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

Authors

Luke T. Izzo[1], Tony Reyes[2], Srijan Meesala[1], Abbie S. Ireland[1], Lisa Earnest-Noble[1], Steven Yang[1], Hari Shankar Sunil[3], Xiao Chun Cheng[1], Nomi Tserentsoodol[1], Sarah B. Hawgood[1], Carolyn Glass[4], Edward F. Patz, Jr.[1,5], Benjamin L. Witt[6], Darren R. Tyson[1], Kathryn A. O'Donnell[3], Trudy G. Oliver[1,2,7*]

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 Pathology, Duke University, Durham, NC, 27710, USA

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

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

7. Lead contact: Trudy G. Oliver

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

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

Abstract

Lung squamous cell carcinoma (LUSC) is a basal-like subtype of lung cancer with limited treatment options. While prior studies have identified tumour-propagating cell states in squamous tumours, the broader landscape of intra-tumoural 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 tumour cells characterized by immunomodulatory gene expression signatures. The tumour hillock-like state is conserved across multiple animal and human-derived models and is present in the majority of human LUSCs as well as head and neck and esophageal squamous tumours. Our findings shed light on the cellular origins of tumour hillock-like states: lung club cells give rise to tumours with luminal hillock-like populations, while basal-like tumour-propagating cells transition into basal hillock-like states, resembling lineage plasticity trajectories of the normal lung. Mechanistically, KLF4 promotes KRT13, a broadly conserved hillock-like state with enrichment of potential therapeutic targets, and resistance to platinum-based chemotherapy. Together, these results

provide molecular insights into the lineage plasticity underlying intra-tumoural heterogeneity within LUSC, offering potential avenues for new therapeutic strategies.

Main

Understanding the cellular complexity of tumours is crucial for improving cancer treatment. Tumours 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-tumoural heterogeneity and the ability of tumour cells to switch between phenotypic states, referred to as plasticity, has recently been highlighted as a hallmark of cancer and a major barrier to cancer treatment[15-17].

Previous studies investigating intra-tumoural heterogeneity in LUSC identified basal-like tumour-propagating cells (TPCs) that can serially transplant the disease[18-23]. The ability of tumour cell states to mimic normal cell fate lineages (so called “lineage plasticity”) 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[30-32], are abundant in the major airways where LUSC 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 LUSC[33,34]. Importantly, squamous tumours 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 normal basal cells have the capacity to give rise to multiple differentiated cell types in the adult airway, including neuroendocrine, tuft, ionocyte, club, ciliated, and newly-identified hillock cells[6,7,35]. Hillock cells were first reported in single cell studies that suggested basal cells can pass through a hillock state before progressing toward a club fate[6]. Normal hillock cells have been identified in multiple tissues including cartilaginous airways of the lung and in the prostate[36], 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 basal hillock cells and protective luminal hillock cells[37]. Basal hillock cells

have a remarkable ability to repopulate the tracheal epithelium after injury. Luminal hillock cells are morphologically and transcriptionally similar to differentiated squamous cells and maintain basal-hillock cell integrity through resistance to a myriad of injuries[37]. The function of normal hillock cells is only beginning to be understood, and whether they have a transcriptional or functional counterpart in cancer remains completely unexplored to our knowledge. Understanding whether hillock-like cells exist in squamous tumours could provide new insights into the transcriptional plasticity driving LUSC progression and resistance to therapy.

Multiple genetically-engineered mouse models (GEMMs) of LUSC have been developed[38-43]. 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 LUSC[30,38]. 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 tumours arise via transdifferentiation[40,44-48]. Importantly, mouse squamous tumours that arise via transdifferentiation highly resemble their human tumour counterparts both molecularly and histologically[18,30,40,49-51]. Due to this transcriptional and histopathological resemblance, mouse models provide useful tools for interrogating intra-tumoural heterogeneity and the molecular mechanisms that regulate cell state in squamous tumours.

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[52-54]. In both the skin and esophageal epithelium, KLF4 is restricted to the suprabasal layer and is necessary for differentiation of specific cell types[55-58]. Apart from its role in normal development, KLF4 has been implicated in multiple cancers—able to adopt either an oncogenic or a tumour suppressive role in a context-dependent manner[52]. Despite KLF4's link to normal squamous differentiation, the function of KLF4 in LUSC 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 tumour heterogeneity, paving the way for novel therapeutic strategies targeting transcriptional plasticity in lung squamous cell carcinoma.

Results

KRT13 is highly enriched in lung squamous cell carcinoma

Recent single cell studies have identified novel lung cell populations[6,7,37], 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 of adenocarcinoma from the lungs of *Kras*$^{G12D/+}$;*Trp53*$^{fl/fl}$ (KP) mice, squamous tumours 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[18,40,49,59-61]. We observed abundant *Krt13* expression restricted specifically to the squamous GEMMs (**Fig. 1a**). To test whether this association was consistent at the protein level, we performed immunohistochemistry (IHC) on tumour-bearing lung tissue from the lung cancer GEMMs. KRT13 expression was highly enriched in squamous but not in adenocarcinoma or SCLC tumours (**Fig. 1b,c**). To determine whether this association was

consistent in human squamous tumours, we examined RNA-seq data derived from human lung adenocarcinoma and squamous tumours from The Cancer Genome Atlas (TCGA). Compared to lung adenocarcinoma, we observed significantly greater *KRT13* expression in human squamous tumours (**Fig. 1d**). Tissue microarrays of human lung cancer (n = 57 total samples, n = 41 containing tumour tissue) stained for KRT13 also had significantly increased expression in human squamous compared to adenocarcinoma samples, with the majority of squamous patient biopsies displaying KRT13 positive cells (16 of 20) (**Fig. 1e,f**). Notably, both mouse and human squamous tumours exhibited intra-tumoural heterogeneity in KRT13 expression (**Fig. 1b,e**). Overall, these data suggest KRT13$^+$ cells are highly enriched in subpopulations of mouse and human squamous tumours.

KRT13 is expressed in a non-basal subpopulation of LUSC

Previous studies have shown functional heterogeneity in LUSC, including a TPC population marked by NGFR$^+$SCA1$^+$ expression with the capacity to serially transplant tumours[18]. Squamous tumours in most GEM models have been shown to arise via transdifferentiation from a mucinous-like state[9,40,44-46]. Due to potential transcriptional and functional heterogeneity, and the heterogeneous expression of KRT13 in squamous tumours (**Fig. 1b,e**), we sought to assess *Krt13* expression at the single cell level in squamous GEMMs via single cell RNA-sequencing (scRNA-seq). We flow-sorted CD45-negative (CD45$^-$) GFP-positive (GFP$^+$) tumour cells from Adenoviral-CMV-Cre-infected (Ad-CMV-Cre) SNL mice utilizing the GFP reporter in the *Sox2-Ires-Gfp* allele. Following sorting, CD45$^-$ GFP$^+$ tumour cells were subject to scRNA-seq (see **Methods**). We obtained 3,663 tumour 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 tumours by IHC. Rather, the majority of CD45$^-$GFP$^+$ tumour cells expressed markers of mucinous adenocarcinoma. To capture more basal-like squamous tumour cells, we flow-sorted Ad-CMV-Cre-infected SNL tumours and isolated CD45$^-$GFP$^+$NGFR$^+$ cells (**Extended Data Fig. 1a**), which are reported to include the TPC population. We obtained 4,363 additional tumour cells via scRNA-seq that were enriched for squamous marker genes (**Fig. 2a-c**). Consistent with previous studies[40,62], mucinous markers *Spdef, Hnf4a*, and *Gkn1* were low in tumour cells that expressed basal-associated squamous markers (**Fig. 2c**). *Krt13* expression was overall low, but present in tumour cell clusters expressing basal markers and absent in cells expressing mucinous adenocarcinoma markers (**Fig. 2d**). Consistently, IHC on serial sections from Ad-CMV-Cre-infected SNL tumours showed that KRT13 was enriched in KRT5$^+$ squamous regions and absent in HNF4A$^+$ mucinous regions (**Fig. 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 (**Fig. 2f**). Together, these data suggest that KRT13$^+$ cells are enriched in a subpopulation of non-mucinous, non-basal squamous tumour cells.

Club-initiated tumours have early and enriched KRT13 expression

Given that normal *Krt13*$^+$ lung cells can act as an intermediate state between basal and club cells[6], and club cells can regenerate basal cell populations[63], we reasoned that club-initiated tumour cells may have the capacity to dedifferentiate to a KRT13$^+$ state. Using adenoviral delivery of Cre controlled by a club cell-specific promoter (CCSP), we induced tumours from club cells in SNL

159 mice to determine whether this would alter the relative abundance or quality of *Krt13*$^{+}$ tumour cell
160 populations by scRNA-seq. Similar to our approach with Ad-CMV-Cre-initiated tumours, we first
161 sorted Adenoviral-CCSP-Cre-driven (Ad-CCSP-Cre) CD45^{-}GFP^{+} tumour cells (CCSP-GFP^{+},
162 "S1") and found a lack of basal-associated squamous marker gene expression. We subsequently
163 performed a second round of sequencing from Ad-CCSP-Cre infected SNL lungs, containing a
164 mixture of sorted CD45^{-}GFP^{+} (CCSP-GFP^{+}, "S2") and CD45^{-}GFP^{+}NGFR^{+} basal-enriched tumour
165 cells (see **Methods**). For subsequent scRNA-seq analysis, cells from both collections were
166 combined with cells derived from Ad-CMV-Cre-infected tumours. Following pre-processing, batch
167 correction, and removal of normal lung and immune cell populations (see **Methods**), we obtained
168 4,012 club cell-derived tumour cells that largely clustered distinctly from Ad-CMV-initiated
169 populations (**Fig. 3a,b**). Sorting based on NGFR expression enriched for cells expressing basal
170 markers genes (**Extended Data Fig. 2a**). We confirmed the presence of both mucinous
171 adenocarcinoma and squamous populations in Ad-CCSP-derived tumour cells (**Fig. 3b,c**),
172 demonstrating that club cells can act as a cell of origin for squamous tumours in the SNL model.
173 *Krt13*$^{+}$ cells were highly enriched in Ad-CCSP- versus Ad-CMV-initiated tumour cells, particularly
174 in clusters that lacked mucinous and basal squamous markers (**Fig. 3b-d**). We assessed
175 expression of another gene, *Tmprss11b*, that correlates with poor survival in NSCLC[64,65].
176 *Tmprss11b* is enriched in a similar population of cells as those expressing *Krt13* and correlated
177 with *Krt13* expression (**Extended Data Fig. 2b,c**). Interestingly, in contrast to Ad-CMV-driven
178 tumours, KRT13 was present in mucinous adenocarcinoma as well as in squamous
179 histopathologies in Ad-CCSP-initiated tumours (**Fig. 3e,f**). Given that prior studies have shown
180 that mucinous adenocarcinoma precedes squamous differentiation[40,46,47], these data suggest
181 KRT13 is enriched and may emerge earlier in club-initiated tumours compared to Ad-CMV-
182 initiated tumours.

183 **KRT13 marks a hillock-like state in squamous tumours**

184 To determine whether *Krt13*$^{+}$ cells exhibit a hillock-like phenotype in LUSC, we generated a
185 "hillock-cell signature" by identifying normal hillock-cell specific genes that were conserved across
186 two independent datasets[6,7] (see **Methods** and **Supplementary Table 1**). We applied the hillock-
187 cell signature to the mouse tumour scRNA-seq data. Tumour cells with a high hillock signature
188 strongly overlapped with *Krt13*-expressing cells and were present in both Ad-CMV and Ad-CCSP-
189 initiated tumours (**Fig. 3d,g**). Additionally, tumour cells with a high hillock score were largely
190 distinct from cells with the highest mucinous and basal cell signatures (**Fig. 3g**) (see **Methods,**
191 **Supplementary Tables 2** and **3**), suggesting that *Krt13* marks a transcriptionally distinct hillock-
192 like state in lung squamous cell carcinoma, as it does in the normal airway epithelium. To query
193 potential similarities between normal lung and airway cell lineages and tumour heterogeneity, we
194 applied additional normal cell type signatures to the scRNA-seq data. Consistent with known
195 relationships between goblet cell phenotypes and mucinous adenocarcinomas, these two
196 signatures largely overlap and cluster together by correlation analysis (**Fig. 3g** and **Extended**
197 **Data Fig. 2d, e**). The club cell signature overlaps most with the hillock signature and partially
198 overlaps with the basal signature (**Extended Data Fig. 2d,e**) in line with known lineage
199 relationships between normal club, hillock, and basal cells[6]. Next, we generated gene sets derived
200 from a recent study[37] to distinguish normal basal hillock and luminal hillock cells from other cancer
201 states (see **Methods**, **Supplementary Tables 4** and **5**). Applying these signatures to the mouse

202 tumour scRNA-seq dataset revealed that *Krt13*$^{+}$ hillock-like tumour cells overlapped most with the
203 luminal hillock score, while the basal hillock score was highest in the basal-like cells (**Fig. 3g** and
204 **Extended Data Fig. 2e**). Together, these data suggest that *Krt13* is enriched in hillock-like cells
205 that may mark both luminal or basal hillock-like populations, which is influenced by cell of origin.

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

208 Next, using marker gene expression and gene signature scores of normal lung and airway cell
209 types, we assigned cell states to Leiden clusters in the mouse tumour scRNA-seq dataset.
210 Tumour cells were assigned to either basal-, basal hillock-, luminal hillock-, or mucinous/goblet-
211 like states (**Fig. 3g-i** and **Extended Data Fig. 3a–d**). In addition, a small population of tumour
212 cells with a tuft cell gene expression program was identified that clustered separately from other
213 cell populations (**Fig. 3g-i**, **Extended Data Fig. 2e**, and **Extended Data Fig. 3a–c,** see
214 **Methods**)[6]. Recent reports have noted tuft-like tumour cells in or near squamous tumours[66-68].
215 Consistently, our lab recently discovered that basal cells can simultaneously give rise to tuft-like
216 SCLC and squamous tumours under specific genetic conditions, and combined SCLC in patients
217 can harbour both tuft-like SCLC and squamous regions[68,69], suggesting a transcriptional
218 relationship between tuft and basal states in lung cancer. These tuft-like cancer cells were
219 enriched in Ad-CMV- compared to Ad-CCSP-driven tumours, suggesting a potential influence of
220 cell of origin (**Extended Data Fig. 3e**).

221 Ad-CCSP-driven tumours contained the majority of luminal-hillock-like cells, whereas Ad-CMV-
222 driven tumours contained the majority of basal-hillock-like cells (**Extended Data Fig. 3e**). Notably,
223 luminal hillock-like tumour cells exhibit intermediate club, goblet, and mucinous signatures
224 compared to the mucinous and basal-like populations (**Fig. 3g-i** and **Extended Data Fig. 3d**).
225 Together, these data suggest that cell of origin impacts the presence of different hillock-like states
226 and that tumour lineage plasticity in SNL squamous tumours may mirror normal developmental
227 programs, i.e. transcriptional relationships between club, hillock, and basal cell states[6,37]
228 (**Extended Data Fig. 3f**).

229 To uncover transcriptional patterns associated with each cell state in an unbiased manner, we
230 identified top differentially-expressed genes (DEGs) in the assigned luminal-hillock, basal-hillock,
231 or basal-like clusters comparing each cell state to all other cells and performed ENRICHR
232 pathway analysis (**Supplementary Table 6**). Luminal hillock-like cells were enriched for
233 signatures of immune pathways, neutrophil degranulation, and cytokine signaling (**Fig. 4a**), with
234 genes driving the signatures including *B2m, Lgals3, and Il1a* (**Supplementary Table 7**). We also
235 assessed gene expression related to immunomodulation including interferon signaling, immune
236 checkpoint and antigen presentation (**Extended Data Fig. 4a**). Luminal hillock-like cells had
237 notable enrichment for interferon regulatory gene expression (*Irf1*, *Irf7*, and *Ifr9*) and immune
238 checkpoint genes including *Cd274* (encoding PD-L1), *Ceacam1*, and *Vtcn1*. Both basal and basal
239 hillock-like clusters were enriched for proliferation-related gene signatures such as translation and
240 metabolic processes (**Fig. 4b**), suggesting luminal-hillock-like cells may be less proliferative
241 compared to basal and basal-hillock-like tumour cells. Consistently, KI67 expression was
242 restricted to TRP63^{+}KRT13^{-} cells and nearly absent from KRT13^{+} regions within squamous

243 tumours by immunostaining (**Fig. 4c**). Additionally, cell cycle-specific gene signatures in scRNA-
244 seq data revealed that the luminal-hillock-cell state has a lower proportion of cells in S and G2/M
245 phases compared to basal and basal-hillock states, consistent with decreased proliferation
246 (**Extended Data Fig. 4b**). Basal-like and basal-hillock-like states have similar gene expression
247 compared to the remaining cells in our data (**Fig. 4b** and **Supplementary Table 6**). To determine
248 differences between the populations, we directly compared the basal-like and basal hillock-like
249 cell states. This analysis revealed that basal hillock-like cells have elevated expression of
250 immunomodulatory signatures and increased expression of tight junction genes compared to
251 basal-like cells, whereas basal-like cells are enriched for translation-related signatures (**Extended**
252 **Data Fig. 4c-d** and **Supplementary Table 6**). Altogether these data suggest that the KRT13^{+}
253 luminal-hillock-like state has relatively lower proliferation than basal-like populations and that the
254 basal-hillock-like population shares characteristics of the basal- and luminal-hillock-like state.

255 To determine if a *KRT13*$^{+}$ hillock-like state is also present within human squamous tumours, we
256 analyzed single cell data from 18 patient tumours (n = 43,126 cells)[70] (**Fig. 4d,e**), excluding normal
257 lung cell types such as immune, endothelial, and fibroblast cells (see **Methods**). We analyzed
258 tumour cells for expression of canonical markers of adenocarcinoma (*NAPSA*, *NKX2-1,* and
259 *SFTPB*) and squamous tumours (*KRT5, PITX1* and *TP63*) (**Fig. 4f,g**); as expected, squamous
260 markers were considerably higher than the rarely detectable adenocarcinoma markers.
261 Consistent with observations in SNL tumours, high *KRT13* expression in human tumour cells was
262 associated with an enriched hillock score (**Fig. 4h,i**). While the luminal hillock-score and basal
263 hillock-score largely overlap, the luminal hillock-score more closely mirrors the distribution of
264 *KRT13* expression (**Fig. 4h,i** and **Extended Data Fig. 4e**). Leiden clustering mainly separated
265 cells by patient, likely due to underlying inter-patient genetic and non-genetic heterogeneity, which
266 has been observed in similar studies comparing patient tumours[4,70-73]. However, this limits the
267 ability to identify cells as distinct cell types via Leiden clustering. To circumvent these limitations,
268 we assigned individual cells as the hillock-like state using cutoff criteria defining high hillock-score
269 expression and low basal-score expression (see **Methods**). Using this single-cell classification,
270 hillock-like cells were detected in 17 of 18 human patient samples ranging from ~1% to 20% of
271 cells per tumour (**Fig. 4j,k** and **Supplementary Table 8**). Similarly, *KRT13*-expressing cells were
272 detected in 16 of 18 human patient samples ranging from ~2% to 50% of cells per tumour
273 (**Extended Data Fig. 4f** and **Supplementary Table 8**). Together, regardless of classification
274 method, hillock-like cells can be detected at the transcriptional level at frequencies similar to those
275 detected at the protein level in human tissue samples (16 of 20 patient samples) (**Fig. 1e**). To
276 ensure that hillock-like cells share similarities between patients and represent a transcriptionally-
277 unique subset across patient tumours, we performed correlation analysis between basal-like and
278 hillock-like cells from each patient. In most cases, hillock-like tumour cells are more
279 transcriptionally similar to hillock-like cells from other patients than basal-like cells within the same
280 tumour—supporting that the hillock-like state represents a unique transcriptional state conserved
281 across patient tumours (**Extended Data Fig. 4g**).

