HHS Public Access

Author manuscript

Cell Stem Cell. Author manuscript; available in PMC 2018 December 07.

Published in final edited form as:

Cell Stem Cell. 2017 December 07; 21(6): 775–790.e9. doi:10.1016/j.stem.2017.10.014.

Injury activates transient olfactory stem cell states with diverse lineage capacities

Levi Gadye[1,*], **Diya Das**[2,3,*], **Michael A. Sanchez**[2,#,*], **Kelly N. Street**[4,9], **Ariane Baudhuin**[2], **Allon Wagner**[5,9], **Michael B. Cole**[6,9], **Yoon Gi Choi**[7], **Nir Yosef**[5,9], **Elizabeth Purdom**[8,9], **Sandrine Dudoit**[3,4,8,9], **Davide Risso**[4,10], **John Ngai**[1,2,7,11], and **Russell B. Fletcher**[2]

[1]Helen Wills Neuroscience Institute, University of California, Berkeley, California 94720

[2]Department of Molecular and Cell Biology, University of California, Berkeley, California 94720

[3]Berkeley Institute for Data Science, University of California, Berkeley, California 94720

[4]Division of Biostatistics, University of California, Berkeley, California 94720

[5]Department of Electrical Engineering and Computer Science, University of California, Berkeley, California 94720

[6]Department of Physics, University of California, Berkeley, California 94720

[7]QB3 Functional Genomics Laboratory, University of California, Berkeley, California 94720

[8]Department of Statistics, University of California, Berkeley, California 94720

[9]Center for Computational Biology, University of California, Berkeley, California 94720

[10]Department of Healthcare Policy and Research, Weill Cornell Medical College, New York, NY 10065

Summary

Tissue homeostasis and regeneration are mediated by programs of adult stem cell renewal and differentiation. However, the mechanisms that regulate stem cell fates under such widely varying conditions are not fully understood. Using single cell techniques, we assessed the transcriptional changes associated with stem cell self-renewal and differentiation and followed the maturation of stem cell-derived clones using sparse lineage tracing in the regenerating mouse olfactory epithelium. Following injury, quiescent olfactory stem cells rapidly shift to activated, transient states unique to regeneration and tailored to meet the demands of injury-induced repair, including

[11]Lead contact and corresponding author: jngai@berkeley.edu.

[*]These authors contributed equally to this work

[#]Present address: Office of the Associate Director, Veterans Affairs Northern California Health Care System, Martinez, CA 94553

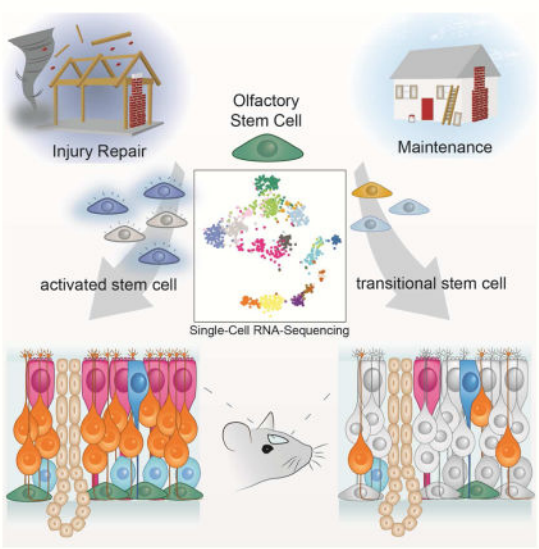
Author Contributions

R.B.F., L.G., M.S., Y.G.C., D.R., E.P., S.D. and J.N. designed the single-cell RNA-seq experiments; D.D., R.B.F., K.N.S., A.W. and M.B.C. analyzed the single-cell RNA-seq data; L.G., R.B.F., M.S., and A.B. designed and performed all other experiments; N.Y., E.P., S.D. and D.R. supervised statistical and computational analyses; L.G., D.D., R.B.F., and J.N. wrote the manuscript with input from the co-authors; J.N. and R.B.F. oversaw all aspects of the study.

barrier formation and proliferation. Multiple cell fates – including renewed stem cells and committed differentiating progenitors – are specified during this early window of activation. We further show that *Sox2* is essential for cells to transition from the activated to neuronal progenitor states. Our study highlights strategies for stem cell-mediated regeneration that may be conserved in other adult stem cell niches.

eTOC Blurb

Gadye et al. use multiple single cell techniques to identify the cell state transitions underlying the stem cell self-renewal and differentiation during injury-induced regeneration of the olfactory epithelium. Olfactory stem cells shift en masse to a transient cell state unique to regeneration in which diverse fates are specified.



Introduction

Tissues that undergo cellular turnover are often capable of robust regeneration, requiring adult stem cell populations to modulate self-renewal and differentiation after the loss of mature cell types both under homeostatic conditions and following injury. A division of labor exists in many tissues, in which actively dividing stem cells support tissue maintenance under normal conditions of tissue homeostasis, while normally quiescent stem cells are recruited to regenerate the tissue following injury (Ito et al., 2005; Wilson et al., 2008; Yan et al., 2012). The mechanisms underlying such dynamic regulation of stem cell proliferation and differentiation remain poorly understood, however.

The mouse olfactory epithelium provides a tractable model system for illuminating the different strategies underlying stem cell-mediated injury-induced repair and homeostatic tissue maintenance. Olfactory neurogenesis is normally sustained over the lifespan of the animal through the differentiation of globose basal cells (GBCs), which are the actively proliferating neurogenic progenitor cells in the niche (Caggiano et al., 1994; Schwob et al.,

1994). Unlike the rest of the nervous system, upon targeted destruction of the sensory neurons or more severe tissue injury, the olfactory epithelium regenerates (Schwob et al., 1995) due mainly to the self-renewal and differentiation of a normally quiescent stem cell, the horizontal basal cell (HBC) (Iwai et al., 2008; Leung et al., 2007).

Recent studies using single cell RNA-sequencing (single-cell RNA-seq) and in vivo lineage-tracing identified early transition states during which cell fates are specified (Fletcher et al., 2017). While these studies revealed the paths that HBCs take when differentiating into olfactory neurons and sustentacular (support) cells under conditions of tissue homeostasis, the cellular and transcriptional mechanisms underlying stem cell fate choice and expansion during regeneration – a coordinated process requiring the rapid production of multiple cell types to reconstitute the epithelium following injury – have yet to be characterized. Using complementary single cell approaches, we trace individual HBC stem cells and their derivatives during injury-induced regeneration and find differences in the mechanisms underlying their activation and specification for tissue repair as compared to homeostatic maintenance.

Results

Cell Fate Determination during Injury-Induced Regeneration in the Olfactory Epithelium

HBCs are usually quiescent under resting conditions but are activated by injury to differentiate and repopulate the epithelium (Figure 1A). We employed clonal lineage tracing of HBCs to determine when different cell fates are acquired during regeneration. After activation of Cre recombinase, severe injury to the olfactory epithelium was induced by administering methimazole (Leung et al., 2007), and animals were sacrificed at 7 and 14 days post-injury (DPI). HBC-derived clones were discriminated by P63 and SOX2 expression coupled with cellular morphology revealed by either the membrane CFP or cytosolic YFP lineage tracer (Figure 1A,B; STAR Methods).

At 7 DPI, sustentacular cells were the predominant cell type formed (46%), followed by renewed HBCs (23%) and neurons (17%) (Figure 1C). At 14 DPI relative to 7 DPI, neurons comprised a significantly higher fraction of labeled cells (51% vs. 17%; $p = 0.03$), whereas the percentage comprising sustentacular cells was significantly reduced (29% vs. 46%; $p = 0.05$) (Figure 1C). Accordingly, in clones containing neurons, the mean number of neurons per clone increased from 5.4 +/− 1.2 at 7 DPI to 9.4 +/− 0.83 at 14 DPI (mean +/− SEM) (Figures 1D and S1), consistent with ongoing neuronal progenitor proliferation and maturation in the second week of regeneration. In clones containing sustentacular cells, the mean number of sustentacular cells per clone decreased slightly from 5.51 +/− 0.5 at 7 DPI to 4.4 +/− 0.4 at 14 DPI (Figure 1D and S1), perhaps due to the ongoing generation of small numbers of sustentacular cell clones later in regeneration.

In the absence of injury, HBCs are mostly quiescent and differentiate only infrequently (Iwai et al., 2008; Leung et al., 2007), a state maintained by the transcription factor p63 (encoded by *Trp63*) (Fletcher et al., 2011). In previous studies we used conditional knockout of *Trp63* in HBCs to stimulate their differentiation under homeostatic, non-injury conditions (Fletcher et al., 2017). Under these conditions, individual HBCs can transdifferentiate into

sustentacular cells without cell division, often giving rise to just a single sustentacular cell (Fletcher et al., 2017). By contrast, multiple sustentacular cells are found in clones from injury-induced regeneration (58 of 69 sustentacular cell-containing clones at 14 DPI), while only a small number of these clones (5 of 69) are composed of a single sustentacular cell and no other cell type (Table S1). Thus, in contrast to the transdifferentiation of HBCs into sustentacular cells observed in the *Trp63* knockout under homeostatic conditions (Fletcher et al., 2017), the generation of sustentacular cells during injury-induced regeneration involves cell proliferation.

Upon conditional knockout of *Trp63* under homeostatic conditions, individual HBCs are typically unipotent, giving rise to a single cell type (Fletcher et al., 2017). To assess the potency of individual HBCs during regeneration, the compositions of HBC-derived clones arising after injury were scored for the presence of at least one cell from each of the three primary olfactory lineages: HBCs (self-renewal); GBCs, immediate neuronal precursors, microvillous cells, and neurons comprising the neuronal lineage; and sustentacular cells (Figure 1E). At both 7 DPI and 14 DPI, clones reflect both unipotent and multipotent stem cell fate choices. The percentage of clones possessing at least one neuronal lineage-fated cell or one sustentacular cell varied little between 7 DPI and 14 DPI (7 DPI: 56% neuronal, 65% sustentacular; 14 DPI: 57% neuronal, 59% sustentacular) (Table S1). However, the composition of clones possessing neuronal lineage-fated cells shifted away from renewing, multipotent clones at 7 DPI toward clones lacking an HBC at 14 DPI (Figure 1E), suggesting that renewed HBCs present in clones at 7 DPI may later differentiate into neurons or sustentacular cells. Consistent with this notion, we observed that the percentage of clones containing renewed HBCs dropped from 61% at 7 DPI to 41% at 14 DPI (Figure 1E). Over 30% of clones contain cells representing both neuronal and sustentacular lineages during regeneration (Figure 1E), compared to ~10% of HBC-derived clones in the absence of injury (Fletcher et al., 2017). Thus, at face value it appears that individual HBCs exhibit multipotency in response to injury, unlike their behavior under homeostatic conditions in the *Trp63* knockout.

Injury Stimulates Symmetric and Asymmetric HBC Cell Divisions Early in Regeneration

To what extent do symmetric and asymmetric cell divisions underlie replenishment of the stem cell pool and production of differentiated progenitors? To address this question, clonal lineage-tracing of HBCs was performed and clones were analyzed at 48 hours post-injury (HPI) (STAR Methods). Immunohistochemistry for P63, which marks HBCs, and SOX2, which marks almost all cells in the olfactory epithelium at 48 HPI, was used to assign either self-renewal or differentiation cell fates of lineage-traced cells (Figure 1F). Of 169 clones analyzed across five animals, 129 cell doublets (corresponding to cells that underwent a single cell division) were identified (Figure 1G,H). Pairs of cells arising from putative symmetric and asymmetric divisions of all types were observed; based on expression of P63 and SOX2, symmetric divisions accounted for 83% of all doublets (51% self-renewing, 25% differentiating, 7% unknown), while asymmetric divisions accounted for just 16% of all doublets (Figure 1H). Moreover, among clones of all sizes, renewed HBCs accounted for 56% of all scored cells at 48 HPI (Figure 1G). These results suggest that HBCs utilize a balance of primarily, but not exclusively, symmetric renewing and differentiating divisions

during regeneration. Moreover, the observation that most doublets appear to be the result of either self-renewing or differentiating symmetric cell divisions led us to hypothesize that cell fates are chosen by 48 HPI and manifest at or during the time of the first HBC cell division.

Single-Cell RNA-Sequencing Identifies Cell States and Lineage Trajectories Specific to Injury-Induced Regeneration

We applied single-cell RNA-seq to HBCs and their descendants to complement the clonal lineage studies described in Figure 1. We wished to test the hypothesis that cell fates are chosen within the first 48 hours following injury by identifying the cell states arising from the activation of HBC stem cells during injury-induced regeneration. HBCs were lineage-traced, olfactory epithelium was harvested, and YFP(+) cells were isolated by fluorescence-activated cell sorting (FACS) in the uninjured state and at time points between 24 HPI and 14 DPI to capture cells in multiple phases of regeneration (Figure 2A,B). We captured and sequenced a total of 831 YFP lineage-traced cells across all conditions; following sequencing, 672 cells subsequently passed quality filters and were included in downstream analyses (STAR Methods, Table S2). The data were then clustered to identify the distinct cell types that arise during injury-induced regeneration (STAR Methods). Clusters were assigned preliminary identities based on known marker gene expression (Figure 2C). To visualize heterogeneity in gene expression, we produced a two-dimensional representation using t-distributed Stochastic Neighbor Embedding (t-SNE; van der Maaten and Hinton, 2008), labeled according to RSEC cluster assignment (Figure 2D,E; STAR Methods). All clusters segregate in the t-SNE plot, confirming the cluster assignments made independently by RSEC, which identified many cell types also found in our previous analysis of differentiating HBCs (Fletcher et al., 2017). Regeneration is unique, however, in the appearance of two novel classes of HBCs that were tentatively assigned as activated HBCs (HBC*1, HBC*2) based on their variable expression of the HBC markers *Trp63*, *Krt5*, and *Krt14*, and high-level expression of genes associated with epidermal differentiation, epithelial stress and wound response, and the cell cycle (see below and Figure 3).

