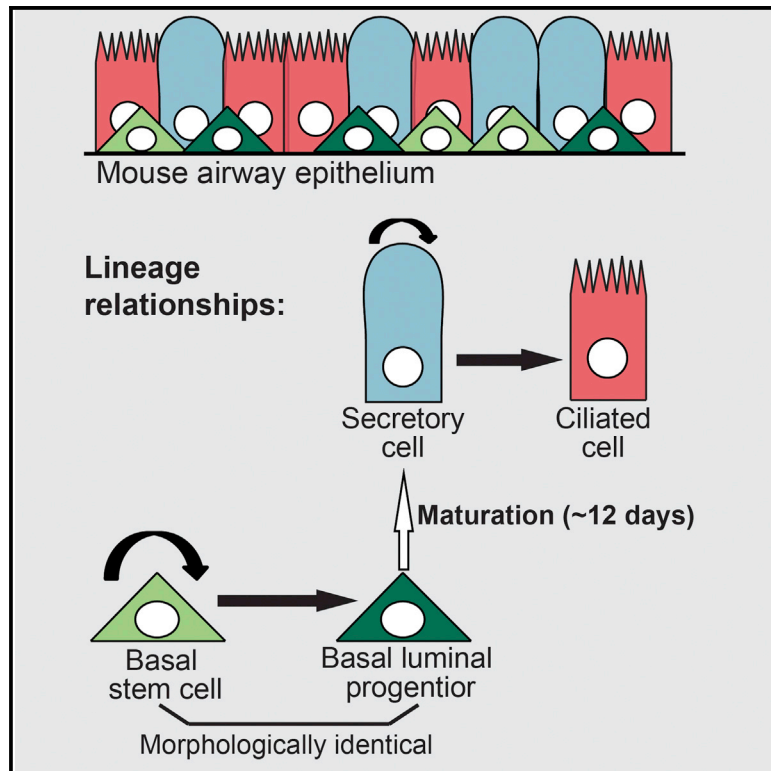


Clonal Dynamics Reveal Two Distinct Populations of Basal Cells in Slow-Turnover Airway Epithelium

Graphical Abstract



Authors

Julie K. Watson, Steffen Rulands, Adam C. Wilkinson, ..., Cédric Blanpain, Benjamin D. Simons, Emma L. Rawlins

Correspondence

e.rawlins@gurdon.cam.ac.uk

In Brief

Using clonal analysis, mathematical modeling, and single-cell qRT-PCR, Watson et al. define the homeostatic tracheal epithelial lineage. The epithelium contains two major, equally distributed subpopulations of basal cells: stem cells and long-lived precursors that are already committed to differentiation.

Highlights

- Clonal analysis for determining homeostatic tracheal epithelial cell hierarchy
- Basal cells comprise two subpopulations: stem cells and luminal precursors
- Luminal secretory cells are short-lived and self-renewing
- Secretory cells are the major steady-state source of new ciliated cells



Clonal Dynamics Reveal Two Distinct Populations of Basal Cells in Slow-Turnover Airway Epithelium

Julie K. Watson,^{1,2,3} Steffen Rulands,^{1,2,4} Adam C. Wilkinson,^{2,5,6} Aline Wuidart,⁷ Marielle Ousset,⁷ Alexandra Van Keymeulen,⁷ Berthold Göttgens,^{2,5,6} Cédric Blanpain,^{7,8} Benjamin D. Simons,^{1,2,4} and Emma L. Rawlins^{1,2,3,*}

¹Wellcome Trust/CRUK Gurdon Institute, University of Cambridge, Cambridge CB2 1QN, UK

²Wellcome Trust-MRC Stem Cell Institute University of Cambridge, Cambridge CB2 3EG, UK

³Department of Pathology, University of Cambridge, Cambridge CB2 3EG, UK

⁴Cavendish Laboratory, Department of Physics, J. J. Thomson Avenue, Cambridge CB3 0HE, UK

⁵Department of Haematology, University of Cambridge, Hills Road, Cambridge CB2 0XY, UK

⁶Cambridge Institute for Medical Research, University of Cambridge, Hills Road, Cambridge CB2 0XY, UK

⁷Institut de Recherche Interdisciplinaire en Biologie Humaine et Moléculaire, Université Libre de Bruxelles, Brussels 1070, Belgium

⁸Welbio, Université Libre de Bruxelles, Brussels 1070, Belgium

*Correspondence: e.rawlins@gurdon.cam.ac.uk

<http://dx.doi.org/10.1016/j.celrep.2015.06.011>

This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

SUMMARY

Epithelial lineages have been studied at cellular resolution in multiple organs that turn over rapidly. However, many epithelia, including those of the lung, liver, pancreas, and prostate, turn over slowly and may be regulated differently. We investigated the mouse tracheal epithelial lineage at homeostasis by using long-term clonal analysis and mathematical modeling. This pseudostratified epithelium contains basal cells and secretory and multiciliated luminal cells. Our analysis revealed that basal cells are heterogeneous, comprising approximately equal numbers of multipotent stem cells and committed precursors, which persist in the basal layer for 11 days before differentiating to luminal fate. We confirmed the molecular and functional differences within the basal population by using single-cell qRT-PCR and further lineage labeling. Additionally, we show that self-renewal of short-lived secretory cells is a feature of homeostasis. We have thus revealed early luminal commitment of cells that are morphologically indistinguishable from stem cells.

INTRODUCTION

The mouse trachea contains three major cell types: TRP63⁺, KRT5⁺ basal cells (BCs); luminal secretory cells (SecCs, mostly Scgb1a1⁺ Club/Clara-like cells); and luminal ciliated cells (CCs) (Rock et al., 2010). Previous population-level lineage tracing using transgenic *Tg(KRT5-CreER)* mice demonstrated that BCs include self-renewing stem cells involved in tracheal growth, homeostasis (at least for up to 16 weeks), and repair (Rock et al., 2009). However, it is not known if BCs are a functionally heterogeneous population. A subset of tracheal BCs (<20%) expressing *Krt14* (*Keratin 14*) was suggested to be a unipotent self-renewing subpopulation at homeostasis (Ghosh et al., 2011).

Similar unipotent BCs have been postulated following injury and in xenografts (Engelhardt et al., 1995; Ghosh et al., 2011; Hong et al., 2004). Other repair studies described an early progenitor (EP) cell as a proliferative KRT8⁺ (luminal type cytokeratin), TRP63⁻ cell derived from BCs and controlled by Notch signaling (Paul et al., 2014; Rock et al., 2011). In development, KRT5⁺ TRP63⁻ cells with basal morphology have recently been described in germline *Notch3* mutants and in embryonic lungs deleted for *Ezh2* (Mori et al., 2015; Snitow et al., 2015), leading to the speculation that these are precursors of luminal cells. Subsequently, an independent study showed that a population of adult BCs (~12% of steady-state total), which express low levels of transcription factors usually found in more differentiated cells, are able to contribute disproportionately to regeneration following injury (Pardo-Saganta et al., 2015). However, none of these studies investigated the adult airway lineage at steady state, leaving key questions unanswered. In particular, is there is an engrained proliferative heterogeneity in the steady-state basal layer? If so, what is the lineage relationship of cells within the basal layer, and how do they connect to the luminal compartments? How do distinct subpopulations of BCs function to maintain normal homeostasis?

Within luminal cells, population lineage-labeling studies had shown that SecCs can self-renew and generate CCs, but their relative contribution to homeostasis was unclear (Rawlins et al., 2009). CCs are post-mitotic, with an average loss-rate of ~6 months in the trachea (Rawlins and Hogan, 2008; Rawlins et al., 2007). Molecular signals controlling the tracheal epithelium are being determined (Brechtuhl et al., 2011; Giangreco et al., 2012; Lu et al., 2013; Paul et al., 2014; Rock et al., 2011; Zhao et al., 2014). However, the lack of a clearly defined epithelial lineage impedes analysis of molecular function at cellular resolution. Human airways have a very similar cell lineage to mouse trachea (Engelhardt et al., 1995; Hackett et al., 2011; Hajj et al., 2007; Teixeira et al., 2013), but the limited resolution for lineage studies in human means that complementary mouse analysis is required to determine the detailed cellular hierarchy. Here, we use clonal lineage labeling, coupled with biophysical

modeling and single-cell molecular analysis, to determine the heterogeneity and functions of BCs and SecCs in the homeostatic mouse tracheal epithelium. We have rigorously obtained quantitative measures of division rates, cell-type abundance, and rates of differentiation/loss. The model that we present thus provides a new experimental and theoretical foundation for studies of airway homeostasis, injury, and disease. Moreover, we reveal an unexpected mechanism of epithelial maintenance in a slowly proliferating tissue: widespread early luminal commitment in cells that are morphologically indistinguishable from stem cells.

RESULTS

Clonal-Level Lineage Analysis of BCs in the Steady-State Tracheal Epithelium Suggests a Proliferative Hierarchy and the Presence of More Than One BC Subpopulation

To study maintenance of the tracheal epithelium, we first tested whether homeostasis was maintained during our time course by analyzing cell proliferation, composition, density, and tracheal size (Figure S1). This confirmed that the tissue was homeostatic for most of the time course, although the proportion of CCs increased by ~30%, and cell density decreased by ~30%, in older animals (1 year post-labeling) consistent with previous data (Wansleben et al., 2014). To label individual *Krt5*⁺ BCs, we used a transgenic mouse line, *Tg(KRT5-CreER)* (Rock et al., 2009), with a *Rosa26*-reporter driving membrane-targeted (farnesylated) EGFP (Rawlins et al., 2009). Exposure of adult (>8 weeks) *Tg(KRT5-CreER);Rosa26R-fGFP* mice to a single low dose of tamoxifen (tmx) resulted in scattered individual lineage-labeled BCs in the distal trachea (from the carina to six cartilage rings above on the dorsal side only; Figure 1A). Negligible labeling was detected in animals without tmx exposure (two clones of one to six cells in two out of four mice at 9–11 months age). Tracheas were harvested at intervals from 0.5 to 74 weeks post-tmx (Figure 1A) and whole-mount immunostained to determine clone size and composition by confocal microscopy (Figure 1B; Table S1). Clonal density varied between mice. However, clones were always more frequently located above the dorsal longitudinal smooth muscle, rather than the cartilage rings. As clones grew, they remained cohesive, suggesting little cell motility at steady state. Moreover, some larger clones were observed to span the muscle-cartilage junction, showing that this is not a compartment boundary.

At 0.5 weeks post-induction, clones consisted predominantly of single BCs (99% one BC; 1% two BCs; $n = 102$ clones, four mice). Rare labeled SecCs and CCs appeared at 3 and 6 weeks post-induction, respectively, and their numbers rose steadily thereafter (Figures 1B, 1F, and 1G), confirming that BCs generate luminal cells. The size distribution of clones became increasingly heterogeneous, but mean clone size increased in a remarkably linear fashion over time (Figures 1C and 1D), consistent with self-renewal of the labeled cells. From 3 weeks post-induction, clones that contained only luminal (secretory and/or ciliated) cells and no BCs emerged, showing that some labeled BCs are lost to differentiation (Figures 1E and 1F). The increasingly heterogeneous clonal composition and the emergence of clones lacking BCs indicated that BCs can divide symmetrically and asymmetrically at steady state, similar to progenitors in the in-

ter-follicular epidermis (Clayton et al., 2007; Mascré et al., 2012). To determine if any of the clonal heterogeneity could be attributed to animal-specific or regional differences in the trachea, we graphed clones separately based on sex or location (Figure S2). Although we observed some small systematic differences in the behavior of clones located over cartilage rings versus dorsal longitudinal muscle, these were statistically insignificant, and we have treated the distal-dorsal trachea as a single region (Supplemental Theory).

Strikingly, the mean number of BCs per clone rose abruptly from an average of one to approximately two cells by 6 weeks post-induction and thereafter remained remarkably constant over the following 68 weeks (Figure 1G). Moreover, most two-cell clones (88%) at 3 weeks post-induction contained precisely two BCs. This was unexpected for a population in homeostasis. By definition, the overall distribution of cell types within a tissue stays constant at homeostasis. Hence, if the transgene targets all BCs in a representative manner, the mean number of BCs per clone should remain at one. The abrupt increase in the average number of BCs per clone, and its near saturation at approximately two BCs per clone over the long-term, indicate that the lineage-labeling assay preferentially targets a subpopulation of BCs, which maintains a second initially unlabeled population (Figures 1H and 1I). We therefore postulated that *Krt5*⁺ BCs contain two discrete populations organized in a hierarchy: a multipotent basal stem cell (BSC; preferentially targeted by the assay) and an additional BC subtype.

Biophysical Modeling of the Behavior of Homeostatic Tracheal BCs

To resolve the cellular hierarchy, proliferation kinetics, and fate potential in the trachea, we used a biophysical modeling approach. We sought the simplest model that could describe the observed distributions of clone sizes and composition and provide testable predictions. To constrain the number of model parameters, we first used the *Tg(KRT5-CreER); Rosa26R-fGFP* clonal assay to infer the dynamics of the BCs alone. Independently, we employed a second lineage-labeling assay that targets SecCs to infer the dynamics of luminal cells alone. Finally, we used the basal and luminal clonal fate data from the *Tg(KRT5-CreER); Rosa26R-fGFP* experiment to challenge the predictions of the model.

Focusing on BCs alone, we investigated whether a model involving two distinct BC subtypes organized in a hierarchy could predict the complex clonal evolution observed. We proposed a model in which a self-renewing basal population (termed BSCs) can divide asymmetrically, giving rise to a BSC and a BC of a second subtype (termed “basal progenitor”), with the two cell types present in approximately equal numbers within the tissue. To account for clonal loss from the basal compartment, we further conjectured that BSCs are also capable of symmetrical cell division, resulting in two BSCs or two basal progenitors. To maintain homeostasis, these two outcomes must occur with equal probability (Figure 2A). From a fit of this stochastic model to the clonal data, we found that a BSC divides, on average, every $11 \pm$ (confidence interval: 4, 4) days. The vast majority of divisions $94 \pm$ (3, 2)% result in asymmetric fate outcome, with the remainder leading to balanced BSC

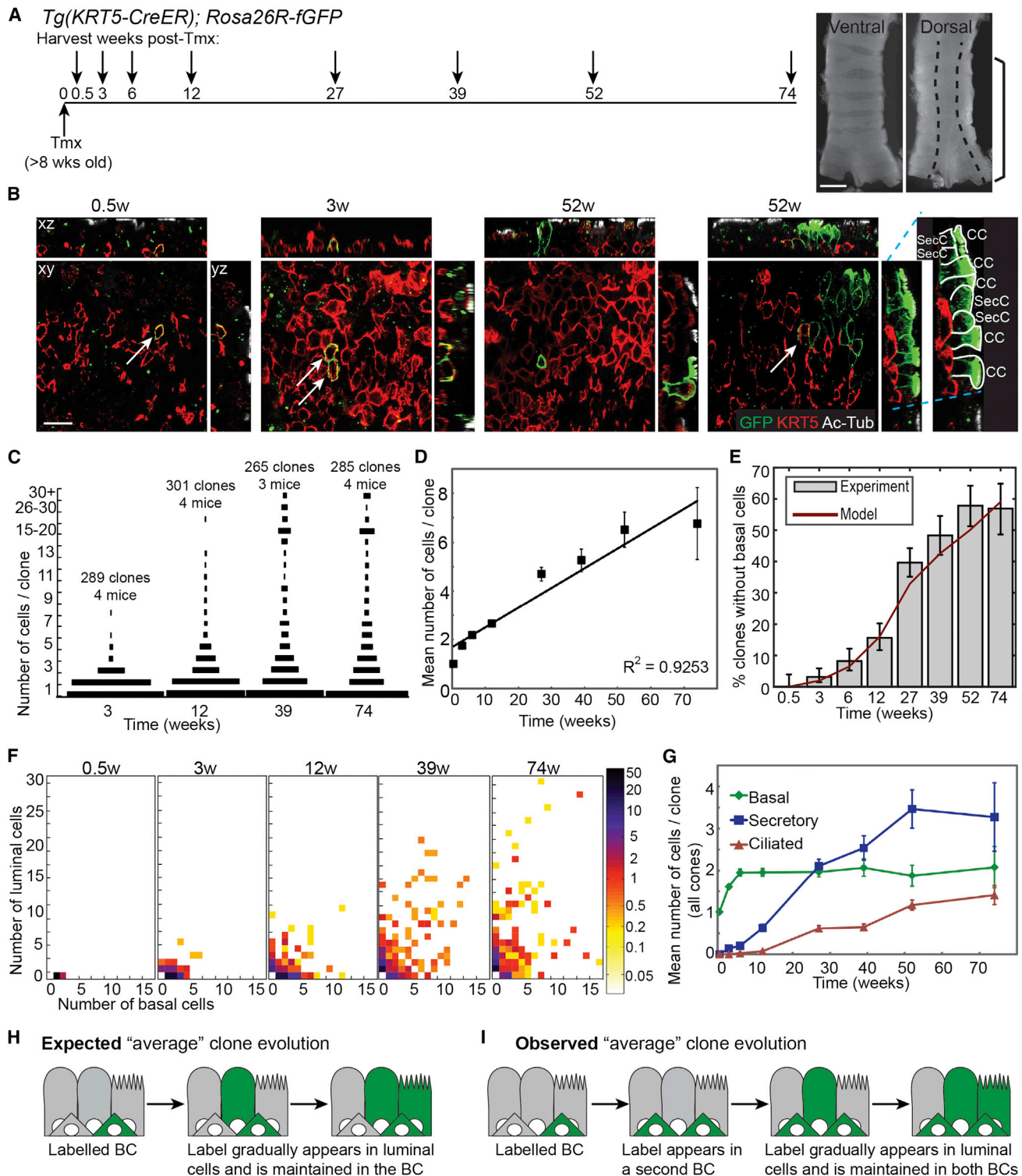


Figure 1. Clonal-Level Lineage Analysis of Basal Cells in the Steady-State Tracheal Epithelium

(A) Schematic of the *Tg(KRT5-CreER); Rosa26R-fGFP* lineage-labeling experiment. Ventral and dorsal trachea images illustrate approximate region analyzed in all experiments (brackets on dorsal side) and location of dorsal longitudinal smooth muscle (dashed line). Scale bar, 1 mm.

(B) Representative single confocal xy planes of the basal area of the epithelium with corresponding xz and yz reconstructions at 0.5, 3, and 52 weeks post-induction. Green, lineage label; red, KRT5 (BCs); white, acetylated tubulin (cilia). BCs were scored as KRT5⁺, and ciliated cells as KRT5⁻, ac-Tub⁺. Secretory cells

(legend continued on next page)

loss/replacement. Basal progenitors are lost, either to further differentiation or death, on average every $11 \pm (7, 4)$ days. With this simple model, we obtained an excellent fit to the entire range of BC clonal data (Figures 2B and 2C; Supplemental Theory).