282 ENRICHR analysis on DEGs in human hillock-like tumour cells revealed enrichment for immune,
283 neutrophil degranulation, and squamous differentiation signatures including keratinization and
284 cornification (**Fig. 4l** and **Supplementary Table 9**)—similar to the patterns observed in mouse
285 *Krt13*$^{+}$ tumour cells (**Fig. 4a**). We then used the transcriptional profile of the human hillock-like

population to perform deconvolution on bulk RNA-seq data of LUSC tumours in TCGA (see **Methods**). Using this approach, we identified a large proportion of LUSCs with cells that resemble the hillock-like population, and many tumours with a high proportion of hillock-like cells (**Fig. 4m**). Deconvolution was used to determine relative proportions of different immune populations in the same tumours. Consistent with our transcriptional predictions and recently published work[65], tumours with high hillock-like proportions were enriched for immunosuppressive myeloid populations including neutrophils and monocytes, while tumours with a low proportion of hillock-like cells were predicted to have tumour suppressive populations including activated T cells and M1 macrophages (**Fig. 4m** and **Supplementary Table 10**). These data are remarkably consistent with recent spatial transcriptomics of SNL tumours where our colleagues found increased macrophage and monocyte populations in *Tmprss11b*-high hillock-like regions[65]. Cell cycle analysis revealed that human hillock-like tumour cells are relatively less proliferative compared to other cell states in LUSC (**Fig. 4n**), consistent with findings in mice (**Fig. 4c** and **Extended Data Fig. 4b**). Together, these data suggest that a more lowly-dividing KRT13$^+$ hillock-like state is present and conserved in both mouse and human LUSC.

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

To address whether the KRT13$^+$ state can arise from the basal-like TPC population, we harvested tumours from Ad-CMV-Cre-infected SNL mice, flow sorted CD45$^-$GFP$^+$NGFR$^+$ tumour cells and plated them in Matrigel to establish organoids (**Fig. 5a**). Importantly, Ad-CMV-derived SNL tumour organoids recapitulate key structures found in SNL squamous tumours *in vivo*, similar to observations by other groups using lung tumour organoid systems[41,43,74-77]. We observed keratinized structures in H&E-stained organoids as well as proliferative KI67$^+$ outer edges coinciding with the basal TPC marker, NGFR (**Fig. 5b,c**). Importantly, a KRT13$^+$ cell population resided internally within the organoids with limited overlap with the ΔNp63$^+$ (also known as P40$^+$) basal layer (**Fig. 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 tumour organoids and sorted cells into two groups based on surface NGFR expression (NGFR^{high} and NGFR^{low}) (**Fig. 5e** and **Extended Data Fig. 5a,** see **Methods**). Sorted cells were replated in Matrigel and allowed to form organoids over 7–10 days in culture. Compared to NGFR^{low} cells, NGFR^{high} cells generated more abundant and larger organoids (**Fig. 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 (**Fig. 5g**).

To determine the transcriptional heterogeneity of tumour 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") (**Fig. 5e** and **Extended Data Fig. 5a**). Importantly, organoids largely transcriptionally recapitulate the heterogeneity observed in squamous SNL tumours, with distinct populations of *Ngfr/Trp63*-high basal-like cells and

Krt13/Tmprss11b-high hillock-like cells (**Fig. 5h-j** and **Extended Data Fig. 5b**). Consistently, there was a subpopulation of hillock-score-enriched cells within organoids, while the basal score was highly expressed across the entire transcriptional space (**Fig. 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[69,78]. Basal hillock and luminal hillock signatures overlapped with the general basal and hillock signatures, respectively (**Fig. 5k** and **Extended Data Fig. 5c**), similar to patterns in SNL tumour data (**Fig. 3g**). Given that mucinous histopathology precedes squamous differentiation in vivo[18,30,40,49-51], we did not expect to observe mucinous-like cells in organoids, and consistently, mucinous markers were extremely rare and low (**Extended Data Fig. 5c**). Outgrowth of NGFR^{high} cells repopulated the entirety of transcriptional space occupied by both freshly-sorted NGFR^{high} and NGFR^{low} cells, suggesting that basal cells can regenerate all captured cell states (**Fig. 5i**). Additionally, *Krt13* expression and the hillock score were restricted to NGFR^{low} (and not NGFR^{high}) cells at the time of sorting (**Extended Data Fig. 5d**). Outgrowth of NGFR^{high} cells led to reappearance of *Krt13* expression and a hillock-like population (**Fig. 5l** and **Extended Data Fig. 5d**), together supporting that hillock-like cells can derive from basal-like TPCs. Consistent with immunostaining for KI67 (**Fig. 5c**), hillock-like organoid cells are less proliferative based on cell-cycle gene expression assignment (**Extended Data Fig. 5e**). Organoid hillock-like cells also upregulate neutrophil degranulation and cornification gene sets (**Extended Data Fig. 5f** and **Supplementary Table 11**), similar to both mouse and human tumour data.

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

Next, we sought to identify master transcriptional regulators driving the hillock-like tumour 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 (**Extended Data Fig. 6a-c** and **Supplementary Table 12**). Basal- and basal hillock-like cells across mouse and human datasets predicted proliferation and basal-lineage transcription factors MYC, MAX, P63 and SOX2 (**Extended Data Fig. 6a**), 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 tumour cells and was also a top predicted regulator of normal airway hillock cells from two independent datasets[6,7] (**Extended Data Fig. 6b, c**). 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[55,79]. Moreover, KLF4 can directly regulate *KRT13* expression in esophageal cancer[80] and can repress basal TPC factors SOX2, TP63 and PITX1[23]. To determine KLF4's expression pattern in lung squamous tumours, we queried our RNA-seq data from lung cancer GEMMs and TCGA data, which showed enrichment of *Klf4/KLF4* in both mouse and human squamous tumours compared to adenocarcinomas (**Fig. 6a,b**).

To determine KLF4 expression and localization in LUSC, we co-stained for basal marker ΔNp63/P40 and KLF4 in mouse SNL tumours. KLF4 exhibited a nuclear and suprabasal-like pattern of expression where KLF4 was closely positioned to, but largely non-overlapping with P40^{+} cells (**Fig. 6c**). KLF4 was similarly located in a suprabasal pattern when co-staining with the basal TPC-marker NGFR in both Ad-CMV- and Ad-CCSP-initiated tumours (**Fig. 6d** and

370 **Extended Data Fig. 6d**). Co-staining for KLF4 and KRT13 showed that KLF4 was highly
371 expressed in a layer of KRT13$^+$ cells (typically in the outermost KRT13$^+$ basal-proximal cells), and
372 absent from the innermost KRT13$^+$ basal-distal cells (**Fig. 6e,f** and **Extended Data Fig. 6d**).
373 Finally, we stained SNL organoids for KLF4 and noted KLF4 expression enriched in a suprabasal
374 location to NGFR$^+$/P40$^+$ cells, with minimal overlap between KLF4 and basal markers (**Fig. 6g**
375 and **Extended Data Fig. 6e**); KLF4 was expressed in the majority of KRT13$^+$ cells, suggesting
376 that organoids have a similar organization as squamous tumours. These data suggest multiple
377 levels of intra-tumoural heterogeneity in squamous lung tumours (**Fig. 6f**) that mimic normal
378 stratified epithelia, similar to findings in the skin and vaginal epithelium[57,81]. These data suggest
379 KLF4 is positioned to be important for the emergence or differentiation of KRT13$^+$ hillock-like lung
380 tumour cells.

381 **KLF4 promotes a KRT13$^+$ hillock-like state**

382 To determine the functional role of KLF4 in the hillock-like tumour state, we sought to deplete
383 KLF4 expression and determine the impact on KRT13 and the hillock-like state in SNL tumour
384 organoids. Multiple attempts to knockout *Klf4* using traditional CRISPR gene-editing were
385 unsuccessful due to organoid death. Thus, we designed a CRISPR-inhibition (CRISPRi) system
386 using a catalytically dead Cas9 enzyme (dCas9) fused to the transcriptional repressor KRAB
387 protein[82]. Compared to non-targeting control sgRNAs (sgNTCi), *Klf4*-targeted CRISPRi constructs
388 (sg*Klf4*i) efficiently decreased KLF4 expression in SNL organoids, especially guide #1 (**Fig. 6h,i**).
389 Analysis of sgNTCi and sg*Klf4*i organoids by co-IF and immunoblot revealed significantly reduced
390 and more diffuse KRT13 staining upon *Klf4* knockdown (**Fig. 6h,i**). The CRISPRi system led to
391 variable KLF4 expression within organoids of the same culture, where KRT13 expression was
392 retained in KLF4-proficient sg*Klf4i* organoids but completely absent in KLF4-deficient sg*Klf4i*
393 organoids (**Extended Data Fig. 6f**). KLF4 suppression did not alter expression of the basal
394 marker KRT5 but increased expression of PITX1 (**Fig. 6i**), a transcription factor that cooperates
395 with SOX2 and TP63 in basal skin TPCs, with a mutually antagonistic relationship with KLF4[23].
396 After seven days of growth following dissociation and plating, KLF4 knockdown with sg*Klf4i-1*
397 slightly but significantly increased organoid frequency with no impact on organoid size (**Fig. 6j**).
398 Accordingly, sg*Klf4i* organoids showed elevated KI67 staining, with KI67$^+$ cells present throughout
399 the organoid when KLF4 was suppressed, whereas KLF4-proficient organoids tended to
400 proliferate specifically in the basal-associated outer layer of cells (**Fig. 6k-m**). These data may be
401 consistent with known functions of KLF4 in regulating cell cycle exit[23,54,83-85]. The lack of increased
402 organoid size despite increased proliferative markers may reflect size constraints in culture
403 conditions. Together, these data suggest that KLF4 is necessary for the transition to a less
404 proliferative KRT13$^+$ hillock-like state in the organoid context.

405 Next, we aimed to determine whether KLF4 is sufficient to promote the KRT13$^+$ hillock-like state.
406 We overexpressed human *KLF4* or *GFP* control in SNL organoids under the control of a
407 doxycycline-inducible tetracycline response element (Tet-ON-TRE-GFP vs KLF4). KLF4
408 overexpression in SNL tumour organoids increased KRT13 levels and substantially reduced
409 organoid growth (**Fig. 6n,o**). After dissociation and re-plating with doxycycline for 7 days,
410 organoids with KLF4 overexpression were smaller and formed at a lower frequency than GFP

controls (**Fig. 6o** and **Extended Data Fig. 6g**). Together, these data suggest KLF4 is both necessary and sufficient for the emergence of the KRT13^{+} state in SNL organoids.

We extended these findings to human squamous lung cancer-derived cell lines that express varying degrees of KLF4 and KRT13 at baseline (**Extended Data Fig. 7a**). Consistent with findings in murine organoids, *KLF4* knockdown in human LUSC led to a decrease in KRT13 (**Fig. 7a** and **Extended Data Fig. 7b**), while overexpression of KLF4 induced KRT13 (**Fig. 7b**). As in SNL organoids, KLF4 overexpression suppressed growth of H226 cells in vitro (**Fig. 7c**). In contrast to tumour organoids where KLF4 loss increased proliferation (**Fig. 6i-m**), KLF4 knockout in H520 cells decreased proliferation (**Extended Data Fig. 7c**), which may be due to unique constraints in these distinct cell populations, which differ by species, genetic alterations, and in vitro growth conditions.

Given that basal cells are implicated as the cell of origin for LUSC[32], that basal hillock cells give rise to luminal hillock cells in the lung airways[37], and that basal TPC-like cells differentiate into hillock-like tumour cells, we asked if KLF4 can modulate KRT13 in normal lung basal cells. We used CRISPR-Cas9 gene editing to knockout *KLF4* in BCi-NS1.1 cells, which naturally express KLF4 and have been shown to undergo differentiation away from a basal state in cell culture conditions[86-88]. (**Fig. 7d** and **Extended Data Fig. 7d**). As in tumour organoids and H520 human LUSC cells, *KLF4* knockout substantially decreased endogenous KRT13 levels especially with guide #3, which showed the most complete DNA editing by surveyor assay (**Fig. 7d** and **Extended Data Fig. 7d**). Similar to findings in H520 cells, *KLF4* knockout in BCi-NS1.1 cells led to reduced proliferation (**Extended Data Fig. 7e**), again suggesting context specific functions in 2D cell culture compared to 3D organoid culture.

To determine if KLF4 is sufficient to promote a KRT13^{+} state in human cells, the TRE-KLF4 overexpression construct was introduced into BEAS-2B cells that are considered a bronchial epithelial cell type that lacks expression of KLF4 and KRT13 under normal culture conditions (**Fig. 7e** and **Extended Data Fig. 7f**). Exposure to doxycycline induced KLF4 and led to striking KRT13 induction (**Fig. 7e**). Additionally, transient exposure of TRE-KLF4 cells to doxycycline was able to induce KRT13 expression, which was maintained for 48 hrs after doxycycline was removed, while KLF4 expression returned to baseline after doxycycline removal (**Extended Data Fig. 7f**). KLF4 overexpression significantly decreased growth of BEAS-2B cells (**Fig. 7f**), which is consistent with the slower proliferating KRT13^{+} hillock-like state observed in LUSC tumours and organoids (**Fig. 4c, 5c, 6k-o** and **7c**). KLF4 overexpression also led to increased P21 (**Fig. 7e**), consistent with previous work[54,83-85].

Transcriptomic analysis of BEAS-2B cells following KLF4 induction revealed significant upregulation of genes associated with the hillock state, including *KRT13* and genes associated with *Krt4/Krt13*+ mouse tracheal cells, and cornification and keratinization-related genes (**Fig. 7g, Extended Data Fig. 7g,h** and **Supplementary Table 13**). Similarly, transcriptomic analysis of SNL tumour organoids with KLF4 silenced (sg*Klf4i*-1) showed suppression of *Krt13* and hillock-associated genes (**Fig. 7h**, **Extended Data Fig. 7i,j,** and **Supplementary Table 13**). Notably, KLF4 overexpression and knockdown did not appear to alter immune-related gene sets, suggesting that KLF4 may regulate a part of the hillock-like program including *Krt13/KRT13* but

452 not all hillock-associated genes (**Extended Data Fig. 7g-j**). To determine if KLF4 induction of
453 *Krt13* and other hillock-associated genes is through direct transcriptional regulation, we performed
454 KLF4 CUT&RUN in SNL tumour organoids. Annotated peaks were enriched for the KLF4 binding
455 motif as expected (**Extended Data Fig. 7k**). A KLF4 bound region was detected directly upstream
456 of the *Krt13* locus (**Fig. 7i**), consistent with a previously annotated GKLF/KLF4-responsive
457 element (GKRE)[80]. Additionally, ENRICHR analysis of the top 500 promoter regions of protein-
458 coding genes bound by KLF4 showed enrichment of genes involved in Rho GTPase signaling
459 (known to regulate squamous epithelial differentiation), neutrophil degranulation, and Ephrin
460 signaling, signatures seen in our single cell analysis across models. The bound promoters were
461 also enriched for genes associated with normal *Krt13*+ hillock cells of the mouse trachea
462 (**Extended Data Fig. 7l-m** and **Supplementary Table 14**).

463 Finally, we generated a conserved 51-gene score between the overlapping genes upregulated in
464 BEAS-2B cells following KLF4 induction and downregulated by *Klf4* knockdown in SNL tumour
465 organoids ("KLF4-up score") and applied this score to our mouse and human single cell datasets
466 (**Extended Data Fig. 7n** and **Supplementary Table 13**, see **Methods**). Consistently, KLF4-
467 induced genes were highly enriched in the hillock-like populations in all three single cell datasets
468 (**Fig. 7j-l**), supporting a role for KLF4 in hillock-like cancer cells. Together, these data suggest
469 that KLF4 promotes KRT13 and a hillock-like state in both normal lung cell lines and tumour
470 contexts.

471 **KLF4 promotes resistance to chemotherapy**

472 Hillock cells of the normal airways are highly resistant to multiple forms of injury including viral
473 infection and physical stress[37]. We hypothesized that hillock-like cells in squamous tumours may
474 possess similar injury-resistant phenotypes. Focusing on standard-of-care chemotherapies
475 cisplatin and paclitaxel, we assessed if the hillock-like state dictates response to clinically-relevant
476 treatments. Organoid cells with and without KLF4 knockdown were treated with varying doses of
477 either cisplatin or paclitaxel as single-cell suspensions, replated with fresh drug, and allowed to
478 form organoids. Cisplatin decreased organoid proliferation within five days, with *sgKlf4i*-organoids
479 displaying increased sensitivity versus *sgNTCi* controls at all doses tested (**Fig. 7m**). Additionally,
480 while control organoids were able to repopulate during extended culture following 5 μM cisplatin,
481 *sgKlf4i* organoids were not (**Extended Data Fig. 8a**). At a lower dose of 2.5 μM cisplatin, *sgKLF4i-*
482 *1* organoids repopulated but exhibited restored expression of KLF4 and KRT13, suggesting
483 escape from KLF4-knockdown may facilitate cisplatin resistance (**Extended Data Fig. 8a,b**).
484 Interestingly, KLF4 knockdown had no effect on organoid response to paclitaxel (**Extended Data**
485 **Fig. 8c**). Similar to findings in mouse organoids, KLF4-knockout increased sensitivity to cisplatin
486 in the human basal cell line BCi-NS1.1, while having no effect on paclitaxel response (**Extended**
487 **Data Fig. 8d-f**). BCi-NS1.1 cells with the most complete KLF4-knockout (*sgKLF4-3,* **Fig. 7d**)
488 exhibited the most profound sensitivity to cisplatin with evidence of apoptosis as measured by
489 cleaved PARP and cleaved Caspase-3 levels (**Extended Data Fig. 8e**).

490 Of note, a top predicted regulator of our hillock-like state is *NFE2L2* (NRF2) (**Extended Data Fig.**
491 **6b, c**), a master regulator of oxidative stress. Cisplatin is known to induce oxidative stress along
492 with DNA damage, whereas this is not a well annotated response to paclitaxel. Thus, we

493 hypothesize that selective sensitization to cisplatin and not paclitaxel may be partially mediated
494 by oxidative stress. Indeed, organoids harbouring KLF4 loss were dramatically sensitized to
495 treatment with hydrogen-peroxide-induced oxidative stress compared to controls (**Fig. 7n**).
496 Further, treatment of cells with the antioxidant N-acetylcysteine (NAC) provided a partial rescue
497 of organoid proliferation in response to cisplatin (in both control and knockdown settings) (**Fig.**
498 **7o**). Recent work has suggested NRF2 and KLF4 cooperate to regulate oxidative stress response
499 genes[89], and NRF2 has been implicated in resistance to numerous therapies[41,90,91], such that
500 future studies should address whether NRF2 cooperates with KLF4 in the hillock-like state.
501 Together, these data suggest that KLF4 supports resistance to chemotherapy, associated with its
502 capacity to drive a hillock-like state.

503 **A conserved hillock-like state is present across squamous tumours with potential**
504 **therapeutic targets**

505 Squamous cell carcinomas (SCCs) can arise in tissues beyond the lung and display similar
506 histologic features[92-94]. Additionally, there are common underlying genetic alterations that promote
507 squamous tumours, especially in the lung, head & neck, and esophageal contexts[92-94]. We
508 therefore queried two previously published SCC datasets, one containing 12,932 cells from 15
509 patients with head & neck SCC (HNSCC)[95] and one containing 44,547 cells from 60 patients with
510 esophageal SCC (ESCC)[96] (**Fig. 8a** and **Extended Data Fig. 9a**), for evidence of the hillock-like
511 state. Both tumour types contained populations of cells that expressed *KRT13*, as well as the
512 basal markers *KRT5* and *TP63*, with minimal overlap between them (**Extended Data Fig. 9b**).
513 The basal score was broadly expressed across the datasets while the hillock score was more
514 restricted to clusters that are lower for basal score expression (**Extended Data Fig. 9c**), similar
515 to patterns seen in LUSC (**Fig. 4i**). Using methods identical to those used for the human LUSC
516 analyses, we assigned cell states to both the HNSCC and ESCC cells using a hillock score cutoff
517 approach (**Fig. 8b** and see **Methods**). Hillock-like cells in both datasets showed evidence of
518 decreased proliferation by cell cycle phase assignment; ENRICHR analysis of upregulated genes
519 in the hillock-like state showed remarkable similarity to LUSC, including innate immune and
520 squamous-related gene signatures (**Extended Data Fig. 9d,e**). Finally, the KLF4-up score was
521 enriched in hillock-like cells in both HNSCC and ESCC (**Fig. 8c**). Together, these findings suggest
522 that a KLF4-driven hillock-like state exists across multiple squamous tumour types and that
523 understanding this cell state should have implications beyond lung cancer.