To gain additional insight into the relative contributions of each experimental time point to each cluster, we labeled the t-SNE plot to highlight the time at which cells were collected for analysis (Figure 2F). By 24 HPI, all cells have shifted to an activated HBC state, with the majority in HBC*1 and a few having progressed to HBC*2. Quite strikingly, few if any cells past 48 HPI contribute to HBC*1 and HBC*2, indicating that these two clusters comprise a transient state. We observe two clusters of cells (HBC1, HBC2) that express the HBC markers *Krt5* and *Krt14* but have slightly lower and more variable expression of the HBC gene *Trp63*. These two clusters are almost exclusively populated by regenerating cells whereas the resting HBC cluster is populated by both uninjured and regenerating cells. Although resting HBCs reappear by 96 HPI, the preceding shift of all HBCs to the activated state indicates that this is a requisite intermediate through which all cells pass prior to self-renewing or differentiating. We also note that most sustentacular cells have formed by 96 HPI, and there are more cells at more differentiated neuronal stages at later time-points post injury, consistent with observations from the clonal lineage tracing (Figure 1).

To map the routes by which activated HBCs contribute to regeneration of the olfactory epithelium following injury, we employed Slingshot, a lineage analysis algorithm that can predict branching lineage trajectories from single-cell RNA-seq data (Fletcher et al., 2017; Street et al., 2017). Because almost all HBCs have transitioned to the HBC*1 stage and none remain in the resting HBC cluster at 24 HPI (Figure 2F), we removed the uninjured HBCs contributing to the resting HBC cluster from our input to Slingshot and selected the activated HBC cluster (HBC*1) as the root of the trajectories. Starting from HBC*1, Slingshot predicts four distinct lineage trajectories (Figure 2G,H), all of which pass through HBC*2 (largely cells from the 48 HPI time point; Figure 2E–H). Importantly, this prediction implies that all lineages – including the sustentacular cell lineage – transit through an actively proliferating phase, consistent with the results of clonal lineage tracing (Figure 1). The lineages branch following this proliferative phase. The neuronal lineage (colored orange, Figure 2G) progresses through GBCs and microvillous cells (GBC/MV), the immediate neuronal precursors (INP1/2,3), and then immature olfactory sensory neurons (iOSN), before terminating in the mature neurons (mOSN). The sustentacular cells (SUS, magenta lineage) mature directly from the HBC*2 cluster, consistent with our previous observation of the close relationship in gene expression between HBCs and sustentacular cells (Fletcher et al., 2017). Two apparently self-renewing lineages extend through HBC1 and terminate either in the resting HBC cluster (HBC) or in HBC2 (Figure 2G,H). Curiously, the HBC2 cluster is heavily populated by cells from the 7 and 14 DPI time points, while resting HBCs have already been regenerated by 96 HPI. It is therefore difficult to distinguish from in silico lineage reconstruction alone whether HBC2 represents an actual endpoint or rather regenerated HBCs that have begun another cycle of differentiation. This issue is addressed through additional clonal lineage tracing in vivo, as described below and in Figure 5.

Activated HBCs Are Unique to Injury-Induced Regeneration

Having identified an activated, transient stem cell state following injury, we next sought to establish more definitively that this state is specific to injury-induced regeneration and not present during differentiation of HBCs in the absence of injury. To compare the HBC states found in these two biological contexts, we used a normalized dataset comprising cells from the present study and cells from a prior analysis of HBC differentiation during homeostatic tissue maintenance (Fletcher et al., 2017). This latter dataset includes cells lineage-traced following conditional knockout of *Trp63*, plus olfactory progenitors and sustentacular cells expressing a $Sox2^{eGFP}$ reporter gene in a *Trp63* wild-type background. We plotted these 1232 cells in two dimensions using t-SNE and colored the HBCs by cluster identity and by experiment (Figures 3A,B). Interestingly, the HBCs segregate by their original cluster assignments, confirming that the transitional HBCs from homeostatic differentiation and the activated HBCs induced by injury represent distinct cell states. Unlike either of the transitional HBC clusters, one cluster of activated HBCs (HBC*2) expresses high levels of cell cycle-associated genes, comparable to the levels observed in GBCs and immediate neuronal precursor cells (Figure 3C). Clusters representing more mature cells in the neuronal and sustentacular cell lineages are populated by cells from both conditions (Figure 3B).

A closer examination of differential gene expression across HBC cell types reveals some notable differences between resting HBCs, activated HBCs in regeneration, and transitional

HBCs from differentiation under homeostatic conditions (Table S3, Figure S3. Resting HBCs express high levels of *Krt5* and *Krt14*, which are also maintained in activated HBCs but largely downregulated in transitional HBCs. Transitional HBCs specifically express high levels of *Sec14l3*, also a marker of sustentacular cells. RNA in situ hybridizations of representative genes enriched in the HBC*1 and HBC*2 clusters validate their transient, regeneration-specific expression as these genes are down-regulated by 96 HPI and not detectable in uninjured tissue (Figure S3). These data provide additional evidence that the activated HBC state is specific to injury-induced regeneration and is distinct from the transitional HBCs found during differentiation in the absence of injury.

Several genes associated with wound response in other epithelial tissues are expressed in the activated HBCs, but not in the early transitional HBCs observed in the *Trp63* knockout under conditions of homeostatic tissue maintenance. To assess the degree of their enrichment in the activated HBCs, we constructed a wound response gene set (*Krt6a*, *Krt16*, *Sprr1a*, *Sprr2a3*, *Krtdap*, *Dmkn*, *Sbsn*, *Hbegf*).

Collectively, these genes are important for cell shape changes and migration, barrier formation, buffering environmental stressors, and inducing proliferation (Bazzi et al., 2007; Henry et al., 2012; Shirakata et al., 2005; Wojcik et al., 2000). This gene set is highly enriched in the HBC*1 and HBC*2 cells (Figures 3D and S4; see also Table S3 and Figure S3), suggesting that the activated HBCs up-regulate these genes in response to injury to mitigate direct exposure to the environment, migrate, elongate, and proliferate to initiate the process of regeneration. One might expect that the cells would also express genes associated with more general cellular stress. Indeed, a set of genes related to the unfolded protein response shows enrichment in the activated HBCs by gene set enrichment analysis (GSEA) (Table S4). However, these genes show only a moderate enrichment in the activated cells (Figure S3), suggesting that transcriptional changes in the activated HBCs represent more than just a response to cellular stress. Rather, activated HBCs engage a specific genetic program associated with the wound response that is both transient and specific to injury.

We employed GSEA to define the functional differences between the different HBC states during regeneration (Table S4). The first activated HBC cluster HBC*1 expresses high levels of genes associated with mTOR signaling and protein translation (Figure S3). We validated this observation using an antibody to the phosphorylated form of S6 (phospho-S6), an indirect downstream target of mTOR signaling that reflects translational activity (Ma and Blenis, 2009). At 24 HPI, phospho-S6 and P63 co-localize, but by 48 HPI, phospho-S6 begins to show enrichment in the apical layers (Figure 3E). GSEA also predicts that the activated HBCs are proliferative, as several of the most enriched gene sets in HBC*2 correspond to active cell cycling (Figure S3). In accord with this prediction, the cell cycle marker Ki67 is detectable at 24 HPI in the regenerating olfactory epithelium, and almost all cells are Ki67(+) at 48 HPI (Figure 3F).

Variability in Gene Expression Reveals Heterogeneity of Activated Stem Cell States Poised for Differentiation

We next sought to determine the extent to which the activated HBCs are heterogeneous and therefore might represent a stage during which multiple cell fates are initially manifested, as

initially suggested by the clonal lineage tracing experiments shown in Figure 1. Although the activated HBCs (HBC*1, HBC*2) are distinguished by several genes at the population level, the expression of these genes is highly variable, indicating heterogeneity at the single cell level. For example, all HBCs express *Krt5* at relatively high levels, but expression of *Trp63* is less consistent. Similarly, genes enriched in HBC*1 cells, such as *Krt6a*, *Sprr1a* (wound response genes), *and Il33* (a gene expressed in sustentacular cells and also upregulated following epidermal injury (Oshio et al., 2017), are expressed in many, but not all activated HBCs (Figure 4B).

This heterogeneity was corroborated using double RNA in situ hybridizations for *Krt5* and *Krt6a, Trp63* and *Sprr1a, and Lgals1* and *Sprr1a* at 24 and 48 HPI (Figures 4C and S4). Furthermore, IL33 protein – a marker of the sustentacular cell lineage – is detectable as early as 24 HPI but becomes restricted to apical P63(−) cells at 48 HPI (Figure 4D), suggesting an early bifurcation between HBC self-renewal and differentiation. Taken together, the heterogeneity in activated HBCs that manifests in more spatially restricted expression domains at 48 HPI suggests that lineage commitment occurs as HBCs transit through the activated state during regeneration.

Approximately two thousand genes are either up- or down-regulated as cells transition from resting HBCs to the HBC*1 stage (Figure S5). Surprisingly, an analysis of the transcription factor networks between cell clusters in the neuronal and sustentacular lineages reveals that activated HBCs express many of the same transcription factors as the resting HBCs (Figures 4E,F and S5), which form a common node of the network. Resting HBCs therefore appear poised to transition to the activated state without a major change in their expression of transcription factors. As in differentiation in the absence of injury (Fletcher et al., 2017), the sustentacular cells also share many of the same transcription factors with the HBCs (Figure 4F). Conversely, during regeneration, the neuronal precursors and neurons occupy a distinct node or subnetwork from the HBCs in a manner similar to that observed during differentiation under homeostatic conditions (Fletcher et al., 2017).

To investigate potential links between the signaling pathways and transcription factors present in HBCs and the robust transcription of genes in the HBC*1 state, we analyzed the promoters of the 1000 genes that showed the highest enrichment in the HBC*1 state relative to the resting HBCs (Table S5). Ets-domain containing transcription factor binding sites (from Ets, Elk and Elf transcription factors) were highly represented (9 of the top 30 motifs). Ets-domain containing transcription factors are established mediators of growth factor signaling, including Egfr signaling (Shwartz et al., 2013), dovetailing with our finding that the ligand *Hbegf* is highly enriched and Egfr signaling activated in the HBC activated states (HBC*1, HBC*2). Furthermore, AP-1 binding sites are also among the sites most enriched in the promoters of genes that are expressed highly in the HBC*1 cells (Table S5). Accordingly, some of the most enriched transcription factors in both resting (HBC) and activated (HBC*1, HBC*2) clusters are the AP-1 transcription factors, which are associated with epidermal wound response and regeneration (Yates and Rayner, 2002; Figure S5).

Olfactory Cell Fates Are Chosen in the Early Activated HBC State

We analyzed the gene expression profiles of the activated HBC clusters to identify putative markers of subsets of activated HBCs. Through this analysis, we found *Hopx*, a transcriptional regulator implicated in intestinal and hair follicle stem cell identity (Takeda et al., 2011), to be expressed in a subset of activated HBCs (Figure S5). We confirmed by RNA in situ hybridization that *Hopx* is expressed in a subset of activated HBCs at 24 HPI but not in resting HBCs in uninjured olfactory epithelium (Figure S5), making it a good candidate for lineage-tracing a subpopulation of activated HBCs.

To test the prediction that cell fate commitment occurs early during the activated HBC state, we carried out a series of lineage-tracing experiments in $Hopx^{CreER}$; $Rosa26^{eYFP}$ animals, in which Cre recombinase activity was induced with tamoxifen either prior to injury or at 24 HPI. Activation of Cre recombinase driven from the *Hopx* locus prior to injury did not result in the labeling of any cells by 14 DPI (data not shown), reflecting the narrow temporal window in which HBCs express *Hopx* and consistent with the absence of *Hopx* expression in resting HBCs (Figure S5). However, induction of $Hopx^{CreER}$ at 24 HPI led to the labeling of differentiated clones at 14 DPI (Figure 5A). Clones traced with $Hopx^{CreER}$ at 24 HPI consisted solely of differentiated cells; there were no renewed HBCs (Figure 5B,C). While all clones consisted of cells from either the neuronal or sustentacular lineages, 26% of clones contained cells from both lineages, in accordance with our analysis of clonal cell doublets indicating that a minority of clones are derived from some type of asymmetric cell fate choice (Figure 1F–H). Importantly, the lack of renewed HBCs in $Hopx^{CreER}$-derived clones during regeneration demonstrates that the cell fate potential of a subset of activated HBCs is restricted to the neuronal and sustentacular cell lineages as early as 24 HPI. These observations dovetail with analysis of clonal cell doublets (Figure 1F–H) and single-cell RNA-seq data (Figure 2), both of which indicate that the decision to self-renew or differentiate manifests during the activated HBC state.

Renewed HBCs Are Competent for Self-Renewal and Differentiation

Our results from both clonal lineage tracing and single-cell RNA-seq confirm that HBCs are renewed during injury-induced regeneration, but raise the question of whether these renewed HBCs are themselves capable of additional rounds of self-renewal and differentiation. HBCs were “late-traced” in *Krt5-CreER;* $Rosa26^{Confetti}$ mice 96 hours following injury, a time by which renewed HBCs appear in the regenerating tissue (Figure 2F). Clones derived from late-traced (renewed) HBCs exhibited a stronger tendency to produce renewed HBCs than clones derived from HBCs traced prior to injury (27% vs. 12% of all cells, and 54% vs. 41% of clones, for clones traced at 96 HPI vs prior to injury) (Figures 5C and S5). Renewed HBCs traced at 96 HPI produced significantly fewer neurons than resting HBCs traced prior to injury (Figure 5D), a difference that could be due in part to the shorter time period of differentiation allowed in these late-tracing experiments compared to our conventional labeling and injury scheme (10 vs. 14 days). Clones traced from 96 HPI were also less likely to contain either neuronal or sustentacular cells compared to clones traced prior to injury (35% vs. 57% neuron-containing clones; 43% vs. 59% sustentacular cell-containing clones; Figure 5C). Thus, renewed HBCs continued to differentiate when traced at 96 HPI, although

their relative rates of fate commitment differed based on the time at which they were lineage-traced.

Our inference of lineage trajectories from single-cell RNA-seq data (Figure 2) revealed some ambiguities in the renewing HBC lineage, which we attribute in part to the multiple rounds of HBC self-renewal and differentiation demonstrated by clonal lineage tracing (Figure 5C,D). Further, the observation that activated HBCs (HBC*1, HBC*2) derived exclusively from cells collected by 48 HPI (Figure 2) implies that renewed HBCs do not transit through the injury-specific activated state when they differentiate later during regeneration (i.e. beginning at 96 HPI). It remains to be determined whether the HBC2 cluster, comprising cells collected later at 7 DPI and 14 DPI (Figure 2), represents renewed HBCs, intermediate stages in a later wave of differentiation, or both. Whatever the case, there appear to be two successive phases of regeneration during the first two weeks following injury: an initial phase that involves the activated state in which the sustentacular cell fate is prioritized and a second wave that produces relatively fewer differentiated cells and bypasses the activated state specific to the initial response to injury.

