KLF4 promotes a KRT13+ hillock-like state in squamous lung cancer

Authors

Luke T. Izzo[1*], Tony Reyes[2*], Srijan Meesala[1], Abbie S. Ireland[1], Steven Yang[1], Hari Shankar Sunil[3], Xiao Chun Cheng[1], Nomi Tserentsoodol[1], Sarah B. Hawgood[1], Edward F. Patz, Jr.[1,4], Benjamin L. Witt[5], Darren R. Tyson[1], Kathryn A. O'Donnell[3], Trudy G. Oliver[1,2,6**]

Affiliations

1. Department of Pharmacology and Cancer Biology, Duke University, Durham, NC, 27710, USA

2. Department of Oncological Sciences, University of Utah, Salt Lake City, UT, 84112, USA

3. Department of Molecular Biology, University of Texas Southwestern Medical Center, Dallas, TX, 75390, USA

4. Department of Radiology, Duke University, Durham, NC, 27710, USA

5. Department of Pathology, University of Utah, Salt Lake City, UT, 84112, USA

6. Lead contact: Trudy G. Oliver

*Co-first authorship

Correspondence: Trudy G. Oliver (tgo@duke.edu)

Keywords: squamous lung cancer, cell fate, hillock, KRT13, KLF4

Summary

Lung squamous cell carcinoma (LUSC) is basal-like subtype of lung cancer with limited treatment options. While prior studies have identified tumor-propagating cell states in squamous tumors, the broader landscape of intra-tumoral heterogeneity within LUSC remains poorly understood. Here, we employ Sox2-driven mouse models, organoid cultures, and single-cell transcriptomic analyses to uncover previously unrecognized levels of cell fate diversity within LUSC. Specifically, we identify a KRT13^{+} hillock-like population of slower-dividing tumor cells characterized by immunomodulatory gene expression signatures. The tumor hillock-like state is conserved across multiple animal models and is present in the majority of human LUSCs as well as head and neck and esophageal squamous tumors. Our findings shed light on the cellular origins of lung hillock-like states: normal club cells give rise to tumors with luminal hillock-like populations, while basal-like tumor-propagating cells transition into basal hillock-like states, resembling homeostatic cellular responses to lung injury. Mechanistically, we identify KLF4 as a key transcriptional regulator of the hillock-like state, both necessary and sufficient to induce KRT13 expression. Together, these results provide new molecular insights into cell fate plasticity that underlies intra-tumoral heterogeneity in LUSC, offering potential avenues for new therapeutic strategies.

Introduction

Understanding the cellular complexity of tumors is crucial for improving cancer treatment. Tumors are composed of diverse cell types, including cancer cells, immune cells, and stromal cells, each contributing to disease progression. Among cancer cells, distinct transcriptional states arise not only from genetic alterations but also from non-genetic factors, such as environmental signals and epigenetic modifications[1-3]. Recent advances in single-cell RNA sequencing (scRNA-seq) have transformed our ability to unravel this molecular complexity, revealing previously unappreciated cancer cell states[4-7]. However, the mechanisms driving these transcriptional states in cancer and their impact on therapy resistance remain poorly understood.

Non-small cell lung cancer (NSCLC) represents approximately 85% of all lung cancer cases in the United States[8]. Large-scale genomic sequencing studies have identified genetic alterations commonly found within NSCLC, including lung squamous cell carcinoma (LUSC) and lung adenocarcinoma (LUAD)[9]. While advances in targeted therapies have significantly improved outcomes for LUAD, the lack of actionable targets in LUSC leaves patients with fewer treatment options, underscoring the urgent need for alternative therapeutic strategies for LUSC[10-12]. There is growing evidence that relying solely on cancer genomic alterations for therapy assignment has limitations, which may be due to epigenetic and transcriptional heterogeneity[3,13,14]. Intra-tumoral heterogeneity within cancer cells and the ability of cells to switch between phenotypic states, referred to as plasticity, has recently been highlighted as a hallmark of malignancy and a major barrier to cancer treatment[15-17].

Previous studies investigating intra-tumoral heterogeneity in LUSC identified basal-like tumor-propagating cells (TPCs) that can serially transplant the disease[18-23]. This ability of tumor cell states to mimic normal cell fate lineages has been observed across various cancer types, including lung cancer, olfactory cancer, and neuroblastoma[24-29]. Basal cells, considered a cell of origin for LUSC[30,31], are abundant in the major airways where squamous lung cancer arises. Basal cells play a central role in maintaining the airway epithelium, responding dynamically to injury and environmental insults such as cigarette smoke, a major risk factor for squamous lung cancer[32,33]. Importantly, squamous tumors share key markers and gene expression profiles with basal cells, including the expression of ΔNp63 and KRT5[8], further supporting the link between basal cells and LUSC pathogenesis. However, whether other cell fates arise from basal-like TPCs in LUSC is largely unexplored.

Lineage-tracing studies show that basal cells have the capacity to give rise to multiple differentiated cell types in the adult lung, including tuft, neuroendocrine, club, ciliated, and newly-identified hillock cells[6,7,34]. Hillock cells were first reported in single cell studies that suggested basal cells pass through a hillock state before progressing toward a club fate[6]. Normal hillock cells have been identified in multiple tissues including the lung and prostate[35], and they are marked by KRT13 expression with signatures of immunomodulation and squamous differentiation[6,7]. Recent work stratified hillock cells into two subsets, proliferative hillock-basal cells and protective hillock-luminal cells[36]. Hillock-basal cells have a remarkable ability to repopulate the tracheal epithelium after injury. Hillock-luminal cells are morphologically and transcriptionally similar to differentiated squamous cells, and maintain basal-hillock cell integrity through resistance to a myriad of

injuries[36]. The function of normal hillock cells is only beginning to be understood, and whether they have any transcriptional or functional counterpart in cancer remains completely unexplored to our knowledge. Understanding whether hillock-like cells exist in squamous tumors could provide new insights into the transcriptional plasticity driving LUSC progression.

Multiple genetically-engineered mouse models (GEMMs) of LUSC have been developed[37-42]. While the basal cell is thought to be the most physiologically relevant cell-of-origin for LUSC, multiple cell types have been shown to give rise to squamous lung cancer[30,37]. A key difference between mouse and human lungs is that basal cells are largely confined to the trachea in the mouse, and most GEMMs rely on delivery of virus to more distal regions of the lung where squamous tumors arise via transdifferentiation[39,43-47]. Importantly, mouse squamous tumors that arise via transdifferentiation highly resemble their human tumor counterparts both molecularly and histologically[18,30,39,48-50]. Due to this transcriptional and histopathological resemblance, mouse models provide useful tools for interrogating intra-tumoral heterogeneity and the molecular mechanisms that regulate cell state in squamous tumors.

Kruppel-like factor 4 (KLF4) is a transcription factor that plays an essential role in proliferation, differentiation, and stem cell pluripotency in a context-specific manner[51-53]. In both the skin and esophageal epithelium, KLF4 is restricted to the suprabasal layer and is necessary for differentiation of specific cell types[54-57]. Apart from its role in normal development, KLF4 has been implicated in multiple cancers—able to adopt either an oncogenic or a tumor suppressive role in a context-dependent manner[51]. Despite KLF4's link to normal squamous differentiation, the function of KLF4 in squamous lung tumors and in the hillock cell state has not been explored. Here, by investigating the cellular origins and transcriptional regulation of the KRT13^{+} hillock-like state in LUSC, this study aims to uncover fundamental mechanisms of tumor heterogeneity, paving the way for novel therapeutic strategies targeting transcriptional plasticity in squamous lung cancer.

Results

KRT13 is highly enriched in squamous lung tumors

Recent single cell studies have identified novel lung cell populations[6,7,36], including KRT13^{+} hillock cells. To determine whether a similar state exists in lung cancer models, we queried *Krt13* expression across multiple lung cancer GEMMs representing the three major histological subtypes of lung cancer: adenocarcinoma (LUAD), squamous (LUSC), and small cell lung cancer (SCLC). We collected multiple samples each of adenocarcinoma from the lungs of *Kras*$^{G12D/+}$;*Trp53*$^{fl/fl}$ (KP) mice, squamous tumors from *Sox2*$^{LSL/LSL}$;*Nkx2-1*$^{fl/fl}$;*Lkb1*$^{fl/fl}$ (SNL) or *Lkb1*$^{fl/fl}$;*Pten*$^{fl/fl}$ (LP) mice, and SCLC from *Rb1*$^{fl/fl}$;*Trp53*$^{fl/fl}$;*Rbl2*$^{fl/fl}$ (RPR2) or *Rb1*$^{fl/fl}$;*Trp53*$^{fl/fl}$;*Myc*$^{T58A/T58A}$ (RPM) mice and performed RNA-seq[18,39,48,58-60]. We observed abundant *Krt13* expression restricted specifically to the squamous GEMMs (**Figure 1A**). To test whether this association was consistent at the protein level, we performed immunohistochemistry on tumor-bearing lung tissue from the subtype-representative GEMMs. KRT13 expression was highly enriched in squamous but not in adenocarcinoma or SCLC tumors (**Figures 1B** and **1C**). To determine whether this association was consistent in human squamous tumors, we examined RNA-seq data derived from human lung adenocarcinoma and squamous tumors from The Cancer

Genome Atlas (TCGA). Compared to lung adenocarcinoma, we observed significantly greater *KRT13* expression in human squamous tumors (**Figure 1D**). In addition, we stained tissue microarrays of human lung cancer (n = 57 total samples) for KRT13 and noted significantly increased KRT13 expression in human squamous compared to adenocarcinoma samples, with the majority of squamous patient biopsies displaying KRT13 positive cells (17 of 25) (**Figures 1E** and **1F**). Notably, both mouse and human squamous tumors exhibited intra-tumoral heterogeneity in KRT13 expression (**Figures 1B** and **1E**). Overall, these data suggest KRT13$^+$ cells are highly enriched in subpopulations of mouse and human squamous tumors.

KRT13 is expressed in a subpopulation of squamous tumors

Previous studies have shown functional heterogeneity in LUSC, as a TPC population marked by NGFR$^+$SCA1$^+$ expression has the capacity to serially transplant tumors[18]. Squamous tumors in most GEM models have been shown to arise via transdifferentiation from a mucinous-like state[9,39,43-45]. Due to potential transcriptional and functional heterogeneity, and the heterogeneous expression of KRT13 in squamous tumors (**Figures 1B** and **1E**), we sought to assess *Krt13* expression at the single cell level in squamous GEMMs. We flow-sorted CD45-negative (CD45$^-$) GFP-positive (GFP$^+$) tumor cells from CMV-Cre-infected SNL mice utilizing the GFP reporter in the *Sox2-Ires-Gfp* allele. Following sorting, CD45$^-$ GFP$^+$ tumor cells were subject to scRNA-sequencing (see **Methods**). We obtained 3,663 tumor cells, but very few of these cells expressed canonical basal-associated squamous markers *Krt5*, *Trp63*, *Pitx1* and *Ngfr* that are known to be present in squamous tumors by immunohistochemistry. Rather, the majority of CD45$^-$GFP$^+$ tumor cells expressed markers of mucinous adenocarcinoma. To capture more basal-like squamous tumor cells, we flow-sorted CMV-Cre-infected SNL tumors and isolated CD45$^-$GFP$^+$NGFR$^+$ cells (**Figure S1A**), which are reported to include the TPC population. We obtained 4,363 additional tumor cells that were now enriched for squamous marker genes (**Figures 2A–C**). Consistent with previous studies[39,61], mucinous markers *Spdef, Hnf4a*, and *Gkn1* were low in the tumor cells that expressed basal-associated squamous markers (**Figure 2C**). Next, we analyzed *Krt13* expression. *Krt13* expression was overall low, but present in tumor cell clusters that expressed basal markers while absent in cells expressing mucinous adenocarcinoma markers (**Figures 2D** and **S1B**). Consistently, IHC on serial sections from CMV-Cre-infected SNL tumors showed that KRT13 was enriched in KRT5$^+$ squamous regions and absent in HNF4A$^+$ mucinous regions (**Figure 2E**). Co-staining for KRT13 and basal marker TRP63 showed that KRT13 was highly expressed in cells that were adjacent to, but distinct from, TRP63-expressing cells (**Figure 2F**). Together, these data suggest that KRT13$^+$ cells are enriched in a subpopulation of non-mucinous, non-basal squamous tumor cells.

Club-initiated tumors have early and enriched KRT13 expression

Given that normal *Krt13*$^+$ cells can act as an intermediate state between basal and club cells[6], and club cells can regenerate basal cell populations[62], we reasoned that club-initiated tumor cells may have the capacity to dedifferentiate to a KRT13$^+$ state. Using Cre controlled by a club cell-specific promoter *Ccsp*, we induced tumors from club cells in SNL mice to determine whether this would alter the relative abundance or quality of *Krt13*$^+$ tumor cell populations. Similar to our approach with CMV-Cre-initiated tumors, we first sorted CCSP-Cre-driven CD45$^-$GFP$^+$ tumor cells

(CCSP-GFP$^+$, S1) and found a lack of basal-associated squamous marker gene expression. We subsequently sorted additional CD45$^-$GFP$^+$ (CCSP-GPF$^+$, S2) and CD45$^-$GFP$^+$NGFR$^+$ basal-enriched tumor cells, which were combined for sequencing (see **Methods**). During scRNA-seq analysis, cells from both collections were combined with cells derived from CMV-Cre-infected tumors. Following pre-processing removal of normal lung and immune cell populations and batch correction (see **Methods**), we obtained 4,012 club cell-derived tumor cells that largely clustered away from CMV-initiated populations (**Figures 3A** and **3B**). We confirmed the presence of both mucinous adenocarcinoma and squamous populations in CCSP-derived tumor cells (**Figures 3B** and **3C**), demonstrating that club cells can act as a cell of origin for squamous tumors in the SNL model. *Krt13*$^+$ cells were highly enriched in CCSP-initiated tumor cells compared to CMV-initiated tumors, particularly in cell clusters that lacked mucinous and basal squamous markers (**Figures 3B-3D**). Additionally, we assessed expression of another gene, *Tmprss11b*, that was previously shown to correlate with poor survival in NSCLC (see **Sunil et al, co-submission**)[63]. *Tmprss11b* is enriched in a similar population of cells as those expressing *Krt13* and correlated with *Krt13* expression (**Figures S2A** and **S2B**). Interestingly, in contrast to CMV-driven tumors, KRT13 was present in mucinous adenocarcinoma as well as in squamous histopathologies in CCSP-initiated tumors (**Figures 3E** and **3F**). Given that prior studies have shown that mucinous adenocarcinoma precedes squamous differentiation[39,45,46], these data suggest KRT13 is enriched and emerges earlier in club-initiated tumors compared to CMV-initiated tumors.

KRT13 marks a hillock-like state in squamous tumors

To determine whether *Krt13*$^+$ cells exhibit a hillock-like phenotype in squamous lung cancer, we generated a “hillock-cell signature” by identifying normal hillock-cell specific genes that were conserved in two independent datasets[6,7] (see **Methods** and **Table S1**). We applied the hillock-cell signature to the mouse tumor scRNA-seq data. Tumor cells with a high hillock signature strongly overlapped with *Krt13*-expressing cells and were present in both CMV and CCSP-initiated tumors (**Figures 3D** and **3G**). Additionally, tumor cells with a high hillock score were largely distinct from cells with the highest mucinous and basal cell signatures (**Figure 3G**) (see **Methods, Tables S2** and **S3**), suggesting that *Krt13* marks a transcriptionally distinct hillock-like state in squamous lung cancer, as it does in the normal lung. To query potential similarities between normal lung cell lineages and tumor heterogeneity, we applied additional normal lung cell type signatures to the scRNA-seq data. Consistent with known relationships between goblet cell phenotypes and mucinous adenocarcinomas, these two signatures largely overlap and cluster together by correlation analysis (**Figures 3G, S2C,** and **S2D**). The club cell signature overlaps with both hillock and basal signatures (**Figure S2D**) in line with known lineage relationships between club, hillock, and basal cells[6]. Next, we generated gene sets to distinguish normal basal-hillock and luminal-hillock cells from other normal lung cells using recently published data[36] (see **Methods, Tables S4** and **S5**). Applying these signatures to the mouse tumor scRNA-seq dataset revealed that *Krt13*$^+$ hillock-like tumor cells overlapped with the luminal-hillock score, while the basal-hillock score was more broadly distributed, including cells with *Krt13* and basal marker gene expression such as *Trp63* and *Krt5* (**Figures 3G** and **S2D**). Together, these data suggest that *Krt13* is enriched in hillock-like cells that may mark both luminal- or basal-hillock-like populations, which is influenced by cell of origin.

Luminal-hillock-like cells have reduced proliferation and express immunomodulatory gene signatures

Next, we assigned cell states to Leiden clusters in the mouse tumor scRNA-seq dataset using marker gene expression and gene signature scores of normal lung cell types. Tumor cells were assigned to either basal-, basal-hillock, luminal-hillock, or mucinous/goblet. In addition, a small population of tumor cells with tuft cell gene expression clustered largely separately from these populations (**Figures 4A, 3G,** and **S3A–D,** see **Methods**)[6]. Recent reports have noted tuft-like tumor cells in or near squamous tumors[64-66]. Consistently, our lab recently discovered that basal cells can simultaneously give rise to tuft-like SCLC and squamous tumors under specific genetic conditions, and that combined SCLC in patients can harbor tuft-like SCLC and squamous regions[66,67], suggesting a transcriptional relationship between tuft and basal states in lung cancer. These tuft-like cancer cells were enriched in CMV-driven tumors compared to CCSP-driven tumors (**Figure S3E**).

CCSP-driven tumors contained the majority of luminal-hillock-like cells, whereas CMV-driven tumors contained the majority of basal-hillock-like cells (**Figure S3E**). These data suggest that cell of origin impacts the presence of different hillock-like states in squamous tumors. To uncover transcriptional patterns associated with each cell state in an unbiased manner, we identified top differentially-expressed genes (DEGs) in the assigned luminal-hillock, basal-hillock, or basal-like clusters and performed ENRICHR pathway analysis (**Table S6**). Luminal hillock-like cells were enriched for signatures of immune pathways, neutrophil degranulation, and cytokine signaling (**Figure 4B**). Both basal- and basal-hillock-like clusters were enriched for proliferation-related gene signatures such as translation and metabolic processes (**Figure 4C**), suggesting luminal-hillock-like cells may be less proliferative compared to basal and basal-hillock-like tumor cells. Consistently, KI67 expression was restricted to TRP63^{+}KRT13^{-} cells within squamous tumors by immunostaining, and nearly absent from KRT13^{+} regions (**Figure 4D**). Additionally, cell cycle-specific gene signatures in scRNA-seq data revealed that the luminal-hillock-cell state has a higher proportion of cells in G0/G1 phase compared to basal and basal-hillock states, consistent with decreased proliferation (**Figure S4A**). Altogether these data suggest that the KRT13^{+} luminal-hillock-like state has relatively lower proliferation than basal-like populations.

To determine if a *KRT13*$^{+}$ hillock-like state is also present within human squamous tumors, we analyzed single cell data from 18 patient tumors[68] (**Figures 4E** and **4F**). We excluded normal lung cell types such as immune, endothelial, and fibroblast cells (n = 43,126, see **Methods**), and analyzed tumor cells for expression of canonical markers of adenocarcinoma (*NAPSA*, *NKX2-1, and SFTPB*) and squamous tumors (*KRT5, PITX1* and *TP63*) (**Figures 4G** and **4H**); as expected, squamous markers were considerably higher than the rarely detectable adenocarcinoma markers. Consistent with observations in SNL tumors, high *KRT13* expression in human tumor cells was associated with an enriched hillock score and low basal score (**Figures 4I** and **4J**). While the luminal hillock-score and basal hillock-score largely overlap, the luminal hillock-score more closely mirrors the distribution of *KRT13* expression (**Figures 4I** and **S4B**). Leiden clustering mainly separated cells by patient, likely due to underlying inter-patient genetic and non-genetic heterogeneity, which is expected. However, this limits the ability to call cells as distinct cell types based on Leiden clustering. To avoid these issues and limit effects of gene dropout, we defined

a cutoff-based assignment of individual cells to a hillock-like state using high hillock-score expression and low basal-score expression (see **Methods**). Using this single-cell classification, hillock-like cells were detected in 17 out of 18 human patient samples ranging from ~1% to 20% of the tumor cells (**Figures 4K, 4L** and **Table S7**). Similarly, *KRT13*-expressing cells were detected in 16 out of 18 human patient samples ranging from ~2% to 50% (**Figure S4C** and **Table S7**). Together, regardless of classification methods, hillock-like cells can be detected at the transcriptional level at frequencies similar to what we detect at the protein level in human tissue samples (17 of 25 patient samples) (**Figure 1E**). ENRICHR analysis on DEGs in human hillock-like tumor cells revealed enrichment for immunomodulatory, neutrophil degranulation, and squamous differentiation signatures including keratinization and cornification (**Figure 4M** and **Table S8**)—similar to the patterns observed in mouse *Krt13*$^+$ tumor cells (**Figure 4B**). Finally, cell cycle analysis revealed that human hillock-like tumor cells are relatively less proliferative compared to other cell states in LUSC (**Figure S4D**), consistent with findings in the SNL mouse model (**Figures 4D** and **S4A**). These data suggest that a more lowly-dividing KRT13$^+$ hillock-like state is conserved in both mouse and human squamous lung tumors.

The hillock-like state can arise from NGFR$^+$ tumor cells

To address whether the KRT13$^+$ state can arise from the basal-like TPC population, we harvested tumors from CMV-Cre-infected SNL mice, flow sorted CD45$^-$GFP$^+$NGFR$^+$ tumor cells and plated them in Matrigel to establish tumor organoids (**Figure 5A**). Importantly, CMV-derived SNL tumor organoids recapitulate key structures found in SNL squamous tumors *in vivo*, similar to observations by other groups using lung tumor organoid systems[40,42,69-72]. We observed keratinized structures in H&E-stained organoids as well as proliferative KI67$^+$ outer edges coinciding with the basal TPC marker, NGFR (**Figures 5B** and **5C**). Importantly, we observed a KRT13$^+$ cell population residing internally within the organoids with limited overlap with the ΔNp63$^+$ (also known as P40$^+$) basal layer (**Figure 5D**). Of note, not all organoids had a KRT13$^+$ cell population, but all organoids contained basal-like cells, consistent with the emergence of KRT13$^+$ cells from basal cells. To confirm the ability of TPC-like cells to regenerate KRT13$^+$ populations *in vitro*, we dissociated live tumor organoids and sorted cells into two groups based on surface NGFR expression (NGFR^{high} and NGFR^{low}) (**Figures 5E** and **S5A**). Sorted cells were replated in Matrigel and allowed to form organoid structures over 7–10 days in culture. Compared to NGFR^{low} cells, NGFR^{high} cells generated more abundant and larger organoids (**Figure 5F**). NGFR^{low} cells were capable of forming organoids, suggesting a less basal-like state may possess plasticity to regenerate an NGFR$^+$ basal-like state. In order to determine if the NGFR^{high} and NGFR^{low} cells formed organoids with similar structures and cell type distributions, we stained organoids for KRT13 and TP63. Both sorted populations were capable of regenerating organoids containing mutually-exclusive TP63$^+$ and KRT13$^+$ cells (**Figure 5G**).

To determine the transcriptional heterogeneity of tumor organoids as well as the ability of basal-like cells to give rise to hillock-like cells, we performed scRNA-seq on freshly dissociated NGFR^{high} and NGFR^{low} organoid cells as well as sorted NGFR^{high} cells that were allowed to regrow as organoids for four weeks in culture (NGFR^{high} outgrowth) (**Figures 5E** and **S5A**). Importantly, organoids largely transcriptionally recapitulate the heterogeneity observed in squamous SNL tumors, with distinct populations of *Ngfr/Trp63*-high basal-like cells and *Krt13/Tmprss11b*-high

hillock-like cells (**Figures 5H-J** and **S5B**). Consistently, there was a subpopulation of hillock-score-enriched cells within organoids, while the basal score was highly expressed across the entire transcriptional space in organoids (**Figure 5K**). The abundance of basal-like states in organoid conditions is not surprising, as other studies have shown that the Matrigel organoid culture conditions support basal-like states[67,73]. Basal-hillock and luminal-hillock signatures overlapped with basal and hillock signatures, respectively (**Figure S5C**), similar to patterns in SNL tumor data (**Figure 3G**). Given that mucinous histopathology precedes squamous differentiation in vivo [18,30,39,48-50], we did not expect to observe mucinous-like cells in organoids, and consistently, mucinous markers were extremely rare and low (data not shown). Outgrowth of NGFR^{high} cells repopulated the entirety of transcriptional space occupied by both freshly-sorted NGFR^{high} and NGFR^{low} cells (**Figure 5I**). Additionally, *Krt13* expression and the hillock score were restricted to NGFR^{low} (and not NGFR^{high}) cells at the time of sorting (**Figure S5D**). Outgrowth of NGFR^{high} cells led to reappearance of *Krt13* expression and a hillock-like population (**Figures 5L** and **S5D**), together supporting that hillock-like cells can derive from basal-like TPC cells. Consistent with immunostaining for KI67 (**Figures 5C** and **5D**), hillock-like organoid cells are in a less proliferative state by gene expression assignment to cell cycle phases (**Figure S5E**). Organoid hillock-like cells also upregulate immunomodulatory, neutrophil degranulation, and cornification gene sets (**Figure S5F** and **Table S9**), similar to both mouse and human tumor data.

