Mechanism of murine epidermal maintenance: Cell division and the voter model

Allon M. Klein,[1] David P. Doupé,[2] Philip H. Jones,[2] and Benjamin D. Simons[1]
[1]*Cavendish Laboratory, Madingley Road, Cambridge CB3 OHE, United Kingdom*
[2]*MRC Cancer Cell Unit, Hutchison-MRC Research Centre, Cambridge CB2 2XZ, United Kingdom*
(Received 3 October 2007; revised manuscript received 30 November 2007; published 10 March 2008)

The dynamics of a genetically labeled cell population may be used to infer the laws of cell division in mammalian tissue. Recently, we showed that in mouse tail skin, where proliferating cells are confined to a two-dimensional layer, cells proliferate and differentiate according to a simple stochastic model of cell division involving just one type of proliferating cell that may divide both symmetrically and asymmetrically. Curiously, these simple rules provide excellent predictions of the cell population dynamics without having to address the cells' spatial distribution. Yet, if the spatial behavior of cells is addressed by allowing cells to diffuse at random, one deduces that density fluctuations destroy tissue confluence, implying some hidden degree of spatial regulation of cell division. To infer the mechanism of spatial regulation, we consider a two-dimensional model of cell fate that preserves the overall population dynamics. By identifying the resulting behavior with a three-species variation of the voter model, we predict that proliferating cells in the basal layer should *cluster*. Analysis of empirical correlations of cells stained for proliferation activity confirms that the expected clustering behavior is indeed seen in nature. As well as explaining how cells maintain a uniform two-dimensional density, these findings present an interesting experimental example of voter-model statistics in biology.

DOI: 10.1103/PhysRevE.77.031907 PACS number(s): 87.17.Ee, 87.23.Cc, 82.45.Jn, 05.70.Fh

I. INTRODUCTION

A major challenge in biology is to determine how proliferating cells behave in developing and adult tissues. To gain insight into the processes of cancer onset, aging, and wound healing, biologists have long recognized that the spatial organization of cells in tissue provides indirect access to the underlying cell behavior. Some tissues, such as the auditory hair cells of the inner ear, are arranged into repeating units containing groups of specialized cells essential for the function of the tissue [1]. In contrast, in other tissues cells do not organize into coherent structures that reflect their cooperative function. Indeed, the arrangement of some cell types appears random [2]. Inferring the rules of cell behavior in these apparently unstructured tissues appears challenging. One may ask, therefore, how cell behavior in such tissues is regulated in the absence of well-defined spatial roles.

In this context, it is interesting to consider the case of mouse tail skin epidermis, for which the laws governing cell behavior have recently been resolved. Mammalian epidermis comprises hair follicles interspersed with interfollicular epidermis (IFE), which consists of sheets of specialized cells known as keratinocytes [3]; see Fig. 1(a). Cells are shed continually from the epidermal surface, and are replaced by proliferating cells in the basal layer, whose progeny may cease proliferating and then migrate through the suprabasal layers before reaching the epidermal surface. Until recently, it was widely assumed that adult tissue is maintained by two different proliferating cell populations. These comprised long-lived, self-renewing stem cells, which have the potential to undergo an unlimited number of cell divisions and which maintain a second population of transit-amplifying (TA) cells, whose proliferative potential is limited [4]. After several rounds of division, it was conjectured that TA cells differentiate, exit the cell cycle, and move out of the basal layer.

In a recent study by Clayton *et al.*, inducible genetic labeling was used to study the mechanism of epidermal maintenance by tracking the fate of a representative sample of cells and their progeny (clones) in normal murine tail epidermis [5]. By analyzing the size distribution of such clones over a period of one year, it was possible to infer that the epidermis is maintained by just one type of progenitor cell. According to the revised model of epidermal maintenance, progenitor cells capable of both symmetric and asymmetric division give rise to a population of nonproliferating cells, which then transfer from the basal layer to the suprabasal layers. More precisely, labeling proliferating cells as type A and differentiated basal layer cells as type B, the basal layer cell population is governed by the stochastic nonequilibrium (Markovian) process

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

involving three adjustable parameters: the overall cell division rate λ, the proportion of cell divisions that are symmetric, $2r$, and the rate of transfer, Γ, of nonproliferating cells from the basal to the suprabasal layers. As well as overturning the accepted paradigm of epidermal homeostasis being achieved by discrete populations of stem and TA cells, the model provides a degree of quantitative predictive rigor that is unusual in the field of cell tissue biology; for example, by quantifying the division and migration rates ($\lambda = 1.1/\text{week}$, $\Gamma = 0.31/\text{week}$) and the branching ratio ($r = 0.08$), as shown in Refs. [5,6]. As a result of its quantitative nature, the model establishes a platform for investigating the role of different cellular constituents (such as gene products) in regulating

cell fate, for example by studying how different genes influence the model parameters.

It is an interesting fact that process (1) is capable of fitting a wide range of clone fate data within a "zero-dimensional" framework, i.e., without having to address the spatial orientation of cells within the basal layer. Yet the observed uniformity of cell density implies a degree of regulation beyond that which can be addressed in the zero-dimensional framework. In particular, when augmented by spatial diffusion, the proposed model leads to "cluster" formation in the two-dimensional system, whereupon local cell densities are predicted to diverge logarithmically [7,8]. In biological terms, such behavior would correspond to a severe disruption of the epidermis, with much of the tissue dying away, leaving only a few isolated and very thick clusters of epidermis. Significantly, the observation that labeled families of cells remain largely *cohesive* [see, for example, Fig. 1(b)], reveals that cell mobility must be small, so that such divergences would be significant within a mammalian lifetime. These divergences cannot be regulated through a local density-dependent mobility.

Thus, the success of the zero-dimensional fit, despite the predicted divergence in two dimensions, leaves us with an interesting challenge that is the focus of the present study: Can we uncover, from the spatial distribution of basal layer cells, the mechanism by which cells regulate a uniform cell density without compromising the integrity of the zero-dimensional fit, as embodied in process (1)?

In order to identify the underlying rules of cell division and differentiation, we shall draw upon the results of two types of experiment. First, we shall revisit the clone fate data used by Clayton *et al.*, and examine the previously discarded spatial distribution of labeled basal layer cells for signatures of underlying regulation. Second, we shall consider the statistics of the entire population of basal layer cells. In particular, by immunostaining basal layer cells for markers of cell proliferation, it is possible to analyze the spatial distribution of all progenitor cells.

Thus, the aim of this paper is to elucidate how the experimental observations constrain any proposed theory of spatial behavior in the basal layer. In summary, we shall show that the dynamics predicted by process (1) are indeed consistent with the constraint on uniform cell density, provided that cell division occurs only upon the migration of a nearby type *B* (i.e., differentiated) cell into the basal layer. Moreover, we confirm that the clone fate data are consistent with a restricted degree of cell mobility, whereby cell motion is *not* diffusive and random. Instead, differentiated cells only migrate laterally as a response to fluctuations in the local density. Finally, to test the validity of the proposed spatial model, we use it to predict that, while maintaining a uniform *total* areal cell density, the population of progenitor cells should cluster over time. By considering the radial correlation function for the spatial distribution of progenitor cells, we find that this prediction is in good qualitative agreement with experiment. Quantitatively, the comparison reveals that the experimental degree of progenitor cell clustering is slightly higher than that expected for the parameter value of $r=0.08$ determined previously through clonal analysis. Although several technical difficulties may challenge the reliability of these quantitative results, we speculate that such excess clustering may be a signature of spatial regulation of cell fate during asymmetric division. These results also shed light on previous observations of clustering of cells undergoing mitosis in the epidermis [9,10]. Such observations have been interpreted in the biological community to be a signature of regulation that leads to coordinated cell division. By contrast, this work shows that the tendency of proliferating cells (and therefore mitoses) to cluster is in fact consistent with cells dividing independently and stochastically—indeed, it is the hallmark of the proposed spatial process.

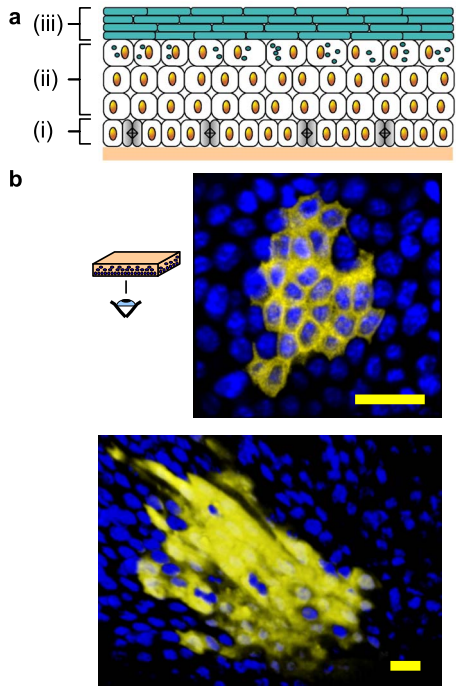

FIG. 1. (Color online) (a) Schematic cross section of murine interfollicular epidermis. Proliferating cells (gray) are confined to the basal layer [labeled (i)]; differentiated cells migrate through the superbasal layers (ii), where they flatten into cornified cells, losing their nuclei and assembling a cornified envelope (green online) (iii), eventually becoming shed at the surface. (b) Two examples of typical clones 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.

This paper is organized as follows. In Sec. II A we develop a phenomenological model of cell behavior that incorporates the experimental constraint on uniform cell density. We identify the proposed model as a variation upon the so-called *monomer-monomer* model of surface catalysis. We then analyze this simpler model in Sec. II B, including an exact solution for the two-point density correlations in the closely related monomer-monomer model. In Sec. III we test the uniform-density model against a range of experimental data, first through a qualitative comparison of the model with the empirical clone shape data (Sec. III A), and then by analyzing the spatial correlations observed for proliferating cells (Sec. III B). We conclude with a discussion of cell clustering in Sec. IV.

