# Epidermal preneoplasia: UVB exposure tilts the stochastic fate of *p53* mutant progenitors toward proliferation

## Supplementary theory

### S-I. CLONE SIZE DISTRIBUTIONS DURING ONGOING INDUCTION

Here we derive Eq. (1) of the main text, together with generalisations of the result to conditions of nonlinear clone induction frequency, (section S-I B), and cases of divergent clone fate (section S-I C). As in the main text, we define the surviving clone size distribution, $g_n(t-t_1)$, as the probability of finding a clone of size $n > 0$ following a single pulse of UVB radiation at some earlier time, $t_1$; the average clone size of this population is then $N(t-t_1) = \sum_{n=1}^{\infty} n g_n(t-t_1)$.

#### A. Simple derivation of clone size distribution

To begin, we first derive results for the simplified case presented in the main text, whereby each clone has precisely the average number of cells, $N(t-t_1)$. Within this approximation, $g_n(t)$ has the simple form $g_n(t) = \delta_{n,N(t)}$, where $\delta_{n,m}$ is the Kronecker delta function. We initially assume that the induction rate is constant during the period of UVB exposure. The clone size distribution after a time $t$ of ongoing UVB radiation is given by the integral:

$$\begin{aligned} P_n(t) &= \frac{1}{t}\int_0^t g_n(t-t_1)dt_1 \\ &\approx \frac{1}{t}\int_0^t \delta[n-N(t_1)]dt_1\,, \end{aligned} \tag{S1}$$

where in the second line we have substituted the Kronecker delta by a Dirac delta function — this substitution is a mathematical approximation, with no biological implications, which treats the clone size $n$ as a continuous (rather than discrete) variable. We then make use of the identity $\delta[n-N(t)] = \delta[\phi_n - t]/|N'(t)|$ to simplify Eq. (S1) [1]. Here $N'(t)$ is the derivative of $N(t)$, and $\phi_n$ denotes the inverse to the function $N(t)$, defined such that $N(\phi_n) \equiv n$. Applying the identity, and assuming that the average clone size grows monotonically (so that $|N'(t)| = N'(t)$), we obtain,

$$\begin{aligned} P_n(t) &\approx \frac{1}{t}\int_0^t \frac{1}{N'(t_1)}\delta[\phi_n - t_1]dt_1 \\ &= \begin{cases} \frac{1}{tN'(\phi_n)} & \text{for } n \le N(t) \\ 0 & \text{for } n > N(t) \end{cases}, \end{aligned} \tag{S2}$$

which is the result given in Eq. (1) of the main text.

To derive the explicit expressions for $P_n(t)$ given in Eq. (2), we evaluate $N'(t)$ and $\phi_n$ for cases of exponential and polynomial growth. For the former, with $N(t) = \exp(\nu t)$, then $N'(t) = \nu N(t)$ and $\phi_n = \ln(n)/\nu$, giving $N'(\phi_n) = \nu n$. Substituting this expression into Eq. (1) directly gives Eq. (2) with $\beta = 1$. Similarly, for the latter, with $N(t) \simeq (\nu t)^\alpha$, it follows that $N'(t) = \alpha N(t)/t$ and $\phi_n \propto n^{1/\alpha}/\nu$, giving $N'(\phi_n) = \alpha\nu n^{(\alpha-1)/\alpha}$. Eq. (2) again follows directly with $\beta = (\alpha-1)/\alpha$.

