Supplementary Theory

In this Supplementary Theory, we provide further details on the modelling approach to study the cellular dynamics of sebaceous gland (SG) morphogenesis, homeostasis, and oncogenic activation. We also provide details on the statistical tests and approaches used to challenge the model versus the dataset.

1 Morphogenesis of the sebaceous gland

As shown in the main text, the SG develops between P2 and P7 from a pool of committed precursors. To address the cellular fate choices and lineage hierarchy involved in this process, we implemented a quantitative clonal lineage tracing strategy inspired by previous studies [1, 2, 3]. Using inducible genetic labelling, based on the Lrig1 promoter, we kept track of the geographical location of the resulting clones, both in terms of hair follicle compartment, but also in a subcompartment manner (for instance, in the case of the SG, we defined three regions based on height: bottom third, middle third and top third, in addition to the SG duct, and assigned clones to each regions). Throughout this section, we use the two-tailed Mann-Whitney test as a statistical test for differences in clonal sizes, given its non-parametric nature (anticipating on the fact that the resulting distributions are non-normal due to the stochastic fate choices performed by cells).

1.1 Clonal dynamics of individual precursors

Using the tracing from P2 to P7 as an indication of the fate choices made by individual cells during morphogenesis, we first wished to test for geographical biases in clone sizes. We thus compared the basal clone sizes in the top and bottom compartments at P7 ($n = 13$ and $n' = 26$, resp.), and found no statistical differences ($P = 0.83$, Mann-Whitney test), arguing against a spatial gradients in morphogenetic potential. This was also true for suprabasal clone sizes ($P = 0.98$, Mann-Whitney test).

Moreover, clone size distributions were consistently broad, arguing against a deterministic and highly regulated developmental sequence. As stochastic fate choices have been observed in other epithelial contexts, we thus sought the simplest possible model,

considering a single equipotent precursor cell (P) population that undergo stochastic fate choice between division and differentiation into suprabasal non-dividing differentiated cells (D). Here, we make the assumption that differentiated basal cells migrate rapidly into the suprabasal layer, so that the vast majority of basal cells are dividing progenitors. We define λ as the division rate of progenitors. In this paradigm, progenitor fate is defined by the Markovian process,

$$P \xrightarrow{\lambda} \begin{cases} P+P & \text{Proba. } 1/2+\Delta \\ D & \text{Proba. } 1/2-\Delta \end{cases}$$

in which the constant fate probabilities are defined intrinsically (i.e. cell-autonomously). We discuss in further depth the assumption of intrinsic or extrinsic cellular fates in the next sections. As we are in an out-of-homeostasis setting, the number of basal precursors P must increase during morphogenesis, which translates into an imbalance towards differentiation Δ.

Formally, defining $P_{n_P,n_D}(t)$ as the probability to find a clone at time t post-induction containing n_P basal precusors and n_D differentiated cells is described by the Master equation:

$$\frac{1}{\lambda}\frac{dP_{n_P,n_D}}{dt} = (1/2 + \Delta)(n_P - 1)P_{n_P-1,n_D} + (1/2 - \Delta)(n_P + 1)P_{n_P+1,n_D-1} \quad (1)$$

Thus, the precursor population conforms to a Galton-Watson birth-death type process, which has been studied extensively by statisticians [4]. For $\Delta = 0$, the process is critical, and starting at $t = 0$ with the initial condition $P_{n_P,n_D}(0) = \delta_{n_P,1}\delta_{n_D,0}$. Note that we neglect suprabasal cell loss here, giving the small timeframe of the observed morphogenetic timecourse. Previous studies have shown that the average number of basal progenitors in surviving clones (i.e. clones with at least one basal cell at a given time) is given by [4, 5]

$$\langle n_P(t) \rangle = 1 + \lambda t/2 \quad (2)$$

This linear increase in average size compensates for the continuous chance loss of clones through differentiation. For a critical process (balanced cell fate), the increase in average basal clone size is simply the inverse of the persisting clone fraction

$$1 - P_0(t) = \sum_{i=1}^{\infty} P_i(t) = \frac{1}{1+\lambda/2t} \quad (3)$$

so that the product of the two is constant, and equal to its initial value.