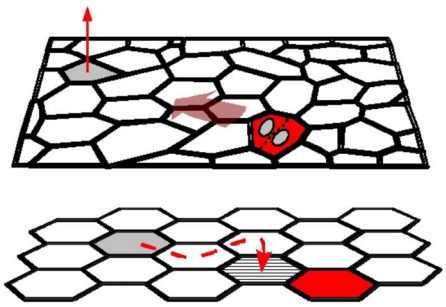

FIG. 2. (Color online) Schematic motivating the proposed lattice model defined in Eq. (2). Top: A cartoon of cells within the basal layer, showing the exit of a cell (light gray) through upward migration into the suprabasal layers, concurrent with the division of a nearby progenitor cell [dark gray (red)]. To ensure continuity of the basal layer, it is postulated that cells rearrange to maintain a uniform cell density (wide arrow). Bottom: In a simple lattice model that captures the essence of the steady-state dynamics, the exit of a cell from the basal layer gives rise to a vacant lattice site (light gray), which rapidly diffuses by exchanging position with adjacent cells (white). Upon coming into contact with a proliferating cell, the latter may divide and replace the vacancy with a daughter cell (striped).

II. A SPATIAL MODEL OF CELL KINETICS

To understand how the experimental observations constrain any proposed theory of spatial behavior in the basal layer, we shall first address the constraint imposed by the observed uniform cell density. To this end, we postulate that proliferating cells divide only upon the migration of a nearby differentiated cell into the suprabasal layers (see Fig. 2). This requirement is purely phenomenological, as it ensures a uniform density without specifying the mechanism by which it is implemented. Indeed, a range of regulatory pathways can be seen to give rise to the same phenomenology, for example by coupling cell division processes to the local stress [11,12] or using short-ranged morphogen gradients as well as feedback by cell-cell communication. Once this initial constraint is accounted for in this section, we shall draw upon the further observations of clone and basal layer morphology in order to identify additional rules governing cell behavior.

In general, it is a challenge to couple the division and migration of basal layer cells while still allowing for some degree of cell compressibility, whereby a dividing cell may compensate for the exit of a nonadjacent cell through lateral motion (see Fig. 2, top). Several approaches have been used in the past to overcome this problem. In the context of tissue development, one may treat the cell tissue as an elastic medium, whereby the local cell density is coupled to the cell division process through the stress in the surrounding tissue [11]. Such an approach is capable of accounting for a range of realistic properties of two-dimensional cell tissue growth, such as the distribution of cell sizes, as well as cell compressibility. Yet, for the simple problem of steady-state tissue maintenance, involving no net growth, it is unlikely that the complexity of the elastic tissue model is required to explain much of the experimental data. A second approach treats the two-dimensional basal layer as a "foam" of cells (a Voronoi tesselation), for which the steady-state condition may be used to relate the local cell topology (or more precisely, the number of nearest neighbors in the basal layer) with the likelihood of division or migration. Applied to the problem of epidermal maintenance, this approach successfully predicts the steady-state topology of epidermal basal layer cells [13], and it identifies that cells with a larger number of neighbors are more likely to undergo division, whereas cells with a small number of neighbors are more likely to migrate into the suprabasal layers. However, it is a challenge to extend this model to allow for two distinct cell populations that are exclusively committed to either division or migration, as described by process (1). Yet a third approach draws upon simulations in which cells are modeled as quasispherical particles that deform during cell division [14]. However, as this approach draws upon a wide range of (uncontrolled) parameters to describe the cell-cell interactions, it is more complex than required for this case.

Therefore, in the following it will be sufficient to use a simpler description of the basal layer, by drawing upon nonequilibrium lattice models discussed in the recent literature. In particular, we shall model the basal layer as a lattice in which each site is occupied by one cell. Cell compressibility is then modeled by a population of lattice vacancies, which are created upon cell migration ($B \to \oslash$) and then diffuse rapidly as compared to the cell kinetic rates (λ,Γ), before annihilating upon the division of an adjacent cell (see Fig. 2). Since this rule-based model is not primarily based on a direct physical representation of individual cells, it may overlook certain physical effects. For example, the migration of type B cells out of the basal layer may be facilitated by mechanical forces exerted by neighboring cells—a situation that is hard to represent with a cellular automaton. However, recalling that the stochastic rules embodied by process (2) have been experimentally verified, it is reasonable to start by considering a similar stochastic process in two dimensions. Later, we shall further justify the use of the lattice model by showing that the lattice geometry does not affect the qualitative behavior of the system [see Sec. II B, paragraph (e)].

Although a significant advantage of the lattice description of the basal layer is the ease by which it may be simulated, a range of analytic results are also made accessible by showing that, in certain limits, the model reduces to a simple two-component model that belongs to the generalized *voter model* universality class [15–17]. The voter model universality class describes lattice processes that undergo phase separation in two dimensions in the absence of surface tension. In the context of basal layer kinetics, this "phase separation" corresponds to the clustering of proliferating cells. Curiously, in the special case $r=1/4$, the basal layer lattice model reduces to the reaction-limited monomer-monomer model of surface catalytic reactions, in which the classical voter model dynamics are augmented by infinite-temperature Kawasaki exchange dynamics [18]. As well as drawing upon a range of existing results afforded by these models, we shall derive here an exact solution for the two-point spatial correlation function of the monomer-monomer model. These results reveal the continuous transition between voter-like and diffusive behavior as the relative rates of symmetric vs asymmet-

ric cell division (or reactant deposition in surface catalysis) are adjusted.

A. The lattice model

As mentioned above, we are interested in constructing a spatial model that recovers the behavior of process (1), and maintains a uniform cell density. To account for the steric repulsion of basal layer cells, we will characterize the basal layer as a two-dimensional lattice, where each site is host to one of the two cell types, or it remains vacant. When the vacancy fraction is very low, then such a lattice description presents a reasonable approximation of the observed near-uniform arrangement of basal layer cells during normal adult skin maintenance. Then, to regulate the cell density, progenitor cells (A) are allowed to divide only when neighboring a site vacancy, while the migration of postmitotic (B) cells from the basal layer leads to the creation of vacancies which are free to diffuse in the basal layer through the displacement of neighboring cells (see Fig. 2). In summary, denoting a site vacancy with the symbol $\oslash$, the lattice model may be written in terms of the nonequilibrium process

$$A\oslash \xrightarrow{\lambda'} \begin{cases} AA, & \text{Prob } r, \\ AB, & \text{Prob } 1/2 - r, \\ BA, & \text{Prob } 1/2 - r, \\ BB, & \text{Prob } r, \end{cases}$$

$$B \xrightarrow{\Gamma} \oslash, \tag{2}$$

$$X\oslash \xrightarrow{\sigma} \oslash X,$$

where the site hopping rate σ reflects the capacity of vacancies to diffuse within the basal layer, and X is used to denote either a type A or a type B cell. To gain some initial insight into the dynamics, and to identify constraints on the parameter space, one may consider the steady-state mean-field cell densities associated with this process. It is straightforward to show that the mean-field equation for the the vacancy fraction $n_{\oslash}$ is given by

$$\partial_t n_{\oslash} = \sigma S \nabla^2 n_{\oslash} + \Gamma(1 - \rho - n_{\oslash}) - \lambda' \rho n_{\oslash}, \tag{3}$$

where ρ is the (constant) A-cell fraction, and S is the area per lattice site, corresponding to the average areal cross section of a basal layer cell. In the following we shall work in units of the average cell cross section ($S=1$). From here one may see that the near-uniform cell density in the basal layer ($n_{\oslash} \ll 1$ and uniformly distributed) constrains us to the region of parameter space $\lambda' \gg \Gamma$, such that any vacancy created through the migration of a type B cell out of the basal layer is rapidly removed upon coming into contact with a proliferating cell. In this limit, the numerical value of the parameter λ' becomes irrelevant, as may be seen by calculating the effective local division rate $\lambda \equiv \lambda' n_{\oslash}$, which takes the value $\lambda = \Gamma(1-\rho)/\rho$, independent of λ'. This relationship between the rate of cell division and migration is identical to that obtained in process (1) [6].

Although the uniform vacancy density is a stable fixed point of the mean-field dynamics [Eq. (3)], one may worry whether fluctuations about the mean-field solution are capable of compromising the uniform density of the basal layer cell lattice in actual manifestations of process (2). To eliminate this concern, one requires rapid dissipation of density fluctuations independently of the cell kinetics, from which we infer that the lattice vacancy population must diffuse rapidly compared to the time scale of cell division and upward migration, viz., $\sigma \gg \Gamma$, as stated earlier. Biologically, this condition corresponds to the assumption that cells are largely incompressible, so that local density fluctuations lead to the rearrangement of cells on a time scale that is significantly faster than that of cell division (cf. Ref. [11]). In Sec. III A we shall show that the empirical clone fate data further constrain the parameter space to the region $\lambda' \gg \sigma$, whereby the exit of a type B cell is compensated for by a nearby cell division.

With these definitions, in the parameter space $\lambda' \gg \sigma \gg \Gamma$ one may see that the spatial model introduces no new relevant parameters compared to process (1). That is, the model behavior depends only on the (known) zero-dimensional parameters (r,λ,Γ), with the contribution of the new parameters (λ',σ) entering through the dimensionless ratios Γ/λ', Γ/σ, and σ/λ', all of which may be made arbitrarily small. But, as a precondition on its validity, does the model also reproduce the observed zero-dimensional basal layer clone size distributions? This is by no means obvious, given the critical (and therefore delicate) nature of process (1) [6]. For example, a high density of progenitor cells in the lattice model may lead to vacancy depletion and jamming, an effect that has no analog in the zero-dimensional system. Therefore, in Sec. III we shall show, using Monte Carlo simulations, that the proposed lattice model indeed succeeds in reproducing the empirical clonal statistics from Ref. [5]. With this basic confidence in the validity of the model, we now proceed to analyze its behavior in more detail.

