1.5 Population genetics of Cancer

In this section, we shall use the general results obtained so far to develop a simple model for progression of cancer. In particular, let us reexamine Haldane's's result for fixation of a newly acquired mutation, $y = 1/(2N)$ in Eq. (1.162),

$$\Pi_1^*(y) = \frac{1 - e^{-s}}{1 - e^{-2Ns}} \,. \tag{1.137}$$

in more detail. 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 ($\sim e^{-2n|s|}$).

We can also inquire about the *rate of near neutral* evolution across the entire gene/genome? The initial mutation can appear in any of the $2N$ chromosome 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 N! Note, 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 ($N_{eff} \sim 10^5$).[8]

1.5.1 Hallmarks of Cancer

Tumors of cancer are formed following uncontrolled division and growth of cells. The important steps for how normal cells transform to such malignant form are nicely 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 cause these changes 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 segments of DNA; *(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 generally implicated with 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.

[8] As the numbers refer to the population bottleneck, the argument applies to fixation of mutations appearing prior to the bottleneck. Different argumentation applied to exponentially increasing populations.

1.5.2 Model of Cancer progression

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 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 to $\mu \approx 10^{-6} - 10^{-8}$. This high mutation rate also creates *passenger mutations* which are not advantageous to cancer. In the following, we develop a mode for the competition between driver and passenger mutations.

Let us focus on single basepair mutations, assume to occur randomly (and independently) over 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, as most mutations are likely to reduce the fitness 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 $f(n_d, n_p) \propto (1+s_d)^{n_d}/(1+s_p)^{n_p}$.[9] 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 other growing population) is the *logistic equation*

$$\frac{dN}{dt} = BN - D(N)N = rN\left(1 - \frac{N}{K}\right) . \tag{1.138}$$

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. To apply the logistic map to 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

$$N(n_d, n_p) = \frac{B(n_d, n_p)}{K} = \frac{(1+s_d)^{n_d}}{K(1+s_p)^{n_p}} , \tag{1.139}$$

and then stop.

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

For the tumor to continue to grow, additional mutations have to occur. The appearance of an extra mutation leads to a new cell line, growing to a maximum size $N(n_d+1, n_p)$ or $N(n_d, n_p+1)$, such that

$$\begin{aligned} \Delta N_d &\equiv N(n_d+1, n_p) - N(n_d, n_p) \approx N(n_d, n_p) s_d \,, \\ \Delta N_p &\equiv N(n_d, n_p+1) - N(n_d, n_p) \approx -N(n_d, n_p) s_p \,. \end{aligned} \tag{1.140}$$

Such incremental growth 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"

$$v = \left\langle \frac{\Delta N}{\Delta t} \right\rangle \equiv v_d - v_p = \Delta N_d R_d + \Delta N_p R_p = N(n_d, n_p)(s_d R_d - s_p R_p) \,, \tag{1.141}$$

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 $\mu T_d N$, and the 'advantageous' mutation is fixed with probability $1 - e^{-s_d} \approx s_d$, leading to $R_d = \mu T_d N s_d$. Correspondingly, 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 together gives

$$v = \mu T_d N^2 s_d^2 - \mu T_p N s_p \,. \tag{1.142}$$