A complete model

April 2019

1 Assumptions and notations

- Three types of mutations can appear on a given genotype G: S (affects the cell survival process), F (affects the cell fate) or M (affects the genomic maintenance);
- An individual gets cancer once one cell carries $k = 3$ mutations, with at least one S and one F;
- The only sources of stochasticity are the mutations appearances and the initial growth of the nodule they initiate (the later growth of the nodule is approximated by a deterministic logistic growth).

2 Population growth

Each population (the initial population as well as any newly arising subpopulation) is modeled as such[1]:

- starts with one individual;
- behaves as a birth-and-death process until it reaches size εK at time τ, where the choice of ε will be discussed later;
- from then on, the population size follows a logistic growth with parameters (r, K), where r corresponds to the difference between the birth and death rates of the birth-and-death process.

On $[0, \tau]$ we will use when it is relevant the approximation of the birth-and-death process by its limiting distribution in the supercritical case, deduced from the branching process theory. Denoting by X_t the population size at time t, and $\Omega_0 = \{\lim_t X_t = 0\}$ its extinction set, we have $e^{-rt}X_t \to V$ almost surely, where $\mathbb{P}(V = 0) = \mathbb{P}(\Omega_0) = d/b$ and $V \mid \Omega_0^c$ is exponentially distributed with parameter r/b.

The constant ε should be chosen in such a way that:

(i) The survival probability of the birth-and-death process is close to the probability of reaching size εK. Using the previous strong convergence, the fact that X_t conditionally on Ω_0 behaves like a subcritical birth-and-death process $\tilde{X}_t$ with birth rate d and death rate b, and Doob's

[1] All the parameters involved might depend on the genotype G of the population, hence will be denoted r_G, K_G etc. when necessary.

martingale inequality, we obtain

$$\begin{aligned}\mathbb{P}\left(\exists t:\, X_t > \varepsilon K\right) &= \frac{d}{b}\mathbb{P}\left(\exists t:\, X_t > \varepsilon K \mid \Omega_0\right) + \frac{r}{b}\mathbb{P}\left(\exists t:\, X_t > \varepsilon K \mid \Omega_0^c\right)\\ &= \frac{d}{b}\mathbb{P}\left(\exists t:\, \tilde{X}_t > \varepsilon K\right) + \frac{r}{b}\\ &\leqslant \frac{d}{b}\lim_n \mathbb{P}\left(\max_{t\in[0,n]} e^{rt}\tilde{X}_t \geqslant \varepsilon K\right) + \frac{r}{b}\\ &\leqslant \frac{d}{b}\frac{1}{\varepsilon K} + \frac{r}{b}.\end{aligned}$$

Hence for εK large enough, this probability is close to the survival probability r/b.

(ii) The approximation of X_t by $e^{rt}V$ is valid when X_t becomes close to εK. This requires V to be of the order of $e^{-r\tau}\varepsilon K$, hence to be much less than εK. We have

$$\mathbb{P}\left(V \leqslant \varepsilon K \mid \Omega_0^c\right) = 1 - \exp\left(-\frac{r\varepsilon K}{b}\right).$$

For εK large enough, this probability is close to 1.

Let f be the probability distribution function of the reaching time τ. Using the approximation $X_t \approx e^{rt}V$ leads to

$$\mathbb{P}\left(\tau \leqslant t\right) \approx \mathbb{P}\left(V \geqslant e^{-rt}\varepsilon K \mid \Omega_0^c\right) = \exp\left(-\frac{re^{-rt}\varepsilon K}{b}\right),$$

hence

$$f\left(t\right) = \frac{r^2 e^{-rt}\varepsilon K}{b}\exp\left(-\frac{re^{-rt}\varepsilon K}{b}\right). \tag{1}$$

(iii) It is reasonable to assume that no mutation occurs during the time-interval $[0,\tau]$. Denoting by ν the mutation rate, we obtain by Jensen's inequality,

$$\begin{aligned}\mathbb{P}\left(\text{mutation during}\,[0,\tau]\right) &= 1 - \mathbb{E}\left(e^{-\nu\int_0^\tau X_t dt}\right)\\ &\leqslant 1 - \mathbb{E}\left(e^{-\nu\varepsilon K\tau}\right)\\ &\leqslant 1 - e^{-\nu\varepsilon K\mathbb{E}(\tau)}.\end{aligned}$$