In general, the cell kinetics in process (2) describe a hard-core nonequilibrium system involving three cell species. Recent progress in the field of nonequilibrium statistical mechanics has resulted in several possible formalisms with which to study such systems [19–21]. However, for the case at hand, these approaches are unnecessarily complex. Rather than analyzing the current microscopic model, it is convenient to recast the cell kinetics into a simpler form that describes the same phenomenology. In particular, for the parameter space of interest ($\lambda' \gg \sigma \gg \Gamma$), it is sufficient to consider a lattice fully occupied by A and B cells, without addressing the population of lattice vacancies. To see this, one may see from Eq. (3) that the vacancy dynamics occur on a fast time scale compared to that of cell migration $(1/\Gamma)$ and division $(1/\lambda)$. Therefore, referring back to the lattice process (2), we may (heuristically) eliminate the vacancy population by replacing the cell division process ($A\oslash \rightarrow XY$) with a direct cell-cell "reaction" process ($AB \rightarrow XY$), and by replacing the A-cell division rate λ' with the effective rate $\lambda(\mathbf{x}) = \lambda' n_{\oslash}(\mathbf{x}) = \Gamma[1-\rho(\mathbf{x})]/\rho(\mathbf{x})$. Here, the local A-cell fraction $\rho(\mathbf{x})$ refers to the A-cell number density coarse grained over the nearby lattice neighborhood. For example,

denoting the number of type A cells on a lattice site as $n_{\mathbf{x}}$, we have $\rho(\mathbf{x})=\Sigma_{\mathbf{x}'}n_{\mathbf{x}'}w(|\mathbf{x}-\mathbf{x}'|)$ where $w(x)$ is some suitably chosen normalized envelope function. Within this framework, one may then replace process (2) with the more simplified form

$$AB \xrightarrow{\lambda(x)} \begin{cases} AA, & \text{Prob } r, \\ AB, & \text{Prob } 1/2-r, \\ BA, & \text{Prob } 1/2-r, \\ BB, & \text{Prob } r. \end{cases} \tag{4}$$

The degree to which this heuristic simplification is indeed justified will be discussed at the end of Sec. II B, together with a quantitative comparison of the behavior of the exact and simplified models. It is already clear that process (4) cannot describe the explicit upward migration of postmitotic cells from the basal layer. However, in the physically relevant limit $\lambda' \gg \sigma$, eliminating the vacancy population has no qualitative effect on the statistics of the progenitor cell compartment, and therefore processes (2) and (4) are expected to result in the same basal layer phenomenology.

Interestingly, process (4) is closely related to the model of monomer-monomer surface catalysis [22,23]. In particular, when the coarse-grained distribution of type A cells is effectively uniform, such that $\lambda(\mathbf{x}) \simeq \text{const}$, then one may identify the symmetric branches of process (4) with the classical zero-temperature voter model, while the asymmetric division channel $AB \rightarrow BA$ describes the Kawaski dynamics of an infinite-temperature Ising spin model. Later it will become clear that for the empirical value of $r=0.08$, the significant contribution of the latter will justify the approximation of near-constant λ.

This analogy provides access to several known results. In Ref. [18], Krapivsky showed that, starting from random initial conditions, the classical voter model (i.e., with $r=1/2$) will lead to a lattice of N sites becoming completely saturated with either type A or type B cells after a time $\lambda t \sim N \ln N$. Moreover, Frachebourg and Krapivski gave an exact solution for the two-point spatial correlations in this case [22], from which they inferred that, in the time leading up to saturation ($\lambda t \ll N \ln N$), the different cell types separate into domains of ever-increasing size, with a typical length scale growing as $L \sim \ln \lambda t$ in units of the lattice spacing (or average cell diameter), and with the density of interfaces c_{AB} between type A and type B cells dropping as $c_{AB} \sim 1/\ln \lambda t$. As the system approaches saturation ($\lambda t \sim N \ln N$), one of the cell types comes to dominate. The classical voter model is an example of domain growth in the absence of surface tension [23]. Therefore, the boundaries between domains rich in A and B cells are completely unstable, leading to strikingly different and irregular domain morphologies, compared to the smooth phase-separated shapes resulting from surface-tension-mediated domain growth.

Qualitatively, the results found for the classical voter model ($r=1/2$) allow us to make several interesting predictions relating to the spatial distribution of A and B cells. In particular, some degree of clustering of proliferating cells is to be expected in adult mice, resulting from the growth of domains rich in progenitor cells. Moreover, the ongoing growth of the domain size L suggests that larger clusters are expected in old vs young epidermis. Yet, to make full contact between process (2) and the empirical data, it becomes necessary to calculate the model properties while allowing for the relatively low value of $r=0.08$ found in the experimental system. Therefore, in the following, we shall extend the analysis of Frachebourg and Krapivsky to obtain results valid for arbitrary r. Indeed, with $r=1/4$, the following analysis results in the exact solution to the reaction-limited monomer-monomer surface catalysis model.

B. Exact solution for two-point correlations

In the following, we will follow the same approach as taken in Refs. [18,22] for the $r=1/2$ case, but we generalize to allow for arbitrary r and different lattice geometries. For completeness, we include here aspects of the solution that were also described in some detail in Ref. [18], such as the master equation and the dynamical equations required to define the problem. We start by identifying type A cells with state 1 and type B cells with state 0, so that a lattice with site index $\mathbf{i}$ may be described in terms of the Ising variables $\Phi=\{n_{\mathbf{i}}\}$, $n_{\mathbf{i}} \in \{0,1\}$. Referring to Ref. [18], the master equation for the probability distribution $P(\Phi,t)$ for the system to occupy state Φ at time t is given by

$$\begin{aligned} \frac{d}{dt}P(\Phi,t) = \frac{\lambda}{2}\sum_{\mathbf{i},\mathbf{e}} \Big[& r[U_{\mathbf{i}}^{(\mathbf{e})}(\hat{F}_{\mathbf{i}}\Phi)P(\hat{F}_{\mathbf{i}}\Phi,t) \\ & + U_{\mathbf{i}}^{(\mathbf{e})}(\hat{F}_{\mathbf{i}+\mathbf{e}}\Phi)P(\hat{F}_{\mathbf{i}+\mathbf{e}}\Phi,t)] \\ & + \left(\frac{1}{2}-r\right)U_{\mathbf{i}}^{(\mathbf{e})}(\hat{F}_{\mathbf{i}}\hat{F}_{\mathbf{i}+\mathbf{e}}\Phi)P(\hat{F}_{\mathbf{i}}\hat{F}_{\mathbf{i}+\mathbf{e}}\Phi,t) \\ & - \left(\frac{1}{2}+r\right)U_{\mathbf{i}}^{(\mathbf{e})}(\Phi)P(\Phi,t) \Big]. \end{aligned} \tag{5}$$

Here, $\{\mathbf{e}\}$ represent the nearest-neighbor lattice vectors ($|\mathbf{e}|=1$), and $U_{\mathbf{i}}^{(\mathbf{e})}(\Phi) \in \{0,1\}$ indicates whether the cells at sites $\mathbf{i}$ and $\mathbf{i}+\mathbf{e}$ are a "reactive" pair, viz.,

$$U_{\mathbf{i}}^{(\mathbf{e})}(\Phi) = n_{\mathbf{i}} + n_{\mathbf{i}+\mathbf{e}} - 2n_{\mathbf{i}}n_{\mathbf{i}+\mathbf{e}}.$$

The spin-flip operator is defined by $\hat{F}_{\mathbf{i}}\Phi=\{n_{\mathbf{j}}$ for all $\mathbf{j} \neq \mathbf{i}$; $1-n_{\mathbf{i}}\}$, so that $\hat{F}_{\mathbf{i}}\Phi$ and $\hat{F}_{\mathbf{i}+\mathbf{e}}\Phi$ correspond to the symmetric division channels and $\hat{F}_{\mathbf{i}}\hat{F}_{\mathbf{i}+\mathbf{e}}\Phi$ corresponds to an asymmetric division in which the location of the type A and B cells is reversed (viz., $AB \rightarrow BA$).

From here, recalling that $\langle n_{\mathbf{i}}n_{\mathbf{j}}\rangle = \Sigma_{\Phi}n_{\mathbf{i}}n_{\mathbf{j}}P(\Phi,t)$, it is simple to show that the two-site correlation function evolves according to the discretized diffusion equation for non-neighboring sites,

$$\frac{d}{dt}\langle n_{\mathbf{i}}n_{\mathbf{j}}\rangle = \frac{\lambda}{2}(\Delta_{\mathbf{i}}+\Delta_{\mathbf{j}})\langle n_{\mathbf{i}}n_{\mathbf{j}}\rangle, \tag{6}$$

where $\Delta_{\mathbf{i}}$ is the discrete Laplacian operator, defined by $\Delta_{\mathbf{i}}n_{\mathbf{i}} = \Sigma_{\mathbf{e}}(n_{\mathbf{i}+\mathbf{e}}-n_{\mathbf{i}})$. However, the diffusion equation is modified for nearest-neighbor correlations, giving

$$\frac{d}{dt}\langle n_{\mathbf{i}} n_{\mathbf{i+e}}\rangle = \frac{\lambda}{2}(\Delta_{\mathbf{i}} + \Delta_{\mathbf{i+e}})\langle n_{\mathbf{i}} n_{\mathbf{i+e}}\rangle + \left(\frac{1}{2} - r\right)\lambda[2\langle n_{\mathbf{i}} n_{\mathbf{i+e}}\rangle - \langle n_{\mathbf{i+e}}\rangle - \langle n_{\mathbf{i}}\rangle], \quad (7)$$

and the on-site moment is trivially $\langle n_{\mathbf{i}} n_{\mathbf{i}}\rangle = \langle n_{\mathbf{i}}\rangle$.

