1.7 Population genetics of Cancer

In this section, the general results obtained so far are used to develop a simple model for progression of cancer. The model relies on two results obtained before. The first is Haldane's's result for fixation of a newly acquired mutation, obtained with $y = 1/(2N)$ in Eq. (**??**),

$$\Pi_1^*(y) = \frac{1 - e^{-s}}{1 - e^{-2Ns}} \,. \tag{1.144}$$

In the following, we shall distinguish between three classes of mutations: **(i)** *Near neutral* mutations with $|s| \ll 1/(2N)$, for which the fixation probability is $1/(2N)$, irrespective of whether the mutation is advantageous or deleterious. **(ii)** *Weakly advantageous* mutations with $1 \gg s \gg 1/(2N)$ in which case $\Pi_1^* \approx 1 - e^{-s} \approx s$. **(iii)** Even *weakly deleterious* mutations with $-1 \ll s \ll -1/(2N)$ are efficiently removed from them population, as the fixation probability is exponentially small ($\sim e^{-2n|s|}$).

Another needed ingredient is *rate of near neutral evolution* across the entire gene/genome. The initial mutation can appear in any of the $2N$ chromosomes at rate μ for a total rate of $2N\mu$. As a near neutral mutation is fixed with probability of $1/(2N)$, the rate at which such mutations are fixed in the population is itself μ, independent of population size! However, that what constitutes a near neutral mutation depends on the size of the population, and is different for say human ($N_{eff} \sim 10^3 - 10^4$) and mouse ($N_{eff} \sim 10^5$).[9]

1.7.1 Hallmarks of Cancer

Cancer tumors arise from the uncontrolled division and growth of cells. The important steps for how normal cells transform to such malignant form are summarized in the classic paper: *Hallmarks of Cancer: The Next Generation*, D. Hanahan and R. Weinberg, Cell **144(5)** 646-74 (2011). Some important steps include uncontrolled division, evasion of apoptosis (programmed cell death), and finally invasion and metastasis. Mutations that can initiate these modified cell behaviors include: *(i)* Single site mutations causing changes in proteins; *(ii)* Chromosomal rearrangements, such as elimination or duplication of a section of DNA, or even scrambling of different parts of DNA; and *(iii)* Mutations that do not affect genes, but modify their level of expression or activity.

The genes implicated in cancer can be roughly separated into two categories: *Oncogenes* which are typically expressed at high levels in tumor cells (even when present as a single copy); and *tumor suppressors* whose inactivation is implicated in disease, such as the p53 protein involved in DNA repair. For the purposes of the model to be developed shortly, both types will be denoted as *drivers*, in contrast to *passenger* mutations, whose appearance does not advance cancer tumors. Of the order of 10^2 such genes have been identified from clinical studies.

Mutation rates are abnormally high in caner cells. Let us recall the earlier estimate of $\mu \sim 2 - 5 \times 10^{-8}$ per basepair in each human generation. Given the roughly 100 cell

---

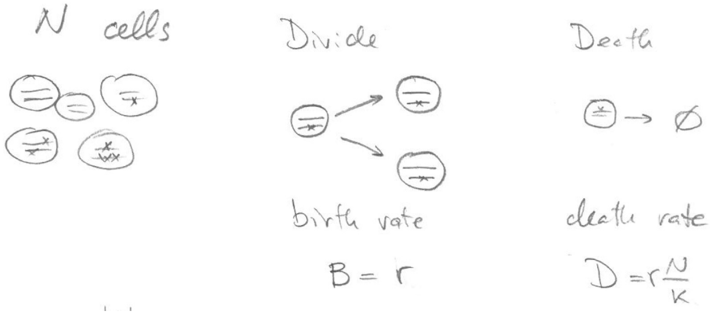
[9] As the numbers refer to the population bottleneck, the argument applies to fixation of mutations appearing prior to the bottleneck. A different reasoning applies to exponentially increasing populations.

divisions per generation from parent to progeny (in the germline, oocyte/sperm), this suggest $\mu \sim 10^{-9} - 10^{-10}$ per basepair in a healthy cell division. For cancer cells this number increases by roughly a factor of 100 to $\mu \approx 10^{-6} - 10^{-8}$. This high mutation rate also creates *passenger mutations* which are not advantageous to cancer.[10] In the following, we develop a mode for the competition between driver and passenger mutations.

1.7.2 Model of Cancer progression

The simple model focuses on single basepair mutations, assumed to occur randomly (and independently) across the entire genome. The *driver target* space T_d, in units of basepairs, is defined as the set of DNA sites whose mutation favors progression of cancer. Since there are roughly 100 driver genes, with 10 to 50 vulnerable sites per gene, we estimate $T_d \sim 5 \times 10^3$. The corresponding target space T_p for passenger mutations should be much larger, directed against the many genes needed for proper functioning of healthy cells. Assuming that there are of the order of 10^4 actively expressed genes within a cell, each with around 10^3 possible sites for (non-synonymous) mutations, leads to an estimate of $T_p \sim 10^5 - 10^7$ basepairs. The rates at which the two types of mutations appear in the cell line are $\mu_d = T_d \times \mu$, and $\mu_p = T_p \times \mu$, respectively.

