Supplementary Information for

A compartment size dependent selective threshold limits mutation accumulation in hierarchical tissues

Dániel Grajzel, Imre Derényi, and Gergely J. Szöllősi

Gergely J. Szöllősi
E-mail: ssolo@elte.hu

This PDF file includes:

www.pnas.org/cgi/doi/10.1073/pnas.1913104117

Supporting Information Text

Sufficient conditions for conservative cell number dynamics

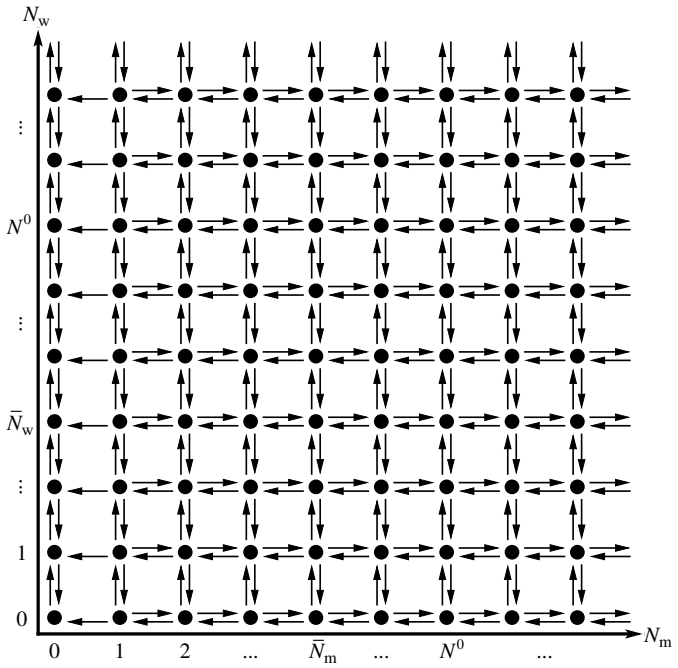

Fig. S1. The state space, where the allowed states $(N_{\mathrm{m}}, N_{\mathrm{w}})$ are denoted by black circles, and the transitions with non-zero rates by black arrows.

Consider the cell number dynamics where (i) both the mutant and wild type transition rates depend only on the number of cells of the given type; and (ii) cell number regulation acts as a multiplicative modifier of these rates and depends only on the total number of cells:

$$
\begin{aligned}
k^+_{\mathrm{m}}(N_{\mathrm{m}}, N_{\mathrm{w}}) &= R^+_{\mathrm{m}}(N_{\mathrm{m}})\rho^+(N_{\mathrm{m}} + N_{\mathrm{w}}),\\
k^-_{\mathrm{m}}(N_{\mathrm{m}}, N_{\mathrm{w}}) &= R^-_{\mathrm{m}}(N_{\mathrm{m}})\rho^-(N_{\mathrm{m}} + N_{\mathrm{w}}),\\
k^+_{\mathrm{w}}(N_{\mathrm{m}}, N_{\mathrm{w}}) &= R^+_{\mathrm{w}}(N_{\mathrm{w}})\rho^+(N_{\mathrm{m}} + N_{\mathrm{w}}),\\
k^-_{\mathrm{w}}(N_{\mathrm{m}}, N_{\mathrm{w}}) &= R^-_{\mathrm{w}}(N_{\mathrm{w}})\rho^-(N_{\mathrm{m}} + N_{\mathrm{w}}),
\end{aligned}
$$

where $R^+_{\mathrm{m}}(N_{\mathrm{m}})$ and $R^-_{\mathrm{m}}(N_{\mathrm{m}})$ indicate the rates of the transitions in the number of mutant cells: $N_{\mathrm{m}} \to N_{\mathrm{m}} + 1$ and $N_{\mathrm{m}} \to N_{\mathrm{m}} - 1$, respectively, and $R^+_{\mathrm{w}}(N_{\mathrm{w}})$ and $R^-_{\mathrm{w}}(N_{\mathrm{w}})$ indicate the rates of the transitions of wild type cells: $N_{\mathrm{w}} \to N_{\mathrm{w}} + 1$ and $N_{\mathrm{w}} \to N_{\mathrm{w}} - 1$. The state space of the tissue compartment under consideration is depicted in Fig. S1, where all the transitions with nonzero rates are indicated by arrows. Starting from an arbitrary population state $N_{\mathrm{m}} = i$, $N_{\mathrm{w}} = j$ the product of the equilibrium constants (i.e., the ratios between to forward and backward rates) along an elementary cycle $(i, j) \to (i, j + 1) \to (i - 1, j + 1) \to (i - 1, j) \to (i, j)$ is unity:

$$
\begin{aligned}
&\frac{k^+_{\mathrm{w}}(i,j)}{k^-_{\mathrm{w}}(i,j+1)} \times \frac{k^-_{\mathrm{m}}(i,j+1)}{k^+_{\mathrm{m}}(i-1,j+1)} \times \frac{k^-_{\mathrm{w}}(i-1,j+1)}{k^+_{\mathrm{w}}(i-1,j)} \times \frac{k^+_{\mathrm{m}}(i-1,j)}{k^-_{\mathrm{m}}(i,j)} =\\
&= \frac{R^+_{\mathrm{w}}(j)\rho^+(i+j)}{R^-_{\mathrm{w}}(j+1)\rho^-(i+j+1)} \times \frac{R^-_{\mathrm{m}}(i)\rho^-(i+j+1)}{R^+_{\mathrm{m}}(i-1)\rho^+(i+j)} \times \frac{R^-_{\mathrm{w}}(j+1)\rho^-(i+j)}{R^+_{\mathrm{w}}(j)\rho^+(i+j-1)} \times \frac{R^+_{\mathrm{m}}(i-1)\rho^+(i+j-1)}{R^-_{\mathrm{m}}(i)\rho^-(i+j)} =\\
&= \left(\frac{R^+_{\mathrm{w}}(j)}{R^-_{\mathrm{w}}(j+1)}\frac{R^-_{\mathrm{w}}(j+1)}{R^+_{\mathrm{w}}(j)}\right) \times \left(\frac{R^-_{\mathrm{m}}(i)}{R^+_{\mathrm{m}}(i-1)}\frac{R^+_{\mathrm{m}}(i-1)}{R^-_{\mathrm{m}}(i)}\right) \times \left(\frac{\rho^+(i+j)}{\rho^-(i+j+1)}\frac{\rho^-(i+j+1)}{\rho^+(i+j)}\frac{\rho^-(i+j)}{\rho^+(i+j-1)}\frac{\rho^+(i+j-1)}{\rho^-(i+j)}\right) =\\
&= (1) \times (1) \times (1) = 1.
\end{aligned}
$$

This is equivalent to Kolmogorov's criterion, and it follows that the cell number dynamics is conservative and described by a potential.

More generally, consider an evolutionary dynamics with an arbitrary number of cell types in a potential. Similarly to the above derivation, it can be shown that if the cell number increasing and decreasing rates for any subset of cell types are multiplied by, respectively, cell number increasing and decreasing rate modifiers that depend only on the total number of cells of the given subset, then the resulting dynamics is also described by a potential. Moreover, the resulting potential is the sum of the two potentials corresponding to the original dynamics and to the rate modifiers. This is because (i) along any path in the state space the product of the equilibrium constants can be factored into the products of the equilibrium constants corresponding to the original dynamics and to the rate modifiers; and (ii) along any elementary cycle both products are unity. As a corollary, if the transition rates of an evolutionary dynamics are factorizable into functions such that each function corresponds to the increase or decrease of the number of cells of a subset of cell types, and each function depends only on that cell number, then the dynamics is conservative and described by a potential.

Derivation of the potential $\Psi(N_\mathrm{m}, N_\mathrm{w})$

For the transition rates considered in the main text, the potential $\Psi(N_\mathrm{m}, N_\mathrm{w})$ corresponding to the dynamics can be calculated as the logarithm of the product of the equilibrium constants from point $(N_\mathrm{m}, N_\mathrm{w})$ to a reference point, e.g., to the lower left state $(1, 0)$ of the reversible part of the state space, up to an additive constant. Taking the path $(N_\mathrm{m}, N_\mathrm{w}) \to (1, N_\mathrm{w}) \to (1, 0)$, this product can be written as:

$$
\begin{aligned}
&\exp\left[\Psi(N_\mathrm{m}, N_\mathrm{w}) - \Psi(1,0)\right] = \\
&= \prod_{i=1}^{N_\mathrm{m}-1} \frac{k_\mathrm{m}^-(i+1, N_\mathrm{w})}{k_\mathrm{m}^+(i, N_\mathrm{w})} \prod_{j=0}^{N_\mathrm{w}-1} \frac{k_\mathrm{w}^-(1, j+1)}{k_\mathrm{w}^+(1, j)} = \\
&= \prod_{i=1}^{N_\mathrm{m}-1} \frac{(i+1)r^-}{i r_\mathrm{m}^+} \prod_{j=0}^{N_\mathrm{w}-1} \frac{(j+1)r^-}{jr^+ + \delta^+} \exp\left[U(N_\mathrm{m}+N_\mathrm{w}) - U(1)\right] = \\
&= \left(\frac{r^-}{r_\mathrm{m}^+}\right)^{(N_\mathrm{m}-1)} N_\mathrm{m} \left(\frac{r^-}{r^+}\right)^{N_\mathrm{w}} \prod_{j=1}^{N_\mathrm{w}} \frac{j}{\delta^+/r^+ + j - 1} \exp\left[U(N_\mathrm{m}+N_\mathrm{w}) - U(1)\right] = \\
&= \left(\frac{r^-}{r_\mathrm{m}^+}\right)^{(N_\mathrm{m}-1)} N_\mathrm{m} \left(\frac{r^-}{r^+}\right)^{N_\mathrm{w}} \frac{\Gamma(N_\mathrm{w}+1)}{\Gamma(\delta^+/r^+ + N_\mathrm{w})/\Gamma(\delta^+/r^+)} \exp\left[U(N_\mathrm{m}+N_\mathrm{w}) - U(1)\right] = \\
&= \left\{\left(\frac{r^-}{r_\mathrm{m}^+}\right)^{N_\mathrm{m}} N_\mathrm{m} \left(\frac{r^-}{r^+}\right)^{N_\mathrm{w}} \frac{\Gamma(N_\mathrm{w}+1)}{\Gamma(\delta^+/r^+ + N_\mathrm{w})} \exp\left[U(N_\mathrm{m}+N_\mathrm{w})\right]\right\} \Bigg/ \left\{\left(\frac{r^-}{r_\mathrm{m}^+}\right)^{1} \frac{\Gamma(1)}{\Gamma(\delta^+/r^+)} \exp\left[U(1)\right]\right\} = \\
&= \left\{\left(\frac{r^-}{r_\mathrm{m}^+}\right)^{N_\mathrm{m}} N_\mathrm{m} \left(\frac{r^-}{r^+}\right)^{N_\mathrm{w}} \frac{\Gamma(N_\mathrm{w}+1)}{\Gamma(N^0(\mathrm{e}^{-\phi}-1) + N_\mathrm{w})} \exp\left[U(N_\mathrm{m}+N_\mathrm{w})\right]\right\} \Bigg/ \left\{\left(\frac{r^-}{r_\mathrm{m}^+}\right)^{1} \frac{\Gamma(1)}{\Gamma(N^0(\mathrm{e}^{-\phi}-1))} \exp\left[U(1)\right]\right\},
\end{aligned}
$$

where Γ represents the gamma function and in the last step we use equations (1) and (5) of the main text. The logarithm of the numerator (between the braces) results in the formula for $\Psi(N_\mathrm{m}, N_\mathrm{w})$ given in Eq. (9) of the main text, while the logarithm of the denominator is identical to $\Psi(1, 0)$.

Mean exit time

The derivation of the mean exit time from the effective potential well near the quasi-stationary state $(\bar{N}_\mathrm{m}, \bar{N}_\mathrm{w})$ to the boundary line corresponding to zero mutants ($N_\mathrm{m} = 0$) is outlined below. Exact analytical formula exists either for continuous systems of arbitrary dimensions (subsections 5.2.7 and 9.3.2 of Gardiner (1)), or discrete systems in one dimension (section 7.4 of Gardiner (1) and Derényi et al. (2)). One can, however, generalize the discrete one-dimensional formula to our discrete two-dimensional system in a straightforward manner.