Sox2 Is Required for Olfactory Neurogenesis Following Injury

Having demonstrated that HBCs quickly transition to an activated state upon injury and then branch to self-renew or differentiate, we wished to gain insight into the mechanisms driving differentiation in the neuronal lineage. *Sox2* is a well-established stem cell transcription factor associated most commonly with proliferative cells (Arnold et al., 2011; Suh et al., 2007). *Sox2* is also expressed in HBCs and GBCs in the uninjured olfactory epithelium and in the activated HBCs (HBC*1, HBC*2) during regeneration (Figure 2A), suggesting that it would be a good candidate for regulating olfactory neurogenesis. Lineage tracing of *Sox2*(+) cells using a $Sox2^{CreER}$ driver demonstrates the presence of a long-term self-renewing *Sox2*(+) GBC in the uninjured olfactory epithelium (Figure S6). Under homeostatic conditions, conditional knockout of *Sox2* in these cells resulted in a dramatic reduction in the formation of neurons and the number of detectable GBCs (Figure S6).

We next examined *Sox2*'s role in regeneration. In the absence of *Sox2*, HBCs are still able to self-renew during regeneration (Figure 6B,D). Quite strikingly, however, the mutant HBCs produced fewer Ki67(+) progenitor cells ($p = 0.02$) and neurons; most lineage-traced differentiated cells were sustentacular cells (Figure 6C,D). There was an increase in the number of EdU label-retaining cells (Figure 6C,D), suggesting a halt of proliferation of the basal progenitors, which include both GBCs and immediate neuronal precursor cells (Figure 3B). A recent report by Packard et al. (2016) also found defects in olfactory neurogenesis in a similar *Sox2* conditional knockout, although the number of GBCs appeared to be unaltered in that study.

Clonal lineage-tracing experiments confirmed that conditional ablation of *Sox2* in the HBCs strongly impedes regenerative neurogenesis. As expected, in the *Sox2* conditional knockout, the number of neurons was dramatically reduced in HBC-derived clones, accounting for just 7% of all counted cells, compared to 51% in wild-type HBC-derived clones (Figure 6E–G). Moreover, 46% of all wild-type clones contained at least one neuron, compared to just 10% of clones in the *Sox2* mutant ($p < 0.001$), confirming the deficiency of neurogenesis in the

absence of *Sox2*. The numbers of renewed HBCs and sustentacular cells were both increased in *Sox2* conditional knockout clones, and 91% of mutant clones contained either exclusively HBCs, sustentacular cells, or HBCs and sustentacular cells (Figure 6G). Thus, in the absence of *Sox2*, HBCs are defective for olfactory neurogenesis but retain their capacity to self-renew and produce sustentacular cells. The persistence of HBC renewal in the absence of *Sox2* contrasts with *Sox2*'s demonstrated role in stem cell maintenance and multipotency in other niches (Pevny and Nicolis, 2010).

Single-Cell RNA-Sequencing Pinpoints the Requirement of *Sox2* to the Formation of GBC Progenitors

The lineage tracing studies described in Figure 6 are insufficient to define precisely where in the HBC-derived lineage Sox2 is acting, since the identities of cells that either accumulate or are depleted in the mutant cannot be unequivocally ascertained using a few immunohistochemical markers. We therefore turned to single-cell RNA-seq to more precisely localize the stage at which the *Sox2* mutant defect manifests. *Sox2* was conditionally knocked out in HBCs, and HBCs and their descendants were isolated following injury and sequenced (Figure 7A). The normalized data from regeneration in the wild type (Figure 2) and knockout backgrounds were together subjected to cluster analysis. All three major lineages – renewing HBCs, sustentacular cells, and neurons – were predicted from this analysis, which also identified three small clusters of activated HBCs (collectively labeled as HBC*) that were placed after HBC*2 and prior to the divergence of the three lineages (Figure 7B–E). The de novo appearance of the HBC* clusters in the combined wild type + *Sox2* mutant analysis may be due to the greater statistical power inherent in analyzing data from a larger number of cells than the wild type time course alone.

Cells derived from *Sox2* conditional knockout HBCs contributed to the sustentacular cell and HBC lineages at levels comparable to wild type (Figure 7C–D). In striking contrast, few mutant cells were found in the neuronal lineage past the HBC* stage, indicating an arrest prior to the formation of GBCs (Figure 7E). The sustentacular cells formed relatively early, with the majority arising by the 96 HPI time-point, as did the renewing HBCs, which develop from the activated state. Thus, *Sox2* is not necessary for the earlier cell fate choice between self-renewal and differentiation. Differential gene expression analysis comparing wild type and *Sox2* knockout cells in each early HBC cluster showed that *Sox2* knockout HBCs are enriched for markers associated with sustentacular cells, including *Il33* and *Pon1* (Figure S7, Table S6). Although it is possible that GBCs form but subsequently fail to proliferate in the *Sox2* mutant, the apparent bias of HBCs toward the sustentacular cell fate suggests that the mutant neurogenic phenotype is due to a failure of activated HBCs to transition into GBCs.

Discussion

Regeneration-Specific Activated Stem Cell States Mediate Tissue Repair Following Injury

Severe injury places demands on a tissue that are not observed under normal conditions of tissue homeostasis. Stem cells not only must respond to the environmental stressors specific to the injury itself, but must also rapidly rebuild the entire structure – a process requiring

robust proliferation and differentiation. Upon injury to the olfactory epithelium, HBCs shift en masse to an activated, transient state characterized by the rapid initiation of the wound response transcriptional program, which includes upregulation of differentiation-associated intermediate filament keratins (*Krt6a*, *Krt16*), members of the *Sprr* gene family (*Sprr1a*, *Sprr2a3*) and the stratified epithelium secreted peptides complex (*Krtdap*, *Dmkn*, *Sbsn*) in the activated HBCs. These genes are associated with epidermal differentiation and are involved in mitigating environmental stressors, promoting cell migration and shape changes, and establishing the barrier function of the epidermis (Bazzi et al., 2007; Henry et al., 2012; Wojcik et al., 2000). Expression of these genes may be important for establishing an initial barrier following destruction of the overlying tissue, given that the olfactory epithelium is directly exposed to the environment. Accordingly, the clonal expansion of HBCs following injury is strongly biased toward reconstituting the population of apical, sustentacular cells prior to establishing the neuronal lineage.

Self-Renewal and Other Cell Fates Are Determined in the Activated HBC State

Although two classes of activated HBCs can be clearly distinguished from resting HBCs by their transcriptomes, cell-to-cell heterogeneity in gene expression exists within each activated HBC class. For example, we found that the transcription factor *Hopx* is expressed in a small subset of activated HBCs by 24 hours following injury. We further demonstrated that activated HBCs lineage-traced with a $Hopx^{CreER}$ driver 24 hours following injury give rise to differentiated cells but not self-renewed HBCs. Since HBC self-renewal occurs within this time frame, these observations indicate that the activated HBC state represents a window during which specific and restricted cell fates are determined. Significantly, the shift of all HBCs by 24 hours following injury to an activated state prior to the later appearance of renewed HBCs at 96 hours (Figure 2) indicates that self-renewal of HBCs during regeneration requires passage through this transient intermediate state. Thus, the apparently symmetric self-renewing cell divisions undertaken by HBCs during the injury response appear to reflect a “round-trip” to and from the activated state, during which the decision is made to self-renew or differentiate. It will be interesting to determine whether self-renewing and differentiating divisions in other stem cell niches similarly proceed through an activated intermediate.

While *Hopx* is expressed in a subset of activated HBCs and is associated with differentiating cell fates, the transcription factor *Sox2* is expressed in both resting and activated HBCs but plays an essential role only in the genesis of the neuronal lineage. A previous report (Packard et al., 2016) as well as the present study demonstrate that *Sox2* is required for the formation of olfactory sensory neurons during injury-induced regeneration. We further show that in the absence of *Sox2*, activated HBCs fail to transition to neuronally-restricted GBCs and instead express genes specific to sustentacular cells, which develop in spite of *Sox2*’s absence.

Cellular Mechanisms of Multipotency in the Regenerating Olfactory Epithelium

HBCs are multipotent and give rise to multiple cell types of the olfactory epithelium under both homeostatic and regenerative conditions. In the absence of injury, multipotency of HBCs at the population level largely arises from unipotent fate choices made by individual

HBCs upon genetic ablation of *Trp63* (Fletcher et al., 2017). During injury-induced regeneration, however, nearly half of all clones derived from individual HBCs comprise cells from more than one lineage, compared to only 10% of HBC-derived clones exhibiting similar multipotency under homeostatic conditions (Fletcher et al., 2017). What, then, are the mechanisms underlying stem cell multipotency in the regenerating olfactory epithelium?

The presence of renewed HBCs alongside differentiated cells in regenerating clones raises the possibility that HBC multipotency arises from successive rounds of HBC self-renewal, such that self-renewed HBCs are competent to adopt different cell fates, including self-renewal. Indeed, late-tracing of renewed HBCs confirms the sustained multipotency of these cells. This situation is reminiscent of regenerating muscle, in which renewed satellite cells retain both their stemness and multipotency (Collins et al., 2005) and are also known to arise from a heterogeneous pool of activated stem cells (Kuang et al., 2007). Likewise, Lgr5(+) stem cells persist as multipotent stem cells following injury to the intestine (Tetteh et al., 2016). Lineage-tracing of *Hopx*-expressing cells further reveals that a subset of individual activated HBCs are multipotent, producing clones containing sibling sustentacular cells and neurons, but not HBCs. These observations indicate that at least some activated HBCs are capable of giving rise to multiple differentiated cell types. The present study suggests that multipotency of HBCs may arise through multiple and conserved cellular mechanisms that together ensure the repair and complete reconstitution of the olfactory epithelium following injury.

STAR Methods

Contact for Reagent and Resource Sharing

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, John Ngai (jngai@berkeley.edu).

Experimental Model Details

Transgenic Mice—Mice containing the following transgenes or modified alleles were used in this study: *Krt5-CreER(T2)* driver (Indra et al., 1999), *Krt5-CrePR* driver (Zhou et al., 2002), $Trp63^{lox}$ conditional knockout allele (Mills et al., 2002), $Sox2^{lox}$ conditional knockout allele (Shaham et al., 2009), $Sox2^{CreER}$ driver (Arnold et al., 2011), $Hopx^{CreER}$ driver (Takeda et al., 2011), $Rosa26^{eYFP}$ reporter (Srinivas et al., 2001), and $Rosa26^{Confetti}$ reporter (Snippert et al., 2010). All experiments were begun on animals at 3–4 weeks of age (P21–P28). We used both male and female mice in our studies; all mice were on a mixed C57BL/6 and 129 background and were assumed to be of normal immune status. We did not employ blinding in our studies. Information about sex, age and genotype of animals used in RNA sequencing experiments is included as metadata in GEO record GSE99251. Animals were housed and maintained according to IACUC guidelines.

Method Details

Fluorescence-Activated Cell Sorting (FACS)—We labeled and isolated cells from the postnatal mouse olfactory epithelium. The *Krt5-CreER* and the $Rosa26^{eYFP}$ transgenes were combined with either the wild-type *Sox2* allele ($Sox2^{+/+}$) or the conditional knockout of

Sox2 (*Sox2*$^{lox/lox}$). This approach allowed us to label HBCs and lineage-trace their descendants. We collected HBCs that were wild type for *Sox2* and the HBC lineage following tamoxifen-induced conditional ablation of *Sox2* after 24 hours, 48 hours, 96 hours, 7 days, and 14 days.

In the HBC lineage tracing experiments, *Krt5-CreER; Rosa26*$^{eYFP/eYFP}$ and *Krt5-CreER; Sox2*$^{lox/lox}$*; Rosa26*$^{eYFP/eYFP}$ mice were injected once with tamoxifen (0.25 mg tamoxifen/g body weight) at P21–23 days of age and sacrificed at the specified times after injection (Figure 2). For each experimental time point, the olfactory epithelium was surgically removed, and the dorsal, sensory portion was dissected and dissociated. The dissociation protocol was identical to that used previously (Fletcher et al., 2017). Tissue from each animal was processed individually in approximately one mL volume. First, approximately 150 units of papain was dissolved in 5 mL of Neurobasal medium with 2.5 mM Cysteine and 2.5 mM ethylenediaminetetraacetic acid (EDTA) and incubated at 37° C for at least 30 minutes to activate the papain. Then, after dissection into Neurobasal medium, the tissue samples were added to an equal volume of the activated papain solution along with DNAase I (100 units/mL) and allowed to incubate with gentle nutation for 25 minutes at 37° C. Following papain digestion and dissociation, the tissue was washed multiple times in phosphate buffered saline containing 10% fetal bovine serum (PBS-FBS). Prior to loading on the cell sorter, cells were kept in PBS-FBS, and cells were sorted on a BD Influx sorter into PBS-FBS. A negative control (a littermate of the same genotype not injected with tamoxifen) was used to ensure proper gating. Propidium iodide was used to identify and select against dead or dying cells; only viable YFP(+) cells were collected.