### B. Clone size distributions with nonlinear clone induction frequency

We now address the possibility that *p53*-mutant clones do not expand directly upon mutation of *p53*, but that additional UVB-induced mutations (or other forms of injury, such as apoptosis of a nearby cell) are required to trigger growth. In this case, the effective rate of clone induction may be super-linear, as seen in multiple-hit models of cancer induction. To incorporate the possibility of non-uniform induction of clones, we modify Eq. (S1) by weighting the distribution $g_n(t-t_1)$ by the rate of clone induction at time $t_1$. In a situation typical of multiple-hit models, the number of clones that have appeared after a time $t$ can be approximated by a power law, $t^a$, where the exponent $a$ gives the number of UVB-induced events required to trigger clone expansion. In this case, the clone size distribution is given by,

$$P_n(t) \;=\; \frac{a}{t^a}\int_0^t g_n(t-t_1)t_1^{a-1}dt_1.$$

Repeating the calculation as in Eqs. (S1), (S2), we find,

$$P_n(t) = \begin{cases} \frac{a(t-\phi_n)^{a-1}}{t^a N'(\phi_n)} & \text{for } n \le N(t) \\ 0 & \text{for } n > N(t) \end{cases} . \tag{S3}$$

This result is similar to that in Eq. (1), but multiplied by an additional $n$-dependent factor $(t-\phi_n)^{a-1}$. When the exponent $a$ is low (say $a \lesssim 4$), this factor varies slowly for $n \ll N(t)$, leaving the behaviour of $P_n(t)$ largely insensitive to the non-linearity in the clone induction rate. Therefore, the power law behaviours reported in the main text (Eqs. (1)-(2)) would provide good approximations of the predicted clone size distributions even for multiple-hit models of clone induction, provided that the number of hits required to induce clone growth is low. (Note that we are referring here to the number of mutations required to induce growth of *p53* mutant clones (PMC); additional mutations may be required to transform a PMC into a tumour).

### C. Clone size distribution with diverging clone fates

We now revisit Eq. (S1) for cases where $g_n(t)$ can no longer be approximated by a delta function, for example as shown in Figure 1D. It is important to note that Eq. (1) is not valid for arbitrary forms of $g_n(t)$. Therefore, we will focus on cases of exponential or power-law growth of the average clone size $N(t)$, which are of particular interest in this paper, and in particular on two general classes of distributions for which variants of Eq. (1) hold true, and which are likely to include most biologically-relevant processes.

We first consider the general class of distributions that maintain a fixed form over time, viz. $g_n(t) \approx h[n - N(t)]$. The distribution function $h(n)$ may represent say a normal distribution with a variance that reflects fluctuations in the behaviour of individual clones about the average $N(t)$. Substituting this form of $g_n(t)$ into Eq. (S1), we find,

$$\begin{aligned} P_n(t) &\approx \frac{1}{t}\int_0^t h[n - N(t_1)]dt_1 \\ &= \frac{1}{t}\int_0^t \left[h(n) * \delta\left[n - N(t_1)\right]\right] dt_1 \\ &= h(n) * \frac{1}{t}\int_0^t \delta[n - N(t_1)]dt_1 \,, \end{aligned}$$

where '$*$' is the convolution operator. This clone size distribution has a similar form to that derived above. The effect of the convolution is to smooth the distribution; such smoothing has no effect on the power-law form of the distribution. However, its effect will be pronounced near $n = N(t)$, where the sharp cut-off shown in Eq. (1) will be replaced by a smooth decay. Note that when the distribution is sharp, $h(n) \approx \delta(n)$, then the original expression for $P_n(t)$ is restored.

Next we consider clone size distributions that *scale* with time, i.e. distributions with a general form,

$$g_n(t) = \frac{1}{N(t)} f\left(\frac{n}{N(t)}\right).$$

Such distributions have the same shape $f(x)$ at all times, however the distribution width grows over time as a result of diverging clone fate (see for example Figure 1D). This general class of distributions represents many (and perhaps all) biological processes in which clone expansion is governed by a single rate-limiting process. Substituting the scaling form of $g_n(t)$ into Eq. (S1), we find,

$$\begin{aligned} P_n(t) &= \frac{1}{t}\int_0^t f[n/N(t_1)]\frac{dt_1}{N(t_1)} \\ &= \frac{1}{t}\int_0^{N(t)/n} \frac{f(x^{-1})}{xN'(t_{nx})}dx \,, \end{aligned} \tag{S4}$$