KLF4 is enriched in a suprabasal hillock-like state in squamous tumors

Next, we sought to identify master transcriptional regulators driving the hillock-like tumor state. We assessed the ENCODE and ChEA transcription factor gene sets in ENRICHR using highly expressed genes across our mouse, human, and organoid single cell datasets (**Figures S6A-C** and **Table S10**). Basal- and basal-hillock-like cells across mouse and human datasets predicted proliferation and known basal-associated transcription factors MYC, MAX, P63 and SOX2 (**Figure S6A**), consistent with their increased proliferation compared to luminal-hillock-like cells. In contrast, KLF4 was a top predicted regulator of both mouse and human hillock-like tumor cells and was also a top predicted regulator of normal lung hillock cells from two independent datasets[6,7] (**Figures S6B** and **S6C**). KLF4 is a promising candidate for regulating the hillock-like state because of its role in multiple squamous epithelial tissues (i.e., esophagus, skin) in promoting cell differentiation away from the basal lamina[54,74]. Moreover, KLF4 can directly regulate *KRT13* expression in esophageal cancer[75] and can repress basal TPC factors SOX2, TP63 and PITX1[23]. To determine KLF4's expression pattern in lung squamous tumors, we queried our RNA-seq data from multiple lung cancer GEMMs and TCGA data, which showed enrichment of *Klf4/KLF4* in both mouse and human squamous tumors compared to adenocarcinomas (**Figures 6A** and **6B**).

To determine KLF4 expression and localization in squamous lung cancer, we co-stained for basal marker ΔNp63/P40 and KLF4 in mouse SNL tumors. KLF4 exhibited a nuclear and suprabasal-like pattern of expression where KLF4 was closely positioned to, but largely non-overlapping with P40^{+} cells (**Figure 6C**). KLF4 was similarly located in a suprabasal pattern when co-staining with the basal TPC-marker NGFR in both CMV- and CCSP-initiated tumors (**Figure 6D**). Co-staining for KLF4 and KRT13 showed that KLF4 was highly expressed in a layer of KRT13^{+} cells (typically in the outermost KRT13^{+} basal-proximal cells), and absent from the innermost KRT13^{+} basal-

distal cells (**Figures 6E** and **6F**). Finally, we stained SNL organoids for KLF4 and noted KLF4 expression enriched in a suprabasal location to NGFR$^+$/P40$^+$ cells, with minimal overlap between KLF4 and basal markers (**Figures 6G** and **6H**); KLF4 was expressed in the majority of KRT13$^+$ cells, suggesting that organoids have a similar organization as squamous tumors. These data suggest multiple levels of intra-tumoral heterogeneity in squamous lung tumors (**Figure 6F**) that mimic normal stratified epithelia, similar to findings in the skin and vaginal epithelium[56,76]. These data suggest KLF4 is positioned to be important for the emergence or differentiation of KRT13$^+$ hillock-like lung tumor cells.

KLF4 is necessary and sufficient to promote the KRT13$^+$ hillock-like state

To determine the functional role of KLF4 in the hillock-like tumor state, we sought to deplete KLF4 expression and determine the impact on KRT13 and the hillock-like state in SNL tumor organoids. Multiple attempts to knockout *Klf4* using traditional CRISPR gene editing were unsuccessful due to organoid death. Thus, we designed a CRISPR-inhibition (CRISPRi) system using a catalytically dead Cas9 enzyme (dCas9) fused to the transcriptional repressor KRAB protein[77]. Compared to non-targeting control sgRNAs (sgNTCi), *Klf4*-targeted CRISPRi constructs (sg*Klf4*i) efficiently decreased KLF4 expression in SNL organoids, especially guide #1 (**Figures 7A, 7B** and **S7A**). Analysis of sgNTCi and sg*Klf4*i organoids by co-IF and immunoblot revealed significantly reduced and more diffuse KRT13 staining upon *Klf4* knockdown (**Figures 7A, 7B** and **S7A**). The CRISPRi system led to variable expression of KLF4 in organoids within the same culture, where KRT13 expression was retained in KLF4-proficient sg*Klf4i* organoids but completely absent in KLF4-deficient sg*Klf4i* organoids (**Figure S7A**). KLF4 suppression did not alter expression of the basal marker KRT5 but increased expression of PITX1 (**Figure 7B**), a transcription factor that cooperates with SOX2 and TP63 in basal skin TPCs, with a mutually antagonistic relationship with KLF4[23]. After seven days of growth following dissociation and plating, KLF4 knockdown with sg*Klf4i-1* slightly but significantly increased organoid formation with no impact on organoid size (**Figure 7C**). Accordingly, sg*Klf4i-1* organoids showed elevated KI67 staining, with KI67$^+$ cells present throughout the organoid when KLF4 was suppressed, whereas KLF4-proficient organoids tended to proliferate specifically in the basal-associated outer layer of cells (**Figures 7D-F**). These data may be consistent with known functions of KLF4 in regulating cell cycle exit[23,53,78-80]. The lack of increased organoid size despite increased proliferative markers may reflect size constraints in culture conditions. Together, these data suggest that KLF4 is necessary for the transition to a less proliferative KRT13$^+$ hillock-like state.

Next, we aimed to determine whether KLF4 is sufficient to promote the KRT13$^+$ hillock-like state. We overexpressed human *KLF4* or *GFP* control in SNL organoids under the control of a doxycycline-inducible tetracycline response element (Tet-ON-TRE-GFP vs KLF4). KLF4 overexpression in SNL tumor organoids increased KRT13 levels and substantially reduced organoid growth (**Figures 7G** and **7H**). After dissociation and re-plating with doxycycline for 7 days, organoids with KLF4 overexpression were smaller and formed at a lower frequency than GFP controls (**Figures 7H** and **S7B**). Together, these data suggest KLF4 is both necessary and sufficient for the emergence of the KRT13$^+$ state.

Given that basal cells are thought to be the cell of origin for squamous lung tumors, that basal-hillock cells give rise to luminal-hillock cells in the normal lung[36], and that basal TPC-like cells differentiate into hillock-like tumor cells, we asked if KLF4 can modulate KRT13 in normal lung basal cells. We evaluated two human basal cell lines, BEAS-2B and BCi-NS1.1, which have been shown to undergo differentiation away from a basal state in cell culture conditions[81-83]. We utilized CRISPR-Cas9 gene editing to knockout *KLF4* in the BCi-NS1.1 cells, which express KLF4 normally. As in tumor organoids, knockout of KLF4 substantially decreased endogenous KRT13 levels especially with guide #3, which showed the most complete DNA editing by surveyor assay (**Figures 7I** and **S7C**). In contrast to tumor organoids where KLF4 loss increased proliferation (**Figure 7F**), KLF4 knockout in normal human basal cells led to reduced proliferation (**Figure S7D**), which may be due to unique constraints in these distinct cell populations, which differ by mouse/human, transformation status, and in vitro growth conditions.

To determine if KLF4 is sufficient to drive a KRT13^{+} state in human basal cells, the TRE-KLF4 overexpression construct was introduced into BEAS-2B cells, which lack expression of both KLF4 and KRT13 under normal culture conditions. Exposure to doxycycline induced KLF4 and led to striking KRT13 induction (**Figure 7J**). Additionally, transient exposure of TRE-KLF4 cells to doxycycline was able to induce KRT13 expression, which was maintained for 48 hours after doxycycline was removed, while KLF4 expression returned to baseline after doxycycline removal (**Figure S7E**). KLF4 overexpression significantly decreased growth of BEAS-2B cells (**Figure 7K**), which is consistent with the slower proliferating KRT13^{+} hillock-like state observed in LUSC tumors and organoids (**Figures 4D, 5C,** and **7C-H**). KLF4 overexpression also led to increased P21 (**Figure 7J**), consistent with previous work suggesting KLF4 regulates cell cycle through P21 in other settings[53,78-80].

Transcriptomic analysis of BEAS-2B cells following KLF4 induction revealed significant upregulation of genes associated with the hillock state (more so than the genes related to the basal state), including *KRT13* and genes associated with *Krt4/Krt13*+ mouse tracheal cells (**Figures 8A** and **8B** and **Table S11**). KLF4 also induced genes related to epidermis development and keratinocyte differentiation (**Figure 8C**). Together, these data suggest that KLF4 is necessary and sufficient to induce KRT13 and a hillock-like state in both normal lung and tumor contexts arising from basal cells. Finally, we generated a gene score from the top-100 upregulated genes ("KLF4-up score") in BEAS-2B cells following KLF4 induction and applied this score to our mouse and human single cell datasets (see **Methods**). Consistently, KLF4-induced genes were highly enriched in the hillock-like populations in all three single cell datasets supporting a role for KLF4 in hillock-like cancer cells (**Figures 8D-F**). Together, these data reveal that KLF4 is necessary and sufficient to induce a KRT13+ hillock-like state in normal basal cells and in lung cancer.

A hillock-like state is present across squamous tumor types

Squamous cell carcinomas (SCCs) can arise in tissues beyond the lung and display similar histologic features[84-86]. Additionally, there are common underlying genetic alterations that promote squamous tumors, especially in the lung, head & neck, and esophageal contexts[84-86]. We therefore queried two previously published squamous cell carcinoma datasets, one containing 12,932 cells from 15 patients with head & neck SCC[87] and one containing 44,547 cells from 60

patients with esophageal SCC[88] (**Figures 8G** and **S8A**), for evidence of the hillock-like state. Both tumor types contained populations of cells that expressed *KRT13*, as well as the basal markers *KRT5* and *TP63*, with minimal overlap between the *KRT13* and basal marker expression (**Figure S8B**). The basal score was broadly expressed across the data while the hillock score was more restricted to clusters that are lower for basal score expression (**Figure S8C**), similar to patterns see in LUSC (**Figure 4J**). Using methods identical to those used for analysis of the human LUSC data, we assigned cell states to both the head & neck and esophageal cells using a hillock score cutoff approach (**Figure 8H** and see **Methods**). Hillock-like cells in both datasets showed evidence of decreased proliferation by cell cycle phase assignment, and differential gene expression analysis of upregulated genes in the hillock-like state using ENRICHR showed remarkable similarity to LUSC, including immune and squamous related gene signatures (**Figure S8D-E**). Finally, we applied the KLF4-up score to these data which revealed enrichment in hillock-like cells in both head & neck and esophageal SCC (**Figure 8I**). Together, these findings highlight that a KLF4-driven hillock-like state exists across multiple squamous tumor types and suggest that understanding this cell state will have implications beyond lung squamous tumors.

Discussion

Here, we integrate data from mouse and human LUSC to uncover a KRT13$^+$ tumor cell state that transcriptionally resembles the recently identified normal luminal- and basal-hillock cell populations in the lung[6,7,36]. Normal hillock cells are an intermediate cell fate between lung basal and club cells, express squamous differentiation markers, and are predicted to have immunomodulatory capacity[6]. Additionally, normal hillock cells can be broken down into two distinct subgroups, one with rapid proliferation that exhibits basal cell characteristics and has the capacity to repopulate the airway epithelium following injury, and a more luminal population that has transcriptional similarities to squamous differentiated epithelium[36]. Consistent with these phenotypes, we observe KRT13$^+$ cells specifically in lung cancers with basal identity and squamous differentiation. Moreover, KRT13$^+$ luminal-hillock-like cells in mouse and human LUSC express signatures of immunomodulation and squamous differentiation while basal-hillock-like tumor cells are more proliferative. Importantly, our data suggest KRT13$^+$ luminal-hillock-like tumor cells act as a relatively slower-dividing subpopulation of squamous tumor cells.

The KRT13$^+$ luminal hillock-like LUSC tumor cells that we identified lack proliferative markers but reside near highly proliferative basal-like TPCs in squamous tumors. We show that isolated TPCs from mouse squamous tumors have the capacity to generate tumor organoids that contain KRT13$^+$ cells *in vitro*, suggesting TPCs give rise to hillock-like tumor cells, reminiscent of how proliferative basal-hillock cells give rise to luminal-hillock cells in the adult lung[6,36]. Additionally, our data suggest that hillock-like cells can de-differentiate into more basal-like cells, at least in the context of tumor organoids. The plasticity between these states within the tumor warrants future investigation as therapies targeting either TPC-like or hillock-like cells may be hindered by cell fate plasticity.

KRT13$^+$ cells express gene signatures related to immunomodulation, which may partly explain why slower-dividing cell states that should be outcompeted are able to survive within squamous tumors, and could ultimately benefit tumor fitness. The LUSC cell states of basal-like, basal hillock-like, and luminal hillock-like, share transcriptional similarities to basal, club, and differentiated hillock cells of the normal lung. Overall, these findings suggest that conserved developmental processes are involved in both normal lung homeostasis and LUSC development.

Club cell-initiated tumors cluster largely distinctly from CMV-induced tumors in single cell analysis, indicating unique transcriptional profiles in squamous tumors depending on cell of origin. Additionally, club cell-initiated tumors harbor KRT13$^+$ cells in both the mucinous and squamous compartments, whereas KRT13$^+$ positivity was restricted to squamous histopathology in the CMV-initiated context. These data suggest a relationship between cell of origin and the hillock-like state, whereby club cell transformation to luminal hillock-like tumor cells may be more permissive than in other cell types that are targeted by CMV-Cre, potentially due to underlying epigenetic states. It will be important for the field to gain a deeper understanding of the relationships between cell of origin and cell states within squamous tumors. The SNL and other squamous GEMMs[37-42] may serve as valuable tools for understanding the origin and function of KRT13$^+$ hillock-like tumor cells.

Our data suggest KLF4 is an important regulator of the hillock-like state. Interestingly, KLF4 can regulate hillock-enriched genes including *KRT13* in different contexts[75,89,90], suggesting these mechanisms may be well conserved. Additionally, KLF4 has also been shown to suppress the cell cycle and drive differentiation[91,92]. The organoid and mouse systems described here may be powerful tools to uncover additional levels of regulation of the hillock cancer state, including the role of transcription factors such as NFE2L2/NRF2, which is another predicted hillock regulator in our analyses (**Figures S6A-C**). NFR2 signaling is altered at the genomic level in squamous lung tumors and can promote therapeutic resistance, making it an additional target of interest related to the hillock-like state[9,40,93-96].

It is established that the immune microenvironment of lung squamous tumors is enriched for neutrophils that are immune-suppressive[39,97]. Studies in both LUSC patients and GEMMs have shown that tumor-associated neutrophils (TANs) can suppress T-cell activation and contribute to an immunosuppressive tumor microenvironment[97-100]. The mechanisms that promote tumor-suppressive states in LUSC remain understudied. Our data suggest that hillock-like cells are enriched for expression of immunomodulatory genes which suggests that distinct cell states within LUSC tumors may play differential roles in regulating the tumor microenvironment. In a companion manuscript, Sunil et al show that in the SNL model, the tumor microenvironment is shaped by the expression of the serine protease *Tmprss11b* which is expressed in the *Krt13*-expressing hillock-like cells and is regulated by KLF4 (**Sunil et al 2025, co-submitted**). TMPRSS11B drives lactate export by promoting MCT4 activity and glycolytic metabolism, which correlates with low-pH regions and increased innate immune infiltration in squamous tumors, providing one example of how the hillock-like state may alter the tumor immune microenvironment (**Sunil et al 2025, co-submitted**)[63]. Together, these data support future investigation into the role of hillock-like cells in squamous tumor immune regulation and their impact on response to immunotherapy.

Squamous cell carcinomas of the upper-aerodigestive tract(lung, esophagus, and head and neck) as well as anogenital tract and skin share remarkable histologic and genetic similarities[84-86]. Our data supporting multiple tumor cell states within LUSC led to the identification of similar KRT13$^+$ hillock-like states within squamous tumors of the esophagus and head and neck. Additionally, work in both healthy and cancerous prostate tissue has identified populations of KRT13$^+$ cells, which have been correlated with poor outcomes in the cancer context[35,101-103]. Prevention of the formation of KRT13$^+$ cells may hinder LUSC progression, and in combination with therapies targeting TPCs, may be an effective strategy to combat tumor plasticity and improve patient outcomes across squamous malignancies.

Altogether, our work highlights the importance of using single cell technologies to uncover heterogeneity in difficult-to-treat tumor types. The transcriptional and cell state heterogeneity within a tumor has the potential to limit the efficacy of standard of care therapies, including cytotoxic chemotherapy that may be more likely to hit rapidly dividing populations like TPCs, or immunotherapy by supporting an immunosuppressive microenvironment. We show that multiple tumor cell fates exist within squamous tumors and provide insight into how KRT13$^+$ hillock-like cells may arise and be regulated through KLF4, hence opening potential avenues for new treatment strategies in LUSC.

Limitations of the Study

Our study provides evidence for non-genetic factors that support intra-tumoral heterogeneity in LUSC and suggest that *Krt13*$^+$ hillock-like cells, regulated by KLF4, may support tumor progression, but we acknowledge limitations to our systems. First, our mouse models and organoids are predominantly derived from the SNL model. *SOX2* amplification is one of the most common genetic events in squamous tumors and SNL squamous tumors are known to recapitulate many of the features of human LUSC, however this model harbors mucinous adenocarcinomas and squamous tumors are likely derived from transdifferentiation from an adenocarcinoma state, which may not reflect the origins of many human tumors. Additionally, reliance on one GEMM limits the ability to assess the impact of genetics on the hillock-like state. We show that the LP model and human tumors also contain KRT13-expressing cells and show similarities transcriptionally between human squamous tumors and the SNL model, but future work should incorporate additional genetic models. Furthermore, using CMV-Cre in the SNL model limits understanding of how cell of origin influences the hillock-like state. Our findings suggest that tumor initiation impacts intra-tumoral heterogeneity, as tumors initiated with a club cell-specific Cre driver show increased KRT13 positivity. Further characterization using diverse Cre drivers will clarify these developmental relationships.

AUTHOR CONTRIBUTIONS

Conceptualization: K.A. O'Donnell, T.G. Oliver

Data Acquisition: L.T. Izzo, T. Reyes, S. Meesala, S. Yang, H.S. Sunil, A.S. Ireland, X.C. Cheng, N. Tserentsoodol, S.B. Hawgood
Data Analysis and Interpretation: L.T. Izzo, T. Reyes, H.S. Sunil, A.S. Ireland, D.R. Tyson, T.G. Oliver
Writing – Original Draft: T. Reyes, L.T. Izzo, T.G. Oliver
Writing – Review and Editing: L.T. Izzo, S. Meesala, A.S. Ireland, T. Reyes, D.R. Tyson, K.A. O'Donnell, T.G. Oliver
Funding Acquisition: K.A. O'Donnell, T.G. Oliver
Resources: E.F. Patz, B.J. Witt
Supervision: K.A. O'Donnell, T.G. Oliver

ACKNOWLEDGEMENTS

We thank members of the Oliver Lab for technical assistance and mouse colony management (K.S. Lucas, D. Soltero, K. Spainhower, R. Uzachenko), and administrative support (X.C. Cheng, L. Houston, L. Youngblood). We appreciate MedGenome staff for sequencing services, and Duke Human Vaccine Institute for use of the tapestation. We acknowledge support from the NIH National Cancer Institute (NCI) under award U24CA213274 (TGO) and R01-CA244841-05 (TGO). T.G.O. received support as a Duke Science & Technology Scholar. We thank the BioRepository & Precision Pathology Center (BRPC), a shared resource of the Duke University School of Medicine and Duke Cancer Institute. The BRPC receives support from the P30 Cancer Center Support Grant (P30 CA014236) and from the Cooperative Human Tissue Network (UM1 CA239755).

DECLARATION OF INTERESTS

TGO has a patent related to SCLC subtypes and a sponsored research agreement with Auron Therapeutics, serves on the scientific advisory board (SAB) for Lung Cancer Research Foundation, and as a consulting editor for *Cancer Research* and *Genes & Development*. All other authors declare no conflicts of interest related to this work.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

GEMMs

Rb1$^{fl/fl}$*;Trp53*$^{fl/fl}$*;Rbl2*$^{fl/fl}$ (RPR2)[60], *Rb1*$^{fl/fl}$*;Trp53*$^{fl/fl}$*;LSL-Myc*^{T58A} (RPM)[59], *Sox2*$^{LSL/LSL}$*;Nkx2-1*$^{fl/fl}$*;Lkb1*$^{fl/fl}$ (SNL)[39], *Lkb1*$^{fl/fl}$*;Pten*$^{fl/fl}$ (LP)[18], and *Kras*$^{G12D/+}$*;Trp53*$^{fl/fl}$ (KP)[58] mice have been previously described. Methods for tumor initiation and monitoring can be found in the listed references. Ad5-Cre with the following promoters was used: RPM (10 CGRP-Cre initiated tumors, 5 CMV-Cre imitated tumors), RPR2 (4 CGRP-Cre initiated tumors, 4 CMV-Cre initiated tumors), KP (CMV-Cre initiated tumors), LP (CMV-Cre initiated tumors), SNL (CMV-Cre and CCSP-Cre initiated tumors). All mice were housed following the regulations set by the Institute Animal Care and Use Committee of the University of Utah and Duke University. At 6-10 weeks of age, anesthetized SNL mice for this study were infected by intranasal instillation with $1x10^8$ plaque forming units of Ad5-CMV-Cre or Ad5-CCSP-Cre adenovirus (University of Iowa). Viruses were

administered in a Biosafety Level 2+ room according to the guidelines from the Institutional Biosafety Committee.

Human Samples

Human squamous tissue microarray (TMA) samples were a gift from Edward Patz, Jr. TMA samples were de-identified and only information on the histopathological subtype scoring by a trained pathologist was provided. Human tissue microarrays from patients with lung cancer were generated following written informed consent from each patient under an Institutional Review Board-approved protocol at Duke University Medical School (Protocol ID: Pro00012914).

Cell Lines

The human basal cell line BEAS-2B was acquired from the Duke Cell Culture Facility (Sigma Catalog #95102433). The BCi-NS1.1 cells, an immortalized human airway basal cell stem/progenitor cell line, was generated at Cornell University by Dr. Ronald Crystal and provided as a gift for research purposes[82]. Both cell lines were grown in 2D adherent cell plates in Bronchial Epithelial Cell Growth Basal Medium (Lonza Catalog # CC- 3171) with Supplements and Growth Factors (Lonza Catalog # CC-4175) along with 0.2% Penstrep (BEBM media). For serial passaging, media was removed and 0.25% Trypsin-EDTA (Gibco catalog #25200056) was added to dislodge cells. HEPES Buffered Saline Solution (Catalog # CC-5024) supplemented with 15% FBS was added to quench the reaction. The cells were centrifuged at 250g for 5 min at 4°C. The supernatant was removed, and cells were resuspended in BEBM media before being plated into tissue culture treated dishes.

Organoids

After sorting from tumors (see **Harvesting tissue for organoid production or single cell sequencing**), single cell suspensions containing SNL tumor cells ($NGFR^{+}GFP^{+}$) were centrifuged at 2000 rpm for 5 min at 4°C. Supernatant was removed and cells were resuspended in Matrigel before plating. Cells were grown in 3D Matrigel domes submerged in organoid media (OCM) [50% DMEM/F12 media (supplemented with 10% fetal bovine serum, 1% penicillin/streptomycin, 1% L-glutamine) + 50% L-WRN conditioned media (containing Wnt-3A, R-spondin 3 and Noggin) + 10uM ROCK inhibitor]. L-WRN conditioned media was generated according to previously published methods[104] using L-WRN cells purchased from ATCC (CRL-3276). For serial passaging or flow sorting, media was removed and cold DMEM media was added to wells before disruption of the Matrigel plug by pipetting. The organoid and Matrigel containing media was centrifuged at 1500 rpm for 5 min at 4°C. Supernatant was removed and organoids were resuspended in PBS before centrifugation at 1500 rpm for 5 min at 4°C. Supernatant was removed and cells were resuspended in 500 uL TrypLe for 15 min at 37°C. Following vigorous pipetting, the reaction was quenched with DMEM and was centrifuged at 1500 rpm for 5 min at 4°C. Cells were then resuspended in Matrigel and plated in 24 well plates. For sorting experiments, single cell suspensions were sorted, centrifuged at 1500 rpm for 5 min at 4°C, then resuspended in Matrigel and plated in 24 well plates.

METHOD DETAILS

Harvesting tissue for organoid production or single cell sequencing

Mice were euthanized to harvest tumor-bearing lung tissue or tracheas. For tumor cell isolation, tumor-bearing lung tissue was removed and put into cold DMEM media (10% FBS, 1% L-glut, 1% Pen/Strep) and chopped into smaller pieces mechanically with fine scissors before further dissociation was performed using the Lung Dissociation kit (Miltenyi Biotech Cat # 130-095-927) at 37°C following the manufacturer's instructions. Dissociated cells were filtered using a 100 µm strainer and collected into conical tubes. Cells were pelleted by centrifugation at 1500 rpm at 4°C for 5 min and then resuspended in ACK lysis buffer to lyse red blood cells. The lysis reaction was quenched with addition of PBS and cells were centrifuged. Supernatant was removed and cells from tumor-bearing lungs were resuspended in FACS buffer (PBS containing 2% FBS) for subsequent flow sorting and single cell sequencing.