Let us first select any of the shortest paths (with only upward and leftward transitions) form $(\bar{N}_\mathrm{m}, \bar{N}_\mathrm{w})$ through $(1, N^0)$ to $(0, N^0)$. The main contribution to the mean exit time (which, in one dimension, is often referred to as "mean first passage time") along this one-dimensional path comes from the product of all backward transition rates (except for the outermost one from $(0, N^0)$ to $(1, N^0)$) divided by the product of all forward transition rates (including the one from $(1, N^0)$ to $(0, N^0)$). This

contribution is independent of the selected path (because the rates correspond to an effective potential, $\Psi(N_\mathrm{m}, N_\mathrm{w})$):

$$\tau_\mathrm{main} = \frac{1}{k_\mathrm{m}^-(1, N^0)} \exp(\Delta\Psi),$$

where

$$\Delta\Psi = \Psi(1, N^0) - \Psi(\bar{N}_\mathrm{m}, \bar{N}_\mathrm{w})$$

is the height of the effective potential barrier against escape.

The mean exit time involves the sum of similar contributions between any pairs of states along the selected path (section 7.4 of Gardiner (1) and Derényi et al. (2)). Only those contributions are significant for which the starting and ending positions are close to the bottom and the top of the effective potential, respectively. The summation for these contributions leads to a correction factor to the main contribution. This correction factor consists of two terms: a sum of the $\exp\{-[\Psi(N_\mathrm{m}, N_\mathrm{w}) - \Psi(\bar{N}_\mathrm{m}, \bar{N}_\mathrm{w})]\}$ Boltzmann weights of the states $(N_\mathrm{m}, N_\mathrm{w})$ along the path near the bottom; and a conceptually similar sum near the top (not detailed here, because its terms cannot be readily expressed by the potential Ψ, but rather only by the products of the ratios of the corresponding transition rates).

Because a typical exit process follows the diagonally oriented potential valley of the state space (see Fig. 2 in the main text), let us restrict the selected path to run along this valley. The correction factor for such a path will depend only on the local "geometry" of the bottom and the top of the effective potential, parallel to the direction of the main valley. In particular, the correction factor at the top (which is the effective width of the potential barrier) can be approximated as

$$C_\mathrm{top}^{||} = \frac{1}{S_\mathrm{m}}.$$

This one-dimensional result can be generalized to two (or any higher) dimensions in analogy to the generalization of the one-dimensional continuous version of the mean exit time (see subsection 9.3.2 of Gardiner (1)): the sum of the above Boltzmann weights should be extended to all states near the bottom of the effective potential:

$$C_\mathrm{bottom} = \sum_i \sum_j \exp\{-[\Psi(i, j) - \Psi(\bar{N}_\mathrm{m}, \bar{N}_\mathrm{w})]\};$$

and the result should be divided by the sum of a different type of Boltzmann weights for all the states along the exit line, $N_\mathrm{m} = 1$ (which gives the effective width of the potential saddle at the barrier):

$$C_\mathrm{top}^{\perp} = \sum_j \exp\{-[\Psi(1, j) - \Psi(1, N^0)]\}.$$

The resulting mean exit time

$$\tau = \tau_\mathrm{main} \frac{C_\mathrm{bottom} C_\mathrm{top}^{||}}{C_\mathrm{top}^{\perp}} = \frac{1}{k_\mathrm{m}^-(1, N^0)} \frac{C_\mathrm{bottom} C_\mathrm{top}^{||}}{C_\mathrm{top}^{\perp}} \exp(\Delta\Psi)$$

is proportional to $\exp(\Delta\Psi)$ and its prefactor, denoted by τ_0, is often referred to as the reciprocal of the attempt frequency. The summations can be approximated by closed formulas using quadratic approximations for the effective potential near the bottom and the top. However, to achieve higher accuracy we executed the summations numerically for displaying the theoretical estimates in the figures of both the main text and the Supplementary Information.

