Kinetics of cell division in epidermal maintenance

Allon M. Klein,[1] David P. Doupé,[2] Phillip H. Jones,[2] and Benjamin D. Simons[1]

[1]*Cavendish Laboratory, Madingley Road, Cambridge CB3 0HE, United Kingdom*
[2]*MRC Cancer Cell Unit, Hutchison-MRC Research Centre, Cambridge CB2 2XZ, United Kingdom*

The rules governing cell division and differentiation are central to understanding the mechanisms of development, aging, and cancer. By utilizing inducible genetic labeling, recent studies have shown that the clonal population in transgenic mouse epidermis can be tracked *in vivo*. Drawing on these results, we explain how clonal fate data may be used to infer the rules of cell division and differentiation underlying the maintenance of adult murine tail-skin. We show that the rates of cell division and differentiation may be evaluated by considering the long-time and short-time clone fate data, and that the data is consistent with cells dividing independently rather than synchronously. Motivated by these findings, we consider a mechanism for cancer onset based closely on the model for normal adult skin. By analyzing the expected changes to clonal fate in cancer emerging from a simple two-stage mutation, we propose that clonal fate data may provide a novel method for studying the earliest stages of the disease.

I. INTRODUCTION

A major challenge in biology is to determine how proliferating cells in developing and adult tissues behave *in vivo*. A powerful technique in solving this problem is clonal analysis, the labeling of a sample of cells within the tissue to enable their fate and that of their progeny to be tracked [1]. This approach gives access to information on proliferation, migration, differentiation (into other cell types), and cell death (apoptosis) of the labeled cell population. The most reliable method of labeling is through genetic modification leading to the expression of a reporter gene in a random sample of cells. Recently it has become possible to activate genetic labeling at a defined time in transgenic mice, enabling the kinetics of labeled cells to be studied with single-cell resolution *in vivo* [2]. From a theoretical perspective, the analysis of clonal fate data presents a challenging "inverse problem" in population dynamics: While it is straightforward to predict the time evolution of a population distribution according to a set of growth rules, the analysis of the inverse problem is more challenging, open to ambiguity and potential misinterpretation.

These principles are exemplified by the mechanism of murine epidermal homeostasis: Mammalian epidermis is organized into hair follicles interspersed with *interfollicular* epidermis (IFE), which consists of layers of specialized cells known as keratinocytes [3] [see Fig. 1(a)]. Proliferating cells are confined to the basal epidermal layer. As they differentiate into specialized skin cells, the basal cells withdraw from the cycle of cell proliferation and then leave the basal layer, migrating towards the epidermal surface from which they are ultimately shed. To maintain the integrity of the tissue, new cells must be generated to replace those lost through shedding. For many years, it has been thought that interfollicular epidermis is maintained by two distinct progenitor cell populations in the basal layer. These comprise long-lived stem cells (*S*) with the capacity to self-renew, and their progeny, known as transit-amplifying cells (TA), which go on to differentiate and exit the basal layer after several rounds of cell division [4]. Stem cells are also found in the hair follicles, but while they have the potential to generate epidermis in circumstances such as wounding, they do not appear to contribute to maintaining normal epidermis [5,6].

The prevailing model of interfollicular homeostasis posits that the tissue is organized into regularly sized "epidermal proliferative units" or EPUs, in which a central stem cell supports a surrounding, clonal, population of transit amplifying cells, which in turn generate a column of overlying differentiated cells [7,8]. Several experimental approaches have been used to attempt to demonstrate the existence of EPUs, but conclusive evidence for their existence is lacking. The EPU model predicts that slowly cycling stem cells should be found in a patterned array in the IFE; cell labeling studies have failed to demonstrate such a pattern [9]. In chimaeric mice the EPU model predicts that the boundaries of mosaicism in the IFE should run along the boundaries of EPUs; instead boundaries were found to be highly irregular [10]. Genetic labeling studies using viral infection or mutation to activate expression of a reporter gene in epidermal cells have demonstrated the existence of long-lived, cohesive clusters of labeled cells in the epidermis, but these clusters do not conform to the predicted size distribution of the EPU [4,9,11–14].

Thus until recently the means by which homeostasis of IFE was achieved has been unclear. However, by exploiting inducible genetic labeling, recent studies have allowed the fate of a representative sample of progenitor cells and their progeny to tracked *in vivo* [15]. As well as undermining the basis of the stem-TA cell hypothesis, the range of clone fate data provide the means to infer the true mechanism of epidermal homeostasis. In particular, these investigations indicate that the maintenance of IFE in the adult system conforms to a remarkably simple birth-death process involving a single progenitor cell compartment. Expanding upon the preliminary theoretical findings of Ref. [15], the aim of this paper is to elucidate in full the evidence for, and the properties of, the model of epidermal maintenance, and to describe the potential of the system as a method to explore early signatures of carcinogenic mutations.
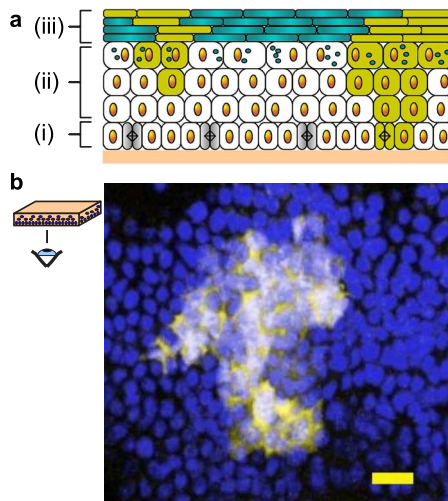
FIG. 1. (Color online) (a) Schematic cross section of murine interfollicular epidermis (IFE) showing the organization of cells within different layers and indicating the architecture of typical labeled clones. Proliferating cells (gray) are confined to the basal layer (labeled i); differentiated cells migrate through the superbasal layers (ii), where they flatten into cornified cells, losing their nuclei and assembling a cornified envelope (green) (iii), eventually becoming shed at the surface. The shaded regions (yellow) indicate two distinct clones, the progeny of single basal layer cells labeled at induction. While the clone on the right retains at least one labeled cell in the basal layer, the clone on the left hand side has detached from the basal layer indicating that all of the cells have stopped proliferating. The former are designated as "persisting clones" and contribute to the clone size distributions, while the latter, being difficult to resolve reliably, are excluded from experimental consideration. (b) Typical example of a clone acquired at a late time point, viewed from the basal layer surface. Cell nuclei are labeled blue; the hereditary clone marker (EYFP) appears yellow. Scale bar: 20 μm.

A. Background: Experimental methodology

To organize our discussion, we begin with an overview of the experimental arrangement, referring to Ref. [15] for technical details of the experimental system. To generate data on the fate of individual labeled cells and their progeny, hereafter referred to as clonal fate data, inducible genetic marking was used to label a sample of cells and their progeny in the epidermis of transgenic mice. The enhanced Yellow Fluorescent Protein (EYFP) label was then detected by confocal microscopy, which enables three-dimensional (3D) imaging of entire sheets of epidermis. Low-frequency labeling of approximately 1 in 600 basal-layer epidermal cells at a defined time was achieved by using two drugs to mediate a genetic event which resulted in expression of the EYFP gene in a cohort of mice. This low efficiency labeling ensures that clones are unlikely to merge (see discussion in Sec. II A). By analyzing samples of mice at different time points it was possible to analyze the fate of labeled clones at single cell resolution *in vivo* for times up to one year post-labeling in the epidermis [see, for example, Fig. 1(b)] [9,15].

With the gradual accumulation of EYFP levels, the early time data (less than two weeks) reveals a small increase in
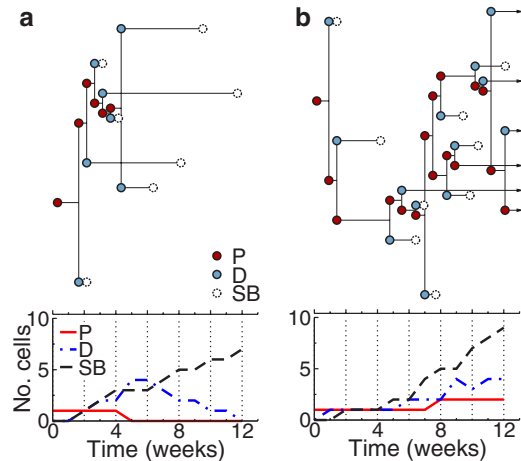

FIG. 2. (Color online) Top: Theoretical lineage for the first 12 weeks post-labeling of (a) a detached clone in which all cells have undergone a transition to terminal differentiation by week 12, and (b) a persisting clone in which some of the cells maintain a proliferative capacity, according to model (2). Circles indicate progenitor cells (P), differentiated cells (D), and suprabasal cells (SB). Note that, because the birth-death process (2) is Markovian, the lifetime of cells is drawn from a Poisson distribution with no strict minimum or maximum lifetime. The statistics of such lineage trees do not change significantly when we account for a latency period between divisions that is much shorter than the mean cell lifetime (see discussion in Sec. II C). Bottom: The total number of proliferating, differentiated, and suprabasal cells for the two clones as a function of time.

the number of labeled clones containing one or two cells. At longer times, clones increase in size while cells within clones begin to migrate through the suprabasal layers forming relatively cohesive irregular columns [see Fig. 1(a)].