Flow sorting tumor cells

Single cell suspensions from tumor-bearing lungs or tumor organoids (see **Harvesting tissue for organoid production or single cell sequencing** and **Organoids**) were stained with a rabbit anti-NGFR (Abcam Cat #ab8875 or ATS Cat # AB- N01AP) primary antibody for 30 min on ice. Cells were then washed with FACS buffer and stained in FACS buffer containing a fluorescent secondary antibody (APC-goat anti-rabbit Invitrogen Cat # A-10931) and a fluorescent conjugated antibody (BV605-rat anti-CD45 Biolegend Cat # 103155, for tumors). Cells were then incubated on ice and in the dark for 20-30 min. Following incubation, cells were washed with FACS buffer and centrifuged. The supernatant was discarded, and cells were resuspended in FACS buffer containing DAPI to detect dead cells. Cells were then sorted using flow cytometry according to the outlined gating scheme on a cell sorter (either BD Bioscience FACSAria or Beckman Coulter MoFlo Astrios). Flow sorted tumor cells were used for single cell RNA-sequencing or for plating of organoids within 30 minutes directly following sorting (see **Single cell RNA-sequencing** or **Organoids** below). For SNL CMV-Cre tumor cells used for sequencing, GFP+ and GFP+/NGFR+ cells were sequenced in two independent experiments. For SNL CCSP-Cre tumor cells used for sequencing, GFP+ cells were used for the first sequencing (S1) and a mixture of GFP+, GFP+/NGFR+, and CD45+ cells were used for the second sequencing (S2).

Ultra-Comp eBeads Plus compensation beads (Invitrogen Cat # 01-3333-41) were used to set compensation for the NGFR primary antibody.

Single cell RNA-sequencing

Single-cell suspensions from flow sorted GFP^+ and $NGFR^+$ GFP^+ SNL tumors or SNL tumor organoids were prepared for sequencing according to the 10x Chromium platform protocols. (https://support.10xgenomics.com/single-cell-gene-expression). The SNL CMV- and CCSP-Cre initiated tumor samples sorted for GFP^+, and the SNL CMV-Cre GFP^+/$NGFR^+$ tumor cells were prepared and sequenced at the Huntsman Cancer Institute High-Throughput Genomics Core at the University of Utah. The SNL CCSP-Cre GFP^+/$NGFR^+$ tumor sample and all organoid samples were prepared in-house at Duke University and sequenced by MedGenome. Libraries were prepared with the Chromium 3' single cell gene expression (GEX) kit V2 chemistry for CMV-GFP^+ (10x Genomics Cat # PN-120237) and V3 chemistry for CMV-$NGFR^+GFP^+$, CCSP-GFP^+, CCSP-$NGFR^+GFP^+$, and all organoid samples (10x Genomics Cat # PN-1000141). In brief, individual cells were barcoded with 16-bp 10x barcodes and cell-specific transcript molecules were tagged

with 10bp unique molecular identifiers (UMIs), according to the manufacturer's instructions. Reverse transcription reagents and 10x gel beads were loaded on Chromium Chip Gs (10x Genomics Cat # PN-1000136) and into the Chromium single-cell controller (10x Genomics Cat # PN-120263) to form GEMs. Within singular GEMs, cDNA generated from captured and barcoded mRNA was synthesized by reverse transcription following the manufacturer's recommendations. Subsequent A tailing, end repair, adaptor ligation, and sample indexing were performed for each sample according to the manufacturer's instructions. The resulting cDNA libraries were assessed on Agilent D1000 or D5000 ScreenTape on an Agilent Technology 2200 TapeStation system and quantified by quantitative PCR using a KAPA Biosystems library quantification kit (Roche Cat # 070960140001). Individual libraries were normalized and sequenced on an Illumina NovaSeq 6000, with the exception of the CMV-GFP^{+} sample that was sequenced on an Illumina HiSeq, with 10x-recommended read lengths and aiming for >20k reads per cell.

10x single Cell RNA-seq data processing

Demultiplexing and data alignment

CellRanger was used to demultiplex and align reads per sample for downstream analysis. Single cell RNA-seq data from SNL samples were first demultiplexed with 10x cellranger mkfastq to create fastq files. For each sample, reads were then aligned to the GRCm38 (mm10) mouse genome. For SNL tumor organoids, reads were aligned to a custom mouse genome containing sequences for *fLuc*, *Venus, Cas9,* and e*GFP* (GRCm38-mm10-2020-A)[67]. Count barcodes and UMIs were generated using cellranger count.

Quality control, data combination, and normalization

All clustering, cell type identification and quality control (QC) was performed in Python (v3.8.8) utilizing Scanpy (v1.10.0) according to current expert recommendations for single cell best practices[105]. All cells sequenced from CMV– or CCSP-Cre initiated tumors or tumor organoids were read into sparse matrix anndata objects created from filtered feature matrices with scanpy.read_ 10X_mtx(), and quality metrics were calculated using scanpy.pp.calculate_qc_metrics(). Low quality cells and potential doublets were excluded by selecting for cells with ≤15% mitochondrial content and the following criteria for respective sample types: SNL primary tumor samples with cells containing >1000 total counts and between 1000 and 6000 genes with at least one read; SNL tumor organoid samples with cells containing >1000 total counts and between 2000 and 8000 genes with at least one read. Human patient data was accessed from Wu et al[68], patients identified as squamous were used, and no additional QC filtering was performed. Head and neck patient data was accessed from Kurten et al[87], utilizing only CD45^{-} samples and were filtered for cells containing >500 total counts and between 500 and 4000 genes with at least one read. Esophageal patient data was accessed from Zhang et al[88], utilizing only CD45^{-} samples annotated as epithelial, and no additional QC filtering was performed. Normalized counts were calculated with scanpy.pp.normalize_total() and a target sum of 10,000. Integrated anndata objects containing multiple scRNA-seq datasets were combined using adata.concatenate() with join='outer'.

Batch correction and clustering

Data were further processed using Scanpy (v1.10.0) and scVI-tools (v0.17.4) and benchmarking standards to minimize batch effects while maintaining true biological variability, particularly across integrated objects[106]. Specifically, we utilized scvi.model.SCVI.setup_anndata() to initialize the integration and clustering of datasets from concatenated anndata objects containing all samples and genes, continuous covariate keys set as percent mitochondrial counts, and batch keys

identifying samples prepared or sequenced at different times. The model was trained with default parameters, an early stopping patience of 20, and a maximum of 400 epochs, using the model.train() function on the top 10,000 variable genes determined using the sc.pp.highly_variable_genes() function. The latent representation of the model was obtained with model.get_latent_representation() and added to the .obsm of each anndata object. Neighbors were then calculated with sc.pp.neighbors() with use_rep set to the .obsm category added from the latent representation. UMAP embedding was performed using sc.tl.umap() with min_dist=0.5. Finally, Leiden clusters were generated with sc.tl.leiden() with resolution set to 1.0 for initial steps. As is required throughout the scVI pipeline, we utilized raw counts for all steps above. Additional rounds of quality control and data filtering were performed per dataset by assessing n_genes_by_counts, total counts, and percent mitochondrial counts per cluster. In general, the model tends to cluster low quality and doublet cells together, so clusters with exceptionally high or low average genes_by_counts, total counts, or mitochondrial content were labeled low-quality and considered for removal from the dataset. Removal was performed after assessing gene expression based on known markers of tumor and normal lung cells, and marker genes per cluster as determined by sc.tl_rank_genes_groups(), to help ensure that biological cells that normally have higher or lower n_genes_by_counts, total_counts, or percent mitochondrial content were not aberrantly filtered out. Additionally, from SNL primary tumor samples and human patient data, we identified and removed non-tumor cell types including immune, endothelial, fibroblast, and normal lung cell types using a panel of common marker genes. For the human patient data, adenocarcinoma clusters were also removed based on expression of genes known to be high in lung adenocarcinomas and not expressed in squamous tumors. Each time a cluster was removed, we ran the scVI pipeline on the new anndata object iteratively through this quality control step until there were no longer any low-quality cell clusters or non-tumor cells in the anndata object. Additionally, each time clusters were removed from a larger anndata object, the pipeline was re-run for optimal clustering. Final Leiden clusters were determined with sc.tl.leiden() resolution set to 1 for the human LUSC patient data (**Figure 4**), 0.75 for SNL tumor organoids (**Figure 5**), 1 for SNL CMV and CCSP cells combined (**Figure 3**), 0.5 for SNL CMV tumor cells (**Figure 2**), 0.5 for both head & neck and esophageal patient data (**Figure 8**). After final leiden clustering, the embeddings determined on highly variable genes were added to anndata files that contained all genes with raw read counts. Counts were then normalized and log transformed using sc.pp.normalize_total() with a target_sum=1e4, followed by sc.pp.log1p() using default settings. UMAPs of individual genes were produced using normalized counts, and all gene scores and DEG analysis was performed on log transformed counts. Source code to reproduce analyses methods is on GitHub (https://github.com/TOliver-Lab/Izzo_Hillock_2025) and publicly accessible upon publication.

Calculation of basal, mucinous adenocarcinoma, tuft, and hillock cell scores

The general hillock-cell score was derived by identifying genes enriched in the *Krt13*$^{+}$ cells from Montoro et al, Nature, 2018[6] and Plasschaert et al, Nature, 2018[7] to create a 65-gene “conserved hillock signature” (genes with conserved significant enrichment versus other normal lung cell types in both independent datasets). The “basal score”, “goblet score”, “club score”, “ciliated score” and “tuft score” are derived from droplet-based sequencing done from isolated epithelial cells from mouse trachea[6]. The top 100 differentially-expressed genes from a ranked list of differentially expressed genes (n = 1,137) from mouse mucinous adenocarcinoma tumors were used to generate the mucinous adenocarcinoma score[61]. The “basal-hillock score” (n=170) was generated from differentially-expressed genes comparing hillock-basal cells to normal pseudostratified lung epithelium (Log2FC>0.5) and the “luminal-hillock score” (n=335) was generated from differentially-expressed genes comparing luminal-hillock cells to normal pseudostratified lung epithelium (Log2FC>0.5)[36] (see **Tables S1-5**). Mouse gene scores or

human gene scores were converted to their respective species orthologs using Gprofiler[107] in Python. Hillock, basal, mucinous adenocarcinoma, basal hillock, luminal hillock, goblet, club, ciliated, and tuft scores were computed for each cell per scRNA-seq library by converting expression data to a AnnData object and utilizing the scanpy.tl.score_genes function in Scanpy on normalized and log transformed data using sc.pp.normalize_total() with a target_sum=1e4, followed by sc.pp.log1p() using default settings.

Calculation of KLF4-up score

The "KLF4-up score" (n=100) was generated using the top 100 upregulated differentially-expressed genes comparing doxycycline treated to no doxycycline control of *TRE-KLF4*-expressing BEAS-2B cells (see **Bulk RNA-seq sample and data processing from BEAS-2B** and **Table S11**). The KLF4-up score was computed for each cell per scRNA-seq library by converting expression data to a AnnData object and utilizing the scanpy.tl.score_genes function in Scanpy.

Calculation of cell cycle phase

Cell cycle analysis was performed using the scanpy.tl.scores_cell_cycle function. S-phase and G2/M-phase gene sets were derived from cc.genes in Seurat[108] and were converted to mouse homologs. Each cell was assigned a cell cycle phase based on gene score expression using the scanpy.tl.scores_cell_cycle function.

Cell state assignment

Cell states for SNL primary tumors were determined by evaluating marker gene expression and gene signature scores in Leiden clusters. Cell state assignments were validated by gene expression per cell state and by evaluation of DEGs from each cell state compared to all other cells (see **Differential gene expression analyses**). For the human patient squamous lung cancer data, cell state was assigned by determining a "hillock-like" cutoff. The z-score for the hillock score and basal score were calculated and cells were determined to be "hillock-like" if the hillock score z-score was >=2 and the basal score z-score was <=3, which excluded highly basal-like cells while maintaining cells with high hillock gene expression compared to the overall cell population.

Differential gene expression analyses

Differentially-expressed genes (DEGs) for all samples was determined using the scanpy.tl.rank_genes_groups function in Scanpy using the Wilcoxon Rank Sum Test on log transformed data. This test compares each group to all other groups to determine top DEGs in a specific cell population. ENRICHR analyses[109-111] were performed on the top 500 upregulated genes (DEGs) in the SNL mouse tumor cell types (basal-like vs all others, basal hillock-like vs all others, and luminal hillock-like vs all others), 500 genes in the human tumor cell type (basal-like vs hillock-like), and 500 genes in organoids (basal-like vs hillock-like). For ENRICHR analysis of normal hillock genes from Montoro et al, Nature, 2018, DEGs for hillock cells were included if the log2FC>=0.5 and all *Krt13/Krt4*-enriched genes from Plasschaert et al, Nature, 2018, were included.

KLF4 modulation in organoids

For lentiviral CRISPR-mediated inhibition, the following guideRNAs targeting *Klf4* were cloned into the dCas9 KRAB vector[77] (Addgene # *71236*) (dCas9 KRAB): sg*Klf4i–1 (5' CGCGCCGCCGACAGCCACAG 3'), sgKlf4i–2 (5' AGTGTCCCCCACCGTTGTCG 3'), and*

sgKlf4i–3 (5' AGTGGCCGCGACAACGGTGG3'), sgNTCi (5' ACCTGATACGTCGTCGCGTA3'). pLV hU6-sgRNA hUbC-dCas9-KRAB-T2a-Puro was a gift from Charles Gersbach (Addgene plasmid # 71236; RRID:Addgene_71236).

Human *KLF4* was overexpressed in tumor organoids under the control of a tetracycline response element (TRE). *KLF4* overexpression was induced in organoids using 1 µg/mL of doxycycline. Sequencing was validated and confirmed through Plasmidsaurus. For knockdown and overexpression, sequence-validated plasmids were used to generate low titer or high-titer lentivirus. 293T cells were transfected with a three-plasmid transfection system including the lentiviral vector, pCMV-dR8.2 dvpr, and pCMV-VSV-G for production of low titer virus or high titer virus.

For infection with high titer lentiviruses, organoid cells were dissociated as described for serial passaging and resuspended in 0.5mL of organoid culture media (OCM) [50% DMEM/F12 media (supplemented with 10% fetal bovine serum, 1% penicillin/streptomycin, 1% L-glutamine) + 50% L-WRN conditioned media (containing Wnt-3A, R-spondin 3 and Noggin) + 10uM ROCK inhibitor] containing polybrene (10 µg/mL) and high-titer lentivirus. Cells were then centrifuged at 300g for 30 min at room temperature within a 24 well plate. Organoid cells were then resuspended in the virus containing media and centrifuged at 2000 rpm for 5 min at 4°C. Virus containing media was removed and stored before organoid cells were resuspended in Matrigel and plated. The virus containing media was then added back to the organoid cells after Matrigel solidified. The cells were then allowed 2-3 days to grow before the media was changed to contain 1 µg/mL of puromycin or 2.5 µg/mL of blasticidin for the dCas9-Krab *Klf4* knockdown and *TRE-KLF4* overexpression constructs, respectively.

Organoid growth assays

To determine differences in organoid growth rates, organoids were first dissociated to single cells (see **Organoids**) then passed through a 35 µm filter to ensure single cells. Organoid cells were then plated in 20 µL Matrigel domes in 48 well tissue culture plates and 200 µL of OCM media per well was added to cover the domes. Media was changed every 3-4 days and for organoids treated with doxycycline (dox), fresh dox was added every 48 hours. Wells were imaged daily using a 4x objective and eight images were taken per well to capture the entire Matrigel dome. Organoids were counted blindly using ImageJ and the number of organoids per well was determined by combining all organoids counted in eight images. Wells in which the entire plug was not captured in eight images were removed.

Formalin fixation of organoids

Media was removed from wells and Matrigel plugs were washed with PBS once before resuspension in 0.5 mL of PBS. Resuspend organoids were fixed using 10% formalin overnight and then transferred to 70% ethanol. Fixed organoids were then centrifuged at 2500 rpm and washed with PBS twice before being placed in 2.5% agarose for embedding. Agarose embedded organoids were then placed in 70% ethanol and submitted for embedding in paraffin.

KLF4 modulation in human basal cells

For lentiviral CRISPR-mediated knockout, the following guideRNAs targeting *KLF4* were cloned into the lentiCRISPRv2 ((Addgene # 52961) (LCV2): sg*KLF4*-1 (5'-GTGGTGGCGCCCTACAACGG -3'), sg*KLF4*-2 (5'- TCCCGGCCGCCCCTATGACC-3') and sg*KLF4-3* (5'- AATTGGAGAGAATAAAGTCC-3'). For overexpression of KLF4 in the two basal cell lines, lentiviral infection of TRE KLF4 was performed.

T7 surveyor assay

Genomic DNA was extracted following the protocol in the Qiagen DNeasy Blood and Tissue Kit (Cat # 69504). Following extraction, the targeted genomic regions were PCR amplified with the following sequence primers:

To isolate the region surrounding the guide binding site for LCV2 *sgKLF4-2*, the following primers were used LCV2 *sgKLF4-2* FWD Primer-1: 5' ACCAAGAGCCACTGAACGAG 3'; LCV2 *sgKLF4-2* REV Primer-1: 5' CCGGATCGGATAGGTGAAGC 3'; LCV2 *sgKLF4-2* FWD Primer-2: 5' GTGTATGCCCGTGGTGCGA 3'; LCV2 *sgKLF4-2* REV Primer-2: 5' GATGGGTCAGCGAATTGGAGA 3'. The expected band sizes were 620 bp and 300 bp for *sgKLF4-2* primer set 1, and 550 bp and 170 bp for *sgKLF4-2* primer set 2.

The PCR conditions used for both primer sets were 30 sec for 98°C, 35 cycles of 5 seconds for 98°C, 10 seconds for 69°C, and 20 seconds for 72°C, 2 minutes for 72°C then held at 4°C.

To isolate the region surrounding the guide binding site for LCV2 *sgKLF4-3*, the following primers were used LCV2 *sgKLF4-3* FWD Primer-1: 5' ATCTTTCTCCACGTTCGCGT 3'; LCV2 *sgKLF4-3* REV Primer-1: 5' TTGATGTCCGCCAGGTTGAA 3'; LCV2 *sgKLF4-3* FWD Primer-2: 5' CCAGCACGTCAGTATGTCGG 3'; LCV2 *sgKLF4-3* REV Primer-2: 5' CGCGTAATCACAAGTGTGGG 3'. The expected band sizes were 1000 bp and 200 bp for *sgKLF4-3* primer set 1, and 250 bp and 1000 bp for *sgKLF4-3* primer set 2.

Q5 high fidelity 2x master mix (NEB Catalog # M0492) was used for the PCR reactions according to the manufacturer protocol. PCR products were then mixed with 6x loading dye and ran on 2% agarose gels. Gels were imaged to confirm correct PCR product size. The PCR products were then gel extracted using the QIAquick gel extraction kit (Qiagen #28706). DNA concentrations were measured on a Biotek Synergy HT plate reader.

200ng of gel-purified DNA was used for endonuclease reaction. The T7 surveyor assay was performed according to the "Determining Genome Targeting Efficiency using T7 Endonuclease I" protocol from NEB. Briefly, 200ng of gel-purified DNA was hybridized in a thermocycler by initial denaturing at 95°C for 5 minutes followed by annealing from 95°C to 25°C. Annealed DNA was then subjected to the T7 endonuclease at 37°C for 15 min. Following the enzymatic reaction DNA was run on 2% gels and imaged.

Immunoblot

Total protein lysates were prepared as previously described, separated via SDS-PAGE, and transferred to a PVDF membrane[112]. Membranes were blocked for 1 h in 5% milk followed by overnight incubation with primary antibodies at 4°C. All primary antibodies were diluted 1:1000 and include: KRT13 (Abcam Cat # 92551); KRT5 (Biolegend Cat #905501); KLF4 (CST Cat # 4038s); CAS9 (Active Motif Cat # 61798); HSP90 (CST Cat # 4877); P21 (CST Cat # 2947s); PITX1 (Novus NBP1-88644); and GFP (CST Cat #2956s). Membranes were washed for three times for 5 min at room temperature in TBS-T. Mouse and rabbit HRP-conjugated secondary antibodies (1:5000; Jackson ImmunoResearch Cat # 711-035-152 and Cat # 115-035-205) were incubated for 1 hr in 5% milk at room temperature followed by washing three times for 5 minutes at room temperature in TBS-T. Membranes were exposed to WesternBright HRP quantum substrate (Advansta) and detected on Hyblot CL film (Denville Scientific, Inc.).

Bulk RNA-seq sample and data processing from GEMMs

5 µm serial sections of FFPE tissue from LP and KP tumor bearing lungs, initiated with CMV-Cre, were stained with H&E and Aniline blue. H&E-slides were used to confirm histology and guide bordering of Aniline-blue-stained serial sections. Using a microdissection scope, bordered tissue samples were dissected from consecutive serial sections using sterile scalpels and razors. 10-14 pieces of each sample isolated were pooled in a microcentrifuge tube containing 100% ethanol. RecoverAll Total Nucleic Acid Isolation Kit for FFPE (Thermo Fisher Cat # AM1975) was used to isolate RNA from FFPE tissues according to the manufacturer's protocol. RNA isolation from ~15 mg flash-frozen SNL (CMV-Cre initiated), RPM (10 CGRP-Cre initiated tumors, 5 CMV-Cre imitated tumors), and RPR2 (4 CGRP-Cre initiated tumors, 4 CMV-Cre initiated tumors) primary tumors was performed using the RNeasy Mini Kit (Qiagen Cat # 74104) with the standard protocol. RNA from SNL, LP, KP, RPM, and RPR2 tumors were subject to library construction with the Illumina TruSeq Stranded mRNA Sample Preparation Kit (Cat # RS-122-2101, RS-122-2102) or the Illumina TruSeq Stranded Total RNA Library Ribo-Zero Gold Prep kit (Cat # RS-122-2301), according to the manufacturer's protocol. LP and KP tumor samples were sequenced on an Illumina HiSeq 2500 instrument, 1x50 bp read lengths. RPR2, RPM, and SNL samples were sequenced on the Illumina NovaSeq 6000, 2x50 bp read lengths or the Illumina HiSeq 2500, 1x50 bp read lengths.

Fastq raw count files were aligned in the R statistical environment. The mouse GRCm38 FASTA and GTF files were downloaded from Ensembl release 94 and the reference database was created using STAR[113] with splice junctions optimized for 50 base pair reads and in two pass mode to output a BAM file sorted by coordinates. Mapped reads were assigned to annotated genes in the GTF file using featureCounts version 1.6.3[114]. The output files from cutadapt, FastQC, Picard CollectRnaSeqMetrics, STAR and featureCounts were summarized using MultiQC[115] to check for any sample outliers. To remove sources of unwanted variation from tumor RNA-seq sample preparation, all non-coding features, histones, and ribosomal RNAs were removed from count matrices for downstream analyses. The featureCount output files were combined into a single raw count matrix. Log2(counts+1)-transformed, normalized intensity values were obtained following DESeq2 analysis in R to obtain gene expression levels for GEMM tumors.

Bulk RNA-seq sample and data processing from BEAS-2B

Total RNA was extracted from BEAS-2B cells with TRE-KLF4 after 48 hours of 250 ng/mL doxycycline treatment using a RNeasy mini kit (Qiagen Cat. No. 74104). RNA was processed by Medgenome and cDNA libraries were prepared using the Illumina TruSeq Stranded Total RNA kit. Libraries were sequenced on a NovaSeq 600/ X Plus with a read length of 100 aiming for at least 30M paired reads per sample. Each sample had over 70M reads and reads were aligned to the reference human genome (GRCh38/hg38) using STAR, raw read counts were estimated using HTSeq, and read counts were normalized using DESeq2. Log2FC of differentially expressed genes between doxycycline and no doxycycline were used to perform GSEA using R and ENRICHR analysis. GSEA analysis was performed on normal hillock and basal scores along with Msigdb hallmark gene sets with a pvalue cutoff of 0.1.

MicroCT imaging

SNL mice were imaged to monitor tumor development using microCT imaging beginning four weeks after tumor initiation by Cre-mediated recombination. Mice were anesthetized with isoflurane and imaged using a small animal Quantum GX2 microCT (Perkin Elmer). Quantum GX2 images were acquired with 2-minute scans at 45 mm resolution, 90 kV, with 88 mA of current.

Data availability

All raw and processed data will be made available through the Gene Expression Omnibus (GEO) upon final publication of this manuscript (GSE-290744, GSE-290745, and GSE-290746). Human lung squamous cell carcinoma data from Wu et al[68] can be found at GEO under the accession code GSE148071, human head and neck squamous cell carcinoma data from Kurten et al[87] can be found at GEO under the accession code GSE164690, and human esophageal squamous cell carcinoma data from Zhang et al[88] can be found at GEO under the accession code GSE160269.

All original code has been deposited on GitHub (https://github.com/TOliver-Lab/Izzo_Hillock_2025) and available on the date of publication. Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

Statistical analyses

Remaining statistical analysis was performed using GraphPad Prism or Python. Error bars show mean ± SD unless otherwise specified. Significance was determined by Student's two-tailed unpaired t-tests with 95% confidence intervals and p values <0.05, by one-way ANOVA correcting for multiple comparisons using Tukey's Honest Significant Difference test, or by two-way ANOVA with a Sidak correction for multiple comparisons. All statistical details are further described in respective figure legends.

SUPPLEMENTAL TABLES

Table S1: Hillock score genes for mouse and human, related to Figure 3.

Table S2: Basal, tuft, goblet, ciliated, and club scores related to Figure 3.

Table S3: Mucinous adenocarcinoma score genes for mouse and human, related to Figure 3.

Table S4: Basal hillock score genes for mouse and human, related to Figure 3.

Table S5: Luminal hillock score genes for mouse and human, related to Figure 3.

Table S6: Differentially expressed genes in SNL tumor cells derived from scRNA-seq data, related to Figure 4.

Table S7: Human patient data hillock-like cells per patient, related to Figure 4.

Table S8: Differentially expressed genes in hillock-like human tumor cells derived from scRNA-seq data, related to Figure 4.

