A complete model

December 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 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

The growth of any cell population (the initial tissue cells as well as any nodule carrying a new driver mutation) is modeled as follows. In both cases, only the stem cells are counted. Note that all the parameters involved in the subsequent description might depend on the genotype of the population. This dependence will be clarified in Section 4. We denote by X_t the size of the stem cell population at time t.

2.1 Growth of a nodule

Initial random phase. The nodule starts with one stem cell. Its size evolves as a birth-and-death process until it reaches size εK at time τ, where the choice of ε is discussed in Section 2.4. The constant K corresponds to the carrying capacity of the nodule. A stem cell renews itself at rate $(1-p)\,b$, asymmetrically differentiates at rate pb, symmetrically differentiates at rate d^s, or dies at rate d. The birth rate of the birth-and-death process is thus $(1-p)\,b$, and its death rate $d^s + d$.

Logistic growth phase. Once it has reached size εK, the growth of the nodule is logistic, namely $\frac{dX_t}{dt} = r\left(1 - \frac{X_t}{K}\right) X_t$, where K is the carrying capacity and r the intrinsic growth parameter, with

$$r = (1-p)\,b - (d^s + d)\,. \tag{1}$$

The fact that the net increase rate per cell and per unit time is $r\left(1 - \frac{X_t}{K}\right)$ comes from the classical assumption that the birth rate per cell and per time unit remains constant equal to $(1-p)\,b$, and the death rate (accounting for natural death and competition) is $d^s + d + ((1-p)\,b - (d^s + d))\,X_t/K$.

2.2 Growth of the initial tissue

The fundamental difference compared to Section 2.1 is that in the initial tissue cell population, the asymmetric division probability is not assumed constant. To account for the fact that there are more asymmetric divisions once the tissue is fully developed than in the development phase, we assume that the asymmetric division probability $p(X_t)$ increases linearly from p to

p_K, reaching its maximal value p_K once $X_t = K$, i.e. $p(X_t) = \frac{p_K - p}{K} X_t + p$. Then, assuming that the birth rate per cell and per time unit is $(1 - p(X_t))\, b$, and that the the death rate is $d^s + d + ((1 - p_K)\, b - (d^s + d))\, X_t/K$, we obtain an increase rate of $r\left(1 - \frac{X_t}{K}\right) X_t$, where r is given by (1). The growth of the population is thus logistic with parameters r and K. We assume that the nodule starts at time 0 with size εK, leading to

$$X_t = \frac{\varepsilon K^2 e^{rt}}{K + \varepsilon K\left(e^{rt} - 1\right)}. \tag{2}$$

2.3 Approximation during the initial growth phase

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 $\Omega_0 = \{\lim_t X_t = 0\}$ the extinction set,

$$\lim_{t \to \infty} e^{-rt} X_t \overset{a.s.}{=} V,$$

where $\mathbb{P}(V = 0) = \mathbb{P}(\Omega_0) = \frac{d^s + d}{(1-p)b}$ and $V \mid \Omega_0^c$ is exponentially distributed with parameter $\frac{r}{(1-p)b}$.

2.4 Choice of the threshold ε

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 $d^s + s$ and death rate $(1 - p)\, b$, and Doob's martingale inequality, we obtain

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

Hence for εK large enough, this probability is close to the survival probability $r/(1-p)\, b$.

(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

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

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

$$\mathbb{P}(\tau \leqslant t) \approx \mathbb{P}\left(V \geqslant e^{-rt} \varepsilon K \mid \Omega_0^c\right) = \exp\left(-\frac{r e^{-rt} \varepsilon K}{(1-p)\, b}\right),$$

hence

$$f(t) = \frac{r^2 e^{-rt} \varepsilon K}{(1-p)\, b} \exp\left(-\frac{r e^{-rt} \varepsilon K}{(1-p)\, b}\right). \tag{3}$$