When fate becomes imbalanced towards division ($\Delta > 0$), the average clone size dependence becomes exponential. However, in both homeostatic and out-of-homeostasis

setting, the basal clone size distribution is expected to rapidly converge towards a scaling behavior, where the chance of finding a clone with n cells at time t post-labelling takes the form $P_n(t) = \frac{1}{\langle n(t)\rangle} f(\frac{n}{\langle n(t)\rangle})$, where $\langle n(t)\rangle$ denotes the average clone size, and the function $f(x) = e^{-x}$.

1.2 Fitting procedure

In order to fit the model to the clonal fate data, in the absence of any prior information on the value that the parameters might have, we made use of the classical method of least-squares, which consists in defining a sum $\mathscr{L}$ of residuals

$$\mathscr{L} = \frac{1}{n}\sum_{i=1}^{n}\left(y^i_{\text{obs}} - \widetilde{y}^i_{\text{model}}(\vec{\theta})\right)^2$$

where y^i_{obs} is a collection of n measured observables, whereas $\widetilde{y}^i_{\text{model}}(\vec{\theta})$ is a collection of predictions from the model using a set of parameters represented by the vector $\vec{\theta}$. The difference, $y^i_{\text{obs}} - \widetilde{y}^i_{\text{model}}(\vec{\theta})$, between a measurement and a prediction is called a residual, and in the case of the SG tracing from P2 to P7, the parameter vector is $\vec{\theta} = (\Delta, \lambda)$.

We then minimise $\mathscr{L}$ with respect to $\vec{\theta}$ to find the optimal parameters describing the dataset, and resort to numerical simulations of the corresponding model, for various values of the fitting parameters, $\vec{\theta}$. In each case described below, we started by performing a coarse parameter sweep in order to determine the neighbourhood of the optimal parameters, and we then performed a much finer parameter sweep to obtain the optimal fitting values. For each set of fitting parameters, or estimators, we simulated at least 50000 clones to build a model two-dimensional probability distribution $P(n_P, n_D)$ of a clone containing n_P basal progenitor cell and n_D suprabasal differentiated cells. However, the dataset was too sparse to sample accurately a 2D probability distribution (for instance at P7, the dataset contained 44 clones ranging until $n_P^{max} = 11$ and $n_D^{max} = 4$. Therefore, we decided to fit instead the 1D probabilities $P(n_P)$ and $P(n_D)$ for a clone to contain resp. n_P and n_D cells. Thus, the observables y^i_{obs} were values of these two distributions (ranging until n_P^{max} and n_D^{max} so that the dataset had $n = n_P^{max} + n_D^{max} + 1$ bins to compare to the model, as $n_P = 0$ is not recorded experimentally). For each parameter sweep, we also performed several replicates to verify that we always converged on the same parameter set.

Once the optimal set of parameters was determined, we complemented the analysis with a confidence interval for each of these parameters, i.e. the confidence that the real parameter values fall within the provided interval. In order to get a confidence interval in a non-parametric fashion, we made use of a bootstrapping method, using a Monte Carlo algorithm for case resampling [6]. This method involves performing the same parameter sweep as the one described before, but for a large number of artificial datasets, which are

samples with replacement of the original dataset. For instance, if 100 clones have been counted for a given data point, we randomly picked 100 clones within these (allowing for replacement, so that the same clone can be picked several times), and determined the optimal parameters for that dataset.

We performed this operation for at least 500 iterations, and thereby built an estimate for the probability distribution of our estimators themselves. Once this distribution had been calculated, it was straightforward to deduce a confidence interval for any confidence value of choice. We also plot confidence intervals (shaded areas on the graphs such as Figure 2d) on each prediction of the model, in order to visualise how the error bars in the fitting parameters extracted influence this prediction. In the remainder of the text, confidence intervals will be stated at the 95% confidence level. In all cases, we performed our parameter fitting on the joint distributions of the basal and suprabasal clone sizes.