Cell Capture and Single-Cell RNA Sequencing

Labeled cells from the olfactory epithelium were subjected to single-cell RNA-seq. Each FACS collection was considered a biological replicate, and at least two biological replicates were collected for each experimental condition. The Fluidigm C1 system was used to capture single cells, lyse them, and produce cDNA (Wu et al., 2014). For each replicate, a litter of animals were given the same treatment, and each transgenic animal was dissected and processed individually prior to loading on the cell sorter. When possible, all cells were obtained from one animal, but if one animal did not provide the minimum number of cells necessary to load the C1 chip (approximately 2000 cells), then cells from additional animals were sorted into the collection tube. Cells from no more than three animals were pooled in any given FACS purification. Each C1 run was considered a single batch. Upon loading of the C1 chip, each capture site was visually inspected at 100 × magnification for fluorescence, debris, and doublets or multiple cells. The following were excluded from future analysis: (1) any capture site scored as having two or more cells (doublets or multiplets), (2) any capture site that included additional debris, and (3) any capture site that did not have an apparently intact, fluorescent cell. The standard Fluidigm C1 Single-Cell auto prep system protocol for cell lysis, cDNA synthesis, and amplification was followed. This incorporates the Clontech SMART-Seq Ultra Low Input RNA reagents (Clontech SMARTer Kit designed for the C1 System) to produce and amplify cDNA. After processing on the C1 chip, 7 μL cell harvest buffer (Fluidigm) was added to each sample, resulting in an approximate final volume of 10 μL. For quality control purposes, any cell with less than 1.7

ng cDNA in 10 μL final volume (quantified using a Qubit fluorometer) was excluded. Illumina Nextera tagmentation-based sequencing library synthesis (Nextera XT DNA Sample Preparation Kit) was performed using Nextera v2 index oligos (Nextera XT DNA Sample Preparation Index Kit). Library size was selected using AMPure XP (Beckman Coulter) beads and confirmed using an Agilent Bioanalyzer. Indexed, single-cell libraries were sequenced on Illumina HiSeq 2500 sequencers to produce 50 nt single-end reads, except for two paired-end read runs (331 cells). Cells were sequenced in multiplex, with approximately 192 cells per lane for most runs. Prior to cell filtering, average sequencing depth was 1.318 million reads per cell (range 0.000622 to 4.657 million).

Clonal Lineage Tracing—*Krt5-CreER; Rosa26*^{Confetti}*, Krt5-CreER; Sox2*$^{lox/lox}$*; Rosa26*^{Confetti,} *and Hopx*^{CreER}; *Rosa26*^{eYFP} transgenic mice were used for clonal lineage tracing of the HBC lineage during regeneration. A dose-response analysis was performed to determine an optimal dose of tamoxifen to achieve sparse labeling with *Krt5-CreER*: 0.025 to 0.05 mg tamoxifen/g body weight administered by intraperitoneal injection, followed 3 days later by induction of injury to the olfactory epithelium via IP injection of methimazole (50 micrograms/g body weight). At this dose of tamoxifen, at 14 days post methimazole administration there were 40 occurrences among 151 (26.5%) clones that expressed different reporters (e.g. mCFP and cYFP) with an edge-to-edge distance of 50 microns or less. Because the reporters localize to different compartments, we could distinguish close or overlapping clones expressing different reporters; we excluded clones that possessed a neighboring clone less than 50 microns away that also expressed the same reporter. We expect that the same percentage of clones were within 50 microns of a clone expressing the same reporter; therefore, we estimate that no more than 27% of our scored clones are contaminated by cells derived from multiple independent labeling events. For all experiments, at least three animals were assayed per time point (Figures S1, S5, S6, and S7).

Tissue was fixed overnight (~16 hours), embedded in tissue freezing medium and frozen, and then sectioned on a cryostat at 40-micron thickness to allow sampling of the majority if not entirety of each clone. Prior to incubation in primary antibody, mild antigen retrieval (incubation for 30 min in 0.01 M sodium citrate, pH 6.0, as the solution cooled from boiling) was applied to slides to expose the P63 antigen without destroying the thick tissue sections. An antibody to GFP was used to visualize the membrane CFP (mCFP), cytosolic YFP (cYFP) and nuclear GFP (nGFP) reporters encoded by *Rosa26*^{Confetti}. Tissue was also labeled with an antibody to SOX2 to visualize the stem and progenitor cells, as well as the sustentacular and microvillous cells, and an antibody directed against P63 was used to visualize the HBC stem cells specifically. We visualized SOX2 immunolabeling with an Alexa-568 conjugated secondary antibody (Invitrogen) and P63 immunolabeling with an Alexa-647 conjugated secondary antibody (Invitrogen). We did not visualize the RFP reporter signal; there was no interference because the antigen retrieval necessary for optimal SOX2 and P63 visualization extinguished the RFP signal. We observed very few nGFP clones and could not judge cell morphology with its restricted localization; therefore, nGFP clones were excluded from our analysis. Identification of lineage-traced cell types was facilitated by cell morphology and position based on mCFP and cYFP expression. Olfactory sensory neurons were identified by their medially located somata and bipolar morphology

highlighted by a thin apical dendrite often terminating in a singular dendritic knob and the absence of SOX2 expression. Basal progenitors were identified by position and SOX2 expression. Microvillous cells were assigned based on the more apical position of their cell bodies, tapered, brush-like apical tufts and SOX2 expression. Sustentacular cells were identified by their apical localization, branched processes that span the epithelium, columnar apical shape, and SOX2 expression. Cells of the Bowman's gland were identified by their organization into multicellular ducts and/or sub-mucosal acini, and absence of SOX2 expression. Clones derived from HBCs that had a conditional knockout of *Sox2* were performed as described above for the *Krt5-CreER; Rosa26*^{Confetti} lineage tracing, except that SOX2 was no longer present in the clones because it was knocked out. Morphology was used in conjunction with P63 expression to define cell identity. For all *Krt5-CreER* clones, confocal z-stacks were obtained using a Zeiss LSM 710 or 780 confocal microscope, and images were processed and quantified using ImageJ (Schneider et al., 2012). For a breakdown of clones by animal and reporter, see Figures S1, S5, and S7.

The *Hopx*^{CreER} lineage tracing was performed in a similar manner to the *Krt5-CreER; Rosa26*^{Confetti} lineage tracing described above; however, the *Rosa26*^{eYFP} reporter was used. But despite using a standard dose of tamoxifen (0.25 mg tamoxifen/g body weight), there were so few clones and they were so sparsely spaced, it was clear that mostly (if not exclusively) clonal events were observed. Because *Hopx* is transiently expressed after injury in a subset of activated HBCs (predominantly HBC*1), to trace *Hopx*(+) cells, animals were injured first with methimazole, and injected with tamoxifen at 24 HPI. Tissue was fixed at two weeks following injury (14 DPI). Clones were visualized as described above, and immunohistochemistry for SOX2 was used to help identify cell types within clones. Five animals were analyzed for *Hopx* clones. Epifluorescence microscopy was used to score clones. As controls, five animals were injected with tamoxifen prior to injury, then injured, and fixed at two weeks following injury (14 DPI), and no clones were observed in any animal (data not shown). For a breakdown of clones by animal, see Figure S5). Clone quantitation is reported in Table S1.

Histology—Tissue was fixed at the indicated stages and time points with 4% paraformaldehyde for 16–18 hours at 4° C, washed with PBS, and decalcified in 10% ethylenediaminetetraacetic acid (EDTA) in PBS at 4° C for a minimum of 3–4 days, washed with distilled H_2O, and equilibrated in 30% sucrose in PBS overnight at 4° C before mounting and freezing in tissue freezing medium (Fisher). Tissue was sectioned at 12-micron thickness on a cryostat. For immunohistochemistry, tissue sections were incubated with PBS containing 0.1% Triton X-100 with primary antibodies diluted in 10% goat or donkey serum overnight at 4° C followed by at least three 20 minute washes in PBS with 0.1% Triton X-100. For the SOX2, NTUB/TuJ1, IL33, and P63 antibodies, we performed antigen retrieval by steaming slides for 15 minutes in 0.01 M sodium citrate followed by a PBS wash prior to the PBS with 0.1% Triton X-100 washes. Detection was performed by incubating samples with Alexa-488, 555, 568, 594, or 647-conjugated secondary antibodies at room temperature for 2 hours, followed by nuclear counterstaining. We used primary antibodies to the following antigens: SOX2, GFP, P63, Ki67, NTUB/TuJ1, Phospho-S6, IL33, NEUROD1, and CASPASE-3. Nuclei were counterstained with either Hoechst 33342

or DAPI, and/or tissue was visualized with brightfield illumination. Fluorescent confocal slices and z-stacks were obtained using Zeiss LSM 710 or 780 confocal microscopes. To label cells in S-phase of mitosis, 5-ethynyl-2′-deoxyuridine (EdU) was administered by intraperitoneal injection of animals (0.05 mg EdU/g of body weight). To detect EdU labeling, standard copper Click-iT chemistry was performed after antibody detection using standard manufacturer recommended protocols with Alexa Fluor conjugated azides (Thermo Fisher Scientific).

For RNA in situ hybridizations, tissue was fixed for a minimum of 24 hours at 4° C. Probes for RNA in situ hybridization were synthesized with either digoxygenin- or fluorescein-labeled UTP. Table S7 includes the oligonucleotides used for PCR amplification of templates for antisense RNA probes. The T7 primer sequence was on the 5-prime end of the reverse sequence oligonucleotide, and T7 RNA polymerase was used for in vitro transcription of probes. In brief, a standard RNA in situ hybridization protocol was used: slides were incubated in a prehybridization buffer for 2 hours. Probes were hybridized overnight at 65° C followed by multiple stringent washes in low salt buffer at 65° C. Subsequently, alkaline phosphatase-conjugated anti-digoxygenin or anti-fluorescein antibodies and BCIP/NBT substrates were used to visualize the hybridized probes. For fluorescence detection, the Perkin-Elmer Tyramide Signal Amplification (TSA) kit was used with additional washes following antibody incubation and fluorophore amplification; both FITC and Cy3 conjugated fluorophores were used. We used the RNAscope (Advanced Cell Diagnostics, Inc.) assay, using RNAscope probes and reagent kits, to detect *Trp63* and *Hopx* mRNA following the manufacturer's instructions on fresh frozen tissue. Confocal slices and z-stacks of fluorescent labeling were obtained using Zeiss LSM 710 or 780 confocal microscopes or by epifluorescence on a Nikon compound scope with a Leica DC500 or DFC500 camera. Brightfield images of colorimetric RNA in situ hybridization were obtained on a Nikon Microphot compound scope and Leica DFC500 camera. For RNA and protein expression profiles, a minimum of biological duplicates were analyzed.

Genetic Mutant Analysis—The $Sox2^{CreER}; Rosa26^{eYFP}$ transgenic was used to lineage trace the Sox2(+) cells in the uninjured olfactory epithelium and tissue was fixed and examined at the indicated hours or days post tamoxifen (HPT, DPT) (Figure S6A,B). Transgenic animals were injected with tamoxifen at P21 to activate CreER and then fixed at the indicated times. At least three biological replicates were examined for each time-point. These time-course images were acquired using epifluorescence on a Nikon Microphot compound microscope and a Leica DFC500 camera and processed and quantified in Adobe Photoshop. The $Sox2^{CreER}$ knock-in transgene was used in conjunction with the conditional $Sox2^{lox}$ allele to ablate the *Sox2* gene to assess the role of *Sox2* in the uninjured olfactory epithelium (Figure S6C–F). P21 animals were injected with tamoxifen and examined at 3-weeks post tamoxifen by immunohistochemistry for the lineage tracer and indicated markers. Four biological replicates were analyzed for each genotype, and a total of 4–5 mm of olfactory epithelium was examined for each biological replicate. To assess the impact of *Sox2* on HBCs under homeostatic conditions, we used the *Krt5-CrePR* driver to conditionally knockout *Sox2* in the HBCs and examined three biological replicates for each condition at three weeks and nine weeks of age (Figure S6G,H). The *Krt5-CrePR* transgene

does not require activation in the olfactory epithelium and is constitutively expressed after birth, commencing at approximately P3 (Fletcher et al., 2011). As a second approach, to assess the role of Sox2 in stem cell maintenance and differentiation in the uninjured olfactory epithelium, we used the *Krt5-CrePR* transgene in conjunction with the *Trp63* and *Sox2* conditional KO alleles. Three biological replicates were examined for each of three genotypes (1. $Sox2^{lox/lox}; Trp63^{lox/+}$; 2. $Sox2^{lox/+}; Trp63^{lox/lox}$; 3. $Sox2^{lox/lox}; Trp63^{lox/lox}$) with the indicated markers (Figure S6I,J). A minimum of 2 mm of olfactory epithelium was examined for each biological replicate. To assess the role of Sox2 during regeneration, the *Krt5-CreER*, $Sox2^{lox}$, $Rosa26^{eYFP}$ transgenic alleles were used. P21 animals were injected with tamoxifen (0.25 mg tamoxifen/g body weight), and injury was induced with methimazole (0.05 mg methimazole/g body weight) three days later. To label cells in S-phase of the cell cycle, 5-ethynyl-2′-deoxyuridine (EdU) (0.05 mg EdU/g of body weight) was injected once per day for three consecutive days at 24, 48, and 72 HPI.

Tissue was collected and fixed at 14 DPI. Immunohistochemistry was performed for the indicated markers, and EdU was visualized as described above. Three biological replicates (n = 3) were analyzed for each of $Sox2^{lox/+}$ and $Sox2^{lox/lox}$, and a minimum of 2 mm of olfactory epithelium were counted for each replicate. All images related to the comparison of controls and *Sox2* conditional knockouts were obtained on confocal microscopes, either Zeiss LSM 710 or 780 models; images were processed and counted in either Adobe Photoshop or NIH ImageJ.

Quantification and Statistical Analysis

Statistical Analysis of Genetic Mutants and Clonal Lineage Tracing—To test for differences in the number of YFP-positive and/or specific marker positive cells between the wild-type or heterozygous and conditional knockout samples, we implemented a negative binomial regression model with the glm function in R. We applied the same approach to different tracing strategies in the clonal analysis. Additionally, to compare percentages of marked cells between genotypes or tracing strategies, we fit the model to the marker-specific or cell-type counts and included the total YFP-positive cell counts as an offset term. This allowed us to account for differences in the number of YFP-positive cells while still modeling the data as counts and allowing for possible overdispersion. All p-values were adjusted for multiple testing using the Benjamini-Hochberg method implemented by the p.adjust function in R and are presented in the text and appropriate figure legends. To be conservative, p-values less than 10E-3 were presented as $p < 0.001$.

Single-Cell RNA-Seq Alignment and Filtering—Reads were aligned to the GRCm38.3 (mm10, patch release 3) mouse genome assembly with Tophat2 (Version 2.1.1; Kim et al., 2013), and low quality reads were removed with Trimmomatic (Version 0.3.2; Bolger et al., 2014). We used RefSeq transcript annotations, which were modified to contain sequences of specific genes used in our analysis (e.g. *CreER* and *eGFP*), and we counted the number of reads aligning to every gene (defined as the union of all splice forms) with featureCounts (Version 1.5.0-p3; Liao et al., 2014). Reads that aligned to more than one gene as well as chimeric fragments were excluded. We also removed the 37 genes that failed to be quantified in at least one sample by Cufflinks (Trapnell et al., 2010).

