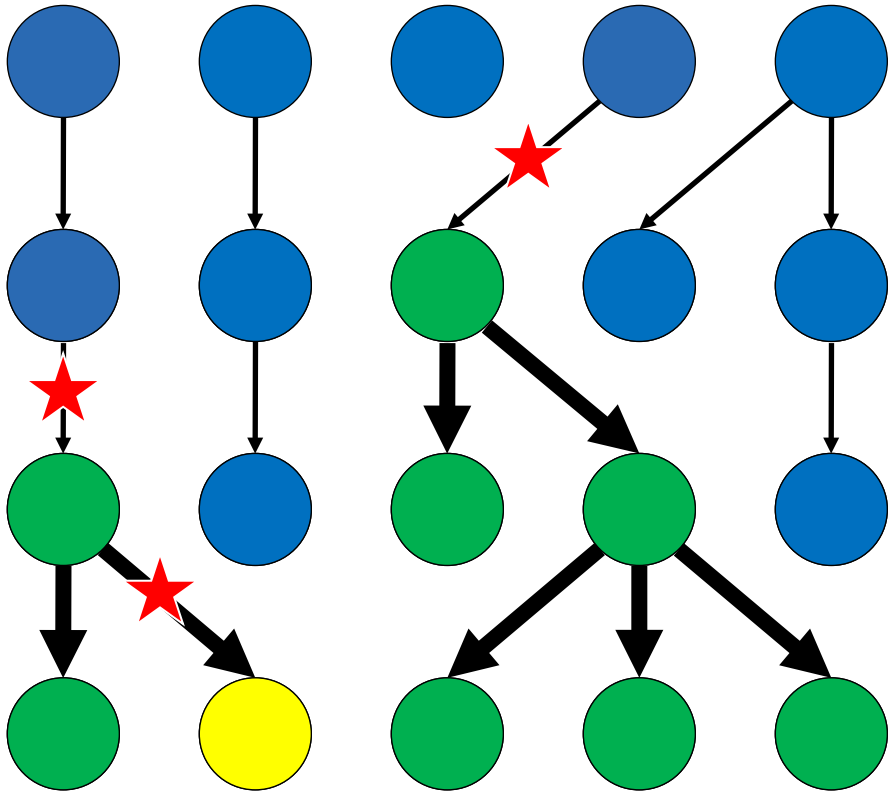
- Let k_i be the mean number of generations before either absorbing state is reached when starting in state i.
- No simple way of calculating k_i, even approximately, is known. We will later use diffusion theory to show that

$$k_1 \approx 2\,[\log(N-1) + \gamma]$$

where γ = 0.5772… is Euler's constant.

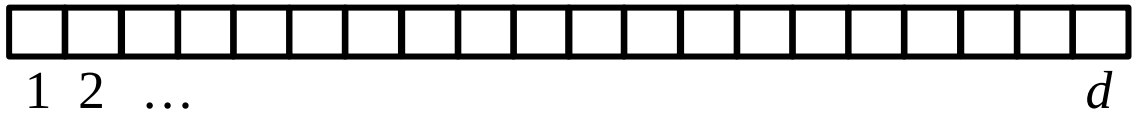
Extensions of the Wright-Fisher process

- Multiple types
- Mutation
- Selection



Wright-Fisher process for accumulating mutations

- We consider a binary genome of length d.



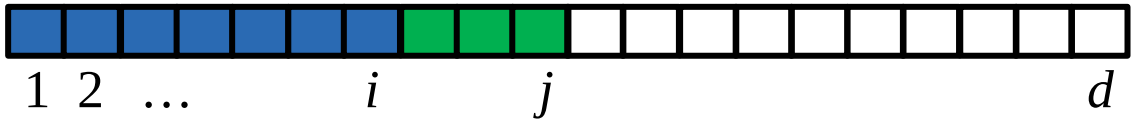
- Each locus (independently) undergoes mutation from 0 to 1 at rate u. We ignore back mutations from 1 to 0.
- Let $X_j(t)$ be the number of cells with j mutations (called *j-cells*) in generation t. Set $x_j(t) = X_j(t) / N$.
- Initially, the population is homogeneously wild type (0-cells), i.e., $X_0(0) = N$, and $X_1(0) = \ldots = X_d(0) = 0$.
- We assume a constant fitness advantage, s, per mutation. Thus, the fitness of a j-cell is proportional to $(1 + s)^j$.

