Cooperation of partially-transformed clones: an invisible force behind the early stages of carcinogenesis

Alessandro Esposito

The Medical Research Council Cancer Unit, University of Cambridge, Hills Road, Cambridge CB2 0XZ, UK

Correspondence should be addressed to A.E. (email: ae275@mrc-cu.cam.ac.uk)

SUPPLEMENTARY METHODS

S.1. Analytical description of tissue organisation

With the basic assumptions described in **section 3.1** and **Fig. 1**, yhe 1-neighbourhood of an individual cell will contain 6 adjacent cells within the plane and 2 polar cells, one at the top and one at the bottom of the reference cell. The 2-neighbourhood will contain 12 cells within the plane, 6 cells in the above and bottom layers and two polar cells. The 3-neighbourhood will have 18 cells within the plane, 24, 12, 6 and 2 in the other layers. By induction, we infer that the k-neighbourhood of an individual cell in such tessellation is made of ks_0 cells within the plane, the sum of is_0 for each of the above and below layers, with i from 1 to k-1, and the two polar cells:

$$C_k = ks_0 + 2\sum_{i=1}^{k-1} is_0 + 2 \qquad \textbf{Eq. S1.1}$$

To simplify, ks_0 (s_0 here is 6) can be brought within the sum, then allowing to simplify to the correspondent triangular number:

$$C_k = -ks_0 + 2\sum_{i=1}^{k} is_0 + 2 = k^2 s_0 + 2 \qquad \textbf{Eq. S1.2}$$

It should be noted that for intercalated layers, the coordination values can be larger.

Another example, this time including a significant constraint in topology, is represented by the same topology, where only three layers are considered.

$$C_1 = s_0 + 2 \qquad \textbf{Eq. S1.3}$$

$$C_{k>1} = ks_0 + 2(k-1)s_0 = s_0(3k-2) \qquad \textbf{Eq. S1.4}$$

S.2. Probability of initiation (power function)

Let's assume the probability of tumour initiation is proportional to a concentration gradient, similar to a morphogen, or an oncogenic mitogen/morphogen field decaying as a power function:

$$p_{nk} = \frac{p_{n0}}{k^l} \qquad \textbf{Eq. S.2.1}$$

Where p_{n0} indicate the probability of tumour initiation when cells are attached (the 1-neighbourhood). Therefore, in the case of 3D hexagonal tessellation, we can derive the factor Cp_n:

$$\Omega = \sum_{k=1}^{\infty} \frac{k^2 s_0 + 2}{k^l} \quad \textbf{Eq. S.2.2}$$

This sum is carried over an infinite neighbourhood, and the validity of the results will be checked numerically (see **Section 3.2** and **Fig. 2**). First, we can expand p_{tn}:

$$\Omega = s_0 \sum_{k=1}^{\infty} k^{2-l} + 2 \sum_{k=1}^{\infty} k^{-l} \quad \textbf{Eq. S.2.3}$$

These series can now be described by Riemann Zeta functions:

$$\Omega = s_0 \zeta(l-2) + 2\zeta(l) \quad \textbf{Eq. S.2.4}$$

Let's now consider the thin 3-layer tissue which tessellation was already discussed. In this case, for one cell:

$$\Omega = s_0 + 2 + \sum_{k=2}^{\infty} \frac{s_0(3k-2)}{k^l} = 2 + \sum_{k=1}^{\infty} \frac{s_0(3k-2)}{k^l} \quad \textbf{Eq. S.2.5}$$

Following the same process described before, we can obtain:

$$\Omega = 2 - 2s_0 \sum_{k=1}^{\infty} \frac{1}{k^l} + 3s_0 \sum_{k=1}^{\infty} \frac{1}{k^{l-1}} \quad \textbf{Eq. S.2.6}$$

And,

$$\Omega = 18\zeta(l-1) - 12\zeta(l) + 2 \quad \textbf{Eq. S.2.7}$$

This describe the probability of transformation for a cell in the middle layer. We can approximate the result over the tissue equal to this value by N/3 (middle layer) and with half contribution for the top and bottom layer resulting in

$$\Omega = 12\zeta(l-1) - 8\zeta(l) + 4/3 \quad \textbf{Eq. S.2.8}$$

S.3. Probability of initiation (exponential function)