Clonal Lineage Analysis of Luminal Cells Shows that SecCs Are Short-Lived and Preferentially Self-Renew at Homeostasis

To independently investigate luminal cell behavior, we analyzed the trachea of 6- and 12-month-old *Scgb1a1-CreER*; *Rosa26R-fGFP* mice (SecC labeling) in which there was a low rate of spontaneous recombination of the reporter in SecCs in the absence of tmx (Figures 2D and 2E). We observed scattered clones throughout the epithelium and focused on the distal-dorsal region used in our BC experiments. Strikingly, the vast majority of clones were small (compare Figures 2F and 1F), even though SecCs divide (Figure S1D), suggesting that the loss rate of SecCs exceeds their rate of self-renewal, necessitating constant replacement by the multipotent BSCs. Nevertheless, we observed labeled CCs, confirming that the SecCs do contribute to the CC lineage at steady state (Figures 2E and 2F; Table S1).

To infer the rules of lineage specification in the luminal cells, we again made use of a simple biophysical modeling scheme, whereby SecCs may divide symmetrically or asymmetrically or are lost through turnover, and CCs are post-mitotic and lost at a rate of once every 6 months (Rawlins and Hogan, 2008). Taking into account the fact that labeling occurs continuously and sporadically, we found that SecCs divide, on average, every $25 \pm (7, 5)$ days with almost all divisions ($93 \pm [2, 2]\%$) leading to symmetrical duplication. The production of SecCs through differentiation of BCs and self-duplication is balanced by a loss rate of once per $14 \pm (2, 2)$ days. From the quality of the fit to the data (Figure 2G), we can infer that, at any given time, 44% of SecCs are derived from the proliferation of a SecC and 56% from differentiation of BCs (Supplemental Theory). Thus SecCs make an important contribution to tracheal homeostasis but do not function as a traditionally defined transit-amplifying population.

Combining the Basal and Luminal Lineage Models Accurately Predicts the Range of the Full *KRT5* Clonal Dataset

With the dynamics of the basal and luminal cells defined separately, we then asked whether a combined model could predict the full range of complex clonal data in the *Tg(KRT5-CreER)*;

Rosa26R-fGFP mice. The simplest combined model is one in which basal progenitors represent cells committed to a luminal (depicted as secretory) fate (called basal luminal precursors [BLPs]) (Figure 2H). Indeed, if these cells differentiate directly to SecCs without division, the combined model accurately predicts the overall detailed clonal variation throughout the entire long-term time course (Figure 3A). Production of new CCs happens rarely at homeostasis and, within the confidence limits of our model, can be entirely accounted for by division of SecCs. However, the resolving power of our clonal analysis is limited for rare events and the direct steady-state production of CCs from the BLPs cannot be altogether ruled out.

Are the BLPs a transit-amplifying population? To meet the traditional definition of a transit-amplifying cell population (Watt and Hogan, 2000), the BLPs would need to self-renew symmetrically at a greater rate than the stem cell in order to increase the pool of undifferentiated cells. From the clonal data, we cannot exclude the possibility that BLPs can self-renew symmetrically. However, our observations of the rate of bromodeoxyuridine (BrdU) incorporation in the steady-state trachea (Figures S1A–S1D) set the overall rate of all BC divisions to 0.09 per day. If we allow BLPs to divide symmetrically in the model, the other parameters are such that we estimate the maximum rate at which BLP division could occur is 0.045 cell divisions per day (Supplemental Theory). This is half the total number of observed cell divisions in the basal layer, suggesting that the BLPs do not divide at a greater rate than the BSCs and are thus not a traditionally defined transit-amplifying population.

A close inspection of the fit to the *Tg(KRT5-CreER)*; *Rosa26R-fGFP* data reveals that the model provides a consistent slight underestimate of single SecC clones (Figure 3A, blue charts). We hypothesized that the excess single SecCs observed represent rare brush and neuroendocrine (NE) cells that are of unknown origin in the trachea (Krasteva et al., 2011; Saunders et al., 2013). Indeed, when we stained *Tg(KRT5-CreER)*; *Rosa26R-fGFP* tracheas for an NE marker, we identified lineage-traced NE cells (Figure 3B). This demonstrates that tracheal NE cells can be derived from BSCs. However, brush and NE cells were indistinguishable from SecCs in our quantitative experiments, and we have not specifically included them in the model.

Independent Clonal and Proliferation Analysis Supports the Tracheal Lineage Model

To further challenge the validity of the model, we repeated part of the BC lineage-tracing time course by using an independent

were scored as $KRT5^-$, $ac-TUB^-$ cells whose apical surface reaches the tracheal lumen (this definition will also include a small number of other cell types such as brush and neuroendocrine cells). See right yz panel for examples of labeled luminal cells. GFP-labeled BCs are indicated by arrows on xy views. Scale bar, 15 μm . (C) Size distributions of clones at 3, 12, 39, and 74 weeks post-induction. Length of bar represents frequency. (D) Plot of the mean number of total cells per clone (all clones included). Error bars represent SEM. (E) Plot of the percentage of clones that do not contain any labeled BCs (gray bars). Error bars represent 95% confidence intervals. Red line represents values predicted by the BC model (see Figures 2A–2C). (F) Heatmaps to show the distribution of basal and luminal cells within each clone at 0.5, 3, 12, 39, and 74 weeks. Colors represent percentage frequency of each clone type on a log scale. (G) Plot of the mean number of cells of each type per clone (all clones included). Green, BCs; blue, secretory cells; red, ciliated cells. Error bars represent SEM. (H and I) Analysis of our lineage-labeling data at a population level (see Figures 1C–1G) is inconsistent with an homogenous BC population (depicted in H) and suggests there are two BC subtypes and, moreover, that our experiments labeled a basal stem cell (BSC) that maintains both the luminal cells and the second BC population (depicted in I). See also Figures S1 and S2.

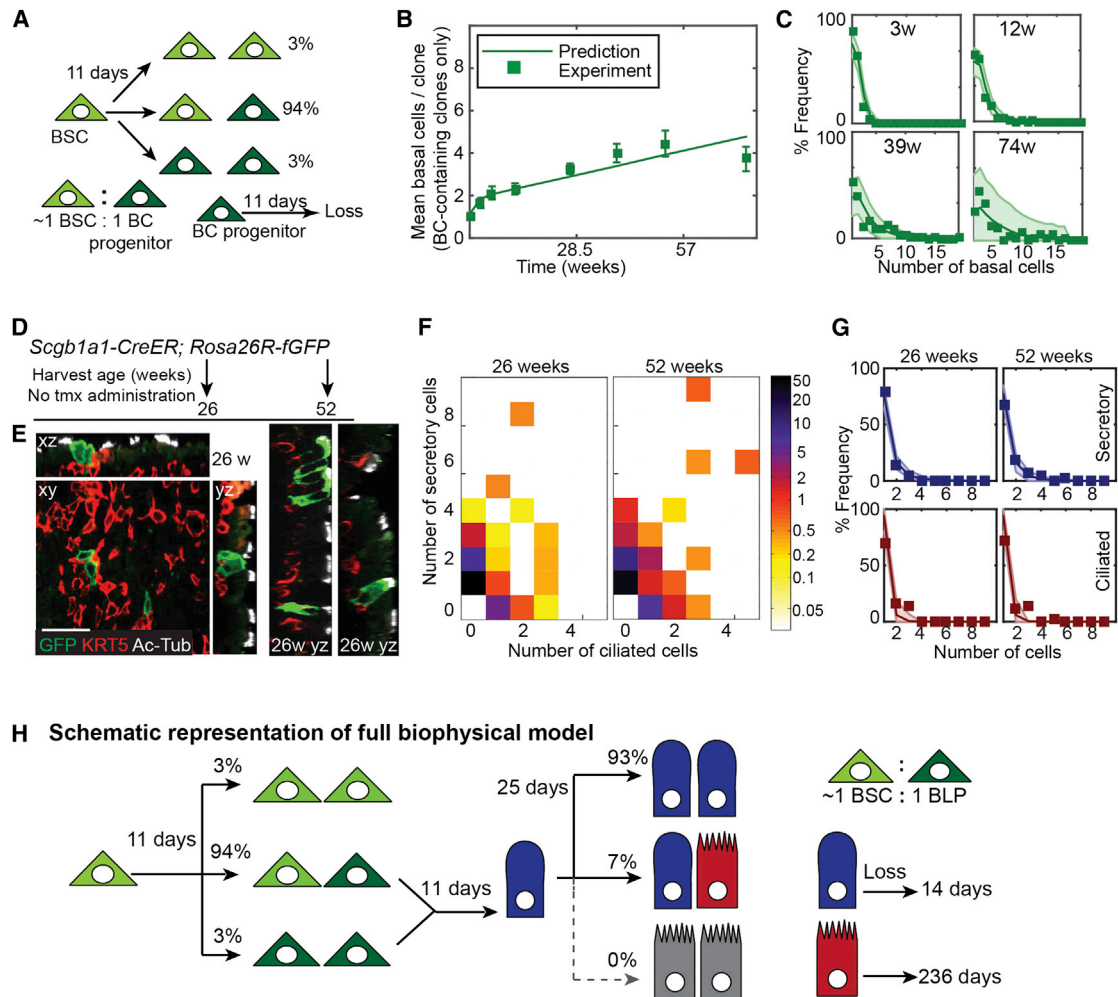


Figure 2. Lineage Model Development for the Basal and Luminal Epithelial Cells

(A) Cellular representation of the mathematical model for BC behavior. Basal stem cells (BSCs) and BC progenitors are present in the tissue in an ~1:1 ratio. BSCs divide symmetrically or asymmetrically every 11 days to either self-renew or produce BC progenitors in a balanced manner. BC progenitors are lost, to either further differentiation or death, at a rate of every 11 days.

(B) Plot showing the experimental observation (boxes) and model prediction (line) of mean number of basal cells per clone (only clones containing BCs are included in this plot and used for the generation of the BC model). Error bars represent SEM.

(C) Percentage frequency distribution of basal cell numbers per clone at 3, 12, 39, and 74 weeks. Boxes, experimental observations; dark line, model prediction of the BC model; shaded area, 95% confidence intervals of the model.

(D) Schematic of *Scgb1a1-CreER*; *Rosa26R-fGFP* lineage-labeling experiment.

(E) Representative xy, yz, and xz confocal sections of three different clones at 26 weeks. (By 26 weeks, the clonal size and composition had already reached a steady state, and these images are representative of the full data range.) Green, lineage label; red, KRT5 (BCs); white, acetylated tubulin (cilia). Scale bar, 25 μ m.

(F) Heatmaps to show the distribution of secretory and ciliated cells observed in all *Scgb1a1-CreER*; *Rosa26R-fGFP* clones over time. Colors represent percentage frequency of each clone type on a log scale.

(G) Frequency distribution of secretory or ciliated cell numbers per clone at 26 and 52 weeks. Blue, secretory; red, ciliated cells. Boxes, experimental observations; dark line, predictions of luminal cell model; shaded area, 95% confidence intervals of the luminal cell model.

(H) Cellular representation of the combined mathematical model. BSCs (light green) divide on average once every 11 days in a balanced manner to produce equal numbers of new BSCs and BLPs (dark green). (Ratio of BSCs: BLPs within the epithelium is ~1:1.) BLPs mature to a luminal cell fate (depicted as secretory, but could be ciliated) once every 11 days. Secretory cells (blue) divide every 25 days to generate two new secretory cells (93%) or one secretory and 1 ciliated (red) cell. The relative proportions of secretory and ciliated cells in the epithelium (~2:1) are maintained by differential rates of loss (every 14 days for secretory cells and every 236 days for ciliated cells).

Krt5-CreER knockin strain (Van Keymeulen et al., 2011) with a *Rosa-confetti* reporter (Snippert et al., 2010) (Figure 3C; Table S1). With the same cell kinetics and fate probabilities, we found that the model reliably predicted the experimental observations

(Figure 3D), even though animals were exposed to a different tmx-induction regimen. Similarly, as an additional consistency check, we used *Tg(Krt8-CreER)*; *Rosa26R-fGFP* animals (Van Keymeulen et al., 2011) with a low dose of tamoxifen to label

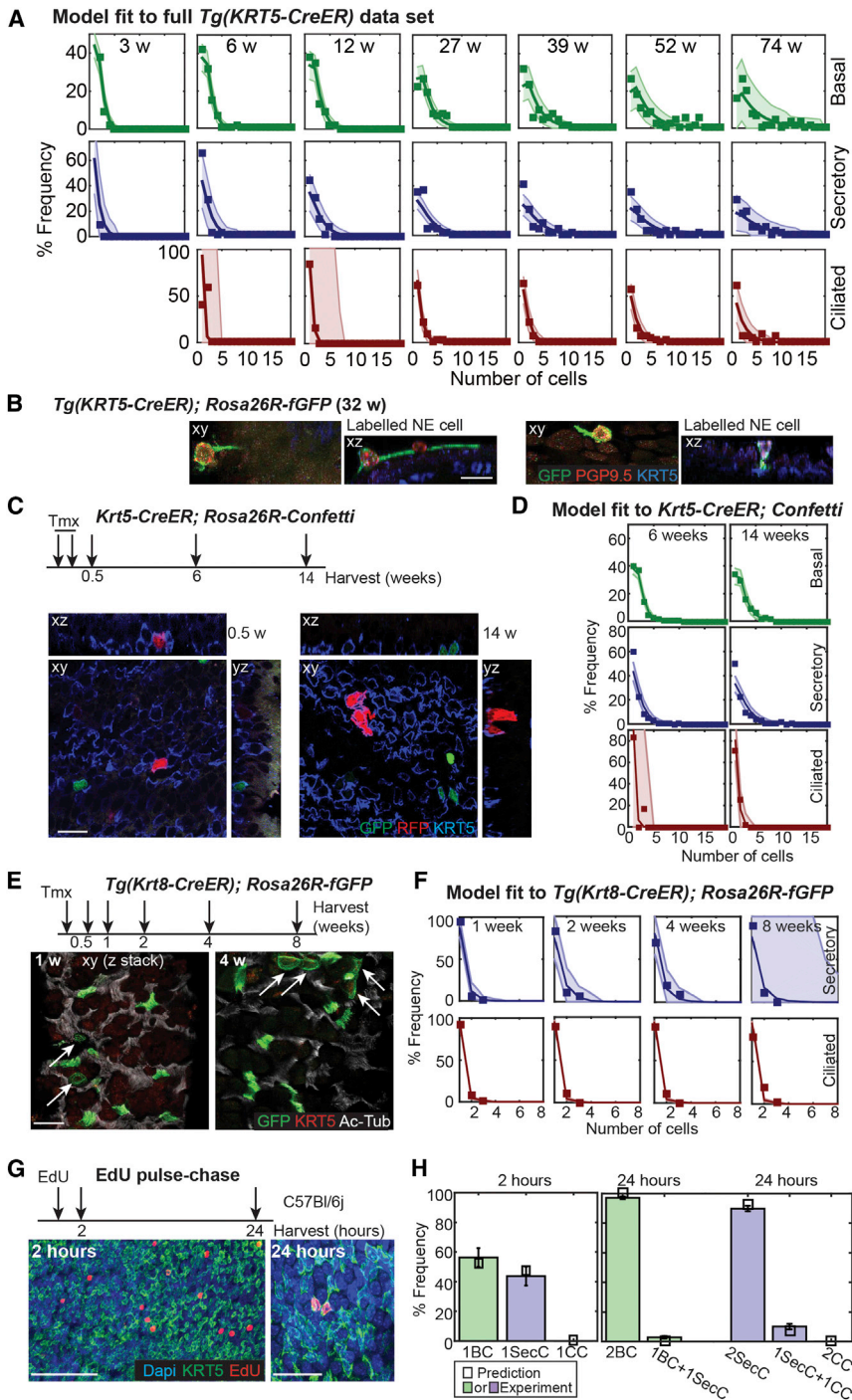


Figure 3. Testing the Tracheal Epithelial Lineage Model

(A) Plots showing the fit of the combined (basal and luminal cells) mathematical model to the full *Tg(KRT5-CreER); Rosa26R-fGFP* dataset (all clones included) throughout the time course. Frequency distribution of basal (green), secretory (blue) and ciliated (red) cell numbers per clone. Boxes, experimental observation; dark line, predictions of combined model; shaded area, 95% confidence intervals of the combined model.

(B) Representative confocal xy and corresponding xz projections of lineage-labeled NE cells in *Tg(KRT5-CreER); Rosa26R-fGFP* trachea harvested 32 weeks post-induction. Green, GFP (lineage-label); red, PGP9.5 (NE cells); blue, KRT5 (BCs). Scale bar, 10 μ m.

(C) *Krt5-CreER; Rosa26R-Confetti* mice were labeled and followed for 14 weeks. Representative single confocal xy planes of the basal area of the whole-mount epithelium with the corresponding xz and yz reconstructions at 0.5 and 14 weeks post-induction. Green, nGFP (lineage label); red, RFP (lineage label); blue, KRT5 (BCs). Scale bar, 15 μ m.

(D) Plots showing the fit of the combined mathematical model to the *Krt5-CreER; Rosa26R-Confetti* dataset at 6 and 14 weeks post-induction. Frequency distribution of basal (green), secretory (blue), and ciliated (red) cell numbers per clone. Boxes, experimental observation; dark line, prediction of combined model; shaded area, 95% confidence intervals of the combined model.

(E) Experimental schematic and confocal projections (z stacks) of apical regions from whole-mount lineage-labeled *Tg(Krt8-CreER); Rosa26R-fGFP* tracheal preparations. Green, GFP (lineage label); red, KRT5 (BCs not visible in these projections, but scored in every image); white, acetylated tubulin (cilia). Scale bar, 20 μ m. Arrows indicate lineage-labeled SecCs. The other GFP⁺ cells are all ciliated cells.

(F) Plots showing the fit of the luminal mathematical model to the *Tg(Krt8-CreER); Rosa26R-fGFP* dataset. Frequency distribution of secretory (blue) and ciliated (red) cell numbers per clone. Boxes, experimental observation; dark line, prediction; shaded area, 95% confidence intervals of the model.

(G) Experimental schematic and single confocal planes of the basal region from whole-mount tracheal preparations showing EdU incorporation at 2 and 24 hr post-exposure in wild-type adults. Green, KRT5 (basal cells). Red, EdU (cells in S phase). Blue, DAPI (nuclei). Scale bar represents 100 μ m (2 hr) or 25 μ m (24 hr).