(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 (3),

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

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

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

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

3 Incidence probability

We assume that an individual gets a given cancer once one of his cells carries three mutations, with at least one of type F and one of type S. We denote by $\mathcal{G}$ the set of the "cancer" genotypes, namely

$$\mathcal{G}=\left\{\text{genotype } G:\ \#G=3, S\in G, F\in G\right\}.$$

Let us consider an individual with genotype G at birth (typically, $G=\emptyset$ or $G=M$) and let T be a given fixed time. We denote by q_G the probability of getting a given cancer after time T, i.e. the probability that the initial tissue cell population of the corresponding organ has no offspring at time T carrying a $\mathcal{G}$-genotype. We assume that this population is logistically growing, with parameters as described in Section 2, starting at time 0 with size εK_G. Then

$$q_G=\begin{cases}0, & G\in\mathcal{G}\\ 1, & \#G=3,\ G\notin\mathcal{G}\\ \exp\left(-\int_0^T\sum_{U\in\{S,F,M\}}\phi_{GU}\left(t\right)dt\right), & \#G<3.\end{cases}\tag{4}$$

where for each mutation U, $\phi_{GU}\left(t\right)$ denotes the appearance rate in the tissue cell population of a GU-nodule of size εK_{GU}, which leads to cancer before time T (i.e. has some offspring at time T carrying a $\mathcal{G}$-genotype).

In order to compute $\phi_{GU}\left(t\right)$, we first define u_G the probability of mutation per cell division, and $\nu_G\left(s\right)$ the appearance rate of any mutation in this population:

$$\nu_G\left(s\right)=\left(\nu_G^a\left(s\right)+\nu_G^s\left(s\right)\right)\frac{\varepsilon K_G^2e^{r_Gs}}{K_G+\varepsilon K_G\left(e^{r_Gs}-1\right)}\tag{5}$$

where the second factor is the deterministic size of the tissue population given by (2), and where

$$\nu_G^a(s) = u_G p_G(s) b_G, \tag{6}$$
$$\nu_G^s(s) = 2u_G(1 - p_G(s)) b_G, \tag{7}$$

account for the mutation rate per cell and per time unit due to errors during asymmetric (resp. symmetric) division. We recall that the asymmetric division probability is in the initial tissue population size-dependent, or equivalently time-dependent, namely

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

Next, let $\pi_{G,U}$ be the probability that a mutation appearing on a genotype G is of type U. If this mutation appears at some time s, let $\rho_{GU(s)}$ (resp. $f_{GU(s)}(t)$) be the probability of reaching size εK_{GU} (the probability density function of the latency period to reach size εK_{GU}) for the GU-nodule. Finally, assuming that this nodule reaches size εK_{GU} at time t, we denote be $q_{GU(s)}(t)$ its probability to have no offspring at time T carrying a $\mathcal{G}$-genotype. Note that these three last quantities depend on the appearance time s of the mutation, because the growth parameters of the GU-nodule depend themselves on the value of the asymmetric division probability $p_G(s)$ in the tissue cell population at that time.

The different growth and mutation parameters for the GU-nodule differ from the parameters of the initial tissue population with genotype G, and depend on the type of the mutation U. The effects of the different mutation types on the parameters will be detailed in Section 4. For now we simply denote with a subscript GU all the parameters of the GU-nodule, and add if necessary the dependence on the appearance time s of the mutation U in brackets, namely b_{GU}, $p_{GU(s)}$, u_{GU}, $r_{GU(s)}$, K_{GU}, $\nu^a_{GU(s)}$, $\nu^s_{GU(s)}$.

The appearance rate $\phi_{GU}(t)$ is then

$$\phi_{GU}(t) = \int_0^t \nu_G(s) \pi_{G,U} \rho_{GU(s)} f_{GU(s)}(t-s) \left(1 - q_{GU(s)}(t)\right) ds, \tag{9}$$

where

- the probability $\rho_{GU(s)}$ can according to (i) be approximated as the survival probability of the birth-and-death process, namely

$$\rho_{GU(s)} = \frac{r_{GU(s)}}{\left(1 - p_{GU(s)}\right) b_{GU}}, \tag{10}$$

- the probability density function $f_{GU(s)}(t)$ is thanks to (3) given by

$$f_{GU(s)}(t) = \frac{r^2_{GU(s)} e^{-r_{GU(s)} t} \varepsilon K}{\left(1 - p_{GU(s)}\right) b_{GU}} \exp\left(-\frac{r_{GU(s)} e^{-r_{GU(s)} t} \varepsilon K_{GU}}{\left(1 - p_{GU(s)}\right) b_{GU}}\right), \tag{11}$$