Making the simplifying assumption that the initial distribution $P(\Phi,0)$ is translationally invariant, then the fraction of type A cells is given by $\langle n_{\mathbf{i}}\rangle = \rho = \text{const}$, and $\langle n_{\mathbf{i}} n_{\mathbf{j}}\rangle$ depends on $(\mathbf{i}-\mathbf{j})$ at all times. Therefore, introducing the correlation function $C_{\mathbf{i}} = \langle n_{\mathbf{j}} n_{\mathbf{j+i}}\rangle$, we can rewrite Eqs. (6) and (7) as follows:

$$\frac{d}{dt}C_{\mathbf{i}} = \lambda\Delta_{\mathbf{i}} C_{\mathbf{i}} - \sum_{\mathbf{e}} \delta_{\mathbf{i},\mathbf{e}}(1-2r)\lambda(\rho - C_{\mathbf{e}}), \quad (8)$$

for $|\mathbf{i}| \geq 1$, subject to the constraint $C_{\mathbf{0}} = \rho = \text{const}$. That is, the correlation function evolves according to a discrete diffusion equation with sink terms at the nearest-neighbor sites and with a fixed boundary condition at the origin. The linear nature of the problem allows one to seek a solution in terms of the relevant Green's function, e.g., $\hat{\Delta}_{\mathbf{i}}^{-1} \equiv G_{(i_x,i_y)}(t) = e^{-4\lambda t} I_{i_x}(2\lambda t) I_{i_y}(2\lambda t)$ for a square lattice $\mathbf{i} = (i_x, i_y)$. For uncorrelated initial conditions, viz., $C_{\mathbf{i}}(t=0) = \rho^2 + \delta_{\mathbf{i},\mathbf{0}}(\rho - \rho^2)$, one may write down a general solution in the form

$$C_{\mathbf{i}}(t) = \rho^2 + (\rho - \rho^2)G_{\mathbf{i}}(t) + \sum_{\mathbf{j}} \int_0^t d\tau\, J_{\mathbf{j}}(\tau) G_{\mathbf{i-j}}(t-\tau), \quad (9)$$

where $J_{\mathbf{i}}(t)$ is the source distribution required to both maintain $C_{\mathbf{0}} = \text{const}$, and also incorporate the sink terms from Eq. (8). Thus, $J_{\mathbf{i}}(t) = 0$ for $|\mathbf{i}| > 1$, $J_{\mathbf{e}} = -(1-2r)\lambda(\rho - C_{\mathbf{e}})$, and $J_{\mathbf{0}} = z\lambda(\rho - C_{\mathbf{e}})$, where z is the number of nearest neighbors, and the initial conditions imply that $C_{\mathbf{e}}$ is the same for all nearest neighbors. For now we will explicitly consider the square lattice ($z=4$); however, the effect of changing lattice geometry will be discussed near the end of this section. Making use of Eq. (9), one may write down a set of self-consistent equations for the source terms, viz.,

$$J_{\mathbf{0}}(t) = 4\lambda\left(\rho - \rho^2 - \int_0^t \{J_{\mathbf{0}}(t-\tau)G_{(0,1)}(\tau) + J_{\mathbf{e}}(t-\tau)[G_{(0,0)}(\tau) + G_{(0,2)}(\tau) + 2G_{(1,1)}(\tau)]\}\right), \quad (10)$$

$$J_{\mathbf{e}}(t) = -\frac{1-2r}{4} J_{\mathbf{0}}(t), \quad (11)$$

which, upon taking the Laplace transform $j(p) = \mathcal{L}[J(t)]$, $g(p) = \mathcal{L}[G(t)]$, gives the expression for the source term,

$$j_{\mathbf{0}}(p) = \frac{4\lambda(\rho - \rho^2)}{p[1 + 4\lambda g_{(0,1)} - (1-2r)\lambda h]}, \quad (12)$$

where we have defined $h(p) = g_{(0,0)}(p) + g_{(0,2)}(p) + 2g_{(1,1)}(p)$ for the square lattice.

Note that, in the classical voter model ($r=1/2$), the sink terms vanish ($j_{\mathbf{e}}=0$), and a single source is located at the origin. The calculation may then proceed exactly as described in Ref. [22]. On the other hand, for the infinite-temperature Kawasaki dynamics ($r=0$), the source and sink terms create no net correlation ($j_{\mathbf{e}} = -j_{\mathbf{0}}/4$), and they serve only to maintain the stationary state $C_{\mathbf{0}} = \rho$, $C_{\mathbf{i}\neq\mathbf{0}} = \rho^2$.

The discussion so far has been exact. We now turn to the long-time asymptotic solution of the correlation function, for which it is sufficient to consider the behavior of $j_{\mathbf{0}}$ and $j_{\mathbf{e}}$ at small p ($p \ll \lambda$). In the following we shall make use of the following expansions:

$$\lim_{p/\lambda\to 0} g_{(0,0)}(p) = \frac{1}{4\pi\lambda}\ln(32\lambda/p) + \mathcal{O}[p\,\ln(p)],$$

and (for the same limit $p/\lambda \to 0$)

$$g_{(0,0)}(p) - g_{(0,1)}(p) = \frac{1}{4\lambda} + \mathcal{O}[p\,\ln(p)],$$

$$g_{(0,0)}(p) - g_{(1,1)}(p) = \frac{1}{\pi\lambda} + \mathcal{O}[p\,\ln(p)],$$

$$g_{(0,0)}(p) - g_{(0,2)}(p) = \frac{1-2/\pi}{\lambda} + \mathcal{O}[p\,\ln(p)].$$

With these expansions, the source terms take the long-time asymptotic values

$$\lim_{p/\lambda\to 0} j_{\mathbf{0}}(p) = \frac{4\pi\lambda(\rho - \rho^2)}{p[\pi(1-2r) + 2r\,\ln(32\lambda/p)]}. \quad (13)$$

The corresponding long-time behavior is therefore

$$\lim_{t\gg 1/\lambda} J_{\mathbf{0}}(t) = \frac{4\pi\lambda(\rho - \rho^2)}{\pi(1-2r) + 2r\,\ln(32\lambda t)}, \quad (14)$$

giving the expected $1/\ln t$ decay. Now the effect of Kawasaki dynamics becomes clear: one may see that reducing r from its maximal value of 1/2 has the opposing effect of weakening the magnitude of the net source ($J_{\mathbf{0}} - 4J_{\mathbf{e}} \propto 2r$) while extending the time over which the source decays, $t_r \sim (t_{r=1/2})^{1/2r}$. In the trivial limit $r=0$, the system becomes stationary as expected, whereas when $r=1/2$ one retrieves the same expression as in Ref. [22].

Taken together, Eqs. (9), (11), and (14) give the solution for the long-time asymptotic behavior of the two-point correlation function. With a mind toward experiment, as well as to comparing with the known results for $r=1/2$, we now summarize the features of this solution.

(a) The density of interfaces c_{AB} between A and B cells drops asymptotically as $c_{AB} \sim 1/2r\,\ln t$.

The density of interfaces between adjacent A and B cells, $c_{AB}(t) \equiv 2[\rho - C_{\mathbf{e}}(t)]$, is an order parameter used to describe the transition from an uncorrelated initial state with $c_{AB} = 2(\rho - \rho^2)$ to the jammed absorbing state $c_{AB}(t\to\infty) = 0$. For $r=1/2$, it was previously shown that $c_{AB} \sim 1/\ln t$.

From its definition, we identify the order parameter to be proportional to the source term, viz., $c_{AB}(t) = J_{\mathbf{0}}/2\lambda$, so that

the long-time asymptotic behavior is found from Eq. (14). This expression reveals the continuous transition between Kawasaki dynamics ($r=0$, $c_{AB}=$const) and voter dynamics ($r=1/2$, $c_{AB}\sim 1/\ln t$). In particular, as $t\to\infty$, the absorbing-state phase transition occurs at $r=0$.

(b) The spatial correlation function $C_{\mathbf{i}}(t)$ decays as $a(t)-b(t)\ln|\mathbf{i}|$ at short distances, and as a Gaussian at long distances.

Away from the origin, where a continuum description suffices, then the correlation function depends only on the distance $x=|\mathbf{i}|$, and all sources appear to be located at the origin, viz., $J_{\mathbf{i}}^{(\mathrm{eff.})}=\delta_{\mathbf{i},\mathbf{0}}(J_{\mathbf{0}}+4J_{\mathbf{e}})$. Replacing the discrete problem with a continuous one simplifies the Green's function, with $\lim_{x\gg 1} G_{\mathbf{i}}(t)\to G(x,t)=e^{-x^2/4\lambda t}/(4\pi\lambda t)$ for a square lattice. As a consequence, the long-time asymptotic correlation function may be approximated as $\lim_{t\gg 1/\lambda, x\gg 1} C_{\mathbf{i}}(t)=C(x,t)$, with

$$C(x,t)=\rho^2-\frac{2r(\rho-\rho^2)}{\pi(1-2r)+2r\ln(32\lambda t)}\mathrm{Ei}\left(-\frac{x^2}{4\lambda t}\right), \quad (15)$$

where $\mathrm{Ei}(x)=-\int_{-x}^{\infty}dt\, e^{-t}/t$ is the exponential integral. From the small-argument expansion of this integral, we find

$$\lim_{t\gg 1/\lambda, x^2\ll\lambda t}\frac{C(x,t)-\rho^2}{\rho-\rho^2}=a(t)-b(t)\ln x,$$

with

$$a(t)=1-\frac{1-2r(\pi-\gamma_e-\ln 8)}{2r\ln(32\lambda t)+\pi(1-2r)},$$

$b(t)=4r/[2r\ln(32\lambda t)+\pi(1-2r)]$, and γ_e the Euler constant. Note that, by definition, the distance x has units of the lattice spacing, which is set by the average cell diameter in the context of the epidermal basal layer tissue.

At long times, $C(1,t)\geq C_{\mathbf{e}}(t)$, so one may infer that the correlation function is concave near the origin. Away from the origin, it is useful to define a correlation length ξ to characterize the short-range correlations, viz.,

$$\xi^{-1}\equiv(\rho-\rho^2)\left.\frac{\partial C}{\partial x}\right|_{x=1}=b, \quad (16)$$

corresponding to the typical size of A-cell rich domains growing as $\xi\sim\ln t$. One may see that, when $r>0$, a variation in r merely adjusts the correlation length by a constant, in contrast to its effect on the order parameter c_{AB}. For $r=0$, the lattice configuration is random and $C(x,t)=\rho^2$, corresponding to an "infinite" correlation length.

(c) The time to saturation T of a lattice of N sites is approximately $T\sim N[\ln N+1/(2r)]/\lambda$.

As mentioned earlier, the classical voter model predicts that any finite-sized system inevitably approaches an absorbing state, in which either the lattice is completely saturated by type A cells or else they have become extinct. Fortunately, the time scale T in which the absorbing state is reached is $T\sim(N\ln N)/\lambda$ for $r=1/2$, which, for a biological system with $1/\lambda\sim 1$ week and $N\gg 1$, far exceeds the lifetime of a mammalian organism.