(H) Plot showing the percentage frequency of types of dividing cells observed (colored bars) compared

to predictions of the combined model (small black boxes). Error bars represent SEM; n = 3 mice at each time. (Note: only single EdU⁺ cells at 2 hr and pairs of EdU⁺ cells at 24 hr were included in the analysis. The full dataset included some groups of more than two cells and some single cells at 24 hr, likely resulting from clone merging or EdU toxicity, respectively; Table S1.)

scattered luminal cells for up to 8 weeks (Figure 3E). These data were also found to be consistent with the quantitative predictions of our model (Figure 3F; Supplemental Theory).

Our model makes strong predictions about the rates and types of cell division in the tracheal epithelium. We tested these by us-

ing nucleotide incorporation assays. We found that BC BrdU incorporation rates provided an estimate of cell cycle times within the range predicted by the model (Supplemental Theory). Moreover, to identify types of cell divisions, we combined 5-ethynyl-2'-deoxyuridine (EdU) incorporation with whole-mount

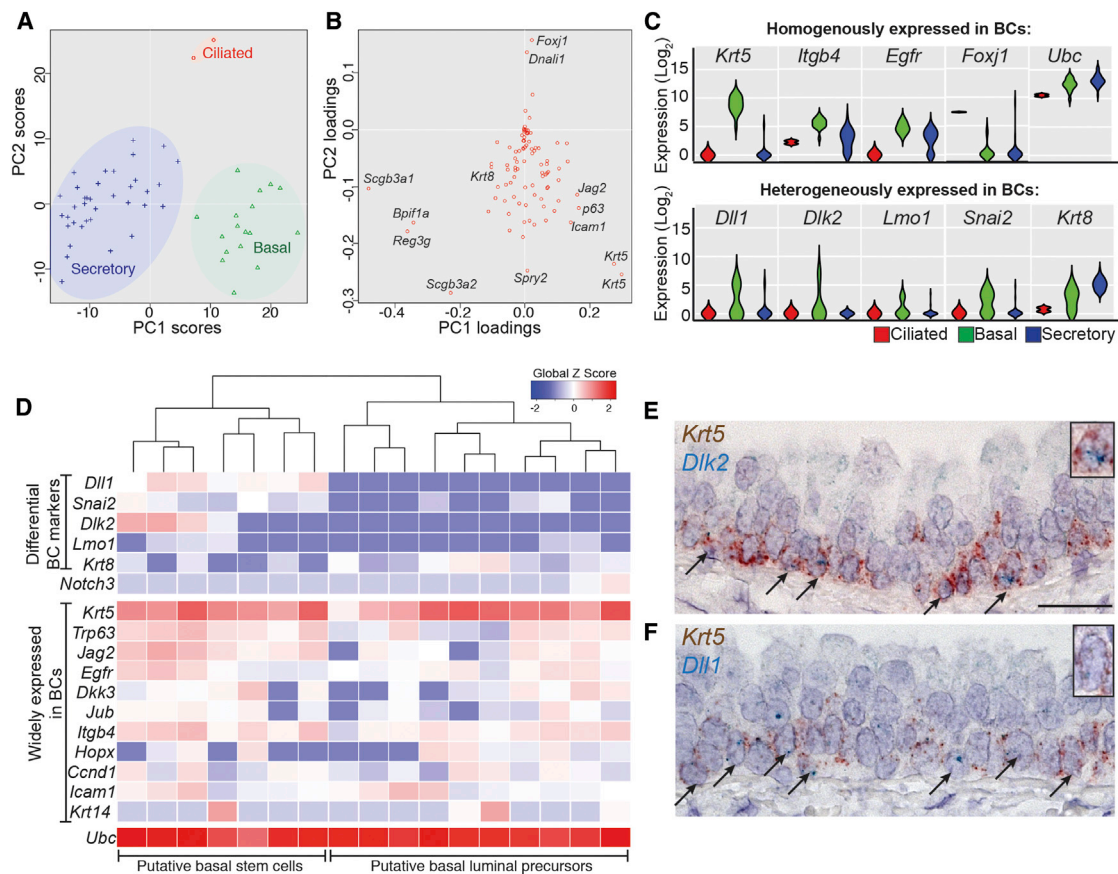


Figure 4. Molecular Analysis Distinguishes Two Subtypes of Basal Cells

(A) Principal component analysis (PCA) of 56 single tracheal epithelial cells from the expression levels of all 92 genes determined by qPCR distinguishes between the major cell lineages.

(B) Principal component loadings indicate the extent to which each gene contributes to the separation of cells along each component in (A).

(C) Violin plots showing \log_2 expression levels of selected genes in BCs (green) or secretory (blue) or ciliated (red) cells. Monophasic (single bulge) plots indicate homogenous gene expression. Biphasic plots (two bulges) indicate two expression levels within the population and show that *Dll1*, *Dlk2*, *Lmo1*, *Snai2*, and *Krt8* are heterogeneously expressed in BCs. Interestingly, *Itgb4* and *Egfr* are heterogeneous within the secretory population but homogenous in the BCs.

(D) Heatmap showing unsupervised hierarchical clustering of the BCs only, based on expression levels of 15 of the BC-enriched genes. BCs split into two major groups: putative stem cells and luminal precursors.

(E and F) Double mRNA in situ hybridization of wild-type adult trachea sections showing *Krt5* (brown, basal cells) and *Dlk2* (blue, E) or *Dll1* (blue, F). Arrows indicate co-expressing cells. Scale bar, 20 μ m.

See also Figure S3.

immunostaining to visualize S phase cells (2 hr post-Edu) and their immediate progeny (24 hr post-Edu; Figure 3G; Table S1). The observations confirmed that in wild-type animals, almost 100% of BC divisions result in two BCs, and 95% of SecC divisions result in two SecCs and 5% one SecC and one CC, in excellent agreement with our predictions (Figure 3H; Supplemental Theory). Indeed, if we take the frequency of CC production from SecCs as the lowest 95% confidence interval from our Edu measurements (7% of all SecC divisions) and use the division rate of SecCs from our model, we find that less than 1% of CC production can potentially originate from BLPs (Supplemental Theory). Thus, our Edu experiments support the model prediction that the major route of new CC production at homeostasis is via the SecCs, rather than direct differentiation of the BLPs.

Single-Cell qRT-PCR Identifies Two Molecularly Distinct Subtypes of BCs

Our experiments support a cellular hierarchy in which there are two distinct BCs. To establish whether these BC populations are molecularly distinct, we performed qRT-PCR on 67 single cells isolated from total epithelium. We analyzed 5 housekeeping, 20 lineage-specific, and 67 genes reportedly enriched in BCs (Hackett et al., 2011; Rock et al., 2009) (Supplemental Experimental Procedures). Cells were grouped by unsupervised hierarchical clustering by using expression levels of all tested genes (Figure S3). This defined three major groups, also seen separated by an independent principal component analysis (Figure 4A). Principal component loadings showed that these represented BCs, SecCs, and CCs (Figure 4B). Similarly, pairwise ANOVA between the groups also showed that each was enriched for expression of

definitive markers (Table S2). Analysis of gene expression levels between individual cells within each group showed that *Dlk2*, *Dll1*, *Lmo1*, *Snai2*, and *Krt8* had biphasic patterns within BCs, indicating heterogeneous expression (Figure 4C). This was in agreement with independent unsupervised hierarchical clustering performed on BCs alone (Figure 4D). In this analysis *Dlk2*, *Dll1*, *Lmo1*, and *Snai2* tended to be expressed together in one BC population, while *Krt8* (a luminal cell marker) was more highly expressed in a second population. The only gene previously reported as differentially expressed in BCs, *Krt14*, was detected in a small number of cells in both subpopulations and is therefore unlikely to distinguish them (Figure 4D). mRNA in situ hybridization confirmed the heterogeneous distribution of *Dlk2* and *Dll1* within BCs (Figures 4E and 4F). These data support our model of two distinct BC subpopulations. Moreover, they suggest that one subpopulation is upregulating luminal markers (*Krt8*), and we hypothesized that this is the BLP.

Low-Level *Krt8* Expression Characterizes a Population of BCs that Are Lost Rapidly from the Tracheal Epithelium

To test the hypothesis that the tracheal epithelium contains a widespread basal *Krt8*⁺ luminal precursor, we used *Krt8-rtTA*; *tetO-H2B-GFP* animals. Two weeks of exposure to doxycycline labeled ~18% of BCs and most SecCs (Figures 5A and 5B; Table S1). Significantly, we found that SecC labeling was maintained after a 6-week chase (Figure 5D), whereas BC labeling was almost absent, consistent with our hypothesis (Figure 5E). Moreover, the loss rate of the GFP⁺ BCs fitted extremely well to an exponential decay curve ($R^2 = 0.998867$) giving a measured loss rate of the BLPs of 16 days, in good agreement with our predicted rate of $11 \pm (7, 4)$ days. Immediately after induction, labeled BCs had 3-fold lower levels of GFP than luminal cells (Figure 5C; Table S1), consistent with a lower expression level of *Krt8* and further supporting our hypothesis that they are a distinct, differentiating BC subpopulation. These data are consistent with our model of two BC subpopulations, one of which is fated as a luminal precursor (Figure 5F).

DISCUSSION

Uncovering the proliferative hierarchy, quantitative fate behavior, and molecular profile of a slow-cycling tissue inevitably requires detailed and long-term studies. By combining the results of long-term lineage tracing using multiple drivers with single-cell gene expression profiling, we have shown that tracheal KRT5⁺ BCs include two subpopulations: stem cells (BSCs) and luminal precursors (BLPs). BSCs maintain epithelial homeostasis by dividing in a balanced manner to self-renew and produce BLPs. BLPs are widespread, long-lived (~2 weeks) luminal precursors, which upregulate the luminal cytokeratin *Krt8* before overt signs of differentiation (Figure 5F). Our data suggest that these cells do not function as a traditionally defined transit-amplifying population. Rather, the existence of a widely distributed, luminal-fated, long-lived precursor, which is morphologically indistinguishable from the stem cell, is a surprising finding and has significant implications for fundamental and reparative biology, and disease initiation, in the airways. In particular, our model suggests that there

must be at least two separate signaling events controlling luminal cell differentiation: specification as a BLP (which may or may not be concurrent with secretory/ciliated specification), and, separately, maturation of that cell into a differentiated luminal cell. We have therefore defined another event (maturation) at which normal homeostasis can be perturbed in disease.

A KRT5⁺, KRT8⁺ parabasal cell population was recently detected in wild-type adults (~25% of the total BC population). The authors suggested that these cells were precursors of the luminal cells and showed that their abundance is regulated by NOTCH3 signaling (Mori et al., 2015). Interestingly, we observed higher *Notch3* mRNA expression in a small number of BLPs that also expressed *Krt8* (Figure 4D). It is likely that the parabasal cells and BLPs (which we propose comprise ~50% of the total BC population) are overlapping cell populations, but their exact relationship is yet to be defined.

Our conclusion that BSCs maintain the airways by a process of stochastic homeostasis is in principle similar to the model defined for the BC population in the mouse inter-follicular epidermis (Clayton et al., 2007). This means that each time a BSC divides, it has a certain probability of making a symmetric or asymmetric division (Figure 2H). The implication of this finding is that an individual BSC can by chance make a large number of symmetric self-renewing divisions (compensated for, at a population level, by symmetric differentiation divisions of other BSCs) leading to neutral drift in the population of clones. This is consistent with the model proposed for human airway BSCs in vivo based on clonal analysis of naturally occurring mitochondrial mutations (Teixeira et al., 2013). Clonal experiments in mice provide us with richer datasets than those available in human samples due to the defined time courses and greater numbers of replicates. This has allowed us to greatly extend the observations of Teixeira et al. and identify the BLPs, the rapid rate of SecC turnover, and the production of NE cells by BSCs. In the future, it will be important to test if these observations are also recapitulated in human airways.

The KRT5⁺ BSC that we characterize achieves perfect self-renewal for at least 17 months. We cannot exclude the existence of an additional, rarely dividing, minority BSC in the trachea (Borthwick et al., 2001). However, our data show that such a cell is not required for homeostatic turnover. A subset (<20%) of BCs express KRT14. These cells typically occur in clusters and have been proposed to be unipotent stem cells at steady state (Ghosh et al., 2011). We detected rare *Krt14*⁺ cells in both the BSC and BLP populations, suggesting that KRT14 is not a reliable marker of stem cell identity. Previous *Tg(KRT14-CreER)* lineage labeling studies in which unipotent BC clones were observed used a time course of ~40 days (Ghosh et al., 2011; Hong et al., 2004) and may not have allowed sufficient time for BLPs to differentiate into luminal cells. We speculate that the clustered distribution of KRT14⁺ BCs and the rapid up-regulation of KRT14 levels post-injury indicate a structural role for this intermediate filament protein.

The trachea does have an additional dividing cell population at steady state: the SecC. We confirm that SecC to BC reversion does not happen at homeostasis (Rawlins et al., 2009). Moreover, we show that SecCs have a short half-life compared to CCs, are not traditionally defined transit-amplifying cells but do preferentially self-renew, such that at any given time, ~44% of

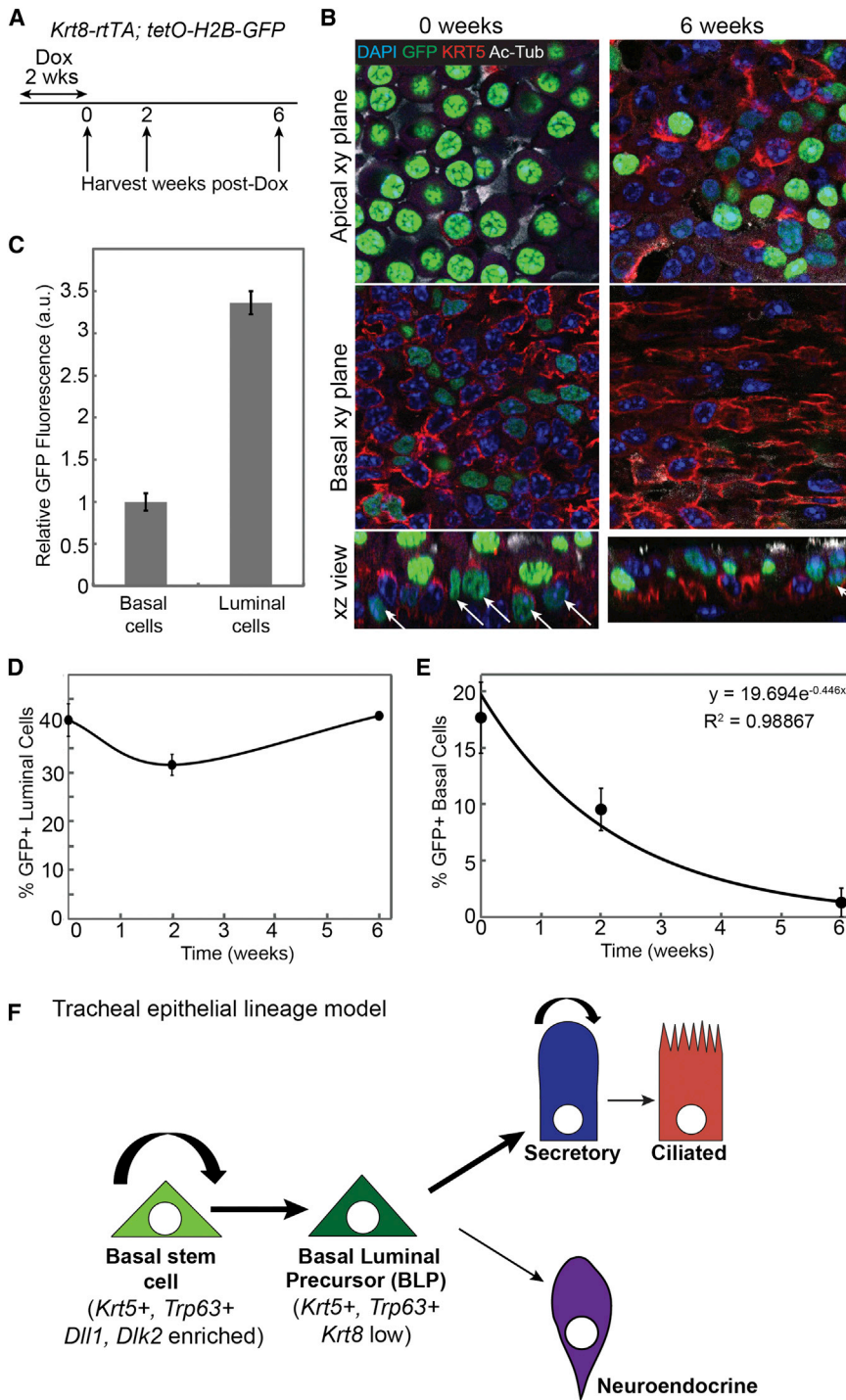


Figure 5. *Krt8* Is Expressed in a Subset of Basal Cells that Are Lost Rapidly from the Basal Layer

(A) Schematic of *Tg(Krt8-rtTA); tetO-H2B-GFP* labeling experiment.

(B) Representative single confocal xy planes from a confocal stack of a basal area and a more apical area of the whole-mount epithelium with a corresponding xz reconstruction at 0 and 6 weeks post-doxycycline. Blue, Hoechst (nuclei); green, H2B-GFP (short-term lineage label); red, KRT5 (BCs); white, acetylated tubulin (cilia). Scale bar, 10 μ m.

(C) Plot of the levels of GFP fluorescence in the basal and luminal cells at 0 weeks post-doxycycline. GFP fluorescence intensity was measured in a.u. and the levels were normalized to 1 in the basal cells. Error bars represent 95% confidence intervals.

(D) Plot of the percentage of GFP⁺ luminal cells over time. Error bars represent 95% confidence intervals. Note that of the luminal cells, only the SecCs are labeled with GFP, resulting in an ~40% labeling efficiency of the entire luminal population.

(E) Plot of the percentage of GFP⁺ basal cells over time. Error bars represent 95% confidence intervals (n = 3 mice at 0 and 2 weeks; n = 2 mice at 6 weeks).

(F) Model of the tracheal epithelial cell lineage. Basal stem cells (BSCs) are enriched in DI-family ligands and divide, mostly via asymmetric division (94%) to produce on average one new BSC and one BLP. The BLPs upregulate *Krt8* expression and differentiate into luminal cells over the course of ~11 days. The simplest model is one in which BLPs do not divide. We cannot exclude the possibility of a small number of BLP divisions, although our modeling suggests that any division of BLPs must be occurring at a rate of <50% of the total basal cell divisions, making them unlikely to function as a transit-amplifying progenitor. The terminally differentiated ciliated cells are produced at an extremely low rate at steady state, and our data suggest that division of SecCs is sufficient to account for all new CC production. Hence, BLPs are depicted as differentiating to secretory cell fate. However, it is possible that they also directly produce ciliated cells at a very low frequency. Our data also show that BSCs can directly produce NE cells. Secretory cells divide at a lower rate than BSCs. They have a short half-life and mostly self-renew (>90% self-renewing divisions), but they do generate new ciliated cells at low frequency.