What is the probability of sampling a j-cell?

$$\theta_j(t) = \sum_{i=0}^{j} P(\text{i-cell} \rightarrow \text{j-cell})$$

$$= \sum_{i=0}^{j} P(\text{i-to-j mutation}) P(\text{i-cell parent})$$

$$= \sum_{i=0}^{j} \binom{d-i}{j-i} u^{j-i} (1-u)^{d-j} \frac{(1+s)^i x_i(t)}{\sum_\ell (1+s)^\ell x_\ell(t)}$$



Multinomial sampling

- The transition probabilities

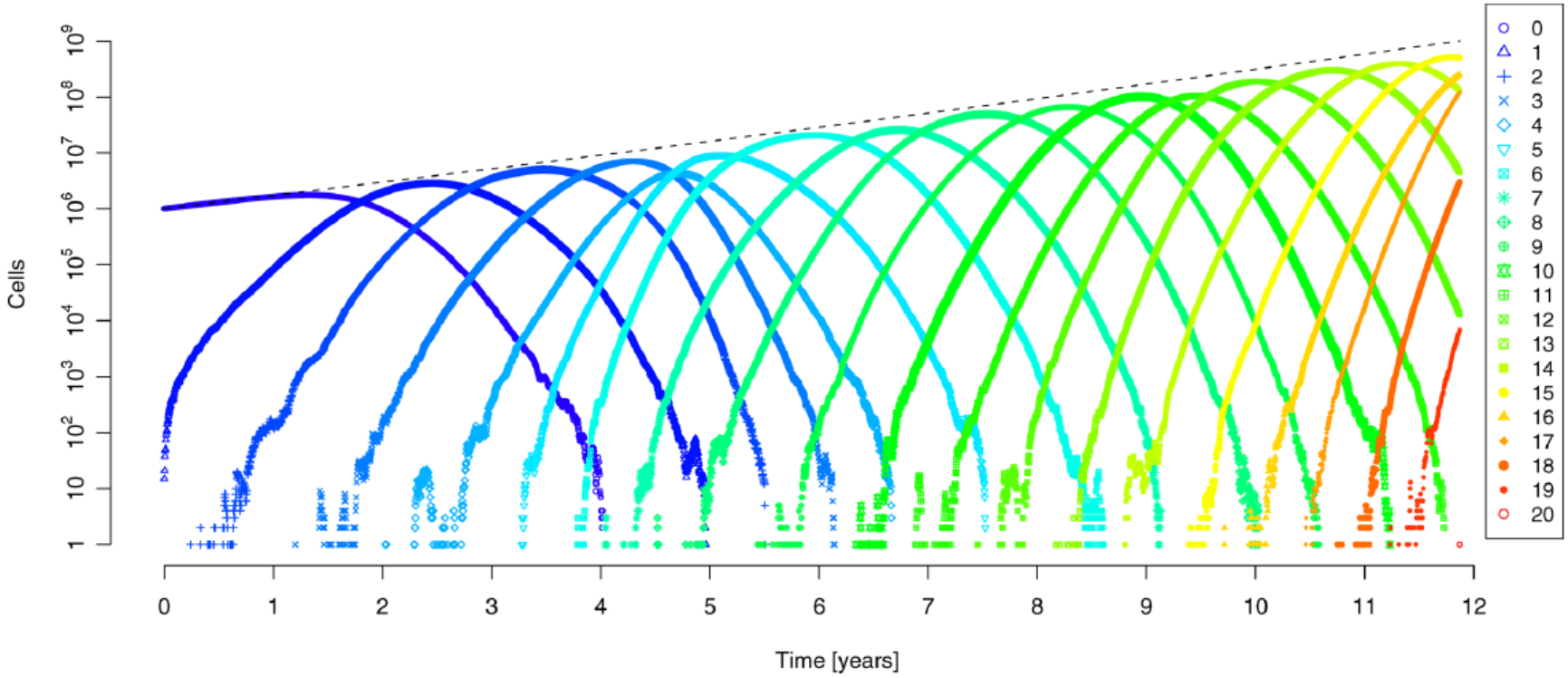
$$P_{m,n} = P[X(t+1) = n \mid X(t) = m]$$

of the Markov chain X(t) = (X_0(t), …, X_d(t)) are given by the multinomial distribution,

$$\frac{(n_0 + \cdots + n_d)!}{n_0! \ldots n_d!} \prod_{j=1}^{d} \theta_j(t)^{n_j}$$