where, to obtain the second line we have made the change of variables $x = N(t)/n$ and, as above, $t_{nx} \equiv N^{-1}(nx)$. For the particular cases of power law or exponential growth, Eq. (2) shows that $N'(\phi_n)$ is itself a power of $n$. As a consequence, $N'(t_{nx})$ is separable, i.e. $N'(t_{nx}) = N'(t_n)N'(t_x)$. Using this property, the clone size distribution is,

$$P_n(t) \;=\; \frac{1}{tN'(t_n)} C(n/N(t))\,, \tag{S5}$$

where $C(n/N(t)) = \int_0^{N(t)/n} \frac{f(x^{-1})}{xN'(t_x)} dx$ defines a 'cutoff' function that asymptotes to a constant for $n \ll N(t)$, and goes to zero for $n \gg N(t)$. Significantly, Eq. (S5) has the same form as Eq. (S2), but now the sharp cutoff of the distribution at $n = N(t)$ is replaced by the smooth function $C(n/N(t))$. Note that this result is general for *all* scaling forms of $g_n(t)$ with exponential or power-law growth of $N(t)$.

## S-II. COMMITTED PROGENITOR (CP) CELL MODEL OF STOCHASTIC FATE

Here we analyse the behaviour of the stochastic CP cell model (see Figure 3A,B), to derive Eqs. (3-5) of the main text. As in the main text, we define the division rate of progenitor cells during UVB radiation as $\lambda$; a fraction, $2r$, of divisions are symmetric (Figure 3A), and the fraction of basal layer cells that are of CP cell type is defined as $\rho$.

### A. Clonal evolution from a single UVB pulse

Restricting attention to the CP cell population, the dynamics of the *p53*-mutant cells is described by the following process (Figure 3B),

$$\mathrm{CP} \overset{2r\lambda}{\mapsto} \begin{cases} \mathrm{CP}+\mathrm{CP} & \text{Pr. } (1+\Delta)/2 \\ \oslash & \text{Pr. } (1-\Delta)/2 \end{cases} . \tag{S6}$$

Note that asymmetric divisions do not affect the size of the progenitor cell pool. For $\Delta = 0$ we recover the balanced (critical) Galton-Watson dynamics of homeostatic tissue, see Refs. [2–4]. For $0 < \Delta \leq 1$, cell division becomes biased towards symmetric self-renewal. Treating (S6) as a Markov process, the properties of the clone size distribution (as measured by the CP cell population alone) can be determined analytically. Setting $\xi = (1-\Delta)/(1+\Delta)$ and

$$\beta(t) = \frac{1-e^{-2r\lambda\Delta t}}{1-\xi e^{-2r\lambda\Delta t}} ,$$

the probability of a labelled CP cell developing into a clone of size $n_{\mathrm{CP}}$ after a time $t$ is given by [5],

$$g_{n_{\mathrm{CP}}}(t) = \begin{cases} \xi\beta(t) & n_{\mathrm{CP}} = 0 \\ [1-\xi\beta(t)][1-\beta(t)][\beta(t)]^{n_{\mathrm{CP}}-1} & n_{\mathrm{CP}} \geq 1 \end{cases} . \tag{S7}$$

From this expression, we obtain the survival probability,

$$p^{\mathrm{surv.}}(t) = 1 - p_{n_{\mathrm{CP}}=0}(t) = 1 - \xi\beta(t) . \tag{S8}$$

As a result we see that, at long times, the clone survival probability of the imbalanced system asymptotes to a non-zero constant,

$$\lim_{t\to\infty} p^{\mathrm{surv.}}(t) = \frac{2\Delta}{1+\Delta} .$$

This reflects the fact that, once clones have grown, the bias towards self-renewal makes their extinction increasingly less probable. Finally, making use of Eq. (S8), we obtain the expression for

the surviving clone size distribution,

$$g^{\text{surv.}}_{n_{\text{CP}}}(t) = \frac{g_{n_{\text{CP}}}(t)}{p^{\text{surv.}}(t)} = [1-\beta(t)][\beta(t)]^{n_{\text{CP}}-1}, \quad n_{\text{CP}} \geq 1\,.$$