For arbitrary r, the time scale T may be estimated by repeating the calculation in Ref. [18]. The saturation condition is $\Sigma_{\mathbf{i}}C_{\mathbf{i}}(T)=N\rho$. Replacing the summation by integration, $\Sigma_{\mathbf{i}}C_{\mathbf{i}}\to\int_0^{\infty}xC(x,t)dx$, we arrive at the asymptotic relation

$$\lim_{\lambda t\gg 1}\sum_{\mathbf{i}}\frac{C_{\mathbf{i}}-\rho^2}{\rho-\rho^2}\simeq\frac{4r\lambda T}{\pi(1-2r)+2r\ln(32\lambda T)},$$

and the result for T follows. Interestingly, this result predicts that even in a reasonably large system and for finite r, when $N\gg e^{\pi/2r}$, the time to saturation is insensitive to the value of r. Indeed, for such systems the crossover to non-voter-like behavior occurs only at very small values of $r\sim 1/\ln N$.

(d) The product of the correlation length and interface density gives a time-invariant characteristic of voter-like coarsening.

It has been previously noted that, in voter-like coarsening, the characteristic length scale of domains is inversely proportional to the interface density c_{AB} [16]. An important implication of this observation is that, at asymptotically long times, one may identify a time-invariant characteristic of the correlations, which we define as $\Omega\equiv(\rho-\rho^2)/(c_{AB}\xi)=2r/\pi$.

To characterize the Ω constant, one may identify its definition as the ratio between the domain size ξ and domain circumference, as calculated from the number of A-B interfaces associated with each domain $c_{AB}\xi^2$. Thus, the domain perimeter has a trivial fractal dimension of 1, with Ω indicating the perimeter roughness, or curvature. Values of $\Omega\simeq 1$ may be associated with cohesive and smooth domains (i.e., with a large area-to-interface ratio), whereas systems with $\Omega\to 0$ have highly fragmented, or rough, domains. Not surprisingly, smaller values of r lead to rougher domains as a result of the Kawasaki dynamics. Yet one may see that, even at the maximum value of $r=1/2$, the voter model predicts rough domains ($\Omega<1/2$)—an observation readily seen in simulations; see, e.g., Ref. [23].

With an eye to the empirical analysis in the next section, let us note that the Ω constant allows us to characterize *static* observations of type A cell correlations, and thus provides a valuable test of whether a given data set is consistent with the long-time behavior of process (4).

(e) The lattice geometry influences only the time scale of clustering, with higher coordination number corresponding to a faster time scale.

For lattice geometries with different coordination numbers $z\neq 4$ (and one site per unit cell), the calculation proceeds as for a square lattice, but now three modifications must be made. First, as mentioned above, the central source term is $J_{\mathbf{0}}=z\lambda(\rho-C_{\mathbf{e}})$, and there are now z sink terms $J_{\mathbf{e}}$ at nearest-neighboring sites. Second, the appropriate Green's function now has the (nonseparable) form

$$G_{\mathbf{i}}(t)=\frac{1}{4\pi^2}\int d^2\mathbf{q}\,\exp\left[i\mathbf{q}\cdot\mathbf{i}-\lambda t\left(z-\sum_{\mathbf{e}}e^{i\mathbf{q}\cdot\mathbf{e}}\right)\right],$$

from which we obtain the general form of the small-p (long-time) expansion of the Laplace transform $\lim_{p/\lambda\to 0} g_{\mathbf{0}}(p)\sim\ln(\lambda/p)+d_{\mathbf{0}}/2\lambda$ and $g_{\mathbf{i}}(p)=g_{\mathbf{0}}(p)-d_{\mathbf{i}}/2\lambda$, where $d_{\mathbf{i}}$ are nu-

merical constants, e.g., $d_e=2/z$. Third, the sum over sink terms in Eq. (10) is revised to reflect the lattice geometry. With these three modifications, one finds that the order parameter c_{AB} takes the general form $c_{AB}=2\pi(\rho-\rho^2)/[\alpha(1-2r)+2r\ln(\beta\lambda t)]$, where α and β are geometry-dependent constants. One is led to conclude that the lattice geometry serves only to rescale the time variable by some constant, $t\to\beta e^{\alpha(1-2r)}t$. It is interesting to contrast this with the more significant effect of modifying the branching ratio r, which instead rescales $\ln t$.

With regard to the experimental system, the apparent insensitivity of the results to details of the lattice geometry reinforces the validity of a lattice-based description of the basal layer. Namely, while it is clear that the basal layer is not a periodic lattice of uniform cell size and coordination number, it may nonetheless be modeled as such.

Finally, because the average coordination number of $z=6$ is expected for the biological system, we have calculated the order parameter exactly for a hexagonal lattice, giving $\alpha^{(\mathrm{hex})}\simeq 0.98\pi$, and $\beta^{(\mathrm{hex})}=51$.

This completes our theoretical discussion of process (4), and of the monomer-monomer surface catalysis model. It now remains to be seen whether the exact model of cell division (2) is indeed described by the properties of the approximate process (4) (with uniform λ). Let us recall that the latter is a reasonable approximation assuming that the density of type A cells is approximately uniform. Thus, although the approximation must fail on the time scale $\lambda T\sim N\ln N$ associated with the jamming transition, the logarithmically slow growth of correlations suggests that at shorter times ($t\ll T$) the degree of progenitor cell clustering should be sufficiently low as to make the analysis self-consistent.

To test whether the model indeed satisfies the expected behavior, in Fig. 3 we compare the spatial correlation of type A cells, as given by the properties derived above, with the results of process (2) as obtained by numerical simulation using a Gillespie-like algorithm (described below in Sec. III A). An example of the cellular automata simulations used for the comparison is shown in Fig. 4, where the distribution of type A cells (black) is shown on a hexagonal lattice. Qualitatively, one may see from this figure that process (2) does indeed give rise to some clustering. By extracting the correlation function from such figures, we obtained the quantitative comparison shown in Fig. 3. In Fig. 3(a), we compare the evolution of the inverse order parameter $1/c_{AB}$ against $\ln t$ for the two models. One may see from the linear behavior of the exact model (red online), that the order parameter indeed shows the expected $1/\ln t$ decay. Equally, we may confirm that the radial correlation function $C(x,t)$ has the functional form $a-b\ln x$, with parameters $(1-a)$ and b both proportional to c_{AB}, as demonstrated by plotting $[C(x,t)-\rho]/c_{AB}$ in Fig. 3(a) (inset). Finally, we may confirm that both models have the same long-time r dependence by plotting Ω against r in Fig. 3(b). Here, one may see that indeed Ω depends linearly on r.

Yet, although Fig. 3 reveals that processes (2) and (4) result in the same functional dependence of the correlation function on r and t, it is striking that the exact model (2) results in a slower growth in correlations compared to the

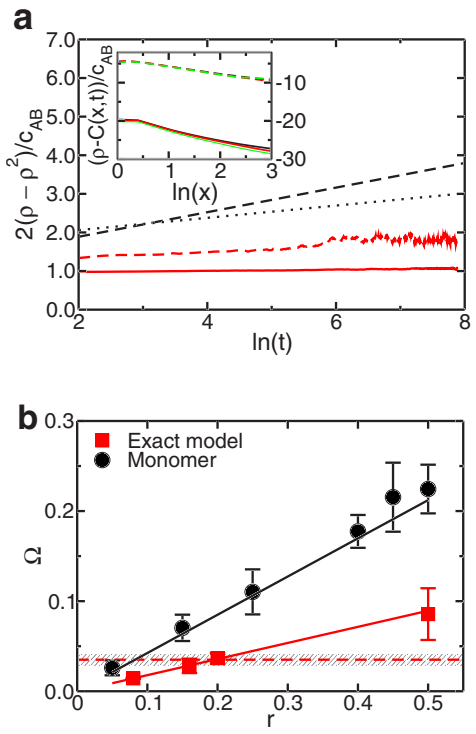

FIG. 3. (Color online) Comparison of the exact basal layer lattice model (2) (red curves) with the simplified model (4) (black). (a) The order parameter $1/c_{AB}$ is plotted against $\ln t$. The curves correspond to the simplified model with $r=1/2$ (dashed black) and $r=1/4$ (dotted), and to the exact model with $r=1/2$ (dashed red) and $r=0.08$ (solid curve). For the simplified model, the analytical expression for c_{AB} was used. For the exact model, the results were obtained from numerical simulation on hexagonal lattices of size $N=1024\times 1024$, with $\rho=0.22$ and setting $\lambda=1$. Inset: The correlation function plotted against logarithm of the distance, $\ln x$, evaluated from numerical simulation and Eq. (17). The time-invariant ratio $[\rho-C(x,t)]/c_{AB}$ is plotted at $\lambda t=50$, 100, 1000 for the classical voter model (dashed), and at $\lambda t=350$, 3500, 7000 for the exact model with $r=1/2$ (solid curves). Consistent with the expected behavior [see Eq. (15)], the curves overlap and are convex near $x=1$, becoming linear at $x>1$. (b) The long-time asymptotic roughness constant Ω plotted against r for both models. Data points correspond to the values calculated using Eq. (17) from numerical simulations such as shown in Fig. 4, using the algorithm described in Sec. III A. Error bars result from fluctuations due to finite-size effects, as evaluated by considering the random variation in Ω over the time of the simulation. The black curve gives the theoretical value of Ω for the simplified (monomer) model on a hexagonal lattice, while the red curve gives the best fit to the numerical results for the exact model.

simplified model, characterized both by a slower decay of c_{AB} seen in Fig. 3(a), and by rougher domains (or lower Ω) seen in Fig. 3(b). How might one explain this change? Referring to the discussion in Sec. II A, let us recall that the simplified model is connected to the exact model by relating the local division rate to the mean-field vacancy density $[\lambda(\mathbf{x})=\lambda' n_{\oslash}(\mathbf{x})]$, with the two models becoming equivalent when $n_{\oslash}(\mathbf{x})=\mathrm{const}$ for an arbitrarily small degree of coarse graining. This condition is satisfied in the limit $\sigma\gg\lambda'$ with the process $A\oslash\to\oslash A$ removed. However, for the physically