Simulations

Two types of simulations were developed to study the cell dynamics and measure the persistence probability of mutants in hierarchically organized tissues:

Explicit kinetics. We performed explicit kinetic Monte Carlo simulations (also known as the "Gillespie algorithm"), where the number of mutants N_m and wild type cells N_w evolved according to the rates described in Eq. (8) of the main text. At each iteration, rates were calculated based on the current value of N_m and N_w, one of the four possible events was chosen proportional to its rate, and N_m or N_w was changed accordingly, while time was increased by the reciprocal of the sum of the four rates. The simulations were started with $N_\mathrm{m} = 0$ and $N_\mathrm{w} = N^0$. At time $t = 20000$, measured in units of $1/(2r^- N^0)$, a single mutant was introduced by setting $N_\mathrm{m} = 1$. The simulations were stopped after reaching time $T = 10^9$. Data used in Fig. 2 of the main text were generated using the explicit kinetic simulation.

Efficiently measuring the persistence probability. The time evolution of the probability distribution $P_{i,j}(t)$ of the number of (i mutant and j wild type) cells is described by the master equation defined by the transition rates given in Eq. (8) of the main text. The inward and outward probability currents into and out of state (i,j) can be written, respectively, as

$$J^{\text{in}}_{i,j}(t) = k^{+}_{\text{m}}(i-1,j)P_{i-1,j}(t) + k^{-}_{\text{m}}(i+1,j)P_{i+1,j}(t) + k^{+}_{\text{w}}(i,j-1)P_{i,j-1}(t) + k^{-}_{\text{w}}(i,j+1)P_{i,j+1}(t),$$
$$J^{\text{out}}_{i,j}(t) = \left[k^{-}_{\text{m}}(i,j) + k^{+}_{\text{m}}(i,j) + k^{-}_{\text{w}}(i,j) + k^{+}_{\text{w}}(i,j)\right] P_{i,j}(t).$$

To efficiently measure the spreading probability (S_{m}) and the mean exit time of escape from the quasi-stationary state (τ), we calculated the steady state probability distributions for the following two modified process:

Process I. The process starts with a single mutant and N^0 wild type cells (green circle in Fig. S2), and each time mutants would either go extinct (red arrows in Fig. S2) or spread and reach their quasi-stationary number $\bar{N}_{\text{m}}$ (blue arrows in Fig. S2), the process is restarted in the $(1, N^0)$ state.

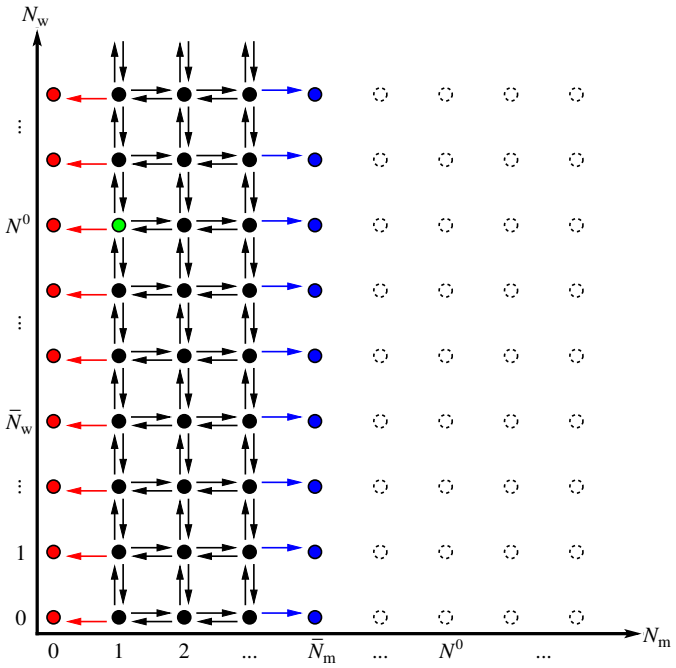

Fig. S2. The state space corresponding to Process I.

The extinction and spreading currents can be defined, respectively, as:

$$J^{\text{ext}}(t) = \sum_{j=0}^{\infty} k^{-}_{\text{m}}(1,j)P_{1,j}(t),$$
$$J^{\text{spread}}(t) = \sum_{j=0}^{\infty} k^{+}_{\text{m}}(\bar{N}_{\text{m}}-1,j)P_{\bar{N}_{\text{m}}-1,j}(t).$$

Thus the master equation of the process for the states with more than 0 but less than $\bar{N}_{\text{m}}$ mutants (solid black and green circles in Fig. S2) is:

$$\frac{\text{d}}{\text{d}t}P_{i,j}(t) = J^{\text{in}}_{i,j}(t) - J^{\text{out}}_{i,j}(t) + \delta_{i,1}\delta_{j,N^0}\left[J^{\text{ext}}(t) + J^{\text{spread}}(t)\right],$$

where the Kronecker delta symbol $\delta_{j,i}$ is 1 if $j = i$, and 0 otherwise.