We implemented a quality control (QC) pipeline that computes an extensive set of quality metrics, relying in part on FastQC (Version 0.3.2) and the Picard suite of alignment metrics (Version 2.5.0 with samtools 1.3.1; Li et al., 2009). We used the open-source R package SCONE (https://bioconductor.org/packages/scone; Version 0.0.7) to perform data-adaptive quality metric-based cell filtering. This yielded the following filtering criteria: any cell with fewer than 100,000 aligned reads or a percentage of aligned reads below 88.8% was filtered out. We identified cells that were non-sensory contaminants by expression of *Reg3g* in uninjured cells and high expression of *Csf1r* combined with low expression of eGFP. We also eliminated doublets by co-expression of known neuronal (*Omp*) and sustentacular cell markers (*Cyp1a2* or *Cyp2g1*). After filtering out low quality cells and removing doublets and non-sensory epithelial contaminants, 672 cells (out of 849) remained. Finally, to ensure stability in subsequent cluster analysis, we retained only those genes having at least 40 reads in at least 5 cells (12,781 genes).

Normalization of Single-Cell RNA-Seq Data—We performed and assessed several normalization schemes using SCONE, based on a set of 9 data-driven performance metrics. These metrics aim to capture two main features of each normalization procedure: the ability to remove unwanted technical variation and the ability to preserve wanted biological variation of interest. The first group of performance metrics includes the correlation of expression measures with factors derived from validated markers of various olfactory epithelium cell types (positive control genes) and the average silhouette width (Rousseeuw, 1987) of the obtained clusters (cluster quality). Metrics in the second group include the correlation of expression measures with factors derived from “housekeeping” negative control genes (obtained from our earlier microarray experiments analyzing injury-induced regeneration of the olfactory epithelium) and the correlation between expression measures and quality control (QC) measures. See the *scone* package vignettes (available at https://github.com/YosefLab/scone) for more details. According to SCONE, full-quantile normalization (Bolstad et al., 2003; Bullard et al., 2010), followed by regression-based adjustment for housekeeping genes, was among the best performing normalizations. We specifically chose one factor of adjustment, as previous experience had demonstrated that removing additional factors removed known biological variation (data not shown); the selected normalization was the top ranked method once this criterion was taken into account. Specifically, principal component analysis (PCA) was applied to the matrix of housekeeping genes, and the first PC was used as a quantitative factor of “unwanted” technical variation across cells. Then, the log-transformed quantile-normalized expression measures (adding a pseudocount of 1 prior to the log transformation) for each gene were modeled as a linear function of this technical covariate. The normalized expression measures were defined as the residuals from the linear model fit, rescaled to have the same mean as the log-transformed quantile-normalized read counts.

Clustering of Single-Cell RNA-Seq Data—We used the clustering framework RSEC (Resampling-based Sequential Ensemble Clustering) to obtain stable and tight cell clusters. The method is implemented in the open-source *clusterExperiment* R package (Version 0.99.3-9001) available at Bioconductor Project (http://bioconductor.org/packages/clusterExperiment). Briefly, RSEC comprises the following steps. Given a base clustering

algorithm, which we chose to be *k-means*, where the parameter *k* determines the number of clusters (Hartigan and Wong, 1979), RSEC creates ensemble clusters by the following steps: 1) repeatedly subsampling the observations (cells or genes, respectively), 2) clustering each set of subsampled observations with *k-means*, and then 3) forming a final clustering determined by clustering samples based on the percentage of subsamples for which two observations were assigned to the same cluster. This procedure was repeated for increasing *k* (the number of clusters), in order to find the cluster that changed the least; this cluster was deemed the most stable and removed, and then the entire clustering procedure was repeated on the remaining data. This sequential strategy follows that of Tseng and Wong (2005) and ensures that outlying clusters do not dominate the clustering; it also alleviates the dependence on the number of clusters *k* in lieu of other parameters that define required cluster similarity and stability. RSEC generates a large collection of such cluster sets by repeating the above procedure for different choices of the parameters defining similarity and stability; it then identifies a *consensus* over the different candidate sets based on the co-clustering of observations. This approach tends to result in a large number of small clusters that sometimes do not differ substantially in terms of gene expression. Hence, the last step of RSEC is to merge closely related clusters that do not exhibit differential expression. The vignette of the *clusterExperiment* package provides additional details on the RSEC algorithm (available at https://www.bioconductor.org/packages/clusterExperiment).

When applied to the first 50 principal components of the entire expression matrix; RSEC found 12 (wild type) and 16 (combined wild type and *Sox2* conditional knockout) stable clusters, which were then manually inspected for the presence of marker genes of known cell types to make preliminary assignments of cell type identities. A final repertoire of 11 (wild type) and 14 (wild type and *Sox2* conditional knockout) clusters was obtained after removing additional contaminants marked by *Fabp5* and low quality samples marked by extremely poor alignment to coding regions (data not shown).

Preliminary assignment of cluster identities were determined as follows, as in our previous characterization of olfactory epithelial clusters (Fletcher et al., 2017). Resting HBCs are characterized by expression of HBC markers *Trp63*, *Krt5*, and *Krt14*. GBCs are marked by expression of *Kit*, *Ascl1*, and cell cycle genes and microvillous cells by *Ascl3* and *Cftr*; these distinct cell types, validated experimentally, were represented by one combined cluster. Intermediate neuronal precursors were represented by two clusters, one cluster containing cells expressing high levels of *Neurod1*, cell cycle genes, and *Lhx2* (denoted INP1/2, to reflect the presence of both INP1 and INP2 cell types), and another expressing lower levels of *Lhx2* but also expressing *Gap43* (INP3). *Gap43* expression is highest in the immature neurons, and the mature neurons express not only high levels of *Omp* but also the cyclic nucleotide-gated channel subunit *Cnga2* and *Gγ13*. The sustentacular cell cluster is a combined cluster of both immature (*Reg3g*) and mature sustentacular cells (*Cyp2g1* and *Cyp1a2*).

t-Distributed Stochastic Neighbor Embedding—To display the relative distances between cells in a lower-dimensional representation of gene expression space, we employed t-distributed stochastic neighbor embedding (t-SNE; van der Maaten, 2014; van der Maaten and Hinton, 2008). t-SNE is a method of dimensionality reduction that excels at representing

distances on multiple scales. We use it here to visualize our data independently from how we generated the cell clusters (for clustering, see RSEC above). We use the Barnes-Hut implementation, as available in the *Rtsne* R package. We set perplexity to 30 with 1000 iterations; varying the perplexity and increasing iterations does not alter our conclusions about the congruence between the t-SNE visualization and our clustering results (data not shown).

Cell Lineages and Developmental Distance—We used a recently developed cell lineage inference algorithm, Slingshot (Street et al., 2017; Version 0.0.3-4, available as an open-source R package *slingshot* at https://github.com/kstreet13/slingshot), to identify lineage trajectories and bifurcations and to order cells along trajectories. Slingshot takes as input a matrix of reduced dimension normalized expression measures (e.g., PCA) and cell clustering assignments. It infers lineage trajectories and branch points by connecting the cluster medoids using a minimum spanning tree (MST) and identifying the starting cluster or root node. Lineages are defined by ordered sets of clusters beginning with the root node and terminating in the most distal cluster(s) with only one connection. Next, principal curves (Hastie and Stuetzle, 1989) are fit to the subsets of cells making up each lineage, providing a smooth, nonlinear summary of each trajectory. Individual cells are then orthogonally projected to each curve and thereby ordered in a space reflecting developmental distance. The ordering provided by Slingshot, analogous to pseudotime, is referred to herein as developmental order.

After an additional merging step, a single cluster containing fewer than 10 cells (<1% of the total population studied) was deemed to be less reliable and not included in the input to Slingshot for the combined wild type and *Sox2* conditional knockout dataset. The wild type dataset produced no clusters with fewer than 10 cells. The cluster representing the earliest activated HBCs (24 hour cells) was chosen as the root node. In addition to assignment of the starting point, *slingshot* allows for the user to specify known end points of the developmental lineage to better guide its identification. Using this feature, we assigned mature sustentacular cells as an endpoint. Note that this assignment only constrains that cluster; all other clusters are allowed to be placed anywhere on a lineage, and additional endpoints can be found beyond this one, as was the case in our data. *slingshot* then generated simultaneous principal curves and cell developmental distances for each lineage. *slingshot* was applied to the first six principal components of the normalized expression matrix. Six components were chosen based on examining the separation of cells along individual PCs (Figure S2E), although the MST was stable across a range of PCs, lending credence to this model (data not shown).

Differential Expression—We used *limma* for differential expression (DE) analysis between clusters within each lineage, applying a one-versus-all approach (i.e., comparing the average of one cluster to the average of all the other clusters in that lineage, Table S3, sheets 2 and 3). For each lineage, the top 500 DE genes (lowest adjusted p-value) for each one-versus-all comparison were retained. For the DE analysis between activated and transitional HBCs presented in Figure 3C and Table S3, we used *limma voom*, applying a pair-wise comparison between a combined cluster of activated HBCs and a combined cluster

of transitional HBCs (Table S3, sheet 1). To identify genes enriched in the HBC*1 cluster relative to resting HBCs for binding site motif enrichment analysis, we performed pair-wise DE analysis between HBC*1 and HBC cell clusters and selected the 1000 genes with the highest log-fold enrichment in HBC*1, using an adjusted p-value cutoff of 0.05 (Table S3, sheet 6).

Gene Set Enrichment Analysis—After clustering the cells, we used *limma* (Version 3.28.19; Smyth, 2004) for DE analysis as implemented in the *clusterExperiment* package. We applied *limma's* implementation of the ROMER rotation test to perform gene set enrichment analysis (GSEA) (Ritchie et al., 2015) with 10^7 rotations contrasting each cluster. Gene sets were taken from the Molecular Signatures Database (MSigDB) at the Broad Institute and included all Hallmark gene sets (H) in addition to the Canonical pathways, KEGG, and Reactome gene sets from the curated (C2) gene sets.

Volcano Plots—We used *clusterExperiment's* wrapper around *limma* for DE analysis across all lineages applying a *pairwise* approach (i.e., comparing the average of one cluster to the average of each other cluster, Table S2). We plotted the −log10 adjusted p-value (Benjamini-Hochberg correction) vs. log2 fold-change in expression for each pairwise comparison corresponding to a transition predicted by Slingshot. Volcano plots highlight genes that are down-regulated or up-regulated with a fold-change greater than two (logFC > 1) and an adjusted p-value less than 0.01.

Transcription Factor Co-Expression Networks—Transcription factor network diagrams were based upon transcription factor genes that were among the 500 most differentially expressed genes within each lineage, using the previously mentioned one-versus-all approach (Table S3). A list of transcription factors was obtained from the Animal Transcription Factor Database. Only transcription factors that had a correlation of at least 0.3 with at least 5 other differentially expressed transcription factors along the lineage were included in the correlation network. This correlation threshold was chosen to be greater than the maximum pairwise correlation after random permutation, as previously done in Treutlein et al. (2016).

Transcription Factor Binding Site Motif Analysis—We used *Homer's* findMotifs function (Heinz et al., 2010) to identify known transcription factor binding site enrichment for the genes that were the most enriched in HBC*1 relative to resting HBCs. This list of differentially expressed genes was derived from pair-wise DE as described above (Table S3, sheet 6). The top 1000 genes ranked by log-fold change (all with adjusted p-values < 0.05) were used as input into *Homer's* findMotifs function. Nine hundred fifty genes were recognized by *Homer*. We analyzed the sequence from 1000 bp upstream of the transcriptional start site to 100 bp downstream. We modified the findMotifs.pl script to calculate adjusted p-values (q-values) to three significant digits. The top 30 conserved known binding sites are presented in Table S5.

Odorant Receptor Expression Analysis—Since odorant receptors (ORs) are typically expressed at one allele per mature olfactory sensory neuron, we chose to investigate transcripts per million (TPM) output from RSEM (Version 1.2.31; Li and Dewey, 2015) to

assess OR expression in the neuronal lineage instead of filtered counts because most olfactory receptors would have been filtered out. For this analysis, we aligned reads to the transcriptome with Bowtie2 (Version 2.2.9; Langmead and Salzberg, 2012) and quantified gene expression with RSEM, while setting the alignment parameters to the ones recommended by RSEM's authors. Analysis of OR expression using Kallisto (Bray et al., 2016) TPMs gave similar results (data not shown).

Data and Software Availability

Accession Numbers—RNA-seq data have been deposited in Gene Expression Omnibus (accession numbers GSE99251 and GSE95601).

Software—Analysis scripts for this dataset can be found at https://github.com/diyadas/HBC-regen. The R software packages scone, clusterExperiment, and slingshot are available on GitHub and Bioconductor as described above.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Hector Nolla and Alma Valeros for their expertise and assistance with fluorescence-activated cell sorting, Shana McDevitt, Karen Lundy, Anett Schmittfull, and Minyong Chung for their tireless efforts on RNA sequencing, Burke Bundy for assistance with cluster computing, and members of the Ngai laboratory for invaluable advice and suggestions over the course of this project. Sarah Tronnes created the graphical abstract. This work was supported by grants from the National Institutes of Health (R01DC007235, U01MH105979, S10RR029668, S10RR027303) and the Siebel Stem Cell Center. R.B.F. was supported by a grant from the National Institute on Aging (K01AG045344). D.D. and K.N.S. were supported by a training grant from the National Human Genome Research Institute (T32HG000047) and L.G. by a training grant from the National Institute of General Medical Sciences (T32GM098218). R.B.F. and D.D. were also supported by training grant TG2-01164 from the California Institute of Regenerative Medicine CIRM Scholars program. D.D. is a fellow of the Berkeley Institute for Data Science, which is funded in part by the Gordon and Betty Moore Foundation (Grant GBMF3834) and the Alfred P. Sloan Foundation (Grant 2013-10-27).

