# Supplementary methods: theory

In the main text, we describe a lineage tracing experiment using inducible Ahcre$^{\mathrm{ERT}}$R26$^{\mathrm{flEYFP/wt}}$ mice. The conclusions of our study hinge on the quantitative insights that can be drawn from the analysis of the clones of labeled cells in this experiment. The aim of this supplementary methods theory section is to expand on the basis of this analysis and its application. In doing so, we will set out in detail the experimental evidence in support of the esophageal progenitor (EP) cell model, how the parameters of this model are constrained by the experimental data, and how the model can be challenged experimentally by drug treatment. In section 1, we will discuss the additional observations that underpin the quantitative analysis of the clonal fate data – the homeostatic nature of esophageal epithelium (EE), and the representativeness of labeling. In section 2 we will describe how the experimental data identifies the dynamics of a single progenitor cell population allowing us to formulate a simple model of EE turnover. In section 3 we will use the clonal fate data to fit the parameters of the modeling scheme. Finally, in section 4, we will challenge the model by considering the effects of drug delivery on tissue.

## 1 Clonal analysis: controls

### 1.1 Homeostasis of tissue turnover in normal EE

To prepare for the quantitative analysis of the clone fate data, it is first necessary to establish whether the EE is in homeostasis over the year long timecourse of the experiment. We found that there was no significant difference in the mitotic index in the basal layer, the proportion of

cells expressing the proliferation associated antigen Ki-67, the number of basal cells per unit area or the ratio of basal cells to differentiated, nucleated suprabasal cells over the year of the experiment (Figs S4A–D). On this basis, **we conclude that the tissue was indeed homeostatic over this period.**

## 1.2 Apoptosis

We also investigated whether apoptosis occurred in normal EE by staining for activated Caspase 3. Positive cells were readily detected in control sections of EE which had been UV irradiated and cultured for 24 hours, but undetectable in 30,000 basal cells in sections of normal EE (Fig. S4E). **We therefore conclude that apoptosis is negligible in normal EE.**

## 1.3 Representativeness of the labeled cell population

Although the analysis of the Rosa26$^{\mathrm{M2rtTA}}$/TetO-HGFP transgenic mice confirms the absence of slow-cycling or quiescent epithelial cells in EE, this does not rule out heterogeneity in the proliferative potential of cycling cells. To reliably interpret the results of the clonal labeling experiment, it is therefore important to confirm that the labeled cell population is representative of the total cell population and therefore capable of revealing proliferative heterogeneity should it exist.

To this end, we first compared the cell composition within clones at the one-year timepoint with that of the surrounding unlabeled epithelium. We found that there was no statistically significant difference in the proportion of cycling cells between the labeled and unlabeled tissue, as determined by immunostaining for the S-M phase marker Geminin as well as the proliferation associated antigen Ki-67 (Figs S5 A, B). Similarly there was no statistically significant difference in the density of basal cells (cell number per unit area), or the ratio of basal cells to nucleated suprabasal cells when comparing cells within and outside labeled clones (Figs S5 C,

D).

Although the cell composition and activity of the clones within the unlabeled tissue at the one-year timepoint confirm that the labeled cell population is representative of the total population, it does not immediately follow that dividing and non-dividing cells are induced in a representative manner, or that there is not a proliferative hierarchy comprising a self renewing stem cell population which supports a population of non self-renewing progenitor cells. Providing the induction protocol includes the self-renewing population, over time, the persistent clones will converge to a stationary and representative ensemble, while clones derived from a non self renewing progenitor cells that were primed or committed to terminal differentiation on induction are transient and will become lost over time. However, since we find no statistically significant variation in the percentage of labeled basal cells over the entire experimental timecourse, from 1 month to 1 year (Fig. 2Bc), we can conclude that either (a) clones derived from a putative non self-renewing progenitor compartment are very rapidly lost (within the first month post-labeling), or (b) that such cells do not exist - i.e. there is only one progenitor cell population - or (c) that basal layer cell types are induced in the exact proportions in which they are present in tissue. Analysis of the detailed clonal fate data below will confirm the veracity of option (b).