Numerically solving the master equation (using Euler's method and setting a large enough upper bound for the number of wild type cells), the probability distribution together with the extinction and spreading currents converge to their steady state values, denoted by $\hat{P}_{i,j}$, $\hat{J}^{\text{ext}}$, and $\hat{J}^{\text{spread}}$, respectively. The probability that a single mutant can spread and avoid early stochastic extinction is, thus:

$$\hat{S}_{\text{m}} = \frac{\hat{J}^{\text{spread}}}{\hat{J}^{\text{ext}} + \hat{J}^{\text{spread}}}.$$

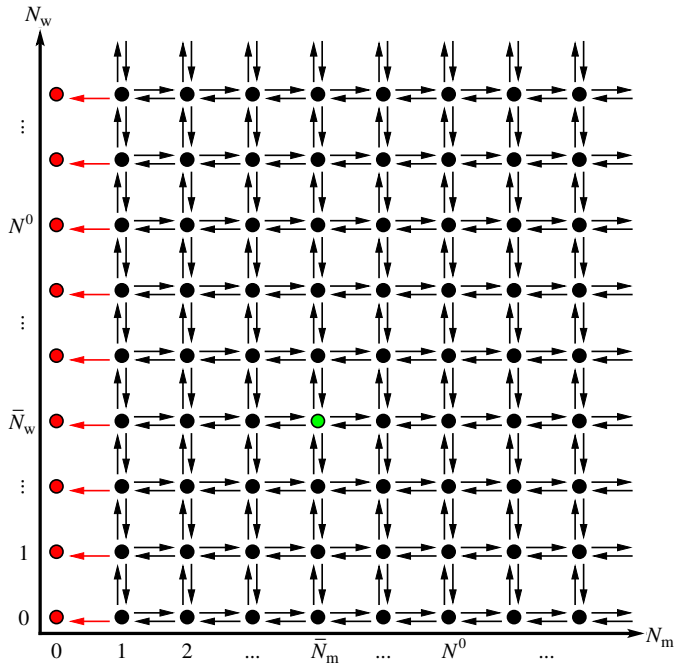

Fig. S3. The state space corresponding to Process II.

Process II. The process starts in the quasi-stationary state (green circle in Fig. S3), i.e., with $\bar{N}_m$ mutants and $\bar{N}_w$ wild type cells, and each time mutants would go extinct (red arrows in Fig. S3) the process is restarted in the quasi-stationary state $(\bar{N}_m, \bar{N}_w)$.

With the extinction current

$$J^{ext}(t) = \sum_{j=0}^{\infty} k_m^-(1,j) P_{1,j}(t)$$

the master equation of the process for all the states with non-zero mutants is:

$$\frac{d}{dt} P_{i,j}(t) = J_{i,j}^{in}(t) - J_{i,j}^{out}(t) + \delta_{i,\bar{N}_m} \delta_{j,\bar{N}_w} J^{ext}(t).$$

Numerically solving the master equation, the probability distribution together with the extinction current converge to their steady state values $\hat{P}_{i,j}$ and $\hat{J}^{ext}$, respectively. The reciprocal of the steady state extinction current provides the mean exit time of the mutants from the quasi-stationary state:

$$\hat{\tau} = \frac{1}{\hat{J}^{ext}}.$$

Persistence probability. With the combination of the above two processes the probability P that a single mutant persists (i.e., first spreads, and then avoids escape) for the lifetime T of the individual can be expressed as:

$$P = \hat{S}_m \exp(-T/\hat{\tau}). \quad [1]$$

Data used for Fig. 3 of the main text and Fig. S4 below were produced by this method.