1.3 Results

Using the method described above on the P7 clone size distributions, we found that the best fits (Figure 2d,e) occurred for

$$\Delta = 0.29^{+0.06}_{-0.1}, \qquad \lambda = 0.4^{+0.05}_{-0.08} d^{-1}$$

which translates into a cell fate choice occurring on average every 1.7 days, and resulting in a symmetric division 79% of the cases, and a differentiation in 21% of the cases. This resulted in a minimum sum of residuals $\mathscr{L}_{min} = 0.0037$.

1.4 Alternative models

In this subsection, we wish to strengthen the validity of our modelling approach by considering alternatives to our model, which can be seen as the simplest option.

Basal differentiated cells As mentioned above, the wide distribution of clone sizes excludes a deterministic model of stem cell fate. However, one could envision a more complex mechanism, for instance as found in oesophagus, where the basal layer consists of both basal precursors P, and basal differentiated cells D_1, on route towards suprabasal stratification D_2 i.e. following the rules:

$$P \overset{\lambda}{\to} \begin{cases} P+P & \text{Pr. } 1/2+\Delta \\ D_1 & \text{Pr. } 1/2-\Delta \end{cases} \qquad D_1 \overset{\Gamma}{\to} D_2$$

This adds a free parameter to the analysis, and would suggest that the basal distribution that we measure consists of $P + D_1$ cells. We thus performed a similar fitting procedure, considering the $\vec{\theta} = (\Delta, \lambda, \Gamma)$ parameter vector instead. One should note that in the limit $\Gamma \gg \lambda$, this model reduces to the previous one (D_1 cells have a very short lifespan). Bounding the possible values of Γ between 0 and 5λ gave an optimal value of the fit for Γ at the upper bound (5λ), arguing that there are very few, if any, basally differentiated cells. When broadening even more the phase space of parameters Γ (we here proceeded in steps of $1/\Gamma$, which are more appropriate to look for very large values of Γ), we found an optimal value for the fit of $\Gamma = 3.1 d^{-1} \approx 8\lambda$. However, the addition of this parameter only infinitesimally lowered the value of the residuals $\mathscr{L}_{min} = 0.0037$ (indeed, the value of $\mathscr{L}_{min}$ became near constant in this region of large Γ, reflecting its lack of influence on the dynamics), so that we could reject the likelihood for a significant fraction of basally differentiated cells (see AIC criterion from section below).

Refractory period after division Next, we sought to incorporate a potential refractory period after division into our model, to check whether this would improve the fit, and in particular, whether this would explain the small deviation from exponential distributions that we see experimentally (i.e. the fact that we observe less small clones than what an exponential distribution would predict - see Figure 2d and Figure S2f). We thus came back to the previous model of a single population of basal (dividing) cells, but incorporated a time T_f post-division below which no subsequent fate choice can occur. Interestingly, and contrary to the model above, this provided a marked improvements of the fits: $\mathscr{L}_{min} = 0.0021$, thus halving the sum of residuals, as this introduced small time-correlations that improved the fits compared to exponential distributions (Figure S2f). This was achieved for best fit parameters of:

$$\Delta = 0.28^{+0.07}_{-0.13}, \qquad \lambda = 0.65^{+0.10}_{-0.2} d^{-1}, \qquad T_f = 1.1^{+0.2}_{-0.05} d$$

To compare this with the previous paradigm, we calculated Akaike's information criteria (AIC) for both models, with the classical formula: $AIC = 2k - nln(\mathscr{L}_{min})$, where k is the number of inferred parameters (resp. 2 and 3 in both models). This gave $AIC_1 = -41.2$ and $AIC_2 = -48.3$ respectively, meaning that the first model without refractory periods was 0.029 times less probable than the model with refractory periods.

One should note that this second model, apart from the rather large refractory period, still yields a qualitatively similar paradigm to the one before, with in particular comparable division rates and imbalances. The effective division time would indeed read $\tau = T_f + ln(2)/\lambda \approx 2.1$ days. In Ref. [7], a refractory period of 3 days was found (for a division time around 4 days), showing that refractory periods can be relatively large *in vivo*. Importantly, this only introduces short-term correlations in the dataset, without changing significantly the key scaling features of longer-term tracings.