References

Arnold K, Sarkar A, Yram MA, Polo JM, Bronson R, Sengupta S, Seandel M, Geijsen N, Hochedlinger K. Sox2(+) adult stem and progenitor cells are important for tissue regeneration and survival of mice. Cell Stem Cell. 2011; 9:317–329. [PubMed: 21982232]

Bazzi H, Fantauzzo KA, Richardson GD, Jahoda CAB, Christiano AM. Transcriptional profiling of developing mouse epidermis reveals novel patterns of coordinated gene expression. Dev Dyn. 2007; 236:961–970. [PubMed: 17330888]

Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014; 30:2114–2120. [PubMed: 24695404]

Bolstad BM, Irizarry RA, Astrand M, Speed TP. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. Bioinforma Oxf Engl. 2003; 19:185–193.

Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. Nat Biotechnol. 2016; 34:525–527. [PubMed: 27043002]

Bullard JH, Purdom E, Hansen KD, Dudoit S. Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. BMC Bioinformatics. 2010; 11:94. [PubMed: 20167110]

Caggiano M, Kauer JS, Hunter DD. Globose basal cells are neuronal progenitors in the olfactory epithelium: A lineage analysis using a replication-incompetent retrovirus. Neuron. 1994; 13:339–352. [PubMed: 8060615]

Collins CA, Olsen I, Zammit PS, Heslop L, Petrie A, Partridge TA, Morgan JE. Stem Cell Function, Self-Renewal, and Behavioral Heterogeneity of Cells from the Adult Muscle Satellite Cell Niche. Cell. 2005; 122:289–301. [PubMed: 16051152]

Fletcher RB, Prasol MS, Estrada J, Baudhuin A, Vranizan K, Choi YG, Ngai J. p63 Regulates Olfactory Stem Cell Self-Renewal and Differentiation. Neuron. 2011; 72:748–759. [PubMed: 22153372]

Fletcher RB, Das D, Gadye L, Street KN, Baudhuin A, Wagner A, Cole MB, Flores Q, Choi YG, Yosef N, et al. Deconstructing Olfactory Stem Cell Trajectories at Single-Cell Resolution. Cell Stem Cell. 2017; 20:817–830. e8. [PubMed: 28506465]

Hartigan JA, Wong MA. Algorithm AS 136: A K-Means Clustering Algorithm. Appl Stat. 1979; 28:100.

Hastie T, Stuetzle W. Principal Curves. J Am Stat Assoc. 1989; 84:502–516.

Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK. Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. Mol Cell. 2010; 38:576–589. [PubMed: 20513432]

Henry J, Toulza E, Hsu CY, Pellerin L, Balica S, Mazereeuw-Hautier J, Serre G, Jonca N, Simon M. Update on the epidermal differentiation complex. Front Biosci. 2012; 17:1517–1532.

Indra AK, Warot X, Brocard J, Bornert JM, Xiao JH, Chambon P, Metzger D. Temporally-controlled site-specific mutagenesis in the basal layer of the epidermis: comparison of the recombinase activity of the tamoxifen-inducible Cre-ER(T) and Cre-ER(T2) recombinases. Nucleic Acids Res. 1999; 27:4324–4327. [PubMed: 10536138]

Ito M, Liu Y, Yang Z, Nguyen J, Liang F, Morris RJ, Cotsarelis G. Stem cells in the hair follicle bulge contribute to wound repair but not to homeostasis of the epidermis. Nat Med. 2005; 11:1351–1354. [PubMed: 16288281]

Iwai N, Zhou Z, Roop DR, Behringer RR. Horizontal Basal Cells Are Multipotent Progenitors in Normal and Injured Adult Olfactory Epithelium. STEM CELLS. 2008; 26:1298–1306. [PubMed: 18308944]

Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. 2013; 14:R36. [PubMed: 23618408]

Kuang S, Kuroda K, Le Grand F, Rudnicki MA. Asymmetric self-renewal and commitment of satellite stem cells in muscle. Cell. 2007; 129:999–1010. [PubMed: 17540178]

Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012; 9:357–359. [PubMed: 22388286]

Leung CT, Coulombe PA, Reed RR. Contribution of olfactory neural stem cells to tissue maintenance and regeneration. Nat Neurosci. 2007; 10:720–726. [PubMed: 17468753]

Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics. 2015; 12:323.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009; 25:2078–2079. [PubMed: 19505943]

Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics. 2014; 30:923–930. [PubMed: 24227677]

Ma XM, Blenis J. Molecular mechanisms of mTOR-mediated translational control. Nat Rev Mol Cell Biol. 2009; 10:307–318. [PubMed: 19339977]

van der Maaten L. Accelerating t-SNE using tree-based algorithms. J Mach Learn Res. 2014; 15:3221–3245.

van der Maaten L, Hinton G. Visualizing Data using t-SNE. J Mach Learn Res. 2008; 9:2579–2605.

Mills AA, Qi Y, Bradley A. Conditional inactivation of p63 by Cre-mediated excision. Genesis. 2002; 32:138–141. [PubMed: 11857801]

Oshio T, Komine M, Tsuda H, Tominaga SI, Saito H, Nakae S, Ohtsuki M. Nuclear expression of IL-33 in epidermal keratinocytes promotes wound healing in mice. J Dermatol Sci. 2017; 85:106–114. [PubMed: 27839630]

Packard AI, Lin B, Schwob JE. Sox2 and Pax6 Play Counteracting Roles in Regulating Neurogenesis within the Murine Olfactory Epithelium. PLoS ONE. 2016; 11(5):e0155167. [PubMed: 27171428]

Pevny LH, Nicolis SK. Sox2 roles in neural stem cells. Int J Biochem Cell Biol. 2010; 42:421–424. [PubMed: 19733254]

Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015; 43:e47–e47. [PubMed: 25605792]

Rousseeuw PJ. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. J Comput Appl Math. 1987; 20:53–65.

Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. Nat Methods. 2012; 9:671–675. [PubMed: 22930834]

Schwob JE, Huard JM, Luskin MB, Youngentob SL. Retroviral lineage studies of the rat olfactory epithelium. Chem Senses. 1994; 19:671–682. [PubMed: 7735846]

Schwob JE, Youngentob SL, Mezza RC. Reconstitution of the rat olfactory epithelium after methyl bromide-induced lesion. J Comp Neurol. 1995; 359:15–37. [PubMed: 8557844]

Shaham O, Smith AN, Robinson ML, Taketo MM, Lang RA, Ashery-Padan R. Pax6 is essential for lens fiber cell differentiation. Development. 2009; 136:2567–2578. [PubMed: 19570848]

Shirakata Y, Kimura R, Nanba D, Iwamoto R, Tokumaru S, Morimoto C, Yokota K, Nakamura M, Sayama K, Mekada E, et al. Heparin-binding EGF-like growth factor accelerates keratinocyte migration and skin wound healing. J Cell Sci. 2005; 118:2363–2370. [PubMed: 15923649]

Shwartz A, Yogev S, Schejter ED, Shilo BZ. Sequential activation of ETS proteins provides a sustained transcriptional response to EGFR signaling. Development. 2013; 140:2746–2754. [PubMed: 23757412]

Smyth GK. Linear Models and Empirical Bayes Methods for Assessing Differential Expression in Microarray Experiments. Stat Appl Genet Mol Biol. 2004; 3:1–25.

Snippert HJ, van der Flier LG, Sato T, van Es JH, van den Born M, Kroon-Veenboer C, Barker N, Klein AM, van Rheenen J, Simons BD, et al. Intestinal Crypt Homeostasis Results from Neutral Competition between Symmetrically Dividing Lgr5 Stem Cells. Cell. 2010; 143:134–144. [PubMed: 20887898]

Srinivas S, Watanabe T, Lin CS, William CM, Tanabe Y, Jessell TM, Costantini F. Cre reporter strains produced by targeted insertion of EYFP and ECFP into the ROSA26 locus. BMC Dev Biol. 2001; 1:4. [PubMed: 11299042]

Street, K., Risso, D., Fletcher, RB., Das, D., Ngai, J., Yosef, N., Purdom, E., Dudoit, S. Slingshot: Cell lineage and pseudotime inference for single-cell transcriptomics; bioRxiv. 2017. p. 128843doi: https://doi.org/10.1101/128843

Suh H, Consiglio A, Ray J, Sawai T, D'Amour KA, Gage FH. In vivo fate analysis reveals the multipotent and self-renewal capacities of Sox2+ neural stem cells in the adult hippocampus. Cell Stem Cell. 2007; 1:515–528. [PubMed: 18371391]

Takeda N, Jain R, LeBoeuf MR, Wang Q, Lu MM, Epstein JA. Interconversion between intestinal stem cell populations in distinct niches. Science. 2011; 334:1420–1424. [PubMed: 22075725]

Tetteh PW, Basak O, Farin HF, Wiebrands K, Kretzschmar K, Begthel H, van den Born M, Korving J, de Sauvage F, van Es JH, et al. Replacement of Lost Lgr5-Positive Stem Cells through Plasticity of Their Enterocyte-Lineage Daughters. Cell Stem Cell. 2016; 18:203–213. [PubMed: 26831517]

Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol. 2010; 28:511–515. [PubMed: 20436464]

Treutlein B, Lee QY, Camp JG, Mall M, Koh W, Shariati SAM, Sim S, Neff NF, Skotheim JM, Wernig M, et al. Dissecting direct reprogramming from fibroblast to neuron using single-cell RNA-seq. Nature. 2016; 534:391–395. [PubMed: 27281220]

Tseng GC, Wong WH. Tight Clustering: A Resampling-Based Approach for Identifying Stable and Tight Patterns in Data. Biometrics. 2005; 61:10–16. [PubMed: 15737073]

Wilson A, Laurenti E, Oser G, van der Wath RC, Blanco-Bose W, Jaworski M, Offner S, Dunant CF, Eshkind L, Bockamp E, et al. Hematopoietic Stem Cells Reversibly Switch from Dormancy to Self-Renewal during Homeostasis and Repair. Cell. 2008; 135:1118–1129. [PubMed: 19062086]

Wojcik SM, Bundman DS, Roop DR. Delayed wound healing in keratin 6a knockout mice. Mol Cell Biol. 2000; 20:5248–5255. [PubMed: 10866680]

Wu AR, Neff NF, Kalisky T, Dalerba P, Treutlein B, Rothenberg ME, Mburu FM, Mantalas GL, Sim S, Clarke MF, et al. Quantitative assessment of single-cell RNA-sequencing methods. Nat Methods. 2014; 11:41–46. [PubMed: 24141493]

Yan KS, Chia LA, Li X, Ootani A, Su J, Lee JY, Su N, Luo Y, Heilshorn SC, Amieva MR. The intestinal stem cell markers Bmi1 and Lgr5 identify two functionally distinct populations. Proc Natl Acad Sci U A. 2012; 109:466–471.

Yates S, Rayner TE. Transcription factor activation in response to cutaneous injury: role of AP-1 in reepithelialization. Wound Repair Regen. 2002; 10:5–15. [PubMed: 11983002]

Zhou Z, Wang D, Wang XJ, Roop DR. In utero activation of K5. CrePR1 induces gene deletion. Genesis. 2002; 32:191–192. [PubMed: 11857819]

Highlights

- Single-cell RNA-seq reveals injury-induced activated olfactory stem cell states.
- Activated olfactory stem cells are transient and unique to regeneration.
- Genes associated with epithelial wound repair are upregulated in activated stem cells.
- Cells in the activated states are heterogeneous and give rise to multiple lineages.

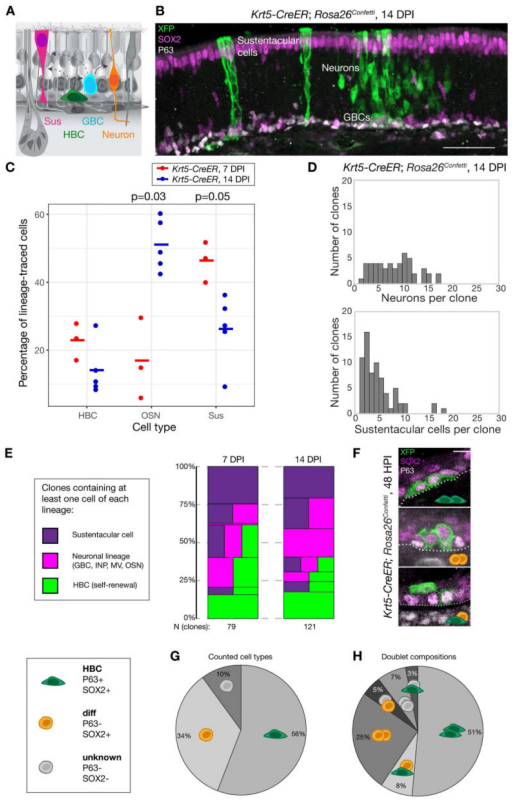


Figure 1. Clonal Analysis of HBC-Derived Cells During Regeneration

(A) Schematic of olfactory cell types produced by HBCs during regeneration. HBC (green), horizontal basal cell; GBC (cyan), globose basal cell; Sus (magenta), sustentacular cell; olfactory sensory neuron (OSN, orange). **(B)** Maximum projection of a 40μm tissue section of regenerating olfactory epithelium at 14 DPI from a *Krt5-CreER; Rosa26*^{Confetti} animal in which Cre was sparsely activated. Reporter localization detected using a GFP antibody coupled with antibodies to SOX2 (expressed by HBCs, GBCs, and sustentacular cells; magenta) and P63 (expressed by HBCs alone; white) were used along with cellular morphology to discriminate cell types and clonal relationships in YFP- and CFP-positive cells. Scale bar, 50 μm. **(C)** Distributions of ratios of HBCs, OSNs, and sustentacular cells with respect to total cells counted across all clones, by animal and by experimental time-point (7 DPI and 14 DPI); the mean, across animals, is indicated by a dash in each condition. P-values were calculated using a negative binomial regression model, and the Benjamini-Hochberg method was used to adjust for multiple testing (STAR Methods). **(D)** Distributions of neurons (top) and sustentacular cells (bottom) per clone in clones possessing at least one of each cell type, respectively, at 14 DPI. **(E)** Ratios of clones scored by lineage representation at 7 DPI and 14 DPI. Clones were binned on the basis of containing at least one cell of one of the three major lineages (HBC, green; neuron, magenta; sustentacular cell, purple). Multi-lineage clones are indicated by different colored rectangles from left to right, while uni-lineage clones are represented by a single colored rectangle. **(F–H)** HBCs were lineage-traced in *Krt5-CreER; Rosa26*^{Confetti} animals; tissue was collected at 48 HPI to assess the fates of HBC-derived doublets. **(F)** Representative maximum projections of 40μm confocal z-stacks show the three most commonly observed doublet compositions (top:

symmetric renewing; middle: symmetric differentiating; bottom: asymmetric) based on expression of marker genes. Basement membrane, dotted line; scale bar, 10 μm. **(G)** Percent of doublet cells that are renewed HBCs, differentiating HBCs, and unknown cells at 48 HPI. **(H)** Doublet compositions at 48 HPI. See Figure S1.