Table S9: Differentially expressed genes in hillock-like tumor organoid cells derived from scRNA-seq data, related to Figure S5.

Table S10: ENCODE and ChEA Consensus TFs from ChIP-X from mouse and human hillock-like and basal-like cells, related to S6.

Table S11: Differentially expressed genes in BEAS-2B cells with TRE-KLF4, related to Figure 8.

References

1 Flavahan, W. A., Gaskell, E. & Bernstein, B. E. Epigenetic plasticity and the hallmarks of cancer. *Science* **357**, eaal2380 (2017). https://doi.org/10.1126/science.aal2380

2 Hinohara, K. & Polyak, K. Intratumoral Heterogeneity: More Than Just Mutations. *Trends Cell Biol* **29**, 569-579 (2019). https://doi.org/10.1016/j.tcb.2019.03.003

3 Ciriello, G. *et al.* Cancer Evolution: A Multifaceted Affair. *Cancer Discov* **14**, 36-48 (2024). https://doi.org/10.1158/2159-8290.CD-23-0530

4 Puram, S. V. *et al.* Single-Cell Transcriptomic Analysis of Primary and Metastatic Tumor Ecosystems in Head and Neck Cancer. *Cell* **171**, 1611-1624 e1624 (2017). https://doi.org/10.1016/j.cell.2017.10.044

5 Li, H. *et al.* Reference component analysis of single-cell transcriptomes elucidates cellular heterogeneity in human colorectal tumors. *Nat Genet* **49**, 708-718 (2017). https://doi.org/10.1038/ng.3818

6 Montoro, D. T. *et al.* A revised airway epithelial hierarchy includes CFTR-expressing ionocytes. *Nature* **560**, 319-324 (2018). https://doi.org/10.1038/s41586-018-0393-7

7 Plasschaert, L. W. *et al.* A single-cell atlas of the airway epithelium reveals the CFTR-rich pulmonary ionocyte. *Nature* **560**, 377-381 (2018). https://doi.org/10.1038/s41586-018-0394-6

8 Chen, Z., Fillmore, C. M., Hammerman, P. S., Kim, C. F. & Wong, K. K. Non-small-cell lung cancers: a heterogeneous set of diseases. *Nat Rev Cancer* **14**, 535-546 (2014). https://doi.org/10.1038/nrc3775

9 Campbell, J. D. *et al.* Distinct patterns of somatic genome alterations in lung adenocarcinomas and squamous cell carcinomas. *Nat Genet* **48**, 607-616 (2016). https://doi.org/10.1038/ng.3564

10 Reck, M. & Rabe, K. F. Precision Diagnosis and Treatment for Advanced Non-Small-Cell Lung Cancer. *N Engl J Med* **377**, 849-861 (2017). https://doi.org/10.1056/NEJMra1703413

11 Herbst, R. S., Morgensztern, D. & Boshoff, C. The biology and management of non-small cell lung cancer. *Nature* **553**, 446-454 (2018). https://doi.org/10.1038/nature25183

12 Oliver, T. G., Patel, J. & Akerley, W. Squamous non-small cell lung cancer as a distinct clinical entity. *Am J Clin Oncol* **38**, 220-226 (2015). https://doi.org/10.1097/COC.0b013e3182a0e850

13 Pisco, A. O. & Huang, S. Non-genetic cancer cell plasticity and therapy-induced stemness in tumour relapse:'What does not kill me strengthens me'. *Br J Cancer* **112**, 1725-1732 (2015). https://doi.org/10.1038/bjc.2015.146

14 Laisne, M., Lupien, M. & Vallot, C. Epigenomic heterogeneity as a source of tumour evolution. *Nat Rev Cancer* (2024). https://doi.org/10.1038/s41568-024-00757-9

15 Hanahan, D. Hallmarks of Cancer: New Dimensions. *Cancer Discovery* **12**, 31-46 (2022). https://doi.org/10.1158/2159-8290.Cd-21-1059

16 Boumahdi, S. & de Sauvage, F. J. The great escape: tumour cell plasticity in resistance to targeted therapy. *Nature Reviews Drug Discovery* **19**, 39-56 (2020). https://doi.org/10.1038/s41573-019-0044-1

17 Gupta, P. B., Pastushenko, I., Skibinski, A., Blanpain, C. & Kuperwasser, C. Phenotypic Plasticity: Driver of Cancer Initiation, Progression, and Therapy Resistance. *Cell Stem Cell* **24**, 65-78 (2019). https://doi.org/10.1016/j.stem.2018.11.011

18 Xu, C. *et al.* Loss of Lkb1 and Pten leads to lung squamous cell carcinoma with elevated PD-L1 expression. *Cancer Cell* **25**, 590-604 (2014). https://doi.org/10.1016/j.ccr.2014.03.033

19 Malanchi, I. *et al.* Cutaneous cancer stem cell maintenance is dependent on β-catenin signalling. *Nature* **452**, 650-653 (2008). https://doi.org/10.1038/nature06835

20 Schober, M. & Fuchs, E. Tumor-initiating stem cells of squamous cell carcinomas and their control by TGF-β and integrin/focal adhesion kinase (FAK) signaling. *Proceedings of the National Academy of Sciences* **108**, 10544-10549 (2011). https://doi.org/doi:10.1073/pnas.1107807108

21 Boumahdi, S. *et al.* SOX2 controls tumour initiation and cancer stem-cell functions in squamous-cell carcinoma. *Nature* **511**, 246-250 (2014). https://doi.org/10.1038/nature13305

22 Choi, J.-E. *et al.* A unique subset of glycolytic tumour-propagating cells drives squamous cell carcinoma. *Nature Metabolism* **3**, 182-195 (2021). https://doi.org/10.1038/s42255-021-00350-6

23 Sastre-Perona, A. *et al.* De Novo PITX1 Expression Controls Bi-Stable Transcriptional Circuits to Govern Self-Renewal and Differentiation in Squamous Cell Carcinoma. *Cell Stem Cell* **24**, 390-404.e398 (2019). https://doi.org/10.1016/j.stem.2019.01.003

24 Huang, Y. H. *et al.* POU2F3 is a master regulator of a tuft cell-like variant of small cell lung cancer. *Genes & Development* **32**, 915-928 (2018). https://doi.org/10.1101/gad.314815.118

25 Rudin, C. M. *et al.* Molecular subtypes of small cell lung cancer: a synthesis of human and mouse model data. *Nat Rev Cancer* **19**, 289-297 (2019). https://doi.org/10.1038/s41568-019-0133-9

26 Finlay, J. B. *et al.* Olfactory neuroblastoma mimics molecular heterogeneity and lineage trajectories of small-cell lung cancer. *Cancer Cell* **42**, 1086-1105 e1013 (2024). https://doi.org/10.1016/j.ccell.2024.05.003

27 Patel, A. S. & Yanai, I. A developmental constraint model of cancer cell states and tumor heterogeneity. *Cell* **187**, 2907-2918 (2024). https://doi.org/10.1016/j.cell.2024.04.032

28 van Groningen, T. *et al.* Neuroblastoma is composed of two super-enhancer-associated differentiation states. *Nature Genetics* **49**, 1261-1266 (2017). https://doi.org/10.1038/ng.3899

29 Boeva, V. *et al.* Heterogeneity of neuroblastoma cell identity defined by transcriptional circuitries. *Nature Genetics* **49**, 1408-1413 (2017). https://doi.org/10.1038/ng.3921

30 Ferone, G. *et al.* SOX2 Is the Determining Oncogenic Switch in Promoting Lung Squamous Cell Carcinoma from Different Cells of Origin. *Cancer Cell* **30**, 519-532 (2016). https://doi.org/10.1016/j.ccell.2016.09.001

31 Sutherland, K. D. & Berns, A. Cell of origin of lung cancer. *Mol Oncol* **4**, 397-403 (2010). https://doi.org/10.1016/j.molonc.2010.05.002

32 Rock, J. R. *et al.* Basal cells as stem cells of the mouse trachea and human airway epithelium. *Proc Natl Acad Sci U S A* **106**, 12771-12775 (2009). https://doi.org/10.1073/pnas.0906850106

33 Goldfarbmuren, K. C. *et al.* Dissecting the cellular specificity of smoking effects and reconstructing lineages in the human airway epithelium. *Nat Commun* **11**, 2485 (2020). https://doi.org/10.1038/s41467-020-16239-z

34 Pardo-Saganta, A. *et al.* Injury induces direct lineage segregation of functionally distinct airway basal stem/progenitor cell subpopulations. *Cell Stem Cell* **16**, 184-197 (2015). https://doi.org/10.1016/j.stem.2015.01.002

35 Henry, G. H. *et al.* A Cellular Anatomy of the Normal Adult Human Prostate and Prostatic Urethra. *Cell Rep* **25**, 3530-3542 e3535 (2018). https://doi.org/10.1016/j.celrep.2018.11.086

36 Lin, B. *et al.* Airway hillocks are injury-resistant reservoirs of unique plastic stem cells. *Nature* **629**, 869-877 (2024). https://doi.org/10.1038/s41586-024-07377-1

37 Ferone, G., Lee, M. C., Sage, J. & Berns, A. Cells of origin of lung cancers: lessons from mouse studies. *Genes Dev* **34**, 1017-1032 (2020). https://doi.org/10.1101/gad.338228.120

38 Pan, Y., Han, H., Labbe, K. E., Zhang, H. & Wong, K. K. Recent advances in preclinical models for lung squamous cell carcinoma. *Oncogene* **40**, 2817-2829 (2021). https://doi.org/10.1038/s41388-021-01723-7

39 Mollaoglu, G. *et al.* The Lineage-Defining Transcription Factors SOX2 and NKX2-1 Determine Lung Cancer Cell Fate and Shape the Tumor Immune Microenvironment. *Immunity* **49**, 764-779 e769 (2018). https://doi.org/10.1016/j.immuni.2018.09.020

40 Jeong, Y. *et al.* Role of KEAP1/NRF2 and TP53 Mutations in Lung Squamous Cell Carcinoma Development and Radiation Resistance. *Cancer Discov* **7**, 86-101 (2017). https://doi.org/10.1158/2159-8290.CD-16-0127

41 Justilien, V. *et al.* The PRKCI and SOX2 oncogenes are coamplified and cooperate to activate Hedgehog signaling in lung squamous cell carcinoma. *Cancer Cell* **25**, 139-151 (2014). https://doi.org/10.1016/j.ccr.2014.01.008

42 Liu, Y. *et al.* Chromosome 3q26 Gain Is an Early Event Driving Coordinated Overexpression of the PRKCI, SOX2, and ECT2 Oncogenes in Lung Squamous Cell Carcinoma. *Cell Rep* **30**, 771-782 e776 (2020). https://doi.org/10.1016/j.celrep.2019.12.071

43 Zhang, H. *et al.* Lkb1 inactivation drives lung cancer lineage switching governed by Polycomb Repressive Complex 2. *Nat Commun* **8**, 14922 (2017). https://doi.org/10.1038/ncomms14922

44 Tata, P. R. *et al.* Developmental History Provides a Roadmap for the Emergence of Tumor Plasticity. *Dev Cell* **44**, 679-693 e675 (2018). https://doi.org/10.1016/j.devcel.2018.02.024

45 Han, X. *et al.* Transdifferentiation of lung adenocarcinoma in mice with Lkb1 deficiency to squamous cell carcinoma. *Nat Commun* **5**, 3261 (2014). https://doi.org/10.1038/ncomms4261

46 Li, F. *et al.* LKB1 Inactivation Elicits a Redox Imbalance to Modulate Non-small Cell Lung Cancer Plasticity and Therapeutic Response. *Cancer Cell* **27**, 698-711 (2015). https://doi.org/10.1016/j.ccell.2015.04.001

47 Gao, Y. *et al.* YAP inhibits squamous transdifferentiation of Lkb1-deficient lung adenocarcinoma through ZEB2-dependent DNp63 repression. *Nat Commun* **5**, 4629 (2014). https://doi.org/10.1038/ncomms5629

48 Mukhopadhyay, A. *et al.* Sox2 cooperates with Lkb1 loss in a mouse model of squamous cell lung cancer. *Cell Rep* **8**, 40-49 (2014). https://doi.org/10.1016/j.celrep.2014.05.036

49 Yuan, G. *et al.* Elevated NSD3 histone methylation activity drives squamous cell lung cancer. *Nature* **590**, 504-508 (2021). https://doi.org/10.1038/s41586-020-03170-y

50 Ji, H. *et al.* LKB1 modulates lung cancer differentiation and metastasis. *Nature* **448**, 807-810 (2007). https://doi.org/10.1038/nature06030

51 Tetreault, M. P., Yang, Y. & Katz, J. P. Kruppel-like factors in cancer. *Nat Rev Cancer* **13**, 701-713 (2013). https://doi.org/10.1038/nrc3582

52 McConnell, B. B. & Yang, V. W. Mammalian Kruppel-like factors in health and diseases. *Physiol Rev* **90**, 1337-1381 (2010). https://doi.org/10.1152/physrev.00058.2009

53 He, Z., He, J. & Xie, K. KLF4 transcription factor in tumorigenesis. *Cell Death Discov* **9**, 118 (2023). https://doi.org/10.1038/s41420-023-01416-y

54 McGinn, J. *et al.* A biomechanical switch regulates the transition towards homeostasis in oesophageal epithelium. *Nat Cell Biol* **23**, 511-525 (2021). https://doi.org/10.1038/s41556-021-00679-w

55 Sen, G. L. *et al.* ZNF750 Is a p63 Target Gene that Induces KLF4 to Drive Terminal Epidermal Differentiation. *Developmental Cell* **22**, 669-677 (2012). https://doi.org/10.1016/j.devcel.2011.12.001

56 Segre, J. A., Bauer, C. & Fuchs, E. Klf4 is a transcription factor required for establishing the barrier function of the skin. *Nat Genet* **22**, 356-360 (1999). https://doi.org/10.1038/11926

57 Tetreault, M. P. *et al.* Esophageal squamous cell dysplasia and delayed differentiation with deletion of kruppel-like factor 4 in murine esophagus. *Gastroenterology* **139**, 171-181 e179 (2010). https://doi.org/10.1053/j.gastro.2010.03.048

58 Jackson, E. L. *et al.* The differential effects of mutant p53 alleles on advanced murine lung cancer. *Cancer Res* **65**, 10280-10288 (2005). https://doi.org/10.1158/0008-5472.CAN-05-2193

59 Mollaoglu, G. *et al.* MYC Drives Progression of Small Cell Lung Cancer to a Variant Neuroendocrine Subtype with Vulnerability to Aurora Kinase Inhibition. *Cancer Cell* **31**, 270-285 (2017). https://doi.org/10.1016/j.ccell.2016.12.005

60 Schaffer, B. E. *et al.* Loss of p130 accelerates tumor development in a mouse model for human small-cell lung carcinoma. *Cancer Res* **70**, 3877-3883 (2010). https://doi.org/10.1158/0008-5472.CAN-09-4228

61 Snyder, E. L. *et al.* Nkx2-1 represses a latent gastric differentiation program in lung adenocarcinoma. *Mol Cell* **50**, 185-199 (2013). https://doi.org/10.1016/j.molcel.2013.02.018

62 Tata, P. R. *et al.* Dedifferentiation of committed epithelial cells into stem cells. *Nature* **503**, 218-+ (2013). https://doi.org/10.1038/nature12777

63 Updegraff, B. L. *et al.* Transmembrane Protease TMPRSS11B Promotes Lung Cancer Growth by Enhancing Lactate Export and Glycolytic Metabolism. *Cell Rep* **25**, 2223-2233 e2226 (2018). https://doi.org/10.1016/j.celrep.2018.10.100

64 Yamada, Y. *et al.* Pulmonary cancers across different histotypes share hybrid tuft cell/ionocyte-like molecular features and potentially druggable vulnerabilities. *Cell Death Dis* **13**, 979 (2022). https://doi.org/10.1038/s41419-022-05428-x

65 Yamada, Y. *et al.* A Tuft Cell-Like Signature Is Highly Prevalent in Thymic Squamous Cell Carcinoma and Delineates New Molecular Subsets Among the Major Lung Cancer Histotypes. *J Thorac Oncol* **16**, 1003-1016 (2021). https://doi.org/10.1016/j.jtho.2021.02.008

66 Jimbo, N. *et al.* The expression of YAP1 and other transcription factors contributes to lineage plasticity in combined small cell lung carcinoma. *J Pathol Clin Res* **10**, e70001 (2024). https://doi.org/10.1002/2056-4538.70001

67 Ireland, A. S. *et al.* Basal cell of origin resolves neuroendocrine–tuft lineage plasticity in cancer. *bioRxiv*, 2024.2011.2013.623500 (2024). https://doi.org/10.1101/2024.11.13.623500

68 Wu, F. *et al.* Single-cell profiling of tumor heterogeneity and the microenvironment in advanced non-small cell lung cancer. *Nat Commun* **12**, 2540 (2021). https://doi.org/10.1038/s41467-021-22801-0

69 Hai, J. *et al.* Generation of Genetically Engineered Mouse Lung Organoid Models for Squamous Cell Lung Cancers Allows for the Study of Combinatorial Immunotherapy. *Clin Cancer Res* **26**, 3431-3442 (2020). https://doi.org/10.1158/1078-0432.CCR-19-1627

70 Tong, X. *et al.* Adeno-to-squamous transition drives resistance to KRAS inhibition in LKB1 mutant lung cancer. *Cancer Cell* **42**, 413-428.e417 (2024). https://doi.org/https://doi.org/10.1016/j.ccell.2024.01.012

71 Shi, R. *et al.* Organoid cultures as preclinical models of non–small cell lung cancer. *Clinical Cancer Research* **26**, 1162-1174 (2020).

72 Kim, M. *et al.* Patient-derived lung cancer organoids as in vitro cancer models for therapeutic screening. *Nature Communications* **10**, 3991 (2019). https://doi.org/10.1038/s41467-019-11867-6

73 Romero, R. *et al.* The neuroendocrine transition in prostate cancer is dynamic and dependent on ASCL1. *Nature Cancer* **5**, 1641-1659 (2024). https://doi.org/10.1038/s43018-024-00838-6

74 Moreci, R. S. & Lechler, T. Epidermal structure and differentiation. *Curr Biol* **30**, R144-R149 (2020). https://doi.org/10.1016/j.cub.2020.01.004

75 He, H. *et al.* Kruppel-like Factor 4 Promotes Esophageal Squamous Cell Carcinoma Differentiation by Up-regulating Keratin 13 Expression. *J Biol Chem* **290**, 13567-13577 (2015). https://doi.org/10.1074/jbc.M114.629717

76 Ali, A. *et al.* Cell Lineage Tracing Identifies Hormone-Regulated and Wnt-Responsive Vaginal Epithelial Stem Cells. *Cell Rep* **30**, 1463-1477.e1467 (2020). https://doi.org/10.1016/j.celrep.2020.01.003

77 Thakore, P. I. *et al.* Highly specific epigenome editing by CRISPR-Cas9 repressors for silencing of distal regulatory elements. *Nat Methods* **12**, 1143-1149 (2015). https://doi.org/10.1038/nmeth.3630

78 Zhang, W. *et al.* The gut-enriched Krüppel-like factor (Krüppel-like factor 4) mediates the transactivating effect of p53 on the p21 WAF1/Cip1 promoter. *J Biol Chem* **275**, 18391-18398 (2000). https://doi.org/10.1074/jbc.C000062200

79 Rowland, B. D., Bernards, R. & Peeper, D. S. The KLF4 tumour suppressor is a transcriptional repressor of p53 that acts as a context-dependent oncogene. *Nature Cell Biology* **7**, 1074-1082 (2005). https://doi.org/10.1038/ncb1314

80 Rowland, B. D. & Peeper, D. S. KLF4, p21 and context-dependent opposing forces in cancer. *Nat Rev Cancer* **6**, 11-23 (2006). https://doi.org/10.1038/nrc1780

81 Prescott, R. A. *et al.* A comparative study of in vitro air-liquid interface culture models of the human airway epithelium evaluating cellular heterogeneity and gene expression at single cell resolution. *Respir Res* **24**, 213 (2023). https://doi.org/10.1186/s12931-023-02514-2

82 Walters, M. S. *et al.* Generation of a human airway epithelium derived basal cell line with multipotent differentiation capacity. *Respir Res* **14**, 135 (2013). https://doi.org/10.1186/1465-9921-14-135

83 Zhao, F. & Klimecki, W. T. Culture conditions profoundly impact phenotype in BEAS-2B, a human pulmonary epithelial model. *J Appl Toxicol* **35**, 945-951 (2015). https://doi.org/10.1002/jat.3094

84 Lin, E. W. *et al.* Comparative transcriptomes of adenocarcinomas and squamous cell carcinomas reveal molecular similarities that span classical anatomic boundaries. *Plos Genet* **13**, e1006938 (2017). https://doi.org/10.1371/journal.pgen.1006938

85 Yan, W., Wistuba, II, Emmert-Buck, M. R. & Erickson, H. S. Squamous Cell Carcinoma - Similarities and Differences among Anatomical Sites. *Am J Cancer Res* **1**, 275-300 (2011).

86 Henick, B. S. *et al.* Squamous cell cancers of the aero-upper digestive tract: A unified perspective on biology, genetics, and therapy. *Cancer Cell* **43**, 178-194 (2025). https://doi.org/https://doi.org/10.1016/j.ccell.2025.01.003

87 Kürten, C. H. L. *et al.* Investigating immune and non-immune cell interactions in head and neck tumors by single-cell RNA sequencing. *Nature Communications* **12**, 7338 (2021). https://doi.org/10.1038/s41467-021-27619-4

88 Zhang, X. *et al.* Dissecting esophageal squamous-cell carcinoma ecosystem by single-cell transcriptomic analysis. *Nature Communications* **12**, 5291 (2021). https://doi.org/10.1038/s41467-021-25539-x

89 Alencar, G. F. *et al.* Stem Cell Pluripotency Genes Klf4 and Oct4 Regulate Complex SMC Phenotypic Changes Critical in Late-Stage Atherosclerotic Lesion Pathogenesis. *Circulation* **142**, 2045-2059 (2020). https://doi.org/10.1161/CIRCULATIONAHA.120.046672

90 Shankman, L. S. *et al.* KLF4-dependent phenotypic modulation of smooth muscle cells has a key role in atherosclerotic plaque pathogenesis. *Nat Med* **21**, 628-637 (2015). https://doi.org/10.1038/nm.3866

91 Tetreault, M. P., Weinblatt, D., Shaverdashvili, K., Yang, Y. & Katz, J. P. KLF4 transcriptionally activates non-canonical WNT5A to control epithelial stratification. *Sci Rep* **6**, 26130 (2016). https://doi.org/10.1038/srep26130

92 Shaverdashvili, K. *et al.* KLF4 activates NFkappaB signaling and esophageal epithelial inflammation via the Rho-related GTP-binding protein RHOF. *PLoS One* **14**, e0215746 (2019). https://doi.org/10.1371/journal.pone.0215746

93 Singh, A. *et al.* NRF2 Activation Promotes Aggressive Lung Cancer and Associates with Poor Clinical Outcomes. *Clinical Cancer Research* **27**, 877-888 (2021). https://doi.org/10.1158/1078-0432.Ccr-20-1985

94 Jiang, C. *et al.* A CRISPR screen identifies redox vulnerabilities for KEAP1/NRF2 mutant non-small cell lung cancer. *Redox Biology* **54**, 102358 (2022). https://doi.org/https://doi.org/10.1016/j.redox.2022.102358

95 DeBlasi, J. M. *et al.* Distinct Nrf2 Signaling Thresholds Mediate Lung Tumor Initiation and Progression. *Cancer Research* **83**, 1953-1967 (2023). https://doi.org/10.1158/0008-5472.Can-22-3848

96 Hellyer, J. A., Padda, S. K., Diehn, M. & Wakelee, H. A. Clinical Implications of KEAP1-NFE2L2 Mutations in NSCLC. *Journal of Thoracic Oncology* **16**, 395-403 (2021). https://doi.org/10.1016/j.jtho.2020.11.015

97 Kargl, J. *et al.* Neutrophils dominate the immune cell composition in non-small cell lung cancer. *Nat Commun* **8**, 14381 (2017). https://doi.org/10.1038/ncomms14381

98 Akbay, E. A. *et al.* Interleukin-17A Promotes Lung Tumor Progression through Neutrophil Attraction to Tumor Sites and Mediating Resistance to PD-1 Blockade. *J Thorac Oncol* **12**, 1268-1279 (2017). https://doi.org/10.1016/j.jtho.2017.04.017

99 Engblom, C. *et al.* Osteoblasts remotely supply lung tumors with cancer-promoting SiglecF(high) neutrophils. *Science* **358** (2017). https://doi.org/10.1126/science.aal5081

100 Park, J. *et al.* Cancer cells induce metastasis-supporting neutrophil extracellular DNA traps. *Sci Transl Med* **8**, 361ra138 (2016). https://doi.org/10.1126/scitranslmed.aag1711

101 Hu, W. Y. *et al.* Isolation and functional interrogation of adult human prostate epithelial stem cells at single cell resolution. *Stem Cell Res* **23**, 1-12 (2017). https://doi.org/10.1016/j.scr.2017.06.009

102 Li, Q. *et al.* Keratin 13 expression reprograms bone and brain metastases of human prostate cancer cells. *Oncotarget* **7**, 84645-84657 (2016). https://doi.org/10.18632/oncotarget.13175