1.5 Coordination of cellular fate choices for SG size regulation

So far, we have described fate choices during SG morphogenesis as a zero-dimension, cell-intrinsic process, where each cell behaved stochastically, without sensing the finite size of the SG. This is partially justified by the fact that clones at P7 are small compared to total SG size (making up on average 9% of a given SG in terms of total cell numbers), so that they do not feel the boundary conditions associated with longer-term monoclonal conversion. However, this still begs the question of how final SG size is regulated during morphogenesis.

A first, more intuitive, hypothesis would be that cell fate, although stochastic, is not intrinsic, as assumed above, but instead depends on the interactions with neighbours and/or a niche, providing spatial information that can regulate SG size correctly. Recently, a study uncovered a mechanism of this class underlying the robustness of fate allocation during gut epithelial homeostasis [8], involving Delta-Notch interactions between neighbours. However, such cell-extrinsic classes of mechanisms would yield the same key predictions as the cell intrinsic model in terms of clone size distribution, making it hard to address from the lineage tracing data alone.

A second class of models would be that all SG precursors cells make independent and stochastic fate choices, but that a temporal signal at P7 shifts their dynamics collectively towards balance. Such "temporal" models would predict a much broader distribution of final SG sizes, as the stochasticity arising from the fate choices of uncorrelated precursors would be "quenched" at P7, without the possibility of corrections. Given that around 11 precursors at P2 give rise to the final SG at P7 (Figure 2j,k and Figure S2d), and that each gives rise to a roughly exponential clone size distribution, one would then predict the SG size follows a broad Gamma distribution.

Strikingly, when simulating 11 independent precursors behaving according to the rules inferred above, we could indeed predict the SG size distribution with high accuracy, in the absence of adjustable parameters (Figure 2l). This strongly argues for the second "temporal" class of model, and for the absence of strong fate coordination between different precursors during the growth phase. As detailed in the main text, ECM rigidity, which increases strongly during the first week of development (Figure 4), could contribute to provide such an external mechanical cue.

1.6 Validating the model

Finally, in order to test the predictions of the model further, we performed H2B experiments which measure how fast nuclear H2B gets diluted in basal cells, as a measure of the cell division rate. Interestingly, assessing dilution between P2 and either P4, P6 and P8, we found an exponential decay with nearly constant slope, and verified that this decay was reproduced by a characteristic division rate of $T_{H2B} = 1.3$ days, close to our inferred

value from the tracing (Figure 2i).

2 Homeostasis of the sebaceous gland

In the following section, we discuss the post-P7 homeostatic cellular dynamics in the SG. As SG size does not evolve after this time point, we can assume that the basal precursors must make fate choices which are balanced, either at the cell or population levels.

2.1 Clonal dynamics of individual precursors

We first make use of the same tracing as the previous section, i.e. induction at P2, but examining longer chase time. To comprehensively study the long-term dynamics of cell fate, we collected samples up to a year: at P23, P56, P90 and P365.

A first possibility for homeostatic regulation would be that cell fate would be balanced at the cellular level through intrinsically asymmetric divisions ($P \rightarrow P + D$), which would translate into clone sizes constant in time, post-P7. However, the data indicated an opposite situation of slow but steady increase in clone sizes, up to the point at 1 year where all SGs were either devoid of clones, or monoclonally converted. This confirmed previous findings using an Lgr6 promoter in the SG [10], and echoes findings in the intestinal crypt [9], indicating that basal precursors compete via symmetric divisions.

In analogy to the section above on SG morphogenesis, we then turned to detailed modelling of individual SG basal cell fate. A key difference with the previous section, as well as with similar models in the interfollicular epidermis [2] or oesophagus [3] is that one must take into account the finite size of the SG compartments at these time scales. One should note that previous modelling of the SG did not take this into account, as they were starting the tracing already at homeostasis and using 22 weeks as latest time point [10], so that the clones were never large enough for finite size effects to strongly influence the dynamics. Importantly indeed, examining our P7 and P23 basal clone size distribution, prior to dominant monoclonal conversion, revealed the expected scaling and exponential dependency (Figure 3m).