From this result, we can obtain the average size of the surviving clones,

$$\begin{aligned} N_{CP}(t) \equiv \langle n_{\text{CP}} \rangle_{\text{surv.}}(t) &= \frac{1}{1-\beta(t)} \\ &= \frac{(1+\Delta)e^{2r\lambda\Delta t} - (1-\Delta)}{2\Delta}, \end{aligned} \tag{S9}$$

i.e. $\beta(t) = 1 - 1/N_{CP}(t)$. Noting that the total average clone size is $N(t) = N_{\text{CP}}(t)/\rho$, we obtain from (S9) the expression given in Eq. (3) of the main text.

Although these results are formally exact, they can be further simplified in the long-time limit. In particular, for $2r\lambda\Delta t \gg 1$, $N_{CP}(t) \gg 1$ and we can use the approximations, $[\beta(t)]^n = (1 - 1/N_{CP}(t))^n \simeq e^{-n/N_{CP}(t)}$ and $(1-\beta)/\beta \simeq 1/N_{CP}$ to obtain

$$g^{\text{surv.}}_{n_{\text{CP}}}(t) \simeq \frac{1}{N_{CP}(t)} \exp\left[-\frac{n_{\text{CP}}}{N_{CP}(t)}\right], \qquad n_{\text{CP}} \geq 1\,, \tag{S10}$$

where $N_{CP}(t)$ is defined above (S9).

### B. On-going induction

We now turn to the clonal properties of the system in the presence of on-going UVB exposure. Substituting Eq. (S7) in to Eq. (S1), the corresponding probability distribution for CP cells at a time $t$ after exposure is given by

$$\begin{aligned} p_{n_{\text{CP}}}(t) &= \frac{1}{t}\int_0^t dt_1 g_{n_{\text{CP}}}(t_1) \\ &= \begin{cases} 1 - \frac{1}{(1+\Delta)r\lambda t}\,\ln[N_{CP}(t)] & n_{\text{CP}} = 0, \\ \frac{1}{(1+\Delta)r\lambda t}\,\frac{(1-1/N_{\text{CP}})^{n_{\text{CP}}}}{n_{\text{CP}}} & n_{\text{CP}} \geq 1 \end{cases}, \end{aligned}$$

From this result, we can extract the clone survival probability, $p^{\text{surv.}}(t) = 1 - p_{n_{\text{CP}}=0}(t) = \ln[N_{CP}(t)]/[(1+\Delta)r\lambda t]$, and the size distribution of surviving (i.e. visible) clones,

$$p^{\text{surv.}}_{n_{\text{CP}}}(t) = \frac{p_{n_{\text{CP}}}(t)}{p^{\text{surv.}}(t)} = \frac{1}{\ln[N_{\text{CP}}(t)]}\,\frac{(1-1/N_{\text{CP}})^{n_{\text{CP}}}}{n_{\text{CP}}}\,. \tag{S11}$$

The average number of CP cells per surviving clone is thus,

$$\langle n_{\text{CP}}(t)\rangle = \sum_{n_{\text{CP}}=1}^{\infty} n_{\text{CP}} P^{\text{surv.}}_{n_{\text{CP}}}(t) = \frac{N_{CP}(t)-1}{\ln[N_{CP}(t)]}.$$