- the probability $q_{GU(s)}(t)$ can be computed similarly as for q_G given by (12), namely

$$q_{GU(s)}(t) = \begin{cases} 0, & GU \in \mathcal{G} \\ 1, & \#GU = 3,\ GU \notin \mathcal{G} \\ \exp\left(-\int_t^T \sum_{V \in \{S,F,M\}} \phi_{GU(s)V}(u)\, du\right), & \#GU < 3. \end{cases} \tag{12}$$

where $\phi_{GU(s)V}(t)$ is the appearance rate in the GU-population of a GUV-nodule of size εK_{GUV} leading to cancer before time T, knowing that the first mutation U appeared on the initial tissue population at time s. The main difference in the computation of $\phi_{GU(s)V}(t)$ compared to $\phi_{GU}(t)$ is that the mutation and growth parameters of a GUV-nodule do not

depend on the appearance time of the second mutation V, since as detailed in Section 2.1 the asymmetric mutation probability remains constant in any newly arising nodule. Still, these parameters are a modification of the GU-nodule parameters (see Section 4), and thus depend on the appearance time s of the first mutation U. Keeping the same notations as in (9), we have

$$\phi_{GU(s)V}(t) = \pi_{GU,V}\rho_{GU(s)V}\left(1 - q_{GU(s)V}(t)\right)\int_0^t \nu_{GU(s)}(u)\, f_{GU(s)V}(t-u)\, du. \tag{13}$$

All the quantities involved here are computed as in (5)-(11), replacing the subscripts accordingly:

$$\begin{aligned}
\nu_{GU(s)}(u) &= \left(\nu^a_{GU(s)} + \nu^s_{GU(s)}\right)\frac{\varepsilon K_{GU}^2 e^{r_{GU(s)}u}}{K_{GU} + \varepsilon K_{GU}\left(e^{r_{GU(s)}u} - 1\right)},\\
\nu^a_{GU(s)} &= u_{GU} p_{GU(s)} b_{GU},\\
\nu^s_{GU(s)} &= 2u_{GU}\left(1 - p_{GU(s)}\right) b_{GU},\\
\rho_{GU(s)V} &= \frac{r_{GU(s)V}}{\left(1 - p_{GU(s)V}\right) b_{GUV}},\\
f_{GU(s)V}(t) &= \frac{r^2_{GU(s)V} e^{-r_{GU(s)V}t}\varepsilon K}{\left(1 - p_{GU(s)V}\right) b_{GUV}} \exp\left(-\frac{r_{GU(s)V} e^{-r_{GU(s)V}t}\varepsilon K_{GUV}}{\left(1 - p_{GU(s)V}\right) b_{GUV}}\right).
\end{aligned}$$

The asymmetric mutation probability in the GU-nodule being constant over time, so are its mutation rates $\nu^a_{GU(s)}$ and $\nu^s_{GU(s)}$. Also,

$$q_{GU(s)V}(t) = \begin{cases} 0, & GUV \in \mathcal{G} \\ 1, & \#GUV = 3,\ GUV \notin \mathcal{G} \\ \exp\left(-\int_t^T \sum_{W\in\{S,F,M\}} \phi_{GU(s)VW}(u)\, du\right), & \#GUV < 3. \end{cases} \tag{14}$$

and we iterate the computation a third time.

Remark 1. In the particular case where the initial genotype G is not empty but carries the inherited mutation M, the third computation (14) is not necessary. We simply have, since no MM-nodule can have offspring with a $\mathcal{G}$-genotype,

$$q_M = \exp\left(-\int_0^T \left(\phi_{MS}(t) + \phi_{MF}(t)\right) dt\right),$$

where $\phi_{MS}(t)$ and $\phi_{MF}(t)$ are given by (9) and only require the computation of $\phi_{MS(s)F}(t)$ and $\phi_{MF(s)S}(t)$.

4 Effect of a mutation

Let us consider a driver mutation $U \in \{S, F, M\}$ appearing on a cell population with a given genotype G. We differentiate two cases:

(A) This mutation appears in the original tissue cell population with some genotype G, in which case its appearance time t influences the mutation and growth parameters of the subsequent nodule. For any parameter θ_G of the tissue cell population, we shall denote by $\theta_{GU(t)}$ (or θ_{GU} when suitable) the corresponding parameter of the subsequent GU-nodule.