524 The ability to differentiate and target specific tumour cell states could allow the ability to study the
525 functional role of individual cell states in disease progression. KRT13 appears to be a marker of
526 the hillock-like state across models and tissues, which may have utility for potential diagnostic
527 purposes. To identify additional clinically-relevant markers, we sought to identify hillock-enriched
528 cell surface markers, which could act as potential targets for antibody drug conjugates (ADCs) or
529 bi-specific T cell engagers (BiTes). We identified *TACSTD2* (encoding TROP2) as a top cell
530 surface protein-encoding gene enriched in hillock-like cells across models (**Fig. 8 d-f**; **Extended**
531 **Fig. 10a-d** and **Supplementary Table 15**) as well as *PSCA* and *CEACAM1*. TROP2 ADCs have
532 recently been approved for treatment of breast cancer and EGFR-mutated lung cancer and are
533 in numerous clinical trials (NCT02574455, NCT03901339, NCT07328490, TROPION-Lung, etc).
534 Finally, *Lgals3* is a top hillock-enriched gene throughout our study (**Extended Data Fig. 4a,**

535 **Supplementary Table 6** and **7**), and while not a cell-surface protein per se, it is a secreted lectin
536 that localizes to the cell surface via interactions with glycosylated receptors (**Fig 8d-f**). We
537 validated LGALS3 to be restricted to the hillock-like compartment in SNL tumours by IF (**Fig. 8g**).
538 LGALS3 has been implicated in neutrophil recruitment and regulation of pro-tumor immunity[97-101],
539 while drugs targeting it have good safety profiles in early-stage clinical trials[102]. Future studies are
540 warranted to determine if these targets can be exploited to enrich for the hillock-like state, evaluate
541 its presence in patient samples, and/or develop novel therapies specifically targeting hillock-like
542 cells.

543 **Discussion**

544 Here, we identify a KRT13^{+} hillock-like tumour cell state in LUSC that resembles recently
545 described hillock states in the normal airway epithelium[37]. Hillock-like tumour cells are present
546 across mouse tumours, tumour organoids, and human LUSC, and we detect similar populations
547 in additional squamous malignancies. Together, our findings support the idea that squamous
548 tumours exhibit lineage plasticity, adopting transcriptional states that mirror normal epithelial cell
549 states, and that these programs contribute to the non-genetic heterogeneity of LUSC.

550 Our data support a model in which hillock-like cells represent a basal-derived state that is more
551 differentiated and slower dividing. In the normal airway, hillock cells are linked to epithelial
552 protection and regeneration after injury[6,37]; in tumours, a similar program may allow a subset of
553 cells to persist under stress while maintaining the capacity to transition between states. This
554 raises key questions for future studies, including whether basal-like cells transition into hillock-like
555 states in vivo, whether hillock-like cells can revert back to basal-like populations after therapy,
556 and which microenvironmental signals promote or restrict these transitions.

557 We also find that cell of origin influences hillock-like programs in the SNL model. Club-initiated
558 tumours more prominently exhibit luminal hillock-like features, whereas basal-enriched contexts
559 (Ad-CMV-Cre and tumour organoids) more readily show basal-hillock-like features, reminiscent
560 of lineage-tracing studies showing that normal hillock cells are intermediate between basal and
561 club cells[6,7]. Together, these findings suggest that lineage history and epigenetic context may
562 shape which cell states are most accessible during tumour evolution. This paradigm, in which the
563 cell of origin impacts lineage trajectories, has also been observed in small cell lung cancer[69] and
564 is therefore likely broadly relevant across cancers.

565 Mechanistically, our data identify KLF4 as a key regulator of the hillock-like state. KLF4 directly
566 binds the *Krt13* locus and regulates a conserved KLF4-driven transcriptional module enriched in
567 hillock-like populations across mouse and human contexts. At the same time, hillock-like cells
568 display broader immunomodulatory features that may not be fully explained by KLF4 alone,
569 suggesting that additional regulators contribute to this program. Another top candidate predicted
570 to regulate the hillock-like state from our computational analyses is NRF2, which is known to
571 regulate oxidative stress responses, the tumour immune microenvironment, and therapeutic
572 resistance in LUSC, all phenotypes of hillock-like cancer cells based on our findings[41,90,91,103].
573 Future work should test whether KLF4 cooperates with NRF2 or other factors to shape the hillock-
574 like phenotype, including immune-related features.

575 The potential clinical relevance of this state is supported by the finding that KLF4 promotes
576 resistance to cisplatin and oxidative stress. A slower-dividing, stress-resistant cell population
577 could survive cytotoxic therapy and contribute to tumour persistence and regrowth through
578 plasticity. In addition, hillock-like programs correlate with immune contexts enriched for myeloid
579 populations in human datasets, which we and colleagues also observe in SNL lung tumours using
580 spatial transcriptomics[65]. TMPRS11B, which is expressed in *Krt13*+ hillock-like cells and is
581 regulated by KLF4[65], drives lactate export by promoting MCT4 activity and glycolytic metabolism,
582 which correlates with low-pH regions and increased innate immune infiltration, providing one
583 example of how the hillock-like state may alter the tumour immune microenvironment[64,65].
584 Together, these findings are consistent with the possibility that specific tumour cell states shape
585 the immunosuppressive microenvironment in distinct ways. Future studies should directly test
586 whether hillock-like tumour cells influence immune recruitment or function, and whether these
587 interactions impact response to immunotherapy.

588 Finally, our analyses highlight potential markers and vulnerabilities enriched in hillock-like cells,
589 including therapeutically tractable surface-associated targets such as TACSTD2 (TROP2),
590 CEACAM1, PSCA, and LGALS3. Together, these observations motivate future work to develop
591 strategies that selectively track, enrich, or target hillock-like tumour cells. Overall, our findings
592 suggest that effective LUSC therapy may require targeting both proliferative basal-like populations
593 and stress-adapted states that support tumour plasticity, therapy resistance, and immune
594 suppression.

595 **Limitations of the Study**

596 Our study suggests non-genetic factors support intra-tumoural heterogeneity in LUSC and that
597 *Krt13*$^{+}$ hillock-like cells may contribute to tumour progression, but we acknowledge limitations to
598 our systems. First, our mouse models and organoids are predominantly derived from the SNL
599 model. *SOX2* amplification is one of the most common genetic events in squamous tumours and
600 SNL squamous tumours are known to recapitulate many features of human LUSC[40]; however,
601 this model harbours mucinous adenocarcinomas, and squamous tumours are likely derived from
602 transdifferentiation from an adenocarcinoma state, which may not reflect the origins of many
603 human tumours. Because we are not yet able to resolve a luminal-hillock-like and basal-hillock-
604 like cell state in human tumour data (partially due to higher levels of heterogeneity limiting
605 clustering-based approaches), we cannot be sure whether human LUSC is predominantly basal
606 and basal-hillock, or contains luminal hillock-like populations as well.

607 Reliance on one GEMM limits the ability to assess the impact of genetics on the hillock-like state.
608 We show that the LP model and human tumours also contain KRT13-expressing cells with
609 transcriptional similarities between other human squamous tumours, but future work should
610 incorporate additional genetic models. Our findings suggest that cell of origin impacts intra-
611 tumoural heterogeneity, as tumours initiated in club cells show increased KRT13 positivity
612 compared to Ad-CMV-Cre-initiated tumours. Given that the cells of origin with Ad-CMV-Cre are
613 unclear, further characterization using more specific Cre drivers (including a basal *Krt5*-Cre) can
614 clarify these developmental relationships.

615 **AUTHOR CONTRIBUTIONS**

616 Conceptualization: L.T. Izzo, K.A. O'Donnell, T.G. Oliver
617 Data Acquisition: L.T. Izzo, T. Reyes, S. Meesala, S. Yang, H.S. Sunil, A.S. Ireland, L. Earnest-
618 Nobel, X.C. Cheng, N. Tserentsoodol, S.B. Hawgood
619 Data Analysis and Interpretation: L.T. Izzo, T. Reyes, H.S. Sunil, A.S. Ireland, C. Glass, D.R.
620 Tyson, T.G. Oliver
621 Writing – Original Draft: T. Reyes, L.T. Izzo, T.G. Oliver
622 Writing – Review and Editing: L.T. Izzo, S. Meesala, A.S. Ireland, T. Reyes, D.R. Tyson, K.A.
623 O'Donnell, T.G. Oliver
624 Funding Acquisition: L.T. Izzo, K.A. O'Donnell, T.G. Oliver
625 Resources: E.F. Patz, B.J. Witt
626 Supervision: K.A. O'Donnell, T.G. Oliver
627

628 **ACKNOWLEDGEMENTS**

629 We thank members of the Oliver Lab for technical assistance and mouse colony management
630 (K.S. Lucas, D. Soltero, K. Spainhower, R. Uzachenko), and administrative support (X.C. Cheng,
631 L. Houston, L. Youngblood), and Dr. Greg Wang's lab for CUT&RUN expertise (Y. Guo). We
632 appreciate MedGenome staff for sequencing services, and Duke Human Vaccine Institute for use
633 of the tapestation. We acknowledge support from the NIH National Cancer Institute (NCI) under
634 award U24CA213274 (TGO), R01-CA244841, and R01-CA187457 (TGO). T.G.O. received
635 support as a Duke Science & Technology Scholar and from the Lung Cancer Initiative of North
636 Carolina. K.A.O. was supported by NCI R01CA273585, P50CA70907, V Foundation (T2021-
637 011), and the Department of Defense (DoD LC190249). L.T.I. received support from the American
638 Cancer Society (PF-25-1366291-01-PFCBI). A.S.I. received support from the NIH NCI (F31
639 CA275295). We thank the BioRepository & Precision Pathology Center (BRPC), a shared
640 resource of the Duke University School of Medicine and Duke Cancer Institute, which receives
641 support from the P30 Cancer Center Support Grant (P30 CA014236) and from the Cooperative
642 Human Tissue Network (UM1 CA239755) and the Duke Cancer Institute Flow Cytometry Core.

643 **DECLARATION OF INTERESTS**

644 TGO has a patent related to SCLC subtypes (US12188095-B2), a sponsored research agreement
645 with Auron Therapeutics, has consulted for Light Horse Therapeutics and Nuage Therapeutics,
646 served on the scientific advisory board (SAB) for Lung Cancer Research Foundation, and as a
647 consulting editor for *Cancer Research* and *Genes & Development*. All other authors declare no
648 conflicts of interest related to this work.

649

650 EXPERIMENTAL MODEL AND SUBJECT DETAILS

651 **GEMMs**

652 *Rb1*$^{fl/fl}$;*Trp53*$^{fl/fl}$;*Rbl2*$^{fl/fl}$ (RPR2)[61], *Rb1*$^{fl/fl}$;*Trp53*$^{fl/fl}$;*LSL-Myc*^{T58A} (RPM)[60], *Sox2-Ires-GFP*$^{LSL/LSL}$;*Nkx2-*
653 *1*$^{fl/fl}$;*Lkb1*$^{fl/fl}$ (SNL)[40], *Lkb1*$^{fl/fl}$;*Pten*$^{fl/fl}$ (LP)[18], and *Kras*$^{G12D/+}$;*Trp53*$^{fl/fl}$ (KP)[59] mice have been
654 previously described. Methods for tumour initiation and monitoring can be found in the listed
655 references. Ad5-Cre with the following promoters was used: RPM (10 Ad-CGRP-Cre initiated
656 tumours, 5 Ad-CMV-Cre initiated tumours), RPR2 (4 Ad-CGRP-Cre initiated tumours, 4 Ad-CMV-
657 Cre initiated tumours), KP (8 Ad-CMV-Cre initiated tumours), LP (6 Ad-CMV-Cre initiated
658 tumours), SNL (4 Ad-CMV-Cre and Ad-CCSP-Cre initiated tumours). All mice were housed
659 following the regulations set by the Institutional Animal Care and Use Committee of the University
660 of Utah and Duke University. At 6-10 weeks of age, anesthetized SNL mice for this study were
661 infected by intranasal instillation with $1x10^8$ plaque forming units of Ad5-CMV-Cre or Ad5-CCSP-
662 Cre adenovirus (University of Iowa). Viruses were administered in a Biosafety Level 2+ room
663 according to the guidelines from the Institutional Biosafety Committee.

664

665 **Human samples**

666 Human lung cancer tissue microarray (TMA) samples were a gift from Edward Patz, Jr. TMA
667 samples were de-identified and only information on the histopathological subtype scoring by a
668 trained pathologist was provided. Human TMAs from patients with lung cancer were generated
669 following written informed consent from each patient under an Institutional Review Board-
670 approved protocol at Duke University Medical School (Protocol ID: Pro00012914). TMA samples
671 were reviewed by an additional pathologist to certify histopathological subtype and provide
672 percent tumour scoring. Sections with zero percent tumour were not analyzed.

673

674 **Cell lines**

675 The human bronchial epithelial cell line BEAS-2B was acquired from the Duke Cell Culture Facility
676 (Sigma Catalog #95102433). The BCi-NS1.1 cells, an immortalized human airway basal cell
677 stem/progenitor cell line, was generated at Cornell University by Dr. Ronald Crystal and provided
678 as a gift for research purposes[87]. Both cell lines were grown in 2D adherent culture plates in
679 Bronchial Epithelial Cell Growth Basal Medium (BEBM, Lonza Catalog # CC- 3171) with
680 Supplements and Growth Factors (Lonza Catalog # CC-4175) along with 0.2% Penicillin-
681 Streptomycin (BEBM media). For serial passaging, media was removed and 0.25% Trypsin-EDTA
682 (Gibco catalog #25200056) was added to dislodge cells. HEPES Buffered Saline Solution
683 (Catalog # CC-5024) supplemented with 15% fetal bovine serum (FBS) was added to quench the
684 reaction. The cells were centrifuged at 250g for 5 min at 4°C. The supernatant was removed, and
685 cells were resuspended in BEBM media before plating into tissue culture treated dishes.

686 Human LUSC cell lines NCI-H226 (H226) and NCI-H520 (H520) were purchased from ATCC
687 (CRL-5826 and HTB-18 respectively). Both cell lines were grown in RPMI media (Gibco #
688 11875093) supplemented with 10% FBS, 1% Penicillin-Streptomycin, and 1% L-glutamine (200
689 mM). For serial passaging, media was removed and 0.25% Trypsin-EDTA (Gibco catalog
690 #25200056) was added to dislodge cells. Cell culture media was used to stop the reaction and
691 cells were plated into tissue culture treated plates.

692

693 **Organoids**

694 After sorting from tumours (see **Harvesting tissue for organoid production or single cell**
695 **sequencing**), single cell suspensions containing SNL tumour cells ($NGFR^{+}GFP^{+}$) were
696 centrifuged at 2000 rpm for 5 min at 4°C. Supernatant was removed and cells were resuspended
697 in Matrigel before plating. Cells were grown in 3D Matrigel domes submerged in organoid culture
698 media (OCM) [50% DMEM/F12 media (supplemented with 10% FBS, 1% penicillin/streptomycin,
699 1% L-glutamine, 50ng/mL Epidermal Growth Factor (EGF) (Thermo Cat # PHG031) and 50ng/mL
700 Fibroblast Growth Factor (FGF) (Thermo Cat # PHG0369)) + 50% L-WRN conditioned media
701 (containing Wnt-3A, R-spondin 3 and Noggin) + 10uM ROCK inhibitor]. L-WRN conditioned media
702 was generated according to previously published methods[104] using L-WRN cells purchased from
703 ATCC (CRL-3276). For serial passaging or flow sorting, media was removed and cold DMEM
704 media was added to wells before disruption of the Matrigel plug by pipetting. The organoid and
705 Matrigel containing media was centrifuged at 1500 rpm for 5 min at 4°C. Supernatant was
706 removed and organoids were resuspended in phosphate buffered saline (PBS) before
707 centrifugation at 1500 rpm for 5 min at 4°C. Supernatant was removed and cells were
708 resuspended in 500 uL TrypLe for 15 min at 37°C. Following vigorous pipetting, the reaction was
709 quenched with DMEM and was centrifuged at 1500 rpm for 5 min at 4°C. Cells were then
710 resuspended in Matrigel and plated in 24 well plates. For sorting experiments, single cell
711 suspensions were sorted, centrifuged at 1500 rpm for 5 min at 4°C, then resuspended in Matrigel
712 and plated in 24 well plates.

713

714 **METHOD DETAILS**

715

716 **Harvesting tissue for organoid production or single cell sequencing**

717 Mice were euthanized to harvest tumour-bearing lung tissue or tracheas. For tumour cell isolation,
718 tumour-bearing lung tissue was removed and put into cold DMEM media (10% FBS, 1% L-glut,
719 1% Pen/Strep) and chopped into smaller pieces mechanically with fine scissors before further
720 dissociation was performed using the Lung Dissociation kit (Miltenyi Biotech Cat # 130-095-927)
721 at 37°C following the manufacturer's instructions. Dissociated cells were filtered using a 100 µm
722 strainer and collected into conical tubes. Cells were pelleted by centrifugation at 1500 rpm at 4°C
723 for 5 min and then resuspended in ammonium chloride potassium (ACK) lysis buffer to lyse red
724 blood cells. The lysis reaction was quenched with addition of PBS and cells were centrifuged.
725 Supernatant was removed and cells from tumour-bearing lungs were resuspended in FACS buffer
726 (PBS containing 2% FBS) for subsequent flow sorting and single cell sequencing.

727 **Flow sorting tumour cells**

728 Single cell suspensions from tumour-bearing lungs or tumour organoids (see **Harvesting tissue**
729 **for organoid production or single cell sequencing** and **Organoids**) were stained with a rabbit
730 anti-NGFR (Abcam Cat #ab8875 or ATS Cat # AB- N01AP) primary antibody for 30 min on ice.
731 Cells were then washed with FACS buffer and stained in FACS buffer containing a fluorescent
732 secondary antibody (APC-goat anti-rabbit Invitrogen Cat # A-10931) and a fluorescent conjugated
733 antibody (BV605-rat anti-CD45 Biolegend Cat # 103155, for tumours). Cells were then incubated
734 on ice and in the dark for 20-30 min. Following incubation, cells were washed with FACS buffer
735 and centrifuged. The supernatant was discarded, and cells were resuspended in FACS buffer
736 containing DAPI to detect dead cells. Cells were then sorted using flow cytometry according to
737 the outlined gating scheme on a cell sorter (either BD Bioscience FACSAria or Beckman Coulter
738 MoFlo Astrios). Flow sorted tumour cells were used for single cell RNA-sequencing or for plating

739 of organoids within 30 minutes directly following sorting (see **Single cell RNA-sequencing** or
740 **Organoids** below). For SNL Ad-CMV-Cre tumour cells used for sequencing, GFP+ and
741 GFP+/NGFR+ cells were sequenced in two independent experiments. For SNL Ad-CCSP-Cre
742 tumour cells used for sequencing, GFP+ cells were used for the first sequencing (S1) and a
743 mixture of GFP+, GFP+/NGFR+, and CD45+ cells were used for the second sequencing (S2).