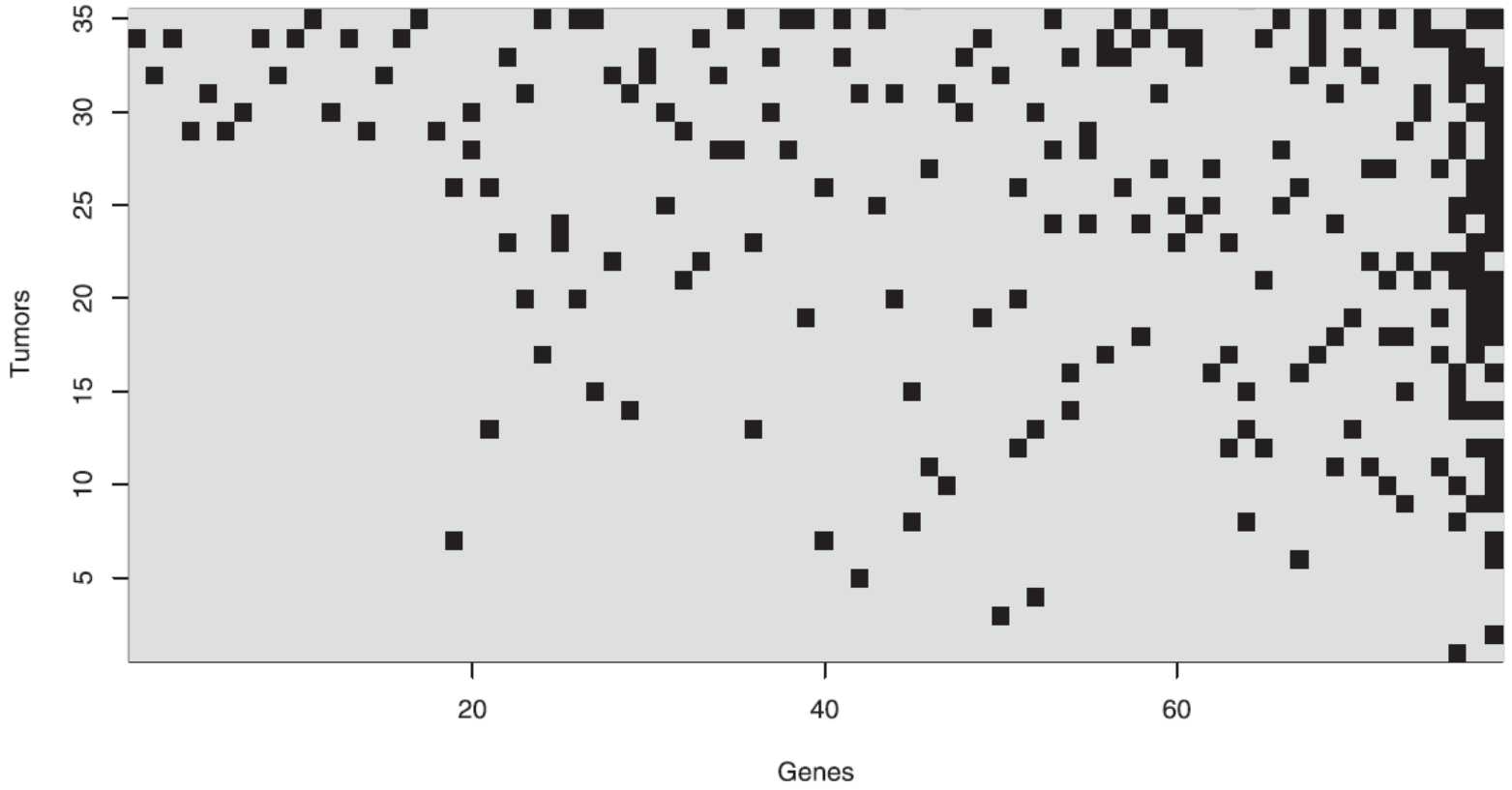
where n = (n_0, …, n_d) and n_0 + … + n_d = N.

Dynamics



- The dynamic behavior is complex, but shows some regularity: mutant waves travel at constant speed.