The loss of nuclei in the cornified layer (Fig. 1) makes determination of the number of cornified layer cells in larger clones by microscopy unreliable. Therefore to identify a manageable population, attention was focused on the population of basal cells in "persisting clones," defined as those labeled clones which retain at least one basal layer cell, such as is exemplified in the theoretical lineage maps in Fig. 2. After two weeks, the density of persisting clones was seen to decrease monotonically indicating that the entire cell population within such clones had become differentiated and the clone detached from the basal layer [shown schematically in Figs. 1(a) and 2(a)]. However, the cell population in persisting clones showed a steady increase in size throughout the entire duration of the experiment.

To what extent are the clone fate data consistent with the orthodox stem-TA cell model of epidermal maintenance? Referring to Fig. 3, one observes an inexorable increase in the average size of an ever-diminishing persisting clone population. This result is incompatible with any model in which the IFE is supported by a population of long-lived stem cells. With the latter, one would expect the number density of persisting clones to reach a nonzero minimum (commensurate with the labeling frequency of stem cells) while the average clone size would asymptote to a constant value characteristic of a single epidermal proliferative unit. We are therefore led
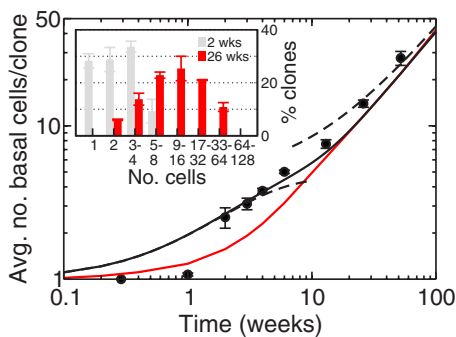
FIG. 3. (Color online) Mean number of basal layer cells in persisting clones. The experimental data (circles) show an inexorable increase in the size of persisting clones over the entire time course of the experiment. The behavior at short times (from 2–6 weeks) and at long times (beyond 13 weeks) follows the two simple analytical approximations described in the main text (lower and upper dashed curves). For times earlier than two weeks (referring to Sec. II D), clones remain approximately one cell in size. The experimental data are consistent with the behavior predicted by process (2) (black line) when it is assumed that only A-type cells are labeled at induction. In contrast, assuming that A- and B-type cells label in proportion to their steady-state population leads to an underestimate of average clone size between two and six weeks (lower curve, red online), as does the assumption that type-B cells label with better efficiency (not shown). Inset: The underlying distribution of basal cells per clone at 2 weeks and 26 weeks post-labeling. The data is binned by cell count in increasing powers of 2.

to abandon, or at least substantially revise, the orthodox stem-TA cell hypothesis and look for a different paradigm for epidermal maintenance.

But, to what extent are the clone fate data amenable to theoretical analysis? Indeed, the application of population dynamics to the problem of cell kinetics has a long history (see, e.g., Refs. [16–19]) with studies of epidermal cell proliferation addressed in several papers [20–23]. However, even in the adult system, where cell kinetics may be expected to conform to a "steady-state" behavior, it is far from clear whether the cell dynamics can be modeled as a simple stochastic process. Regulation due to environmental conditions could lead to a highly nonlinear or even nonlocal dependence of cell division rates. Indeed, *a priori*, it is far from clear whether the cell kinetics can be considered as Markovian, i.e., that cell division is both random and independent of the past history of the cell. Therefore instead of trying to formulate a complex theory of cell division, taking account of the potential underlying biochemical pathways and regulation networks [20], we will follow a different strategy looking for signatures of steady-state behavior in the experimental data and evidence for a simple underlying mechanism for cell fate. Intriguingly, such evidence is to be found in the scaling properties of the clone size distribution [15].

B. Scaling

To identify scaling characteristics, it is necessary to focus on the basal layer clone size distribution $P_n(t)$, which describes the probability that a labeled progenitor cell develops
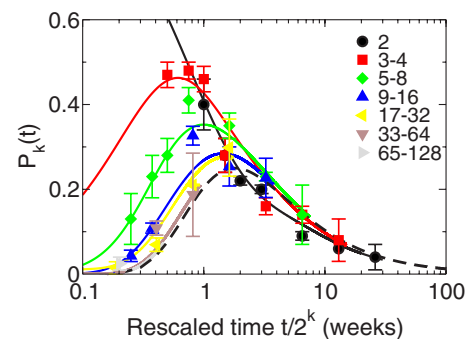

FIG. 4. (Color online) Time dependence of the grouped size distribution of *persisting* clones, $\mathcal{P}_k^{\text{pers.}}(t)$, plotted as a function of the rescaled time coordinate $t/2^k \mapsto t$. The data points show measurements [extracted from data such as shown in Fig. 3(inset), given fully in Ref. [15]], while the solid curves show the probability distributions associated with the nonequilibrium process (2) for the basal-layer clone population as obtained by a numerical solution of the master equation (3). (Error bars refer to standard error of the mean.) At long times, the data converge onto a universal curve (dashed line), which one may identify with the form given in Eq. (7). The rescaling compresses the time axis for larger clones, so that the large-clone distributions appear to converge much earlier onto the universal curve.

into a clone with a total of n basal layer cells at a time t after labeling. (Note that, in general, the total number of cells in the suprabasal layers of a clone may greatly exceed the number of basal layer cells.) With this definition, $P_0(t)$ describes the "extinction" probability of a clone, i.e., the probability that *all* of the cells within a labeled clone have migrated into the suprabasal layers. To make contact with the experimental data, it is necessary to eliminate from the statistical ensemble the extinct clone population (which is difficult to monitor experimentally) and single-cell clones (whose contribution to the total ensemble is compromised by the seemingly unknown relative labeling efficiency of proliferating and postmitotic cells at induction), leading to a reduced distribution for "persisting" clones,

$$P_{n \geqslant 2}^{\text{pers.}}(t) \equiv \frac{P_n(t)}{1 - P_0(t) - P_1(t)}.$$

Then, to consolidate the data and minimize fluctuations due to counting statistics, it is further convenient to *bin* the distribution in increasing powers of 2,

$$\mathcal{P}_k^{\text{pers.}}(t) = \sum_{n=2^{k-1}+1}^{2^k} P_{n \geqslant 2}^{\text{pers.}}(t),$$

i.e., $\mathcal{P}_1^{\text{pers.}}(t)$ describes the probability of having two cells per clone, $\mathcal{P}_2^{\text{pers.}}(t)$ describes the probability of having three to four cells per clone, and so on. Referring to Fig. 4, one may see that, after an initial transient behavior, the clone size distribution asymptotes in time to the simple scaling form,

$$\mathcal{P}_k^{\text{pers.}}(t) = f(2^k/t). \qquad (1)$$

This striking observation brings with it a number of important consequences: As well as reinforcing the inapplica-

bility of the stem cell–TA cell hypothesis, such behavior suggests that epidermal maintenance must conform to a simple model of cell division. The absence of further characteristic time scales, beyond that of an overall proliferation rate, motivates the consideration of a simple kinetics in which *only one process dictates the long-time characteristics of clonal evolution*.

Moreover, from the scaling observation one can also deduce two additional constraints: First, in the long-time limit, the average number of basal layer cells within a *persisting* clone *increases linearly with time*, viz.

$$P_{n \geqslant 2}^{\text{pers.}}(t) \simeq \frac{d}{d(2^k)} \mathcal{P}_k^{\text{pers.}}(t) = \frac{1}{t} f'(2^k/t),$$

$$\langle n \rangle_{\text{pers.}} \equiv \sum_{n \geqslant 2}^{\infty} n P_{n \geqslant 2}^{\text{pers.}}(t) \simeq \int_0^{\infty} dn \frac{n}{t} f'(n/t) \propto t.$$

Second, if we assume that labeled progenitor cells are representative of *all* progenitor cells in the epidermis, and that the population of clones with only one basal layer cell is not "extensive" [i.e., $\lim_{t \to \infty} P_1(t) = 0$], this means that, in the long-time limit, the clone persistence probability must scale as $1 - P_0(t) \propto 1/t$ such that

$$\langle n \rangle = \sum_n n P_n(t) \overset{!}{=} \rho,$$

where the constant ρ is given by the fraction of proliferating cells in the basal layer. Without this condition, one is led to conclude that the labeled population of basal layer cells either grows or diminishes, a behavior incompatible with the (observed) steady-state character of the adult system.