Let's now assume the oncogenic field decays as an exponential function:

$$p_{nk} = p_{n0} e^{-(k-1)k_c^{-1}} \quad \textbf{Eq. S.3.1}$$

Where k_c is a decay constant expressed in terms of k-neighbourhood for simplicity. If two cells are in contact, the probability of initiation will be p_{n0} as per definition of p_{n0}. When cells are at a k_c+1 distance, this probability is 1/e lower, *i.e.* ~30% lower. In the case of 3D hexagonal tessellation, the factor Cp_n can be now expressed as:

$$\Omega = \sum_{k=1}^{\infty} (k^2 s_0 + 2) e^{-(k-1)k_c^{-1}} \quad \textbf{Eq. S.3.2}$$

Or the sum of the series:

$$\Omega = 2 \sum_{k=1}^{\infty} e^{-(k-1)k_c^{-1}} + s_0 \sum_{k=1}^{\infty} k^2 e^{-(k-1)k_c^{-1}} \quad \textbf{Eq. S.3.3}$$

The first series converges to:

$$\sum_{k=1}^{\infty} e^{-(k-1)k_c^{-1}} = \frac{e^{k_c^{-1}}}{e^{k_c^{-1}} - 1} \quad \textbf{Eq. S.3.4}$$

The second series can be represented as:

$$\sum_{k=1}^{\infty} k^2 e^{-(k-1)k_c^{-1}} = e^{k_c^{-1}} \sum_{k=1}^{\infty} k^2 e^{-k/k_c} = \frac{e^{2k_c^{-1}}\left(e^{k_c^{-1}}+1\right)}{\left[e^{k_c^{-1}}-1\right]^3} \qquad \textbf{Eq. S.3.5}$$

Therefore,

$$\Omega = e^{k_c^{-1}} \frac{(2+s_0)e^{2k_c^{-1}}+(4+s_0)e^{k_c^{-1}}+2}{\left[e^{k_c^{-1}}-1\right]^3} \qquad \textbf{Eq. S.3.6}$$

With $s_0 = 6$, once again to confirm mathematical consistency, $\lim_{k_c \to 0} \Omega = 8$, as in the case where only adjacent cells are important. Shallower decays will again increase this value (see **Fig. 2**).

S.4 Probability of initiation (generalisation)

We have characterised the oncogenic field in relation to typical descriptions of morphogenic gradients [22]. While relevant for specific cases, steady-state concentration gradients of shared resources in space, generated by passive diffusion and linear or non-linear degradation, can adopt different shapes. One useful analytical description is represented by concentrations that decay as the product of exponential and power-law functions, for instance as:

$$p_{nk} = \frac{p_{n0}}{k^l} e^{-(k-1)k_c^{-1}} \qquad \textbf{Eq. S.4.1}$$

With the same formalism and strategies described in **Sections S.2** and **S.3** we can show that, for a three-dimensional tissue:

$$\Omega = \sum_{k=1}^{\infty} (k^2 s_0 + 2) k^{-l} e^{-(k-1)k_c^{-1}} \qquad \textbf{Eq. S.4.2}$$

This analytical representation of Ω can be expressed as sums of polylogarithm functions:

$$\Omega = e^{k_c^{-1}} \left[2Li_l\left(e^{-k_c^{-1}}\right) + s_0 Li_{l-2}\left(e^{-k_c^{-1}}\right)\right] \qquad \textbf{Eq. S.4.3}$$

This representation converges to those shown in **Sections S.2** and **S.3** in the cases where k_c is very large or where *l* is very small, respectively, *i.e.* when the power-law or the exponential decay components are negligible. The case l=1 represents an oncogenic field induced by continuous point-sources in an unconstrained three-dimensional space in the presence of linear degradation. In this geometry:

$$\Omega = e^{k_c^{-1}} \left[-2\log\left(1 - e^{-k_c^{-1}}\right) + s_0 \frac{e^{-k_c^{-1}}}{\left(1-e^{-k_c^{-1}}\right)^2}\right] \qquad \textbf{Eq. S.4.4}$$

S.5. Cell-autonomous time-horizon (discrete-time Markov chain)