SecCs within the epithelium are derived from a pre-existing SecC. This is analogous to prostate and breast epithelia (Chua et al., 2014; Ousset et al., 2012; Rios et al., 2014; Van Keymeulen et al., 2011) where luminal progenitors are recognized as a cell of origin for some cancers (Bai et al., 2013; Molyneux et al., 2010; Wang et al., 2013, 2014). Thus, we may need to consider multiple potential initiating cells for airway carcinogenesis.

Why does the tracheal epithelium contain a widespread, long-lived luminal precursor population that is located in the basal position? We speculate that post-injury, BLPs may either rapidly differentiate or revert to stem cell fate and thus function as a basally positioned reserve progenitor that is not exposed to inhaled agents. Significantly, a recent publication has identified a population of steady-state BCs (~12% of the total) that

express low levels of transcription factors usually found in more differentiated luminal cells. Rarely, these BCs were also KRT8⁺ (Pardo-Saganta et al., 2015). Post-injury, the abundance of luminal transcription factor-positive BCs increased to ~50% of the total BCs, and they were subsequently observed to proliferate. Our model has BLPs present as ~50% of the total steady-state BC population. It is highly likely that the ~12% of transcription-factor-positive homeostatic BCs identified by Pardo-Saganta et al. (2015) correspond to the most differentiated of the homeostatic BLPs that we identified. If this hypothesis is correct, the post-injury role of BLPs is indeed to both rapidly differentiate and proliferate. However, exact testing of this hypothesis will require more specific tools for lineage-labeling the BLP population. Given that the BLPs are a subpopulation of BCs that, at homeostasis, are undergoing a slow differentiation process to luminal fate, a unique transcription factor or other specific molecular marker may be hard to find.

Are BLPs a unique aspect of airway stem cell biology, or could other epithelia also contain fate-committed cells that are morphologically indistinguishable from the stem cells? A recent report using prospective isolation techniques suggests that there is a hierarchy of BC organization in the mouse esophagus, with stem cells being molecularly distinct from more committed suprabasal precursors (DeWard et al., 2014). Similarly, we note that the enteroblast (EB) cell in the *Drosophila* midgut can only be distinguished from the stem cell by expression of *DI* (Micchelli and Perrimon, 2006; Ohlstein and Spradling, 2006, 2007) and occupies a similar position in the cellular hierarchy to tracheal BLPs. We therefore speculate that uncoupling division of the stem/progenitor from morphological differentiation of the progeny may be a previously undetected general phenomenon in epithelia, with steady-state rates of cell maturation being controlled by cell-type-specific mechanisms. Our strategy of long-term in vivo lineage tracing, coupled with single-cell molecular analysis, should prove widely applicable for the fine dissection of homeostatic lineage potential in epithelia such as the esophagus, prostate, mammary gland, and skin, where no obvious morphological stem cell niches exist.

EXPERIMENTAL PROCEDURES

See Supplemental Experimental Procedures for a full description of all materials and methods used.

Animals

All experiments were performed under license PPL80/2326. *Tg(KRT5-CreER)* transgenic (Rock et al., 2009), *Krt5-CreER^{T2}* knockin (Van Keymeulen et al., 2011), *Scgb1a1-CreER* (Rawlins et al., 2009), *Tg(Krt8-CreER)* (Van Keymeulen et al., 2011), *Rosa26R-fGFP* (Rawlins et al., 2009), *Rosa-confetti* (Snippert et al., 2010), and *tetO-H2B-GFP* (Tumbar et al., 2004) mice have been described. *Krt8-rtTA* transgenic mice were generated by using a fragment of the murine *Krt8* gene. Males and females >8 weeks old were used in all experiments. Wild-type mice were C57Bl/6J.

Lineage Tracing

Low-frequency activation of the reporter was achieved by a single intraperitoneal injection of tamoxifen (Sigma-Aldrich, T5648) at a dose of 25 μg/g body weight in *Tg(KRT5-CreER)*; *Rosa26R-fGFP* or 13 μg/g body weight in *Tg(Krt8-CreER)*; *Rosa26R-fGFP* mice or with two doses of 5 mg tmx per mouse spaced 48 hr apart in *Krt5^{CreER/+}*; *Rosa-confetti* mice. Doxycycline was administered to *Krt8-rtTA*; *tetO-H2B-GFP* mice in food at a dose of 10

g/kg (SAFE-DIETS) for 2 weeks. BrdU was given intraperitoneally at 30 μg/g body weight and EdU at 50 μg per mouse.

Whole-Mount Immunostaining

For *Rosa26R-fGFP* and wild-type mice, tracheas were fixed overnight in 4% paraformaldehyde at 4°C. Primary antibodies were anti-GFP (chicken, 1:1,000; Abcam, AB13970), anti-KRT5 (rabbit, 1:500; Covance, PRB-160P), anti-acetylated tubulin (mouse, 1:1,000; Sigma, T7451), and anti-PGP9.5 (guinea pig, 1:500; Neuromics, GP14104). Secondary antibodies were Alexa Fluor conjugates (Life Technologies, 1:2,000). Samples were processed to 97% TDE (2'2'-thiodiethanol) for mounting. For mice carrying *tetO-H2B-GFP* or *Rosa26R-confetti*, the whole-mount protocol was adapted to enable direct visualization of native fluorescence. Anti-GFP staining was omitted, and samples were mounted in Glycergel (Dako) + 2.5% DABCO.

Section Immunostaining

8- to 10-μm cryosections were stained with: anti-acetylated tubulin (mouse, 1:1,000; Sigma, T7451), anti-BrdU (mouse, 1:500; Sigma, B8434), anti-Ecad (rat, 1:3000; Life Technologies, 13-1900), anti-Krt5 (rabbit, 1:500; Covance, PRB-160P), anti-T1α (1:1,000; DSHB, 8.1.1), and anti-Scgb1a1 (rabbit, 1:500; Santa Cruz, sc25555). Antigen retrieval was used for BrdU (2 N HCl 30 min 37°C, 0.5% trypsin 5 min, room temperature).

mRNA In Situ Hybridization

Trachea were formalin-fixed for 24 hr at room temperature and paraffin embedded. 5-μm sections were processed for RNA in situ with the RNA Scope 2-plex Detection Kit (Chromogenic) according to the manufacturer's standard protocol (Advanced Cell Diagnostics). RNAscope probes were *Krt5* (NM 027011.2, region 666–2,086), *Dlk2* (NM 023932.3, region 267–1,279), and *Dll1* (NM 007865.3, region 888–1,883).

Microscopy and Image Scoring

z stacks of the full epithelial thickness were acquired at an optical resolution of 1,024 × 1,024 with an optical z slice every 1 μm. Clones were scored manually by looking through the entire z depth of the tracheal epithelium in FV viewer or LAS AF software to score the identity of all labeled cells. For *tetO-H2B-GFP* samples, z stacks were acquired at an optical resolution of 1,024 × 1,024, with a z slice every 0.38 μm. Fluorescence intensity was assessed in Fiji, using the Gurdon Institute Imaging Facility's plugin, ObjectScan.

Cryosections for analysis of cellular composition/density were imaged on an Olympus FV1000, using a 100× oil objective (numerical aperture [NA] 1.4). The length of the basement membrane in each image was measured in Fiji. Density was calculated as the number of cells present per μm of basement membrane. Cryosections for BrdU analysis were imaged on a Zeiss AxioImager compound microscope, using a 20× air objective (NA 0.8) and counted in Fiji.

Single-Cell qRT-PCR

The distal tracheal epithelium was peeled away from the underlying mesenchyme following a brief dispase digest and dissociated to single cells as described previously (Rock et al., 2009). Unsorted epithelial cells were loaded into a Fluidigm C1 machine on a 10- to 17-μm chip at a concentration of ~400 cells/μl for cell capture, lysis, cDNA synthesis, and target pre-amplification. 67 single cells were used for subsequent qRT-PCR on a 96.96 Fluidigm Dynamic array using a Biomark qPCR machine using TaqMan gene expression assays (Life Technologies). Data analysis was performed in the Fluidigm Singular Analysis Toolset 3.0 in R.

Modeling

See Supplemental Experimental Procedures and Supplemental Theory for all details of model construction.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, Supplemental Theory, three figures, and two tables and can be found with this article online at <http://dx.doi.org/10.1016/j.celrep.2015.06.011>.

AUTHOR CONTRIBUTIONS

J.K.W. designed and performed experiments, analyzed data, and wrote the manuscript. S.R. designed and executed the biophysical modeling and wrote the manuscript. A.C.W. and B.G. assisted with qRT-PCR experimental design and analysis and edited the manuscript. A.W., M.O., A.V.K., and C.B. provided mice and trachea samples and assisted with experimental design. B.D.S. led the biophysical modeling and edited the paper. E.L.R. conceived and led the project, performed experiments, analyzed data, and wrote and edited the manuscript.

ACKNOWLEDGMENTS

This study was supported by the Medical Research Council (G0900424 to E.R.), European Union grant EuroSyStem (200720; FP7/2008), the Newton Trust (to E.R.), the Wellcome Trust (098357/Z/12/Z to B.D.S.), and core grants from the Wellcome Trust (092096) and Cancer Research UK (C6946/A14492).

Received: February 6, 2015

Revised: April 25, 2015

Accepted: June 2, 2015

Published: June 25, 2015

REFERENCES

- Bai, F., Smith, M.D., Chan, H.L., and Pei, X.H. (2013). Germline mutation of *Brca1* alters the fate of mammary luminal cells and causes luminal-to-basal mammary tumor transformation. *Oncogene* 32, 2715–2725.
- Borthwick, D.W., Shahbazian, M., Krantz, Q.T., Dorin, J.R., and Randell, S.H. (2001). Evidence for stem-cell niches in the tracheal epithelium. *Am. J. Respir. Cell Mol. Biol.* 24, 662–670.
- Brechbuhl, H.M., Ghosh, M., Smith, M.K., Smith, R.W., Li, B., Hicks, D.A., Cole, B.B., Reynolds, P.R., and Reynolds, S.D. (2011). β -catenin dosage is a critical determinant of tracheal basal cell fate determination. *Am. J. Pathol.* 179, 367–379.
- Chua, C.W., Shibata, M., Lei, M., Toivanen, R., Barlow, L.J., Bergren, S.K., Baddani, K.K., McKiernan, J.M., Benson, M.C., Hibshoosh, H., and Shen, M.M. (2014). Single luminal epithelial progenitors can generate prostate organoids in culture. *Nat. Cell Biol.* 16, 951–961, 1–4.
- Clayton, E., Doupé, 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, 185–189.
- DeWard, A.D., Cramer, J., and Lagasse, E. (2014). Cellular heterogeneity in the mouse esophagus implicates the presence of a nonquiescent epithelial stem cell population. *Cell Rep.* 9, 701–711.
- Engelhardt, J.F., Schlossberg, H., Yankaskas, J.R., and Dudus, L. (1995). Progenitor cells of the adult human airway involved in submucosal gland development. *Development* 121, 2031–2046.
- Ghosh, M., Brechbuhl, H.M., Smith, R.W., Li, B., Hicks, D.A., Titchner, T., Runkle, C.M., and Reynolds, S.D. (2011). Context-dependent differentiation of multipotential keratin 14-expressing tracheal basal cells. *Am. J. Respir. Cell Mol Biol.* 45, 403–410.
- Giagreco, A., Lu, L., Vickers, C., Teixeira, V.H., Groot, K.R., Butler, C.R., llieva, E.V., George, P.J., Nicholson, A.G., Sage, E.K., et al. (2012). β -Catenin determines upper airway progenitor cell fate and preinvasive squamous lung cancer progression by modulating epithelial-mesenchymal transition. *J. Pathol.* 226, 575–587.
- Hackett, N.R., Shaykhiyev, R., Walters, M.S., Wang, R., Zwick, R.K., Ferris, B., Witover, B., Salit, J., and Crystal, R.G. (2011). The human airway epithelial basal cell transcriptome. *PLoS ONE* 6, e18378.
- Hajj, R., Baranek, T., Le Naour, R., Lesimple, P., Puchelle, E., and Coraux, C. (2007). Basal cells of the human adult airway surface epithelium retain transit-amplifying cell properties. *Stem Cells* 25, 139–148.
- Hong, K.U., Reynolds, S.D., Watkins, S., Fuchs, E., and Stripp, B.R. (2004). In vivo differentiation potential of tracheal basal cells: evidence for multipotent and unipotent subpopulations. *Am. J. Physiol. Lung Cell. Mol. Physiol.* 286, L643–L649.
- Krasteva, G., Canning, B.J., Hartmann, P., Veres, T.Z., Papadakis, T., Mühlfeld, C., Schliecker, K., Tallini, Y.N., Braun, A., Hackstein, H., et al. (2011). Cholinergic chemosensory cells in the trachea regulate breathing. *Proc. Natl. Acad. Sci. USA* 108, 9478–9483.
- Lu, L., Teixeira, V.H., Yuan, Z., Graham, T.A., Endesfelder, D., Kolluri, K., Al-Juffali, N., Hamilton, N., Nicholson, A.G., Falzon, M., et al. (2013). LRIG1 regulates cadherin-dependent contact inhibition directing epithelial homeostasis and pre-invasive squamous cell carcinoma development. *J. Pathol.* 229, 608–620.
- Mascreé, G., Dekoninck, S., Drogat, B., Youssef, K.K., Brohéé, S., Sotiropoulou, P.A., Simons, B.D., and Blanpain, C. (2012). Distinct contribution of stem and progenitor cells to epidermal maintenance. *Nature* 489, 257–262.
- Micchelli, C.A., and Perrimon, N. (2006). Evidence that stem cells reside in the adult *Drosophila* midgut epithelium. *Nature* 439, 475–479.
- Molyneux, G., Geyer, F.C., Magnay, F.A., McCarthy, A., Kendrick, H., Natrajan, R., Mackay, A., Grigoriadis, A., Tutt, A., Ashworth, A., et al. (2010). BRCA1 basal-like breast cancers originate from luminal epithelial progenitors and not from basal stem cells. *Cell Stem Cell* 7, 403–417.
- Mori, M., Mahoney, J.E., Stupnikov, M.R., Paez-Cortez, J.R., Szymaniak, A.D., Varelas, X., Herrick, D.B., Schwob, J., Zhang, H., and Cardoso, W.V. (2015). Notch3-Jagged signaling controls the pool of undifferentiated airway progenitors. *Development* 142, 258–267.
- Ohlstein, B., and Spradling, A. (2006). The adult *Drosophila* posterior midgut is maintained by pluripotent stem cells. *Nature* 439, 470–474.
- Ohlstein, B., and Spradling, A. (2007). Multipotent *Drosophila* intestinal stem cells specify daughter cell fates by differential notch signaling. *Science* 315, 988–992.
- Ousset, M., Van Keymeulen, A., Bouvencourt, G., Sharma, N., Achouri, Y., Simons, B.D., and Blanpain, C. (2012). Multipotent and unipotent progenitors contribute to prostate postnatal development. *Nat. Cell Biol.* 14, 1131–1138.
- Pardo-Saganta, A., Law, B.M., Tata, P.R., Villoria, J., Saez, B., Mou, H., Zhao, R., and Rajagopal, J. (2015). Injury induces direct lineage segregation of functionally distinct airway basal stem/progenitor cell subpopulations. *Cell Stem Cell* 16, 184–197.
- Paul, M.K., Bisht, B., Darmawan, D.O., Chiou, R., Ha, V.L., Wallace, W.D., Chon, A.T., Hegab, A.E., Grogan, T., Elashoff, D.A., et al. (2014). Dynamic changes in intracellular ROS levels regulate airway basal stem cell homeostasis through Nrf2-dependent notch signaling. *Cell Stem Cell* 15, 199–214.
- Rawlins, E.L., and Hogan, B.L. (2008). Ciliated epithelial cell lifespan in the mouse trachea and lung. *Am. J. Physiol. Lung Cell. Mol. Physiol.* 295, L231–L234.
- Rawlins, E.L., Ostrowski, L.E., Randell, S.H., and Hogan, B.L. (2007). Lung development and repair: contribution of the ciliated lineage. *Proc. Natl. Acad. Sci. USA* 104, 410–417.
- Rawlins, E.L., Okubo, T., Xue, Y., Brass, D.M., Auten, R.L., Hasegawa, H., Wang, F., and Hogan, B.L. (2009). The role of *Scgb1a1+* Clara cells in the long-term maintenance and repair of lung airway, but not alveolar, epithelium. *Cell Stem Cell* 4, 525–534.
- Rios, A.C., Fu, N.Y., Lindeman, G.J., and Visvader, J.E. (2014). In situ identification of bipotent stem cells in the mammary gland. *Nature* 506, 322–327.
- Rock, J.R., Onaitis, M.W., Rawlins, E.L., Lu, Y., Clark, C.P., Xue, Y., Randell, S.H., and Hogan, B.L. (2009). Basal cells as stem cells of the mouse trachea and human airway epithelium. *Proc. Natl. Acad. Sci. USA* 106, 12771–12775.
- Rock, J.R., Randell, S.H., and Hogan, B.L. (2010). Airway basal stem cells: a perspective on their roles in epithelial homeostasis and remodeling. *Dis. Model. Mech.* 3, 545–556.
- Rock, J.R., Gao, X., Xue, Y., Randell, S.H., Kong, Y.Y., and Hogan, B.L. (2011). Notch-dependent differentiation of adult airway basal stem cells. *Cell Stem Cell* 8, 639–648.