Although the manifestation of scaling behavior in the clone size distributions gives some confidence that the mechanism of cell fate in IFE conforms to a simple nonequilibrium process, it is nevertheless possible to conceive of complicated, multicomponent, models which could asymptote to the same long-time evolution. To further constrain the possible theories, it is helpful to draw on additional experimental observations [15]: First, immunostaining of clones with a total of two cells (using the proliferation marker Ki67 and, separately, the replication licensing factor cdc6) reveals that a single cell division may generate either one proliferating and one nonproliferating daughter through asymmetric division, or two proliferating daughters, or two nonproliferating daughters (cf. [24–26]). Second, three-dimensional imaging of the epidermis reveals that only 3% of mitotic spindles lie perpendicular to the basal layer indicating that divisions may be considered to be confined to the basal layer, confirming the results of earlier work that indicates a dividing basal cell generates two basal layer cells [25].
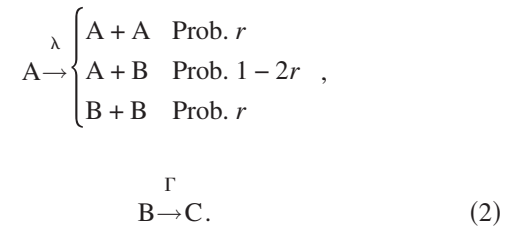
This completes our preliminary discussion of the experimental background and phenomenology. In summary, the clone fate data reveal a behavior wholly incompatible with any model based on the concept of long-lived self-renewing stem cells. The observation of long-time scaling behavior motivates the consideration of a simple model based on a stochastic nonequilibrium process and is indicative of the

labeled cells being both a representative (i.e., self-sustaining) population and in steady state. In the following, we will develop a theory of epidermal maintenance which encompasses all of these observations.

II. THEORY OF EPIDERMAL MAINTENANCE

A. Model

Taken together, the range of clonal fate data and the observation of symmetric and asymmetric division are consistent with a remarkably simple model of epidermal homeostasis involving only one proliferating cell compartment and engaging just three adjustable parameters: the overall cell division rate λ; the proportion of cell divisions that are symmetric, r; and the rate of transfer Γ, of nonproliferating cells from the basal to the suprabasal layers. To maintain the total proliferating cell population, a constraint imposed by the steady-state assumption, we have used the fact that the division rates associated with the two channels of symmetric cell division must be equal. Denoting the proliferating cells as type A, differentiated basal layer cells as type B, and suprabasal layer cells as type C, the model describes the nonequilibrium process,

$$A \xrightarrow{\lambda} \begin{cases} A + A & \text{Prob. } r \\ A + B & \text{Prob. } 1 - 2r \\ B + B & \text{Prob. } r \end{cases},$$

$$B \xrightarrow{\Gamma} C. \tag{2}$$

Finally, the experimental observation that the total basal layer cell density remains approximately constant over the time course of the experiment leads to the additional constraint that

$$\Gamma = \frac{\rho}{1 - \rho} \lambda,$$

reducing the number of adjustable parameters to just 2.

By ignoring processes involving the shedding of cells from the surface of the epidermis, the applicability of the model to the consideration of the *total* clone size distribution is limited to appropriately short time scales (up to six weeks post-labeling). However, if we focus only on the clone size distribution associated with those cells which occupy the basal layer, the model can be applied up to arbitrary times. In this case, the transfer process must be replaced by one in which $B \xrightarrow{\Gamma} \varnothing$. In either case, if we treat all instances of cell division and cell transfer as independent stochastic events, a point that we shall revisit later, then the time evolution associated with the process (2) can be cast in the form of a master equation. Defining $P_{n_A, n_B}(t)$ as the probability of finding n_A type A cells and n_B type B cells in a given clone after some time t, the probability distribution evolves according to the master equation:

$$\partial_t P_{n_A,n_B} = r\lambda[(n_A - 1)P_{n_A-1,n_B} - n_A P_{n_A,n_B}]$$
$$+ r\lambda[(n_A + 1)P_{n_A+1,n_B-2} - n_A P_{n_A,n_B}]$$
$$+ (1 - 2r)\lambda[n_A P_{n_A,n_B-1} - n_A P_{n_A,n_B}]$$
$$+ \Gamma[(n_B + 1)P_{n_A,n_B+1} - n_B P_{n_A,n_B}]. \tag{3}$$

If we suppose that the basal layer cells label in proportion to their population, the latter must be solved subject to the boundary condition $P_{n_A,n_B}(0) = \rho\delta_{n_A,1}\delta_{n_B,0} + (1-\rho)\delta_{n_A,0}\delta_{n_B,1}$. Later, in Sec. II D, we will argue that the clone size distribution is compatible with a labeling efficiency which favors A over B type cells. Either way, by excluding single cell clones from the distribution, this source of ambiguity may be safely eliminated. Although the master equation (and its total cell number generalization) is not amenable to exact analytic solution, its properties can be inferred from the consideration of the A cell population alone for which an explicit solution may be derived.

When considered alone, A type cells conform to a simple set of rate laws,

$$A \xrightarrow{2r\lambda} \begin{cases} A + A & \text{Prob. } 1/2, \\ \varnothing & \text{Prob. } 1/2, \end{cases} \tag{4}$$

an example of a Galton-Watson process, long known to statisticians (see, e.g., Ref. [27]). In this case, the probability distribution, which is related to that of the two-component model through the relation $p_{n_A}(t) = \sum_{n_B=0}^{\infty} P_{n_A,n_B}(t)$, can be solved analytically. (Here, we have used a lower case p to discriminate the probability distribution from its two-component counterpart.) For an initial distribution $p_{n_A}(0) = \delta_{n_A,1}$ it may be shown that [27]

$$p_{n_A}(t) = \left(1 + \frac{1}{r\lambda t}\right)^{-(n_A+1)} \times \begin{cases} 1 & n_A = 0, \\ \dfrac{1}{(r\lambda t)^2} & n_A > 0. \end{cases} \tag{5}$$

From this system and its associated dynamics, one can draw several key implications.

1. Epidermis is maintained through an ever-decreasing clonal population

Starting with a single labeled cell, the Galton-Watson process predicts that the persistence probability of the resulting clone (i.e., in this case, the probability that the clone retains at least one proliferating cell), is given by

$$p_{n_A>0} \equiv 1 - p_0(t) = \frac{1}{1 + r\lambda t},$$

i.e., as with the experiment, the persistence probability of a clone decays monotonically, asymptoting to the form $1 - p_0(t) \propto 1/t$ at time scales $t \gg 1/r\lambda$, the time scale for symmetric division. Applied to the experimental system, this suggests that labeled clones continue to detach from the basal layer indefinitely. At the same time, defining

$$p_{n_A>0}^{\text{pers.}}(t) = \frac{p_{n_A}(t)}{1 - p_0(t)},$$

as the size distribution of *persisting* clones, the mean number of basal layer cells in a persisting clone grows steadily as

$$\langle n_A \rangle_{\text{pers.}} \equiv \sum_{n=1}^{\infty} n_A p_{n_A>0}^{\text{pers.}}(t) = 1 + r\lambda t,$$

such that the overall cell population remains constant, viz. $\langle n_A \rangle \equiv \sum_{n=0}^{\infty} n_A p_{n_A}(t) = 1$, i.e., the continual extinction of clones is compensated by the steady growth of persisting clones such that the average number of proliferating cells remains constant: given enough time, all cells would derive from the same common ancestor, the hallmark of the Galton-Watson process [38].

This linear increase in clone size may lead one to worry about neighboring clones coalescing. Fortunately, the continual extinction of clones ensures that the fraction of clones conjoined with their neighbors remains small and of same order as the initial labeling density [39]. The fact that this fraction is constant is again indicative of the steady-state condition maintained throughout the experiment.

2. Larger clones begin to exhibit the stability of the macroscopic system

If, at some instant, a clone is seen to have, say, N_A proliferating cells then, after a further time t, its size will fluctuate as

$$\frac{\langle (n_A - \langle n_A \rangle)^2 \rangle^{1/2}}{\langle n_A \rangle} = \sqrt{\frac{2r\lambda t}{N_A}}.$$

Thus clones (as defined by the A cell population) will maintain an approximately stable number of cells providing $t \ll N_A/r\lambda$. For larger clones this time may exceed the lifetime of the system. At the limit where macroscopic sections of the basal layer are considered, the statistical fluctuations are small. The increased stability of larger clones also explains the surprising prediction that, given enough time, all clones eventually become extinct [viz. $\lim_{t\to\infty} p_{n>0}(t) = 0$]. Calculated explicitly, the extinction probability for a clone of size $N_A \gg 1$ scales as $p_0(t) \approx e^{-N_A/r\lambda t}$ [27] approaching unity at long times. However, because this extinction probability is small when $t \ll N_A/r\lambda$, a large enough clone may easily persist beyond the lifetime of the system.