The waiting time to cancer

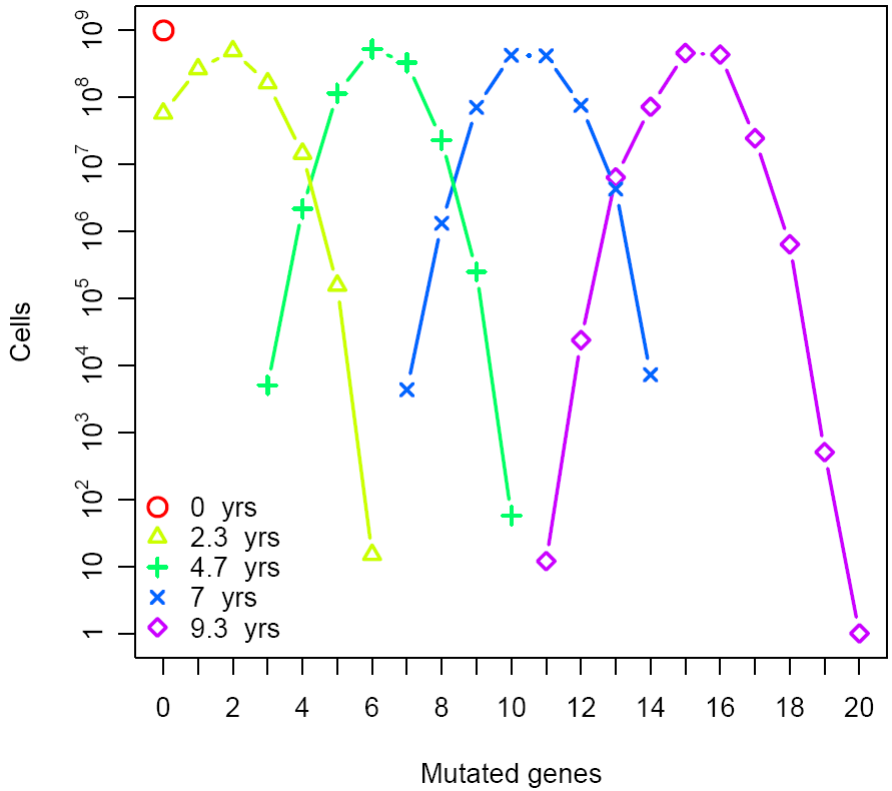


- How long does it take until the first cell with any 20 out of 100 mutations occurs?

Approximating the average waiting time

- We assume:
  - each mutant wave has a Gaussian shape
  - waves travel at constant speed
  - $s \ll 1$, so $(1+s)^j \approx 1 + sj$
- Ansatz:

$$x_j(t) = Ae^{-\frac{(j-vt)^2}{2\sigma^2}}$$



Determining $x_j(t) = Ae^{-\frac{(j-vt)^2}{2\sigma^2}}$

- $A \approx 1/\sqrt{2\pi}\sigma$ by the case of continuous j.
- The two unknowns, v and σ, are determined by decoupling clonal expansion (driven by selection) and generation of new types (by mutation).
- Clonal expansion is governed by the replicator equation

$$\dot{x}_j = s x_j \left[j - \sum_{i=0}^{d} i\, x_i(t) \right]$$

- Substituting the expression for $x_j(t)$ yields the relation

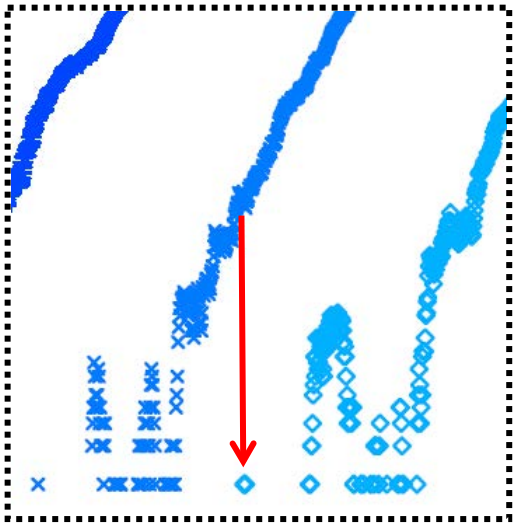
$$v = s\sigma^2$$

Generating a new mutant

- Let τ be the average time it takes to produce a new mutant.
- The velocity $v = 1/\tau$ is found by solving

$$\frac{1}{N} = x_{j+1}(\tau) = ud \int_0^\tau x_j(t)dt$$

where we use the fact that initially x_j grows exponentially.