- Saunders, C.J., Reynolds, S.D., and Finger, T.E. (2013). Chemosensory brush cells of the trachea. A stable population in a dynamic epithelium. *Am. J. Respir. Cell Mol. Biol.* *49*, 190–196.
- Snippert, H.J., van der Flier, L.G., Sato, T., van Es, J.H., van den Born, M., Kroon-Veenboer, C., Barker, N., Klein, A.M., van Rheenen, J., Simons, B.D., and Clevers, H. (2010). Intestinal crypt homeostasis results from neutral competition between symmetrically dividing Lgr5 stem cells. *Cell* *143*, 134–144.
- Snitow, M.E., Li, S., Morley, M.P., Rathi, K., Lu, M.M., Kadzik, R.S., Stewart, K.M., and Morrisey, E.E. (2015). Ezh2 represses the basal cell lineage during lung endoderm development. *Development* *142*, 108–117.
- Teixeira, V.H., Nadarajan, P., Graham, T.A., Pipinikas, C.P., Brown, J.M., Falzon, M., Nye, E., Poulosom, R., Lawrence, D., Wright, N.A., et al. (2013). Stochastic homeostasis in human airway epithelium is achieved by neutral competition of basal cell progenitors. *eLife* *2*, e00966.
- Tumbar, T., Guasch, G., Greco, V., Blanpain, C., Lowry, W.E., Rendl, M., and Fuchs, E. (2004). Defining the epithelial stem cell niche in skin. *Science* *303*, 359–363.
- Van Keymeulen, A., Rocha, A.S., Ousset, M., Beck, B., Bouvencourt, G., Rock, J., Sharma, N., Dekoninck, S., and Blanpain, C. (2011). Distinct stem cells contribute to mammary gland development and maintenance. *Nature* *479*, 189–193.
- Wang, Z.A., Mitrofanova, A., Bergren, S.K., Abate-Shen, C., Cardiff, R.D., Califano, A., and Shen, M.M. (2013). Lineage analysis of basal epithelial cells reveals their unexpected plasticity and supports a cell-of-origin model for prostate cancer heterogeneity. *Nat. Cell Biol.* *15*, 274–283.
- Wang, Z.A., Toivanen, R., Bergren, S.K., Chambon, P., and Shen, M.M. (2014). Luminal cells are favored as the cell of origin for prostate cancer. *Cell Rep.* *8*, 1339–1346.
- Wansleben, C., Bowie, E., Hotten, D.F., Yu, Y.R., and Hogan, B.L. (2014). Age-related changes in the cellular composition and epithelial organization of the mouse trachea. *PLoS ONE* *9*, e93496.
- Watt, F.M., and Hogan, B.L. (2000). Out of Eden: stem cells and their niches. *Science* *287*, 1427–1430.
- Zhao, R., Fallon, T.R., Saladi, S.V., Pardo-Saganta, A., Villoria, J., Mou, H., Vinnarsky, V., Gonzalez-Celeiro, M., Nunna, N., Hariri, L.P., et al. (2014). Yap tunes airway epithelial size and architecture by regulating the identity, maintenance, and self-renewal of stem cells. *Dev. Cell* *30*, 151–165.

Cell Reports

Supplemental Information

Clonal Dynamics Reveal Two Distinct Populations of Basal Cells in Slow-Turnover Airway Epithelium

**Julie K. Watson, Steffen Rulands, Adam C. Wilkinson, Aline Wuidart, Marielle Ousset,
Alexandra Van Keymeulen, Berthold Göttgens, Cédric Blanpain, Benjamin D. Simons,
and Emma L. Rawlins**

Figure S1, related to Figure 1. Analysis of cell types, proliferation rates and density indicates that the tracheal epithelium is at homeostasis for most of the period under study. (A) Timecourse of experiment. Animals used in homeostasis experiments were all *Tg(KRT5-CreER); Rosa26R-fGFP* and > 8 weeks old at time of tmx exposure. (B) Representative image of epithelium from BrdU incorporation experiment. Red: KRT5 (basal cells), green: BrdU (cell in S-phase). (C, D) Histograms depicting the proportion of (C) basal and (D) secretory cells that have incorporated BrdU following a 2-hour pulse, at 0.5, 12 and 52 weeks in the dorsal tracheal epithelium adjacent to the dorsal longitudinal muscle (blue), or the cartilage (red). (E-J) Representative images of epithelia from cellular composition and density experiments. (E) 0.5 weeks, (F) 3 weeks, (G) 52 weeks adjacent to the longitudinal smooth muscle. (H) 0.5 weeks, (I) 3 weeks, (J) 52 weeks adjacent to the cartilage/inter-cartilage regions of mesenchyme. White: T1 α (basal cells); purple: Scgbl1 (secretory cells); blue: acetylated-tubulin (ciliated cells); green: E-cadherin (lateral cell membranes). (K, L) Histogram depicting (K) the cellular composition and (L) cellular density of the epithelium at 0.5, 12 and 52 weeks, in muscle and cartilage regions. (M) Plot of the mean length of the trachea from the carina to the 6th cartilage ring over time. (N) Plot of the mean width of the trachea at the 3rd cartilage ring above the carina over time. All error bars = 95% confidence intervals. Scale bars = 10 μ m. See Table S1 for all raw data.

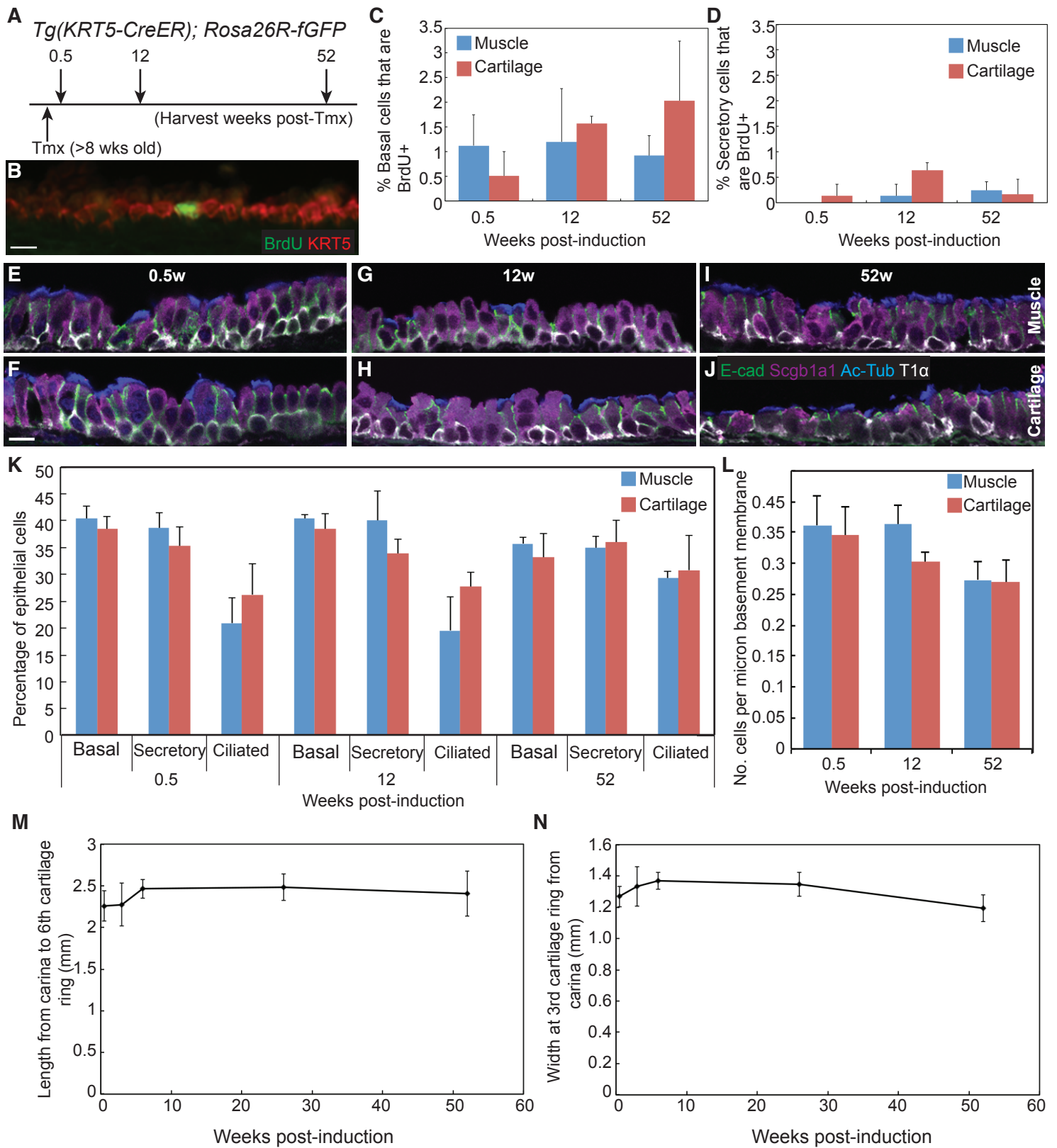
Figure S2, related to Figure 1. Attempts to stratify the *Tg(KRT5-CreER); Rosa26R-fGFP* basal cell clonal data indicate that the dorsal-distal mouse trachea is a homogenous region with one epithelial cell hierarchy. (A) Plot of the mean number of total cells per clone, split into male (blue) and female (red) mice. (B) Plot of the mean number of total cells per clone, with the data split into 3 broad proximal-distal regions based on distance from the carina. Most distal (blue), 2nd most distal (red), 3rd most distal (green). (C) Plot of the mean number of total cells per clone, split into clones over the dorsal longitudinal muscle (blue), or over the cartilage region (red). (D, E) Plot of the mean number of cells of each type per clone (green = basal; blue = secretory; red = ciliated), split into

clones located above dorsal longitudinal muscle (D) and clones located above the cartilage region (E). The graphs in (C-E) suggest that there are small systematic differences in the behaviour of the clones located over the dorsal muscle versus the cartilage, but this was not statistically significant using a Mann-Witney U-test (Supplemental Theory). Error bars in all plots = 95% confidence intervals, with n set to the number of mice.

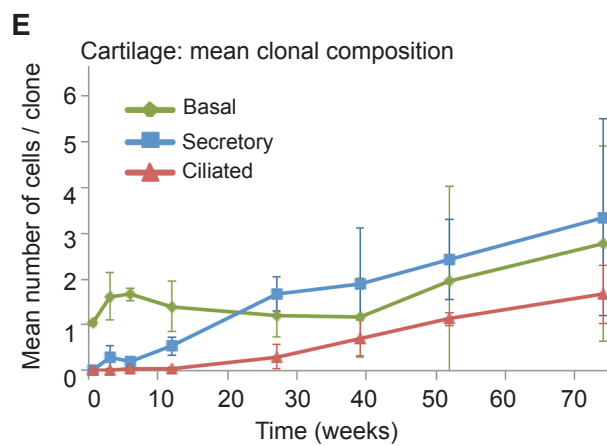
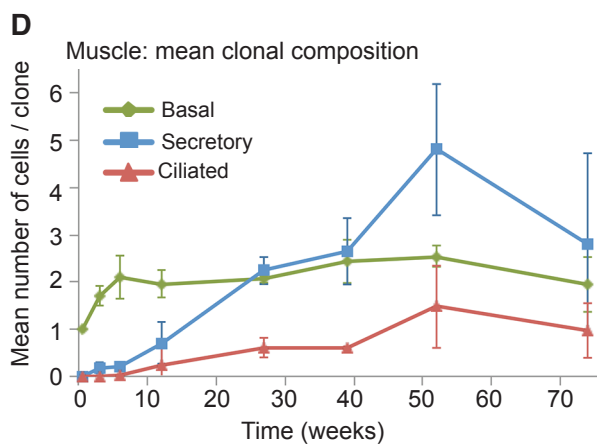
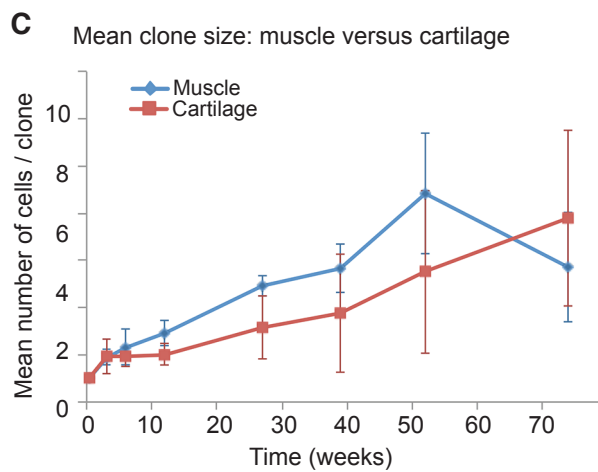
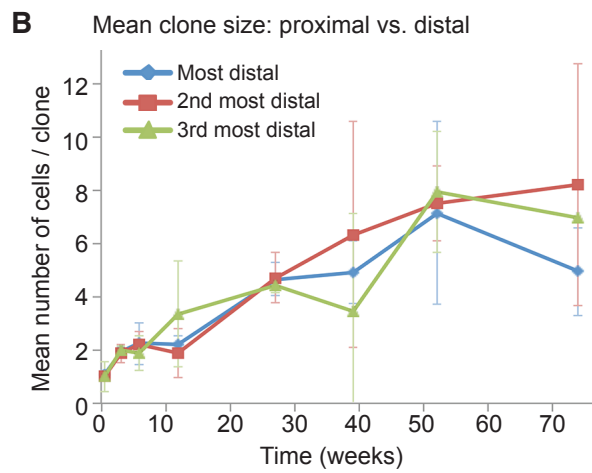
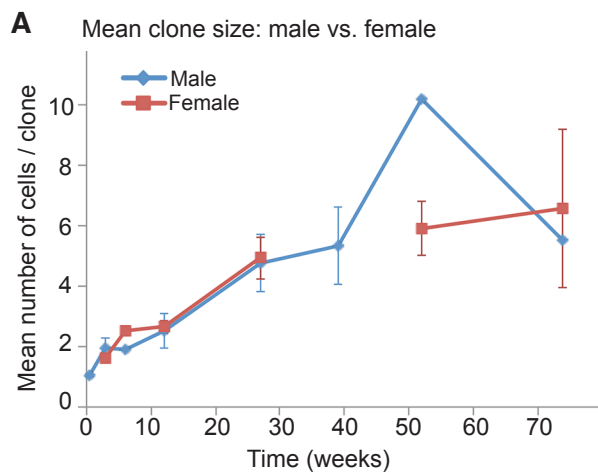
Figure S3, related to Figure 4. Unsupervised hierarchical clustering of all single tracheal epithelial cells which passed quality controls, based on 96 independent qPCR reactions covering 92 genes. The three clusters of cells are labelled ciliated, basal and secretory. These identities were determined by mapping back onto an independent principal component analysis (see Figure 4A, B), and also by pair-wise ANOVA between each group which showed that each group was enriched for definitive markers of the basal, secretory or ciliated epithelial cells (see Table S3).

Table S1, related to Figures 1, 2, 3 and 5. Raw data for all of the clonal analysis, the EdU data set and the *K8-H2B-GFP*.

Table S2, related to Figures 4 and S3. Genes differentially expressed between the three major tracheal epithelial cell groups based on pairwise ANOVA analysis of single cell RT-qPCR data.



Watson et al. Sup Fig 1



Supplementary Fig. 2 Watson et al.

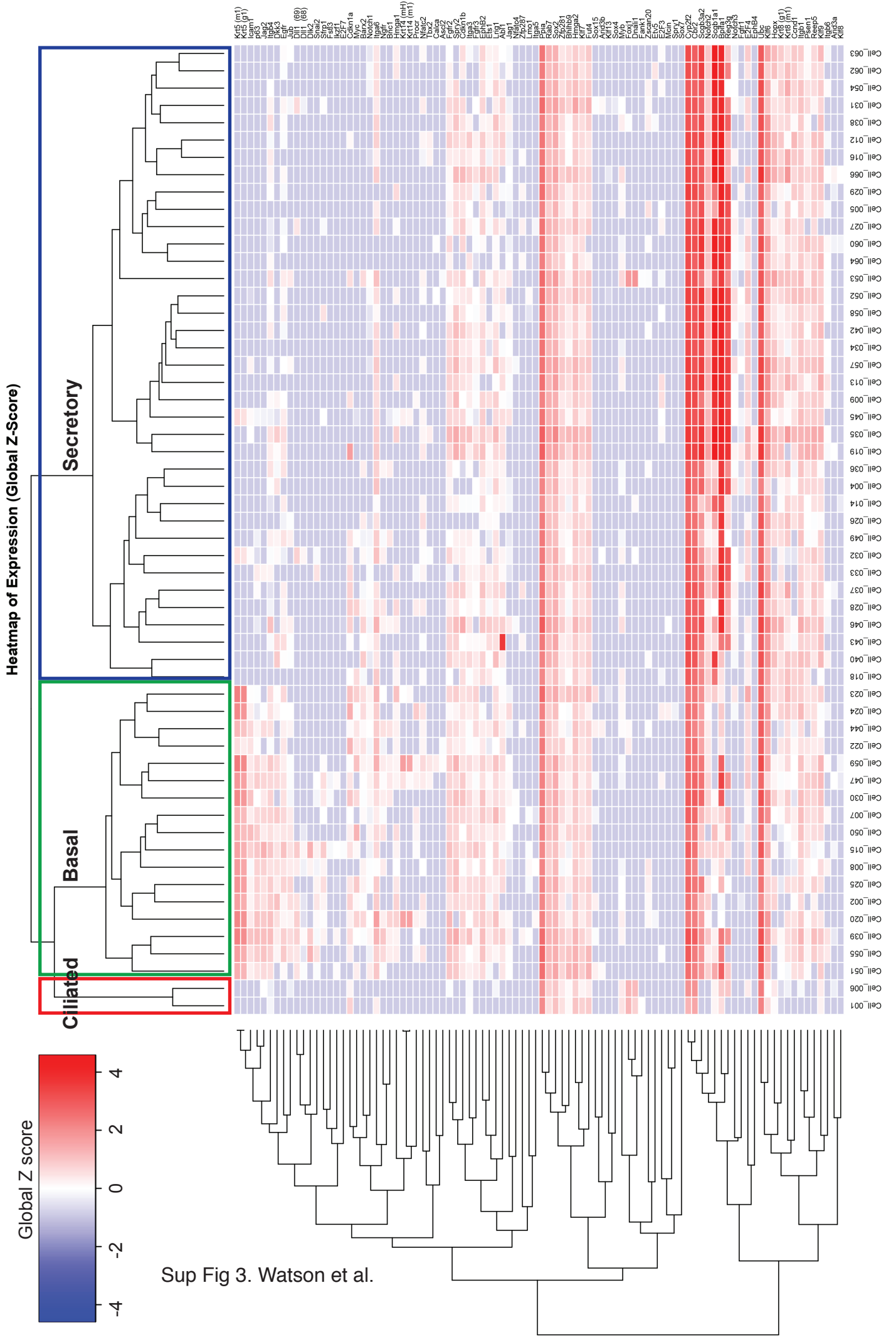


Table S2. Genes differentially expressed between the three major tracheal epithelial cell groups based on pairwise ANOVA analysis of single cell RT-qPCR data. ANOVA was used to compare the mean Log₂ expression level of each gene between the basal, secretory and ciliated cell groups defined in Figure 4A. Genes with an ANOVA p value <0.05 and a >1.2x Log₂ fold-change in mean expression level between two of the cell groups are listed.