3. The properties of the proliferating cell population dictate the behavior of the entire clone size distribution

At asymptotically long times, one may show [40] that the full probability distribution for finding $n = n_A + n_B$ cells within a persisting clone scales in proportion to $p_{n_A}^{\text{pers.}}(t)$, viz.

$$\lim_{t \gg 1/r\lambda} P_{n>0}^{\text{pers.}}(t) = \frac{\rho}{r\lambda t} \exp\left[-\frac{\rho n}{r\lambda t}\right], \tag{6}$$

and so

$$\lim_{t \gg 1/r\lambda} \mathcal{P}_k^{\text{pers.}}(t) \simeq \exp\left[-2^k\frac{\rho}{2r\lambda t}\right] - \exp\left[-2^k\frac{\rho}{r\lambda t}\right], \tag{7}$$

i.e., the probability distribution acquires the scaling form found empirically. Referring to Eq. (1), we can therefore deduce the form of the scaling function,
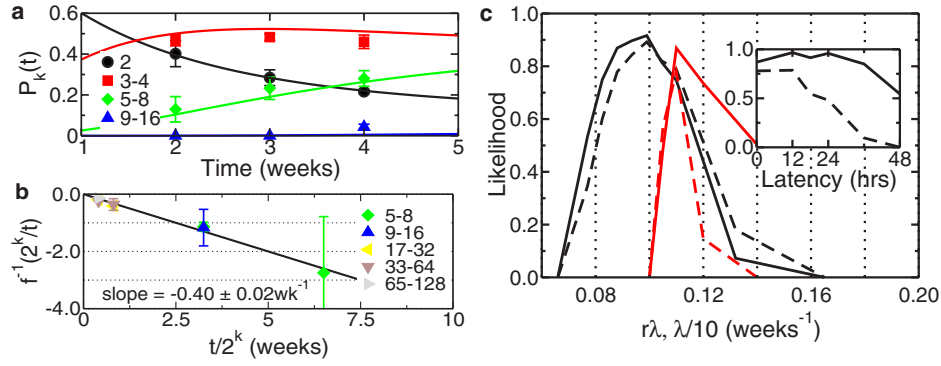
FIG. 5. (Color online) (a) Fit of Eq. (9) to the short-time clone size distributions. The data are optimally fitted by Eq. (9) using the value $\lambda = 1.1/$week and the empirical value $\rho = 0.22$ (solid lines show fit). To ensure integrity of the analysis, data for times earlier than week 2 have been excluded (see Sec. II D). (b) Linearization of the long-time asymptotic data using the "inverse" scaling function $f^{-1}(2^k/t)$ for $t \geqslant 13$ weeks and $k \geqslant 3$ (see main text). (c) Likelihood of the overall division rate λ (red) and the symmetric division rate $r\lambda$ (black), as assessed from a χ^2 test of the numerical solution to Eq. (3) [29]. A fit to the basal-layer clone size distribution alone (solid curve) is less discriminatory than a simultaneous fit to both the basal-layer and total clone size distributions (dashed). The likelihood of $r\lambda$ is shown for the optimal value of λ, and vice versa. Inset: Referring to Sec. II C, the likelihood is plotted against the duration of a latency period ($\tau_{\text{min.}}$) immediately following cell division, and assuming that division events are otherwise independent (see main text).

$$f(x) = \exp[-\rho x/2r\lambda] - \exp[-\rho x/r\lambda]. \qquad (8)$$

As a result, at long times, the average basal layer population of persisting clones becomes proportional to the average number of proliferating cells per clone, $\langle n \rangle_{\text{pers.}} = (1 + r\lambda t)/\rho$, a behavior consistent with that seen in experiment (see Fig. 3).

4. The creation and transfer of differentiated cells dictates the short-time behavior of the clone size distribution

In fitting the model to the data (see below), we will find that the rates λ and Γ at which differentiated cells are created and then transferred into the superbasal region are significantly larger than the rate of symmetric division $r\lambda$, which dictates the long-time behavior of the clone size distribution. In this case, at early times ($t \lesssim 1/\Gamma$), the clone size distributions are dominated by the differentiation and transfer rates, which remain prominent until the population of labeled differentiated cells associated with each proliferating cell reaches its steady-state value of $(1-\rho)/\rho$. One may therefore infer that, at short times, the mean number of basal layer cells in clones arising from proliferating cells is given by

$$\lim_{t \ll 1/\Gamma} \langle n \rangle_{\text{pers.}} = 1/\rho - (1/\rho - 1)e^{-\Gamma t},$$

and that the early-time clone size distribution is Poisson-distributed, viz.

$$\lim_{t \ll 1/\Gamma} P_{n \geqslant 2}^{\text{pers.}}(t) = \frac{(\langle n \rangle_{\text{pers.}} - 1)^{n-1}}{(e^{\langle n \rangle_{\text{pers.}}-1} - 1)(n-1)!}. \qquad (9)$$

B. Fit to the data

With these insights it is now possible to attempt a fit of the model to the data. Referring to Fig. 5, one may infer the rate of cell division λ from the short-time data, and the symmetric division rate $r\lambda$ from the long-time scaling data. In particular, taking the fraction of proliferating cells in the

basal layer to be $\rho = 0.22$, a figure obtained experimentally by immunostaining using Ki67 [15], a fit of Eq. (9) to the short-time data [Fig. 5(a)] is consistent with a transfer rate of $\Gamma = 0.31/$week which, in turn, implies a rate of cell division of $\lambda = 1.1/$week. Furthermore, by plotting the long-time, large-k, size distributions in terms of the "inverse" to the scaling function,

$$f^{-1}(2^k/t) \equiv [2 \ln(\{1 - [1 - f(2^k/t)]^{1/2}\}/2)]^{-1}$$
$$= [2 \ln(\{1 - [1 - \mathcal{P}_k^{\text{pers.}}(t)]^{1/2}\}/2)]^{-1},$$

the data converge onto a linear plot [Fig. 5(b)]. The resulting slope takes the value $-r\lambda/\rho$, from which we may infer the symmetric division rate $r\lambda = 0.09 \pm 0.01/$week, and $r = 0.08 \pm 0.01$.

These figures compare well with an optimal fit of the *entire* basal layer clone size distribution (Fig. 4), obtained by numerically integrating the master equation (3). The fitting procedure is shown in Fig. 5(c) (solid curves), where the likelihood of the model is evaluated for a range of values of λ and $r\lambda$, as assessed from a χ^2 test of the model solution [29]. One may see that the likelihood is maximized with an overall division rate of $\lambda = 1.1/$week and a symmetric division rate in the range $r\lambda = 0.1 \pm 0.01/$week, thus confirming the validity of the asymptotic fits. Moreover, the corresponding fit of both the basal layer distribution and the *total* clone size distribution, including both basal and suprabasal cells, is equally favorable [Fig. 5(c), dashed]. Thus in the following sections we shall use the asymptotically fitted value of $r = 0.08$, however, any choice of the parameter in the range $r = 0.08-0.10$ gives similar results.

Although the comparison of the experimental data with the model leaves little doubt in its validity, it is important to question how discerning is the fit. By itself, the observed increase in the size of persisting clones is sufficient to rule out any model based on long-lived self-renewing stem cells, the basis of the orthodox EPU model. However, could one construct a more complicated model, which would still yield
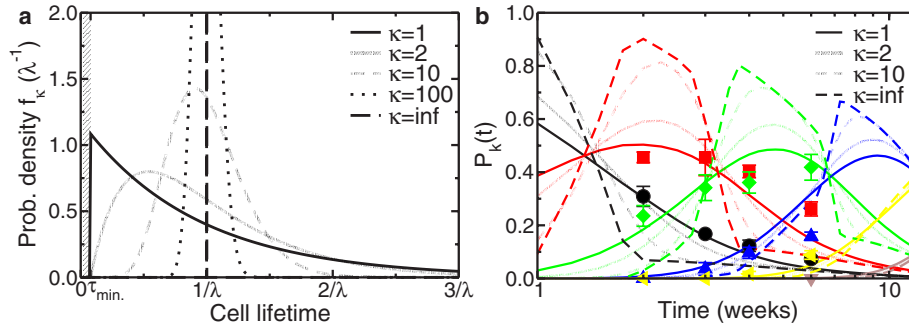
FIG. 6. (Color online). (a) Examples of progenitor cell cycle-time distributions with the same average cycle time $1/\lambda$ ($\lambda = 1.1$/week), and with a latency period of $\tau_{min.} = 12$ h introduced between consecutive cell divisions (hashed region). The case $\kappa = 1$ corresponds to a model of independent cell division such as that assumed in Sec. II A, but now accounting for an initial latency period. The case $\kappa \to \infty$ (black dashed) corresponds to all cells having an exact cell cycle time of $1/\lambda$. Note that small values of κ allow for both very short and very long cycle times. (b) Using Monte Carlo simulations of process (2), the clone size distributions predicted by each of the different cycle-time distributions in (a) are compared with the empirical data. Data points show the size distribution of persisting clones including suprabasal layer cells over the first 6 weeks post-labeling (extracted from data given fully in Ref. [15]; for legend see Fig. 4), and the theoretical curves correspond to the same legend as in (a). All of the models give an optimal fit with the same value of $\lambda = 1.1$/week, $r = 0.08$.

a similar fit? Certainly, providing the long-time evolution is controlled by a single rate-determining process, the incorporation of further short-lived proliferating cell compartments (viz. transit-amplifying cells) would not affect the observed long-time scaling behavior. However, it seems unlikely that such generalizations would provide an equally good fit to the short-time data.