The mutational process illustrated in this work can be modelled as a discrete-time Markov process (see also **Sup. File 'firstpassageproblem_v2.nb'** or **'firstpassageproblem_v2.pdf'** in the GitHub repository ***alesposito/CloE-PE*** [50] for the Mathematica Notebook used in this work and the **peer-review open documentation** for related discussion). Each cell is described by four states: wild-type (W), mutant X, mutant Y, and double-mutant (XY). At any given time, the transition matrix between these states is:

$$T = \begin{pmatrix} 1-(x+y)\rho_0 - xy\rho_0^2 & x\rho_0 & y\rho_0 & xy\rho_0^2 \\ 0 & 1-y\rho_0 & 0 & y\rho_0 \\ 0 & 0 & 1-x\rho_0 & x\rho_0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \qquad \textbf{Eq. S.5.1}$$

where the probability of acquiring the mutation *X* or *Y* are independent and directly proportional to the mutational rate ρ_0 with proportionality constants *x* and *y*, respectively. Initially, the system is described by the state vector $S_0 = (1 \quad 0 \quad 0 \quad 0)$, *i.e.* all cells are wild-type. The probability of observing a double-mutant XY at day t (with $t \in \mathbb{N}$) is:

$$p_{XY}(t) = 1 - \frac{(1-x\rho_0)^t}{1+x\rho_0} - \frac{(1-y\rho_0)^t}{1+y\rho_0} + \frac{(1-xy\rho_0^2)(1-x\rho_0-y\rho_0-xy\rho_0^2)^t}{(1+x\rho_0)(1+y\rho_0)} \qquad \textbf{Eq. S.5.2}$$

As $\rho_0 \ll 1$, $p_{XY}(t)$ can be well-approximated with a second-order (or the order matching the number of mutations for **Eq. S.5.9**) element of a Taylor series:

$$p_{XY}(t) \approx \mathrm{t}^2 xy\rho_0^2 \qquad \textbf{Eq. S.5.3}$$

The probability of not observing any XY mutant in a population of N cells will be therefore $(1-\mathrm{t}^2 xy\rho_0^2)^N$, and the probability of observing a double-mutant after t days will be thus:

$$P_{XY}(t) \approx 1-(1-\mathrm{t}^2 xy\rho_0^2)^N \approx 1-\mathrm{e}^{-\mathrm{t}^2 Nxy\rho_0^2} \qquad \textbf{Eq. S.5.4}$$

The probability $d_{XY}(t)$ to observe a first XY mutant can be then evaluated by differentiating **Eq. S.5.4**.

$$d_{XY}(t) = \frac{\partial P_{XY}}{\partial \mathrm{t}} = 2\mathrm{t}Nxy\rho_0^2 \mathrm{e}^{-\mathrm{t}^2 Nxy\rho_0^2} \qquad \textbf{Eq. S.5.5}$$

The expectation for the average latency of the first double-mutant XY can be then evaluated as:

$$\langle t_{XY} \rangle = \int_0^\infty t d_{XY}(t) dt = \frac{1}{2\rho_0}\sqrt{\frac{\pi}{xyN}} \qquad \textbf{Eq. S.5.6}$$

For a two-hits model, the cell-autonomous time-horizon (t_a) and the time at which the first CD clone might appear can be then described by **Eq. S.5.6** with x=1, and y=1 or y=Ω, respectively.

$$t_a = \langle t_{AB} \rangle = \frac{1}{2\rho_0}\sqrt{\frac{\pi}{N}} \qquad \textbf{Eq. S.5.7}$$

$$\langle t_{CD} \rangle = \frac{1}{2\rho_0}\sqrt{\frac{\pi}{\Omega N}} = t_a \Omega^{-0.5} \qquad \textbf{Eq. S.5.8}$$

We note that the scaling factor $\Omega^{-0.5}$ in **Eq. S.5.8** is the factor $(xy)^{-1/2}$ in **Eq. S.5.6** with x=1 and y=Ω. In the Mathematica notebook we also show that for three mutations (X, Y, Z) the scaling factors is $(xyz)^{-1/3}$. We infer that if we define an average or apparent oncogenic field effect Ω, the scaling factor between the first passage time of a clones cooperating by paracrine effects and clone accruing a similar number of mutations within a single cell would be of the form:

$$\Omega^{-d/m} \qquad \textbf{Eq. S.5.9}$$

in which *m* is the number of mutations required for transformation and *d* is the number of mutations which effect is mediated through, for example, diffusible molecules.