103 Liu, S. *et al.* Keratin 13 Is Enriched in Prostate Tubule-Initiating Cells and May Identify Primary Prostate Tumors that Metastasize to the Bone. *PLoS One* **11**, e0163232 (2016). https://doi.org/10.1371/journal.pone.0163232

104 Miyoshi, H. & Stappenbeck, T. S. In vitro expansion and genetic modification of gastrointestinal stem cells in spheroid culture. *Nature Protocols* **8**, 2471-2482 (2013). https://doi.org/10.1038/nprot.2013.153

105 Heumos, L. *et al.* Best practices for single-cell analysis across modalities. *Nature Reviews Genetics* **24**, 550-572 (2023). https://doi.org/10.1038/s41576-023-00586-w

106 Luecken, M. D. *et al.* Benchmarking atlas-level data integration in single-cell genomics. *Nature Methods* **19**, 41-50 (2022). https://doi.org/10.1038/s41592-021-01336-8

107 Kolberg, L. *et al.* g:Profiler—interoperable web service for functional enrichment analysis and gene identifier mapping (2023 update). *Nucleic Acids Research* **51**, W207-W212 (2023). https://doi.org/10.1093/nar/gkad347

108 Tirosh, I. *et al.* Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* **352**, 189-196 (2016). https://doi.org/doi:10.1126/science.aad0501

109 Chen, E. Y. *et al.* Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* **14**, 128 (2013). https://doi.org/10.1186/1471-2105-14-128

110 Kuleshov, M. V. *et al.* Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res* **44**, W90-97 (2016). https://doi.org/10.1093/nar/gkw377

111 Xie, Z. *et al.* Gene Set Knowledge Discovery with Enrichr. *Curr Protoc* **1**, e90 (2021). https://doi.org/10.1002/cpz1.90

112 Oliver, T. G. *et al.* Caspase-2-mediated cleavage of Mdm2 creates a p53-induced positive feedback loop. *Mol Cell* **43**, 57-71 (2011). https://doi.org/10.1016/j.molcel.2011.06.012

113 Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21 (2013). https://doi.org/10.1093/bioinformatics/bts635

114 Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923-930 (2014). https://doi.org/10.1093/bioinformatics/btt656

115 Ewels, P., Magnusson, M., Lundin, S. & Kaller, M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32**, 3047-3048 (2016). https://doi.org/10.1093/bioinformatics/btw354

Figure 1

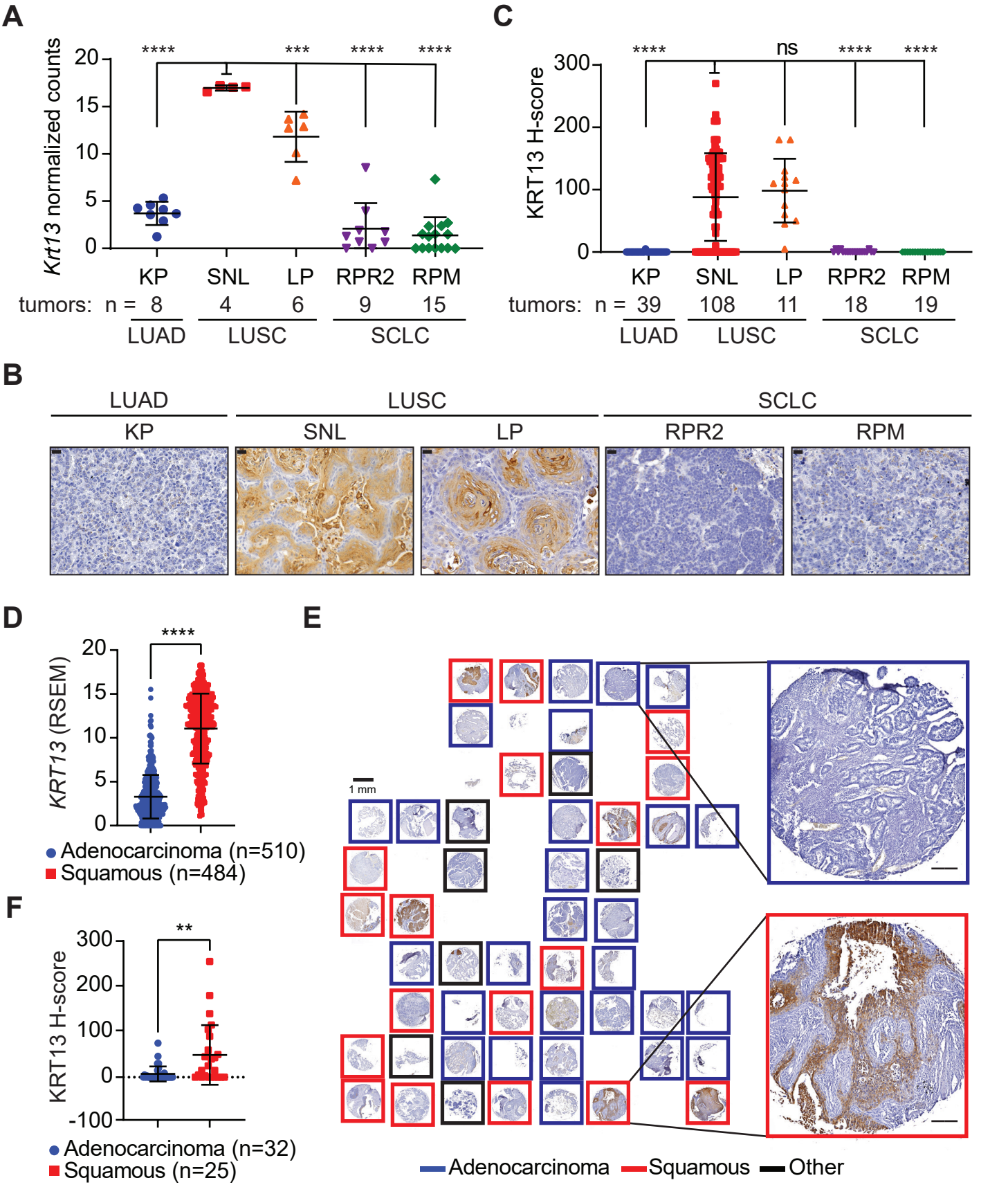
A
Krt13 normalized counts
20
15
10
5
0
****
***
****
****
KP
SNL
LP
RPR2
RPM
tumors: n = 8
4
6
9
15
LUAD
LUSC
SCLC
C
KRT13 H-score
300
200
100
0
****
ns
****
****
KP
SNL
LP
RPR2
RPM
tumors: n = 39
108
11
18
19
LUAD
LUSC
SCLC
B
LUAD
LUSC
SCLC
KP
SNL
LP
RPR2
RPM
D
KRT13 (RSEM)
20
15
10
5
0
****
Adenocarcinoma (n=510)
Squamous (n=484)
E
1 mm
Adenocarcinoma
Squamous
Other
F
KRT13 H-score
300
200
100
0
-100
**
Adenocarcinoma (n=32)
Squamous (n=25)

Figure 1. KRT13 is highly enriched in squamous tumors. A) *Krt13* expression shown as log2 normalized counts by RNA-seq from lung tumors in indicated GEMMs grouped according to histological type. Tumors were collected from the following number of mice for each genotype: RPR2 (n=8), RPM (n=12), SNL (n=4), LP (n=5), KP (n=6). Statistical significance determined by one-way analysis of variance (ANOVA) and p-values are for all pairwise comparisons to SNL. Lung squamous cell carcinoma (LUSC); lung adenocarcinoma (LUAD); small cell lung cancer (SCLC). *Sox2*$^{LSL/LSL}$*;Nkx2-1*$^{fl/fl}$*;Lkb1*$^{fl/fl}$ (SNL); *Lkb1*$^{fl/fl}$*;Pten*$^{fl/fl}$ (LP); *Rb1*$^{fl/fl}$*;Trp53*$^{fl/fl}$*;Myc*$^{T58A/T58A}$ (RPM); *Rb1*$^{fl/fl}$*;Trp53*$^{fl/fl}$*;Rbl2*$^{fl/fl}$ (RPR2); *Kras*$^{G12D/+}$*;Trp53*$^{fl/fl}$ (KP). **B)** Representative immunohistochemistry (IHC) for KRT13 in indicated GEMMs. Scale bar, 20 µm. **C)** H-score quantification of IHC in **1B** (see **Methods**). Number of tumors indicated in figure. Statistical significance was determined by one-way ANOVA and p-values are for all pairwise comparisons to SNL. **D)** *KRT13* expression shown as RSEM from human lung adenocarcinoma and squamous cell carcinoma from TCGA. Number of tumors indicated in figure. Two-tailed t-test. **E)** Representative KRT13 IHC in human lung tumor tissue microarrays (red = squamous cell carcinoma; blue = adenocarcinoma; black = NSCLC/large cell). Scale bar, 1 mm. Higher magnification insets for adenocarcinoma (blue) and squamous cell carcinoma (red); scale bar, 100 µm. **F)** H-score quantification of IHC in **1E** with number of tumors indicated. Two-tailed t-test. **For all panels**, each point represents an individual tumor, and error bars represent mean +/- SD. ns = not significant; **p ≤ 0.01; ***p ≤ 0.001; ****p ≤ 0.0001.

Figure 2

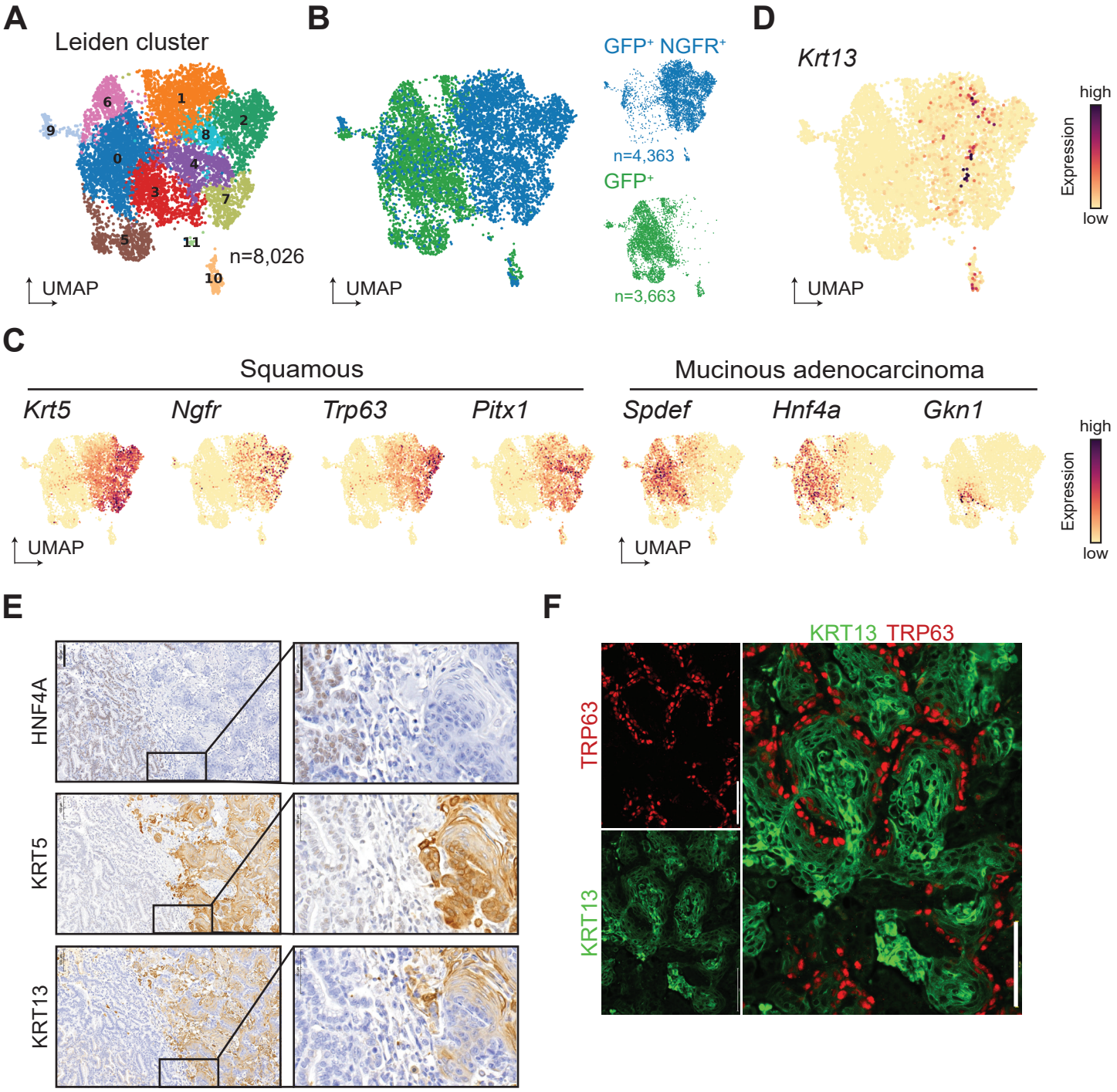
A
Leiden cluster
0
1
2
3
4
5
6
7
8
9
10
11
n=8,026
UMAP
B
GFP+ NGFR+
n=4,363
GFP+
n=3,663
UMAP
D
Krt13
high
Expression
low
UMAP
C
Squamous
Mucinous adenocarcinoma
Krt5
Ngfr
Trp63
Pitx1
Spdef
Hnf4a
Gkn1
high
Expression
low
UMAP
UMAP
E
HNF4A
KRT5
KRT13
F
KRT13 TRP63
TRP63
KRT13

Figure 2. KRT13 is expressed in a subpopulation of squamous tumors. A) ScRNA-seq of SNL-CMV tumor cells labeled by Leiden cluster in UMAP space. **B)** SNL-CMV tumor cells labeled by sorted population (GFP^+ or $GFP^+/NGFR^+$) in UMAP space as in **2A**. **C)** Expression of indicated squamous or mucinous adenocarcinoma gene markers in UMAP space from cells in **2B**. **D)** ScRNA-seq expression of *Krt13* projected on UMAP space from cells in **2B**. **E)** Representative IHC from serial sections in SNL-CMV tumor-bearing lungs harboring both squamous cell carcinoma and mucinous adenocarcinoma. Representative of n = 6 SNL-CMV mice. Scale bar, 100 µm for left column; 50 µm for high-magnification insets on right column. **F)** Representative immunofluorescence co-staining for KRT13 (green) and TRP63 (red) in SNL-CMV tumors. Representative of n = 4 SNL-CMV mice. Scale bars, 75 µm.
See also Figure S1.

Figure 3

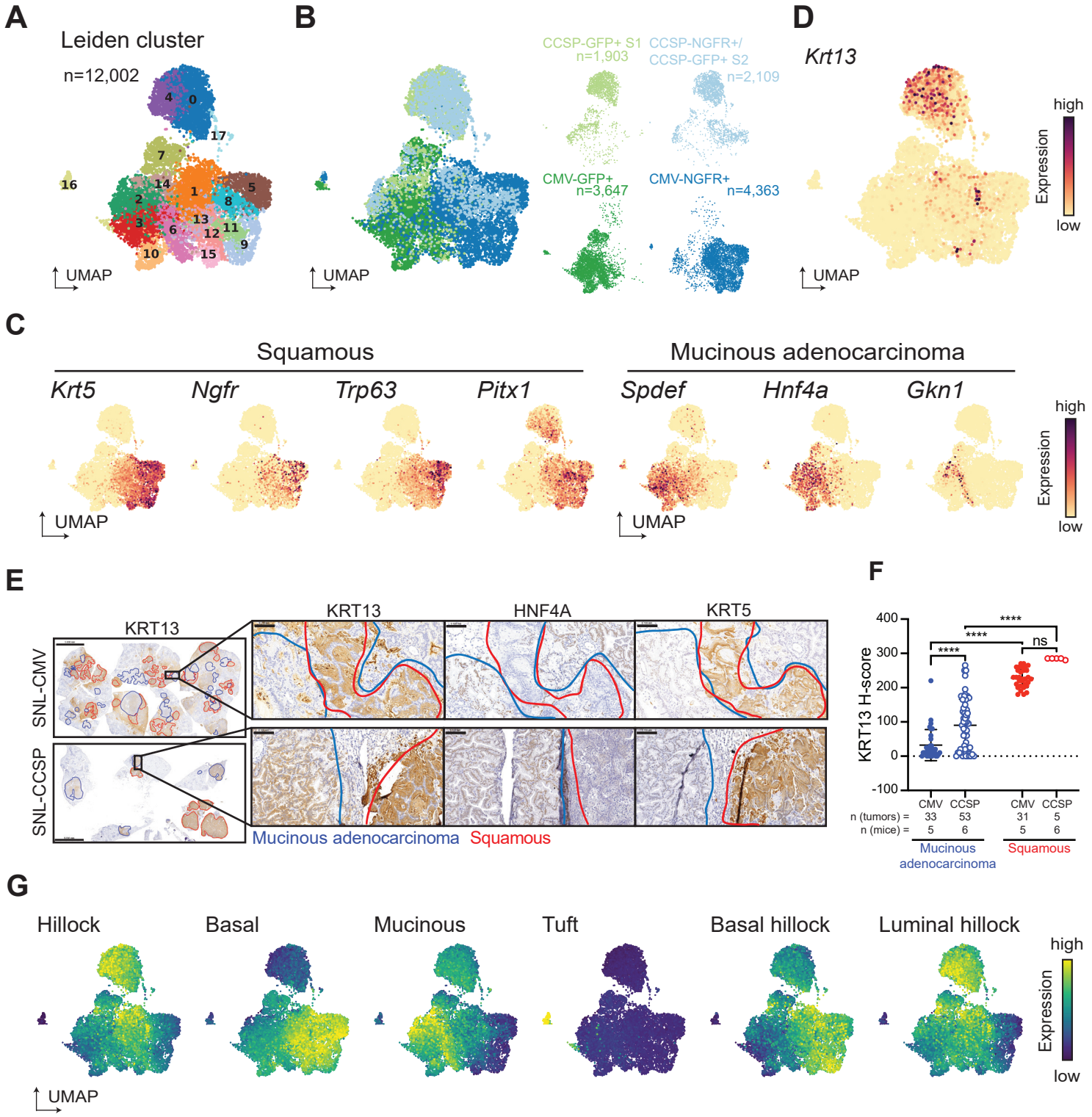
A
Leiden cluster
n=12,002
UMAP
B
CCSP-GFP+ S1
n=1,903
CCSP-NGFR+/
CCSP-GFP+ S2
n=2,109
CMV-GFP+
n=3,647
CMV-NGFR+
n=4,363
UMAP
D
Krt13
high
Expression
low
UMAP
C
Squamous
Mucinous adenocarcinoma
Krt5
Ngfr
Trp63
Pitx1
Spdef
Hnf4a
Gkn1
high
Expression
low
UMAP
UMAP
E
KRT13
KRT13
HNF4A
KRT5
SNL-CMV
SNL-CCSP
Mucinous adenocarcinoma
Squamous
F
KRT13 H-score
400
300
200
100
0
-100
****
****
****
ns
CMV
CCSP
CMV
CCSP
n (tumors) =
33
53
31
5
n (mice) =
5
6
5
6
Mucinous adenocarcinoma
Squamous
G
Hillock
Basal
Mucinous
Tuft
Basal hillock
Luminal hillock
high
Expression
low
UMAP

Figure 3. Club-initiated tumors have early and enriched KRT13 expression. A) ScRNA-seq UMAP of primary CMV- and CCSP-SNL tumor cells labeled by Leiden clusters. **B)** CMV- and CCSP-initiated SNL tumor cells labeled by sorted population (GFP^+ or $GFP^+/NGFR^+$) in UMAP space as in **3A**. **C)** Expression of indicated squamous or mucinous adenocarcinoma gene markers in UMAP space from cells in **3B**. **D)** scRNA-seq expression of *Krt13* projected on UMAP space from cells in **3B**. **E)** Representative KRT5, HNF4A, and KRT13 IHC for CMV- and CCSP-initiated SNL tumors. Squamous (red) and mucinous adenocarcinoma (blue) tumor lesions are outlined. Scale bar, 5 mm for left panels; 2 mm for high-magnification insets on serial sections in right three panels. Representative of n = 5 SNL-CMV and n = 6 SNL-CCSP mice. **F)** H-score quantification of IHC in **3E**. Two-way ANOVA with a Sidak correction for multiple comparisons. Not all comparisons are shown. Each point represents an individual tumor, and error bars represent mean +/- SD. ns = not significant; ****$p \leq 0.0001$. **G)** Gene scores applied to SNL tumor cells projected onto UMAP space from **3A** (see **Methods** and **Tables S1-S5**).
See also Figure S2.

Figure 4

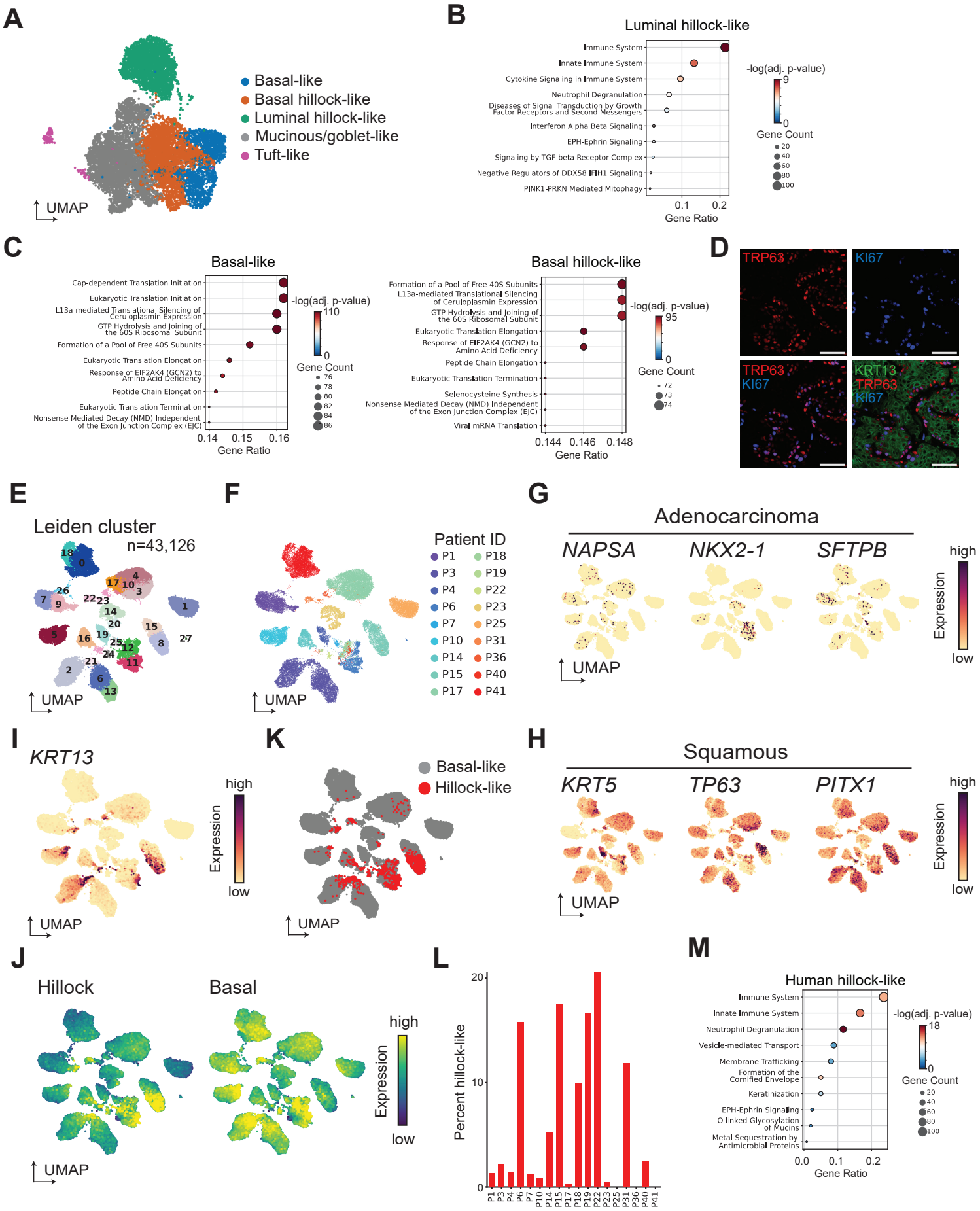
A
Basal-like
Basal hillock-like
Luminal hillock-like
Mucinous/goblet-like
Tuft-like
UMAP
B
Luminal hillock-like
Immune System
Innate Immune System
Cytokine Signaling in Immune System
Neutrophil Degranulation
Diseases of Signal Transduction by Growth Factor Receptors and Second Messengers
Interferon Alpha Beta Signaling
EPH-Ephrin Signaling
Signaling by TGF-beta Receptor Complex
Negative Regulators of DDX58 IFIH1 Signaling
PINK1-PRKN Mediated Mitophagy
Gene Ratio
-log(adj. p-value)
Gene Count
C
Basal-like
Cap-dependent Translation Initiation
Eukaryotic Translation Initiation
L13a-mediated Translational Silencing of Ceruloplasmin Expression
GTP Hydrolysis and Joining of the 60S Ribosomal Subunit
Formation of a Pool of Free 40S Subunits
Eukaryotic Translation Elongation
Response of EIF2AK4 (GCN2) to Amino Acid Deficiency
Peptide Chain Elongation
Eukaryotic Translation Termination
Nonsense Mediated Decay (NMD) Independent of the Exon Junction Complex (EJC)
Gene Ratio
Basal hillock-like
Formation of a Pool of Free 40S Subunits
L13a-mediated Translational Silencing of Ceruloplasmin Expression
GTP Hydrolysis and Joining of the 60S Ribosomal Subunit
Eukaryotic Translation Elongation
Response of EIF2AK4 (GCN2) to Amino Acid Deficiency
Peptide Chain Elongation
Eukaryotic Translation Termination
Selenocysteine Synthesis
Nonsense Mediated Decay (NMD) Independent of the Exon Junction Complex (EJC)
Viral mRNA Translation
Gene Ratio
D
TRP63
KI67
TRP63
KI67
KRT13
TRP63
KI67
E
Leiden cluster
n=43,126
UMAP
F
Patient ID
P1
P3
P4
P6
P7
P10
P14
P15
P17
P18
P19
P22
P23
P25
P31
P36
P40
P41
UMAP
G
Adenocarcinoma
NAPSA
NKX2-1
SFTPB
high
Expression
low
UMAP
I
KRT13
high
Expression
low
UMAP
K
Basal-like
Hillock-like
UMAP
H
Squamous
KRT5
TP63
PITX1
high
Expression
low
UMAP
J
Hillock
Basal
high
Expression
low
UMAP
L
Percent hillock-like
M
Human hillock-like
Immune System
Innate Immune System
Neutrophil Degranulation
Vesicle-mediated Transport
Membrane Trafficking
Formation of the Cornified Envelope
Keratinization
EPH-Ephrin Signaling
O-linked Glycosylation of Mucins
Metal Sequestration by Antimicrobial Proteins
Gene Ratio
-log(adj. p-value)
Gene Count

Figure 4. KRT13 marks a hillock-like state in squamous tumors. A) Cell type classification based on gene and gene signature score expression from cells in **3A**. **B-C)** Gene set enrichment analysis comparing SNL tumor cells in **3A** by assigned cell type. The ENRICHR gene set used was 'Reactome Pathways 2024'. **D)** Representative immunofluorescence (IF) co-staining for KRT13 (green), TRP63 (red), and KI67 (blue) in SNL squamous tumors. Representative of n = 4 SNL-CMV mice. Scale bars, 50 µm. **E)** UMAP of scRNA-seq from lung squamous tumor cells from 18 patients[68] labeled by Leiden cluster. **F)** UMAP from **4E** labeled by patient (P) sample ID (n = 18). **G-I)** scRNA-seq expression of adenocarcinoma (*NAPSA, NKX2-1, SFTPB*), squamous (*TP63, KRT5, PITX1*), and hillock markers (*KRT13*) in UMAP space from cells in **4E**. **J)** Human hillock score and human basal score (**Tables S1** and **S2**) applied to tumor cells as in **4E** with score projected in UMAP. **K)** Cell type classification based on hillock score expression from cells in **4E**. **L)** Percent of basal- and hillock-like cells by patient ID as in **4F**. **M)** Gene set enrichment analysis comparing human tumor cells in **4E** and **4K** by assigned cell type. The ENRICHR gene set used was 'Reactome Pathways 2024'.