By (1),

$$\begin{aligned}\mathbb{E}\left(\tau\right) &= \frac{r^2\varepsilon K}{b}\int_{-\infty}^{+\infty} te^{-rt}\exp\left(-\frac{re^{-rt}\varepsilon K}{b}\right)dt\\ &= -\frac{1}{r}\int_0^{+\infty}\ln\left(\frac{bs}{r\varepsilon K}\right)e^{-s}ds\\ &= -\frac{1}{r}\ln\left(\frac{b}{r\varepsilon K}\right)\int_0^{+\infty}e^{-s}ds - \frac{1}{r}\int_0^{+\infty}\ln\left(s\right)e^{-s}ds\\ &= \frac{1}{r}\left(\ln\left(\frac{r\varepsilon K}{b}\right) + \gamma\right)\end{aligned}$$

where $\gamma = -\int_0^{+\infty}\ln\left(s\right)e^{-s}ds \approx 0.58$ is Euler's constant. Hence

$$\mathbb{P}\left(\text{mutation during}\,[0,\tau]\right) \leqslant 1 - e^{-\frac{\nu\varepsilon K}{r}\left(\ln\left(\frac{r\varepsilon K}{b}\right)+\gamma\right)},$$

which for $\nu\varepsilon K$ small enough is close to 0.

3 Effect of a mutation

Let us consider a population of stem cells with genotype G in which during its initial growth phase a cell renews itself at rate $(1-p_G)\,b_G$, asymmetrically differentiates at rate $p_G b_G$, symmetrically differentiates at rate d_G^s, or dies at rate d_G. We denote by u_G the probability of mutation per cell division. It is known that the appearance of a mutation on this population has an effect on the parameters b_G, p_G and u_G, depending on the type of mutation.

In our model, the growth of this cell population after its initial random phase is logistic. The effect of a new mutation on (b_G, p_G, u_G) should thus translate into an effect on the parameters (r_G, K_G) of the logistic growth, as well as on ν_G^a (resp. ν_G^s), the mutation rate per cell and per time unit due to errors during asymmetric (resp. symmetric) division. These effects are summarized in Table 1, and the reasoning is detailed in Sections 3.2-3.4.

Note that we have

$$r_G = (1-p_G)\,b_G - (d_G^s + d_G)\,, \tag{2}$$

$$\nu_G^a = u_G p_G b_G, \tag{3}$$

$$\nu_G^s = 2u_G\,(1-p_G)\,b_G. \tag{4}$$

Even though only the stem cells are counted, the carrying capacity K_G of the nodule should take into account the potentially increased number of differentiated cells. It seems indeed reasonable to assume a common competition intensity between both cell types, inversely proportional to the total carrying capacity of the nodule

$$K_G\,(1+\lambda_G)\,, \tag{5}$$

where λ_G is the ratio between the differentiated cells and the stem cells carrying capacities in the nodule.

initial genotype G	genotype GS	genotype GF	genotype GM
b_G	$(1+\alpha)\,b_G$	b_G	b_G
p_G	p_G	$(1-\beta)\,p_G$	p_G
u_G	u_G	u_G	$(1+\gamma)\,u_G$
r_G	$(1+\alpha)\,r_G$	$\left(1+\frac{\beta p_G}{1-p_G}\right) r_G$	r_G
K_G	$(1+\alpha)\,K_G$	$\left(1+\frac{\beta\lambda_G}{1+(1-\beta)\lambda_G}\right) K_G$	K_G
ν_G^a	$(1+\alpha)\,\nu_G^a$	$(1-\beta)\,\nu_G^a$	$(1+\gamma)\,\nu_G^a$
ν_G^s	$(1+\alpha)\,\nu_G^s$	$\left(1+\frac{\beta p_G}{1-p_G}\right) \nu_G^s$	$(1+\gamma)\,\nu_G^s$

Table 1: Effects of the appearance of a mutation on a cell population with genotype G.