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

746 **Single cell RNA-sequencing**

747 Single-cell suspensions from flow sorted GFP^+ and $NGFR^+$ GFP^+ SNL tumours or SNL tumour
748 organoids were prepared for sequencing according to the 10x Chromium platform protocols.
749 (https://support.10xgenomics.com/single-cell-gene-expression). The SNL Ad-CMV- and Ad-
750 CCSP-Cre initiated tumour samples sorted for GFP^+, and the SNL Ad-CMV-Cre GFP^+/$NGFR^+$
751 tumour cells were prepared and sequenced at the Huntsman Cancer Institute High-Throughput
752 Genomics Core at the University of Utah. The SNL Ad-CCSP-Cre mixture of GFP+, GFP+/NGFR+
753 (S2) tumour sample and all organoid samples were prepared in-house at Duke University and
754 sequenced by MedGenome. Libraries were prepared with the Chromium 3' single cell gene
755 expression (GEX) kit V2 chemistry for Ad-CMV-GFP^+ (10x Genomics Cat # PN-120237) and V3
756 chemistry for Ad-CMV-$NGFR^+GFP^+$, Ad-CCSP-GFP^+, Ad-CCSP-$NGFR^+GFP^+$, and all organoid
757 samples (10x Genomics Cat # PN-1000141). In brief, individual cells were barcoded with 16-bp
758 10x barcodes and cell-specific transcript molecules were tagged with 10bp unique molecular
759 identifiers (UMIs), according to the manufacturer's instructions. Reverse transcription reagents
760 and 10x gel beads were loaded on Chromium Chip Gs (10x Genomics Cat # PN-1000136) and
761 into the Chromium single-cell controller (10x Genomics Cat # PN-120263) to form GEMs. Within
762 singular GEMs, cDNA generated from captured and barcoded mRNA was synthesized by reverse
763 transcription following the manufacturer's recommendations. Subsequent A tailing, end repair,
764 adaptor ligation, and sample indexing were performed for each sample according to the
765 manufacturer's instructions. The resulting cDNA libraries were assessed on Agilent D1000 or
766 D5000 ScreenTape on an Agilent Technology 2200 TapeStation system and quantified by
767 quantitative PCR using a KAPA Biosystems library quantification kit (Roche Cat #
768 070960140001). Individual libraries were normalized and sequenced on an Illumina NovaSeq
769 6000, with the exception of the Ad-CMV-GFP^+ sample that was sequenced on an Illumina HiSeq,
770 with 10x-recommended read lengths and aiming for >20k reads per cell.

771 **10x single Cell RNA-seq data processing**

772 **Demultiplexing and data alignment**

773 CellRanger was used to demultiplex and align reads per sample for downstream analysis. Single
774 cell RNA-seq data from SNL samples were first demultiplexed with 10x cellranger mkfastq to
775 create fastq files. For each sample, reads were then aligned to the GRCm38 (mm10) mouse
776 genome. For SNL tumour organoids, reads were aligned to a custom mouse genome containing
777 sequences for *fLuc*, *Venus, Cas9,* and e*GFP* (GRCm38-mm10-2020-A)[69]. Count barcodes and
778 UMIs were generated using cellranger count.

779 **Quality control, data combination, and normalization**

780 All clustering, cell type identification and quality control (QC) was performed in Python (v3.8.8)
781 utilizing Scanpy (v1.10.0) according to current expert recommendations for single cell best
782 practices[105]. All cells sequenced from Ad-CMV- or Ad-CCSP-Cre initiated tumours or tumour

783 organoids were read into sparse matrix anndata objects created from filtered feature matrices
784 with scanpy.read_ 10X_mtx(), and quality metrics were calculated using
785 scanpy.pp.calculate_qc_metrics(). Low quality cells and potential doublets were excluded by
786 selecting for cells with ≤15% mitochondrial content and the following criteria for respective sample
787 types: SNL primary tumour samples with cells containing >1000 total counts and between 1000
788 and 6000 genes with at least one read; SNL tumour organoid samples with cells containing >1000
789 total counts and between 2000 and 8000 genes with at least one read. Human patient data was
790 accessed from Wu et al[70], patient tumours identified as squamous were used, and no additional
791 QC filtering was performed. Head and neck patient data was accessed from Kurten et al[95], utilizing
792 only $CD45^-$ samples and were filtered for cells containing >500 total counts and between 500 and
793 4000 genes with at least one read. Esophageal patient data was accessed from Zhang et al[96],
794 utilizing only $CD45^-$ samples annotated as epithelial, and no additional QC filtering was performed.
795 Normalized counts were calculated with scanpy.pp.normalize_total() and a target sum of 10,000.
796 Integrated anndata objects containing multiple scRNA-seq datasets were combined using
797 adata.concatenate() with join='outer'.
798
799 **Batch correction and clustering**

800 Data were further processed using Scanpy (v1.10.0) and scVI-tools (v0.17.4) and benchmarking
801 standards to minimize batch effects while maintaining true biological variability, particularly across
802 integrated objects[106]. Specifically, we utilized scvi.model.SCVI.setup_anndata() to initialize the
803 integration and clustering of datasets from concatenated anndata objects containing all samples
804 and genes, continuous covariate keys set as percent mitochondrial counts, and batch keys
805 identifying samples prepared or sequenced at different times. The model was trained with default
806 parameters, an early stopping patience of 20, and a maximum of 400 epochs, using the
807 model.train() function on the top 10,000 variable genes determined using the
808 sc.pp.highly_variable_genes() function. The latent representation of the model was obtained with
809 model.get_latent_representation() and added to the .obsm of each anndata object. Neighbors
810 were then calculated with sc.pp.neighbors() with use_rep set to the .obsm category added from
811 the latent representation. UMAP embedding was performed using sc.tl.umap() with min_dist=0.5.
812 Finally, Leiden clusters were generated with sc.tl.leiden() with resolution set to 1.0 for initial steps.
813 As is required throughout the scVI pipeline, we utilized raw counts for all steps above. Additional
814 rounds of quality control and data filtering were performed per dataset by assessing
815 n_genes_by_counts, total counts, and percent mitochondrial counts per cluster. In general, the
816 model tends to cluster low quality and doublet cells together, so clusters with exceptionally high
817 or low average genes_by_counts, total counts, or mitochondrial content were labeled low-quality
818 and considered for removal from the dataset. Removal was performed after assessing gene
819 expression based on known markers of tumour and normal lung cells, and marker genes per
820 cluster as determined by sc.tl_rank_genes_groups(), to help ensure that biological cells that
821 normally have higher or lower n_genes_by_counts, total_counts, or percent mitochondrial content
822 were not aberrantly filtered out. Additionally, from SNL primary tumour samples and human
823 patient data, we identified and removed non-tumour cell types including immune, endothelial,
824 fibroblast, and normal lung cell types using a panel of common marker genes. For the human
825 patient data, adenocarcinoma clusters were also removed based on expression of genes known
826 to be high in lung adenocarcinomas and not expressed in squamous tumours. Each time a cluster
827 was removed, we ran the scVI pipeline on the new anndata object iteratively through this quality
828 control step until there were no longer any low-quality cell clusters or non-tumour cells in the
829 anndata object. Additionally, each time clusters were removed from a larger anndata object, the
830 pipeline was re-run for optimal clustering. Final Leiden clusters were determined with sc.tl.leiden()
831 resolution set to 1 for the human LUSC patient data (**Fig. 4**), 0.75 for SNL tumour organoids (**Fig.
832 5**), 1 for SNL Ad-CMV and Ad-CCSP cells combined (**Fig. 3**), 0.5 for SNL Ad-CMV tumour cells

(**Fig. 2**), 0.5 for both head & neck and esophageal patient data (**Fig. 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/TGOliver-Lab/Izzo_Hillock_2025) and publicly accessible upon publication.

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

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 tumours were used to generate the mucinous adenocarcinoma score[62]. The "basal hillock score" (n=170) was generated from differentially-expressed genes comparing basal hillock 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)[37] (see **Supplementary Tables 1-5**). 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.

CIBERSORTx

CIBERSORTx is a tool developed by the Alizadeh and Newman Labs at Stanford University that uses single-cell RNA-sequencing expression data input to estimate abundances of defined cell types in a bulk cell population, such as a bulk tumor[108,109]. We utilized the interactive user interface of CIBERSORTx at https://cibersortx.stanford.edu and the Cell Fraction module. A custom signature matrix file was built from a single cell reference sample file that included normalized single cell transcriptomic data from human squamous cell carcinoma tumors[70] (n=100 randomly sampled cells per patient, n=18 patient tumors analyzed). Each cell was defined as "hillock-like" or "non-hillock-like" based on threshold expression values of the "hillock-cell score" and "basal score", described above (hillock-like=z-score hillock_score>=2, z-score basal_score<=3). Bulk RNA-seq data from TCGA of LUSC tumors (n=538 tumor) were used as the input mixture file and subject to deconvolution with the number of permutations set to 50. Results of cell-type imputation were sorted by tumors with the highest fraction of "hillock-like" cells to the lowest fraction. Independent deconvolution of the LUSC TCGA tumors was also performed using the published leukocyte, LM22_Cibersort signature matrix[108] to deconvolve each LUSC tumor into 22 immune cell types. Correlations of immune cell fractions (**Supplementary Table 10**) and KLF4-fraction with the KRT13-like fraction in each bulk LUSC tumor were assessed using simple linear regression tests in Graphpad Prism where CIBERSORTx graphs were also generated.

Calculation of KLF4-up score

881 The “KLF4-up score” (n=51) was generated through the intersection of genes upregulated when
882 comparing doxycycline treated to no doxycycline control of *TRE-KLF4*-expressing BEAS-2B cells
883 (converted to mouse orthologs) and downregulated comparing *sgKlf4i-1* and *sgNTCi* SNL tumour
884 organoids (see **Bulk RNA-seq sample and data processing from BEAS-2B and SNL tumour**
885 **organoids** and **Supplementary Table 13**). The KLF4-up score was computed for each cell per
886 scRNA-seq library by converting expression data to a AnnData object and utilizing the
887 scanpy.tl.score_genes function in Scanpy.

888 **Calculation of cell cycle phase**

889 Cell cycle analysis was performed using the scanpy.tl.scores_cell_cycle function. S-phase and
890 G2/M-phase gene sets were derived from cc.genes in Seurat[71] and were converted to mouse
891 homologs. Each cell was assigned a cell cycle phase based on gene score expression using the
892 scanpy.tl.scores_cell_cycle function. Gene scores were used for statistical comparison between
893 assigned cell states to compare between predicted populations in S and G2/M phases.

894 **Cell state assignment and correlation analysis**

895 Cell states for SNL primary tumours were determined by evaluating marker gene expression and
896 gene signature scores in Leiden clusters. Cell state assignments were validated by gene
897 expression per cell state and by evaluation of DEGs from each cell state compared to all other
898 cells (see **Differential gene expression analyses**). For the human patient lung squamous cell
899 carcinoma, esophageal squamous cell carcinoma, and head and neck squamous cell carcinoma
900 data, cell state was assigned by determining a “hillock-like” cutoff. The z-score for the hillock score
901 and basal score were calculated, and cells were determined to be “hillock-like” if the hillock score
902 z-score was >=2 and the basal score z-score was <=3, which excluded highly basal-like cells
903 while maintaining cells with high hillock gene expression compared to the overall cell population.
904 Correlation analysis was performed using the sc.tl.dendrogram and the sc.pl.correlation_matrix
905 functions in scanpy with default parameters using the Pearson method.

906 **Differential gene expression analyses**

907 Differentially-expressed genes (DEGs) for all samples was determined using the
908 scanpy.tl.rank_genes_groups function in Scanpy using the Wilcoxon Rank Sum Test on log
909 transformed data. This test compares each group to all other groups to determine top DEGs in a
910 specific cell population. ENRICHR analyses[110-112] were performed on the top 500 upregulated
911 genes (DEGs) in the SNL mouse tumour cell types (basal-like vs all others, basal hillock-like vs
912 all others, and luminal hillock-like vs all others), 500 genes in the human tumour cell type (basal-
913 like vs hillock-like), and 500 genes in organoids (basal-like vs hillock-like). For ENRICHR analysis
914 of normal hillock genes from Montoro et al, Nature, 2018, DEGs for hillock cells were included if
915 the log2FC>=0.5 and all *Krt13/Krt4*-enriched genes from Plasschaert et al, Nature, 2018, were
916 included.

917 **Differential gene expression analyses for cell surface protein encoding genes**

918 Differentially-expressed genes (DEGs) for all samples was determined using the
919 scanpy.tl.rank_genes_groups function in Scanpy using the Wilcoxon Rank Sum Test. Genes were
920 then filtered using the CSPA mouse or human cell surface protein list[113] combined with the human
921 in silico cell surface protein atlas (all predicted surface proteins as defined by SURFY)[114] for
922 human or the SURFY list converted to mouse orthologs for mouse.

923 **CUT&RUN sample processing and analysis**

924 SNL tumour organoids were collected into a single cell suspension as described in **Organoids**
925 and resuspended in OCM media. CUT&RUN was performed using the Cutana CUT&RUN
926 protocol version 2.1 (EpiCypher). In brief, 500,000 organoid cells per CUT&RUN reaction were
927 washed in Wash Buffer with 0.05% digitonin, incubated with activated ConA beads (EpiCypher
928 Cat # 21-1401), and subject to antibody binding overnight (anti-KLF4, R&D Systems AF3158, 0.5
929 µg; anti-H3K27Ac, CST 8173S, 0.5 µg; IgG, CST 2729S, 0.5 µg). pAG-MNase was used to digest
930 the chromatin, the released DNA fragments were purified using an NEB Monarch Spin PCR &
931 DNA Cleanup Kit (NEB # T1135). 5 ng of DNA per reaction was used as input for library
932 preparation using the NEBNext Ultra II DNA Library Prep Kit (NEB # E7645S). Libraries were
933 sequenced on a NovaSeq instrument with 150 bp paired end reads targeting 5 to 10M reads per
934 sample. We obtained over 5M reads for the anti-KLF4 CUT&RUN.

935 Raw sequencing read data was input into the nf-core/cutandrun pipeline (version 3.2.2) with
936 default parameters and aligned to the mouse mm10 genome[115,116]. Peaks were called using both
937 SEACR[117] then were annotated using HOMER[118,119] with the mm10 reference mouse genome.
938 SEACR annotated peaks were filtered by protein-coding and promoter-TSS annotations before
939 ENRICHR analysis. The standard nf-core/cutandrun .bigWig output files were used in IGV for
940 genome track visualization.

941 **KLF4 modulation in organoids**

942 For lentiviral CRISPR-mediated inhibition, the following guideRNAs targeting *Klf4* were cloned
943 into the dCas9 KRAB vector[82] (Addgene # *71236*) (dCas9 KRAB): sg*Klf4i–1 (5'*
944 *CGCGCCGCCGACAGCCACAG 3'), sgKlf4i–2 (5' AGTGTCCCCCACCGTTGTCG 3'), and*
945 *sgNTCi (5' ACCTGATACGTCGTCGCGTA3').* pLV hU6-sgRNA hUbC-dCas9-KRAB-T2a-Puro
946 was a gift from Charles Gersbach (Addgene plasmid # 71236; RRID:Addgene_71236).

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

954 For infection with high titer lentiviruses, organoid cells were dissociated as described for serial
955 passaging and resuspended in 0.5mL of organoid culture media (OCM) [50% DMEM/F12 media
956 (supplemented with 10% fetal bovine serum, 1% penicillin/streptomycin, 1% L-glutamine) + 50%
957 L-WRN conditioned media (containing Wnt-3A, R-spondin 3 and Noggin) + 10uM ROCK inhibitor]
958 containing polybrene (10 µg/mL) and high-titer lentivirus. Cells were then centrifuged at 300g for
959 30 min at room temperature within a 24 well plate. Organoid cells were then resuspended in the
960 virus containing media and centrifuged at 2000 rpm for 5 min at 4°C. Virus containing media was
961 removed and stored before organoid cells were resuspended in Matrigel and plated. The virus
962 containing media was then added back to the organoid cells after Matrigel solidified. The cells
963 were then allowed 2-3 days to grow before the media was changed to contain 1 µg/mL of
964 puromycin or 2.5 µg/mL of blasticidin for the dCas9-Krab *Klf4* knockdown and *TRE-KLF4*
965 overexpression constructs, respectively.

966 **Organoid growth assays**

967 To determine differences in organoid growth rates, organoids were first dissociated to single cells
968 (see **Organoids**) then passed through a 35 µm filter to ensure single cells. Organoid cells were

969 then plated in 20 µL Matrigel domes in 48 well tissue culture plates and 200 µL of OCM media
970 per well was added to cover the domes. Media was changed every 3-4 days and for organoids
971 treated with doxycycline (doxy) (TRE-KLF4 and TRE-GFP organoids only), fresh doxy was added
972 every 48 hrs. Wells were imaged daily using a 4x objective and eight images were taken per well
973 to capture the entire Matrigel dome. Organoids were counted blindly using ImageJ and the
974 number of organoids per well was determined by combining all organoids counted in eight images.
975 Wells in which the entire plug was not captured in eight images were removed.

976 **Organoid drug treatment assays**

977 To determine differences in organoid growth rates with drug treatments, organoids were first
978 dissociated to single cells (see **Organoids**) then passed through a 35 µm filter to ensure single
979 cells. Single cell suspensions were treated with cisplatin, paclitaxel, or hydrogen peroxide at the
980 indicated doses for 15 minutes while rotating in OCM. Organoid cells were then plated in 20 µL
981 Matrigel domes in 48 well tissue culture plates and 200 µL of OCM media per well containing drug
982 was added to cover the domes. Media was changed after 3 days to OCM without drug. For
983 extended cisplatin treatment experiments, on day 7 after cisplatin treatment, individual wells were
984 treated with TrypLE to dissolve the Matrigel, organoids were spun down, then replated in 20 µL.
985 Confluency was determined using Incucyte live cell imaging (Sartorius) using the GFP
986 fluorescence to indicate live organoids (GFP is expressed by all cells in SNL tumour organoids).
987 For N-acetylcysteine (NAC) rescue experiments, single cell suspensions were treated with
988 cisplatin or vehicle for 15 minutes then plated in Matrigel. Following solidification of Matrigel, OCM
989 media containing cisplatin or vehicle with the addition of NAC was added and incubated with cells
990 for the first 3 days before media was changed to OCM without drug or NAC.

991 **Cell line growth assays and drug treatments**

992 For NCI-H520, BEAS-2B, and BCi-NS1.1 cells, cells were plated in 96 well plates and confluency
993 was monitored using Incucyte live cell imaging (Sartorius). For NCI-H226, cells were plated in 96
994 well plates and cell growth was determined compared to control conditions using the CellTiter-
995 Glo luminescence-based ATP detection assay (Promega #G7570). For cell lines expressing *TRE-*
996 *GFP* or *TRE-KLF4*, doxycycline was added on the day of plating and every 2-3 days when media
997 was refreshed. For drug treatments, either cisplatin or paclitaxel were added 1 day after plating
998 at the indicated doses.

999 **Formalin fixation of organoids**

1000 Media was removed from wells and Matrigel plugs were washed with PBS once before
1001 resuspension in 0.5 mL of PBS. Resuspend organoids were fixed using 10% formalin overnight
1002 and then transferred to 70% ethanol. Fixed organoids were then centrifuged at 2500 rpm and
1003 washed with PBS twice before being placed in 2.5% agarose for embedding. Agarose embedded
1004 organoids were then placed in 70% ethanol and submitted for embedding in paraffin.

1005 **KLF4 modulation in cell lines**

1006 For lentiviral CRISPR-mediated knockout, the following guideRNAs targeting *KLF4* were cloned
1007 into the lentiCRISPRv2 ((Addgene # 52961) (LCV2): sg*KLF4*-2 (5'-
1008 TCCCGGCCGCCCCTATGACC-3') and sg*KLF4-3* (5'- AATTGGAGAGAATAAAGTCC-3'). For
1009 overexpression of KLF4, lentiviral infection of *TRE-KLF4* was performed.