At this stage it is convenient to turn to the *entire* cell population, including non-progenitor cells. We make the approximation that the fraction of CP cells per clone is approximately constant, such that the actual number of cells counted is $n = n_{\rm CP}/\rho$. This approximation should be accurate in large clones (i.e. when $n \gg 1/\rho$), where fluctuations in the fraction of progenitor cells are expected to be small. Using this relation, for times $2r\lambda\Delta t \gg 1$ the size distribution of surviving clones reduces to the form given in Eq. (4) of the main text,

$$P_n(t) \simeq \frac{1}{n}\frac{\exp\left[-n/N(t)\right]}{\ln[N(t)]}, \qquad n \geq 1\,, \tag{S12}$$

where, as above, $N(t) = N_{\rm CP}(t)/\rho$. Similarly, the average clone size simplifies to,

$$\langle n(t)\rangle \simeq \langle n_{CP}(t)\rangle/\rho \simeq \frac{N(t) - 1/\rho}{\ln[\rho N]},$$

which is the same result presented in the main text (the small additive factor of $-1/\rho$ was dropped from the numerator in the main text).

Finally, to expose the exponential tail of the clone size distribution, we calculate the first incomplete moment for the CP cell model,

$$\mu_1(n,t) = \sum_{m=1}^{n} mP_m(t) \simeq \langle n(t)\rangle \left(1 - e^{-\rho n/N(t)}\right)\,. \tag{S13}$$

Eq. (S13) predicts that the quantity $[1-\mu_1(n,t)/\mu_1(\infty,t)]$ should correspond to simple exponential decay with a decay constant $\rho/N(t)$, consistent with the plots shown in Figure 3C,D, and in Supplementary Figures S1,S2.

### C. Relation to the Luria-Delbrück model

The CP cell model for PMC growth belongs to a class of models known as *Luria-Delbrück* models. These models have a long history, starting with the original study by Luria and Delbrück on the appearance of phage resistant clones in bacterial populations [6]. In this section we provide the relationship between our work and the substantial body of literature that exists characterizing these models (see Ref. [7] and references therein). However, understanding this relationship is not necessary for following the paper; we provide this section for the interest of biostatisticians familiar with the literature.

A generalised Luria-Delbrück model describes the behaviour of a population of cells, A, which multiply (A$\rightarrow$ 2A) at a rate $\alpha_1$, and die (ie become committed, A$\rightarrow$0) at a rate $\beta_1$. Mutant progenitor cells, $A^*$, multiply ($A^* \rightarrow 2A^*$) at a rate $\alpha_2$ and die ($A^* \rightarrow 0$) at a rate $\beta_2$. Mutations

($A \rightarrow A^*$) arise at a rate $\nu$. This model is characterized in Ref. [8], where the authors show that stochastic cell division and loss lead to extra-Poissonian variation in the size of the mutant cell populations.

To make contact with the literature, we re-express our model in terms of this generalised notation. In the absence of UV, we have $\nu = 0$, $\alpha_1 = \beta_1$, and $\alpha_2 = \beta_2$. That is, no mutations occur and all cells follow a critical birth-death process. In the presence of UV, we have $\nu > 0$, $\alpha_1 = \beta_1$, and $\alpha_2 > \beta_2$. The latter leads to exponential growth of the mutant population while the wildtype population remains constant. To make contact with the model parameters defined above, we have: $\alpha_1 = \beta_1 = r\lambda$, $\alpha_2 = r\lambda(1+\Delta)$, and $\beta_2 = r\lambda(1-\Delta)$. In the absence of UV radiation we have $\Delta = 0$, and $\Delta > 0$ in the presence of UV. This change of variables maps the CP cell model onto the generalised Luria-Delbrück model.

### D. Clone fate after end of UVB exposure

In this section, we consider the implications of the cessation of UVB exposure on the long-term survival of the mutated clone population. In the absence of UVB, the tissue eventually returns to homeostatic turnover in which no new *p53*-mutant clones are produced. Within the confines of the CP cell model, *p53* mutant clones (PMC) may continue to expand with some imbalance $\Delta > 0$, or else they may behave as normal CP cells (with $\Delta = 0$).