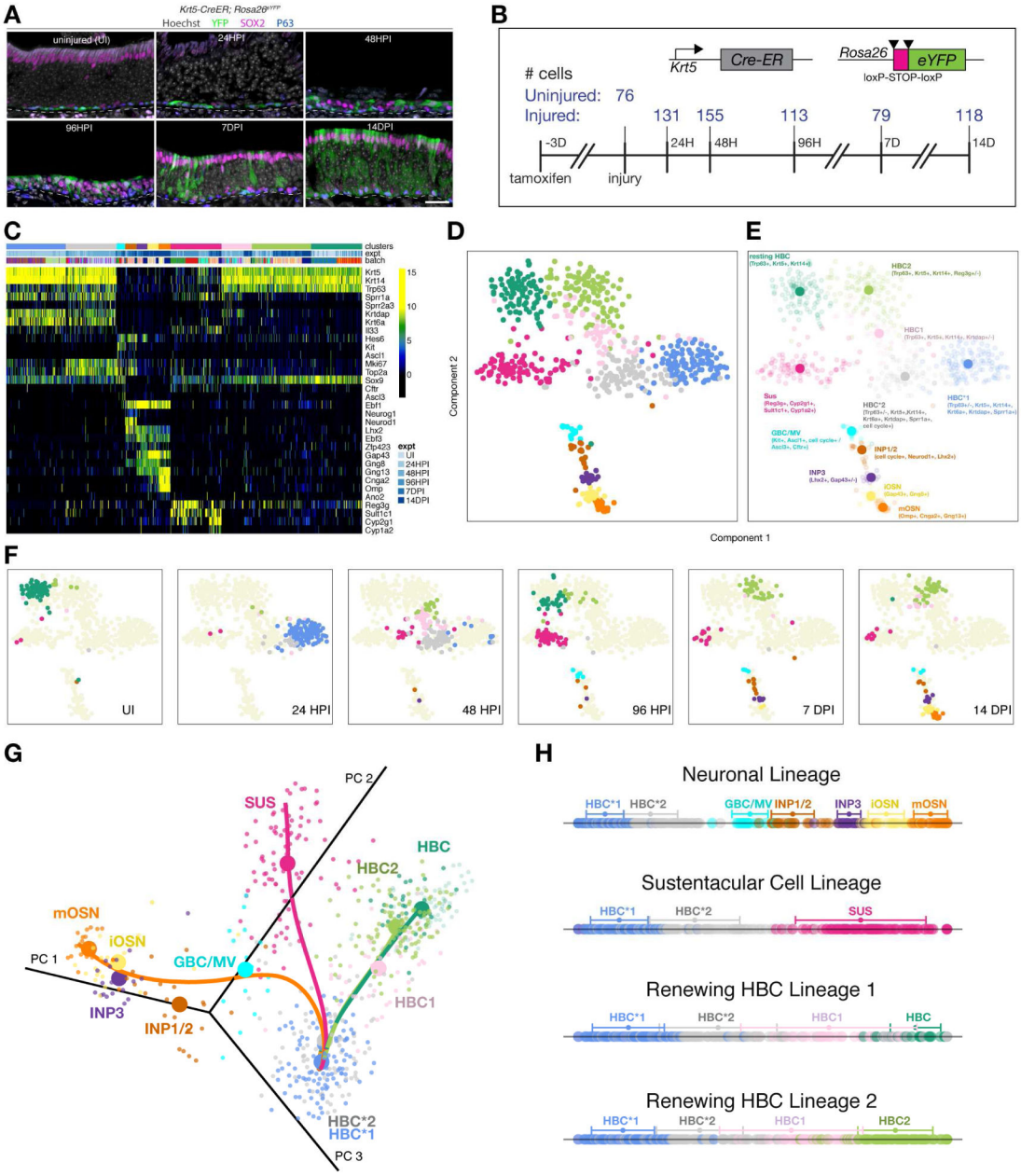


Figure 2. Single-Cell RNA-Sequencing Identifies Distinct Cell States in the Olfactory Stem Cell Trajectory Following Injury

(A) Immunohistochemistry for the HBC lineage tracer YFP (green), SOX2 (magenta) and P63 (blue), shows basal uninjured HBCs and differentiation of HBCs during injury-induced regeneration. Scale bar, 50 μm. **(B)** YFP(+) cells were collected by FACS at the indicated times following methimazole administration from mice carrying the *Krt5-CreER; Rosa26*^{eYFP} transgenes. **(C)** Heatmap of 32 marker genes for various cell types in the adult olfactory epithelium. The 11 cell clusters and experimental time points (expt) are indicated across the top. Batch refers to each biological replicate or FACS run (see STAR Methods). **(D, E)** t-SNE plots of cells, colored by cluster assignments. Cluster medoids and the marker

genes used to provisionally assign cluster identity are visualized in (E). **(F)** t-SNE plots as in (D), colored by cluster assignment for each experimental time point. **(G)** Three-dimensional PCA plot of single cell gene expression with branching lineage trajectories inferred by Slingshot, using the HBC*1 stage as a starting point. Neuronal (orange), sustentacular cell (magenta) and renewing HBC (green) lineages are predicted. **(H)** Projection of cells onto their respective lineage trajectories reflects the developmental order and distance. See Figure S2.

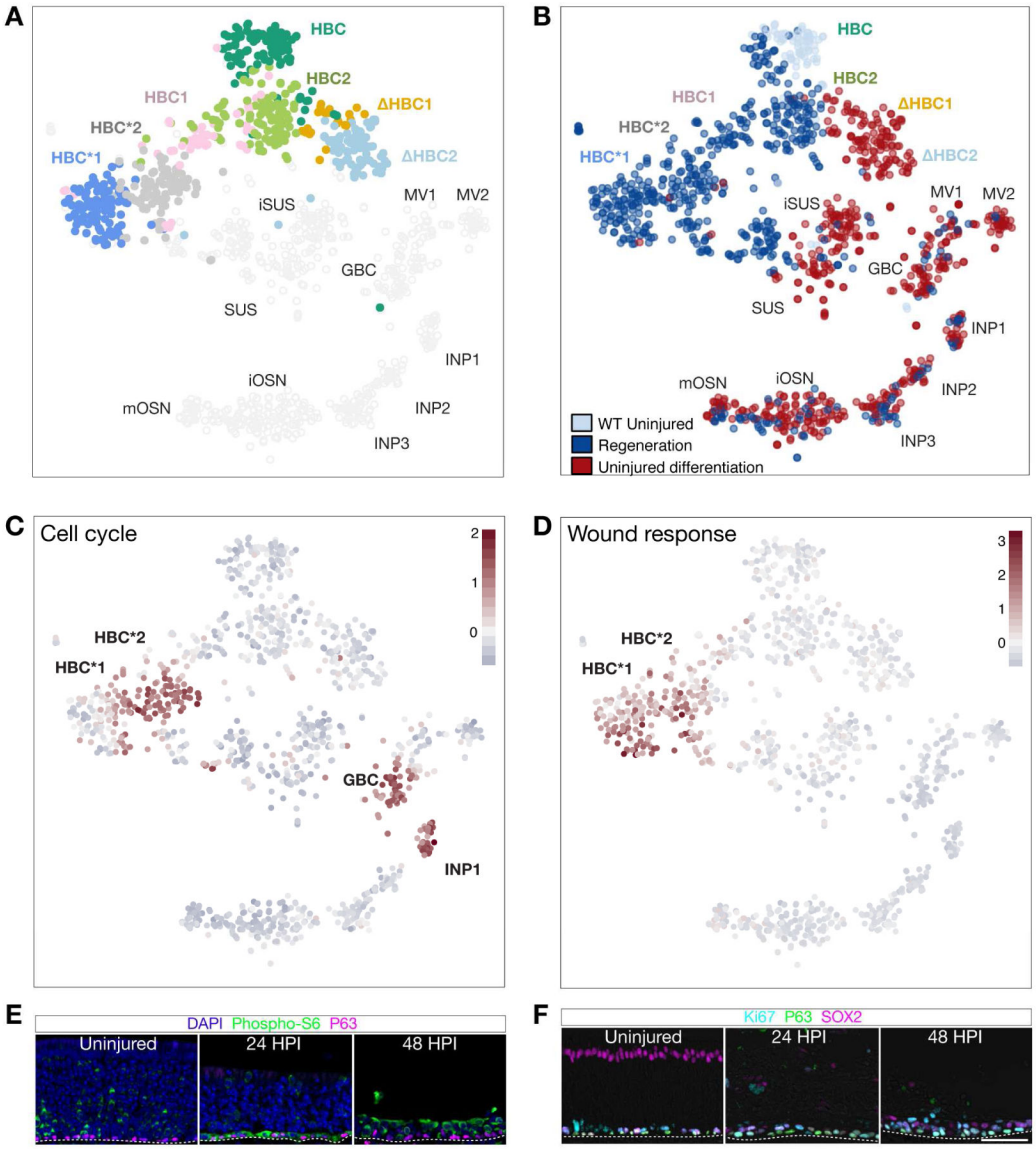


Figure 3. Activated HBCs are Distinct Cell States Unique to Injury-Induced Regeneration

(A) Normalized data from both the regeneration experiment (described in Figure 2B) and a separate time-course for differentiation induced by conditional knockout of *Trp63* under homeostatic conditions (Fletcher et al. 2017) are plotted in two dimensions using t-SNE, with HBC-specific clusters highlighted. **(B)** The same t-SNE plot as in (A) but colored by experimental origin of the cells; wild-type (WT) uninjured cells are shared between the experiments. **(C)** Visualization of normalized expression of 40 cell cycle-associated genes (see Table S7). **(D)** Visualization of normalized expression of 8 wound response genes (*Krt6a*, *Krt16*, *Sprr1a*, *Sprr2a3*, *Krtdap*, *Dmkn*, *Sbsn*, *Hbegf*). **(E)** Immunohistochemistry for phospho-S6 (a readout of protein translation) and P63 in the uninjured olfactory epithelium and at 24 and 48 hours following injury. **(F)** Immunohistrochemistry for Ki67 (a marker of cell proliferation), P63 and SOX2 in the uninjured epithelium and at 24 and 48 hours following injury. Scale bar, 50 μm.

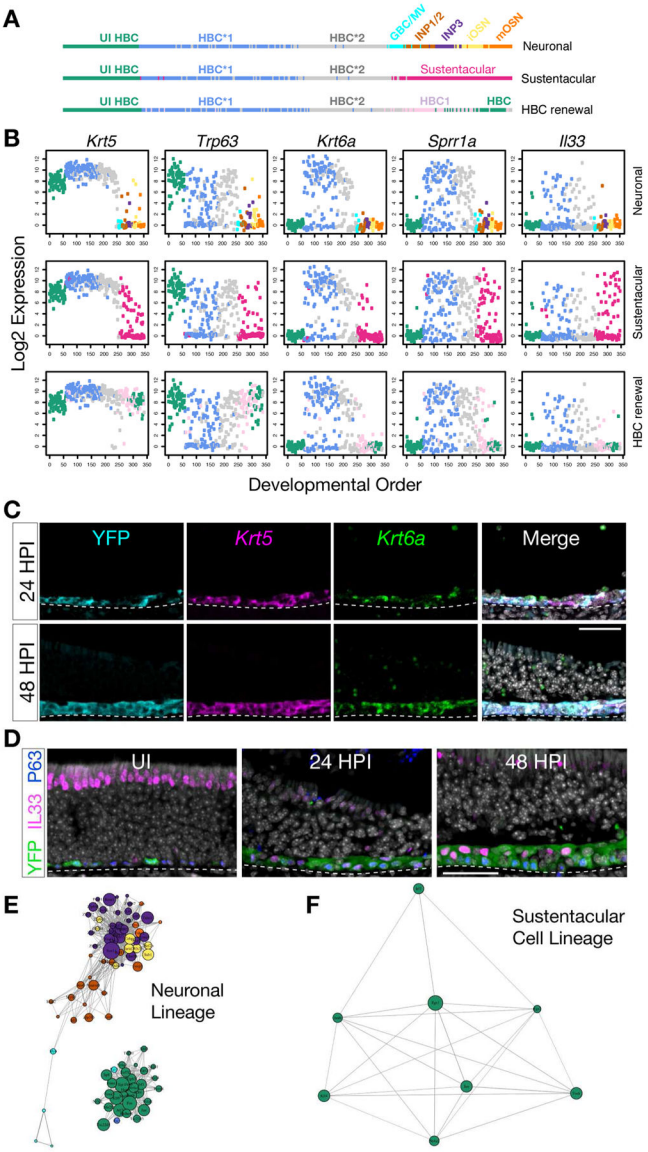


Figure 4. Activated HBCs are Heterogeneous

(A) Developmental ordering of cells in the lineages inferred by Slingshot. Uninjured HBCs that comprise the resting HBC cluster, which were not given as input to Slingshot, were appended to the beginning of each lineages as they are the de facto starting point prior to injury. **(B)** Selected genes associated with resting and activated HBCs are plotted for each cell in the lineage order. **(C)** Localization of *Krt5* and *Krt6a* in uninjured olfactory epithelium and at 24 and 48 HPI by RNA in situ hybridization. Cells lineage-traced by the *Krt5-CreER; Rosa26*^{eYFP} transgene were localized using an anti-GFP antibody. **(D)** Immunohistochemical localization of IL33, a marker of a subset of activated HBCs, and P63 in olfactory epithelium of *Krt5-CreER; Rosa26*^{eYFP} mice in uninjured tissue and at 24 and 48 HPI. **(E, F)** Connectivity graphs of the most differentially expressed transcription factors for the neuronal (E) and sustentacular cell (F) lineages. The lines connecting transcription factors represent correlated expression throughout the lineage and the transcription factor

nodes are colored by the cluster in which average expression is highest; the size of each node indicates magnitude of expression. See Table S3 for a list of the transcription factors displayed in each lineage graph.