(B) This mutation appears in a nodule, that is to say not in the original tissue cell population. In this case, its appearance time of the mutation has no effect on the parameters of the subsequent. Denoting by G the genotype of the first nodule, then for any parameter θ_G we shall denote by θ_{GU} the corresponding parameter of the subsequent GU-nodule.

The effect of the different types of mutation on the parameters b_G, p_G, u_G, which translates into an effect on the mutation and growth parameters r_G, K_G, ν_G^a and ν_G^s, are summarized in Table 1 (case (A)) and Table 2 (case (B)). The reasoning is detailed in Sections 4.1-4.3.

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

$$K_G\left(1+\lambda_G\right), \tag{15}$$

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

In what follows we will often make use of the fact that the death rates d^s and d (assumed constant through all genotypes) are negligible compared to $(1-p_G)\,b_G$. Indeed, 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. Hence we shall when necessary make the approximation

$$(1-p_G)\,b_G-(d^s+d)\approx(1-p_G)\,b_G. \tag{16}$$

We also introduce the notation (useful in case (A))

$$r_G(t)=(1-p_G(t))\,b_G-(d^s+d). \tag{17}$$

	$U=S$	$U=F$	$U=M$
b_{GU}	$(1+\alpha)\,b_G$	b_G	b_G
$p_{GU(t)}$	$p_G(t)$	$(1-\beta)\,p_G(t)$	$p_G(t)$
u_{GU}	u_G	u_G	$(1+\gamma)\,u_G$
$r_{GU(t)}$	$(1+\alpha)\,r_G(t)$	$\left(1+\frac{\beta p_G(t)}{1-p_G(t)}\right) r_G(t)$	$r_G(t)$
K_{GU}	$(1+\alpha)\,K_G$	$\left(1+\frac{\beta\lambda_G}{1+(1-\beta)\lambda_G}\right) K_G$	K_G
$\nu^a_{GU(t)}$	$(1+\alpha)\,\nu^a_G(t)$	$(1-\beta)\,\nu^a_G(t)$	$(1+\gamma)\,\nu^a_G(t)$
$\nu^s_{GU(t)}$	$(1+\alpha)\,\nu^s_G(t)$	$\left(1+\frac{\beta p_G(t)}{1-p_G(t)}\right) \nu^s_G(t)$	$(1+\gamma)\,\nu^s_G(t)$

Table 1: Effects of a driver mutation $U\in\{S,M,F\}$ appearing at time t on the initial tissue cell population with genotype G (case (A)).

	$U=S$	$U=F$	$U=M$
b_{GU}	$(1+\alpha)\,b_G$	b_G	b_G
p_{GU}	p_G	$(1-\beta)\,p_G$	p_G
u_{GU}	u_G	u_G	$(1+\gamma)\,u_G$
r_{GU}	$(1+\alpha)\,r_G$	$\left(1+\frac{\beta p_G}{1-p_G}\right) r_G$	r_G
K_{GU}	$(1+\alpha)\,K_G$	$\left(1+\frac{\beta\lambda_G}{1+(1-\beta)\lambda_G}\right) K_G$	K_G
ν^a_{GU}	$(1+\alpha)\,\nu^a_G$	$(1-\beta)\,\nu^a_G$	$(1+\gamma)\,\nu^a_G$
ν^s_{GU}	$(1+\alpha)\,\nu^s_G$	$\left(1+\frac{\beta p_G}{1-p_G}\right) \nu^s_G$	$(1+\gamma)\,\nu^s_G$

Table 2: Effects of a driver mutation $U\in\{S,M,F\}$ appearing on a cell population with genotype G, if the population is not the initial tissue cells (case (B)).

4.1 Effect of a type S mutation

The appearance of a type S mutation induces an increase of the division rate, namely $b_{GS} = (1+\alpha)\, b_G$. This increase of the division rate results in an increase of the total carrying capacity (15), 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 $K_{GS} = (1+\alpha)\, K_G$.