The appearance of mutations modifies the fitness (reproductive capacity) of the cell, which we shall denote by $f(n_d, n_p)$ in the presence of n_d driver and n_p passanger mutations. We shall posit that each driver mutation *independently* increases fitness by a factor of $(1+s_d)$, while each passenger mutation decreases it by $(1-s_p) \approx 1/(1+s_p)$, for an overall contribution of $f(n_d, n_p) \propto (1+s_d)^{n_d}/(1+s_p)^{n_p}$.[11] We shall further assume $s_d \gg |s_p| > 0$, i.e. the (frequent) passenger mutations are nearly neutral, while the (rare) driver mutations are advantageous. Indeed, recent experiments suggest $s_d \sim 10^{-1}$.



A commonly used model to describe the increase in the number of tumor cells (or any

---

[10] Density of mutations is highly non-uniform across the genome, and appears to correlate with expression (more highly expressed regions have more mutations). As a very rough estimate, there are between 30–300 non-synonymous mutations in a typical cancer, out of which 2–5 are estimated to be within driver genes.

[11] The combined effect of distinct mutations on fitness is termed *epistasis*. The model used here corresponds to *multiplicative epistasis*.

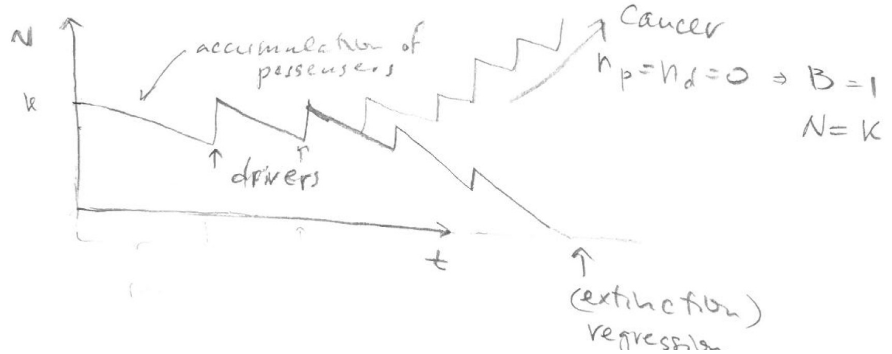
other growing population) is the *logistic equation*

$$\frac{dN}{dt} = BN - D(N)N = rN\left(1 - \frac{N}{K}\right) . \tag{1.145}$$

Cells initially grow exponentially at a rate $r = B$ (birth rate), which is then limited by the competition for resources captured by the (death) rate $D(N) = rN/K$. The growing population then saturates at $\lim_{t\to\infty} N(t) = K$, known as the carrying capacity. In fact, for our purposes, we only use the result that the saturation value is proportional to the birth rate B. For the model of cancer cells, we shall assume a (normalized) mutation-dependent birth rate of $B(n_d, n_p) = (1 + s_d)^{n_d}/(1 + s_p)^{n_p}$, but a mutation-independent death rate of $D = N/K$. Within this model, the cell line will grow to a maximum size of

$$N(n_d, n_p) = B(n_d, n_p)K = \frac{(1 + s_d)^{n_d}}{(1 + s_p)^{n_p}}K\,, \tag{1.146}$$

and then stop.



For the tumor to continue to grow, additional mutations have to occur. The appearance of an extra mutation leads to a new cell line, which (in the absence of competition for resources with previous cell lines) grows to a maximum size of $N(n_d+1, n_p)$ or $N(n_d, n_p+1)$, such that

$$\begin{aligned} \Delta N_d &\equiv N(n_d + 1, n_p) - N(n_d, n_p) \approx N(n_d, n_p)s_d\,, \\ \Delta N_p &\equiv N(n_d, n_p + 1) - N(n_d, n_p) \approx -N(n_d, n_p)s_p\,. \end{aligned} \tag{1.147}$$

Such incremental growth (assuming that the time for fixation is faster than that for appearance of a new mutation) can either eventually stop (if fitness continues to decrease with accumulation of passenger mutations), or grow unbounded (if driver mutations dominate). As an indicator of the possible outcomes, we examine the average "velocity"

$$v = \left\langle \frac{\Delta N}{\Delta t} \right\rangle \equiv v_d - v_p = \Delta N_d R_d + \Delta N_p R_p = N(n_d, n_p)(s_d R_d - s_p R_p)\,, \tag{1.148}$$

where R_d and R_p are the rates at which new driver or passenger mutations are fixed in the population, each being a product of the rate of appearance of the mutation and the probability of its fixation. The probability of a new driver mutation in a population of size $N = N(n_d, n_p)$ is $\mu T_d N$, where T_d is the target size for driver mutations. Similarly a passenger mutation occurs with probability $\mu T_p N$, where T_p is the corresponding target size. (It is reasonable to expect $T_p \gg T_d$.)