However, for the P56, P90 and P365 time points, a significant portions of the clones (all clones at P365) were spanning entire glands, so that not only the finite size of SGs, but also the size distribution of SG themselves, had to be taken into account. For that, we made use of the findings from the first section: we simulated SG size distributions by considering 11 independent precursors at P2 (with one being labelled), with Galton-Watson dynamics.

For each simulation, at P7, we "freeze" the SG total size, and build a two-dimensional rectangular lattice of cells, with dimensions NxM such that the total basal cell number

approximates optimally the number calculated from the P2-P7 simulation of the 11 precursors. We then position the P7 clone (according to its size in this given simulation) on the lattice, and let the simulation evolve from P7 to P365, according to loss-replacement kinetics with division rate λ_h (see Figure S3i,j for a model schematics). Post-P7, we ran the simulation according to classical loss-replacement dynamics: each basal cell can divide symmetrically, with rate λ_h, causing another basal cell to migration suprabasally, with periodic boundary conditions. Given the fact that suprabasal cells make up half the number of basal cells, we assume that suprabasal cells are organised in a Mx$(N/2)$ grid, and that any basal cell differentiation causes the suprabasal cell above to be lost (see Figure S3j for a schematic).

One should therefore note that this model has few adjustable parameters: the division rate λ_h, and the length scale for loss/replacement dynamics, i.e. when a given cell divides, where the corresponding lost cell is located. There are two limits to this: nearest neighbour loss-replacement, where a dividing cell expels one of its direct neighbours, and global competition, where a dividing cell expels any other cell in the tissue. These two scenarii only differ mildly from on another, as we show on Figure S3q. This is consistent with previous theoretical analyses of lineage-tracing datasets[5, 11]. Indeed, although cell-intrinsic and cell-extrinsic models of stem cell renewal do give very different predictions for one-dimensional tissues, notably as a clone can only grow on its boundary in cell-extrinsic models (where the former predicts exponential clone size distributions and linear clone size growth in time, while the latter predicts gaussian clone size distributions and clone size growing as square root of time), both models make very similar predictions for two-dimensional tissues (and rigorously undistinguishable in three-dimensions). In two-dimensions, both models are expected to give rise to linear (or quasi-linear) monoclonal conversion, as well as exponential clone size distributions for intermediate time scales, explaining the similarity of outcomes displayed on Figure S3q. For the same reason, we found that positioning clones in the simulation at P7 on the NxM grid either via random position picking for each cell, or as a coherent clonal structure (isotropic or anisotropic), to give very similar outcomes (one should note that in the limit of global competition, relative positions of cells do not matter, and the way of initialising the clone at P7 becomes rigorously irrelevant). As detailed below, global competition yields a slightly more rapid monoclonal conversion, fitting better with the data, so that we choose this model in the following.

2.2 Fit of the homeostatic division rate

Similarly to the previous section on morphogenesis, one could fit the division rate versus the time evolution of the basal and suprabasal clones. However, to constrain better the dynamics, we adopt here the complimentary approach of first measuring independently the division rate, in order to make predictions on the clonal evolution subsequently.

For this, we make use of an H2B dilution experiments, as detailed in the main text (starting at P49 and traced for two weeks). We simulate the clonal dynamics detailed above, and perform again a least-square fit analysis with bootstrapping to determine the homeostatic division rate λ_h.

This leads the best fit for values of

$$\lambda_h = 0.2^{+0.02}_{-0.02} d^{-1}$$

which is equivalent to a division every 3.5 days, close to the value observed for the interfollicular epidermis in mouse [2]. On Figure 3j, we show the data, together with the model predictions (thick and thin lines indicate resp. the best fit and confidence intervals). Interestingly, in homeostatic tracings from adulthood such as the one performed by Ref. [10], the SG basal clone size for intermediate time is expected to grow linearly with slope $r\lambda_h$, where $1-2r$ quantifies the possibility of intrinsically asymmetric divisions. Using the first 40 days on this tracing (to remain at intermediate time scales where growth is linear), together with our division rate of $\lambda_h = 0.2d^{-1}$ would give a value of r slightly above 0.5, arguing again in the homeostasis setting for the absence of intrinsically asymmetric divisions, a point we will come back to below using our own dataset. One should also note that incorporating the small, experimentally measured division rate gradient along the top-bottom axis of the SG (Figure S3o) in the model (by assuming a linear 25% gradient in division rates) did not change significantly either the clone size distributions or the average clone size evolution (Figure S3q).