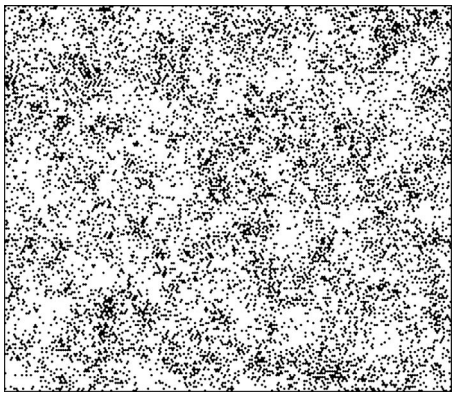

FIG. 4. Cellular automaton simulation of process (2), showing the distribution of progenitor cells (black hexagons) on a lattice of $N=200\times200$, and using the experimental branching ratio $r=0.08$ and progenitor cell fraction $\rho=0.22$. The frame shown corresponds to an evolution time of $t=30/\lambda$, where $t=0$ corresponds to random initial conditions. White areas are fully occupied by postmitotic cells.

meaningful case $\sigma\ll\lambda'$ one can no longer treat the vacancy density as uniform, leading to quantitative (but not qualitative) differences in the behavior of the two models. As shown schematically in Fig. 5, the vacancy population has the effect of accelerating cell division on the edge of smooth progenitor cell clusters, as a result of the higher concentration of nearby postmitotic cells migrating into the superbasal layers. Conversely, cell division on rough cluster edges is slowed down. In total, rough *A*-*B* interfaces remain stable over a longer period of time, leading to the observed differences between the two models.

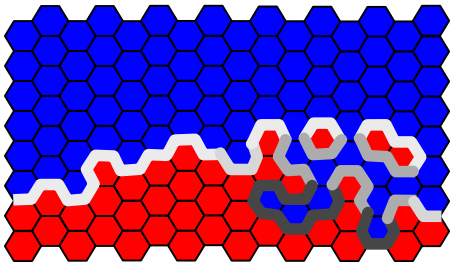

FIG. 5. (Color online) Schematic demonstrating the variation in the effective division rate in process (2), accounting for the quantitative variation in behavior between processes (2) and (4). The reactive interface between progenitor cells [light gray (red)] and postmitotic cells [dark gray (blue)] is indicated by the thick gray line, with a faster division rate shown in light gray and slower division rate shown in dark gray. On the left, a smooth interface results in faster cell division, as few progenitors on the boundary must compensate for the migration of many postmitotic cells in the bulk. On the right, the "rough" interface results in a variation between fast and slow division rates, due to restricted access of vacancies into the rough interfacial regions. Referring to the discussion at the end of Sec. II B, this leads to increased stability of rough compared to smooth interfaces.

III. EMPIRICAL ANALYSIS

We are now in a position to turn to the experimental analysis of the IFE. In order to identify the underlying rules of cell division and differentiation, one can envision two types of experiment.

(a) First, one may track the fate of *individual cells* and their progeny, and then look for a cell kinetic description compatible with their observed behavior. Such an approach was used with considerable success in the definition of the zero-dimensional process (1) through clonal analysis, and we shall extend it to the analysis of the spatial process in Sec. III A.

(b) Second, in Sec. III B we shall consider the statistics of the *entire population* of basal layer cells. In particular, by immunostaining basal layer cells for markers of cell proliferation, it is possible to analyze the spatial distribution of all progenitor cells within the layer. Then, referring to the theoretical discussion in Sec. II, one may look for a signature of the underlying cell kinetics in the spatial correlation of proliferating cells.

Note that the two types of experiment give access to independent aspects of cell behavior: The former probes the temporal evolution of cell lineages, whereas the latter reveals the static basal layer morphology. As such, the experiments provide a significant degree of mutual verification of any proposed theory of cell behavior. Yet even the best of such experiments leave room for some ambiguity: For example, it is far from clear what importance should be assigned to the embryonic development of the IFE in predetermining the spatial distribution of cells. Moreover, one may in principle conceive of regulatory pathways that leave no signature on the spatial distribution of cells, or which may not be distinguished from an independent stochastic process. Thus, in the following we will look for the simplest possible model of cell behavior that succeeds in capturing the known biological constraints.

A. Clonal analysis

To begin our analysis of the empirical data, we start by considering the fate of individual labeled cells and their progeny, hereafter referred to as clonal fate data. In an extensive experiment reported elsewhere [5], the low-frequency labeling of approximately 1 in 600 basal layer epidermal cells at a defined time was achieved by a drug-inducible genetic event, which resulted in expression of the enhanced yellow fluorescent protein (EYFP) gene in a cohort of mice. At intervals, the EYFP label was detected by confocal microscopy, which enables three-dimensional (3D) imaging of entire sheets of epidermis. By analyzing samples of mouse epidermis at different time points, it was possible to analyze the fate of labeled basal layer clones at single cell resolution *in vivo* for times up to one year post-labeling in the epidermis [see, for example, Fig. 1(b)] [2,5].

The distribution of cells within a clone constitutes a valuable data set, which one may analyze for signatures of the underlying rules of cell division. By scoring the total number of labeled cells n in each clone at progressive time points, it was possible to access the evolution of the full clone size

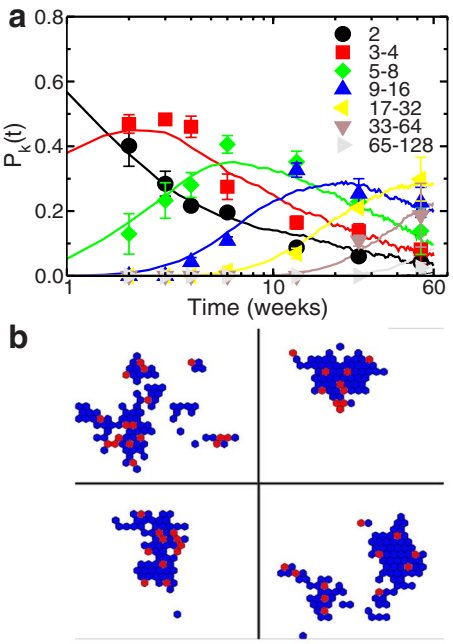

FIG. 6. (Color online) (a) Comparison of the empirical clone size distribution (data points) to predictions of process (2) (solid curves), as obtained from Monte Carlo simulations of 10^4 labeled clones. The distributions are plotted in terms of the probability $\mathcal{P}_k(t)$ for a clone to have between $2^{k-1}+1$ and 2^k basal layer cells at time t after labeling, normalized to include only clones with two or more cells in the basal layer ($k \geq 1$). The empirical data are reproduced from Ref. [5]. (b) Examples of the basal layer structure of several large late-stage clones evolving according to process (2), starting from a uniform random distribution of unlabeled cells, with one single type-A cell labeled at $t=0$. Light and dark gray (red and blue online) hexagons indicate sites occupied by type A and B cells, respectively. White areas are populated by unlabeled cells. Each frame corresponds to the progeny of one initially labeled cell.

distribution, as given by the probability $P_{n>0}(t)$ for finding a clone with $n>1$ basal layer cells at a time t postlabeling, shown in Fig. 6(a). The properties of the size distribution were used to infer the laws summarized in process (1) [5,6]. In particular, it was shown that, in the long-time limit ($t \gg 1/r\lambda$), the basal layer clone size distribution conforms to the scaling form

$$\lim_{r\lambda t \gg 1} P_{n>0}(t) = \frac{\rho}{r\lambda t} \exp\left(\frac{\rho n}{r\lambda t}\right).$$

As discussed in Sec. II, we must first establish, for any proposed model of spatial behavior, that the zero-dimensional clone size distributions are faithfully reproduced.

To this end, we conducted multiple simulations of process (2) as an asynchronous cellular automaton evolving on a hexagonal lattice of $N=60\times 60$ sites over a period corresponding to $T=60$ weeks in the experimental system (recall that $\lambda=1.1/\text{week}$ [5]). At $t=0$, the lattice was fully occupied by randomly placed type A and type B cells. A single, randomly chosen, type A cell was assigned a hereditary "label" at the start of each simulation. In effect, each such simulation mimics the evolution of one labeled clone, so that repeated simulations may be used to sample the full clonal statistics, viz., Monte Carlo sampling. In particular, by tracking the number of labeled cells as a function of time over 10^4 such clone simulations, we could compare the clonal statistics predicted by process (2) with those expected from the zero-dimensional process (1). In keeping with the empirical fit from the previous section, we used a B -cell migration rate of $\Gamma=0.31/\text{week}$, and an initial A -cell fraction of $\rho=0.22$. The "fast" rates of cell division and hole diffusion were set to be $\lambda'=10^4/\text{week}$ and $\sigma=200/\text{week}$, although the precise values are unimportant provided that $\lambda' \gg \sigma \gg \Gamma$, as discussed in Sec. II. The results are plotted in Fig. 6(a), where we compare the clone size distribution $P_{n>0}(t)$ to the empirical data. One may see that the fit is remarkably good, and within the current empirical resolution is indistinguishable from the predictions of the zero-dimensional model (cf. Ref. [6]). In summary, the fit provides a first validation of process (2) as a viable model of spatial behavior in the basal layer.

We may now reconsider the clone fate data with a view to studying spatial structure. Unfortunately, although the clone size data were stored for all clones, images showing their spatial structure were retained only for a small number of clone samples. Therefore we are not in a position to conduct a comprehensive *quantitative* analysis of clone shape evolution. Nevertheless, a striking qualitative feature of clones is that they remain largely cohesive, as shown in the example in Fig. 1(b). (Indeed, without this property the very enterprise of clonal analysis would have proved difficult.) Therefore, to challenge the validity of the proposed lattice model of cell division, we test, using additional Monte Carlo simulations, the ability of the model to produce cohesive clones over the one-year time period of the experiment. Representative results are shown in Fig. 6 for large clones at 60 weeks postlabeling. When $\lambda' \gg \sigma \gg \Gamma$, clones were seen to remain largely cohesive throughout the period of the simulation. We may therefore conclude that, at least with respect to the existing clonal fate data, process (2) presents a reasonable phenomenological description of the spatial behavior of basal layer cells. In particular, the cohesive nature of clones may be completely explained in terms of an independent stochastic process, with no evidence for further forms of regulation.