1010 **T7 surveyor assay**

1011 Genomic DNA was extracted following the protocol in the Qiagen DNeasy Blood and Tissue Kit
1012 (Cat # 69504). Following extraction, the targeted genomic regions were PCR amplified with the
1013 following sequence primers:

1014 To isolate the region surrounding the guide binding site for LCV2 *sgKLF4-2*, the following primers
1015 were used LCV2 *sgKLF4-2* FWD Primer-1: 5' ACCAAGAGCCACTGAACGAG 3'; LCV2 *sgKLF4-*
1016 *2* REV Primer-1: 5' CCGGATCGGATAGGTGAAGC 3'; LCV2 *sgKLF4-2* FWD Primer-2: 5'
1017 GTGTATGCCCGTGGTGCGA 3'; LCV2 *sgKLF4-2* REV Primer-2: 5'
1018 GATGGGTCAGCGAATTGGAGA 3'. The expected band sizes were 620 bp and 300 bp for
1019 *sgKLF4-2* primer set 1, and 550 bp and 170 bp for *sgKLF4-2* primer set 2.

1020 The PCR conditions used for both primer sets were 30 sec for 98°C, 35 cycles of 5 seconds for
1021 98°C, 10 seconds for 69°C, and 20 seconds for 72°C, 2 minutes for 72°C then held at 4°C.

1022 To isolate the region surrounding the guide binding site for LCV2 *sgKLF4-3*, the following primers
1023 were used LCV2 *sgKLF4-3* FWD Primer-1: 5' ATCTTTCTCCACGTTCGCGT 3'; LCV2 *sgKLF4-3*
1024 REV Primer-1: 5' TTGATGTCCGCCAGGTTGAA 3'; LCV2 *sgKLF4-3* FWD Primer-2: 5'
1025 CCAGCACGTCAGTATGTCGG 3'; LCV2 *sgKLF4-3* REV Primer-2: 5'
1026 CGCGTAATCACAAGTGTGGG 3'. The expected band sizes were 1000 bp and 200 bp for
1027 *sgKLF4-3* primer set 1, and 250 bp and 1000 bp for *sgKLF4-3* primer set 2.

1028 Q5 high fidelity 2x master mix (NEB Catalog # M0492) was used for the PCR reactions according
1029 to the manufacturer protocol. PCR products were then mixed with 6x loading dye and ran on 2%
1030 agarose gels. Gels were imaged to confirm correct PCR product size. The PCR products were
1031 then gel extracted using the QIAquick gel extraction kit (Qiagen #28706). DNA concentrations
1032 were measured on a Biotek Synergy HT plate reader.

1033 200ng of gel-purified DNA was used for endonuclease reaction. The T7 surveyor assay was
1034 performed according to the "Determining Genome Targeting Efficiency using T7 Endonuclease I"
1035 protocol from NEB. Briefly, 200ng of gel-purified DNA was hybridized in a thermocycler by initial
1036 denaturing at 95°C for 5 minutes followed by annealing from 95°C to 25°C. Annealed DNA was
1037 then subjected to the T7 endonuclease at 37°C for 15 min. Following the enzymatic reaction DNA
1038 was run on 2% gels and imaged.

1039 **Immunoblot**

1040 Total protein lysates were prepared as previously described, separated via SDS-PAGE, and
1041 transferred to a PVDF membrane[120]. Membranes were blocked for 1 h in 5% milk followed by
1042 overnight incubation with primary antibodies at 4°C. All primary antibodies were diluted 1:1000
1043 and include: KRT13 (Abcam Cat # 92551); KRT5 (Biolegend Cat #905501); KLF4 (CST Cat #
1044 4038s); CAS9 (Active Motif Cat # 61798); HSP90 (CST Cat # 4877); P21 (CST Cat # 2947s);
1045 PITX1 (Novus NBP1-88644); and GFP (CST Cat #2956s). Membranes were washed for three
1046 times for 5 min at room temperature in TBS-T. Mouse and rabbit HRP-conjugated secondary
1047 antibodies (1:5000; Jackson ImmunoResearch Cat # 711-035-152 and Cat # 115-035-205) were
1048 incubated for 1 hr in 5% milk at room temperature followed by washing three times for 5 minutes
1049 at room temperature in TBS-T. Membranes were exposed to WesternBright HRP quantum
1050 substrate (Advansta) and detected on Hyblot CL film (Denville Scientific, Inc.).

1051 **Immunohistochemistry (IHC)**

1052 Tissue from autochthonous mouse models was prepared by harvesting lungs inflated with 1X
1053 PBS. Individual lobes were separated and fixed in 10% neutral buffered formalin for 24 h at room

1054 temperature (RT), washed in PBS and transferred to 70% EtOH. Formalin-fixed paraffin
1055 embedded (FFPE) tissue was sectioned at 4–5 µm. For staining, tissue were dewaxed,
1056 rehydrated, and subjected to high-temperature antigen retrieval by boiling for 20 min in 0.01 M
1057 citrate buffer at pH 6.0 in a pressure cooker. Slides were quenched of endogenous peroxide in
1058 3% H_2O_2 for 15 min and blocked in 5% goat serum in PBS/0.1% Tween-20 (PBS-T) for 1 h. Slides
1059 were incubated with primary antibodies in blocking buffer (5% goat serum or SignalStain antibody
1060 diluent, Cell Signaling Technology, Cat #8112). Slides were washed 3x with PBS-T then
1061 incubated for 1 hr with secondary antibody. For all non-CST antibodies, the secondary antibody
1062 was HRP-conjugated (Vector Labs) used at a 1:200 dilution in PBS-T. Alternatively, CST primary
1063 antibodies were detected using 150 µL of SignalStain Boost IHC Detection Reagent (CST
1064 cat#8114). DAB staining was then performed, and slides were coverslipped with the automated
1065 Epredia ClearVue coverslipper. The primary antibodies include: 1:250 KRT13 (Abcam Cat #
1066 92551); 1:2000 KRT5 (BioLegend Cat # 905501); 1:500 HNF4α (CST Cat # 3113). For manual
1067 H-score quantification, whole slides were scanned using a Pannoramic Midi II automatic digital
1068 slide scanner (3DHistech). IHC quantification included all SNL tumours from all collected lung
1069 lobes, including squamous and adenocarcinoma tumours. H-score was determined by rating
1070 staining intensity per antibody on a scale from 0-3 then multiplying each intensity by the
1071 percentage of tumour cells that express the marker. For example, a tumour with 50% positive
1072 cells with high intensity of 3 and 20% positive cells with moderate intensity of 2 has a 190 H-
1073 score.

1074 For human tissue TMA IHC, staining was performed as described above. H-scoring was
1075 performed on the entire TMA spot. Any spots scored as containing no tumour were excluded from
1076 analysis.

1077 **Immunofluorescence**

1078 FFPE lung tissue (see **IHC**) and organoids (see **Formalin fixation of organoids**) were prepared
1079 as described earlier. For staining, tissue was dewaxed, rehydrated, and subjected to high-
1080 temperature antigen retrieval by boiling for 20 min in 0.01 M citrate buffer at pH 6.0 in a pressure
1081 cooker. Slides were blocked in 5% goat serum in PBS/0.1% Tween-20 (PBS-T) for 1 hr or if
1082 mouse primary antibodies were used, slides were blocked in blocking agent from the M.O.M. kit
1083 (Vector Labs Cat # BMK-2202) for 1 hr. Slides were washed 3x with PBS-T then incubated for 1
1084 hr with secondary antibody in the dark. Then slides were washed with PBS-T and incubated for 1
1085 hr with DAPI in PBS-T before slides were coverslipped using Fluoromout-G (SouthernBiotech Cat
1086 # 0100-01). The primary antibodies include: 1:250 KRT13 (Abcam Cat # 92551); 1:400 TRP63
1087 (R&D Systems Cat # AF1916); 1:100 P40 (Biocare Cat # ACI3066); 1:200 KI67 (BD Cat #
1088 556003); 1:400 NGFR (CST Cat # 8238); 1:500 KLF4 (R&D System Cat # AF3158); 1:200
1089 LGALS3 (CST Cat # 89572). Quantification of percent positive cells was performed on non-
1090 overlapping images using CellProfiler or through blinded manual counting[121].

1091 **Bulk RNA-seq sample and data processing from GEMMs**

1092 5 µm serial sections of FFPE tissue from LP and KP tumour bearing lungs, initiated with Ad-CMV-
1093 Cre, were stained with H&E and Aniline blue. H&E-slides were used to confirm histology and
1094 guide bordering of Aniline-blue-stained serial sections. Using a microdissection scope, bordered
1095 tissue samples were dissected from consecutive serial sections using sterile scalpels and razors.
1096 10-14 pieces of each sample isolated were pooled in a microcentrifuge tube containing 100%
1097 ethanol. RecoverAll Total Nucleic Acid Isolation Kit for FFPE (Thermo Fisher Cat # AM1975) was
1098 used to isolate RNA from FFPE tissues according to the manufacturer's protocol. RNA isolation
1099 from ~15 mg flash-frozen SNL (CMV-Cre initiated), RPM (10 CGRP-Cre initiated tumours, 5 Ad-
1100 CMV-Cre imitated tumours), and RPR2 (4 CGRP-Cre initiated tumours, 4 Ad-CMV-Cre initiated

1101 tumours) primary tumours was performed using the RNeasy Mini Kit (Qiagen Cat # 74104) with
1102 the standard protocol. RNA from SNL, LP, KP, RPM, and RPR2 tumours were subject to library
1103 construction with the Illumina TruSeq Stranded mRNA Sample Preparation Kit (Cat # RS-122-
1104 2101, RS-122-2102) or the Illumina TruSeq Stranded Total RNA Library Ribo-Zero Gold Prep kit
1105 (Cat # RS-122-2301), according to the manufacturer's protocol. LP and KP tumour samples were
1106 sequenced on an Illumina HiSeq 2500 instrument, 1x50 bp read lengths. RPR2, RPM, and SNL
1107 samples were sequenced on the Illumina NovaSeq 6000, 2x50 bp read lengths or the Illumina
1108 HiSeq 2500, 1x50 bp read lengths.

1109 Fastq raw count files were aligned in the R statistical environment. The mouse GRCm38 FASTA
1110 and GTF files were downloaded from Ensembl release 94 and the reference database was
1111 created using STAR[122] with splice junctions optimized for 50 base pair reads and in two pass
1112 mode to output a BAM file sorted by coordinates. Mapped reads were assigned to annotated
1113 genes in the GTF file using featureCounts version 1.6.3[123]. The output files from cutadapt,
1114 FastQC, Picard CollectRnaSeqMetrics, STAR and featureCounts were summarized using
1115 MultiQC[124] to check for any sample outliers. To remove sources of unwanted variation from
1116 tumour RNA-seq sample preparation, all non-coding features, histones, and ribosomal RNAs
1117 were removed from count matrices for downstream analyses. The featureCount output files were
1118 combined into a single raw count matrix. Log2(counts+1)-transformed, normalized intensity
1119 values were obtained following DESeq2 analysis in R to obtain gene expression levels for GEMM
1120 tumours.

1121 **Bulk RNA-seq sample and data processing from BEAS-2B**

1122 Total RNA was extracted from BEAS-2B cells with TRE-KLF4 after 48 hours of 250 ng/mL
1123 doxycycline treatment using a RNeasy mini kit (Qiagen Cat. No. 74104). RNA was processed by
1124 Medgenome and cDNA libraries were prepared using the Illumina TruSeq Stranded Total RNA
1125 kit. Libraries were sequenced on a NovaSeq 600/ X Plus with a read length of 100 aiming for at
1126 least 30M paired reads per sample. Each sample had over 70M reads and reads were aligned to
1127 the reference human genome (GRCh38/hg38) using STAR, read counts per gene were estimated
1128 using HTSeq, and read counts were normalized using DESeq2. Log2FC of differentially
1129 expressed genes between doxycycline and no doxycycline were used to perform GSEA using the
1130 python package GSEApy[125] and to perform ENRICHR analysis.

1131 **Bulk RNA-seq sample and data processing from SNL tumour organoids**

1132 Total RNA was extracted from SNL tumour organoids after digestion to single cell suspension
1133 using a RNeasy mini kit (Qiagen Cat. No. 74104). RNA was processed by the Duke Genomics
1134 Core and cDNA libraries were prepared using the Watchmaker mRNA library prep kit
1135 (Watchmaker Genomics 7BK0001). Libraries were sequenced on a NextSeq 1000 with a read
1136 length of 100 aiming for at least 30M paired reads per sample. Reads were aligned to a custom
1137 mouse genome containing GFP (GRCm38) using STAR, read counts per gene were estimated
1138 using the summarizeOverlaps function in the GenomicAlignments package of Biocondutor.
1139 Log2FC of differentially expressed genes between *sgNTCi* and *sgKlf4i-1* were used to perform
1140 GSEA using the python package GSEApy[125] and to perform ENRICHR analysis.

1141 **MicroCT imaging**

1142 SNL mice were imaged to monitor tumour development using microCT imaging beginning four
1143 weeks after tumour initiation by Cre-mediated recombination. Mice were anesthetized with
1144 isoflurane and imaged using a small animal Quantum GX2 microCT (Perkin Elmer). Quantum
1145 GX2 images were acquired with 2-minute scans at 45 mm resolution, 90 kV, with 88 mA of current.

Material availability

Human LUSC and LUAD tissue used in this study are not available due to scarcity. Tumour derived organoids have limitations on availability due to derivation from primary tumours but may be made available upon reasonable request.

Data availability

All raw and processed data have been made available through the Gene Expression Omnibus (GEO) under the superseries GSE291730. Human lung squamous cell carcinoma data from Wu et al[70] can be found at GEO under the accession code GSE148071, HNSCC data from Kurten et al[95] can be found at GEO under the accession code GSE164690, and ESCC data from Zhang et al[96] can be found at GEO under the accession code GSE160269.

Code availability

All original code has been deposited on GitHub (https://www.github.com/TGOliver-lab/Izzo_Hillock_2026) and is present on Zenodo (https://doi.org/10.5281/zenodo.18260411) which will be available at the time 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 were 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 Mann-Whitney U-test, by one-way ANOVA correcting for multiple comparisons using Tukey's Honest Significant Difference test, by two-way ANOVA with a Sidak correction for multiple comparisons, or by Kruskal-Wallis test followed by the Dunn's test with Bonferroni correction for multiple comparisons. All statistical details are further described in respective figure legends.

SUPPLEMENTARY TABLES

Supplementary Table 1: Hillock score genes for mouse and human, related to Fig. 3.

Supplementary Table 2: Basal, tuft, goblet, ciliated, and club score genes, related to Fig. 3.

Supplementary Table 3: Mucinous adenocarcinoma score genes for mouse and human, related to Fig. 3.

Supplementary Table 4: Basal hillock score genes for mouse and human, related to Fig. 3.

Supplementary Table 5: Luminal hillock score genes for mouse and human, related to Fig. 3.

Supplementary Table 6: Differentially expressed genes in SNL tumour cells derived from scRNA-seq data, related to Fig. 4 and Extended data Fig. 4.

1181 **Supplementary Table 7: Reactome pathway enrichment in luminal hillock-like SNL primary**
1182 **tumour cells, related to Fig. 4.**

1183 **Supplementary Table 8: Human patient data hillock-like cells per patient, related to Fig. 4.**

1184 **Supplementary Table 9: Differentially expressed genes in hillock-like human tumour cells**
1185 **derived from scRNA-seq data, related to Fig. 4.**

1186 **Supplementary Table 10: Immune cell population correlation to the hillock-like cell state**
1187 **by CIBERSORTx deconvolution, related to Extended Data Fig. 4.**

1188 **Supplementary Table 11: Differentially expressed genes in hillock-like tumour organoid**
1189 **cells derived from scRNA-seq data, related to Extended Data Fig. 5.**

1190 **Supplementary Table 12: ENCODE and ChEA Consensus TFs from ChIP-X from mouse**
1191 **and human hillock-like and basal-like cells, related to Extended Data Fig. 6.**

1192 **Supplementary Table 13: Differentially expressed genes in BEAS-2B cells with TRE-KLF4**
1193 **and SNL tumour organoids with sgKlf4i-1, related to Fig. 7.**

1194 **Supplementary Table 14: SEACR peaks annotated using HOMER from KLF4 CUT&RUN in**
1195 **SNL tumour organoids, related to Fig. 7.**

1196 **Supplementary Table 15: Differentially regulated genes encoding for cell surface proteins,**
1197 **related to Fig. 8.**

1198

1199 **References**

1200 1 Flavahan, W. A., Gaskell, E. & Bernstein, B. E. Epigenetic plasticity and the hallmarks of
1201 cancer. *Science* **357**, eaal2380 (2017). https://doi.org/10.1126/science.aal2380
1202 2 Hinohara, K. & Polyak, K. Intratumoral Heterogeneity: More Than Just Mutations. *Trends*
1203 *Cell Biol* **29**, 569-579 (2019). https://doi.org/10.1016/j.tcb.2019.03.003
1204 3 Ciriello, G. *et al.* Cancer Evolution: A Multifaceted Affair. *Cancer Discov* **14**, 36-48
1205 (2024). https://doi.org/10.1158/2159-8290.CD-23-0530
1206 4 Puram, S. V. *et al.* Single-Cell Transcriptomic Analysis of Primary and Metastatic Tumor
1207 Ecosystems in Head and Neck Cancer. *Cell* **171**, 1611-1624 e1624 (2017).
1208 https://doi.org/10.1016/j.cell.2017.10.044
1209 5 Li, H. *et al.* Reference component analysis of single-cell transcriptomes elucidates
1210 cellular heterogeneity in human colorectal tumors. *Nat Genet* **49**, 708-718 (2017).
1211 https://doi.org/10.1038/ng.3818
1212 6 Montoro, D. T. *et al.* A revised airway epithelial hierarchy includes CFTR-expressing
1213 ionocytes. *Nature* **560**, 319-324 (2018). https://doi.org/10.1038/s41586-018-0393-7
1214 7 Plasschaert, L. W. *et al.* A single-cell atlas of the airway epithelium reveals the CFTR-
1215 rich pulmonary ionocyte. *Nature* **560**, 377-381 (2018). https://doi.org/10.1038/s41586-
1216 018-0394-6
1217 8 Chen, Z., Fillmore, C. M., Hammerman, P. S., Kim, C. F. & Wong, K. K. Non-small-cell
1218 lung cancers: a heterogeneous set of diseases. *Nat Rev Cancer* **14**, 535-546 (2014).
1219 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. *Gene Dev* **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 Bairakdar, M. D. *et al.* Learning the cellular origins across cancers using single-cell chromatin landscapes. *Nature Communications* **16**, 8301 (2025). https://doi.org/10.1038/s41467-025-63957-3
33 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
34 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
35 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
36 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
37 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
38 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
39 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
40 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
41 Jeong, Y. *et al.* Role of KEAP1/NRF2 and TP53 Mutations in Lung Squamous Cell Carcinoma Development and Radiation Resistance. *Cancer Discovery* **7**, 86-101 (2017). https://doi.org/10.1158/2159-8290.Cd-16-0127
42 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
43 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
44 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