An 'advantageous' mutation is fixed with probability $1 - e^{-s_d} \approx s_d$, leading to $R_d = \mu T_d N s_d$.[12] Conversely, the near neutral passenger mutations are fixed with probability $1/N$, while appearing at rate $\mu T_p N$, resulting in $R_p = \mu T_p N/N = \mu T_p$. Putting these results in Eq.(**??**) gives

$$v = \mu T_d N^2 s_d^2 - \mu T_p N s_p \,. \tag{1.149}$$

The borderline between regression and progression of the tumor occurs when $v(N^*) = 0$, at

$$N^* = \frac{s_p T_p}{s_d^2 T_d} \,. \tag{1.150}$$

There is an effective barrier for tumor progression at small sizes due to accumulation of passengers. We shall use the following estimates: $s_p \sim 10^{-3} - 10^{-4}$, $s_d \approx 0.1$, $T_d \approx 1 - 5 \times 10^3$, and $T_p \approx 10^6 - 10^7$, leading to $N^* \sim 10^2 - 10^3$ cells. Assuming that these cells grow on top of a substrate (as is the case in melanoma), this number corresponds to a size of around 1mm. Oncologists typically consider biopsies for growths larger than around half a centimeter.

Note that the mutation rate μ cancels out in the above estimate for N^*. It is, however, relevant to the long time fate of tumors grown beyond this critical size. To arrive at this conclusion, let us first consider the balance of mutation and selection in a simple model of two alleles, mutating to each other at the same rate μ. The wild-type and mutant are assumed to have fitness values of $1 - s$ and 1, respectively. At a frequency x of the fitter mutants, the mean fitness of the population is $\overline{f} = x + (1 - x)(1 - s)$. The frequency of the mutant in the subsequent generation is given by

$$x' = \frac{\mu(1-x)(1-s)}{\overline{f}} + \frac{(1-\mu)x \times 1}{\overline{f}} = \frac{\mu(1-x)(1-s) + (1-\mu)x}{x + (1-x)(1-s)} \,; \tag{1.151}$$

the first term arising from mutation of wild-types and the second from un-mutated mutants. The steady state solution occurs for $x' = x = x_\infty$, where

$$s x_\infty^2 + \left[2\mu - s(1+\mu)\right] x_\infty - \mu(1-s)^2 = 0 \,. \tag{1.152}$$

In the limit, where μ and s are much smaller than one, but of the same order, after dropping terms, we arrive at

$$x_\infty = 1 - \frac{\mu}{s} \,, \qquad \text{with} \qquad \overline{f}_\infty = 1 - \mu \,. \tag{1.153}$$

This surprising result, that the average fitness in steady state does not depend on the gain in fitness s, is known as the *Muller–Haldane principle.* This conclusion remains true in a more general model in which at each generation any number of weakly deleterious mutations

---

[12] The use of results based on "Fisher–Wright" binomial selection is certainly debatable for cancer cells. An alternative *Moran process* leads to a fixation rate of $s/(1 + s)$, which is similar for small s.

can appear, with fitness cost of $(1-s)^k \approx 1-sk$ for $0 < s \ll 1$. For $\mu \ll 1$, it is reasonable to assume that the probability of appearance of k mutations in each generation is Gaussian distributed with mean μ. It can then be shown (see problems) that the balance of mutations and selections results in steady state, where the fraction of cells with k mutations in the population is Poisson distributed, i.e. $x_k = e^{-\langle k \rangle} \langle k \rangle^k / k!$, with $\langle k \rangle = \mu/s$. The average fitness in this steady state is again (to lowest order in s) $\overline{f}_\infty = 1 - s \sum_k k p_k = 1 - s \langle k \rangle = 1 - \mu$.

Let us apply the above result to our model of passengers and drivers: Given the target space of passengers, their effective mutation rate is μT_p, leading to an average (steady-state) number $\langle k \rangle = \mu\ T_p/s_p$, and a corresponding overall loss in average fitness of $\mu\ T_p$. A 'productive' driver mutation must confer sufficient fitness advantage to overcome this loss to become competitive, i.e. requiring $s_d > \mu\ T_p$. Thus continued growth of the tumor through accumulation of additional drivers needs $\mu < \mu^* = s_d/T_p$, creating a new obstacle to progression of cancer. Hence, rather counterintuitively, a large mutation rate proves a barrier to expansion of cancer tumors.