More importantly, it is crucial to emphasize that the current experimental arrangement would be insensitive to the presence of a small, quiescent, long-lived stem cell population. Yet, such a population could play a crucial role in *nonsteady* state dynamics such as that associated with wound healing or development. We are therefore led to conclude that the range of clone fate data for normal adult IFE are consistent with a simple (indeed, the simplest) nonequilibrium process involving just a single progenitor cell compartment.

C. Stochastic behavior of cell division

At this stage, it is useful to reflect upon the sensitivity of the model to the stochasticity assumption applied to the process of cell division. Clearly, the scaling behavior [Eq. (6)] depends critically on the statistical independence of successive cell divisions; each cell division results in symmetric or asymmetric cell fate with relative probabilities as detailed in Eq. (4). But, to what extent would the findings above be compromised if the cell *cycle time*, i.e., the time between consecutive cell divisions, were not determined by an independent stochastic process? This question may have important ramifications, because the assumption of independent cell division, used in formulating the master equation (3), introduces a manifestly unphysical behavior by allowing cells to have arbitrarily short cycle times. Moreover, although a wide distribution of cell cycle times has been observed for human keratinocytes *in vitro* [30], it is possible that keratinocytes *in vivo* may divide in *synchrony*, giving a cell cycle-time distribution narrowly centered about the

mean $(1/\lambda)$. In the following, we shall address both of these points: First, we shall show that, up to some potential latency period (the time delay before a newly divided cell is able to divide again), consecutive cell divisions occur independently as an asynchronous Poisson process. Second, while the data are insufficient to detect a latency period of 12 h or less between consecutive cell divisions, the data do discriminate against a period lasting longer than 24 h.

To investigate the degree to which the model is sensitive to the particular cell cycle-time distribution, let us revisit the original model of independent cell division with several variations: First, we introduce a latency period of τ_{min} immediately following cell division, in which daughter cells cannot divide. This biologically motivated constraint renders a more complicated yet more realistic model of cell division than the idealized system studied in the previous section. Motivated by observations of the minimal cycle time of (human) keratinocytes [30], where a latency period of $\tau_{min} \simeq 10$ h was observed *in vitro*, we shall here consider the range of latency periods of up to 48 h. Second, we compare the empirical clone size distributions with a model where all progenitor cells have a cycle time of exactly $1/\lambda$, i.e., where cells within each clone divide in perfect synchrony. Finally, we shall investigate a range of intermediate models with different distributions of progenitor cell cycle time [see Fig. 6(a)].

Technically, the resulting clone size distributions may be evaluated through Monte Carlo simulations of the nonequilibrium process (2) with the cycle time τ of each proliferating cell selected at random from a gamma distribution of the form

$$f_\kappa(\tau) = \begin{cases} 0 & \tau < \tau_{min.} \\ \dfrac{\kappa^\kappa (\tau - \tau_{min.})^{\kappa-1}}{\bar{\tau}^\kappa \Gamma(\kappa)} e^{-\kappa(\tau - \tau_{min.})/\bar{\tau}} & \tau \geq \tau_{min.} \end{cases},$$

where $\bar{\tau} = 1/\lambda - \tau_{min.}$ is the average time to division following the initial latency period $\tau_{min.}$, and κ is the "shape parameter"

of the Gamma distribution. In particular, the choice of shape parameter $\kappa=1$ corresponds to the exponential distribution which characterizes the independent cell cycle-time distribution, whereas $\kappa\rightarrow\infty$ describes the case in which all A cells have an exact cycle time of $1/\lambda$ [see Fig. 6(a)]. Then, to reflect the assumption that initially labeled, spatially separated, progenitor cells have uncorrelated cell cycles, the time to the initial division event post-labeling is adjusted by a random time $\tau\in[0,1/\lambda]$. Finally, for an unbiased comparison of the models, we optimize the value of λ for each model separately against the empirical data, while keeping $r\lambda$ =const. to ensure an optimal fit of the long-time data, as discussed below.

The resulting clone size distributions are shown in Fig. 6(b), where the case of independent division following a 12-h latency ($\kappa=1$) and the exact cycle-time case ($\kappa\rightarrow\infty$) are compared to the empirical *total* clone size distribution, which includes both basal and suprabasal (type C) cells, over the first 6 weeks post-labeling. Two intermediate cases are also shown for comparison ($\kappa=2,10$). Focusing first on the results for the case $\kappa=1$, which bears closest resemblance to the Markovian model analyzed using the master equation (3), one may see by inspection that the quality of the fit to the data remains good even when the effects of a latency period between cell divisions is taken into account. More rigorously, a likelihood analysis reveals that the two cases are statistically indistinguishable [see Fig. 5(c), inset], which indicates that the duration of a latency period of $\tau_{\text{min.}}\lesssim 12$ h is beyond the current empirical resolution. However, referring to Fig. 5(c) (inset), a similar analysis of longer latency periods reveals that for periods of $\tau_{\text{min.}}\gtrsim 24$ h, the fit to the data is significantly poorer.

Turning next to the predicted basal-layer clone size distributions at late times ($t\gtrsim\rho/r\lambda$) (not shown), one may see that all of the proposed distributions asymptotically converge: Starting with exactly one cell, then the moment-generating function $G(q,s)=\Sigma_{n=0}^{\infty}p_n(s)q^n$ associated with the A cell population distribution $p_n(s)$ after s cell cycles satisfies the recursion relation [31]:

$$G(q,s+1)-G(q,s)=r[G(q,s)-1]^2,$$

which asymptotes to the continuous master equation $\lim_{s\gg1}\partial_s G(q,s)=r[G(q,s)-1]^2$, with the relative magnitude of the leading-order correction dropping off as $1/s$. But with $s=\lambda t$, this equation is simply the master equation for the moment-generating function associated with the original model, Eq. (5), and so the two models converge. One may therefore conclude that, beyond the first several weeks of the experiment ($t\gg1/\lambda$), the fit to the data is sensitive only to the average cycle time of progenitor cells. With this in mind, we note that for the case of perfectly synchronous cell division, an optimal (albeit poor) numerical fit was obtained when $\lambda=1.2$/week, a figure that compares well with the fit for the independent case. It appears therefore that *the predicted average cell division rate* (λ) *is insensitive to the shape of the cell cycle distribution.*

Finally, let us turn to the early time behavior ($t\sim1/\lambda$), where the predicted distributions are distinct. Referring to

Fig. 6(b), one may see, at 2–4 weeks post-labeling, that relatively large clones (five to eight cells) appear earlier than expected by a model assuming synchronous division, and that, compared with the same model, a sizeable proportion of small clones (e.g., two cells) lingers on for far longer than expected. The same behavior is observed for the basal layer clone size distribution (not shown). One may therefore infer that cell division conforms to a model of *independent* rather than *synchronous* division, allowing for some progenitor cells to divide unusually early, and for others to remain quiescent for an unusually long period of time.

In summary, we have established that, following division, progenitor cells do not divide for a period that is likely to last up to 12 h, and not more than 24 h. After this latency period, the data are consistent with cells switching to a mode of independent, asynchronous, cell division. These results shed light on why the simple model of independent cell division presented in Sec. II A succeeds in producing such a remarkable fit to the data.

D. Labeling efficiency and EYFP accumulation in basal cells

Although the integrity of the fit of the model to the data provides some confidence in its applicability to the experimental system, its viability as a model of epidermal homeostasis rests on the labeled clone population being representative of all cells in the IFE. Already, we have seen that the model, and by inference, the labeled clone population, has the capacity to self-renew. However, the slow accumulation of EYFP after induction, together with the question of the relative labeling efficiency of the two basal layer cell types, leaves open the question of the very short-time behavior. Accepting the validity of the model, we are now in a position to address this regime.

In doing so, it is particularly useful to refer to the time evolution of clone size as measured by the average number of basal cells in a persisting clone. As expected from the scaling analysis discussed in Sec. I B, a comparison of the experimental data with that predicted by the proposed cell kinetic model shows a good agreement at long times (Fig. 3). However, comparison of the data at intermediate time scales provides significant new insight. In particular, if we assume *equal* labeling efficiency of progenitor and differentiated cells, i.e., that both cell types label in proportion to their steady-state population [shown as the lower (red) curve in Fig. 3], then there is a substantial departure of the predicted curve from the experimental data for times of between two and six weeks. Intriguingly, if we assume that differentiated cells simply do not label, then the agreement of the data with theory is excellent from two weeks on! We are therefore led to conclude that, at least from two weeks, all labeled clones derive from progenitor cells labeled at induction.