To estimate the fraction of surviving clones following cessation of UVB, we make use of Eq. (S11), which gives the distribution in the number of progenitor cells per clone upon cessation of UVB treatment, and Eq. (S8), which gives the survival probability of individual CP cells. With these equations, and defining $t_{\text{exp.}}$ as the total duration of UVB treatment, the fraction of surviving clones at some later time $t$ is given by

$$P^{\text{surv.}\cdot}(t) = 1 - \sum_{n_{\text{CP}}=1}^{\infty} p_{n_{\text{CP}}}^{\text{surv.}\cdot}(t_{\text{exp.}})[1 - p^{\text{surv.}\cdot}(t - t_{\text{exp.}})]^{n_{\text{CP}}} \,. \tag{S14}$$

This equation captures the fact that a clone of size $n_{\text{CP}}$ will survive unless all of its cells are lost, an event that may occur at time $t$ with probability $[1-p^{\text{surv.}\cdot}(t-t_{\text{exp.}})]^{n_{\text{CP}}}$, where $p^{\text{surv.}\cdot}(t)$ is defined in Eq. (S8) above. Substituting in the explicit form of $p_{n_{\text{CP}}}^{\text{surv.}\cdot}(t)$, we find,

$$\begin{aligned} P^{\text{surv.}\cdot}(t) &= 1 - \sum_{n_{\text{CP}}=1}^{\infty} \frac{1}{n_{\text{CP}} \ln[N_{\text{CP}}(t_{\text{exp.}})]} \left[ (1 - p^{\text{surv.}\cdot}(t - t_{\text{exp.}})) \left( 1 - \frac{1}{N_{\text{CP}}(t_{\text{exp.}})} \right) \right]^{n_{\text{CP}}} \\ &= \frac{1}{\ln[N_{\text{CP}}(t_{\text{exp.}})]} \ln\left[1 + p^{\text{surv.}\cdot}(t - t_{\text{exp.}})(N_{\text{CP}}(t_{\text{exp.}}) - 1)\right] \\ &= \frac{1}{\ln[\rho N(t_{\text{exp.}})]} \ln\left[1 + p^{\text{surv.}\cdot}(t - t_{\text{exp.}})(\rho N(t_{\text{exp.}}) - 1)\right] \,. \end{aligned} \tag{S15}$$

Therefore, one may see that the fraction of surviving clones depends on the average size of the oldest clones induced, $N_{\mathrm{CP}}$, as wells as the survival probability of each individual *p53* mutant progenitor cell after recovery of homeostasis. To evaluate the number of clones that will ultimately survive assuming no further mutations, we may draw upon our prior result from Eq. (S8), viz. $p^{\mathrm{surv.}}(t \to \infty) = 2\Delta/(1+\Delta)$, where $\Delta$ now describes the imbalance in homeostasis. Substituting this expression into (S15), we directly recover Eq. (5) in the main text.

---


[1] G. Arfken, Mathematical Methods for Physicists, 3rd ed., Academic Press, Orlando, FL, 1985.

[2] E. Clayton, D. P. Doupé, A. M. Klein, D. J. Winton, B. D. Simons, and P. H. Jones, A single type of progenitor cell maintains normal epidermis, Nature 446, 185 (2007).

[3] A. M. Klein, D. P. Doupé, P. H. Jones, and B. D. Simons, Kinetics of cell division in epidermal maintenance, Phys. Rev. E 76, 021910 (2007).

[4] A. M. Klein, D. P. Doupé, P. H. Jones, and B. D. Simons, Mechanism of murine epidermal maintenance: Cell division and the Voter model, Phys. Rev. E 77, 031907 (2008).

[5] N. T. J. Bailey, The Elements of Stochastic Processes, J. Wiley & Sons, New York, 1964.

[6] S.E. Luria, M. Delbück, Mutations of bacteria from virus sensitivity to virus resistance, Genetics 28, 491 (1943).

[7] Q. Zheng, Progress of a half century in the study of the Luria-Delbrück distribution, Math. Biosci. 162, 1 (1999).

[8] A. Dewanji, E.G. Luebeck and S.H. Moolgavkar, A generalized Luria-Delbrück model, Mathematical biosciences 197, 140–42 (2005).