45 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

46 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

47 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

48 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

49 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

50 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

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

52 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

53 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

54 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

55 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

56 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

57 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

58 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

59 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

60 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

61 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

62 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

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

1371 64 Updegraff, B. L. *et al.* Transmembrane Protease TMPRSS11B Promotes Lung Cancer
1372 Growth by Enhancing Lactate Export and Glycolytic Metabolism. *Cell Rep* **25**, 2223-
1373 2233 e2226 (2018). https://doi.org/10.1016/j.celrep.2018.10.100
1374 65 Sunil, H. S. *et al.* TMPRSS11B promotes an acidified microenvironment and immune
1375 suppression in squamous lung cancer. *EMBO Reports* **0**, 1-34 (2025).
1376 https://doi.org/10.1038/s44319-025-00631-1
1377 66 Yamada, Y. *et al.* Pulmonary cancers across different histotypes share hybrid tuft
1378 cell/ionocyte-like molecular features and potentially druggable vulnerabilities. *Cell Death*
1379 *Dis* **13**, 979 (2022). https://doi.org/10.1038/s41419-022-05428-x
1380 67 Yamada, Y. *et al.* A Tuft Cell-Like Signature Is Highly Prevalent in Thymic Squamous
1381 Cell Carcinoma and Delineates New Molecular Subsets Among the Major Lung Cancer
1382 Histotypes. *J Thorac Oncol* **16**, 1003-1016 (2021).
1383 https://doi.org/10.1016/j.jtho.2021.02.008
1384 68 Jimbo, N. *et al.* The expression of YAP1 and other transcription factors contributes to
1385 lineage plasticity in combined small cell lung carcinoma. *J Pathol Clin Res* **10**, e70001
1386 (2024). https://doi.org/10.1002/2056-4538.70001
1387 69 Ireland, A. S. *et al.* Basal cell of origin resolves neuroendocrine–tuft lineage plasticity in
1388 cancer. *Nature* **647**, 257-267 (2025). https://doi.org/10.1038/s41586-025-09503-z
1389 70 Wu, F. *et al.* Single-cell profiling of tumor heterogeneity and the microenvironment in
1390 advanced non-small cell lung cancer. *Nat Commun* **12**, 2540 (2021).
1391 https://doi.org/10.1038/s41467-021-22801-0
1392 71 Tirosh, I. *et al.* Dissecting the multicellular ecosystem of metastatic melanoma by single-
1393 cell RNA-seq. *Science* **352**, 189-196 (2016). https://doi.org/doi:10.1126/science.aad0501
1394 72 Tirosh, I. & Suva, M. L. Cancer cell states: Lessons from ten years of single-cell RNA-
1395 sequencing of human tumors. *Cancer Cell* **42**, 1497-1506 (2024).
1396 https://doi.org/10.1016/j.ccell.2024.08.005
1397 73 Chan, J. M. *et al.* Signatures of plasticity, metastasis, and immunosuppression in an
1398 atlas of human small cell lung cancer. *Cancer Cell* **39**, 1479-1496.e1418 (2021).
1399 https://doi.org/https://doi.org/10.1016/j.ccell.2021.09.008
1400 74 Hai, J. *et al.* Generation of Genetically Engineered Mouse Lung Organoid Models for
1401 Squamous Cell Lung Cancers Allows for the Study of Combinatorial Immunotherapy.
1402 *Clin Cancer Res* **26**, 3431-3442 (2020). https://doi.org/10.1158/1078-0432.CCR-19-1627
1403 75 Tong, X. *et al.* Adeno-to-squamous transition drives resistance to KRAS inhibition in
1404 LKB1 mutant lung cancer. *Cancer Cell* **42**, 413-428.e417 (2024).
1405 https://doi.org/https://doi.org/10.1016/j.ccell.2024.01.012
1406 76 Shi, R. *et al.* Organoid cultures as preclinical models of non–small cell lung cancer.
1407 *Clinical Cancer Research* **26**, 1162-1174 (2020).
1408 77 Kim, M. *et al.* Patient-derived lung cancer organoids as in vitro cancer models for
1409 therapeutic screening. *Nature Communications* **10**, 3991 (2019).
1410 https://doi.org/10.1038/s41467-019-11867-6
1411 78 Romero, R. *et al.* The neuroendocrine transition in prostate cancer is dynamic and
1412 dependent on ASCL1. *Nature Cancer* **5**, 1641-1659 (2024).
1413 https://doi.org/10.1038/s43018-024-00838-6
1414 79 Moreci, R. S. & Lechler, T. Epidermal structure and differentiation. *Curr Biol* **30**, R144-
1415 R149 (2020). https://doi.org/10.1016/j.cub.2020.01.004
1416 80 He, H. *et al.* Kruppel-like Factor 4 Promotes Esophageal Squamous Cell Carcinoma
1417 Differentiation by Up-regulating Keratin 13 Expression. *J Biol Chem* **290**, 13567-13577
1418 (2015). https://doi.org/10.1074/jbc.M114.629717
1419 81 Ali, A. *et al.* Cell Lineage Tracing Identifies Hormone-Regulated and Wnt-Responsive
1420 Vaginal Epithelial Stem Cells. *Cell Rep* **30**, 1463-1477.e1467 (2020).
1421 https://doi.org/10.1016/j.celrep.2020.01.003

1422 82 Thakore, P. I. *et al.* Highly specific epigenome editing by CRISPR-Cas9 repressors for
1423 silencing of distal regulatory elements. *Nat Methods* **12**, 1143-1149 (2015).
1424 https://doi.org/10.1038/nmeth.3630
1425 83 Zhang, W. *et al.* The gut-enriched Krüppel-like factor (Krüppel-like factor 4) mediates the
1426 transactivating effect of p53 on the p21 WAF1/Cip1 promoter. *J Biol Chem* **275**, 18391-
1427 18398 (2000). https://doi.org/10.1074/jbc.C000062200
1428 84 Rowland, B. D., Bernards, R. & Peeper, D. S. The KLF4 tumour suppressor is a
1429 transcriptional repressor of p53 that acts as a context-dependent oncogene. *Nature Cell*
1430 *Biology* **7**, 1074-1082 (2005). https://doi.org/10.1038/ncb1314
1431 85 Rowland, B. D. & Peeper, D. S. KLF4, p21 and context-dependent opposing forces in
1432 cancer. *Nat Rev Cancer* **6**, 11-23 (2006). https://doi.org/10.1038/nrc1780
1433 86 Prescott, R. A. *et al.* A comparative study of in vitro air-liquid interface culture models of
1434 the human airway epithelium evaluating cellular heterogeneity and gene expression at
1435 single cell resolution. *Respir Res* **24**, 213 (2023). https://doi.org/10.1186/s12931-023-
1436 02514-2
1437 87 Walters, M. S. *et al.* Generation of a human airway epithelium derived basal cell line with
1438 multipotent differentiation capacity. *Respir Res* **14**, 135 (2013).
1439 https://doi.org/10.1186/1465-9921-14-135
1440 88 Zhao, F. & Klimecki, W. T. Culture conditions profoundly impact phenotype in BEAS-2B,
1441 a human pulmonary epithelial model. *J Appl Toxicol* **35**, 945-951 (2015).
1442 https://doi.org/10.1002/jat.3094
1443 89 Wang, Z. *et al.* Nrf2 Drives Epigenetic Reprogramming and Acts as the Master
1444 Regulator of KLF4 Expression and Activity in Arsenic-Induced Transformation.
1445 *Advanced Science* **12**, e00221 (2025).
1446 https://doi.org/https://doi.org/10.1002/advs.202500221
1447 90 Torrente, L. & DeNicola, G. M. Targeting NRF2 and Its Downstream Processes:
1448 Opportunities and Challenges. *Annu Rev Pharmacol Toxicol* **62**, 279-300 (2022).
1449 https://doi.org/10.1146/annurev-pharmtox-052220-104025
1450 91 Pillai, R., Hayashi, M., Zavitsanou, A. M. & Papagiannakopoulos, T. NRF2: KEAPing
1451 Tumors Protected. *Cancer Discov* **12**, 625-643 (2022). https://doi.org/10.1158/2159-
1452 8290.Cd-21-0922
1453 92 Lin, E. W. *et al.* Comparative transcriptomes of adenocarcinomas and squamous cell
1454 carcinomas reveal molecular similarities that span classical anatomic boundaries. *Plos*
1455 *Genet* **13**, e1006938 (2017). https://doi.org/10.1371/journal.pgen.1006938
1456 93 Yan, W., Wistuba, II, Emmert-Buck, M. R. & Erickson, H. S. Squamous Cell Carcinoma -
1457 Similarities and Differences among Anatomical Sites. *Am J Cancer Res* **1**, 275-300
1458 (2011).
1459 94 Henick, B. S. *et al.* Squamous cell cancers of the aero-upper digestive tract: A unified
1460 perspective on biology, genetics, and therapy. *Cancer Cell* **43**, 178-194 (2025).
1461 https://doi.org/https://doi.org/10.1016/j.ccell.2025.01.003
1462 95 Kürten, C. H. L. *et al.* Investigating immune and non-immune cell interactions in head
1463 and neck tumors by single-cell RNA sequencing. *Nature Communications* **12**, 7338
1464 (2021). https://doi.org/10.1038/s41467-021-27619-4
1465 96 Zhang, X. *et al.* Dissecting esophageal squamous-cell carcinoma ecosystem by single-
1466 cell transcriptomic analysis. *Nature Communications* **12**, 5291 (2021).
1467 https://doi.org/10.1038/s41467-021-25539-x
1468 97 Snarr, B. D. *et al.* Galectin-3 enhances neutrophil motility and extravasation into the
1469 airways during Aspergillus fumigatus infection. *PLoS Pathog* **16**, e1008741 (2020).
1470 https://doi.org/10.1371/journal.ppat.1008741
1471 98 Gordon-Alonso, M., Hirsch, T., Wildmann, C. & van der Bruggen, P. Galectin-3 captures
1472 interferon-gamma in the tumor matrix reducing chemokine gradient production and T-cell

tumor infiltration. *Nature Communications* **8**, 793 (2017). https://doi.org/10.1038/s41467-017-00925-6

99 Girotti, M. R., Salatino, M., Dalotto-Moreno, T. & Rabinovich, G. A. Sweetening the hallmarks of cancer: Galectins as multifunctional mediators of tumor progression. *J Exp Med* **217** (2020). https://doi.org/10.1084/jem.20182041

100 Rabinovich, G. A., Gabrilovich, D. & Sotomayor, E. M. Immunosuppressive strategies that are mediated by tumor cells. *Annu Rev Immunol* **25**, 267-296 (2007). https://doi.org/10.1146/annurev.immunol.25.022106.141609

101 Vuong, L. *et al.* An Orally Active Galectin-3 Antagonist Inhibits Lung Adenocarcinoma Growth and Augments Response to PD-L1 Blockade. *Cancer Research* **79**, 1480-1492 (2019). https://doi.org/10.1158/0008-5472.Can-18-2244

102 Hirani, N. *et al.* Target inhibition of galectin-3 by inhaled TD139 in patients with idiopathic pulmonary fibrosis. *European Respiratory Journal* **57**, 2002559 (2021). https://doi.org/10.1183/13993003.02559-2020

103 Hammerman, P. S. *et al.* Comprehensive genomic characterization of squamous cell lung cancers. *Nature* **489**, 519-525 (2012). https://doi.org/10.1038/nature11404

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 Newman, A. M. *et al.* Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods* **12**, 453-457 (2015). https://doi.org/10.1038/nmeth.3337

109 Newman, A. M. *et al.* Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat Biotechnol* **37**, 773-782 (2019). https://doi.org/10.1038/s41587-019-0114-2

110 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

111 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

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

113 Bausch-Fluck, D. *et al.* A Mass Spectrometric-Derived Cell Surface Protein Atlas. *PLOS ONE* **10**, e0121314 (2015). https://doi.org/10.1371/journal.pone.0121314

114 Bausch-Fluck, D. *et al.* The in silico human surfaceome. *Proceedings of the National Academy of Sciences* **115**, E10988-E10997 (2018). https://doi.org/doi:10.1073/pnas.1808790115

115 Cheshire, C. *et al. nf-core/cutandrun: nf-core/cutandrun v3.2.2 Iridium Ibis*. (Zenodo, 2024).

116 Ewels, P. A. *et al.* The nf-core framework for community-curated bioinformatics pipelines. *Nature Biotechnology* **38**, 276-278 (2020). https://doi.org/10.1038/s41587-020-0439-x

1522 117 Meers, M. P., Tenenbaum, D. & Henikoff, S. Peak calling by Sparse Enrichment
1523 Analysis for CUT&RUN chromatin profiling. *Epigenetics & Chromatin* **12**, 42 (2019).
1524 https://doi.org/10.1186/s13072-019-0287-4
1525 118 Duttke, S. H., Chang, M. W., Heinz, S. & Benner, C. Identification and dynamic
1526 quantification of regulatory elements using total RNA. *Genome Res* **29**, 1836-1846
1527 (2019). https://doi.org/10.1101/gr.253492.119
1528 119 Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime
1529 cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* **38**, 576-
1530 589 (2010). https://doi.org/10.1016/j.molcel.2010.05.004
1531 120 Oliver, T. G. *et al.* Caspase-2-mediated cleavage of Mdm2 creates a p53-induced
1532 positive feedback loop. *Mol Cell* **43**, 57-71 (2011).
1533 https://doi.org/10.1016/j.molcel.2011.06.012
1534 121 Stirling, D. R. *et al.* CellProfiler 4: improvements in speed, utility and usability. *BMC*
1535 *Bioinformatics* **22**, 433 (2021). https://doi.org/10.1186/s12859-021-04344-9
1536 122 Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21
1537 (2013). https://doi.org/10.1093/bioinformatics/bts635
1538 123 Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for
1539 assigning sequence reads to genomic features. *Bioinformatics* **30**, 923-930 (2014).
1540 https://doi.org/10.1093/bioinformatics/btt656
1541 124 Ewels, P., Magnusson, M., Lundin, S. & Kaller, M. MultiQC: summarize analysis results
1542 for multiple tools and samples in a single report. *Bioinformatics* **32**, 3047-3048 (2016).
1543 https://doi.org/10.1093/bioinformatics/btw354
1544 125 Fang, Z., Liu, X. & Peltz, G. GSEApy: a comprehensive package for performing gene set
1545 enrichment analysis in Python. *Bioinformatics* **39** (2022).
1546 https://doi.org/10.1093/bioinformatics/btac757

1547

Figure 1

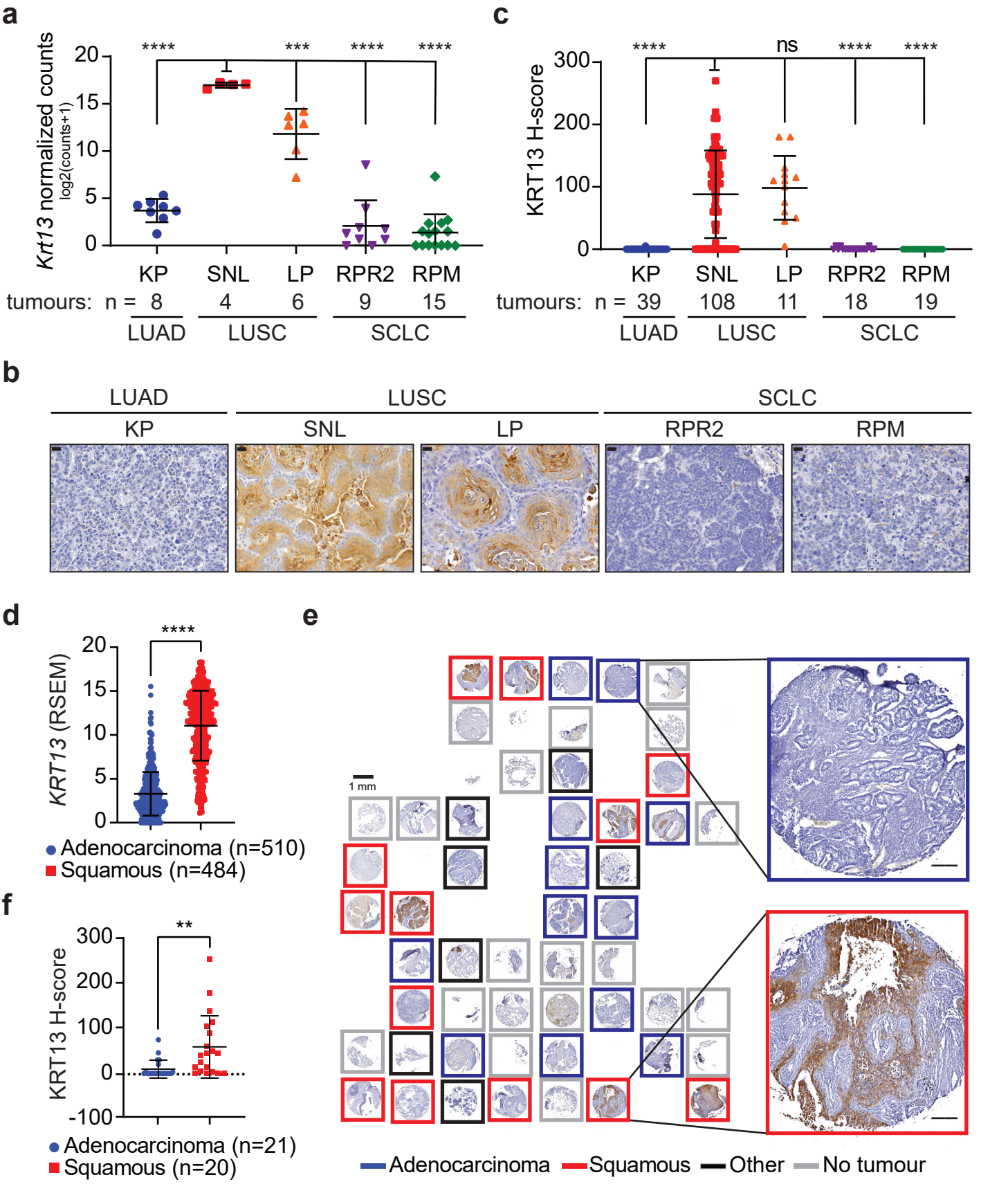
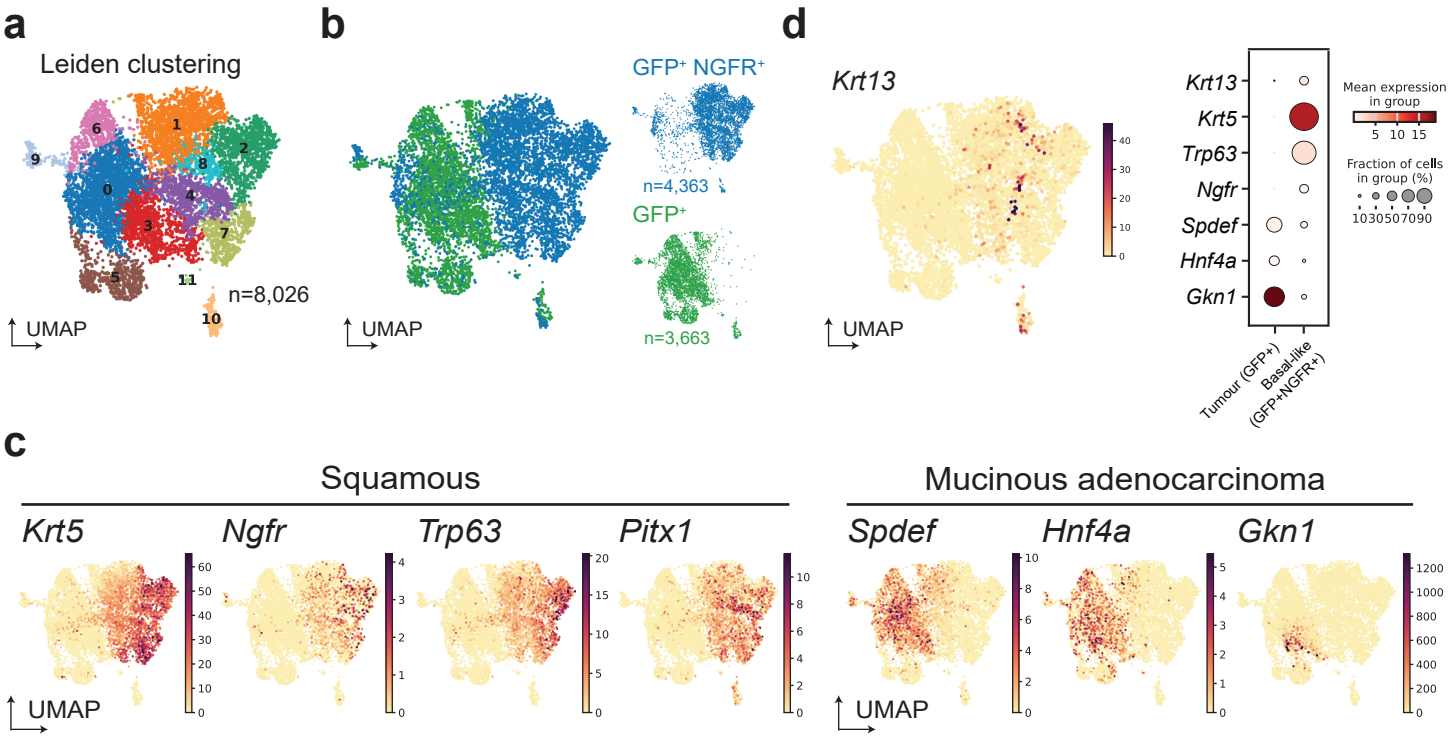
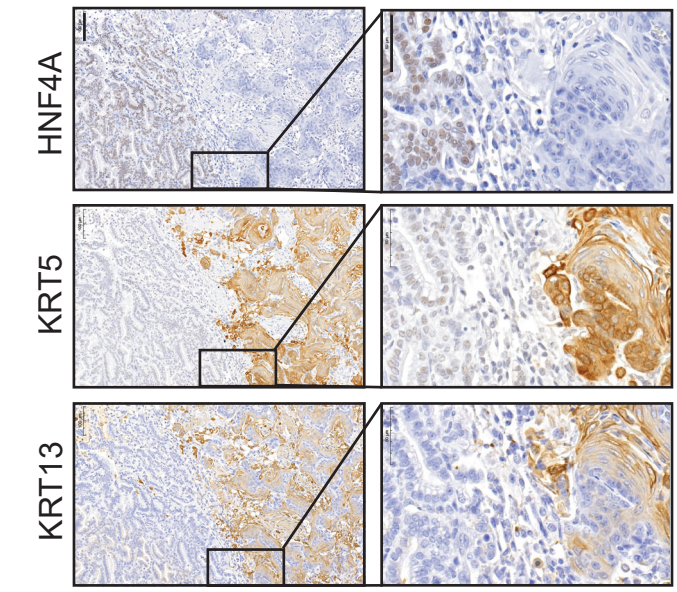
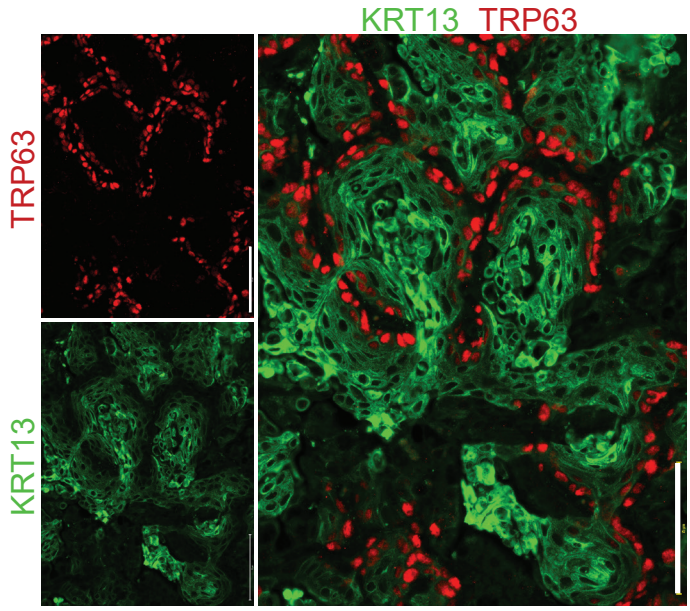