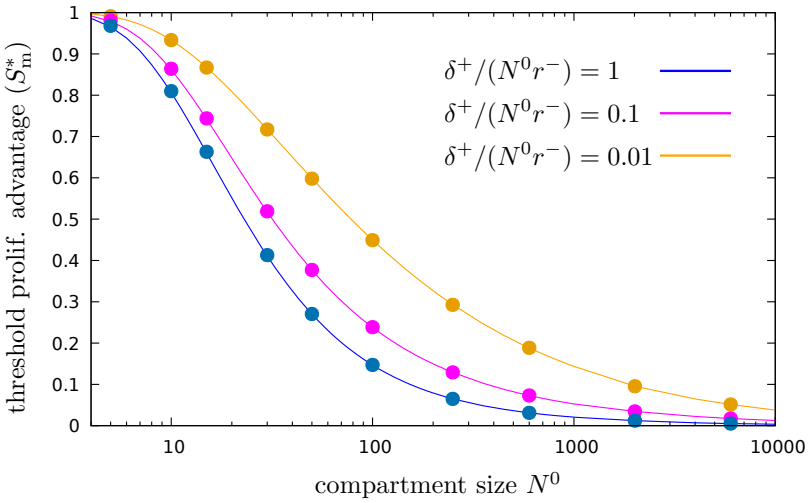

Fig. S4. The threshold spreading factor for varying strength of washing out, corresponding to $\delta^+/(N^0 r^-) = 1$, 0.1, and 0.01, i.e., to 100%, 10%, and 1% of cells being produced by differentiation from below instead of self-renewal or, equivalently, $\phi = 1 - ln[\delta^+/(N^0 r^-)] \approx -\infty$, -0.1, and -0.01, respectively, for $\beta = 10$ and $T = 10^9$. Similarly to Fig. 3b in the main text the threshold spreading factor separates the plot into two distinct regimes: below the curve the persistence probability is close to zero, mutations cannot accumulate; while above the curve mutants that avoid early stochastic extinction, which occurs with probability S_m, will persist in the tissue during the lifetime of the individual, and can accumulate further mutations leading to neoplastic progression. Continuous lines are theoretical estimates based on the mean exit time approximation, while points indicate explicit numerical simulations.

References

1. C Gardiner. *Handbook of stochastic methods for physics, chemistry and natural sciences*. Springer. Berlin, 2004.
2. Imre Derényi, Denis Bartolo, and Armand Ajdari. Effects of intermediate bound states in dynamic force spectroscopy. *Biophysical journal*, 86(3):1263–1269, 2004.
3. Christopher S Potten and Markus Loeffler. Stem cells: attributes, cycles, spirals, pitfalls and uncertainties. lessons for and from the crypt. *Development*, 110(4):1001–1020, 1990.
4. Carmen Pin, Alastair JM Watson, and Simon R Carding. Modelling the spatio-temporal cell dynamics reveals novel insights on cell differentiation and proliferation in the small intestinal crypt. *PloS one*, 7(5):e37115, 2012.
5. Peter Buske, Jörg Galle, Nick Barker, Gabriela Aust, Hans Clevers, and Markus Loeffler. A comprehensive model of the spatio-temporal stem cell and tissue organisation in the intestinal crypt. *PLoS computational biology*, 7(1):e1001045, 2011.
6. Nick Barker, Johan H Van Es, Jeroen Kuipers, Pekka Kujala, Maaike Van Den Born, Miranda Cozijnsen, Andrea Haegebarth, Jeroen Korving, Harry Begthel, Peter J Peters, et al. Identification of stem cells in small intestine and colon by marker gene lgr5. *Nature*, 449(7165):1003, 2007.
7. Christopher S Potten, M Kellett, Stephen A Roberts, DA Rew, and GD Wilson. Measurement of in vivo proliferation in human colorectal mucosa using bromodeoxyuridine. *Gut*, 33(1):71–78, 1992.
8. Rafael Bravo and David E Axelrod. A calibrated agent-based computer model of stochastic cell dynamics in normal human colon crypts useful for in silico experiments. *Theoretical Biology and Medical Modelling*, 10(1):66, 2013.
9. Pierre Nicolas, Kyoung-Mee Kim, Darryl Shibata, and Simon Tavaré. The stem cell population of the human colon crypt: analysis via methylation patterns. *PLoS computational biology*, 3(3):e28, 2007.