Basal cell population-enriched transcripts (p value from comparison of BC vs. secretory populations)	Secretory cell population-enriched transcripts (p value from comparison of BC vs. secretory populations)	Ciliated cell population-enriched transcripts (p value from comparison of BC vs. ciliated populations)
<i>Krt5</i> * (0)	<i>Bpif1a</i> (2.08 x 10 ⁻¹¹)	<i>Foxj1</i> * (1.83 x 10 ⁻⁷)
<i>Trp63</i> * (0)	<i>Scgb1a1</i> (1.78 x 10 ⁻¹⁰)	<i>Dnai1</i> * (2.64 x 10 ⁻⁷)
<i>Jag2</i> * (0)	<i>Cyp2f2</i> (3.93 x 10 ⁻⁹)	<i>Fank1</i> * (2.30 x 10 ⁻⁴)
<i>Icam1</i> * (0)	<i>Krt8</i> (1.19 x 10 ⁻⁷)	
<i>Myc</i> * (1.10 x 10 ⁻⁷)	<i>Cbr2</i> (1.54 x 10 ⁻⁷)	
<i>Igb4</i> (2.84 x 10 ⁻⁷)	<i>Reg3g</i> (2.18 x 10 ⁻⁷)	
<i>Notch1</i> (1.92 x 10 ⁻⁶)	<i>Scgb3a2</i> (1.19 x 10 ⁻⁶)	
<i>Ngfr</i> * (2.22 x 10 ⁻⁶)	<i>Notch2</i> (1.29 x 10 ⁻⁵)	
<i>Jub</i> * (4.81 x 10 ⁻⁶)	<i>Notch3</i> (6.42 x 10 ⁻⁵)	
<i>Snai2</i> * (2.87 x 10 ⁻⁵)	<i>Hopx</i> (2.12 x 10 ⁻³)	
<i>Fgfr2</i> (5.2 x 10 ⁻⁵)	<i>Fgfr1</i> (3.02 x 10 ⁻²)	
<i>Sfrp1</i> * (6.58 x 10 ⁻⁵)		
<i>Egfr</i> (1.09 x 10 ⁻⁴)		
<i>Fgfr3</i> (1.49 x 10 ⁻⁴)		
<i>Procr</i> * (1.75 x 10 ⁻⁴)		
<i>Sox15</i> * (2.81 x 10 ⁻⁴)		
<i>Dkk3</i> (3.27 x 10 ⁻⁴)		
<i>Cdkn1a</i> (7.84 x 10 ⁻⁴)		
<i>Dlk2</i> * (5.91 x 10 ⁻³)		
<i>Barx2</i> (6.08 x 10 ⁻³)		
<i>Hmgal</i> (7.07 x 10 ⁻³)		
<i>Bnc1</i> * (8.05 x 10 ⁻³)		
<i>Dll1</i> * (9.16 x 10 ⁻³)		
<i>Krt14</i> * (2.56 x 10 ⁻²)		
<i>Iga6</i> (3.76 x 10 ⁻²)		
<i>Lmo1</i> * (2.32 x 10 ⁻²)		

EXPERIMENTAL PROCEDURES

Animals. All experiments were approved by University of Cambridge local ethical review committees and conducted according to Home Office project license PPL80/2326. Mouse strains *Tg(KRT5-CreER)* transgenic (Rock et al., 2009), *Krt5-CreER^{T2}* knock-in (Van Keymeulen et al., 2011), *Scgbl1a1-CreER* (Rawlins et al., 2009), *Tg(Krt8-CreER)* (Van Keymeulen et al., 2011), *Rosa26R-fGFP* (Rawlins et al., 2009), *Rosa-confetti* (Snippert et al., 2010), *tetO-H2B-GFP* (Tumbar et al., 2004) have been described previously. Strains used for long-term lineage-tracing: *Tg(KRT5-CreER)*, *Scgbl1a1-CreER*, *Rosa26R-fGFP* were all maintained on a C57Bl/6J inbred background; other strains were at least N2 C57Bl/6J. All animals used for experiments were >8 weeks old. For transgenic experiments a mixture of males and females were used and no gender-specific differences observed. Wildtype C57Bl/6J inbred mice were used for wholemount EdU analysis, mRNA in situ hybridisation and single cell RT-qPCR.

Generation of *Krt8-rtTA* mice. The rtTA fragment from pTetON Advanced plasmid preceded by the β -globin intron and followed by a SV40 polyA signal was subcloned into pBluescript II SK+. The 3.5-kb sequence upstream of the ATG codon of the murine *Krt8* gene, obtained from the BAC clone RO-254K21 (BACPAC Resources Center, Children's Hospital Oakland Research Institute) using the forward primer 5'-GGTGGATCACTTGCCCCCTCCGTTTG-3' and the reverse primer 5'-GGGACAGCGCCCAGCGAAGGCC-3', was cloned upstream of the β -globin intron. The resulting *Krt8-rtTA* fragment of 5.1 kb was released from the backbone by NotI and KpnI digestion and was microinjected into fertilized oocytes to generate transgenic mice (in the transgenic facility of the Université catholique de Louvain, Brussels, Belgium). Seven transgenic founders were first identified by PCR, out of 42 mice born. Expression profiles of

the *Krt8-rtTA* founders were screened with reporter *tetO-H2B-GFP* mice. Five founders expressed GFP in cells expressing the endogenous *Krt8*, and one founder was used throughout this study.

Lineage-tracing. Low frequency activation of the reporter allele was achieved by a single intraperitoneal injection of tamoxifen (Sigma, T5648) dissolved in vegetable oil at a dose of 25 μg per gram body weight in *Tg(KRT5-CreER); Rosa26R-fGFP* mice, or 13 μg per gram body weight in *Tg(Krt8-CreER); Rosa26R-fGFP* mice, or with 2 intraperitoneal doses of 5mg tmx per mouse spaced 48h apart in *Krt5^{CreER/+}; Rosa-confetti*. Doxycycline was administered to *Krt8-rtTA; tetO-H2B-GFP* mice in food at a dose of 10g/kg (SAFE-DIETS) for 2 weeks followed by chase periods of up to 6 weeks.

BrdU/EdU Labelling. BrdU was administered intraperitoneally at a dose of 30 $\mu\text{g/g}$ body weight and tracheas harvested 2 hours post-injection. EdU was administered by intraperitoneal injection at a dose of 50 μg per mouse and tracheas were harvested at 2 hours or 24 hours post-injection. EdU was detected using the Click-it EdU kit (Life Technologies) in wholemount preparations.

Wholemount immunostaining

For mice carrying the *Rosa26R-fGFP* locus and wildtype mice, dissected tracheas were fixed overnight in 4% paraformaldehyde at 4°C followed by PBS washes. Blood vessels and connective tissue were removed manually and trachea were cut into dorsal and ventral halves longitudinally. Tracheal halves were permeabilized in 1% Triton X-100 in PBS for 1 hour at room temperature, followed by PBS washes, blocking overnight, and primary antibody incubation for 2 nights at 4°C in block (block: 1% BSA, 5% Normal Donkey Serum, 0.2%

Triton X-100 in PBS). Washes were performed in PBTx (PBS, 0.2% Triton X-100) and secondary antibodies incubated for 2 nights at 4°C in 5% Normal Donkey Serum, 0.2% Triton X-100 in PBS. DAPI (4',6-diamidino-2-phenylidole) or Hoechst were diluted 1:5,000 in 0.2% Triton X-100 in PBS and incubated for 1 hour at room temperature followed by 0.2% Triton X-100 in PBS washes. Samples were then washed in PBS followed by incubation in increasing concentrations (10%, 25%, 50% for 2h each and 97% overnight) of TDE (2'2'-thiodiethanol) in 0.25x PBS at 4°C and mounting in 97% TDE under a raised coverslip. The following primary antibodies were used: anti-GFP (chicken, 1:1000, Abcam, AB13970), anti-KRT5 (rabbit, 1:500, Covance, PRB-160P), anti-acetylated tubulin (mouse, 1:1000, Sigma, T7451), anti-Pgp9.5 (guinea pig, Neuromics, GP14104). All secondary antibodies were Alexa Fluor conjugated from Life Technologies and used at 1:2000 dilution.

For mice carrying *tetO-H2B-GFP* or *Rosa26R-confetti* loci, the wholemount protocol was adapted to enable direct visualisation of native fluorescence. Anti-GFP staining was omitted, and samples were mounted in Glycergel (Dako) + 2.5% DABCO, rather than TDE. For *Rosa26R-confetti* samples, anti-acetylated-tubulin was detected using an anti-mouse Alexafluor-405 secondary antibody (1:200, Jackson ImmunoResearch), anti-rabbit KRT5 with an Alexafluor-647 secondary antibody (Life Technologies, 1:2000) and DAPI/Hoechst staining was omitted.

Section Immunostaining

For cryoembedding tracheas were fixed in 4% paraformaldehyde at 4°C for 4 hours. Samples were washed in PBS, sucrose protected, embedded in OCT (Optimum Cutting Temperature Compound, Tissue Tek) and sectioned at 8-10 µm. The following primary antibodies were used: anti-acetylated tubulin (mouse, 1:1000, Sigma, T7451), anti-BrdU (mouse, 1:500,

Sigma, B8434), anti-Ecad (rat, 1:3000, Life Technologies, 13-1900), anti-Krt5 (rabbit, 1:500, Covance, PRB-160P), anti-T1 α (1:1000, DSHB, 8.1.1), anti-Scgb1a1 (rabbit, 1:500, Santa Cruz, sc25555). Antigen retrieval was used for BrdU staining by warming samples to 37°C in 2N HCl for 30 minutes, and then incubating in 0.5% trypsin for 5 minutes at room temperature. All secondary antibodies were Alexa Fluor conjugated from Life Technologies and used at 1:2000 dilution. Slides were mounted in Fluoromount (Sigma).

mRNA in situ hybridisation

Trachea were formalin-fixed 24h at room temperature and processed for paraffin embedding via an ethanol gradient. Tissue sections were cut at 5 μ m and processed for RNA in situ detection using the RNA Scope 2-plex Detection Kit (Chromogenic) according to the manufacturer's standard protocol (Advanced Cell Diagnostics). RNAscope probes used were: *Krt5* (NM 027011.2, region 666-2086) which was detected using the Fast Red detection reagent; and *Dlk2* (NM 023932.3, region 267-1279) and *Dll1* (NM 007865.3, region 888-1883) which were detected using the Green detection reagent.

Microscopy and Image scoring

For wholemount samples from mice carrying the *Rosa26R-fGFP* or *Rosa26R-confetti* locus, the study area was first imaged using a 20x lens to identify the locations of clones. Images of clones were then acquired using a 60x oil objective (NA 1.35 or 1.4) on either an Olympus FV1000, or a Leica Sp8. Z-stacks of the full epithelial thickness were acquired at an optical resolution of 1,024 x 1,024 with an optical z slice every 1 μ m. Clones were scored manually by looking through the entire z depth of the tracheal epithelium in FV viewer or LAS AF software to score the identity of all labelled cells. Cells were scored based on the following criteria: basal cells – KRT5⁺; ciliated cells – acetylated tubulin⁺ cilia on their apical surface;

secretory cells – luminal cells with no cilia, in addition auto-fluorescent secretory granules were present in these cells. (This definition of secretory cells means that any labelled brush or neuroendocrine cells, which are present in the trachea at low numbers and can only be distinguished by specific staining, would have been scored as secretory). Cells were scored as within 1 clone if they were within <1 cell diameter of each other.

For wholemount samples from mice carrying the *tetO-H2B-GFP* locus, randomly selected locations within the study area were imaged using a 100x oil objective (NA 1.4) on a Leica SP8. Z-stacks were acquired at an optical resolution of 1,024 x 1,024, with a z-slice every 0.38 μm . Fluorescence intensity was assessed in Fiji, using the Gurdon Institute Imaging Facility's plugin, ObjectScan. Nuclei were defined based on Hoechst staining, and the total GFP signal within the nuclear volume measured. Object maps were checked manually to exclude any nuclei that were mapped incorrectly. Cell types were scored as for the clonal analysis experiments, but with GFP positivity indicating KRT8 expression.

For wholemount samples from wildtype mice used for EdU analysis, the study area was imaged on an Olympus FV1000, using a 40x oil objective (NA 1.3). Z-stacks were acquired at an optical resolution of 1,024 x 1,024, with a z-slice every 1 μm . Cell types were scored as for the clonal analysis experiments, but with EdU incorporation marking cells in S-phase, or their immediate progeny.

Cryosections used for analysis of cellular composition and density were imaged on an Olympus FV1000, using a 100x oil objective (NA 1.4). Z-stacks were acquired at an optical resolution of 1,024 x 1,024, with a z-slice every 0.45 μm . The numbers of cells of each type were assessed using FV1000 software. Cells were scored based on the following criteria:

basal cells – T1 α ⁺; ciliated cells – acetylated tubulin⁺ cilia on their apical surface; secretory cells – Scgb1a1⁺ cells. The length of the basement membrane in each image was measured in Fiji. Density was calculated as the number of cells present per μm of basement membrane.

Cryosections used for BrdU analysis were captured on a Zeiss AxioImager compound microscope, using a 20x air objective (NA 0.8) and counted in Fiji. Cells were scored based on the following criteria: basal cells in S-phase – KRT5⁺BrdU⁺; basal cells not in S-phase – KRT5⁺BrdU⁻; secretory cells in S-phase – KRT5⁻BrdU⁺. The total number of secretory cells was not counted directly but was inferred as being equal to the number of basal cells, based on the cellular composition experiments, which were performed on tracheal sections from the same mice.

Single cell RT-qPCR. The distal trachea was isolated (from the carina up to ~5th cartilage ring), the epithelium was peeled away from the underlying mesenchyme following a brief dispase digest and dissociated to single cells as described (Rock et al., 2009). Unsorted epithelial cells were loaded into a Fluidigm C1 machine on a medium (10-17 μm) chip at a concentration of ~400 cells/ μl for cell capture, lysis, cDNA synthesis and target pre-amplification. 67 single cells were used for subsequent RT-qPCR on a 96.96 Fluidigm Dynamic array using a Biomark qPCR machine. Taqman gene expression assays (Life Technologies) were used for expression analysis (Supplemental Experimental Procedures). Data analysis was performed in the Fluidigm Singular Analysis Toolset 3.0 in R. Briefly, cells with aberrant transcript levels were excluded using the Outlier analysis function based on the median levels of 23 highly expressed genes, and also by manual curation based on expression of <4 of the 5 housekeeping genes included in the analysis. 56 single cells passed

these quality controls and Principle Component Analysis, unsupervised hierarchical clustering, violin plots and ANOVAs were all performed using Singular.

Modelling. See theoretical supplement for all details of model construction. Note that for the analysis of the EdU data (Figure 3H) only single EdU⁺ cells at 2 hours and pairs of EdU⁺ cells at 24 hours were included in the comparison to the model. The full data set included some groups of >2 cells and some single cells at the 24 hour time point, likely resulting from clone merging or EdU toxicity respectively. The complete data set is available in Table S1.

Rawlins, E.L., Okubo, T., Xue, Y., Brass, D.M., Auten, R.L., Hasegawa, H., Wang, F., and Hogan, B.L. (2009). The role of Scgb1a1⁺ Clara cells in the long-term maintenance and repair of lung airway, but not alveolar, epithelium. *Cell stem cell* *4*, 525-534.

Rock, J.R., Onaitis, M.W., Rawlins, E.L., Lu, Y., Clark, C.P., Xue, Y., Randell, S.H., and Hogan, B.L. (2009). Basal cells as stem cells of the mouse trachea and human airway epithelium. *Proc Natl Acad Sci U S A* *106*, 12771-12775.

Snippert, H.J., van der Flier, L.G., Sato, T., van Es, J.H., van den Born, M., Kroon-Veenboer, C., Barker, N., Klein, A.M., van Rheenen, J., Simons, B.D., *et al.* (2010). Intestinal crypt homeostasis results from neutral competition between symmetrically dividing Lgr5 stem cells. *Cell* *143*, 134-144.

Tumbar, T., Guasch, G., Greco, V., Blanpain, C., Lowry, W.E., Rendl, M., and Fuchs, E. (2004). Defining the epithelial stem cell niche in skin. *Science* *303*, 359-363.

Van Keymeulen, A., Rocha, A.S., Ousset, M., Beck, B., Bouvencourt, G., Rock, J., Sharma, N., Dekoninck, S., and Blanpain, C. (2011). Distinct stem cells contribute to mammary gland development and maintenance. *Nature* *479*, 189-193.

Supplemental Experimental Procedures, Table 1. List of 96 Taqman assays used for single cell RT-qPCR.