With this in mind, we may now turn to the average clone size as inferred from the data at two days and one week. Here one finds that the model appears to substantially overestimate the clone size. Indeed, Fig. 3 suggests that the average clone size is pinned near unity until beyond the first week post-labeling, i.e., the relative population of single-cell clones is significantly *larger* than expected at one week, yet

falls dramatically to the theoretical value at two weeks. Referring again to the slow accumulation of EYFP, can one explain the over-representation of single-cell clones at one week post-labeling? At one week, two-cell clones are observed soon after cell division, and thus express lower concentrations of EYFP compared to single-cell clones. As a result they may be underrepresented. At later times, all labeled clones become visible as EYFP concentration grows, explaining the coincidence of experiment and theory at two weeks. It follows, of course, that the size distributions at later time points would be unaffected by slow EYFP accumulation. However, a full explanation of this effect warrants further experimental investigation, and is beyond the scope of this paper.

III. MANIFESTATION OF MUTATIONS IN CLONAL DISTRIBUTIONS

Having elucidated the mechanism of normal skin maintenance, it is interesting to address its potential as a predictive tool in clonal analysis. Conceptually, the action of mutations, drug treatments, or other environmental changes to the tissue can effect the nonequilibrium dynamics in a variety of ways: First, a revision of cell division rates or "branching ratios" (i.e., symmetric vs asymmetric) of *all* cells may drive the system towards either a new nonequilibrium steady-state or towards a nonsteady state evolution resulting in atrofication or unconstrained growth of the tissue. (The development of closed non steady-state behavior in the form of limit cycles seems unfeasible in the context of cellular structures.) Second, the stochastic revision of cell division rates or branching ratios of *individual* cells may lead to cancerous growth or extinction of a subpopulation of clones. The former may be referred to as a "global perturbation" of the cell division process while the second can be referred to as "local." In both cases, one may expect clonal analysis to provide a precise diagnostic tool in accessing cell kinetics. To target our discussion to the current experimental system, in the following we will focus on the action of a local perturbation in the form of a carcinogenic mutation, reserving discussion of a global perturbation, and its ramifications for the study of drug treatment, to a separate publication.
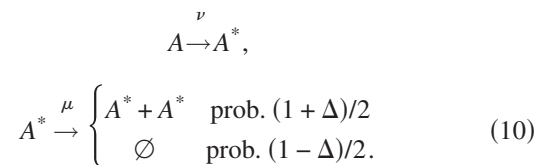
Let us then consider the action of a local perturbation involving the activation of a cancer gene in a small number of epidermal cells, which leads to the eventual formation of tumors. In the experimental system, one can envisage the treatment coinciding with label induction, for example, by simultaneously activating the EYFP and the cancer gene. In this case, clonal fate data should simply reflect a modified model of cell proliferation leading to the eventual failure of the steady-state model of tissue maintenance.

A. A simple model of carcinogenesis

To quantify the process of cancer onset, we start by establishing the simplest possible changes to process (2) which may be associated with tumor growth. Cancer is widely held to be a disease caused by genetic instability that is thought to arise when a progenitor cell undergoes a series of mutations

[32–34]. As a result, cells within the mutant clone prefer to proliferate, on average, over processes leading to terminal differentiation or death. In this investigation we shall consider a "simple" cancer resulting from *two* rate-limiting mutations: Referring to our proposed labeling experiment, the controlled induction of a cancer-causing mutation during label induction defines the first mutation; a second, rate-limiting step then occurs with the stochastic occurrence of a second cancer causing mutation. Examples of the first type of mutation may be genes that affect the ability of a cell to respond to genetic changes of the cell, e.g., *p53*, while the second mutation may be of a gene that affects clone fate such as the *Ras* oncogene [33]. We may therefore distinguish between "stage-1" mutated cells, which maintain the steady-state, and "stage-2" cells, which have the capacity for tumor formation.

The resulting process of cell proliferation is set by three parameters: The overall rate of mutation ν from a stage-1 A cell into a cancerous stage-2 cell; the division rate μ of the stage-2 cells; and the degree of imbalance Δ between their stochastic rate of proliferation and differentiation. In summary, focusing on the proliferating cell compartment only, and denoting the stage-2 mutated cells as type A^*, then the revised cell proliferation model includes the additional nonequilibrium processes

$$A \overset{\nu}{\rightarrow} A^*,$$

$$A^* \overset{\mu}{\rightarrow} \begin{cases} A^* + A^* & \text{prob. } (1+\Delta)/2 \\ \varnothing & \text{prob. } (1-\Delta)/2. \end{cases} \tag{10}$$

The rate ν may be interpreted as the mean rate with which a stage-1 cell acquires an additional mutation necessary to activate a second oncogene. The mutated cells then give rise, on average, to an exponentially growing cell lineage with growth rate $\Delta\mu$.

This nonequilibrium process was originally addressed by Kendall, who predicted the distribution in the number of tumors detected at time t after mutation [35]. His focus on tumor statistics may reflect the experimental limitations in clonal analysis at the time: Until recently it was not possible to reliably detect clones at all, let alone to count the number of cells per clone. Experimentally, however, the clone size distributions are a more efficient measure of cell kinetics than the tumor number distributions, because they result in a far richer data set, and are accessible within weeks rather than months. We shall therefore extend Kendall's approach to predict the clone size distributions at times far earlier than tumor appearance.

B. Clonal behavior during early-stage cancer

To familiarize ourselves with the modified model, consider the evolution of the average clone size with time. Focusing on the proliferating cell compartment with n type A cells and n^* type A^* cells in a clone, the relevant mean-field equations are

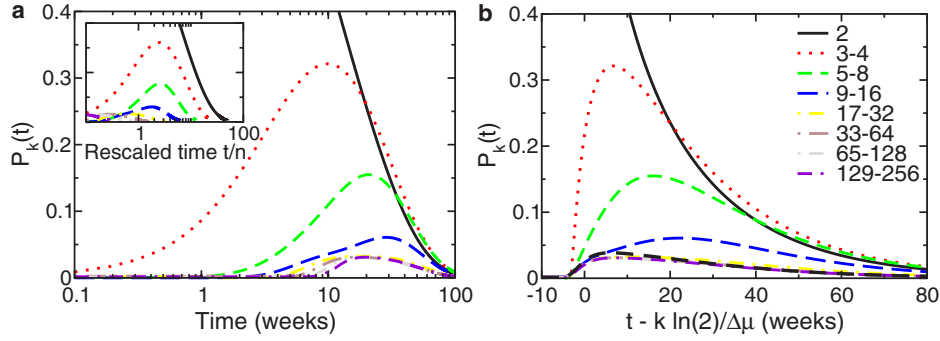$$\partial_t \langle n \rangle = -\nu \langle n \rangle,$$

FIG. 7. (Color online) (a) The total number of basal layer cells per labeled clone during the onset of cancer according to process (10). The figure was plotted by numerically integrating Eq. (11) using the empirical value $r\lambda = 0.088/$week found for normal skin, together with hypothetical values of the cancer growth parameters $\nu = 0.1r\lambda$, $\mu = 10r\lambda$, and $\Delta = 0.5$. To compare with normal skin, the predicted clone size distributions are replotted against the rescaled time coordinate $t/2^k \mapsto t$ in (a) inset. In contrast with Fig. 4, here the curves fail to converge. In (b), the same curves are shown converge onto the universal form given in Eq. (11) (dashed) when they are plotted against a new rescaled time $t \mapsto t'_k = t + k \ln 2/\Delta\mu$. Note that the large-clone distributions converge rapidly, whereas the distributions for smaller clones are affected by the non-negligible contribution of noncancerous (A) cells to the small-clone size distribution.

$$\partial_t \langle n^* \rangle = \nu \langle n \rangle + \Delta\mu \langle n^* \rangle,$$

which give the expected shift from linear growth of clones in normal skin to that of exponential growth, $\langle n + n^* \rangle = (\nu e^{\Delta\mu t} + \Delta\mu e^{-\nu t})/(\nu + \Delta\mu)$. More interestingly, referring to the master equation below, one may show that the variance in clone size also changes qualitatively: Whereas for normal skin the rms variance in clone size grows as $t^{1/2}$, here the variance in the long-time limit is *finite*,

$$\lim_{t \to \infty} \frac{\langle (n^* - \langle n^* \rangle)^2 \rangle^{1/2}}{\langle n^* \rangle} = \sqrt{1 + \Delta^{-1}}.$$

That is, the relative broadening of the clone size distribution observed in normal skin is halted by the introduction of an exponentially growing cell population.