2.3 Results and model validation

Firstly, we wish to start with predictions which are independent on the exact details of either the division rate, or the nature of loss-replacement. For this, we turn to the clone size distributions at all time points (P7, P23, P56, P90 and P365), and compare them to the model predictions by matching each experimental data point not the corresponding simulated time (as was done in Figure 3m), but to simulated time points that best fit the average clone size at each time point (Figure S3p). Effectively, this enables us to test how good the predicted shapes of the distribution match the data, after having fitted the averages. Importantly, this provided consistently good predictions: as mentioned above, P7 and P23 distributions are close to exponential as predicted by the model, whereas P56 and P90 display a first exponential part for small clone sizes (which do not feel the boundary condition from the finite size of the SG), followed by a tail which matches the SG size distribution for large clone sizes. For P365, the clone size distribution trivially matches the SG size distribution, hence the goodness of the fit is not surprising. This is a strong indication of SG homeostasis being maintained by a single functional progenitor population.

Secondly, we wish to test whether the model can recapitulate the evolution of basal clone sizes throughout the 1-year tracing. Assuming that the division time instantly drops from every 1.7 days to 3.5 days at P7 was found to under-estimate the clone size evolution. However, modifying slightly the model, by incorporating a monotonous decrease of the division rate between the two values for 4 weeks post-P7 (i.e. during the growth phase of the rest of the epidermis) drastically improved the fit for the basal clone size, as shown on Figure 3k. One should note that an alternative explanation for monoclonal conversion occurring slightly faster than expected from the adult division rate could be that the division rate varies in time, and for instance with the hair cycle, whereas we have assumed it to be constant within a whole year. This would need to be addressed via live-imaging methods.

Thirdly, we found the model provided an excellent fit for the full time course of the average suprabasal clone size (Figure 3k, red), which was more surprising, as we haven't made any complex assumption on the suprabasal migration and loss rate. This argues in favour of a theory where suprabasal cell number, possibly due to finite size constraints, are capped and slave to the dynamics of the basal dynamics. Furthermore, the theory also provided a good fit for the evolution of the clonal persistence, as shown on Figure 3n, which was characterised as expected by a first phase of stark decay, followed by a plateau arising from the monoclonal conversion.

Fourthly, to challenge the model further, we performed additional short-term tracing experiments from P56 to P65, using both Lrig1 and K14 promoters to enhance the generality of the findings (Figure 3o). We first tested whether the distribution of basal clone sizes were the same in both cases ($n = 46$ clones for K14, $n = 23$ clones for Lrig1), and found no statistically significant differences ($P = 0.35$, two-tailed Mann-Whitney test), providing an additional test of our theory of a single equipotent population maintaining the SG. One should note that both the mean clone size and exponential distribution are also consistent with the Lgr6 tracing from Ref. [10]. Importantly, both distributions (Figure 3q) are consistent with our model predictions (with the homeostatic division rate $\lambda_h = 0.2d^{-1}$ determined above). This provides additional confidence in our model.

3 Oncogenic activation in the SG

In a third section, we discuss the effect of oncogenic KRAS activation at different time points in the SG, both at the cellular and gland level.

3.1 Clonal dynamics post P7

We then performed KRAS activation combined with short-term (9 days) lineage tracing during adulthood (at P56, to match the adult homeostatic tracing from the previous

section), to follow its effect on clonal dynamics.

We then performed the same least-square fitting procedure as in the previous sections, with the parameter vector $\vec{\theta} = (\Delta, \lambda)$ (again, we neglected suprabasal cell loss at these short time scales), and found that the best fit (Figure 5k,m,p and Figure S4o,p) occurred for values of:

$$\Delta = 0.23^{+0.1}_{-0.1}, \qquad \lambda = 0.22^{+0.05}_{-0.05} d^{-1}$$

Interestingly, this thus revealed that the cellular division rate was only weakly affected by adult KRAS activation, and that clone size increase proceeded via an imbalance towards symmetric division. This results in a minimum sum of residuals $\mathscr{L}_{min} = 0.002$. It is worth noting that the level of imbalance inferred here is rather close to the one inferred in the WT sebaceous gland during morphogenesis (within error bars), which could suggest that this represents an attractor of the system in the absence of strong finite-size constraints.