# 2 Clonal analysis: defining the model of EP cell dynamics

## 2.1 Scaling and equipotency

To isolate and quantify the behavior of proliferating cells, we scored the number of basal cells in clones containing two or more cells (i.e. clones in which a proliferative cell was labeled at the start of the experiment) from 3 days to one year post-induction, by confocal imaging of esophageal wholemounts (Fig. 2A). Previously, it has been shown that in a similar homeostatic stratified squamous epithelium, interfollicular epidermis, a simple statistical characterization

can be used to both establish the equipotency of a self-renewing cell population, and elucidate the pattern of cell fate (18, 19). In particular, if self-renewal involves the stochastic loss and replacement of such progenitors, clones derived from these cells will undergo a process of "neutral drift" in which ongoing clonal loss is compensated by the expansion of adjacent clones. However complex is the underlying dynamics, such processes lead inexorably to the long-term scaling of the surviving clone size distribution, in which the chance of finding a clone with a size larger than some multiple of the average remains constant over time.

Applied to the current dataset, the convergence of the basal layer clone size distribution onto a scaling form (Fig. S6A) both confirms the functional equivalence of the self-renewing progenitor cell population, and shows that the balance between their proliferation and differentiation is achieved on a population basis, and not at the level of individual cell divisions, i.e., in the course of turnover, some cells may undergo terminal division leading to loss while others may undergo symmetric duplication. Indeed, such behavior is consistent with the observed heterogeneity in the clonal composition at short times, which reveals that, soon after induction, all possible permutations of two-cell clones can be found (i.e. two basal cells, one basal and one suprabasal, and two suprabasal cells). The same degree of heterogeneity is apparent in larger clones (Figs S6B, C).

Taken together, the results above suggest that, as in interfollicular epidermis, EE is maintained by a single cycling progenitor cell population in which cell division can lead to all three possible fate outcomes; two daughters that go on to divide, two cells that exit cycle and then stratify out of the basal layer, or one daughter that goes on to divide and one that differentiates (18, 19). To assess the validity of this model, and to quantify the respective rates and probabilities, we must turn now to a more detailed analysis of the short-time data, prior to scaling, where signatures of the detailed dynamics can be elucidated. To prepare for this analysis, we must embed the progenitor cell dynamics into a parameterization that includes the stratification

and loss of terminally differentiated cells.

## 2.2 Parameterizing EP dynamics

To be concrete, we will suppose that the basal layer comprises a single population of cycling esophageal progenitor cells (with proportion $\rho$), and their differentiating progeny (with proportion $1-\rho$), which remain in the basal layer without dividing until they stratify. If we assume that the progenitors divide with an average rate $\lambda$, and the differentiating cells stratify at an average rate $\gamma$, to achieve homeostasis, it follows that $\lambda\rho = (1-\rho)\gamma$, i.e. the rate at which differentiated cells are generated in the basal layer must be perfectly matched by the rate at which they stratify into the suprabasal cell layer.

To further specify the model, we must also define the distribution of cell cycle and stratification times. Here, for simplicity, we will suppose that division and stratification are both uncorrelated between successive events leading, in both cases, to an exponential (Poisson) time distribution, with averages set by the rates $\lambda$ and $\gamma$, respectively. While this assumption is surely unsafe (after all, cell division must be accompanied by a small refractory period before further division is possible), any degree of synchrony in the cell cycle or stratification timings will be rapidly erased from the clonal record over the timecourse. In particular, we expect features due to potential synchrony to be lost within experimental error bars after ca. 1–2 rounds of division (20).