These observations may already provide a crude method for identifying carcinogenesis through clonal analysis. To do better, it becomes necessary to solve for the full size distribution by extending the master equation (3) to include process (10). If we neglect the fate of differentiated cells, then the master equation now describes the evolution of the probability $P_{n,n^*}(t)$ for finding n type A cells and n^* type A^* cells in a clone,

$$\partial_t P_{n,n^*} = r\lambda[(n-1)P_{n-1,n^*} - nP_{n,n^*}] + r\lambda[(n+1)P_{n+1,n^*} - nP_{n,n^*}] + \nu[(n+1)P_{n+1,n^*-1} - nP_{n,n^*}]$$
$$+ \frac{1+\Delta}{2}\mu[(n^*-1)P_{n,n^*-1} - n^*P_{n,n^*}]$$
$$+ \frac{1-\Delta}{2}\mu[(n^*+1)P_{n,n^*+1} - n^*P_{n,n^*}],$$

subject to the experimental boundary condition $P_{n,n^*}(0) = \delta_{1,0}$ corresponding to exactly one "stage-1" cell per clone at $t=0$. As for the case of normal skin, we shall later be interested in the distribution of persistent clones, defined as

$$\mathcal{P}_{2^k}^{(\text{canc.})}(t) = \sum_{N=2^{k-1}+1}^{2^k} \sum_{n=0}^{N} \frac{P_{n,N-n}(t)}{1 - P_{0,0}(t) - P_{1,0}(t) - P_{0,1}(t)}.$$

While it is not possible to solve Eq. (11) analytically, progress may be made when we allow for the widely accepted view that tumors are *monoclonal*, that is they arise from a single "stage-2" mutated cell [33]. This assumption conveniently limits us to the parameter space $\nu \ll \Delta\mu$, for which an approximate long-time solution for the full clone size distribution may be found.

Referring to the Appendix for details, we find that the binned clone size distribution takes the long-time asymptotic scaling form,

$$\mathcal{P}_k^{(\text{canc.})}(t) \simeq \mathcal{N}\left[\mathcal{I}_{\beta,a}\left(\frac{1}{2\phi_k(t)}\right) - \mathcal{I}_{\beta,a}\left(\frac{1}{\phi_k(t)}\right) \right], \quad (11)$$

where $\phi_k(t) = (1 + \Delta^{-1})e^{\Delta\mu t}/2^k$,

$$\mathcal{I}_{\beta,a}(x) = \int_1^\infty d\zeta \frac{\zeta^{1-\beta}e^{-x\zeta}}{(1 + a\zeta^{-\beta})^2},$$

$\mathcal{N} = \frac{4r\lambda\chi^2}{\Delta\mu(\chi + \nu/2r\lambda)}$, $\chi^2 = \left(\frac{\nu}{2r\lambda}\right)^2 + 2\frac{\Delta\nu}{(1+\Delta)r\lambda}$, $\beta = 2\chi r\lambda/\Delta\mu$, and $a = \frac{2r\lambda\chi - \nu}{2r\lambda\chi + \nu}$. Despite its apparent complexity, this distribution is characterized by a simple scaling behavior: Referring to Fig. 7(a), the predicted clone size distributions are plotted using the scaling appropriate to the normal (unperturbed) system (cf. Fig. 4). In this case, it is apparent that the scaling $t \mapsto t/2^k$ fails. By contrast, from the expression for $\phi_k(t)$, it is clear that the size distributions should scale according to the time translation, $t \mapsto t'_k = t + k \ln 2/\Delta\mu$ as confirmed by the results shown in Fig. 7(b). Further consideration of the size distribution exposes several additional features, which may provide further access to the new model parameters:

(i) The long-time distribution decays with a rate $\beta\Delta\mu$: Expanding $\mathcal{I}_{\beta,a}(x)$ for small x gives us the asymptotic form of the universal decay curve. For $\beta < 1$, consistent with the monoclonicity requirement $\Delta\mu \gg \nu$, we find

$$\lim_{t \gg \Delta\mu} \mathcal{P}_k^{(\text{canc.})}(t) = \mathcal{N}\Gamma(-\beta)(2^{-\beta}-1)\phi_k(t)^{-\beta}, \quad (12)$$

where $\Gamma(x)$ denotes the Gamma function. This expression allows us to estimate β from the rescaled clone size distributions, providing access to the cell division and mutation parameters of the observed cells.

(ii) The probability of tumor formation is finite: This is a well-known feature of the simple noncritical birth-death process (10) [27]. Referring to the Appendix, we find that the probability p_T for any given clone to survive and form a tumor is finite,

$$p_T = 1 + \frac{\nu}{2r\lambda} - \sqrt{\left(\frac{\nu}{2r\lambda}\right)^2 + \frac{2\Delta\nu}{r\lambda(1+\Delta)}}.$$

As a result, the onset of cancer will halt the steady decrease in the density of labeled clones that is a hallmark of the unperturbed system.

These properties, and especially the change in scaling behavior, allow the onset of early-stage cancer to be identified from observations of clones less than 100 cells in size. This may provide a dramatic improvement both in speed and accuracy over current experimental models, which rely on much later observations of tumors (or hyperplasias) in order to deduce the cell kinetics at early stages.

IV. CONCLUSIONS

To summarize, we have shown that the range of clone fate data obtained from measurements of murine tail epidermis are consistent with a remarkably simple stochastic model of cell division and differentiation involving just one proliferating cell compartment. These findings overturn a long-standing paradigm of epidermal fate which places emphasis on a stem cell supported epidermal proliferative unit. As well as providing significant insight into the mechanism of epidermal homeostasis, these results suggest the utility of inducible genetic labeling as a means to resolve the mechanism of cell fate in other tissue types, and as a means to explore quantitatively the effects of drug treatment and mutation.

To conclude, we note that the analysis above has focused on the dynamics of the clonal population without regard to the spatial characteristics. Indeed, we have implicitly assumed that any model capable of describing the cell size distributions will also succeed in maintaining the near-uniform areal cell density observed in the basal layer. However, it is known that, when augmented by spatial diffusion, a simple Galton-Watson birth-death process leads to "cluster" formation in the two-dimensional system whereupon local cell densities diverge logarithmically [36,37]. Significantly, these divergences cannot be regulated through a density-dependent mobility. Understanding how the Galton-Watson process emerges from a two-dimensional reaction-diffusion type process represents a significant future challenge.

From a practical perspective, there is also the significant question of how the cell kinetic model might be generalized to describe other forms of epidermis. In particular, it is not feasible to repeat these experiments *in vivo* in human epidermis, a system of obvious medical significance. Therefore it may be of great interest to determine, in future studies, the extent to which our results compare with the behavior found in other systems.

Last, our analysis of the cancer system referred to the relatively simple case of a two-stage mutation. It is, of course, well known that tumor formation is usually the result of multiple mutations. Understanding whether clonal fate data can be used to probe the kinetics of *multistage* mutation remains an interesting future challenge.

APPENDIX: CLONE SIZE DISTRIBUTIONS IN THE TWO-STAGE CANCER MODEL

To derive the clone size distribution given in Eq. (11), we start by quoting the known result for the probability distribution $\Pi_{n^*}(t;\tau)$ of finding n^* stage-2 A^* cells at time t starting from a single A^* cell at time τ [27],

$$\Pi_{n^*}(t;\tau) = \begin{cases} \dfrac{(1-\Delta)(1-e^{-\Delta\mu(t-\tau)})}{(1+\Delta)-(1-\Delta)e^{-\Delta\mu(t-\tau)}} & \text{for } n^* = 0 \\[3mm] \left(\dfrac{2\Delta}{(1+\Delta)(1-e^{-\Delta\mu(t-\tau)})}\right)^2 e^{-\Delta\mu(t-\tau)}\left(1 - \dfrac{2\Delta e^{-\Delta\mu(t-\tau)}}{(1+\Delta)-(1-\Delta)e^{-\Delta\mu(t-\tau)}}\right)^{n^*-1} & \text{for } n^* \geq 1. \end{cases} \quad (A1)$$

When $\Delta\mu(t-\tau) \gg 1$, this distribution asymptotes to the form

$$
\Pi_{n^*}(t;\tau) \simeq
\begin{cases}
\dfrac{(1-\Delta)}{(1+\Delta)} & \text{for } n^* = 0 \\[2ex]
\left(\dfrac{2\Delta}{(1+\Delta)}\right)^2 e^{-\Delta\mu(t-\tau)} & \\[2ex]
\times \exp\left(-n^* \dfrac{2\Delta}{(1+\Delta)} e^{-\Delta\mu(t-\tau)}\right) & \text{for } n^* \geq 1.
\end{cases}
\tag{A2}
$$

From the value of Π_0 we see that even when a cell has mutated, it is not guaranteed to result in a tumor: This will only occur with a probability of $f = 1 - \Pi_0(t\to\infty) = (2\Delta)/(1+\Delta)$. The value of f plays an important role in determining the statistics of tumor formation, as will be seen below.

We now make two approximations: First, we take the long-time clone size distribution to be dominated by the statistics of A^* cells. This is a safe assumption at times $t \gtrsim 1/\nu$ and $t \gg 1/\Delta\mu$, as may be seen by considering the behavior of the mean-field equations in Sec. III B. This approximation allows us to focus on the size distribution of A^* cells only, $p_{n^*}(t)$, which is related to the full clone size distribution by the sum $p_{n^*}(t) = \sum_{n=0}^{\infty} P_{n,n^*}(t)$. Second, we assume that the entire population of type A^* cells in each clone arises from the first mutated cell that gives rise to a stable, exponentially growing lineage of cells. This corresponds to the condition $\nu \ll \Delta\mu$, as discussed in the main text.