In case (A), for a mutation S appearing at time t, we have $p_{GS(t)} = p_G\,(t)$. Using (16), it follows that

$$r_{GS(t)} = (1+\alpha)\,(1-p_G\,(t))\, b_G - (d^s+d) \approx (1+\alpha)\, r_G\,(t)\,.$$

From (6)-(7),

$$\nu^a_{GS(t)} = (1+\alpha)\, u_G p_G\,(t)\, b_G = (1+\alpha)\, \nu^a_G\,(t)\,,$$
$$\nu^s_{GS}\,(t) = 2\,(1+\alpha)\, u_G p_G\,(t)\, b_G = (1+\alpha)\, \nu^s_G\,(t)\,.$$

In case (B), we have $p_{GS} = p_G$, which with the same computation as in case (A) leads to $r_{GS} \approx (1+\alpha)\, r_G$, $\nu^a_{GS} = (1+\alpha)\, \nu^a_G$ and $\nu^s_{GS} = (1+\alpha)\, \nu^s_G$.

4.2 Effect of a type F mutation

The appearance of a type F mutation induces a decrease of asymmetric division probability, say via a multiplicative factor $1-\beta$. This decrease biases the asymptotic proportions of the stem cells versus the differentiated cells, namely $\lambda_{GF} = (1-\beta)\,\lambda_G$. However, the total carrying capacity (15) of the nodule remains unchanged, namely $K_G\,(1+\lambda_G) = K_{GF}\,(1+\lambda_{GF})$, leading to

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

In case (A), for a mutation F appearing at time t, we have $p_{GF(t)} = (1-\beta)\, p_G\,(t)$. Defining $\tilde{\beta} = \beta p_G\,(t) \,/\, (1-p_G\,(t))$ and using (16), this implies

$$r_{GF(t)} = \big(1+\tilde{\beta}\big)\,(1-p_G\,(t))\, b_G - (d^s+d) \approx \big(1+\tilde{\beta}\big) r_G\,(t)\,.$$

From (6)-(7), we have

$$\nu^a_{GF(t)} = u_G\,(1-\beta)\, p_G\,(t)\, b_G = (1-\beta)\, \nu^a_G\,(t)\,,$$

$$\nu^s_{GF(t)} = 2u_G\,(1-(1-\beta)\, p_G\,(t))\, b_G = \big(1+\tilde{\beta}\big)\nu^s_G\,(t)\,.$$

In case (B), $p_{GF} = (1-\beta)\, p_G$. Defining $\tilde{\beta} = \beta p_G / \,(1-p_G)$, we obtain with the same computation $r_{GF} \approx \big(1+\tilde{\beta}\big) r_G$, $\nu^a_{GF} = (1-\beta)\, \nu^a_G$ and $\nu^s_{GF} = \big(1+\tilde{\beta}\big)\nu^s_G$.

4.3 Effect of a type M mutation

The appearance of a type M mutation induces an increase of the mutation probability, namely $u_{GM} = (1+\gamma)\, u_G$. The carrying capacity thus remains unchanged, $K_{GM} = K_G$.

In case (A), for a mutation M appearing at time t, we have $p_{GM(t)} = p_G\,(t)$, hence $r_{GM(t)} = r_G\,(t)$. Moreover, by (6)-(7),

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

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

In case (B), $p_{GM} = p_G$, hence $r_{GM} = r_G$. We obtain with the same computation as in case (A) $\nu^a_{GM} = (1+\gamma)\, \nu^a_G$ and $\nu^s_{GM} = \big(1+\gamma\big)\nu^s_G$.

5 Influence of the mutation probability on the incidence probability

Let us compute P_u the probability of getting cancer for an individual by age T, as a function of u the individual mutation probability per cell division. If we assume that the individual does not carry any driver mutation at birth, then $P_u = 1 - q$ as defined in (4) for $G = \emptyset$.

Since only the driver mutations of type M influence the mutation probability u (see Tables 1 and 2), we can write for any genotype G, $u_G = (1+\gamma)^{\delta_G} u$, where δ_G is the number of type M driver mutations in the sequence G. As a consequence, one can factor the appearance rate of any mutation on any genotype G by u, and write $\nu_G(t) = u\tilde{\nu}_G(t)$ (or $\nu_{G(s)}(t) = u\tilde{\nu}_{G(s)}(t)$ etc.)