Alongside the division and stratification rates, we must further specify the probabilities for the respective fate outcomes following division. Since both the labeled cell number and their composition is found to be conserved over the time course, any process of cell division must be consistent with homeostatic turnover. Moreover, since, through scaling, EPs are seen to function as an equipotent cell population, we will therefore suppose that proliferation and differentiation is finely balanced so that, with probability, $r$, EP division results in two dividing cells, with

probability, $1-2r$, one dividing and one non-dividing basal later cell, and with probability, $r$, two non-dividing basal layer cells. Although we cannot rule out a small contribution arising from perpendicular cell divisions resulting in the placement of one of the daughters directly into the suprabasal layer, given such divisions are observed to be rare, the effect of such "orientationally asymmetric" divisions would again be impossible to resolve. Similarly, we will neglect apoptosis, which was found to be negligible in the basal layer (Fig. S4E).

In summary, the time-evolution of the basal layer population, along with the clonal fate data, is therefore fully characterized by three adjustable parameters, the division rate, $\lambda$, the stratification rate, $\gamma$, and the fraction of divisions that lead to symmetrical fate, $r$. The progenitor cell fraction is then fixed by the rates $\rho = \gamma/(\lambda+\gamma)$. To fully characterize the behavior of the *total* clone size, including suprabasal as well as basal cells, we must include a further parameter, $\mu$, which defines the average rate at which suprabasal cells are shed from the tissue (once again, we will suppose for simplicity that the corresponding distribution of shedding times is Poisson). Fortunately, this additional parameter can be related directly to the division and stratification rates through the ratio of nucleated suprabasal to basal layer cells, $m$, which can be measured directly. Then, since the "flux" of differentiated cells stratifying from the basal layer must be perfectly compensated by the flux of differentiated cells that are shed, we must have $m\mu = \rho\lambda = \gamma\lambda/(\gamma+\lambda)$, thereby providing a relation linking $\mu$ with $\gamma$ and $\lambda$.

Taken together, the EP cell model dynamics can be cast as a critical (i.e. balanced) continuous time Markovian branching process,

$$\mathrm{EP} \xrightarrow{\lambda} \begin{cases} \mathrm{EP}+\mathrm{EP} & \mathrm{Pr.}\ r \\ \mathrm{EP}+\mathrm{T_B} & \mathrm{Pr.}\ 1-2r \\ \mathrm{T_B}+\mathrm{T_B} & \mathrm{Pr.}\ r \end{cases}, \qquad \begin{aligned} \mathrm{T_B} &\xrightarrow{\gamma} \mathrm{T_S}, \\ \mathrm{T_S} &\xrightarrow{\mu} \text{loss}, \end{aligned}$$

where $\mathrm{EP}$ represents the progenitor, $\mathrm{T_{B/S}}$ differentiated cells in the basal/suprabasal layer, and the rates $\lambda$, $\gamma$, and $\mu$ are defined above. As discussed above, we suppose that EP cell fate follows a pattern of intrinsic or cell-autonomous regulation in which the fate outcome following division

is uncorrelated with the behavior of neighboring cells. In previous studies, we have shown that, in the two-dimensional geometry pertinent to EE, the long-term clonal dynamics is largely insensitive to this choice (30).

This completes the specification of the cellular dynamics as a generalized branching-type process. In principle, we could further refine the modeling scheme to include aspects of the spatial regulation. However, our aim here is to establish and challenge the simplest model which is consistent with the observed range of clonal fate data. Indeed, we will find that, by itself, this model provides a faithful description of the cellular dynamics over the timecourse of the experiment. To assess the integrity of the model, and to fit the three adjustable parameters, we will make use of a Bayesian approach.