See also Figures S3 and S4 and Tables S1-8.

Figure 5

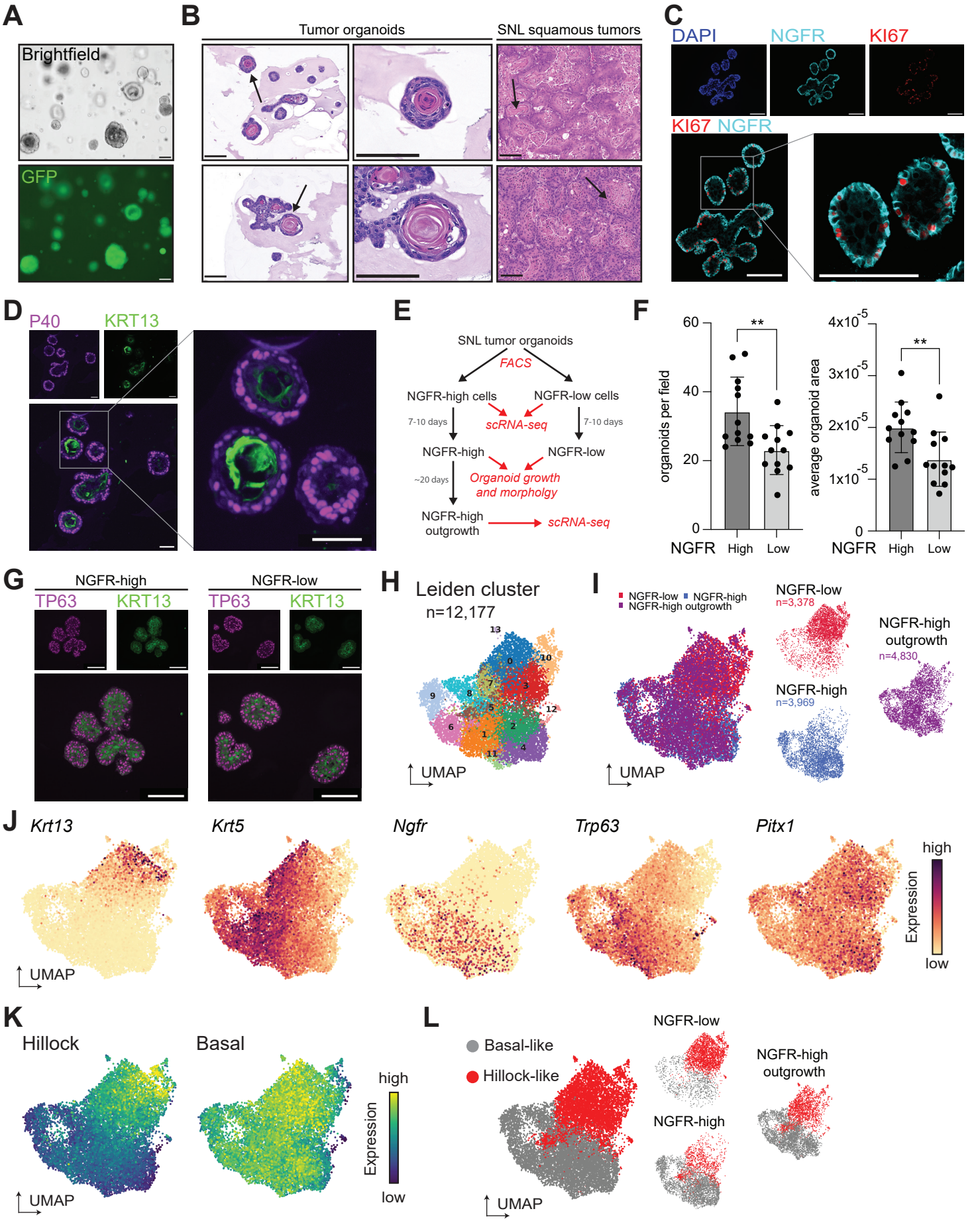
A
Brightfield
GFP
B
Tumor organoids
SNL squamous tumors
C
DAPI
NGFR
KI67
KI67 NGFR
D
P40
KRT13
E
SNL tumor organoids
FACS
NGFR-high cells
NGFR-low cells
7-10 days
scRNA-seq
7-10 days
NGFR-high
NGFR-low
~20 days
Organoid growth and morpholgy
NGFR-high outgrowth
scRNA-seq
F
organoids per field
average organoid area
**
**
NGFR High Low
NGFR High Low
G
NGFR-high
NGFR-low
TP63
KRT13
TP63
KRT13
H
Leiden cluster
n=12,177
UMAP
I
NGFR-low
NGFR-high
NGFR-high outgrowth
NGFR-low
n=3,378
NGFR-high
n=3,969
NGFR-high outgrowth
n=4,830
UMAP
J
Krt13
Krt5
Ngfr
Trp63
Pitx1
UMAP
Expression
high
low
K
Hillock
Basal
UMAP
Expression
high
low
L
Basal-like
Hillock-like
NGFR-low
NGFR-high
NGFR-high outgrowth
UMAP

Figure 5. NGFR$^+$ tumor propagating cells can regenerate KRT13$^+$ hillock cells. A) Representative brightfield and immunofluorescence images of tumor organoids established from GFP$^+$NGFR$^+$ SNL tumor cells. Scale bar, 100 µm. **B)** Representative H&E staining of SNL tumor organoids (n = 2 independent cultures) and SNL squamous tumors (n = 6 mice). Arrows indicate keratinized structures. Scale bars, 100 µm. **C)** Representative immunofluorescence (IF) co-staining for KI67 (red) and NGFR (cyan) in SNL tumor organoids (n = 3). Scale bar, 100 µm. **D)** Representative IF co-staining for P40 (purple) and KRT13 (green) in SNL tumor organoids (n = 3). Scale bar, 50 µm. **E)** Experimental scheme for organoid sorting and scRNA-seq. **F)** Number of organoids and average organoid area after SNL tumor organoid cells were sorted by NGFR status and replated in Matrigel and allowed to grow for 10 days. Each point represents an image taken at 4x magnification, and error bars represent mean +/- SD. Two-tailed t-test, **p ≤ 0.01. **G**) Representative IF co-staining for TP63 (purple) and KRT13 (green) from SNL tumor organoids allowed to grow for 10 days after sorting. Representative of n = 2 experiment with 6 different 20x fields per experiment. **H)** Leiden clustering of scRNA-seq performed on SNL tumor organoids as in **5I**. **I)** scRNA-seq of SNL tumor organoids at time of NGFR sorting (NGFR-high and NGFR-low) and NGFR-high cells after four weeks of growth in organoid conditions ("NGFR-high outgrowth") in **5H**. **J)** Expression of indicated hillock (*Krt13*) or basal (*Ngfr*, *Krt5*, *Trp63, Pitx1*) genes in UMAP space from cells in **5H**. **K)** scRNA-seq expression of the hillock score and basal score projected on UMAP space from cells in **5H. L)** Cell type classification by Leiden cluster based on gene and gene signature score expression from cells in **5H**. (see **Methods** and **Table S1**).
See also Figure S5.

Figure 6

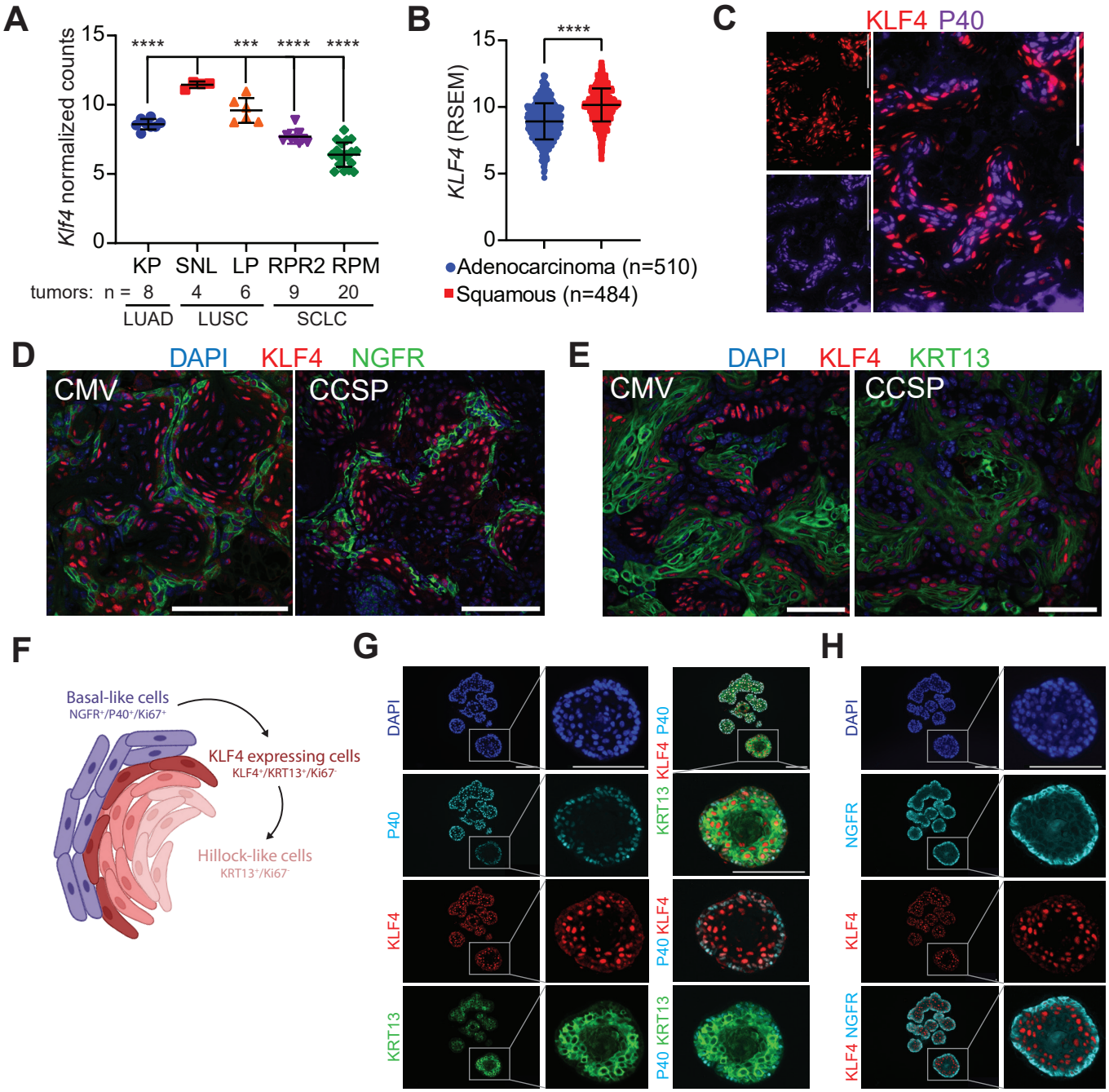
A
Klf4 normalized counts
15
10
5
0
****
***
****
****
KP
SNL
LP
RPR2
RPM
tumors: n = 8 4 6 9 20
LUAD
LUSC
SCLC
B
KLF4 (RSEM)
15
10
5
0
****
Adenocarcinoma (n=510)
Squamous (n=484)
C
KLF4 P40
D
DAPI KLF4 NGFR
CMV
CCSP
E
DAPI KLF4 KRT13
CMV
CCSP
F
Basal-like cells
NGFR+/P40+/Ki67+
KLF4 expressing cells
KLF4+/KRT13+/Ki67-
Hillock-like cells
KRT13+/Ki67-
G
DAPI
P40
KLF4
KRT13
KLF4 P40
KRT13
P40 KLF4
P40 KRT13
H
DAPI
NGFR
KLF4
KLF4 NGFR

Figure 6. KLF4 regulates the hillock-like state. A) Log2 normalized counts of *Klf4* from RNA-seq of lung tumors from indicated GEMMs with number of tumors indicated in the figure. Statistical significance determined by one-way ANOVA. **B)** RSEM expression of *KLF4* from bulk human RNA-seq data of lung adenocarcinoma and squamous cell carcinoma from TCGA with number of samples indicated in the figure. Two-tailed t-test. **C)** Representative immunofluorescence (IF) co-staining for KLF4 (red) and dNP63/P40 (purple) in squamous tumors from CMV-initiated SNL tumors. Representative of n = 2 mice. Scale bar, 100 µm. **D)** IF staining for NGFR (green) and KLF4 (red) in squamous tumors from CCSP- and CMV-initiated SNL tumors (white labels). Representative of n = 3 mice for CCSP-Cre and n= 2 mice for CMV-Cre. Scale bars, 100 µm. **E)** Representative IF co-staining for KRT13 (green) and KLF4 (red) in squamous tumors from CCSP- and CMV-initiated SNL tumors. Representative of n = 2 mice for CCSP-Cre and n= 3 mice for CMV-Cre. Scale bars, 50 µm. **F)** Schematic of a squamous tumor with basal-like cells, KLF4-expressing suprabasal cells, and KRT13-expressing hillock-like cells. Made with Biorender. **G)** Representative IF co-staining for KRT13 (green), KLF4 (red), and dNP63/P40 (cyan) in SNL tumor organoids. Scale bars, 100 µm. Representative of n = 3 experiments at with least 5 different 20x fields per experiment. **H)** Representative IF co-staining for KLF4 (red) and NGFR (cyan) in SNL tumor organoids. Scale bars, 100 µm. Representative of n = 2 experiments at with least 4 different 20x fields per experiment. **For all panels**, each data point represents an individual tumor and error bars represent mean +/- SD; ***p ≤ 0.001; ****p ≤ 0.0001. DAPI (blue) labels nuclei.
See also Figure S6.

Figure 7

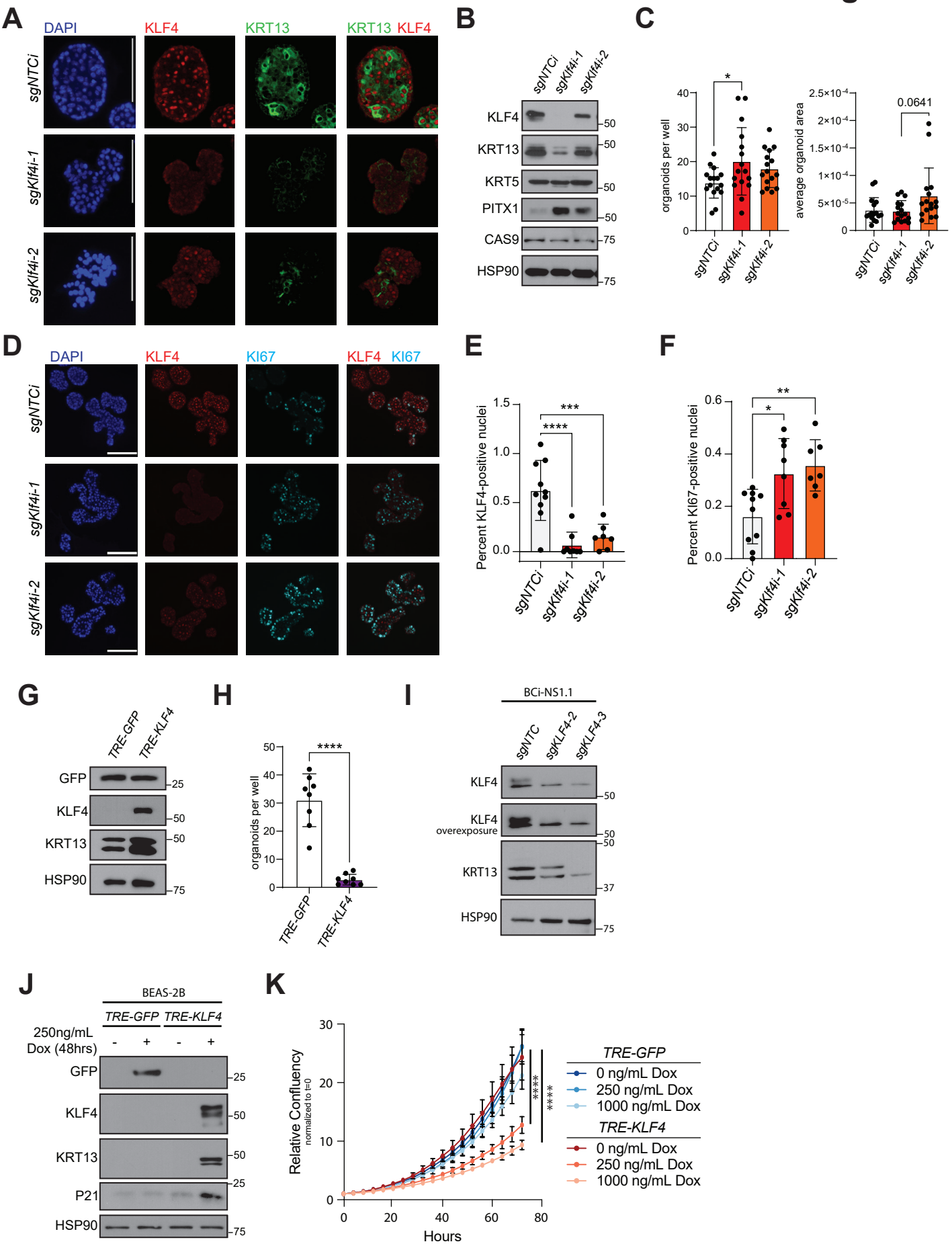
A
DAPI
KLF4
KRT13
KRT13 KLF4
sgNTCi
sgKlf4i-1
sgKlf4i-2
B
sgNTCi
sgKlf4i-1
sgKlf4i-2
KLF4
KRT13
KRT5
PITX1
CAS9
HSP90
50
50
50
50
75
75
C
organoids per well
*
average organoid area
0.0641
sgNTCi
sgKlf4i-1
sgKlf4i-2
D
DAPI
KLF4
KI67
KLF4 KI67
sgNTCi
sgKlf4i-1
sgKlf4i-2
E
Percent KLF4-positive nuclei
***
****
sgNTCi
sgKlf4i-1
sgKlf4i-2
F
Percent KI67-positive nuclei
**
*
sgNTCi
sgKlf4i-1
sgKlf4i-2
G
TRE-GFP
TRE-KLF4
GFP
KLF4
KRT13
HSP90
25
50
50
75
H
organoids per well
****
TRE-GFP
TRE-KLF4
I
BCi-NS1.1
sgNTC
sgKLF4-2
sgKLF4-3
KLF4
KLF4
overexposure
KRT13
HSP90
50
50
50
37
75
J
BEAS-2B
TRE-GFP
TRE-KLF4
250ng/mL
Dox (48hrs)
-
+
-
+
GFP
KLF4
KRT13
P21
HSP90
25
50
50
25
75
K
Relative Confluency
normalized to t=0
Hours
****
****
TRE-GFP
0 ng/mL Dox
250 ng/mL Dox
1000 ng/mL Dox
TRE-KLF4
0 ng/mL Dox
250 ng/mL Dox
1000 ng/mL Dox

Figure 7. KLF4 is necessary and sufficient to induce a KRT13^{+} hillock-like state.
A) Representative IF co-staining for KRT13 (green) and KLF4 (red) in SNL tumor organoids with KLF4 knockdown using CRISPRi. Scale bars, 100 µm. Representative of n = 2 experiments at least 4 different 20x fields per experiment. **B)** Immunoblot for indicated proteins from SNL tumor organoids after 10 days of growth following dissociation to single cells and replating in Matrigel. **C)** Number and average size of organoids after 7 days of growth following dissociation to single cells and replating in Matrigel. One-way ANOVA. **D)** Representative IF co-staining for KI67 (cyan) and KLF4 (red) in SNL tumor organoids with KLF4 knockdown using CRISPRi. Scale bars, 100 µm. Representative of n = 2 experiment with at least 7 different 20x fields per sample. **E-F)** Quantification of nuclei staining positive for KLF4 (**E**) or KI67 (**F**) from **7D**. One-way ANOVA. **G)** Immunoblot for indicated proteins from SNL organoids using the TRE-GFP or TRE-KLF4 vectors. Cells were grown for 10 days and 1µg/mL doxycycline was added every 48 hrs. **H)** Number and average size of SNL tumor organoids after 7 days of growth following dissociation to single cells and replating in Matrigel and treated with doxycycline as in **7G**. Two-tailed t-test. **I)** Immunoblot for indicated proteins in human BCi-NS1.1 cell lines expressing non-targeting control (sgNTC) or sgKLF4. **J)** Immunoblot for indicated proteins in human BEAS-2B cell lines with TRE-GFP or -KLF4 overexpression constructs with (+) or without (-) doxycycline exposure for 48 hrs. **K)** Cell growth determined by confluency using an Incucyte live-cell imaging system (Sartorius) normalized to initial plating density of BEAS-2B cells grown with varying doses of doxycycline (Dox). One-way ANOVA. **For all panels with bar graphs**, each data point represents a 4x (organoid growth)/20x (IF quantification) microscopy field of view, and error bars represent mean +/- SD; *$p \leq 0.05$; **$p \leq 0.01$;***$p \leq 0.001$; ****$p \leq 0.0001$. **For all other panels,** DAPI (blue) marks nuclei. HSP90 serves as loading control.
See also Figure S7.