a
Krt13 normalized counts
log2(counts+1)
****
***
****
****
KP
SNL
LP
RPR2
RPM
tumours: n = 8 4 6 9 15
LUAD
LUSC
SCLC
c
KRT13 H-score
****
ns
****
****
tumours: n = 39 108 11 18 19
b
LUAD
LUSC
SCLC
d
KRT13 (RSEM)
****
Adenocarcinoma (n=510)
Squamous (n=484)
e
1 mm
Adenocarcinoma
Squamous
Other
No tumour
f
KRT13 H-score
**
Adenocarcinoma (n=21)
Squamous (n=20)

Figure 1. KRT13 is highly enriched in lung squamous cell carcinoma. a) *Krt13* expression shown as log2 normalized counts by RNA-seq from lung tumours in indicated GEMMs grouped according to histological type. Tumours 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 tumours 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 tumours indicated in figure. Two-tailed t-test. **e)** Representative (1 of 2 TMA slides) KRT13 IHC in human lung tumour tissue microarrays (red = squamous cell carcinoma; blue = adenocarcinoma; black = NSCLC/large cell, grey = section without tumour tissue or not enough material to score). 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 tumours indicated. Two-tailed t-test. **For all panels**, each point represents an individual tumour, and error bars represent mean +/- SD. ns = not significant; **p ≤ 0.01; ***p ≤ 0.001; ****p ≤ 0.0001.

Figure 2

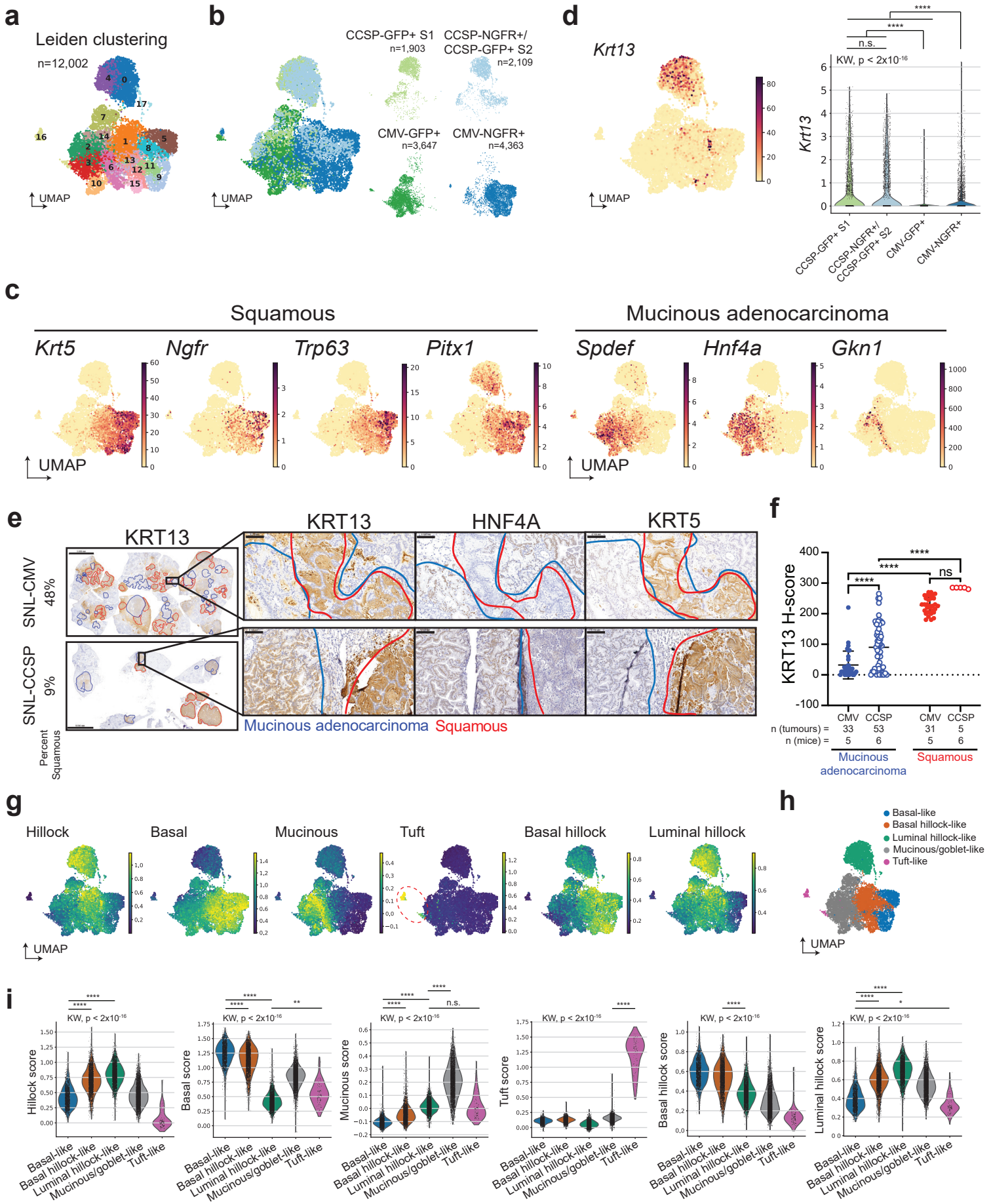

a
Leiden clustering
n=8,026
UMAP
b
GFP+ NGFR+
n=4,363
GFP+
n=3,663
UMAP
d
Krt13
UMAP
Krt13
Krt5
Trp63
Ngfr
Spdef
Hnf4a
Gkn1
Mean expression in group
Fraction of cells in group (%)
Tumour (GFP+)
Basal-like (GFP+NGFR+)
c
Squamous
Mucinous adenocarcinoma
Krt5
Ngfr
Trp63
Pitx1
Spdef
Hnf4a
Gkn1
UMAP


e
HNF4A
KRT5
KRT13


f
KRT13 TRP63
TRP63
KRT13

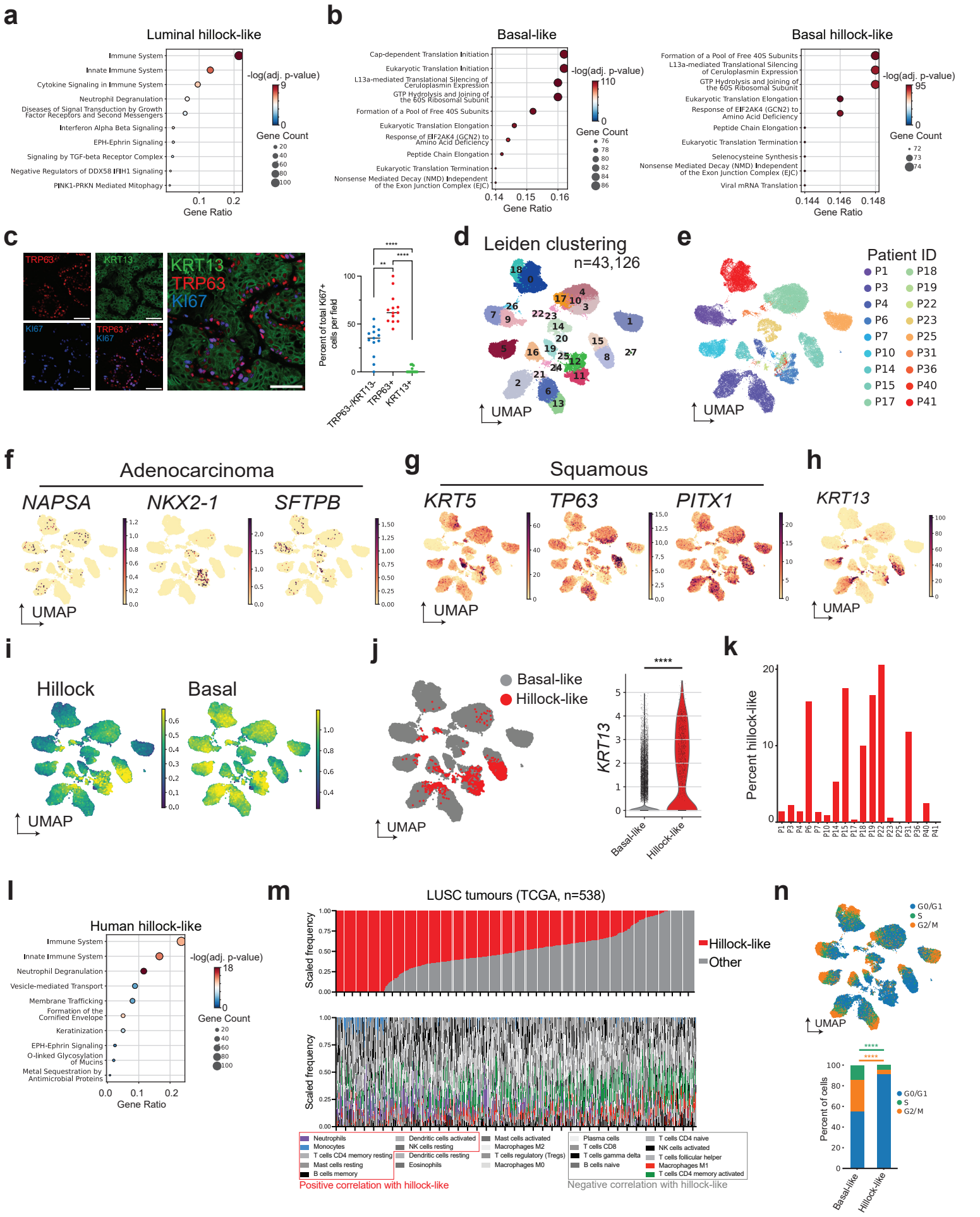
1570 **Figure 2. KRT13 is expressed in a non-basal subpopulation of LUSC. a)** ScRNA-seq of Ad-
1571 CMV-SNL tumour cells labeled by Leiden clustering in UMAP space. **b)** Ad-CMV-SNL tumour
1572 cells labeled by sorted population (GFP^+ or $GFP^+/NGFR^+$) in UMAP space as in **2a**. **c)** Expression
1573 of indicated squamous or mucinous adenocarcinoma gene markers in UMAP space from cells in
1574 **2a**. **d)** ScRNA-seq expression of *Krt13* projected on UMAP space from cells in **2a** and dotplot of
1575 gene expression from scRNA-seq separated by flow sort. Dot size indicates percent of cells
1576 expressing each transcript and the color indicates the average expression per cluster indicated in
1577 the figure. **e)** Representative IHC from serial sections in Ad-CMV-SNL-tumour-bearing lungs
1578 harbouring both squamous cell carcinoma and mucinous adenocarcinoma. Representative of n =
1579 6 Ad-CMV-SNL mice. Scale bar, 100 μm for left column; 50 μm for high-magnification insets on
1580 right column. **f)** Representative immunofluorescence co-staining for KRT13 (green) and TRP63
1581 (red) in Ad-CMV-SNL tumours. Representative of n = 4 Ad-CMV-SNL mice. Scale bars, 75 μm.
1582 **See also Extended Data Fig. 1.**
1583
1584

Figure 3


a
Leiden clustering
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
UMAP
KW, p < 2x10-16
n.s.
CCSP-GFP+ S1
CCSP-NGFR+/ CCSP-GFP+ S2
CMV-GFP+
CMV-NGFR+
c
Squamous
Mucinous adenocarcinoma
Krt5
Ngfr
Trp63
Pitx1
Spdef
Hnf4a
Gkn1
UMAP
e
KRT13
KRT13
HNF4A
KRT5
SNL-CMV
48%
SNL-CCSP
9%
Percent Squamous
Mucinous adenocarcinoma
Squamous
f
KRT13 H-score
ns
CMV
CCSP
n (tumours) =
n (mice) =
Mucinous adenocarcinoma
Squamous
g
Hillock
Basal
Mucinous
Tuft
Basal hillock
Luminal hillock
UMAP
h
Basal-like
Basal hillock-like
Luminal hillock-like
Mucinous/goblet-like
Tuft-like
UMAP
i
Hillock score
Basal score
Mucinous score
Tuft score
Basal hillock score
Luminal hillock score
KW, p < 2x10-16
n.s.
Basal-like
Basal hillock-like
Luminal hillock-like
Mucinous/goblet-like
Tuft-like

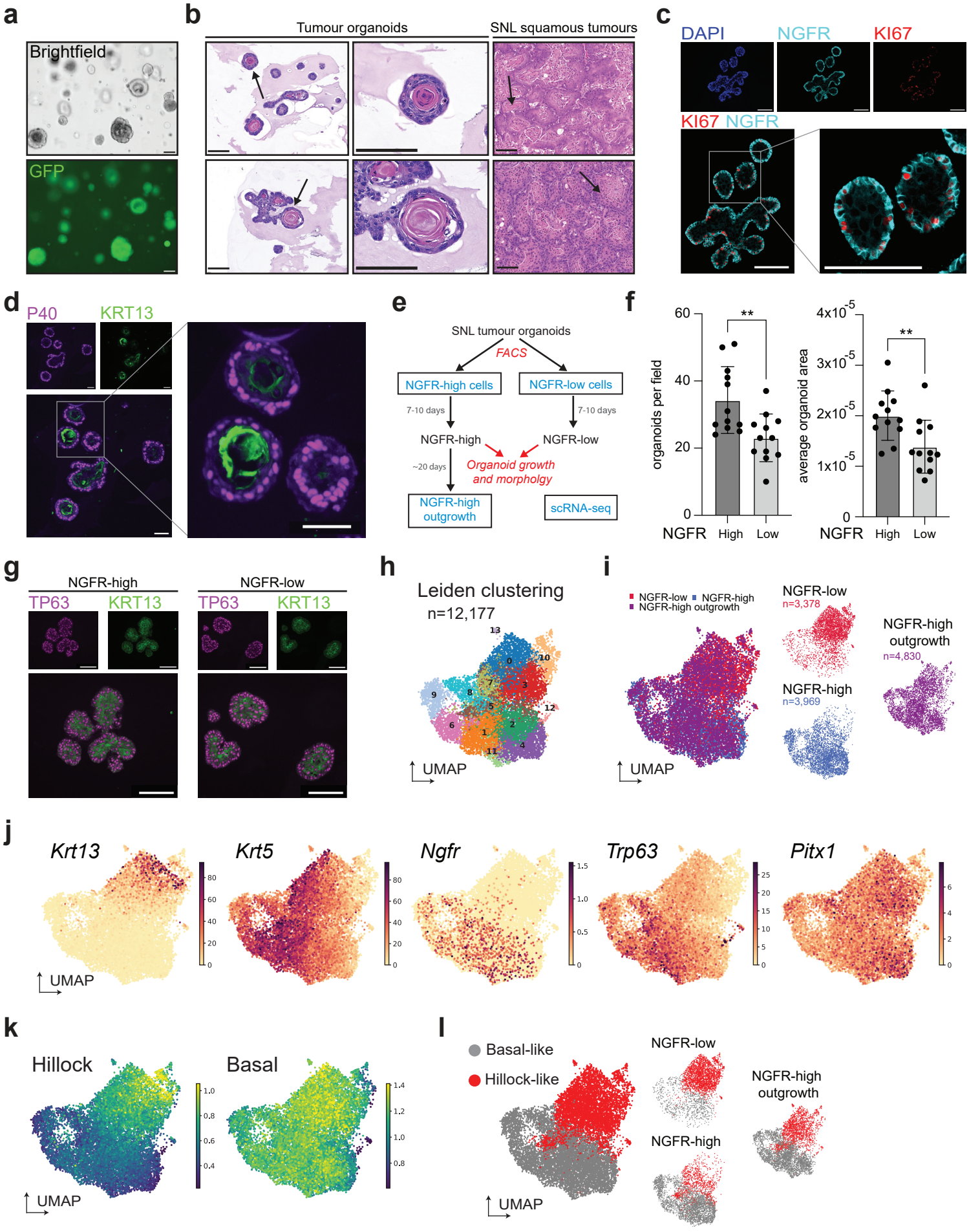
1585 **Figure 3. Club-initiated tumours have early and enriched KRT13 expression. a)** ScRNA-seq
1586 UMAP of primary Ad-CMV- and Ad-CCSP-SNL tumour cells labeled by Leiden clustering. **b)** Ad-
1587 CMV- and Ad-CCSP-initiated SNL tumour cells labeled by sorted population (GFP^+ or
1588 $GFP^+/NGFR^+$) in UMAP space as in **3a**. **c)** Expression of indicated squamous or mucinous
1589 adenocarcinoma gene markers in UMAP space from cells in **3a**. **d)** scRNA-seq expression of
1590 *Krt13* projected on UMAP space from cells in **3a** and violin plots for *Krt13* expression by sample.
1591 Kruskal-Wallis followed by Dunn's test with Bonferroni correction, not all comparisons shown. **e)**
1592 Representative KRT5, HNF4A, and KRT13 IHC for Ad-CMV- and Ad-CCSP-initiated SNL
1593 tumours. Squamous (red) and mucinous adenocarcinoma (blue) tumour lesions are outlined.
1594 Scale bar, 5 mm for left panels; 2 mm for high-magnification insets on serial sections in right three
1595 panels. Representative of n = 5 Ad-CMV-SNL and n = 6 Ad-CCSP-SNL mice. **f)** H-score
1596 quantification of IHC in **3e**. Two-way ANOVA with a Sidak correction for multiple comparisons.
1597 Not all comparisons are shown. Each point represents an individual tumour, and error bars
1598 represent mean +/- SD. ns = not significant; ****$p \leq 0.0001$. **g)** Gene scores applied to SNL tumour
1599 cells projected onto UMAP space from **3a** (see **Methods** and **Supplementary Tables 1-5**). **h)**
1600 Cell type classification based on gene and gene signature score expression from cells in **3a**. **i)**
1601 Violin plots of gene score expression by cell type assignment. Kruskal-Wallis followed by Dunn's
1602 test with Bonferroni correction, all comparison $p < 0.0001$ unless otherwise shown.
1603 **See also Extended Data Fig. 2 and 3.**
1604
1605