# 3 Clonal analysis: quantitative modeling

## 3.1 Scaling

According to the EP cell paradigm, following induction, a differentiated basal or suprabasal cell would progressively stratify, lose its cell nucleus and be shed, providing only a transient contribution to the clonal dynamics. By contrast, a clone derived from an EP cell would progressively undergo chance expansion or contraction according to the fate choice of the constituent cells. As tissue turns over, the gradual extinction of some clones due to chance commitment to terminal differentiation will be perfectly compensated by the expansion of other clones. Analysis of the branching process shows that, over time, neutral drift dynamics leads to a progressive (linear) increase in the average size of the surviving clones, while the surviving fraction falls proportionally such that the total density of labeled cells remains constant (Fig. 2B) (18). In this regime, the basal layer clone size distribution (Fig. S6A) acquires a scaling behavior (Fig. S6A) described above. In particular, the chance of finding a clone with more than $n$ basal layer cells, takes the form of an exponential $\exp[-n/\langle n(t)\rangle]$ where, according to the parameters of

the model, the average basal layer clone size (i.e. the average size of the clone "footprint" on the basal layer) is given by $\langle n(t) \rangle \simeq t/\tau$ with $1/\tau = \rho/r\lambda$.

Although the long-term scaling behavior provides the means to extract the characteristic timescale $\tau$ (see below), the rapid convergence to this regime does not allow the individual parameters $\rho$, $r$, and $\lambda$ to be inferred reliably from the basal layer clone size distribution alone. Therefore, to properly validate the model, and effect a reliable fit of the parameters, we consider the full range of clonal fate data, short-term and long-term, basal and suprabasal.

## 3.2 Parameter estimation

The data itself is acquired in the form of a set of frequencies $D = \{f_{bn}(t)\}$ describing the number of observations of clones with $b$ basal cells and $n$ suprabasal cells at time $t$; an example is Fig. S6B. The stochastic nature of our model will predict probabilities $p_{bn}(t)$ for a single cell to turn into a clone at time $t$ with $b$ basal and $n$ suprabasal cells, based on the parameters $\rho$, $r$ and $\lambda$. One might approach the fit with a least-squares method to optimize the parameters. However, as a significant portion of our data lies in the tails with large $b$ and $n$, where the counts are small, we would have to employ complex methods such as ad-hoc binning or continuity corrections. As such, we have elected to use a more Bayesian approach, which allows us to analyze the experiment with no manipulation of the data.

To fit the model to the range of experimental data, we implement a basic algorithm involving Bayes' theorem for updating prior probabilities in the presence of data. More precisely, the probability that the observed clonal fate data $D$ is described by the model is specified by the proportionality,

$$P\left(\lambda, r, \rho | D\right) \propto P\left(D | \lambda, r, \rho\right) P\left(\lambda, r, \rho\right),$$

where the posterior distribution $P(D|\lambda, r, \rho)$ denotes the probability of obtaining the data $D$ given the parameters $\lambda$, $r$, and $\rho$, multiplied by the likelihood that the parameters are given by

those same values. The constant of proportionality may be calculated *a posteriori* by imposing the normalisation, $\int_0^\infty d\lambda \int_0^1 d\rho \int_0^{\frac{1}{2}} dr\, P(\lambda, r, \rho|D) = 1$. In our case, the posterior distributions will turn out to be approximately Gaussian, so we may characterize the distribution by its first two moments and treat them as a point estimate for the parameters, and a credible interval covering approximately $68\%$ (using one standard deviation) or $95\%$ (using two).

From Bayes' equation above, it is apparent that we need to specify a prior distribution $P(\lambda, r, \rho)$ and a likelihood function $P(D|\lambda, r, \rho)$. We chose the maximum entropy prior as we have no further cogent information, which corresponds to $\rho$ uniform on the interval $[0, 1]$, $r$ uniform on the interval $\left[0, \frac{1}{2}\right]$, and $\lambda$ log-uniform. Crucially, since we have sufficient data, the likelihood function is sharply peaked and only significant on a small support, and thus the posterior distribution is dominated by the likelihood function, and so any reasonable prior (one which does not fluctuate strongly over that narrow support) would give quantitatively similar results.