3.1 Parameter values

For a mutation-free genotype $G = \emptyset$, the aforementioned parameters have the following order of magnitude, depending on the tissue: $p_G \in [0.6, 0.95]$, b_G is a division every 8 months to every 4 days, d_G is once every two years, and d_G^s is relatively small. The mutation probability u_G is around 10^{-6} per cell division, and the carrying capacity K_G is of the order of 10^7 to 10^8 *(or is it $K_G\,(1+\lambda_G)$ the total carrying capacity, including the differentiated cells?).*

This implies in particular that the growth rate (2) can be approximated by

$$r_G = (1-p_G)\,b_G. \tag{6}$$

3.2 Effect of a type S mutation

The appearance of a type S mutation induces an increase of the division rate, namely $b_{GS} = (1+\alpha)\, b_G$. It ensues from (6) that

$$r_{GS} = (1-p_G)\,(1+\alpha)\, b_G = (1+\alpha)\, r_G.$$

This increase of the division rate results in an increase of the total carrying capacity (5), which becomes $(1+\alpha)\, K_{GS}\,(1+\lambda_{GS})$. Yet it does not affect the relative dynamics between the differentiated and stem cell population, hence $\lambda_{GS} = \lambda_G$, leading to

$$K_{GS} = (1+\alpha)\, K_G.$$

3.3 Effect of a type F mutation

The appearance of a type F mutation induces a decrease of asymmetric division probability, namely $p_{GF} = (1-\beta)\, p_G$. Let us define $\tilde{\beta} = \beta p_G / (1-p_G)$. We deduce from (6) that

$$r_{GF} = (1-(1-\beta)\, p_G)\, b_G = (1-p_G)\left(1+\tilde{\beta}\right) b_G = \left(1+\tilde{\beta}\right) r_G.$$

and from (3)-(4) that

$$\nu^a_{GF} = u_G\,(1-\beta)\, p_G b_G = (1-\beta)\, \nu^a_G,$$
$$\nu^s_{GF} = 2u_G\,(1-(1-\beta)\, p_G)\, b_G = \left(1+\tilde{\beta}\right)\nu^s_G.$$

Moreover, the decrease of the asymmetric division probability biases the asymptotic proportions of the stem cells versus the differentiated cells, namely $\lambda_{GF} = (1-\beta)\,\lambda_G$. However, the total carrying capacity (5) of the nodule remains unchanged, namely $K_G\,(1+\lambda_G) = K_{GF}\,(1+\lambda_{GF})$, leading to

$$K_{GF} = \frac{1+\lambda_G}{1+(1-\beta)\,\lambda_G} K_G.$$

3.4 Effect of a type M mutation

The appearance of a type M mutation induces an increase of the mutation probability, namely $u_{GM} = (1+\gamma)\, u_G$, which by (3)-(4) implies that

$$\nu^a_{GM} = (1+\gamma)\, u_G p_G b_G = (1+\gamma)\, \nu^a_G,$$

$$\nu^s_{GM} = 2\,(1+\gamma)\, u_G\,(1-p_G)\, b_G = (1+\gamma)\, \nu^s_G.$$

4 Incidence probability

Let T_k be the first appearance time of a cell carrying $k = 3$ mutations (with at least one F and one S). For any fixed time T, any genotype G and any $t \leqslant T$, we define

$$q_G(t) = \mathbb{P}\,(\text{a cell population of size } \varepsilon K_G \text{ at time } t \text{ with genotype } G$$
$$\text{has no offspring carrying at time } T \text{ some genotype } G' \text{ with } \#G' \geqslant 3,\ S \in G',\ F \in G'\ ).$$

We are then interested in

$$\mathbb{P}_G\,(T_k > T) = q_G\,(0)\,,$$

with for instance $G = \emptyset$ or $G = M$.

Let us consider some mutation $U \in \{S, F, M\}$ potentially appearing on a population with genotype G. We assume that this population is logistically growing, starting at time 0 with size

εK_G. Let us denote by $\nu_G(t)$ the appearance rate of any mutation in this population, and by $p_{G,U}$ the probability that this mutation is of type U. We shall set for instance

$$p_{G,U} = 0 \quad \text{if} \quad \begin{cases} M \in G,\ U = M, \\ \#G = 3. \end{cases}$$

Let $\mu_{GU}(t)$ then denote the appearance rate in this population of a nodule of size εK_{GU}, with genotype GU. Letting f_{GU} be the probability density function of the latency period as defined in (1), and using approximation (i), we have

$$\mu_{GU}(t) = \frac{p_{G,U} r_{GU}}{b_{GU}} \int_0^t \nu_G(s) f_{GU}(t-s)\, ds, \tag{7}$$

where

$$\nu_G(t) = \frac{(\nu_G^a + \nu_G^s)\varepsilon K_G^2 e^{r_G t}}{K_G + \varepsilon K_G (e^{r_G t} - 1)}. \tag{8}$$