Yet to what extent is the observation of cohesiveness sensitive to modifications of the basal layer lattice model? To test the degree to which cohesiveness constrains the model, we allowed for a degree of cell mobility within the basal layer, by introducing the additional exchange process $AB \rightarrow BA$ with rate $\gamma \sim \Gamma$. The exchange process is a reasonable candidate for cell behavior, motivated by the observation that keratinocytes in culture are highly motile, which leads one to postulate whether cells *in vivo* are capable of independent lateral migration in the basal layer. We found that, for any non-small value of γ (i.e., $\gamma \gtrsim \Gamma$), the progeny of labeled clones rapidly dispersed, as shown in Fig. 7. We are led to conclude that epidermal cells *in vivo* move only in response to a local density gradient resulting from cell division and migration. Moreover, a second investigation in which the hard-core mobility σ was made larger than λ' (and keeping $\gamma=0$) again led to a loss of clone cohesiveness. We are therefore led to conclude that, within the framework of the non-

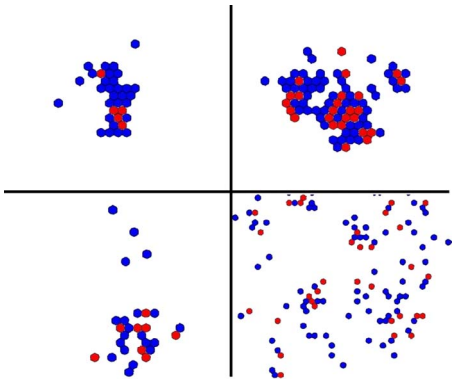

FIG. 7. (Color online) Loss of cohesiveness with increasing relative hopping rate σ/λ' shown through examples of late-stage clone simulations. Light and dark gray (red and blue online) hexagons indicate sites occupied by type A and B cells, respectively, from the same clone. White areas are populated by unlabelled cells. The simulations used the same parameter set as described above for the clones in Fig. 6, but using the parameter values $\sigma=20$/week and 200/week with $\lambda'=10\,000$/week for the top left and top right clones, respectively; $\sigma=200$/week and 2000/week with $\lambda'=200$/week for the bottom left and bottom right clones. For the latter, which is clearly unphysical, the scale has been reduced twofold to demonstrate the wide dispersion of labeled cells. The examples demonstrate that clones are cohesive when $\sigma/\lambda' \ll 1$, but dispersive otherwise.

interacting lattice model, progenitor cells undergo division rather than lateral migration as a response to a local drop in density.

In conclusion, the observation of clone cohesiveness imposes a severe constraint on cell behavior, which allows us to rule out several variations of the basic lattice model. Yet, at least qualitatively, there is no clear evidence either for or against additional regulation of cell division in the clone shape data. So to what extent is the lattice model truly capable of shedding new light on the mechanism of cell fate regulation? To do better, we must look for quantitative data for comparison. Fortunately, such data can be found in the readily available spatial distribution of progenitor cells.

B. Correlation analysis

As mentioned earlier, an intriguing feature of the noninteracting lattice model is the prediction of clustering of proliferating cells. How does this prediction compare with experiment? Before we turn to considering new results for mouse tail skin, it is interesting to first consider results presented in past work. In a detailed study of proliferating cells in hamster cheek epidermis by Gibbs and Casarett [9], a subset of progenitor cells were labeled using a radioactive marker for DNA synthesis (S phase). By measuring the number of unlabeled cells separating consecutive labeled ones in one-dimensional basal layer cross sections, it was possible to access the full radial distribution for the separation between adjacent cells undergoing S-phase synthesis. Remarkably, while the large-interval distribution decayed exponentially as expected for an uncorrelated random distribution, the probability of finding a nearby cell in S phase was significantly higher at short distances, indicating that proliferating cells were clustered. A second study addressing the distribution of S-phase cells in mouse esophagus also revealed identical qualitative results [10].

Not surprisingly, in the absence of the intuition afforded by the voter model, such clustering has been interpreted in the biological community as evidence of an underlying regulatory process, which leads to the synchronous division of nearby cells [10]. On the other hand, an independent analysis of clone fate data in Ref. [6] indicates that cell division within clones occurs independently. It is therefore satisfying to note that the tendency of proliferating cells to cluster *is* in fact consistent with independent division—indeed it is the hallmark of voter-model dynamics.

To extract the spatial correlation between progenitor cells in mouse tail skin, we analyzed confocal micrographs of basal layer cross sections of IFE that were stained for the proliferation marker Ki67, such as the one shown in Fig. 8(a). Cells bright in Ki67 are designated as type A (progenitor) cells, whereas Ki67-dull cells were designated as type B (i.e., differentiated) cells, with no capacity to divide. Using image analysis software (IMAGEJ), the coordinates of each Ki67-bright cell were extracted. This data allows a full statistical analysis of the spatial distribution of progenitor cells. In particular, we shall focus on the radial correlation function

$$C(x) = \left\langle \frac{1}{A}\int_0^{2\pi}\frac{d\theta}{2\pi}\int_A d\mathbf{x}' n(\mathbf{x}')n(\mathbf{x}'+\mathbf{x}(x,\theta)) \right\rangle, \quad (17)$$

where A denotes the area of each sample, $n(\mathbf{x})$ denotes the areal density of proliferating cells at position $\mathbf{x}$, and the angular brackets $\langle\cdot\rangle$ indicate averaging over all basal layer samples.

Using Eq. (17), the aim of the correlation analysis is to assess whether basal layer progenitor cells in adult mice do indeed cluster, and if so, to assess whether the experimental correlation function is consistent with the predictions made in Sec. II B. In principle, one may look to the data for signatures of the expected spatial dependence $C(x,t) \sim a(t) - b(t)\ln x$ (for a given value of t), as well as for evidence of increased clustering with time, viz., $c_{AB}(t) \sim 1/(2r \ln t)$. For the latter, however, the temporal analysis is difficult to implement due to the sensitivity required to resolve the $1/\ln t$ decrease in c_{AB} at long times. In particular, the predicted increase in clustering over a biologically relevant period of $\lambda t \sim 10^1 - 10^2$ cell cycles corresponds to a decrease in c_{AB} of $<10\%$, whereas variations in the efficiency of Ki67 labeling between different mice introduce systematic errors of the same order. Therefore, in the following we shall restrict ourselves to the quantitative analysis of tissue samples taken from a single adult mouse (aged 8 weeks).

To incorporate the empirical coordinates of proliferating cells into Eq. (17), we replace the product $n(\mathbf{x}')n(\mathbf{x}'+\mathbf{x})$ with the sum $\Sigma_{i,j}\delta(\mathbf{x}'-\mathbf{R}_i)f(\mathbf{x}'+\mathbf{x}-\mathbf{R}_j)$, where $\delta(\mathbf{x})$ is the Dirac delta function, $\{\mathbf{R}_i\}$ is the set of all progenitor cell coordinates, and $f(\mathbf{X})$ is a two-dimensional Gaussian envelope with width w. With this substitution, the integrals in Eq. (17) can

be solved exactly. Then, explicitly accounting for the averaging procedure over samples of variable size, we obtain the expression

$$C(x) = \left\langle \frac{\rho}{2\pi w^2 N_A(x)} \sum_{i=1}^{N_A} \sum_{j\neq i}^{N_A(x)} \exp\left(-\frac{x^2+R_{ij}^2}{2w^2}\right) I_0\left(-\frac{xR_{ij}}{w^2}\right) \right\rangle. \tag{18}$$

Here, the sum i,j over all progenitor cell coordinates arises from the empirical expression for $n(\mathbf{x}')n(\mathbf{x}'+\mathbf{x})$ given above, and we have defined $R_{ij} \equiv |\mathbf{R}_i - \mathbf{R}_j|$. The prefactor, exponential factor, and modified Bessel function (I_0) result from solving the integrals in Eq. (17). The total number of progenitor cells found in each sample is N_A, and $N_A(x)$ is the number of progenitor cells at a distance x or more from the sample edges. As defined, the correlation function avoids errors resulting from edge effects by averaging each sample over the $N_A(x)$ progenitor cells that are unaffected by the finite sample size. To correctly average over different samples, as indicated by $\langle\cdot\rangle$, the sample results are weighted by $N_A(x)$, e.g., $C^{(1+2)}(x) = [N_A^{(1)}(x)C^{(1)}(x) + N_A^{(2)}(x)C^{(2)}(x)]/[N_A^{(1)}(x) + N_A^{(2)}(x)]$.

Making use of Eq. (18), the experimental correlation between progenitor cells was averaged over ten samples of approximately 30×30 cells each; see Fig. 8(b). Remarkably, the data indeed show a significant degree of progenitor cell clustering, in good agreement with the linear decay in $\ln x$ expected from Eq. (15). In real terms, one may infer from the value of the nearest-neighbor correlation $C(1)$ that each progenitor cell is in contact, on average, with approximately *two* adjacent progenitor cells, compared to 1.3 progenitor cells expected for an uncorrelated random distribution (at $\rho = 0.22$).

For a more careful test of the theory, one may extract from Fig. 8(b) the order parameter $c_{AB} = 0.29 \pm 0.02$, and the correlation length of $\xi = 14.7 \pm 0.7$ cells, from which we find the empirical roughness constant $\Omega = 0.04 \pm 0.01$. Repeating the analysis with samples taken from different mice results in the same value of c_{AB}, but with values of Ω in the range $\Omega = 0.02-0.04$. Referring to Fig. 3(b), where the empirical value of Ω is compared to the model predictions (dashed), we see that the model indeed recovers the correct order of magnitude of the roughness constant. Qualitatively, then, it appears that the model is consistent with the observed clustering. Let us emphasize that this fit requires *no additional parameters*, and is purely a result of mapping the cell kinetics onto a lattice.

IV. DISCUSSION AND CONCLUSIONS

To summarize, we have demonstrated that the zero-dimensional model of cell division is consistent with the maintenance of the basal layer at uniform density, and we have shown that the size distribution and qualitative shapes of labeled clones are consistent with a simple stochastic model of cell division and differentiation on a two-dimensional lattice. These results explain why the recently discovered zero-dimensional model of cell behavior gives such an excellent fit to the clone fate data despite taking no account of additional regulatory pathways. Significantly, despite the many forms of cell fate regulation known to exist in development and adult tissue, the only extracellular regulation required to understand the existing observations of clone fate in normal IFE is *steric*, viz., the coupling of cell division to the local cell density.