Turning to the likelihood function, since each observed clone is independent, it is a multinomial distribution with a countable number of outcomes:

$$P(D|\lambda, r, \rho) = \prod_t \frac{[\sum_{bn} f_{bn}(t)]!}{\prod_{bn} [f_{bn}(t)!]} \prod_{bn} [p_{bn}(t)]^{f_{bn}(t)},$$

where $p_{bn}(t)$ denote the expected probabilities that a clone contains $b$ basal cells and $n$ suprabasal cells at time $t$ post-induction. In the following, we will suppress the time index $t$ for notational concision where it is considered unambiguous. To eliminate potential ambiguities due to uncertain induction frequencies of the two cell types, we conditioned the data on having at least two cells, which removes the events where a differentiated cell is induced. Moreover, to focus on a defined population, we further consider only clones which retain at least one basal layer cell. Thus we consider the probabilities,

$$p_{bn} \mapsto \tilde{p}_{bn} = \begin{cases} 0 & b = 0 \text{ or } b + n < 2, \\ \frac{p_{bn}}{1 - p_{10} - \sum_n p_{0n}} & \text{otherwise.} \end{cases}$$

To determine the probabilities $p_{bn}$, we consider the more fundamental distribution $p_{lmn}$ describing the (unconditioned) probability of finding a clone with $l$ progenitors, $m$ differentiated basal layer cells and $n$ suprabasal cells. From these probabilities we can determine $p_{bn}$ through the relation, $p_{bn} = \sum_{l+m=b} p_{lmn}$. The probabilities $p_{lmn}$ obey a continuous time Markov process, with the time-evolution given by the Master equation,

$$\begin{aligned}\frac{d}{dt}p_{lmn} = \lambda \left[r(l-1)p_{(l-1)mn} + (1-2r)lp_{l(m-1)n} + r(l+1)p_{(l+1)(m-2)n} - lp_{lmn}\right] \\ +\gamma \left[(m+1)p_{l(m+1)(n-1)} - mp_{lmn}\right] - \mu np_{lmn}.\end{aligned}$$

with the initial conditions $p_{lmn} = \delta_{l1}\delta_{m0}\delta_{n0}$.

Although this is a straightforward set of coupled differential equations, direct numerical solution is prohibitively computationally expensive because each element is coupled to all the others: attempts to truncate the $lmn$-space causes errors that are hard to control. Instead, to develop the computational scheme, it is helpful to package the equations into the form of a generating function, $G(x,y,z,t) = \sum_{lmn} p_{lmn}(t)x^l y^m z^n$, which obeys the differential equation,

$$\begin{aligned}\frac{\partial G}{\partial t} &= \lambda \left[rx^2 + (1-2r)xy + ry^2 - x\right]\frac{\partial G}{\partial x} + \gamma\left(z-y\right)\frac{\partial G}{\partial y} + \mu(1-z)\frac{\partial G}{\partial z}, \\ &= \lambda\left[f_x(x,y,z) - x\right]\frac{\partial G}{\partial x} + \gamma\left[f_y(x,y,z) - y\right]\frac{\partial G}{\partial y} + \mu\left[f_z(x,y,z) - z\right]\frac{\partial G}{\partial z}.\end{aligned}$$

with $G(x,y,z,t=0) = x$ and

$$f_x(x,y,z) = rx^2 + (1-2r)xy + ry^2, \qquad f_y(x,y,z) = z, \qquad f_z(x,y,z) = 1$$

which can be recognized as the probability generating functions for individual divisions of each cell type.

The partial differential equation for $G$ may be solved by the method of characteristics. Introducing the auxiliary equations,

$$\frac{\partial}{\partial t}F_v(\mathbf{v},t) = a_v\left[f_v\left(F_x,F_y,F_z\right) - v\right]$$

where $\mathbf{v} = (x, y, z)$, $v$ ranges over the set $\{x, y, z\}$, $a_v$ ranges over $\{\lambda, \gamma, \mu\}$ and $F_v(\mathbf{v}, t = 0)$ ranges over $\{x, y, z\}$, we can write the solution to $G$ as $G(x, y, z, t) = F_x(x, y, z, t)$. In fact, we can recognize the auxiliary function $F_v$ as the probability generating functions for the clone size distribution starting with a cell of type $v$.