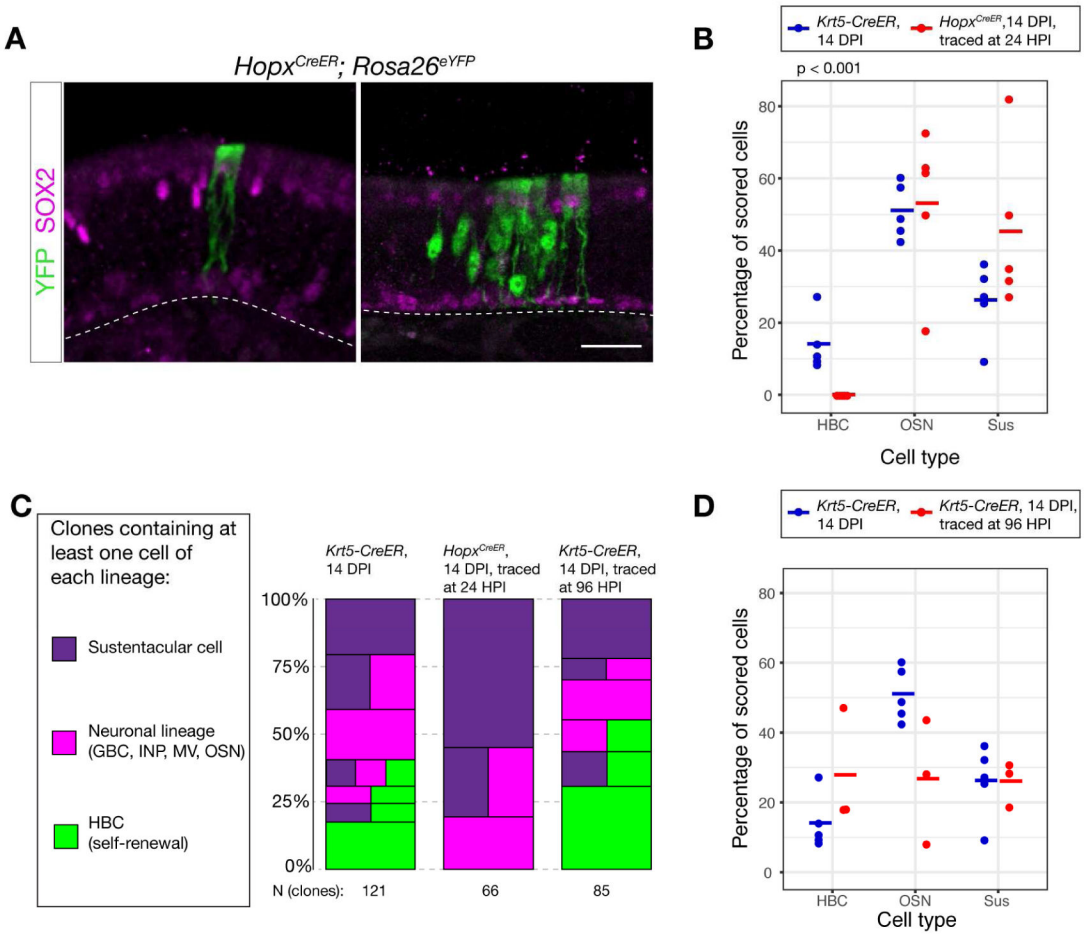


Figure 5. Cell Fate Choice Occurs Early During the Activated, Injury-Response State

(A) Representative maximum projections of 40 μm confocal z-stacks containing lineage-traced clones derived from *Hopx*(+) HBC*1 cells. Lineage tracing was induced in *Hopx*^{CreER}*; Rosa26*^{eYFP} animals at 24 HPI, and tissue was collected at 14 DPI for analysis. SOX2 antibody was used to identify sustentacular cells, HBCs, and GBCs. Scale bar, 25 μm. **(B)** Composition of *Hopx*^{CreER}-derived clones, traced at 24 HPI and scored at 14 DPI (red) compared to clones derived from *Krt5*(+) HBCs traced prior to injury and scored at 14 DPI (blue). **(C)** Ratios of clones scored by lineage representation: *Krt5-CreER; Rosa26*-derived clones traced prior to injury (left), 24 HPI traced *Hopx*^{CreER} derived clones at 14 DPI (middle), and *Krt5-CreER; Rosa26*-derived clones traced at 96 HPI (right). Multi-lineage clones are indicated by different colored rectangles in the horizontal direction, while uni-lineage clones are represented by a single colored rectangle. **(D)** Composition of *Krt5-CreER; Rosa26*^{eYFP}-derived clones, traced at 96 HPI and scored at 14 DPI (red), compared to clones derived from *Krt5*(+) HBCs traced prior to injury and scored at 14 DPI (blue). P-values were calculated using a negative binomial regression model, and the Benjamini-Hochberg method was used to adjust for multiple testing (STAR Methods).

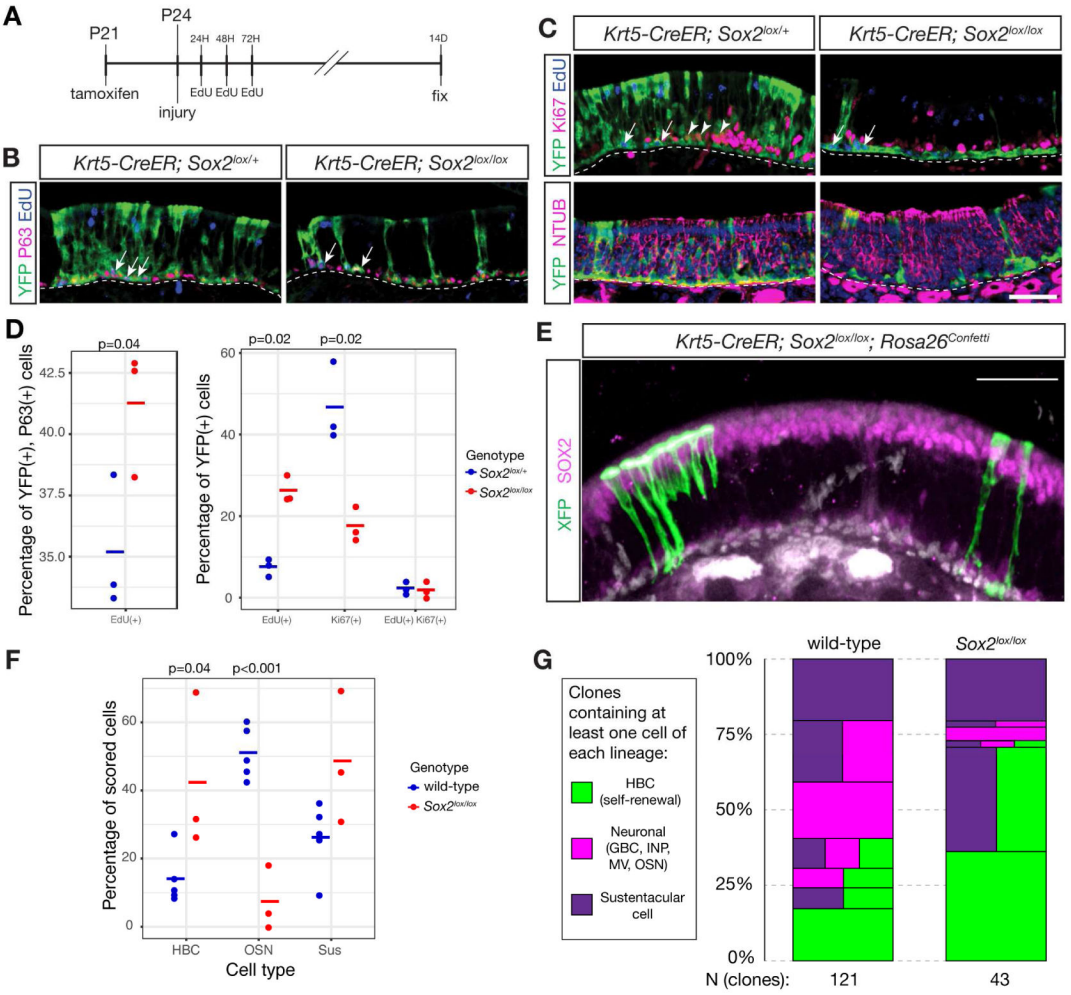


Figure 6. *Sox2* Is Required For Neurogenesis Following Injury to the Olfactory Epithelium

(A) Timeline of experimental treatments described in (B–D). **(B)** Re-formed HBCs (lineage-traced YFP(+), EdU(+), P63(+) cells), in both control and *Sox2* conditional knockout samples, are indicated with arrows. **(C)** Proliferative progenitors (top panels, arrowheads) or neurons (bottom panels) following injury identified in control or *Sox2* conditional knockout tissue. **(D)** In the mutant there is a modest but significant increase in the percentage of lineage-traced self-renewed HBCs (EdU(+), P63(+) cells, p = 0.04; left panel); the percentage of lineage-traced suprabasal cells that retain the EdU label is significantly increased (right panel, p = 0.02) while the percentage of proliferative (Ki67(+)) cells is significantly decreased (p = 0.002; n = 3 for each genotype). **(E)** Representative maximum projection of a 40 μm confocal z-stack containing lineage-traced clones derived from *Sox2* conditional knockout HBCs. Lineage tracing was induced in $Krt5^{CreER}$; $Sox2^{lox/lox}$; $Rosa26^{Confetti}$ HBCs three days prior to injury, and tissue was collected at 14 DPI. Tissue was stained and analyzed as in Figure 1. **(F)** Total cell type ratios in *Sox2* knockout HBC-derived clones (43 clones across three animals) at 14 DPI, compared to wild-type (121 clones, five animals). Self-renewal of HBCs increases in the *Sox2* mutant (p = 0.04), but the production of neurons is significantly reduced (p < 0.001). **(G)** Composition of *Sox2*

knockout HBC-derived clones at 14 DPI, compared to the wild type. P-values were calculated using a negative binomial regression model, and the Benjamini-Hochberg method was used to adjust for multiple testing; see STAR Methods for details. Scale bars, 50 μm. See Figure S6.

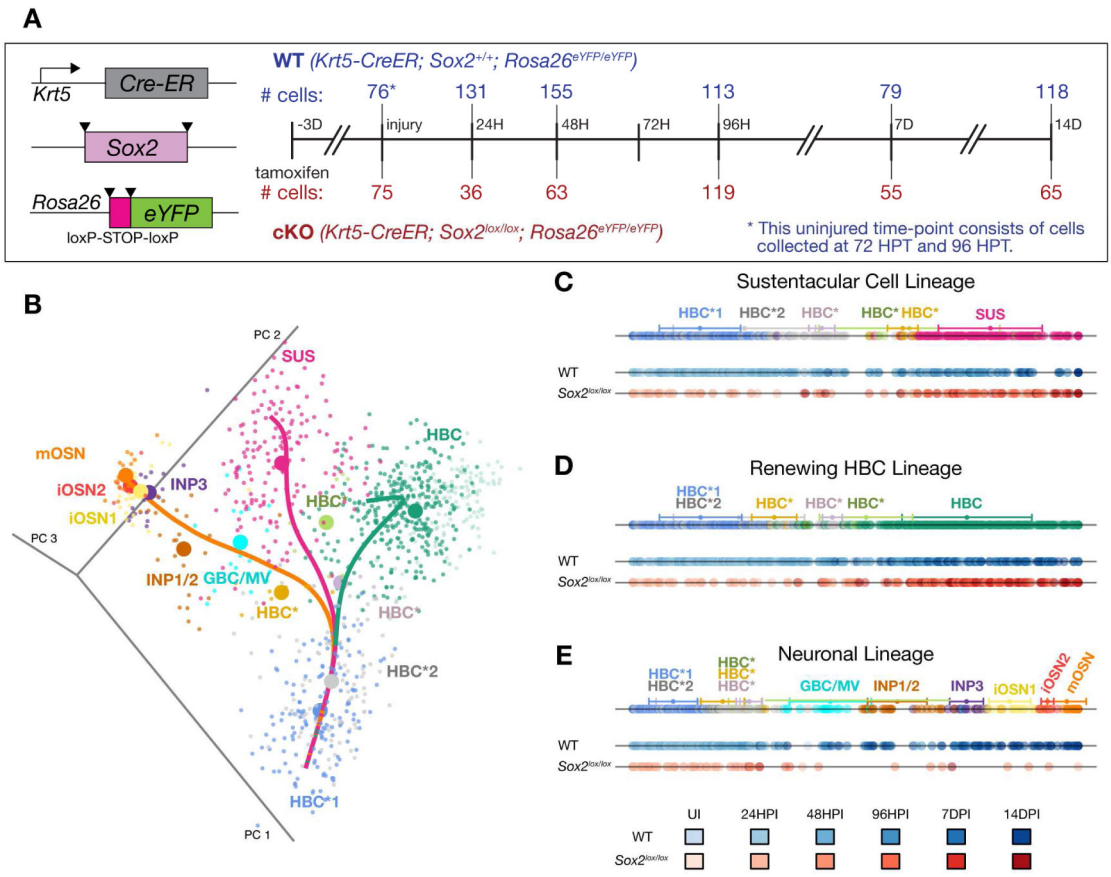


Figure 7. Single-Cell RNA-Sequencing Pinpoints the *Sox2* Conditional Knockout Defect to the Transition from the Activated HBCs to GBCs

(A) YFP(+) cells were collected by FACS at the indicated times following methimazole administration from mice carrying the *Krt5-CreER; Rosa26*^{eYFP} transgenes and either the *Sox2*$^{+/+}$ (WT) or *Sox2*$^{lox/lox}$ (cKO) alleles. **(B)** Three-dimensional representation of single cell gene expression with branching lineage trajectories inferred by Slingshot starting with the HBC*1 cluster (neuronal, orange; sustentacular cell, magenta; renewing HBC, green). **(C–E)** Projection of cells onto their respective lineages reflects developmental order and distance for the sustentacular cell lineage (C), the renewing HBC lineage (D) and the neuronal lineage (E). Cells are colored by experimental time point below each line plot, with darker colors representing later time points.