The speed of adaptation

- For the velocity, v, we eventually find

$$v \approx \frac{2s \log N}{\log \left[s/(ud)\right]^2}$$

- This is (an approximation of) the speed of adaptation in an asexual population evolving according to the Wright-Fisher process.
- The Moran process leads to very similar approximations.
- Essentially, $v \propto s$.

The waiting time to cancer

- The average time it takes until the first cell with k mutations appears is approximately

$$\tau_k \approx \frac{k \log \left[s/(ud)\right]^2}{2s \log N}$$

- Thus, $\tau_k \propto k / s$. The waiting time is linear in k.
- The selective advantage, s, has a much larger impact on the waiting time than the mutation rate, u, or the population size, N.

Approximation versus simulation

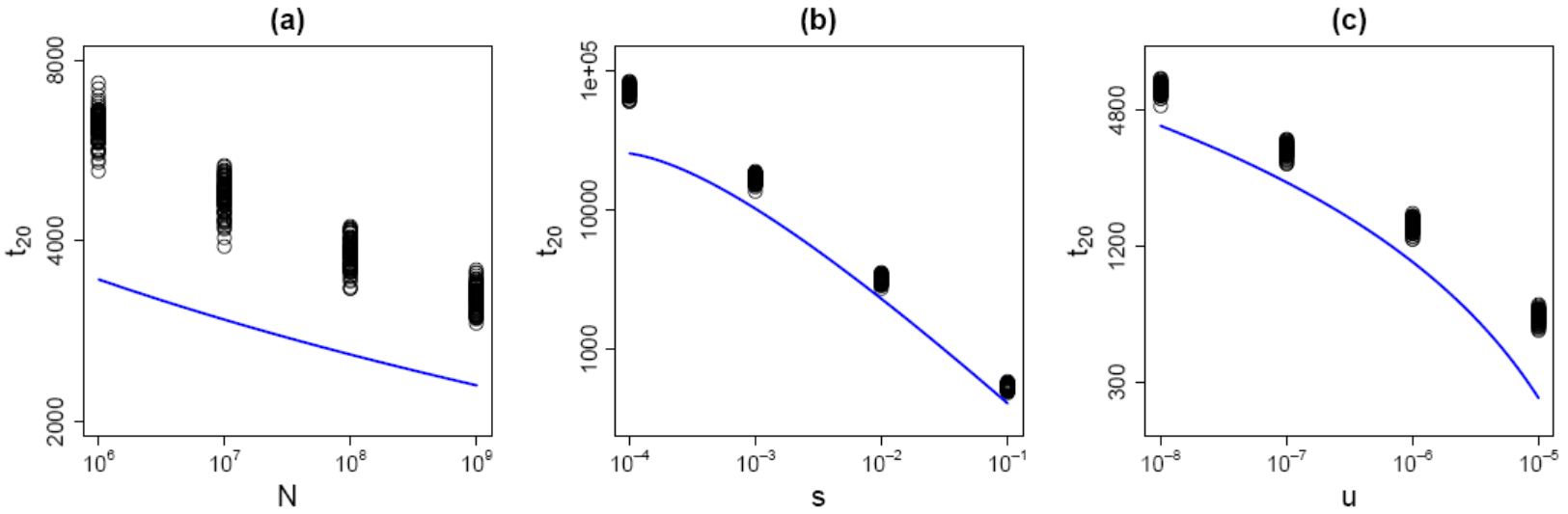


FIG. 2: Expected waiting time for a cell with 20 mutantions, t_{20}, as a function of (a) the population size N, (b) the selective advantage s per mutation, and (c) the per-locus mutation rate u. The circles are the results of 100 independent simulations at each parameter set. We always assumed $d = 100$ sensitive loci, and set $N = 10^9$ in (b) and (c), $s = 0.01$ in (a) and (c), and $u = 10^{-7}$ in (a) and (b). The solid curves correspond to the analytic approximation (17).

Summary

- Cancer progression is driven by the accumulation of many mutations in the genome.
- The waiting time to cancer depends on the mutation rate and the fixation rate (i.e., the rate of evolution), and on the number of available mutational pathways.
- The Wright-Fisher process is a stochastic process that describes the evolution of an asexual population and, in particular, the (neutral) allele sampling variance.
- The waiting time to cancer is dominated by the selective advantage per mutation.

- Exercises: #13, #14, #15, #17

References

- Beerenwinkel N, Schwarz RF, Gerstung M, Markowetz F. Cancer Evolution: Mathematical Models and Computational Inference. *Systematic Biology* 64(1):e1–e25, 2015.