Figure 8

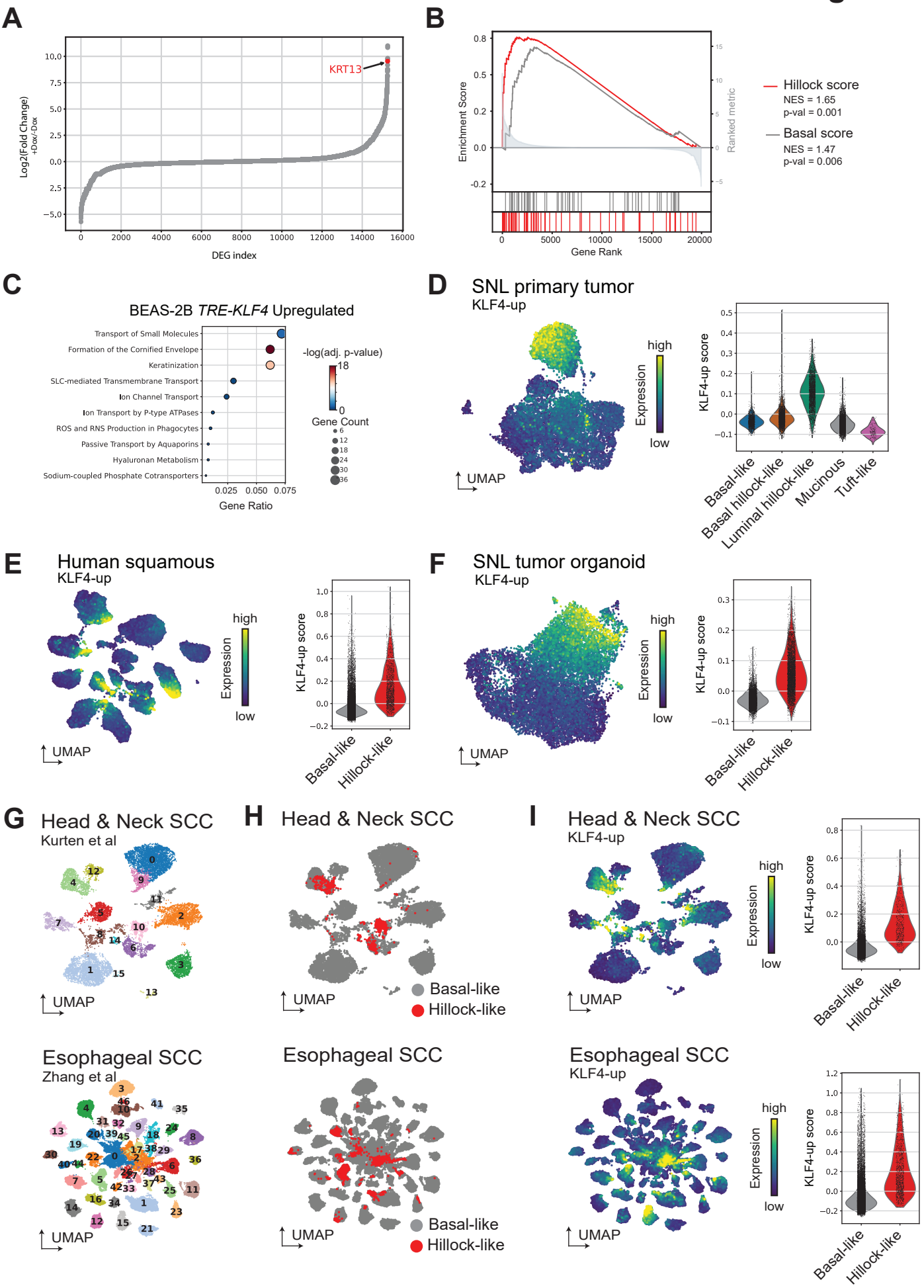
A
KRT13
Log2(Fold Change) +Dox/-Dox
DEG index
B
Enrichment Score
Ranked metric
Gene Rank
Hillock score
NES = 1.65
p-val = 0.001
Basal score
NES = 1.47
p-val = 0.006
C
BEAS-2B TRE-KLF4 Upregulated
Transport of Small Molecules
Formation of the Cornified Envelope
Keratinization
SLC-mediated Transmembrane Transport
Ion Channel Transport
Ion Transport by P-type ATPases
ROS and RNS Production in Phagocytes
Passive Transport by Aquaporins
Hyaluronan Metabolism
Sodium-coupled Phosphate Cotransporters
Gene Ratio
-log(adj. p-value)
Gene Count
D
SNL primary tumor
KLF4-up
high
Expression
low
UMAP
KLF4-up score
Basal-like
Basal hillock-like
Luminal hillock-like
Mucinous
Tuft-like
E
Human squamous
KLF4-up
F
SNL tumor organoid
KLF4-up
G
Head & Neck SCC
Kurten et al
H
Head & Neck SCC
Basal-like
Hillock-like
I
Head & Neck SCC
KLF4-up
Esophageal SCC
Zhang et al
Esophageal SCC
Esophageal SCC
KLF4-up

Figure 8. KLF4 regulates a hillock-like gene expression signature.

A) Waterfall plot of differentially expressed genes from RNA-sequencing of TRE-KLF4 BEAS-2B cells treated +/- Dox as in **7J**. **B)** GSEA for the general hillock score and basal score derived from RNA-sequencing of TRE-KLF4 BEAS-2B cells treated +/- Dox as in **7J**. **C)** Gene set enrichment analysis of the top 100 upregulated genes by TRE-KLF4 in BEAS-2B cells treated +/- Dox as in **7J**. The ENRICHR gene set used was 'Reactome Pathways 2024'. **D-F)** KLF4-up score applied to SNL tumor cells (D), human squamous tumor cells (E), and organoid cells (F) projected onto UMAP space from **3A, 4E,** and **5H** with score expression violin plots by cell state assignment. (see **Methods** and **Tables S11**). **G)** UMAP of scRNA-seq from head & neck tumor cells from 15 patients[87] (Kurten et al) or from esophageal tumors from 60 patients[88] (Zhang et al) labeled by Leiden cluster. **H)** Cell type classification of head & neck SCC or esophageal SCC single cell data based on hillock score expression from cells in **8G**. **I)** UMAP of KLF4-up score applied to head & neck and esophageal SCC cells from **8G** with score expression in violin plots grouped by cell state assignment.

See also Figure S8.

Figure S1

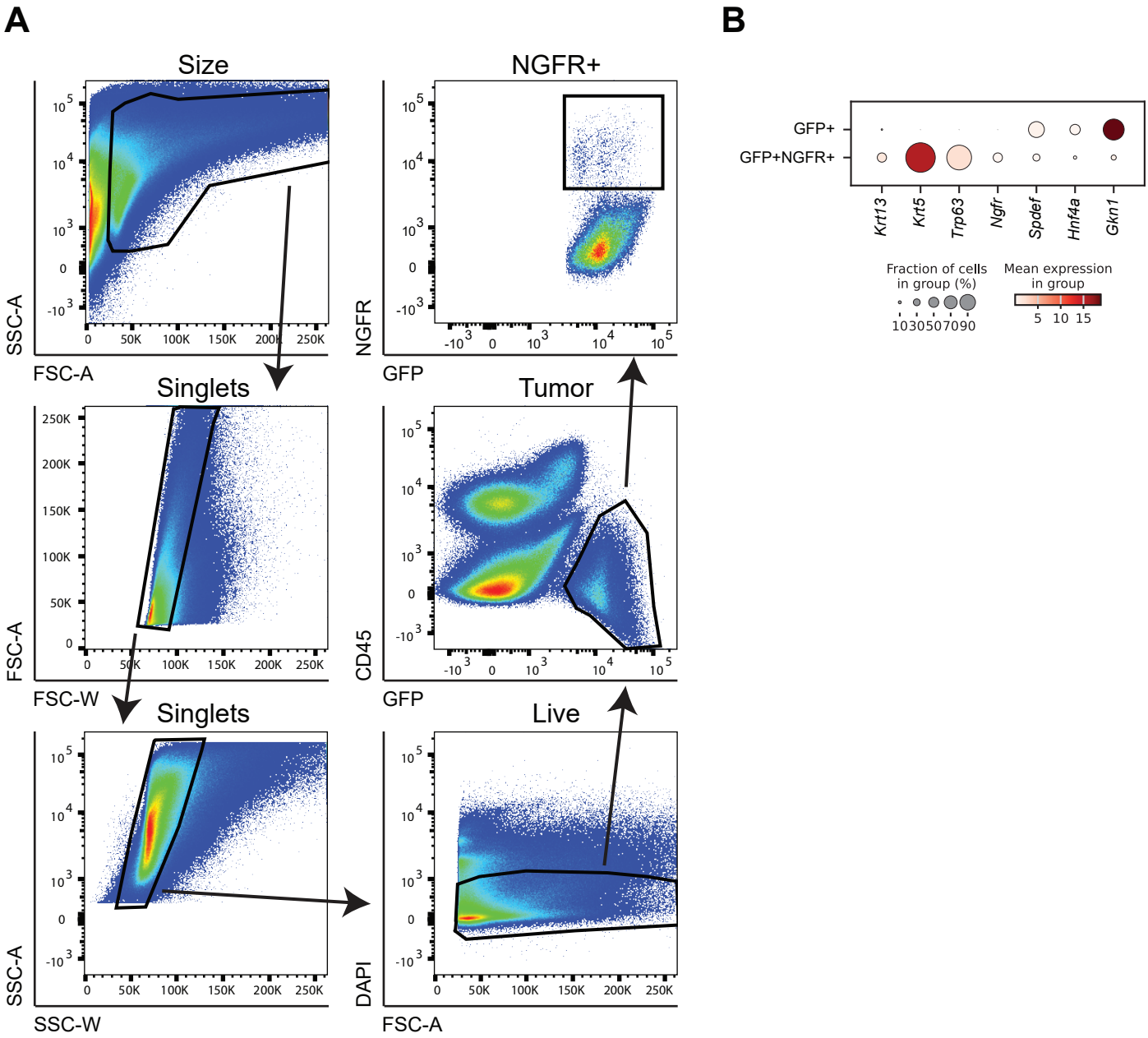
A
Size
SSC-A
FSC-A
NGFR+
NGFR
GFP
Singlets
FSC-A
FSC-W
Tumor
CD45
GFP
Singlets
SSC-A
SSC-W
Live
DAPI
FSC-A
B
GFP+
GFP+NGFR+
Krt13
Krt5
Trp63
Ngfr
Spdef
Hnf4a
Gkn1
Fraction of cells in group (%)
10 30 50 70 90
Mean expression in group
5 10 15

Figure S1. NGFR$^+$ tumor cells sorted from SNL tumors for scRNA-seq, related to Figure 2. A) Representative flow cytometry plots for gating strategy of DAPI$^-$CD45$^-$GFP$^+$NGFR$^+$ SNL tumor cells. SSC: side scatter, FSC: forward scatter, A: Area, W: Width. Arrows in upper left graph flow down, to the right, and up back to the upper right panel, indicating progressive gating during the flow sort. **B)** Scanpy dot plot of scRNA-seq gene expression from cells in **2A** by sorting identity. Dot size indicates percent of cells expressing each transcript and the color indicates the average expression per cluster indicated in the figure.

Figure S2

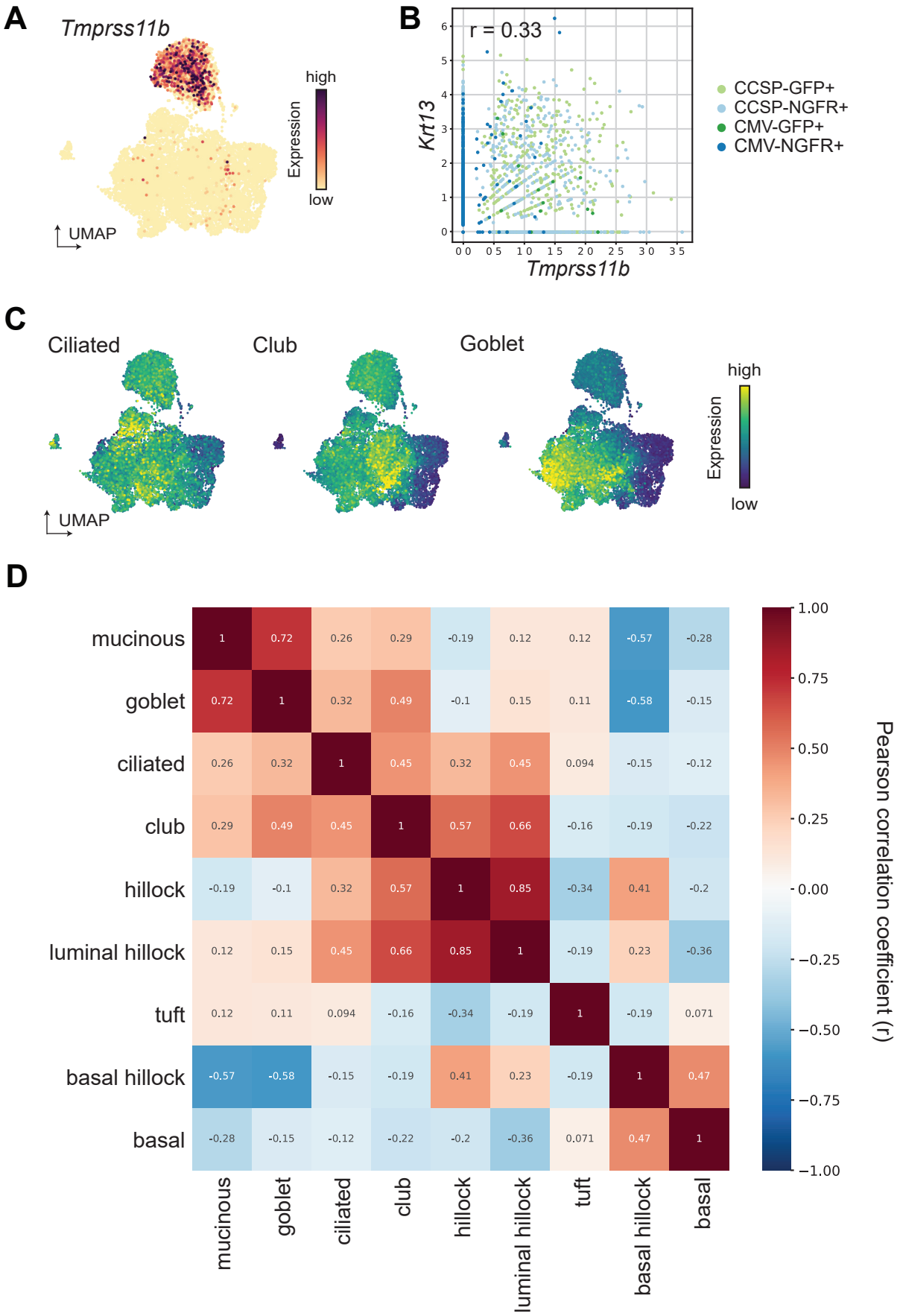
A
Tmprss11b
high
Expression
low
UMAP
B
r = 0.33
Krt13
Tmprss11b
CCSP-GFP+
CCSP-NGFR+
CMV-GFP+
CMV-NGFR+
C
Ciliated
Club
Goblet
high
Expression
low
UMAP
D
mucinous
goblet
ciliated
club
hillock
luminal hillock
tuft
basal hillock
basal
Pearson correlation coefficient (r)

Figure S2. *Krt13* and *Tmprss11b* correlate with hillock score, related to Figure 3. A) scRNA-seq expression of *Tmprss11b* projected on UMAP space from cells in **3A**. **B)** Correlation between Krt13 and Tmprss11b expression in cells from 3A. **C)** Gene scores applied to SNL tumor cells projected onto UMAP space from **3A** (see **Methods** and **Tables S2**). **D)** Correlation analysis between gene scores in SNL tumor single cell data in **3A** with the Pearson correlation coefficient shown (r) shown.

Figure S3

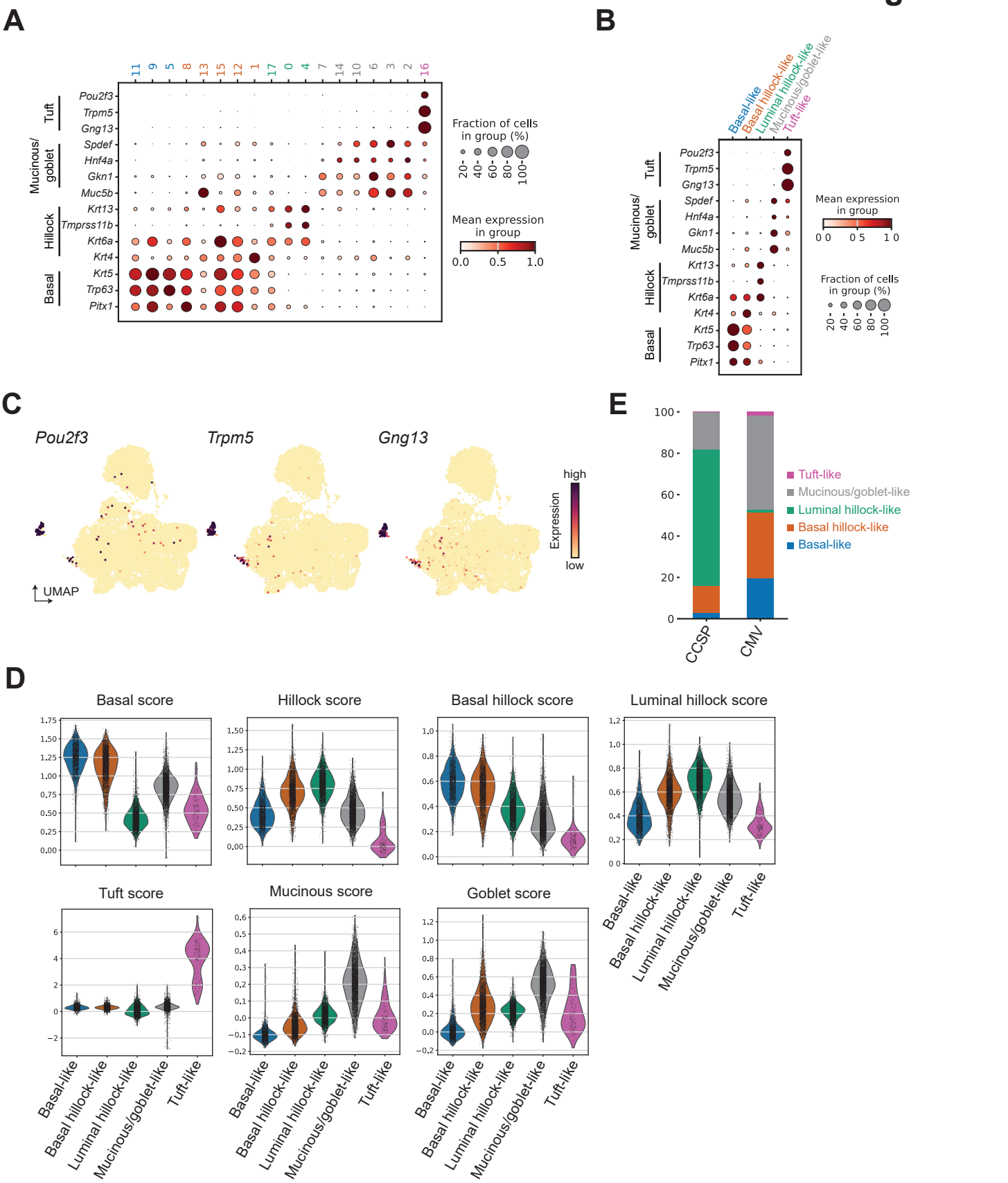
A
11 9 5 8 13 15 12 1 17 0 4 7 14 10 6 3 2 16
Tuft
Mucinous/ goblet
Hillock
Basal
Pou2f3
Trpm5
Gng13
Spdef
Hnf4a
Gkn1
Muc5b
Krt13
Tmprss11b
Krt6a
Krt4
Krt5
Trp63
Pitx1
Fraction of cells in group (%)
20- 40- 60- 80- 100-
Mean expression in group
0.0 0.5 1.0
B
Basal-like
Basal hillock-like
Luminal hillock-like
Mucinous/goblet-like
Tuft-like
Mean expression in group
0.0 0.5 1.0
Fraction of cells in group (%)
20- 40- 60- 80- 100-
C
Pou2f3
Trpm5
Gng13
high
Expression
low
UMAP
E
100
80
60
40
20
0
Tuft-like
Mucinous/goblet-like
Luminal hillock-like
Basal hillock-like
Basal-like
CCSP
CMV
D
Basal score
Hillock score
Basal hillock score
Luminal hillock score
Tuft score
Mucinous score
Goblet score
Basal-like
Basal hillock-like
Luminal hillock-like
Mucinous/goblet-like
Tuft-like

Figure S3. Defining cell states by transcriptional profiles in SNL tumors, **related to Figure 4. A)** Scanpy dot plot of scRNA-seq gene expression by Leiden cluster. Dot size indicates percent of cells expressing each transcript and the color indicates the average expression per cluster indicated in figure. Text coloring indicates Leiden cluster grouping for cell type assignment. **B)** Scanpy dot plot of scRNA-seq gene expression by assigned cell type. Dot size indicates percent of cells expressing each transcript and the color indicates the average expression per cluster indicated in figure. **C)** scRNA-seq expression of tuft marker genes projected on UMAP space from cells in **3A**. **D)** Violin plots of gene score expression by cell type assignment. **E)** Stacked bar plot showing percent of cells occupying each cell state assignment by sorted group from **3A**.

Figure S4

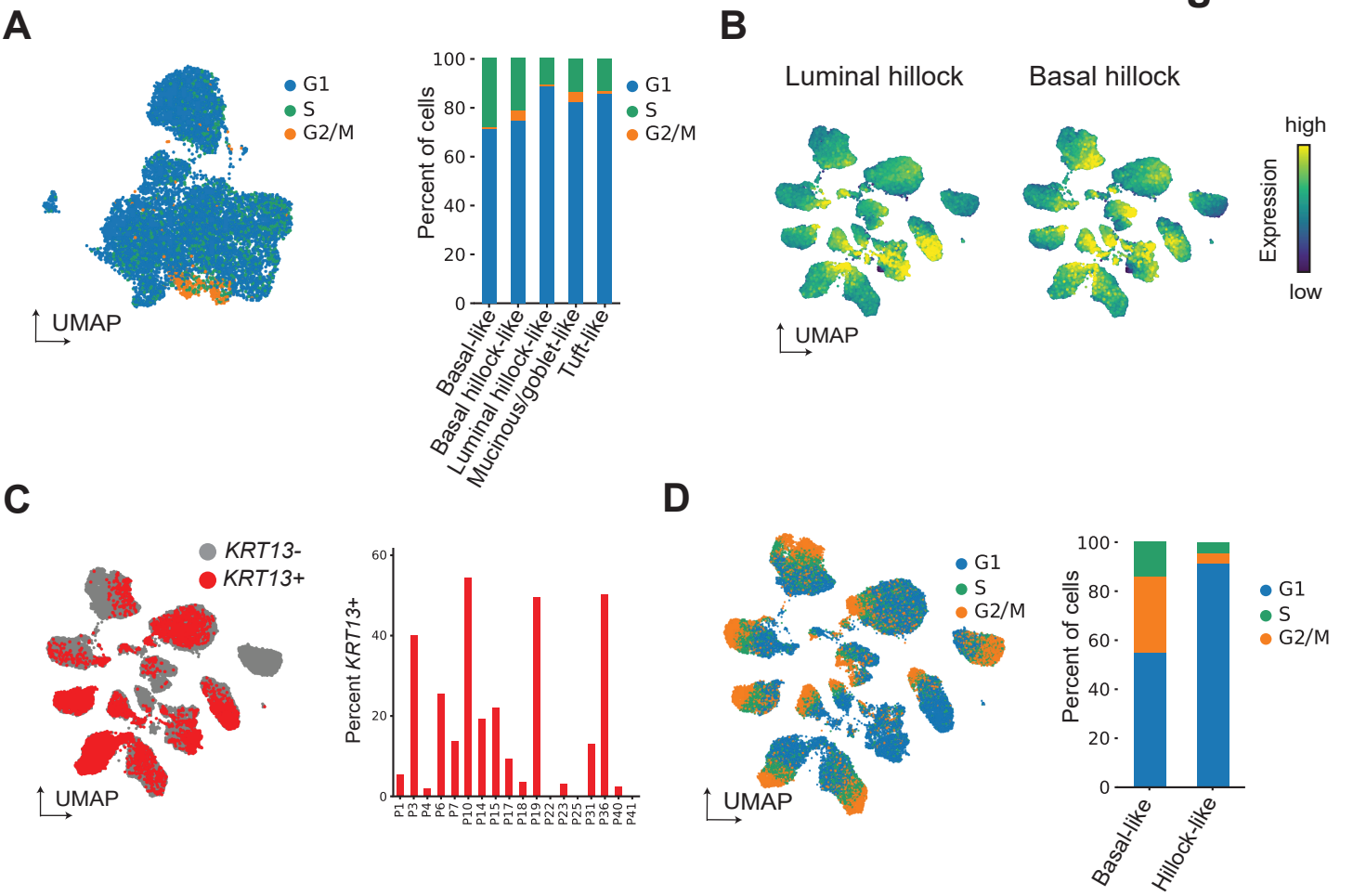
A
G1
S
G2/M
UMAP
Percent of cells
100
80
60
40
20
0
Basal-like
Basal hillock-like
Luminal hillock-like
Mucinous/goblet-like
Tuft-like
B
Luminal hillock
Basal hillock
high
Expression
low
UMAP
C
KRT13-
KRT13+
UMAP
Percent KRT13+
60
40
20
0
P1
P3
P4
P6
P7
P10
P14
P15
P17
P18
P19
P22
P23
P25
P31
P36
P40
P41
D
G1
S
G2/M
UMAP
Percent of cells
100
80
60
40
20
0
Basal-like
Hillock-like

Figure S4. Hillock state in squamous tumors is relatively non-proliferative, **related to Figure 4. A)** Cell cycle assignments applied to cells projected onto UMAP space from **3A** (see **Methods**) and stacked bar plot showing percent of cells occupying each cell cycle phase by cell state assignment as in **3A**. **B)** Luminal hillock and basal hillock scores applied to cells projected onto UMAP space from **4E** (see **Methods** and **Tables S4** and **S5**). **C)** UMAP showing assignment of *KRT13* state from **4E** and *KRT13* assignment per patient tumor (n=18) as a percentage of all tumor cells per sample. **D)** Cell cycle assignments applied to cells projected onto UMAP space from **4E** (see **Methods**) and stacked bar plot showing percent of cells occupying each cell cycle phase by cell state assignment as in **4E**. **See also Table S7.**