Finally, the coefficients $p_{lmn}$ may be extracted from $G$ by complex analysis. Indeed, it is possible to directly extract $p_{bn}$ by considering $G(x, x, z, t)$. Defining

$$H(x, y, t) = G(x, x, y, t) = \sum_{bn} p_{bn}(t) x^b y^n,$$

we see that $H$ is analytic in both $x$ and $y$ on the complex unit disc $|x|, |y| \leq 1$, and so there are no poles within it. Defining $\mathcal{C}$ to be an anticlockwise contour around the unit circle, we can recover $p_{bn}$ by considering residues at zero:

$$p_{bn}(t) = \left(\frac{1}{2\pi i}\right)^2 \oint_{\mathcal{C}} dx \oint_{\mathcal{C}} dy \frac{H(x, y, t)}{x^{b+1} y^{n+1}}.$$

Note that, up to a minus sign, it is just a multi-dimensional inverse $Z$-transform. The integral can be approached by a sum

$$p_{bn}(t) = \lim_{M \to \infty} \sum_{j=1}^{M} \sum_{k=1}^{M} H\left(e^{2\pi i j/M}, e^{2\pi i k/M}, t\right) e^{-2\pi i j b/M} e^{-2\pi i k n/M}$$

the truncation of which at finite $M$ approximates the integral and is in the form of a two-dimensional discrete Fourier transform. This allows the use of industrial Fast Fourier Transforms to evaluate all $p_{bn}$ for $M$ being a power of two simultaneously. We use adaptive oversampling to estimate the error in the truncation, and use it to bound the errors. This gives a robust black box algorithm which gives point estimates and credible intervals for arbitrary clonal data sets.

Before finally carrying out the parameter estimation, it is important to confirm that the degree of inter-mouse variation is sufficiently small to justify collating data from different mice.

For this purpose, we made use of the long-term scaling behavior of the clone size distribution. Although, generically, the dynamics of the basal layer cells depends independently on the three adjustable parameters, $\rho$, $\lambda$ and $r$, in the scaling limit, the basal layer clone size depends only on the only on the combination, $1/\tau = r\lambda/\rho$, a measure of the rate of progenitor cell loss and replacement. As a result, this parameter can be estimated with a smaller uncertainty, allowing it to be inferred on a mouse by mouse basis for each timepoint. Inferring this parameter from the basal layer clone size distribution, as well as the total clone size at shorter times, the comparison shown in Figs S11 and S12 reveal that, although there is some degree of inter-mouse variation, the amalgamation of data from different mice at the same timepoint can be justified.

Using the full clone data (Figs 2B, 3B, S6B and C) we computed a posterior distribution on $\rho$, $\lambda$ and $r$. This distribution is narrow and Gaussian-like, so we can describe it using its first two moments. Specifically, the mean is a point estimate for the parameters, and twice the standard deviation is an estimate for the $95\%$ credible interval; the results are shown in Fig. 2C. We show a comparison between the fitted parameters and the detailed basal and suprabasal cell number distribution in Figs 3B, S6B and C. We used the point estimates to computed a set of probabilities, with which we can compute $95\%$ likelihood intervals for $f_{bn}$ knowing the total number of clones observed (which we have called plausible intervals due to a lack of standard nomenclature); these are plotted as the error bars in the model, and the graphs form a visual significance test with the null hypothesis being the model with the parameters as estimated. We draw attention to the fact that these do not contain the uncertainties in the parameters themselves. Furthermore, Fig. S6D contains a detailed comparison with the much larger basal clone size data set, which shows the expected deviations from scaling at shorter times; again the error regions refer to the sampling error only. Lastly, in Fig. 2Ba and 2Bb we show the comparisons with density and average size of basal clones; the error bars in are the mouse to mouse variation. Given that we can fit the entire distributions at all times the average

size also fits (unsurprisingly). However it should be noted that we can make a prediction of the density which fits within the experimental error, with only one parameter, the induction efficiency, estimated by the labeling frequency at time zero (Fig. 2Ba).