3.2 Clonal dynamics during morphogenesis

We then performed KRAS activation combined with lineage tracing from P2, to follow its effect in the phase of morphogenesis.

Tracing until P7 At P7, the effect was relatively mild, with roughly 60% increase in basal clone size (Figure 5i,m). Here, as in the entire KRAS section, we assume, based on the literature, that KRAS recombination is much more efficient than Confetti recombination, so that the overwhelming majority of cells in the SG are KRAS-activated, whereas only a small clonal subset are labelled by Confetti. Thus, we model this dynamics as a uniform pool of KRAS cells, as opposed to a mixture of competing KRAS and wild-type (WT) cells.

This is in fact an assumption that we can experimentally challenge, as it makes two key testable predictions:

- one should not see an "anomalous" early drop in persistence - which would be associated with outcompeted WT cells.
- the clone size distributions should still display the same scaling as in wild-type situation, as opposed to the more complex bimodal distributions one would observe for mixtures of two cell populations.

As we will see in the subsection below, these predictions hold true in the KRAS dataset (Figure 5l,m), validating the assumption of very efficient KRAS recombination. One cannot rule out the possibility of WT cells being expelled and lost so rapidly between P2 and P7 that they cannot be observed. However, this would still mean that our dataset

consists of the dynamics of the remaining KRAS cells. An intriguing alternative would be one where some WT cells remain within the SG, but behave exactly like KRAS neighbours (for instance if KRAS cells eliminate the niche constrain for them as well, and if that would be sufficient to convert WT cells to the imbalanced level seen in KRAS cells). Ruling out this fascinating possibility is not possible within our modelling scheme, and would require live-imaging datasets and/or experimental tools to assess the recombination status of SG cells.

We then perform the same least-square fitting procedure as in the WT morphogenesis section, with the same parameter vector $\vec{\theta} = (\Delta, \lambda)$, and found that the best fit (Figure 5k,m) occurred for values of:

$$\Delta = 0.25^{+0.05}_{-0.05}, \qquad \lambda = 0.62^{+0.1}_{-0.12} d^{-1}$$

which corresponds to a very similar value for the imbalance compared to wild-type, but a slightly accelerated division time, of on average one division every 1.1 day. A key signature for the fact that the division rate is the most affected here comes from the fact that both basal and suprabasal clone sizes increase for the KRAS mutant, in similar proportions. If the increase in basal clone size for KRAS arises from larger Δ, this would result, on the short-term (P7), in smaller suprabasal clone sizes. Thus, contrary to the adult KRAS activation, neonatal KRAS activation seems to act rather via an increase of the division rate. Note that as for the homeostasis tracing, this type of lineage-tracing based on fixed samples does not allow us to conclude on the cell-intrinsic or cell-extrinsic nature of cell fate regulation upon KRAS activation.

However, as imbalance is conserved compared to WT morphogenesis and adult KRAS activation (Figure 5p), it is tempting to speculate that this imbalance represents a constraint of the system, i.e. that the system cannot go above it during morphogenesis as it is already at the upper bound. Understanding how these regulated fate choices are implemented molecularly by a given cell (in the absence of finite size constraint) would be a next key question arising from this analysis.