With these two approximations, the probability of finding a labeled clone containing $n^* > 0$ mutated cells is given by

the population distribution of the *first surviving cell lineage* of A^* cells,

$$
p_{n^*}(t) \simeq \mathcal{N} \int_0^t \Pi_{n^*}(t-\tau) \sum_{m=1}^{\infty} (1-f)^{m-1} r_m(\tau) d\tau, \tag{A3}
$$

where $\mathcal{N}$ is some normalization constant, and we have introduced the probability $r_m(\tau)d\tau$ for the mth lineage of mutated cells within a given clone to be created during the interval $\tau \to \tau + d\tau$ through the mutation process $A \to A^*$. The weight factor $(1-f)^{m-1}$ gives the probability that the first $m-1$ cell lineages of A^* cells within a clone will become extinct—a situation necessary to make the mth cell line relevant to the distribution according to the monoclonal approximation.

The rates $r_m(\tau)$ may be accessed by considering the probability $w_{n,m}(t)$ that a clone containing n type A cells at time t also contains m *independent lineages* of mutated A^* cells, each arising from a separate mutation event. (Later we shall treat the evolution of each of these cell lines post-creation.) To solve for $w_{n,m}(t)$ we must introduce its moment-generating function $G(q,Q^*;t) \equiv \sum_{n=0}^{\infty}\sum_{m=0}^{\infty} w_{n,m}(t) q^n (Q^*)^m$, which evolves [from Eq. (11)] according to the dynamical equation

$$
\dot{G} = [r\lambda(q-1)^2 + \nu(Q^* - q)]\partial_q G. \tag{A4}
$$

Solving this equation subject to the initial condition of one stage-1 (A) cell per clone, we find the solution

$$
G(q,Q^*;t) = \xi_{Q^*} - \frac{2\xi_{Q^*}}{1 + \{\xi_{Q^*} + [q-1-(\nu/2r\lambda)]\}/\{\xi_{Q^*} - [q-1-(\nu/2r\lambda)]\}e^{-2\xi_{Q^*}r\lambda t}} + 1 + \frac{\nu}{2r\lambda}, \tag{A5}
$$

with $\xi_{Q^*} \equiv \sqrt{\left(\frac{\nu}{2r\lambda}\right)^2 + \frac{\nu}{r\lambda}(1-Q^*)}$. Equation (A5) describes the evolution of a single A cell as it proliferates and eventually gives rise to a set of internal lines of mutated cells.

Before we proceed to find $p_{n^*}(t)$, note that setting $q=1$, $Q^* = 1 - f$ in Eq. (A5) gives us the result (quoted in the main text) for the asymptotic fraction p_T of clones in which all mutated cell lines become extinct,

$$
p_T = 1 + \frac{\nu}{2r\lambda} - \sqrt{\left(\frac{\nu}{2r\lambda}\right)^2 + \frac{\nu f}{r\lambda}}.
$$

On the other hand, setting $q=1$ only in Eq. (A5) gives the moment-generating function for (yet another) distribution $W_m(t) = \sum_{n=0}^{\infty} w_{n,m}(t)$ of a clone containing m independent

lines of A^* cells irrespective of the number of normal cells in the clone. Finally, noting that $\dot{W}_m(t) = r_m - r_{m+1}$ then gives

$$
r_m(t) = -\sum_{n=0}^{m-1} \dot{W}_m(t),
$$

which we may substitute into Eq. (A3) to find (for $n^* > 0$)

$$
p_{n^*}(t) \simeq -\mathcal{N} \int_0^t d\tau \Pi_{n^*}(t-\tau) \dot{G}(1, 1-f; \tau). \tag{A6}
$$

From this expression, simplified by the large-n^* approximation ($\sum_{n^*} \simeq \int dn$), we obtain the final form of the binned size distribution given in Eq. (11).

[1] J. D. Clarke and C. Tickle, Nat. Cell Biol. **1**, E103 (1999).

[2] J. Jonkers and A. Berns, Nat. Rev. Cancer **2**, 251 (2002).

[3] E. Fuchs, Nature (London) **445**, 834 (2007).

[4] C. S. Potten, Int. Rev. Cytol. **69**, 271 (1981).

[5] M. Ito, Y. Liu, Z. Yang, J. Nguyen, F. Liang, R. J. Morris, and G. Cotsarelis, Nat. Med. (N.Y.) **11**, 1351 (2005).

[6] V. Levy, C. Lindon, B. D. Harfe, and B. A. Morgan, Dev. Cell **9**, 855 (2005).

[7] I. C. Mackenzie, Nature (London) **226**, 653 (1970).

[8] C. S. Potten, Cell Tissue Kinet. **7**, 77 (1974).

[9] K. M. Braun *et al.*, Development (Cambridge, U.K.) **130**, 5241 (2003).

[10] G. H. Schmidt, M. A. Blount, and B. A. Ponder, Development (Cambridge, U.K.) **100**, 535 (1987).

[11] S. Ghazizadeh and L. B. Taichman, EMBO J. **20**, 1215 (2001).

[12] T. Kameda, A. Nakata, T. Mizutani, K. Terada, H. Iba, and T. Sugiyama, Exp. Cell Res. **283**, 167 (2003).

[13] S. Ro and B. Rannala, EMBO Rep. **5**, 914 (2004).

[14] S. Ro and B. Rannala, Exp. Dermatol. **14**, 838 (2005).

[15] E. Clayton, D. P. Doupé, A. M. Klein, D. J. Winton, B. D. Simons, and P. H. Jones, Nature (London) **446**, 185 (2007).

[16] M. Loeffler and H. E. Wichmann, Cell Tissue Kinet. **13**, 543 (1980).

[17] M. Loeffler and B. Grossmann, J. Theor. Biol. **150**, 175 (1991).

[18] R. A. White and N. H. A. Terry, Math. Comput. Modell. **32**, 113 (2000).

[19] I. Roeder, Curr. Opin. Hematol. **13**, 222 (2006).

[20] N. J. Savill, Cell Prolif **36**, 1 (2003).

[21] D. R. Appleton, N. A. Wright, and P. Dyson, J. Theor. Biol. **65**, 769 (1977).

[22] C. S. Potten, H. E. Wichmann, and M. Loeffler, Cell Tissue Kinet. **15**, 305 (1982).

[23] G. D. Weinstein, J. L. Mccullough, and P. Ross, J. Invest. Dermatol. **82**, 623 (1984).

[24] T. Lechler and E. Fuchs, Nature (London) **437**, 275 (2005).

[25] I. H. Smart, Br. J. Dermatol. **82**, 276 (1970).

[26] W. Zhong, J. N. Feder, M. M. Jiang, L. Y. Jan, and Y. N. Jan, Neuron **17**, 43 (1996).

[27] N. T. J. Bailey, *The Elements of Stochastic Processes* (J. Wiley & Sons, New York, 1964), Chap. 8.6.

[28] H. W. Watson and F. Galton, Journal of the Anthropological Institute of Great Britain and Ireland **4**, 138 (1875).

[29] W. H. Press, S. A. Teukolsky, W. Vetterling, and B. P. Flannery, *Numerical Recipes in C: The Art of Scientific Computing*, 2nd ed. (Cambridge University Press, Cambridge, UK, 1993), Chap. 15.

[30] R. Dover and C. S. Potten, J. Cell. Sci. **89**, 359 (1988).

[31] S. Karlin and H. M. Taylor, *A First Course in Stochastic Processes*, 2nd edition (Academic Press, Inc., San Diego, 1975), Chap. 8.

[32] E. R. Fearon, S. R. Hamilton, and B. Vogelstein, Science **238**, 193 (1987).

[33] E. R. Fearon and B. Vogelstein, Cell **61**, 759 (1990).

[34] A. Whittemore and J. B. Keller, SIAM Rev. **20**, 1 (1978).

[35] D. G. Kendall, Biometrika **47**, 13 (1960).

[36] B. Houchmandzadeh, Phys. Rev. E **66**, 052902 (2002).

[37] W. R. Young, A. J. Roberts, and G. Stuhne, Nature (London) **412**, 328 (2001).

[38] Curiously, Galton and Watson first dealt with this question amidst the concern that aristocratic surnames were becoming "extinct" in Victorian Britain [28]!

[39] The fraction of clones in contact with their neighbors is estimated by assuming that all clones are randomly and independently distributed, and by noting the empirical observation that clones are oblique in shape, so that the distribution in clone area has the same form as $P_{n>0}^{\text{pers.}}(t)$.

[40] To obtain Eq. (6), we treat n_A, n_B as continuous variables in Eq. (3) (a good approximation at large values of n). Then, making the ansatz that the B-cell population remains slave to the A-cell population viz. $n_A = \rho n$, $n_B = (1-\rho)n$, the master equation simplifies to the approximate form

$$\partial_t P = \frac{r\lambda}{\rho}(n\partial_n^2 P + 2\partial_n P),$$

which is solved by $P_n(t) = (\rho/r\lambda t)^2 \exp(-\rho n/r\lambda t)$, leading to Eq. (6).