Category	Gene name	Taqman Assay ID
Housekeeping	<i>Abll</i>	Mm00802029_m1
	<i>Ppia</i>	Mm02342429_g1
	<i>Rab7</i>	Mm01183732_g1
	<i>Reep5</i>	Mm00492230_m1
	<i>Ubc</i>	Mm02525934_g1
All columnar cells	<i>Krt8</i>	Mm04209403_g1
	<i>Krt8</i>	Mm00835759_m1
Secretory cell specific	<i>Bpifa1</i>	Mm00465064_m1
	<i>Cbr2</i>	Mm01246806_g1
	<i>Cyp2f2</i>	Mm00484087_m1
	<i>Fut4</i>	Mm00487448_s1
	<i>Reg3g</i>	Mm01181783_g1
	<i>Scgb1a1</i>	Mm00442046_m1
	<i>Scgb3a2</i>	Mm00504412_m1
Putative ciliated cell specific	<i>Dnal1</i>	Mm00613749_m1
	<i>Fank1</i>	Mm00482221_m1
	<i>Foxj1</i>	Mm01267279_m1
	<i>Mcin</i>	Mm01308202_m1
	<i>Myb</i>	Mm00501741_m1
NE cell specific	<i>Ascl1</i>	Mm04207567_g1
	<i>Calca</i>	Mm00801463_g1
BC-specific, or previously reported as enriched in BCs	<i>Arid3a</i>	Mm00492251_m1
	<i>Arid3b</i>	Mm01179426_m1
	<i>Ascl2</i>	Mm01268891_g1
	<i>Barx2</i>	Mm01176203_m1
	<i>Bhlhb9</i>	Mm01701581_s1
	<i>Bnc1</i>	Mm01324337_m1
	<i>CEBPg</i>	Mm01266786_m1
	<i>EphB2</i>	Mm01181021_m1
	<i>EphB4</i>	Mm01201157_m1
	<i>Fstl3</i>	Mm01168957_m1
	<i>Hmgal</i>	Mm04213075_u1
	<i>Hmgal2</i>	Mm04183367_g1
	<i>Hopx</i>	Mm00558630_m1
	<i>Icam1(CD54)</i>	Mm00516023_m1
	<i>Ikzf1</i>	Mm01187881_m1
	<i>Itga3</i>	Mm00442910_m1
	<i>Itga5</i>	Mm00439820_m1
	<i>Itga6</i>	Mm00434375_m1
	<i>Itgb1</i>	Mm01253230_m1
	<i>Itgb4</i>	Mm01266840_m1
	<i>Itgb6</i>	Mm01269869_m1
	<i>Jub</i>	Mm00495049_m1
	<i>Klf6</i>	Mm00516184_m1
	<i>Klf7</i>	Mm00728361_s1

	<i>Klf8</i>	Mm01353192_m1
	<i>Klf9</i>	Mm00495172_m1
	<i>Klf13</i>	Mm00727486_s1
	<i>Krt5</i>	Mm01305291_g1
	<i>Krt5</i>	Mm00503549_m1
	<i>Krt14</i>	Mm00516876_m1
	<i>Krt14</i>	Mm00516870_mH
	<i>Lmo1</i>	Mm00475438_m1
	<i>Myc</i>	Mm00487804_m1
	<i>Nfatc2</i>	Mm00477776_m1
	<i>Nfatc4</i>	Mm00452373_g1
	<i>Ngfr</i>	Mm01309638_m1
	<i>Snai2</i>	Mm00441531_m1
	<i>Sox2</i>	Mm03053810_s1
	<i>Sox4</i>	Mm00486317_s1
	<i>Sox7</i>	Mm00776876_m1
	<i>Sox15</i>	Mm00839542_g1
	<i>Tbx2</i>	Mm00436915_m1
	<i>Trp63</i>	Mm00495791_m1
	<i>Zfp287</i>	Mm01334842_m1
	<i>Zfp213</i>	Mm01309214_m1
	<i>Zscan20</i>	Mm00625080_m1
	<i>Zfp281</i>	Mm01296016_s1
Cell cycle regulators	<i>Ccnd1 (CycD)</i>	Mm00432359_m1
	<i>Cdkn1a (p21)</i>	Mm04205640_g1
	<i>Cdkn1b (p27)</i>	Mm00438168_m1
	<i>E2F3</i>	Mm01138833_m1
	<i>E2F4</i>	Mm00514160_m1
	<i>E2F7</i>	Mm00618098_m1
Notch signalling	<i>Dll1</i>	Mm01279269_m1
	<i>Dll1</i>	Mm01279268_m1
	<i>Dlk2</i>	Mm01281511_g1
	<i>Jag1</i>	Mm00496902_m1
	<i>Jag2</i>	Mm01325629_m1
	<i>Notch1</i>	Mm00435249_m1
	<i>Notch2</i>	Mm00803062_m1
	<i>Notch3</i>	Mm01345646_m1
	<i>Psen1</i>	Mm01297739_m1
RTK signalling	<i>Egfr</i>	Mm00433023_m1
	<i>Ets1</i>	Mm01175819_m1
	<i>Etv4</i>	Mm01245872_m1
	<i>Etv5</i>	Mm00465816_m1
	<i>Fgfr1</i>	Mm00438930_m1
	<i>Fgfr2</i>	Mm01269930_m1
	<i>Fgfr3</i>	Mm00433294_m1
	<i>Lrig1</i>	Mm00456116_m1
	<i>Spry1</i>	Mm00839969_g1
	<i>Spry2</i>	Mm00442344_m1

Wnt signalling	<i>Dkk3</i>	Mm00443800_m1
	<i>Procr (CD201)</i>	Mm00440993_mH
	<i>Sfrp1</i>	Mm00489161_m1

Supplementary Note

In this Supplementary Note we present further details of the biophysical and statistical approach used to infer the lineage hierarchy and proliferation kinetics of progenitors in the trachea.

1 Stochastic Modelling

To define cell fate behaviour in the trachea we extend the biophysical modelling approach introduced in Refs. ^{1,2}. In this model, stem cells and their differentiating progenitor cell progeny function as equipotent cell populations in which their proliferation kinetics and fate behaviour follow defined statistical dependencies. These stochastic “rules” summarize the product of the complex gene regulatory processes that regulate cell fate behaviours. For simplicity, following Refs. ^{1,2}, we consider a model based on intrinsic (cell-autonomous) regulation. However, in the two-dimensional geometry of the tracheal epithelium, models of fate behaviour based on external (cell extrinsic) regulation are difficult to discriminate from intrinsically regulated systems. The clonal analysis described here therefore leaves the question of the underlying regulatory dynamics unspecified.

In the following we make use of this stochastic modelling approach to infer the simplest model which is capable of describing the complex range of clonal dynamics observed in experiment. First, we analyse the dynamics in the basal layer making use of the *Tg(KRT5-CreER); Rosa26R-fGFP* line. We find that the basal population consists of two subtypes - a dividing basal stem cell (BSC) population and transient luminal precursors (BLPs). Independently, we then employ lineage

tracing data from *Scgbl1a1-CreER; Rosa26R-fGFP* mice to infer the dynamics in the luminal layer. We find that secretory cells (SecCs) are lost at a surprisingly high rate. Finally, we compare these findings with the full clonal data from *Tg(KRT5-CreER); Rosa26R-fGFP* mice, as well as *Krt5-CreER; Rosa26R-Confetti* and *K8-CreER; Rosa26R-fGFP* mice, and independent BrdU and EdU measurements.

1.1 Maximum likelihood estimation

The statistical approach we employ to disclose the dynamical rules that underlie fate determination in the trachea is based on the idea of *Bayesian inference*³. By applying this approach we seek the model that has the highest probability to describe the given experimental outcome. In other words, we aim to calculate the probability $P(\Theta|E)$ of a model (defined by dynamical rules and a set of parameters $\Theta = (\Theta_1, \Theta_2, \dots)$) taking into account the experimental evidence E . According to *Bayes' Theorem*, this probability is proportional to the product of the probability of obtaining the experimental evidence if we assume a certain model to be true, $P(E|\Theta)$, and our *a priori* belief in a certain model, $P(\Theta)$:

$$P(\Theta|E) = \frac{P(E|\Theta)P(\Theta)}{\int_{\Theta} d\Theta P(E|\Theta)P(\Theta)}. \quad (1)$$

The integral in the denominator ensures normalisation. We define the likelihood function $\mathcal{L}_E(\Theta) \equiv P(E|\Theta)$ by treating $P(E|\Theta)$ as a function of the models, Θ , for a given set of experimental evidence, E . The principle of maximum likelihood then states that, if we do not have any *a priori* belief in the validity of any model, i.e. the distribution $P(\Theta)$ is uniform, the model with the highest probability is given by the maximum of the likelihood function $\mathcal{L}_E(\Theta)$.

In our case, the experimental evidence is given by the frequency $f_{n_b, n_s, n_c, t}$ of observations of clones with given numbers of basal, secretory, and ciliated cells (n_b, n_s, n_c) at a given time point t after induction. Since these measurements are statistically independent, the likelihood function factorises into single observations of clones, $\mathcal{L}_{f_{n_b, n_s, n_c, t}}(\Theta) = \mathcal{N}^{-1} \prod_t \prod_{n_b, n_s, n_c} P(n_b, n_s, n_c, t | \Theta)^{f_{n_b, n_s, n_c, t}}$. The prefactor is determined by normalisation, $\mathcal{N}^{-1} = \left(\sum_{n_b, n_s, n_c, t} f_{n_b, n_s, n_c, t} \right)! / \prod_{n_b, n_s, n_c, t} f_{n_b, n_s, n_c, t}!$, where the product and sum go over all measured time points t and all cellular compositions (n_b, n_s, n_c) . Hence, the likelihood function takes the form of a multinomial distribution,

$$\mathcal{L}_E(\Theta) = \frac{\left(\sum_{n_b, n_s, n_c, t} f_{n_b, n_s, n_c, t} \right)!}{\prod_{n_b, n_s, n_c, t} f_{n_b, n_s, n_c, t}!} \prod_t \prod_{n_b, n_s, n_c} P(n_b, n_s, n_c, t | \Theta)^{f_{n_b, n_s, n_c, t}}, \quad (2)$$

How do we obtain the likelihood function $\mathcal{L}_E(\Theta)$? The dynamics of the stochastic models we consider here is described by *Master equations* of the form

$$\frac{d}{dt} P(n, t | \Theta) = \sum_{n'_b, n'_s, n'_c=0}^{\infty} \mathcal{W}_{n', n} P(n', t | \Theta) - \mathcal{W}_{n, n'} P(n, t | \Theta), \quad (3)$$

determining the time evolution of the probability $P(n, t | \Theta)$ to find the clonal composition $n = (n_b, n_s, n_c)$ at time t . The quantities $\mathcal{W}_{n, n'}$ are the transition rates between states with different cellular composition and therefore encode the details of the model Θ under consideration. In the simplest cases, these equations can be solved analytically, such that $P(n_b, n_s, n_c, t | \Theta)$ is explicitly known. Generally, however, Master equations can only be solved numerically. Here, we employ stochastic simulations using *Gillespie's algorithm*⁴, which allows us to compute single realisations of the stochastic dynamics. By aggregating these realisations into histograms we are able to obtain an approximation to the probability density $P(n_b, n_s, n_c, t | \Theta)$. However, in lineage tracing experiments it is not feasible to measure the number of clones that were lost before the mice are analysed.

We therefore effectively measure a conditional (persistence) probability distribution $P_{n>0}^p(n, t|\Theta)$, which is obtained by dividing P by the probability of finding a clone of size larger than zero,

$$P_{n>0}^p(n, t|\Theta) = \frac{P(n, t|\Theta)}{1 - P(0, t|\Theta)}. \quad (4)$$

These solutions of the Master equations give the probability to find any stochastic fate of a single clone. By substituting a particular experimental outcome into these solutions we therefore obtain the value of the likelihood function for a given model Θ and the logarithm of the likelihood function is given by

$$\log \mathcal{L}_E(\Theta) = \log \mathcal{N}^{-1} + \sum_t \sum_{n_b, n_s, n_c} N_{n_b, n_s, n_c, t} \log P_{n>0}^p(n_b, n_s, n_c, t|\Theta). \quad (5)$$

In principle, to calculate $\log \mathcal{N}^{-1}$ we could made use of *Stirling's approximation*, $\log n! \approx n \log n - n$, for the numerator. However, the term $\log \mathcal{N}^{-1}$ is independent of the model Θ . Since we are here only interested in the ratios of likelihood functions (or differences in the log likelihoods) we henceforth neglect the normalisation factor and report the unnormalised likelihood function $\tilde{\mathcal{L}}_E(\Theta) = \mathcal{N} \mathcal{L}_E(\Theta)$. To determine the maximum of the likelihood function we first searched parameter space using a simulated annealing algorithm to roughly estimate the location of the most likely parameters. Once the location of the maximum likelihood parameters had been narrowed down, we rasterised parameter space surrounding these optimal parameters to precisely determine the best fit parameters and to calculate credible intervals.

1.2 Uncertainty and credibility intervals

The maximum likelihood estimation is subject to various sources of uncertainty, most importantly finite sample sizes of experimental and numerical data. In addition, there is a significant variability

between mice. To account for this uncertainty we give parameter estimates with margins denoting the uncertainty, which is associated with these values. Specifically, in the spirit of a likelihood ratio test, we consider all parameter values as acceptable whose likelihood is above a given threshold, α , such that the resulting credible region (CR) is

$$\text{CR} = \left\{ \Theta \mid \frac{\mathcal{L}_E(\Theta)}{\mathcal{L}_E(\Theta^*)} > \alpha \right\}. \quad (6)$$

It has been shown that this approach is most powerful in classifying sets of credible parameters³.

We here choose $\alpha = 0.05$, such that the credible region includes parameter values whose likelihood is larger than 5% of the maximum likelihood value. The dynamics of cells in the trachea is defined by several parameters. Hence, for a specific value of one of these parameters, Θ_i , we give credible intervals in the form $\Theta_i^* \pm (\max_{\text{CR}}\{\Theta_i\} - \Theta_i^*, \Theta_i^* - \min_{\text{CR}}\{\Theta_i\})$.

1.3 Regional variability

Clones were regionally specified according to whether they were located over cartilage rings or over the dorsal longitudinal muscle. To test if the clonal dynamics in these two regions differs significantly, we calculated the probability that for each time point and cell type the resulting probability distributions are equal. To calculate this probability we employed a Mann-Witney U-test. The resulting p-values are given in Table 1. From 24 different statistical tests for all time points and cell types, only three indicated statistically significant differences (p-value < 0.05). Two of these three involve mice analysed at 1 year after induction. Inspection of the average clone

Time [days]	3	21	42	85	188	275	366	517
Basal	-	0.22	0.02	0.40	0.42	0.28	0.002	0.22
Secretory	-	0.44	0.45	0.15	0.07	0.18	0.60	0.66
Ciliated	-	-	0.20	0.75	0.92	0.16	0.01	0.33

Table 1: P-values for comparing clone sizes observed in regions overlying cartilage rings and clones overlying dorsal longitudinal muscle band.

sizes for each time point (Fig. 1g) reveals that the 1 year time point is likely not representative for the overall clonal dynamics. Since most deviations between the different regions in the trachea can therefore be explained by statistical variation it is reasonable to assume that there are no significant differences in the mechanisms of fate choice. In the following we therefore pool the data for both regions.

1.4 Clonal merging

To analyse the clonal fate data it is important to take into account possible ambiguity due to clonal mergers. In the following we estimate the frequency of mergers in regions overlying the dorsal longitudinal muscle band in which clones have the highest induction frequency and are therefore particularly prone to merging events. For muscle regions we find roughly 40 patches in an area of $630 \mu\text{m} \times 630 \mu\text{m}$, which translates to a typical distance between clones of about $a \approx 100 \mu\text{m}$. With this, the distance r between neighbouring clones follows a Rayleigh distribution, $G(r) = 2r/a^2 \exp[-(r/a)^2]$. The probability of merging of two clones of size R is then given by the probability that a randomly chosen clone overlaps with its nearest neighbour. This

probability is just the cumulative distribution, i.e. the probability that their distance is smaller than the sum of their radii, $P_R^{\text{merge}} = \int_0^R G(r) dr = 1 - \exp[-(R/a)^2]$. Since a single cell covers an area of approximately $s = 120 \mu\text{m}^2$ we can re-express the probability of merging as $1 - \exp(-ns/a^2)$. For an average clone size of 10 cells this yields a probability of merging of roughly 10%, indicating that clonal mergers are rare but not negligible. However, this simple estimation does not take into account the fact that large clones occur at a lower frequency than small clones. In the absence of mergers we expect the clone sizes to be exponentially distributed. Mergers have a larger clone size than expected from an exponential distribution and the clone size above which mergers become relevant is characterized by the deviation of the empirical clone size distribution from the exponential form. Following this approach we therefore neglected all clones that had more than 19 cells of a given type.

For the *Scgblal-CreER* experiment, the clonal density was considerably lower, with on average 37 clones per mm^2 in muscle areas and 29 clones per mm^2 in cartilage areas. While the exact form of the distribution of clone sizes is hard to read from the data, we again approximated this distribution by an exponential with an average clone size of $\langle n \rangle \approx 1.5$, roughly independent of time. Since luminal cells have a similar area in the xy plane as basal cells ($120\mu\text{m}^2$) we find that the frequency of mergers is roughly 2% in both areas. For the *Scgblal-CreER* dataset we may therefore neglect the influence of merging events.

1.5 Inference of the best fitting model

We are now in a position to employ statistical inference in order to estimate the model, which describes the experimental data with the highest likelihood. Since the trachea consists of several cell types, each potentially exhibiting a variety of different behaviors, the parameter space we have to search for the best model is high-dimensional, which renders a simultaneous estimation of the maximum likelihood parameters impossible. To make progress we therefore employ different lineage tracing experiments to independently infer the dynamics in the basal and the luminal layer. Moreover, the range of possible models is greatly constrained by the requirement that the lung tissue is homeostatic in adult mice.

1.5.1 Basal layer

We start our analysis with the basal layer. In the *Tg(KRT5-CreER); Rosa26R-fGFP* line, basal cells are genetically labelled. This label is, upon cell division, inherited by all progeny of these cells. If we neglect the loss of clones we can interpret the average number of basal cells in analysed clones as the average number of basal cells in *all* clones. In homeostasis, we would expect that once a basal cell is labelled the number of basal cells, averaged over all clones, remains constant at a value of unity. What we find, however, is an increase of the average number of basal cells to a value of two, cf. Fig. 1g. This behaviour is not consistent with a single basal population in a homeostatic tissue. Rather, this increase in the average number of basal cells strongly suggests that the basal population consists of at least two cell types of which only one is labeled.

To describe the dynamics in the basal layer we therefore consider two cell populations: basal stem cells (BSCs), denoted by A , and basal luminal precursors (BLPs), denoted by B . The existence of two basal populations will later be independently supported by our statistical analysis. Further, as noted in the main text, we unambiguously identified this second population via the genetic marker K8. In adult tissue, self-renewal of stem cells must be tightly balanced with differentiation to maintain homeostasis. In principle, there are two possible ways in which this balance can be achieved: on the level of single cells by asymmetric divisions, or on the population level, if symmetric divisions and differentiation happen at exactly the same rate. Taking into account both modes of division we are left with the following possible dynamical rules for basal stem cells:



where homeostasis dictates that $\alpha_1 = \alpha_3$.

What are the dynamical rules governing the behaviour of the second basal population? As discussed below, we can rule out asymmetric divisions of the second basal population from EdU experiments. Also, to maintain homeostasis, basal progenitors must differentiate at the same rate that they are created and leave the basal layer. For now, we leave the details of this differentiation process to a later discussion. We note that differentiation via asymmetric division does not change the number of basal progenitors and that any direct differentiation processes can be effectively

summarised by loss of basal progenitors:



If N_b is the total number of basal cells in a sufficiently large frame of tissue, the loss of basal progenitors per unit time and per basal cell is given by the rate of loss of single basal progenitors, β , times the fraction of basal progenitors in the tissue, b , where the concentration b is defined such that $N_b \cdot b$ is the total number of basal progenitors. This overall loss rate of progenitors, βb , must exactly match the overall production rate of progenitors through differentiation of basal stem cells, $a(2\alpha_1 + \alpha_2)$. In homeostasis, we therefore require

$$\beta = \frac{a}{b} [2\alpha_1 + \alpha_2], \quad (11)$$

where a/b denotes the relative abundances of basal stem cells and basal progenitors.