We then obtain

$$q_G(t) = \begin{cases} 0, & \#G = 3,\ S \in G \text{ and } F \in G \\ 1, & \#G = 3,\ S \notin G \text{ or } F \notin G \\ \exp\left(-\int_t^T \sum_{U \in \{S,F,M\}} \mu_{GU}(s-t)(1 - q_{GU}(s))\, ds\right), & \#G < 3. \end{cases} \tag{9}$$

For instance,

$$\mathbb{P}_M(T_k > T) = q_M(0) = \exp\left(-\int_0^T \left[\mu_{MS}(t)(1 - q_{MS}(t)) + \mu_{MF}(t)(1 - q_{MF}(t))\right] dt\right),$$

with

$$q_{MS}(t) = \exp\left(-\int_t^T \mu_{MSF}(s-t)\, ds\right),$$

$$q_{MF}(t) = \exp\left(-\int_t^T \mu_{MFS}(s-t)\, ds\right).$$

5 Relative risk

For a given initial genotype G, let us consider the effect of an exposure inducing an increase of the probability of mutation per cell division by a factor $x > 1$: $u_G^e = x u_G$. By (3)-(4) and (8)

$$\nu_G^e(t) = x \nu_G(t),$$

which implies that for any genotype G' corresponding to the initial genotype G with added mutations,

$$\mu_{G'}^e(t) = x \mu_{G'}(t). \tag{10}$$

For any genotype of the form $U_1U_2U_3$, we have by (9) $q^e_{U_1U_2U_3}(t) = q_{U_1U_2U_3}(t)$ since $\#U_1U_2U_3 = 3$. From (10), $\mu^e_{U_1U_2U_3}(t) = x\mu_{U_1U_2U_3}(t)$, therefore

$$\begin{aligned} q^e_{U_1U_2}(t) &= \exp\left(-x \int_t^T \sum_{U_3 \in \{S,F,M\}} \mu_{U_1U_2U_3}(s-t)(1 - q_{U_1U_2U_3}(s))\, ds\right) \\ &= q_{U_1U_2}(t)^x. \end{aligned} \tag{11}$$

It follows that

$$q^e_{U_1}(t) = \exp\left(-x\int_t^T \sum_{U_2\in\{S,F,M\}} \mu_{U_1U_2}(s-t)\left(1-q_{U_1U_2}(s)^x\right)ds\right),$$

which using

$$1-q \leqslant 1-q^x \leqslant x(1-q) \tag{12}$$

leads to

$$q_{U_1}(t)^{x^2} \leqslant q^e_{U_1}(t) \leqslant q_{U_1}(t)^x \tag{13}$$

Let us now consider the hazard ratio for these given exposure and genotype G. We have

$$\mathrm{HR} = \frac{\sum_{U\in\{S,F,M\}} \mu^e_{GU}(t)\left(1-q^e_{GU}(t)\right)}{\sum_{U\in\{S,F,M\}} \mu_{GU}(t)\left(1-q_{GU}(t)\right)}. \tag{14}$$

- For a mutation-free genotype $G=\emptyset$, it comes from (10) and (13) that

$$\frac{\sum_{U_1} x\mu_{U_1}(t)\left(1-q_{U_1}(t)^x\right)}{\sum_{U_1}\mu_{U_1}(t)\left(1-q_{U_1}(t)\right)} \leqslant \mathrm{HR} \leqslant \frac{\sum_{U_1} x\mu_{U_1}(t)\left(1-q_{U_1}(t)^{x^2}\right)}{\sum_{U_1}\mu_{U_1}(t)\left(1-q_{U_1}(t)\right)}$$

  which together with (12) implies that

$$x \leqslant \mathrm{HR} \leqslant x^3. \tag{15}$$

- For $G=M$, we similarly obtain

$$\mathrm{HR} = \frac{\sum_{U_2} x\mu_{MU_2}(t)\left(1-q_{MU_2}(t)^x\right)}{\sum_{U_2}\mu_{MU_2}(t)\left(1-q_{MU_2}(t)\right)}$$

  which by (12) leads to

$$x \leqslant \mathrm{HR} \leqslant x^2. \tag{16}$$