Note that for any sequence $U_1U_2U_3$ and any appearance time s for U_1, the probability $q_{U_1(s)U_2U_3}(t)$ is either 0 or 1 (see (14)), and does not depend on u. We can thus write $\phi_{U_1(s)U_2U_3}(t) = u\tilde{\phi}_{U_1(s)U_2U_3}(t)$, where the second factor does not depend on u. As a consequence, by (14)

$$\begin{aligned} q_{U_1(s)U_2}(t) &= \exp\left(-u\int_t^T \sum_{U_3\in\{S,F,M\}} \tilde{\phi}_{U_1(s)U_2U_3}(v)\,dv\right) \\ &=: \exp\left(-uM_{U_1U_2,s,t}\right) \\ &= 1 - uM_{U_1U_2,s,t} + u\varepsilon_u^{U_1U_2,s,t} \end{aligned} \tag{18}$$

where $K_{U_1U_2,s,t}$ does not depend on u and where for all $s,t\in[0,T]$, $0 \leqslant \varepsilon_u^{U_1U_2,s,t} \leqslant \frac{1}{2}uM_{U_1U_2,s,t}^2 \leqslant \frac{1}{2}uM_{U_1U_2,0,0}^2$, hence in particular $\lim_{u\to 0}\varepsilon_u^{U_1U_2,s,t} = 0$ uniformly on $s,t\in[0,T]$. From (13) it appears that one can write $\phi_{U_1(s)U_2}(t) = u\tilde{\varphi}_{U_1(s)U_2}(t)\left(1 - q_{U_1(s)U_2}(t)\right)$, where $\varphi_{U_1(s)U_2}(t)$ does not depend on u. It follows from (12) and (18) that

$$\begin{aligned} q_{U_1(s)}(t) &= \exp\left(-u\int_t^T \sum_{U_2\in\{S,F,M\}} \tilde{\varphi}_{U_1(s)U_2}(v)\left(1 - q_{U_1(s)U_2}(v)\right)dv\right) \\ &= \exp\left(-u^2\int_t^T \sum_{U_2\in\{S,F,M\}} \tilde{\varphi}_{U_1(s)U_2}(v)\left(M_{U_1U_2,s,v} - \varepsilon_u^{U_1U_2,s,v}\right)dv\right) \\ &=: \exp\left(-u^2\left(M_{U_1,s,t} - \eta_u^{U_1,s,t}\right)\right) \end{aligned}$$

where $M_{U_1,s,t}$ does not depend on u and where for all $s,t\in[0,T]$

$$0 \leqslant \eta_u^{U_1,s,t} \leqslant \int_0^T \sum_{U_2\in\{S,F,M\}} \tilde{\varphi}_{U_1(0)U_2}(v)\,\varepsilon_u^{U_1U_2,s,v}dv.$$

The uniform convergence of $\varepsilon_u^{U_1U_2,s,t}$ implies that $\lim_{u\to 0}\eta_u^{U_1,s,t} = 0$ uniformly on $s,t\in[0,T]$. Consequently, one can write

$$q_{U_1(s)}(t) = 1 - u^2M_{U_1,s,t} + u^2\varepsilon_u^{U_1,s,t} \tag{19}$$

where $\lim_{u\to 0}\varepsilon_u^{U_1,s,t} = 0$ uniformly on $s,t\in[0,T]$. Finally, by (4) and (19),

$$\begin{aligned} q &= \exp\left(-u^3\int_0^T \sum_{U_1\in\{S,F,M\}} \int_0^t \tilde{\nu}(s)\,\pi_{U_1}\rho_{U_1(s)}f_{U_1(s)}(t-s)\left(M_{U_1,s,t} - \varepsilon_u^{U_1,s,t}\right)dsdt\right) \\ &= 1 - u^3\mathcal{K} + u^3\varepsilon_u, \end{aligned}$$

where $\mathcal{K}$ does not depend on u and $\lim_{u\to 0}\varepsilon_u = 0$.

The probability of getting cancer before age T is thus

$$P_u \overset{u\to 0}{\sim} u^3\mathcal{K} \tag{20}$$

where $\mathcal{K}$ does not depend on u.

With the same reasoning we prove that for an individual born with a type M mutation (in which case $P_u = 1 - q_M$), this probability becomes

$$P_u \overset{u\to 0}{\sim} u^2\mathcal{K}. \tag{21}$$