For the basal layer we are left with three parameters to be determined by statistical inference: α_1 , α_2 , and a/b . Performing the maximum likelihood estimation we find that the best fitting parameters correspond to a log likelihood of $\log \mathcal{L}_E(\Theta^*) = -2077$, with rates

$$\begin{aligned} \alpha_1 &= 0.0029 \pm (0.0006, 0.0008) \text{ d}^{-1}, \\ \alpha_2 &= 0.085 \pm (0.05, 0.025) \text{ d}^{-1}, \\ \beta &= 0.095 \pm (0.062, 0.037) \text{ d}^{-1}, \\ a/b &= 1.05 \pm (0.25, 0.25) \text{ d}^{-1}. \end{aligned} \quad (12)$$

The uncertainty for α_2 is much larger than for α_1 since, after a short time, the clonal dynamics approaches a scaling regime, where the distribution of the size of persistent basal clones takes a

simple scaling form ²,

$$\lim_{t \gg 1/\alpha_1} P_{n>0}^{\text{pers}} = (\alpha_1 t)^{-1} f [(\alpha_1 t)^{-1} n]. \quad (13)$$

This is independent of the parameter α_2 which must therefore be determined from the early time points alone.

In summary, we find that basal stem cells divide on average every $11 \pm (4, 4)$ d. Of these divisions, the vast majority, $93.6 \pm (2.7, 2.3)\%$, result in asymmetric fate outcome. A basal progenitor differentiates on average every $11 \pm (7, 4)$ d.

Having determined the optimal parameters defining a model for the basal layer involving two basal populations we now ask whether the basal dynamics can also be described by a model involving a homogeneous basal population. To this end we consider a single population, A , undergoing symmetric and asymmetric divisions: $A \rightarrow AA$ and $A \rightarrow A\emptyset$, where the second process summarises all loss of basal cells through differentiation or cell death. In homeostasis, both processes must be precisely balanced and happen at the same rate, α . Again, asymmetric divisions do not change the number of basal cells and may therefore be neglected. Statistical inference reveals that, given a homogeneous basal population, the maximum likelihood parameter is $\alpha = 0.016 \pm (0.0007, 0.0018) \text{ d}^{-1}$ with a corresponding log likelihood of $\log \tilde{\mathcal{L}}_E(\Theta^*) = -2325$. Comparing this result to the maximum likelihood estimate obtained for the heterogeneous basal population (-2077), the difference corresponds to a Bayes factor of $K = 5 \cdot 10^{107}$ or $2 \ln K = 496$, which is highly decisive in favour of a heterogeneous basal population.

While our simple model provides an excellent prediction of the clonal data, we cannot rule out

the possibility of a more complex dynamics in the basal layer. Indeed, many more complicated models yield clone size distributions which quickly relax to the same scaling form that arises from our simple model ². However, we can make progress by making use of the fact that the overall rate of cell divisions is constrained by BrdU measurements to roughly 0.09 d^{-1} . By keeping the overall rate of cell divisions fixed we first asked whether reversibility, i.e. the “de-differentiation” of BLPs to BSCs ($B \rightarrow A$) could capture the clonal data equally well. Our stochastic simulations suggested that the rate of such reversibility processes, would they exist, must be lower than 0.001 d^{-1} . It therefore seems unlikely that BLPs de-differentiate to BSCs, at least under conditions of normal homeostasis.

We then asked whether the clonal data is compatible with proliferating BLPs, again fixing the overall frequency of cell divisions in the basal layer. To this end, we extended our model by a process involving symmetric divisions of BLPs, $B \rightarrow BB$, while adjusting the rate of loss of BLPs such that homeostasis is maintained. From the clonal data we cannot rule out symmetric divisions of BLPs. However, their rate is constrained to less than 0.045 d^{-1} , which is less than 50% of the total number of cell divisions in the basal layer. In particular, this suggests that the BLPs are not a traditional transit amplifying cell population. This view is supported by H2B-GFP measurements, as discussed below.

1.5.2 Luminal layer

The relatively large degree of uncertainty in the rate of differentiation from the basal layer, β , makes straightforwardly extending the basal dynamics to the luminal layer impossible. We therefore performed independent experiments making use of a *Scgblal* promoter, which labels secretory cells and their progeny. With SeCs induced sporadically at a constant rate, this translates to a mathematical theory of branching processes with immigration⁵, which predicts that the observed clone size distributions should be stationary. Indeed from the clonal data, we find that the probability distributions of cell numbers in the luminal layer are essentially time-independent.

We first note that there are no clones involving basal cells and we can therefore exclude the possibility of de-differentiation of secretory cells in homeostasis. The probability distribution $\tilde{P}(n_s, n_c, t)$ of finding n_s secretory cells and n_c ciliated cells at a time t after induction can then be obtained by integrating over the time course,

$$\tilde{P}(n_c, n_s, t) = \int_0^t d\tau P(n_s, n_c, t - \tau) Q(t - \tau), \quad (14)$$

where $P(n_s, n_c, t - \tau)$ denotes the distribution of clones at a time $t - \tau$ after induction and $Q(t - \tau)$ is the probability that a clone is induced at a time $t - \tau$. We here set $Q(t - \tau) = 1/t$, i.e. clones are continuously induced with the same probability at each time point.

To begin we focus on secretory cells, S , alone. The range of possible dynamics is constrained by a remarkably low loss rate at the end of the lineage: a ciliated cell is, on average, lost once every 6 months, translating to a loss rate of $\gamma = 0.0038 \text{ d}^{-1}$, which is by a factor of 20 smaller

than the rate of cell divisions found for the basal layer. This already restricts the rate of potential divisions of ciliated cells, as well as the total rates differentiation to ciliated cells to less than 0.0038 d^{-1} . Hence, within a statistical analysis these rates can not be distinguished from 0 and we may therefore neglect them in the ensuing analysis. Since we know from experiments that ciliated cells do not divide we end up with the following dynamical processes for the luminal layer:



Fortunately, homeostasis again constrains the number of free parameters. In order for the population of secretory cells to remain constant, their total production rate, $s\sigma_1 + b\beta$, must equal their total loss rate, $s(\sigma_3 + \sigma_4)$, i.e.

$$\sigma_3 = \frac{b}{s}\beta + \sigma_1 - \sigma_4. \quad (20)$$

The ratio of basal progenitors to secretory cells can not be directly inferred from cell counting as the morphology of basal stem cells and basal progenitors is indistinguishable. However, with the ratio of all basal stem cells and basal progenitors, $a/b \approx 1.05$ and the ratio of basal cells to secretory cells, $(a + b)/s \approx 1$, we can rewrite b/s in terms of known quantities, namely the ratio of basal stem cells to progenitors and the ratio of basal cells to secretory cells,

$$\frac{b}{s} = \frac{b}{a+b} \frac{a+b}{s} = \frac{a+b}{s} \frac{1}{a/b + 1}. \quad (21)$$

Since σ_3 should be greater or equal to zero, the condition of homeostasis not only fixes the value of σ_3 but also constrains the loss rate of secretory cells,

$$\sigma_4 \leq \frac{b}{s}\beta + \sigma_1. \quad (22)$$

Stationarity of the ciliated cell population requires that

$$\sigma_2 = \frac{c}{s}\gamma - 2\sigma_3, \quad (23)$$

and making use of Eq. (20), we obtain

$$\sigma_2 = \frac{c}{s}\gamma - 2 \left[\frac{b}{s}\beta + \sigma_1 - \sigma_4 \right] \quad (24)$$

Since σ_2 must be greater than zero we obtain another constraint on the value of the loss of secretory cells,

$$\sigma_4 \geq \frac{b}{s}\beta + \sigma_1 - \frac{\gamma c}{2s} \quad (25)$$

Together with Eq. (22) this constrains possible values of σ_4 to an interval of length $\gamma c/(2s)$. Indeed, since it has been found that 50% of ciliated cells are lost within six months, i.e. $\gamma \approx 0.0038 \text{ d}^{-1}$, the width of this interval is much smaller than the actual value of σ_4 and its uncertainty, which is mostly determined by the uncertainty in β . Since according to EdU measurements there is no indication for symmetric, differentiating divisions of secretory cells we fixed the value of σ_4 such that σ_3 vanishes, namely to the upper end of this interval,

$$\sigma_4 \approx \frac{b}{s}\beta + \sigma_1 \quad (26)$$

We finally obtain

$$\begin{aligned}\sigma_2 &\approx \frac{c}{s}\gamma, \\ \sigma_3 &\approx 0, \\ \sigma_4 &\approx \frac{a+b}{s} \frac{1}{a/b+1} \beta + \sigma_1,\end{aligned}\tag{27}$$

with the only remaining parameter to be determined by statistical inference being the rate of symmetric divisions of secretory cells, σ_1 . Already at this point we can therefore say that the rate of asymmetric divisions and the rate of symmetric differentiation is small, while the loss of secretory cells compensates for potential symmetric divisions and the differentiation of basal cells.

To determine the rate of symmetric divisions of secretory cells, σ_1 , we performed stochastic simulations to obtain its maximum likelihood estimate. We find that the value of maximum likelihood is $\sigma_1 = 0.037 \pm (0.010, 0.008)$. Hence, a secretory cell divides on average every $25 \pm (7, 5)$ d. $93 \pm (2, 2)\%$ of these divisions are symmetric resulting in two secretory cells, while $8.6 \pm (2.9, 1.7)\%$ lead to asymmetric fate outcome. The proliferative capacity of secretory cells is compensated by a large loss rate: a secretory cell is lost on average every $14 \pm (2, 2)$ d.

1.5.3 The full lineage

Having defined the dynamics in the basal layer and the luminal layer separately we now establish how these two layers are connected. Since BrdU measurements reveal that our model of the basal layer encompassed the majority if not all of the possible cell divisions, further proliferation of basal progenitors seems unlikely. It also seems unlikely that basal cells are lost at a significant rate. We

therefore propose that basal progenitors differentiate to secretory cells at a rate $\beta = 0.095 \text{ d}^{-1}$. Indeed, by assuming this simplest possible link between the basal layer and the luminal layer were able to obtain an excellent fit to the full clonal data from the *Tg(KRT5-CreER); Rosa26R-fGFP* mice (Fig. 2e and S4).

Interestingly, close inspection of the fits for the secretory cells reveals that we slightly, but consistently, underestimate the frequency of clones with a single secretory cell. Note that the number of secretory cells in a given clone was defined as the number of cells that were not identified as basal or ciliated cells. We therefore speculate that the abundance of singlets of secretory cells can be explained by the the existence of another cell type in the lineage, which are likely neuroendocrine or brush cells. The existence of an additional (but infrequent) persistent cell type also explains why clonal loss is overall rare (cf. Fig. 1g), while our model predicts a relatively significant loss.

1.6 Deviation from homeostasis

Measuring the cellular composition in mice of different ages we find that the ratio of ciliated cells to secretory cells, c/s , changes from 0.83 in young mice to about 0.5 in mice 1 year year after injection. We now ask what effect such a deviation from homeostasis has on the values of the model parameters. In the framework of our model, such a deviation is caused by an imbalance in the rates of symmetric divisions, leading to an exponential departure from the homeostatic cellular composition. Since this departure is exponential, we expect that its influence on the model parameters is only logarithmic, i.e. very weak. Importantly, according to our analysis, the value of c/s only affects the dynamics in the luminal layer. We therefore performed the maximum like-

likelihood estimation for the maximum and the minimum value of c/s . For $c/s = 0.83$ we obtained $\sigma_1 = 0.036 \pm (0.011, 0.008)$, while for $c/s = 0.5$ we found $\sigma_1 = 0.040 \pm (0.013, 0.008)$. Both results are in good agreement with the maximum likelihood value of $0.037 \pm (0.0105, 0.008)$ for $c/s = 0.75$, which we chose to fit our model. Hence, the change in cellular compositions does not significantly affect the estimation of parameters.

2 Consistency checks

While our model faithfully predicts the full clone size distributions and their cellular composition over time, it is important to check the validity of our results by independent experiments.

2.1 *Krt5-CreER; Rosa26R-Confetti* and *K8-CreER; Rosa26R-fGFP* experiments

To test whether our model could predict the outcome of independent lineage tracing experiments we made use of *Krt5-CreER; Rosa26R-Confetti* and *Krt8-CreER; Rosa26R-fGFP* experiments mice. For the *Krt5-CreER; Rosa26R-Confetti* line we found excellent agreement between the model predictions and the observed distributions of clone sizes (Fig. 3d). In the *Krt8-CreER; Rosa26R-fGFP* line both, SecCs and CCs were induced at a rather high frequency. As a result, labelled cells at 4 d post induction occurred not only in singlets, but also in doublets and triplets, even for non-dividing CCs. To take account for these mergers we drew the initial conditions of our simulations randomly from the clone size distribution given by the 4d post induction time point. We also observed a small degree of background labelling, particularly at late time points, which is supported by the fact that clone size distributions change little over time⁵. Therefore, not too long

after induction the clone size distribution should be dominated by clones from the initial labelling. Indeed, our model does a good job at predicting these early clones. By contrast, at late time points clones dominate that had been induced by background labelling at earlier time points. Assuming background induction at the same relative frequency between SecCs and CCs as initial labelling our model excellently predicts the long time clonal dynamics (Fig. 3f).

2.2 BrdU experiments

We can challenge these results using independent nucleotide incorporation assays. From our measured steady-state BrdU incorporation rates (Supplemental S1) we can infer the fraction of cells in S-phase. The measured fraction $f \approx 1.3\%$ relates to the cell cycle time of dividing cells, T_c via the time cells spend in S-phase, which we take to be around 7 hours,

$$T_c = \frac{T_S}{2f} \approx 11 \text{ days} . \quad (28)$$

We divided by 2 as, according to the modelling results, only about 50% of the basal cells undergo divisions. In the stochastic model the average cell cycle time is $(2\alpha_1 + \alpha_2)^{-1} \approx 11 \text{ d}$, in excellent agreement with the BrdU measurements.

2.3 EdU experiments

Having confirmed the overall rate of cell divisions in the basal layer we now ask whether our model is also able to predict the distribution of different division modes. To identify the types of cell divisions made in the wildtype trachea at steady-state we combined EdU incorporation with wholemount immunostaining to visualize S-phase cells (2 hours post-EdU, Fig. 2h and i, left) and

their immediate progeny (24 hours post-EdU, Fig. 2i, right). Based on our model we can calculate the predicted ratios of labelled single cells 2h after EdU incorporation. To this end we note that the fraction of basal cells among all cells in S-phase is given by the fraction of dividing basal cells in the whole population divided by the fraction of dividing cells in the tissue. In the whole population the fraction of dividing basal cells is equal to fraction of basal cells, which are capable to divide, r_b , multiplied by the probability that a basal cells is in S-phase, $T_s \lambda_b$. Here, T_s is the time a cell spends in S-phase (typically 6-8h) and $\lambda_b = 2\alpha_1 + \alpha_2$ is the total division rate of basal cells. Since, according to the model, secretory cells are the only other dividing cell type, the total fraction of dividing cells in the tissue is $T_s(r_b \lambda_b + r_s \lambda_s)$, where r_b denotes the fraction of secretory cells and $\lambda_s = \sigma_1 + \sigma_2$ is their total division rate. Therefore, we predict that after 2h of all labelled singlets $r_b \lambda_b / (r_b \lambda_b + r_s \lambda_s)$ are basal cells, $r_s \lambda_s / (r_b \lambda_b + r_s \lambda_s)$ are secretory cells and none are ciliated cells. Indeed, we find an excellent agreement with the EdU measurements (Fig. 2i).

Since all average division times are on the order of weeks we expect that 24h post EdU the cells have undergone exactly one division. The fact that at this time point we observe a considerable fraction of secretory cell singlets strongly suggests that EdU is toxic to secretory cells. We therefore focus our analysis on the relative fractions of different types of doublets in the basal layer and the luminal layer. According to our model, dividing cells in the basal layer will always produce two daughter cells in the basal layer, but no combinations of basal cells and secretory cells. We find that this prediction is in excellent agreement with the relative fractions of these division modes obtained from the EdU experiments (Fig. 2i). We now turn to the luminal layer. Here, our model predicts that 93% of divisions of secretory cells generate two secretory daughter cells, 7% produce

a secretory cell and a ciliated cell and none leads to a pair of ciliated cells. Again, this is in almost perfect agreement with the ratios we obtained from the EdU experiments.

With these results we can now ask whether there is a possibility for CC production through BLPs. To answer this we note that in homeostasis the loss of ciliated cells must be balanced by production from different sources, $\gamma c = \sigma_2 s + \epsilon b$. Here, ϵ denotes the production rate of CCs by BLPs. We now consider the minimum percentage of asymmetric divisions of SCs consistent with the EdU experiment, which we define as the lower bound of the 95% confidence intervals (6.6%). With the overall rate of SC divisions from our model we find that the minimum rate of CC production by SCs allowed by the EdU experiments is $\sigma_{2,\min} \approx 0.0026$ 1/d. With this, the maximum rate of CC production by basal cells is

$$\epsilon_{\max} = d/b\gamma - c/b\sigma_{2,\min} = 4.2 \cdot 10^{-4} \text{ d}. \quad (29)$$

This means that by the EdU experiments and the overall rate of SC divisions from the model the rate of CC production by BLPs is constrained to less than once per 6 years, corresponding to less than 1% of the total CC production.

In summary, the results from the EdU experiments excellently agree with our model predictions. This strongly suggests that our modelling framework contains the most important modes of cell divisions. Together with the BrdU experiments, which fix the overall rate of cell division in the basal layer, these results reassure us that there are no relevant further division modes, such as divisions of basal progenitors, or divisions of basal stem cells which contribute to the luminal layer.

2.4 H2B-GFP experiments

Using *Krt8-H2B-GFB* marking BLPs we then asked whether there was evidence for proliferation of BLPs. To this end we counted the frequencies of GFP positive cells at several time points after induction. Our model predicts that K8 expressing basal cells do not divide and the rate of loss of labelled cells therefore depends linearly on their concentration, $p(t)$ and the rate of loss of BLPs from the basal layer, $\dot{p}(t) = -\beta p(t)$. Therefore, the percentage of labelled cells should decrease exponentially with a rate corresponding to the rate of loss of cells from the basal layer, β , viz. $p(t) \propto \exp(-\beta t)$. Indeed, fitting an exponential to the H2B-GFP time course gives an average lifetime of BLPs of 16 d, which is in good agreement with the predicted lifetime of 11 d. We therefore believe that BLPs do not divide significantly in homeostasis.

References

1. Klein, A. M. & Simons, B. D. Universal patterns of stem cell fate in cycling adult tissues. *Development* **138**, 3103–11 (2011).
2. Klein, A., Doupé, D., Jones, P. & Simons, B. Kinetics of cell division in epidermal maintenance. *Phys. Rev. E* **76**, 021910 (2007).
3. Box, G. E. P. & Tiao, G. C. *Bayesian inference in statistical analysis*, vol. 40 (John Wiley & Sons, 2011).
4. Gillespie, D. T. Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.* **81**, 2340–2361 (1977).
5. Bailey, N. T. J. *The Elements of Stochastic Processes with Applications to the Natural Sciences* (John Wiley & Sons, 1990).