Longer tracing Finally, we analysed the results of tracing KRAS cells from P2 to longer time points (where the wild-type SG should be homeostatic). Unlike shorter time points, one cannot fit independently both Δ and λ in this setting from the clone sizes or clone sizes distribution, as the clone size distributions are expected to converge towards exponentials (a feature well-recapitulated in the data, see Figure 5m), whereas the basal clone size is expected to grow exponentially as $e^{2\Delta\lambda t}$, making only the product of the two parameters accessible from a fitting procedure. Importantly, the scaling behavior of clone size distributions at longer time points demonstrates that the KRAS data can be explained by a model of a single population of imbalanced progenitors on the long term (although short-term priming and differences among KRAS progenitor could occur). Note that we

do observe small deviations to the exponential scaling at both P7 and P23, in analogy to Section 1 on morphogenesis, but this can again be explained in a straightforward manner by introducing a small refractory period following cell division (see Figure S4o,p for best-fit models of the P7 and P23 basal clone size distribution with the addition of a 24h refractory period). We note however that introducing the analysis of a fraction of cells remaining wild-type in the KRAS tracing systematically worsened the fit (Figure S4q), by introducing an early drop at small clone sizes in the clone size distributions, not consistent with the observed rarity of small clones (i.e. of a size comparable to wild-type at the same time point) in the data.

Morever, given the findings above, we can make the simplifying assumption that KRAS cells adopt an homeostatic value for their division rate post-P7 ($\lambda_h = 0.2d^{-}1$). From the P23 average basal clone size, one can then infer an imbalance post-P7 of $\Delta = 0.2$. Interestingly, running such a simulation predicts that contrary to the WT case, KRAS clonal persistence should rapidly reach a plateau (Figure 5l), as clone sizes reach values at which stochastic clonal loss becomes vanishingly small. Interestingly, this was verified in the persistence data for the KRAS tracing (Figure 5l), characterised by a constant persistence post-P7. This confirms experimentally the role of the finite-size constraint of the SG in promoting cell competition at homeostasis, and that KRAS activation liberates cells from the finite size constraint to alleviate the cell-cell competition.

References

[1] Clayton, E., Doupe, D. P., Klein, A. M., Winton, D. J., Simons, B. D., and Jones, P. H. (2007). A single type of progenitor cell maintains normal epidermis. Nature, 446(7132), 185-189.

[2] Sanchez-Danes, A., Hannezo, E., Larsimont, J.C., Liagre, M., Youssef, K.K., Simons, B.D. and Blanpain, C., 2016. Defining the clonal dynamics leading to mouse skin tumour initiation. Nature, 536(7616), pp.298-303.

[3] Doupe, D.P., Alcolea, M.P., Roshan, A., Zhang, G., Klein, A.M., Simons, B.D. and Jones, P.H., 2012. A single progenitor population switches behavior to maintain and repair esophageal epithelium. Science, 337(6098), pp.1091-1093.

[4] Bailey, N. T. J. The Elements of Stochastic Processes, J. Wiley and Sons, New York, 1964, Chap. 8.6.

[5] Klein, A.M., Doupe, D.P., Jones, P.H. and Simons, B.D., (2007). Kinetics of cell division in epidermal maintenance. Physical Review E, 76(2), p.021910.

[6] Manly, B.F.J. (2006). Randomization, bootstrap and Monte Carlo methods in biology. Vol. 70. CRC Press.

[7] Rompolas, P., Mesa, K.R., Kawaguchi, K., Park, S., Gonzalez, D., Brown, S., Boucher, J., Klein, A.M. and Greco, V., 2016. Spatiotemporal coordination of stem cell commitment during epidermal homeostasis. Science, 352(6292), pp.1471-1474.

[8] Toth, B., Ben-Moshe, S., Gavish, A., Barkai, N. and Itzkovitz, S., 2017. Early commitment and robust differentiation in colonic crypts. Molecular systems biology, 13(1), p.902.

[9] Lopez-Garcia, C., Klein, A.M., Simons, B.D. and Winton, D.J., 2010. Intestinal stem cell replacement follows a pattern of neutral drift. Science, 330(6005), pp.822-825.

[10] Fullgrabe, A., Joost, S., Are, A., Jacob, T., Sivan, U., Haegebarth, A., Linnarsson, S., Simons, B.D., Clevers, H., Toftgard, R. and Kasper, M., 2015. Dynamics of Lgr6+ progenitor cells in the hair follicle, sebaceous gland, and interfollicular epidermis. Stem cell reports, 5(5), pp.843-855.

[11] Klein A. and Simons B.D., 2011, Universal patterns of stem cell fate in cycling adult tissues. Development 138, 3103-3111