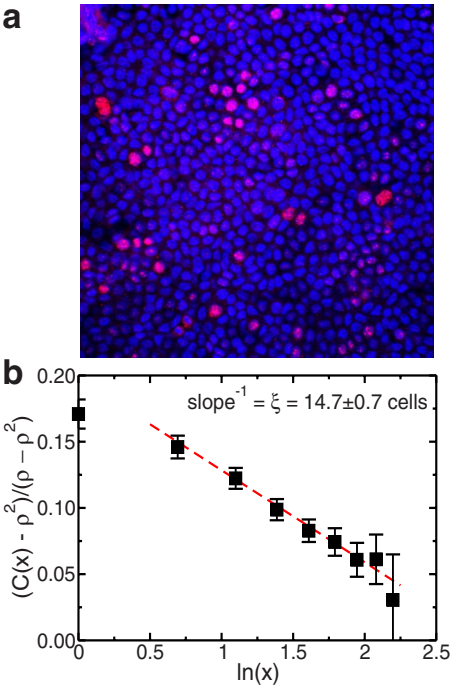

FIG. 8. (Color online) (a) Confocal micrograph of whole mounted mouse tail skin IFE, showing the two-dimensional basal layer immunostained for the nuclear marker DAPI [dark gray (blue)], and the proliferation marker Ki67 [light gray (red)]. The area surrounding the stained nuclei is occupied by unstained cell cytoplasm (black). (b) The empirical radial correlation function $C(x)$, as defined in Eq. (18). Data points show results obtained by analysis of the Ki67-stained epidermal whole mounts exemplified in (a), taken from a mouse aged 8 weeks; the dashed line shows the fit to the analytical form of the correlation function predicted by process (2), $C(x,t) = a(t) - \xi^{-1}(t)\ln x$, with x in units of the average cell diameter [see Eqs. (15) and (16)], and where the single time point t is fixed by the age of the mouse and by initial conditions (see discussion in Sec. IV). From the fitted slope one may extract the correlation length $\xi(t)$ defined in Eq. (16), giving $\xi = 14.7$ cell diameters.

Beyond the success of the model in explaining the observed clone fate data, we have also identified that the degree of progenitor cell clustering, as measured by Ki67 staining, is in good qualitative agreement with the predictions of the spatial process. In particular, there are two features that allow us to characterize the spatial process. First, the empirical roughness constant $\Omega = 0.04 \pm 0.01$ has the expected order of magnitude predicted by the model [Fig. 3(b)], and, second, the correlation function is in excellent agreement with the expected decay form $a(t) - b(t)\ln x$ (at fixed t), as seen in Fig. 8.

Taken together, these results have two important implications for future investigations of epidermal cell fate regula-

tion. First, by demonstrating that the zero-dimensional process (1) is indeed capable of maintaining a uniform total basal layer cell density, the spatial process consolidates the proposed stochastic model as a robust platform for investigating biochemical constituents in future work. For example, by over- or underexpressing specific genes and then studying the resulting change in the empirical parameters (r,λ,ρ), one may attempt to identify the role of each constituent in regulating cell behavior. Second, the model introduces a set of spatial measures (such as the roughness constant Ω) that yield further information in such investigations of the biochemistry, beyond that which may be obtained through empirical evaluation of the zero-dimensional parameters alone.

Beyond the qualitative features of the dynamics, one may ask whether there are any implications to the quantitative features of the correlation analysis. Unfortunately, as mentioned earlier, variations in the efficiency of Ki67 labeling prevent us from generating the comprehensive statistics necessary to accurately quantify the universal values of the system parameters. Nevertheless, taken at face value, it appears that the empirical value of Ω obtained from the sample analyzed in Sec. III B is consistent with process (2) only when one imposes a branching ratio $r=0.19\pm0.04$, which differs significantly from the value of $r=0.08$ established through clonal analysis [see Fig. 3(b)]. This value is further consistent with the observed AB interface concentration of $c_{AB}=0.29$ and the correlation length of $\xi=14.7$ cells, which correspond to a long evolution time of $\lambda t\approx10^4-10^6$ for $r=0.08$, but a realistic biological time scale of $\lambda t\approx10^1-10^2$ for $r=0.19$.

Although it is possible that the discrepancy in the inferred value of r results from errors in progenitor cell classification as discussed in Sec. III B, the difference is large enough to call into question the reliability of the correlation analysis. Beyond the issue of progenitor cell labeling efficiency, there is also the more basic question of whether Ki67 is at all effective as a marker of progenitor cells. In particular, it is widely accepted that Ki67 is a marker of cell growth as well as proliferation, which may imply that differentiated cells remain Ki67-bright for some time after division, or that progenitor cells may fail to continuously express Ki67 [24,25]. However, assuming that Ki67 is indeed a faulty marker, it becomes difficult to explain why the analysis nevertheless results in an excellent fit to the $a(t)-b(t)\ln x$ decay form of correlations (at fixed t). We may also challenge our assumption of using "random initial conditions" to model steady-state maintenance. If initial conditions are at all relevant, then the observed clustering may be a signature of development and growth processes that are no longer active in the adult system. To test the importance of initial conditions, we have compared the order parameter c_{AB} between adult mice aged 8 and 60 weeks (corresponding to young and old adult mice). If clustering is inherent in the initial conditions, then one would expect c_{AB} to grow over time as the memory of initial conditions erodes. On the other hand, as mentioned in Sec. III B, steady-state maintenance results in $<10\%$ change in the order parameter over this time period, which would be undetectable given the systematic errors in Ki67 labeling. Indeed, we find no change from $c_{AB}=0.29\pm0.02$ at 60 weeks, so there is no indication that the observed clustering is a signature of tissue development.

Finally, it is interesting to speculate whether additional forms of regulation may be capable of reconciling the higher value of $r=0.19$ found for the spatial process with the lower value of $r=0.08$ found from the zero-dimensional analysis. We propose a simple revision of the model that is capable of reconciling the spatial and zero-dimensional data: Referring to process (2), if the two channels of asymmetric division in process (2) occur with different probabilities (viz., $P_{A\varnothing\rightarrow AB}\neq P_{A\varnothing\rightarrow BA}\neq 1/2-r$, and $P_{A\varnothing\rightarrow AB}+P_{A\varnothing\rightarrow BA}=1-2r$), then the effective value of the branching ratio r is effectively renormalized in the spatial process, while leaving it unchanged in zero dimensions. To reconcile the two empirical values of r, it is simple to show that one requires $P_{A\varnothing\rightarrow AB}\simeq 7P_{A\varnothing\rightarrow BA}\simeq(1-2r)/8$. Thus, one might speculate that the increase in clustering is associated with a spatial asymmetry in daughter cell fate during asymmetric division. It would be an interesting challenge to devise an experiment with which to test this asymmetry.

ACKNOWLEDGMENTS

D.P.D., P.H.J., and B.D.S. acknowledge the financial support of the U.K. Medical Research Council.

[1] J. Lewis and A. Davies, J. Neurobiol. **53**, 190 (2002).
[2] K. M. Braun, C. Niemann, U. B. Jensen, J. P. Sundberg, V. Silva-Vargas, and F. M. Watt, Development **130**, 5241 (2003).
[3] E. Fuchs, Nature (London) **445**, 834 (2007).
[4] C. S. Potten, Cell Tissue Kinet. **7**, 77 (1974).
[5] E. Clayton, D. P. Doupé, A. M. Klein, D. J. Winton, B. D. Simons, and P. H. Jones, Nature (London) **446**, 185 (2007).
[6] A. M. Klein, D. P. Doupé, P. H. Jones, and B. D. Simons, Phys. Rev. E **76**, 021910 (2007).
[7] B. Houchmandzadeh, Phys. Rev. E **66**, 052902 (2002).
[8] W. R. Young, A. J. Roberts, and G. Stuhne, Nature (London) **412**, 328 (2001).
[9] S. J. Gibbs and G. W. Casarett, J. Dent. Res. **51**, 30 (1970).
[10] I. L. Cameron, D. G. Gosslee, and C. Pilgrim, J. Cell. Comp. Physiol. **66**, 431 (1965).
[11] B. I. Shraiman, Proc. Natl. Acad. Sci. U.S.A. **102**, 3318 (2005).
[12] L. Hufnagel, A. A. Teleman, H. Rouault, S. M. Cohen, and B. I. Shraiman, Proc. Natl. Acad. Sci. U.S.A. **104**, 3835 (2007).
[13] B. Dubertret and N. Rivier, Biophys. J. **73**, 38 (1997).
[14] D. Drasdo, R. Kree, and J. S. McCaskill, Phys. Rev. E **52**, 6635 (1995); J. Galle, M. Loeffler, and D. Drasdo, Biophys. J. **88**, 62 (2004).
[15] T. M. Liggett, *Interacting Particle Systems* (Springer, New York, 1985).
[16] I. Dornic, H. Chaté, J. Chave, and H. Hinrichsen, Phys. Rev.

Lett. **87**, 045701 (2001).
[17] O. Al Hammal, H. Chaté, I. Dornic, and M. A. Muñoz, Phys. Rev. Lett. **94**, 230601 (2005).
[18] P. L. Krapivsky, Phys. Rev. A **45**, 1067 (1992).
[19] G. Ódor, Rev. Mod. Phys. **76**, 663 (2004).
[20] F. van Wijland, Phys. Rev. E **63**, 022101 (2001).
[21] S.-C. Park and J.-M. Park, Phys. Rev. E **71**, 026113 (2005).
[22] L. Frachebourg and P. L. Krapivisky, Phys. Rev. E **53**, R3009 (1996).
[23] H. Hinrichsen, Physica A **369**, 1 (2006).
[24] D. C. Brown and K. C. Gatter, Histopathology **40**, 2 (2002).
[25] T. Scholzen and J. Gerdes, J. Cell. Physiol. **182**, 311 (2000).