# 4 Clonal analysis: challenging the model

## 4.1 Effects of atRA treatment

It is straightforward to implement the same inference algorithm to address the atRA treated tissue. Wit the assumption that atRA pre-treatment establishes a new steady-state EP cell dynamics, the resulting fit of the model to the experimental data is shown in Fig. 3D, while the resulting model predictions are shown alongside the experimental data in Fig. 3C. Once again, the fits reveal a close agreement of the model with the experimental data.

With the parameters for normal and atRA treated tissue in hand, it is then possible to predict the outcome when atRA is applied after induction. In particular, if we assume that, following atRA treatment, the division and stratification rates immediately adjust from their normal to atRA treated steady-state values, we can predict the resulting clone dynamics. Indeed, when compared to the measured basal clone distribution, we find that the predictions offer a favourable fit (Fig. S7C). By contrast, comparison with the joint suprabasal and basal distribution shows significant deviations (Fig. S7D). This departure can be easily understood as the effect of atRA treatment on cells which have already stratified and left the basal layer is more complex, and there may be some time-lag before the entire tissue changes to the new steady-state behavior. Nevertheless, these results suggests that, as far as the basal layer is concerned, the application of atRA treatment can be well-approximated as an instantaneous change of the parameters of the EP cell dynamics.

## 4.2 Single cell clones

Finally, the development of the EP model relied upon the observation of long-term scaling behavior of the basal layer clone size distribution which suggested that tissue maintenance involves only a single equipotent progenitor cell population. However, by focusing on clones with a total size greater than one, we eliminated the potential signature of a second very slow-cycling or quiescent cell population. Although the existence of such a population in EE was ruled out by the H2B-GFP assay, with the predictions of the EP model, we can do back and question whether the dynamics of the single-cell clones are consistent with the ansatz of the modeling scheme.

Since, for times in greatly in excess of the cell stratification time, $1/\gamma$, differentiated cells labeled at induction will have been lost from the basal layer, we can use the long-term evolution of the single-cell basal clones to challenge the predicted model dynamics. Once again, the results, shown in Fig. S13, reveal an excellent agreement between theory and experiment over the year long timecourse. In particular, we can conclude that the persistence of single-cell clones merely reflects the set of clones which, by chance, have expanded and the sister cells have been shed leaving behind a single labeled cell.

# 5 Conclusion

In summary, motivated by parallel studies of tissue turnover in interfollicular epidermis, we have shown that the range of clonal fate data is consistent with a simple model dynamics in which a single population of epidermal progenitors give rise to differentiated cells that stratify and become shed at the surface of tissue. In this model, EP cell division results in balanced stochastic fate where, on average, following division, one cell remains in cycle and the other commits to differentiation. The model is fully-defined by just four adjustable parameters: the

EP cell division rate $\lambda$, the ratio of symmetric to asymmetric cell divisions controlled by the parameter $r$, the stratification rate $\gamma$, and the loss rate of cells from tissue $\mu$. Indeed, the latter parameter is constrained and therefore fixed by the ratio of basal to suprabasal cells, which is easily and accurately measured. In the long-term scaling regime, the dynamics is fully-specified by a single parameter reflecting the ratio of the remaining three adjustable parameters.

By adopting a Bayesian approach, we have shown that this simple model provides an excellent agreement with the range of complex clonal fate data (covering both basal and suprabasal cells) engaging more than 140 independent data points (Figs 3B, S6B, C and D). Moreover, we have shown that this model can capture the effect of atRA treatment through a straightforward adjustment of the average EP cell division rate. The predictive power of the model has also been tested both through the clone survival curve (Fig. 2B) as well as a second schedule of atRA treatment.