Figure S5

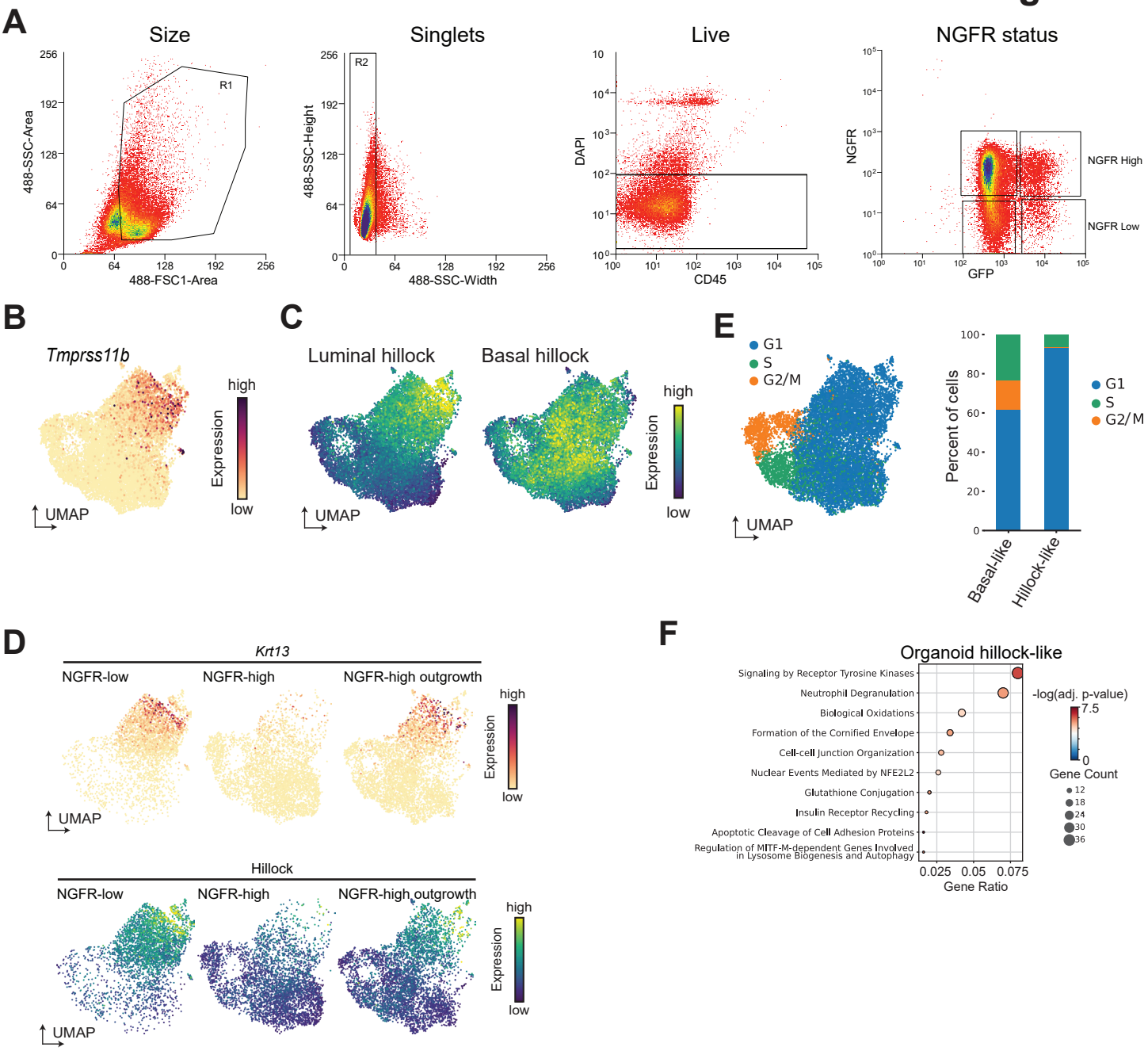
A
Size
Singlets
Live
NGFR status
R1
R2
488-SSC-Area
488-FSC1-Area
488-SSC-Height
488-SSC-Width
DAPI
CD45
NGFR
GFP
NGFR High
NGFR Low
B
Tmprss11b
high
Expression
low
UMAP
C
Luminal hillock
Basal hillock
high
Expression
low
UMAP
E
G1
S
G2/M
UMAP
Percent of cells
Basal-like
Hillock-like
D
Krt13
NGFR-low
NGFR-high
NGFR-high outgrowth
high
Expression
low
UMAP
Hillock
NGFR-low
NGFR-high
NGFR-high outgrowth
high
Expression
low
UMAP
F
Organoid hillock-like
Signaling by Receptor Tyrosine Kinases
Neutrophil Degranulation
Biological Oxidations
Formation of the Cornified Envelope
Cell-cell Junction Organization
Nuclear Events Mediated by NFE2L2
Glutathione Conjugation
Insulin Receptor Recycling
Apoptotic Cleavage of Cell Adhesion Proteins
Regulation of MITF-M-dependent Genes Involved in Lysosome Biogenesis and Autophagy
Gene Ratio
-log(adj. p-value)
Gene Count

Figure S5. NGFR$^+$ cells give rise to KRT13$^+$ hillock-like cells, related to Figure 5. A) Gating strategy for NGFR-high and low cells from dissociated SNL tumor organoids cells in **5F-L**. **B)** *Tmprss11b* expression projected onto UMAP space from **5H**. **C)** Luminal hillock and basal hillock scores applied to cells projected onto UMAP space from **5H** (see **Methods** and **Tables S4** and **S5**). **D)** Split UMAPs by sample showing *Krt13* and hillock score expression in cells from **5H**. **E)** Cell cycle assignments applied to cells projected onto UMAP space from **5H** (see **Methods**) and stacked bar plot showing percent of cells occupying each cell cycle phase by cell state assignment as in **5H**. **F)** Gene set enrichment analysis comparing organoid cells in **5L** by assigned cell type. The ENRICHR gene set used was 'Reactome Pathways 2024'.

Figure S6

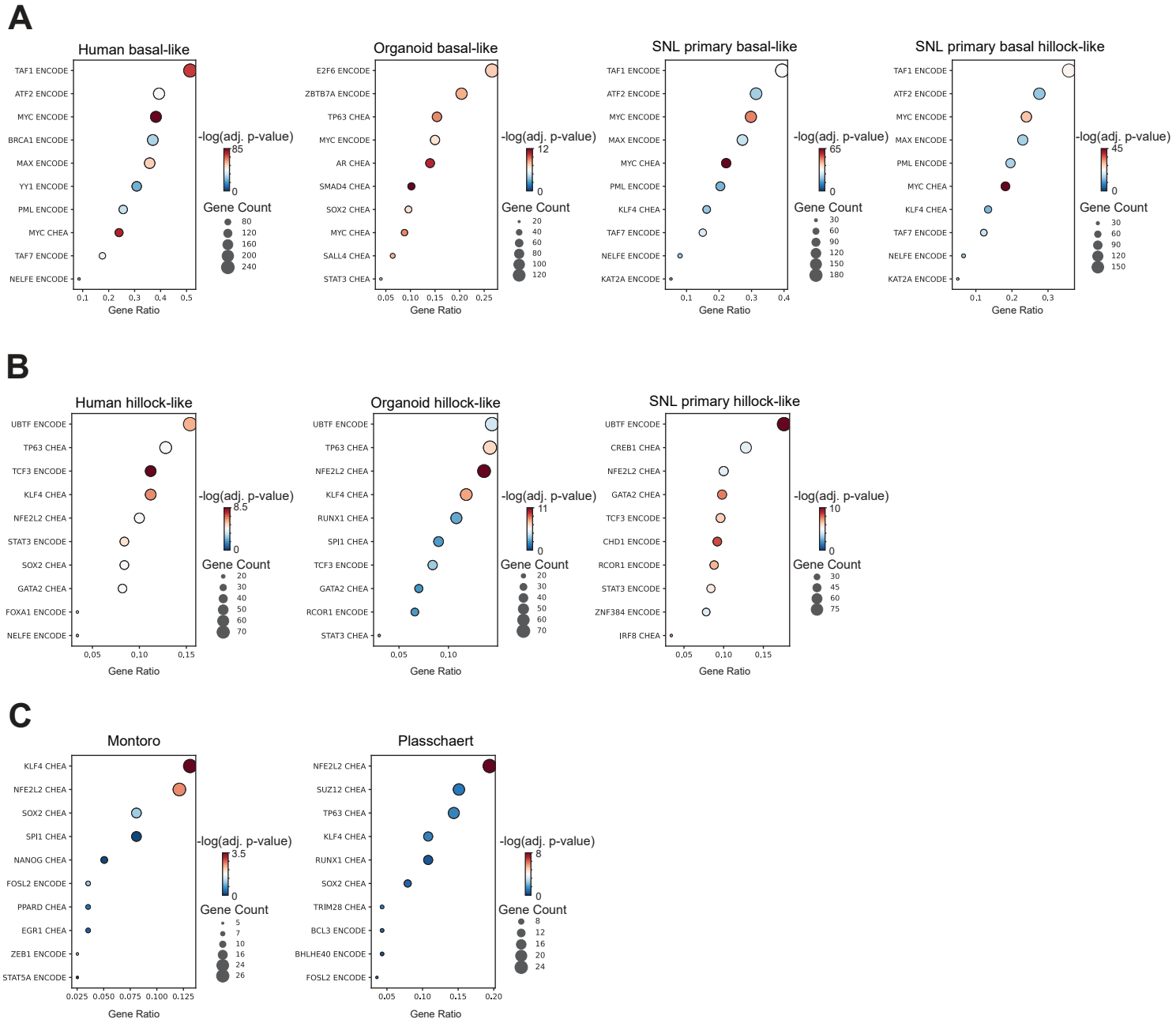
A
Human basal-like
TAF1 ENCODE
ATF2 ENCODE
MYC ENCODE
BRCA1 ENCODE
MAX ENCODE
YY1 ENCODE
PML ENCODE
MYC CHEA
TAF7 ENCODE
NELFE ENCODE
Gene Ratio
-log(adj. p-value)
Gene Count
Organoid basal-like
E2F6 ENCODE
ZBTB7A ENCODE
TP63 CHEA
MYC ENCODE
AR CHEA
SMAD4 CHEA
SOX2 CHEA
MYC CHEA
SALL4 CHEA
STAT3 CHEA
SNL primary basal-like
TAF1 ENCODE
ATF2 ENCODE
MYC ENCODE
MAX ENCODE
MYC CHEA
PML ENCODE
KLF4 CHEA
TAF7 ENCODE
NELFE ENCODE
KAT2A ENCODE
SNL primary basal hillock-like
TAF1 ENCODE
ATF2 ENCODE
MYC ENCODE
MAX ENCODE
PML ENCODE
MYC CHEA
KLF4 CHEA
TAF7 ENCODE
NELFE ENCODE
KAT2A ENCODE
B
Human hillock-like
UBTF ENCODE
TP63 CHEA
TCF3 ENCODE
KLF4 CHEA
NFE2L2 CHEA
STAT3 ENCODE
SOX2 CHEA
GATA2 CHEA
FOXA1 ENCODE
NELFE ENCODE
Organoid hillock-like
UBTF ENCODE
TP63 CHEA
NFE2L2 CHEA
KLF4 CHEA
RUNX1 CHEA
SPI1 CHEA
TCF3 ENCODE
GATA2 CHEA
RCOR1 ENCODE
STAT3 CHEA
SNL primary hillock-like
UBTF ENCODE
CREB1 CHEA
NFE2L2 CHEA
GATA2 CHEA
TCF3 ENCODE
CHD1 ENCODE
RCOR1 ENCODE
STAT3 ENCODE
ZNF384 ENCODE
IRF8 CHEA
C
Montoro
KLF4 CHEA
NFE2L2 CHEA
SOX2 CHEA
SPI1 CHEA
NANOG CHEA
FOSL2 ENCODE
PPARD CHEA
EGR1 CHEA
ZEB1 ENCODE
STAT5A ENCODE
Plasschaert
NFE2L2 CHEA
SUZ12 CHEA
TP63 CHEA
KLF4 CHEA
RUNX1 CHEA
SOX2 CHEA
TRIM28 CHEA
BCL3 ENCODE
BHLHE40 ENCODE
FOSL2 ENCODE

Figure S6. Identifying candidate regulators of the hillock cell state, related to Figure 6. A) ENRICHR analysis for predicted regulators of the basal-like state in human basal-like cells, organoid basal-like cells, SNL tumor basal-like and basal hillock-like cells using 'ENCODE and ChEA Consensus TFs from ChIP-X'. **B)** ENRICHR analysis for predicted regulators of the hillock-like state in human hillock-like cells, organoid hillock-like cells, and SNL tumor luminal hillock-like cells using 'ENCODE and ChEA Consensus TFs from ChIP-X'. **C)** ENRICHR analysis for predicted regulators of the hillock state in normal lung tissue from the indicated publications (Montoro et al, Nature, 2018; Plasschaert et al, Nature, 2018) using 'ENCODE and ChEA Consensus TFs from ChIP-X' (see **Methods - Differential gene expression analyses** and **Table S10**).

Figure S7

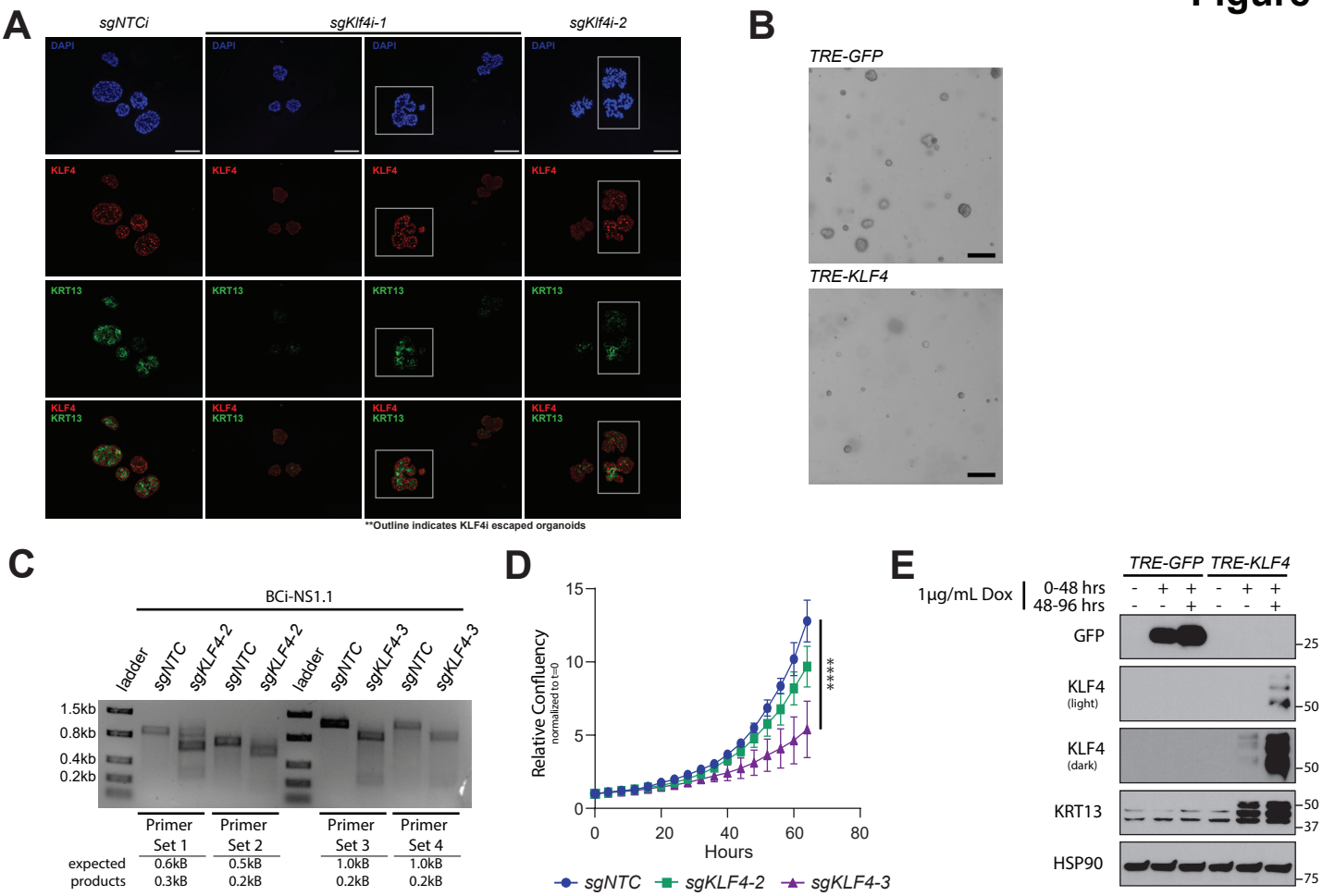
A
sgNTCi
sgKlf4i-1
sgKlf4i-2
DAPI
KLF4
KRT13
KLF4
KRT13
**Outline indicates KLF4i escaped organoids
B
TRE-GFP
TRE-KLF4
C
BCi-NS1.1
ladder
sgNTC
sgKLF4-2
sgNTC
sgKLF4-2
ladder
sgNTC
sgKLF4-3
sgNTC
sgKLF4-3
1.5kb
0.8kb
0.4kb
0.2kb
Primer Set 1
Primer Set 2
Primer Set 3
Primer Set 4
expected products
0.6kB
0.3kB
0.5kB
0.2kB
1.0kB
0.2kB
1.0kB
0.2kB
D
Relative Confluency
normalized to t=0
Hours
****
sgNTC
sgKLF4-2
sgKLF4-3
E
TRE-GFP
TRE-KLF4
1µg/mL Dox
0-48 hrs
48-96 hrs
GFP
KLF4 (light)
KLF4 (dark)
KRT13
HSP90
25
50
50
50
37
75

Figure S7. KLF4 regulates the KRT13^{+} hillock-like-state, related to Figure 7. A) Representative IF co-staining for KRT13 (green) and KLF4 (red) in SNL tumor organoids with KLF4 knockdown using CRISPRi. Scale bars, 100 µm. Images are uncropped coinciding to representative images shown in **7A**. White boxes indicate *sgKlf4i* expressing organoids with escaped KLF4 expression. DAPI (blue) marks nuclei. **B)** Representative brightfield images of SNL tumor organoids infected with TRE-GFP or -KLF4 treated with 1µg/mL doxycycline every 48 hours for 7 days. Scale bars, 200 µm. **C)** T7 surveyor assay for *KLF4* gene editing in human BCi-NS1.1 cells. Expected fragment sizes after T7 cutting of mismatched (edited) DNA are shown below. **D)** Cell growth determined by confluency normalized to initial plating density of BCi-NS1.1 cells. One-way ANOVA. **E)** Immunoblot for indicated proteins in BEAS-2B cells infected with TRE-GFP or KLF4 treated with (+) or without (-) doxycycline (Dox) for the indicated times. HSP90 serves as loading control.

Figure S8

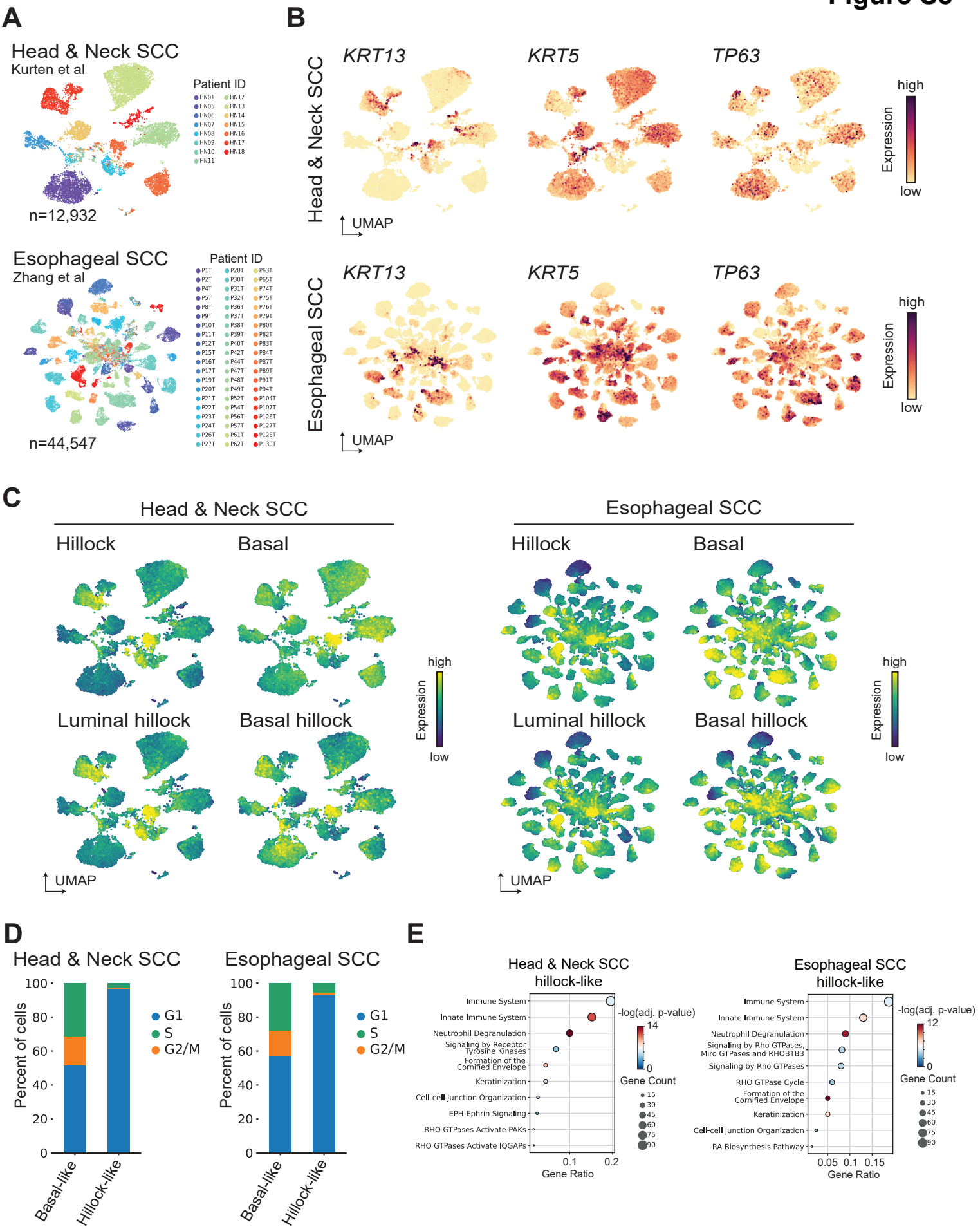
A
Head & Neck SCC
Kurten et al
Patient ID
HN01 HN05 HN06 HN07 HN08 HN09 HN10 HN11 HN12 HN13 HN14 HN15 HN16 HN17 HN18
n=12,932
Esophageal SCC
Zhang et al
Patient ID
P1T P2T P4T P5T P8T P9T P10T P11T P12T P15T P16T P17T P19T P20T P21T P22T P23T P24T P26T P27T P28T P30T P31T P32T P36T P37T P38T P39T P40T P42T P44T P47T P48T P49T P52T P54T P56T P57T P61T P62T P63T P65T P74T P75T P76T P79T P80T P82T P83T P84T P87T P89T P91T P94T P104T P107T P126T P127T P128T P130T
n=44,547
B
Head & Neck SCC
KRT13
KRT5
TP63
UMAP
Expression
high
low
Esophageal SCC
KRT13
KRT5
TP63
UMAP
Expression
high
low
C
Head & Neck SCC
Hillock
Basal
Luminal hillock
Basal hillock
UMAP
Expression
high
low
Esophageal SCC
Hillock
Basal
Luminal hillock
Basal hillock
UMAP
Expression
high
low
D
Head & Neck SCC
Percent of cells
100
80
60
40
20
0
G1
S
G2/M
Basal-like
Hillock-like
Esophageal SCC
Percent of cells
100
80
60
40
20
0
G1
S
G2/M
Basal-like
Hillock-like
E
Head & Neck SCC
hillock-like
Immune System
Innate Immune System
Neutrophil Degranulation
Signaling by Receptor Tyrosine Kinases
Formation of the Cornified Envelope
Keratinization
Cell-cell Junction Organization
EPH-Ephrin Signaling
RHO GTPases Activate PAKs
RHO GTPases Activate IQGAPs
0.1
0.2
Gene Ratio
-log(adj. p-value)
14
0
Gene Count
15
30
45
60
75
90
Esophageal SCC
hillock-like
Immune System
Innate Immune System
Neutrophil Degranulation
Signaling by Rho GTPases, Miro GTPases and RHOBTB3
Signaling by Rho GTPases
RHO GTPase Cycle
Formation of the Cornified Envelope
Keratinization
Cell-cell Junction Organization
RA Biosynthesis Pathway
0.05
0.1
0.15
Gene Ratio
-log(adj. p-value)
12
0
Gene Count
15
30
45
60
75
90

Figure S8. Hillock-like cells are present in squamous tumors of the esophagus and head & neck, related to Figure 8. A) UMAP of scRNA-seq from human head & neck tumor cells from 15 patients[87] or from esophageal tumors from 60 patients[88] labeled by patient (HN, or P) sample ID. **B)** scRNA-seq expression of squamous (*TP63, KRT5)* or hillock markers (*KRT13*) in UMAP space from cells in **8G**. **C)** Human hillock score, basal score, luminal hillock score, and basal hillock score (**Tables S1-4**) applied to tumor cells as in **S8A** with score projected in UMAP. **D)** Cell cycle assignments applied to cells from **8G** (see **Methods**) shown as a stacked bar plot with percent of cells occupying each cell cycle phase by cell state assignment as in **8H**. **E)** Gene set enrichment analysis comparing human tumor cells in **8H** by assigned cell type. The ENRICHR gene set used was 'Reactome Pathways 2024'.