Figure 4


a
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
b
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)
Basal hillock-like
Viral mRNA Translation
c
TRP63
KRT13
KI67
Percent of total Ki67+ cells per field
TRP63-/KRT13-
TRP63+
KRT13+
d
Leiden clustering
n=43,126
UMAP
e
Patient ID
P1
P3
P4
P6
P7
P10
P14
P15
P17
P18
P19
P22
P23
P25
P31
P36
P40
P41
f
Adenocarcinoma
NAPSA
NKX2-1
SFTPB
g
Squamous
KRT5
TP63
PITX1
h
KRT13
i
Hillock
Basal
j
Basal-like
Hillock-like
k
Percent hillock-like
l
Human hillock-like
m
LUSC tumours (TCGA, n=538)
Scaled frequency
Hillock-like
Other
Neutrophils
Monocytes
T cells CD4 memory resting
Mast cells resting
B cells memory
Dendritic cells activated
NK cells resting
Dendritic cells resting
Eosinophils
Mast cells activated
Macrophages M2
T cells regulatory (Tregs)
Macrophages M0
Plasma cells
T cells CD8
T cells gamma delta
B cells naive
T cells CD4 naive
NK cells activated
T cells follicular helper
Macrophages M1
T cells CD4 memory activated
Positive correlation with hillock-like
Negative correlation with hillock-like
n
G0/G1
S
G2/M
Percent of cells

1606 **Figure 4. KRT13 marks a hillock-like state in squamous tumours. a,b)** Gene set enrichment
1607 analysis comparing SNL tumour cells in **3a** by assigned cell type. The ENRICHR gene set used
1608 was 'Reactome Pathways 2024'. **c)** Representative immunofluorescence (IF) co-staining for
1609 KRT13 (green), TRP63 (red), and KI67 (blue) in SNL squamous tumours and quantification of
1610 percent of total Ki67+ cells by co-expression with other marker genes. Representative of n = 4
1611 Ad-CMV-SNL mice. Scale bars, 50 µm. One-way ANOVA. **d)** UMAP of scRNA-seq from lung
1612 squamous tumour cells from 18 patients[70] labeled by Leiden clustering. **e)** UMAP from **4d** labeled
1613 by patient (P) sample ID (n = 18). **f-h)** scRNA-seq expression of adenocarcinoma (*NAPSA, NKX2-*
1614 *1, SFTPB*), squamous (*TP63, KRT5, PITX1*), and hillock markers (*KRT13*) in UMAP space from
1615 cells in **4d**. **i)** Human hillock score and human basal score (**Supplementary Tables 1** and **2**)
1616 applied to tumour cells as in **4d** with score projected in UMAP. **j)** Cell type classification based on
1617 hillock score expression from cells in **4d** and KRT13 expression by cell type. Mann-Whitney U
1618 test; ****$p \leq 0.0001$. **k)** Percent of basal- and hillock-like cells by patient ID as in **4j**. **l)** Gene set
1619 enrichment analysis comparing human tumour cells in **4d** and **4j** by assigned cell type. The
1620 ENRICHR gene set used was 'Reactome Pathways 2024'. **m)** CIBERSORTx deconvolution of
1621 human LUSC tumours ordered from most to least hillock-like (top); CIBERSORTx deconvolution
1622 of leukocyte populations in LUSC tumours ordered same as above (bottom), legend is ordered
1623 by correlation with the hillock-like state (see **Methods** and **Supplementary Table 10**). **n)** Cell
1624 cycle assignments applied to human tumour cells in UMAP space as in **4d** (see **Methods**), and
1625 stacked bar plot showing percent of cells occupying each cell cycle phase by cell state assignment
1626 as in **4j.** Statistics shown compare G2/M (orange) or S-phase (green) gene scores between
1627 assigned cell states. Mann-Whitney U-Test. ****$p \leq 0.0001$.
1628 **See also Extended Data Fig. 4 and Supplementary Tables 1-10.**
1629
1630

Figure 5


a
Brightfield
GFP
b
Tumour organoids
SNL squamous tumours
c
DAPI
NGFR
KI67
KI67 NGFR
d
P40
KRT13
e
SNL tumour organoids
FACS
NGFR-high cells
NGFR-low cells
7-10 days
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 clustering
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
k
Hillock
Basal
UMAP
l
Basal-like
Hillock-like
NGFR-low
NGFR-high
NGFR-high outgrowth
UMAP

1631 **Figure 5. The hillock-like state can arise from NGFR$^+$ basal-like tumour cells. a)**
1632 Representative brightfield and immunofluorescence images of tumour organoids established from
1633 GFP$^+$NGFR$^+$ SNL tumour cells. Scale bar, 100 µm. **b)** Representative H&E staining of SNL tumour
1634 organoids (n = 2 independent cultures) and SNL squamous tumours (n = 6 mice). Arrows indicate
1635 keratinized structures. Scale bars, 100 µm. **c)** Representative immunofluorescence (IF) co-
1636 staining for KI67 (red) and NGFR (cyan) in SNL tumour organoids (n = 3). Scale bar, 100 µm. **D)**
1637 Representative IF co-staining for P40 (purple) and KRT13 (green) in SNL tumour organoids (n =
1638 3). Scale bar, 50 µm. **e)** Experimental scheme for organoid sorting and scRNA-seq. Populations
1639 labeled in blue were used for scRNA-seq. **f)** Number of organoids and average organoid area
1640 after SNL tumour organoid cells were sorted by NGFR status and replated in Matrigel and allowed
1641 to grow for 10 days. Each point represents an image taken at 4x magnification, and error bars
1642 represent mean +/- SD. Two-tailed t-test, **$p \leq 0.01$. **g**) Representative IF co-staining for TP63
1643 (purple) and KRT13 (green) from SNL tumour organoids allowed to grow for 10 days after sorting.
1644 Representative of n = 2 experiments with 6 different 20x fields per experiment. **h)** Leiden
1645 clustering of scRNA-seq performed on SNL tumour organoids as in **5e**. **i)** scRNA-seq of SNL
1646 tumour organoids at time of NGFR sorting (NGFR-high and NGFR-low) and NGFR-high cells after
1647 four weeks of growth in organoid conditions ("NGFR-high outgrowth") in **5h**. **j)** Expression of
1648 indicated hillock (*Krt13*) or basal (*Ngfr*, *Krt5*, *Trp63, Pitx1*) genes in UMAP space from cells in **5h**.
1649 **k)** scRNA-seq expression of the hillock score and basal score projected on UMAP space from
1650 cells in **5h. l)** Cell type classification by Leiden clustering based on gene and gene signature score
1651 expression from cells in **5h**. (see **Methods** and **Supplementary Table 1 and 2**).
1652 **See also Extended Data Fig. 5.**
1653
1654

Figure 6

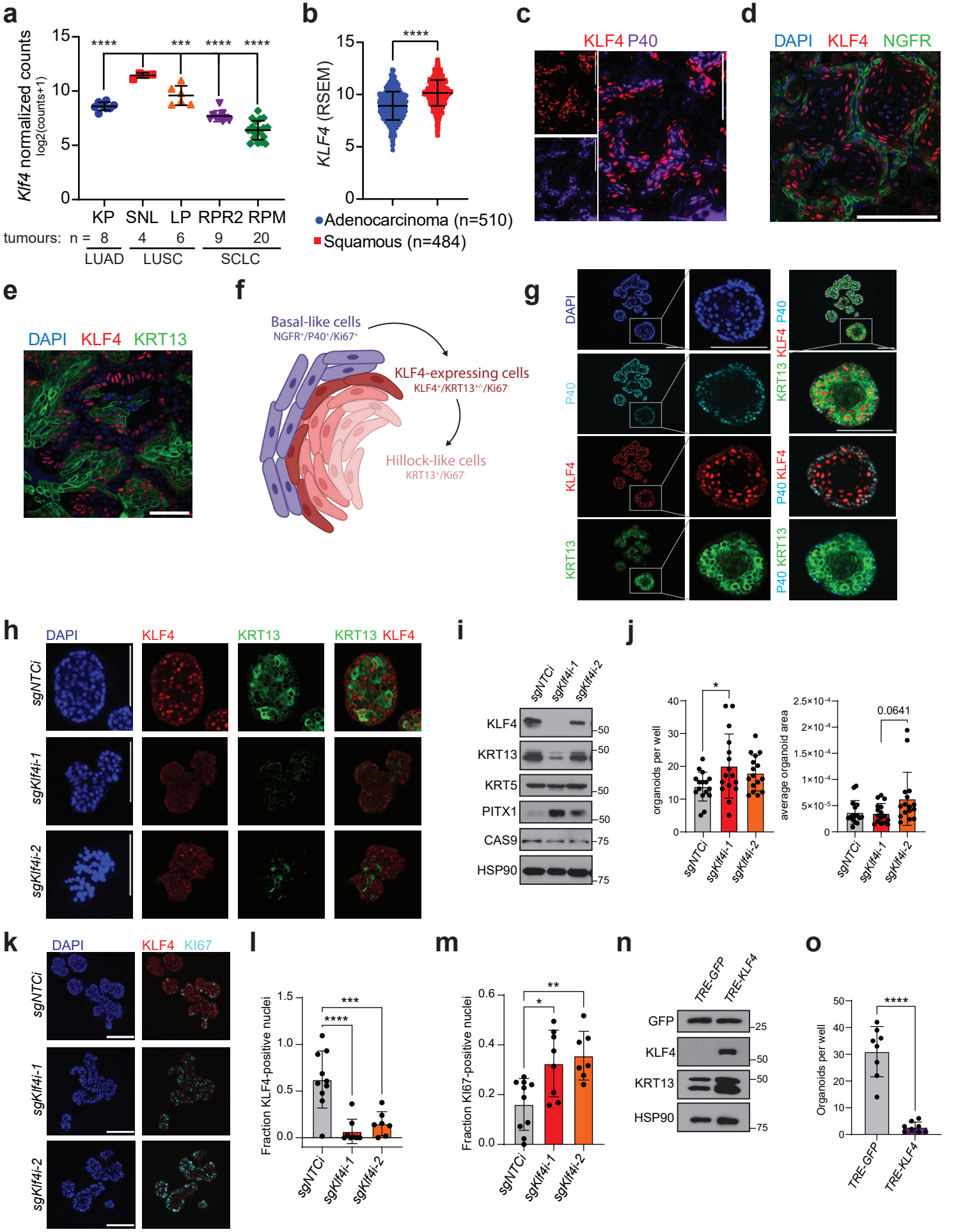

a
Klf4 normalized counts
log2(counts+1)
15
10
5
0
****
***
****
****
KP
SNL
LP
RPR2
RPM
tumours: n = 8 4 6 9 20
LUAD
LUSC
SCLC
b
KLF4 (RSEM)
15
10
5
0
****
Adenocarcinoma (n=510)
Squamous (n=484)
c
KLF4 P40
d
DAPI KLF4 NGFR
e
DAPI KLF4 KRT13
f
Basal-like cells
NGFR+/P40+/Ki67+
KLF4-expressing cells
KLF4+/KRT13+/-/Ki67-
Hillock-like cells
KRT13+/Ki67-
g
DAPI
P40
KLF4
KRT13
P40 KLF4
KRT13
P40 KLF4
P40 KRT13
h
DAPI
KLF4
KRT13
KRT13 KLF4
sgNTCi
sgKlf4i-1
sgKlf4i-2
i
sgNTCi
sgKlf4i-1
sgKlf4i-2
KLF4
KRT13
KRT5
PITX1
CAS9
HSP90
50
50
50
50
75
75
j
organoids per well
40
30
20
10
0
*
sgNTCi
sgKlf4i-1
sgKlf4i-2
average organoid area
2.5×10-4
2×10-4
1.5×10-4
1×10-4
5×10-5
0
0.0641
sgNTCi
sgKlf4i-1
sgKlf4i-2
k
DAPI
KLF4 KI67
sgNTCi
sgKlf4i-1
sgKlf4i-2
l
Fraction KLF4-positive nuclei
1.5
1.0
0.5
0.0
***
****
sgNTCi
sgKlf4i-1
sgKlf4i-2
m
Fraction KI67-positive nuclei
0.6
0.4
0.2
0.0
**
*
sgNTCi
sgKlf4i-1
sgKlf4i-2
n
TRE-GFP
TRE-KLF4
GFP
KLF4
KRT13
HSP90
25
50
50
75
o
Organoids per well
50
40
30
20
10
0
****
TRE-GFP
TRE-KLF4

Figure 6. KLF4 is enriched in a suprabasal hillock-like state in squamous tumours and regulates KRT13 expression. a) Log2 normalized counts of *Klf4* from RNA-seq of lung tumours from indicated GEMMs with number of tumours 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 tumours from Ad-CMV-initiated SNL tumours. Representative of n = 2 mice. Scale bar, 100 μm. **d)** IF staining for NGFR (green) and KLF4 (red) in squamous tumours from Ad-CMV-initiated SNL tumours. Representative of n= 2 mice. Scale bars, 100 μm. **e)** Representative IF co-staining for KRT13 (green) and KLF4 (red) in squamous tumours from Ad-CMV-initiated SNL tumours. Representative of n= 3 mice. Scale bars, 50 μm. **f)** Schematic of a squamous tumour 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 tumour organoids. Scale bars, 100 μm. Representative of n = 3 experiments with at least 5 different 20x fields per experiment. **h)** Representative IF co-staining for KRT13 (green) and KLF4 (red) in SNL tumour organoids with KLF4 knockdown using CRISPRi. Scale bars, 100 μm. Representative of n = 2 experiments at least 4 different 20x fields per experiment. **i)** Immunoblot for indicated proteins from SNL tumour organoids after 10 days of growth following dissociation to single cells and replating in Matrigel. **j)** Number and average size of organoids after 7 days of growth following dissociation to single cells and replating in Matrigel. One-way ANOVA. **k)** Representative IF co-staining for KI67 (cyan) and KLF4 (red) in SNL tumour organoids with KLF4 knockdown using CRISPRi. Scale bars, 100 μm. Representative of n = 2 experiment with at least 7 different 20x fields per sample. **l-m)** Quantification of nuclei staining positive for KLF4 (**l**) or KI67 (**m**) from **6k**. One-way ANOVA. **n)** 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. **o)** Number and average size of SNL tumour organoids after 7 days of growth following dissociation to single cells and replating in Matrigel and treated with doxycycline as in **6n**. Two-tailed t-test. **For all panels**, each data point represents an individual tumour or a 4x (organoid growth)/20x (IF quantification) microscopy field of view and error bars represent mean +/- SD; *p ≤ 0.05;**p ≤ 0.01; ***p ≤ 0.001; ****p ≤ 0.0001. DAPI (blue) labels nuclei. HSP90 serves as loading control for immunoblots.

See also Extended Data Fig. 6.

Figure 7

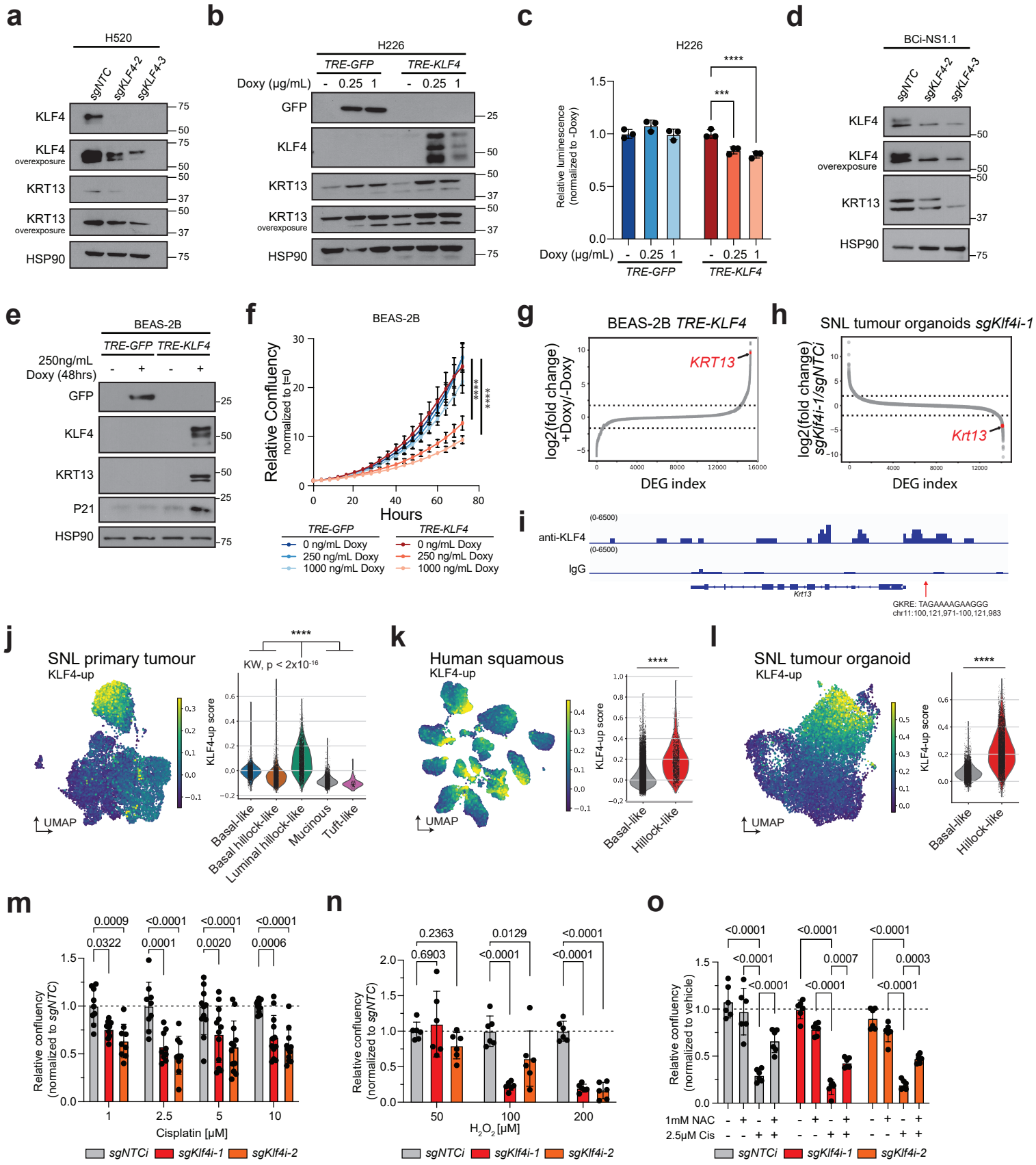
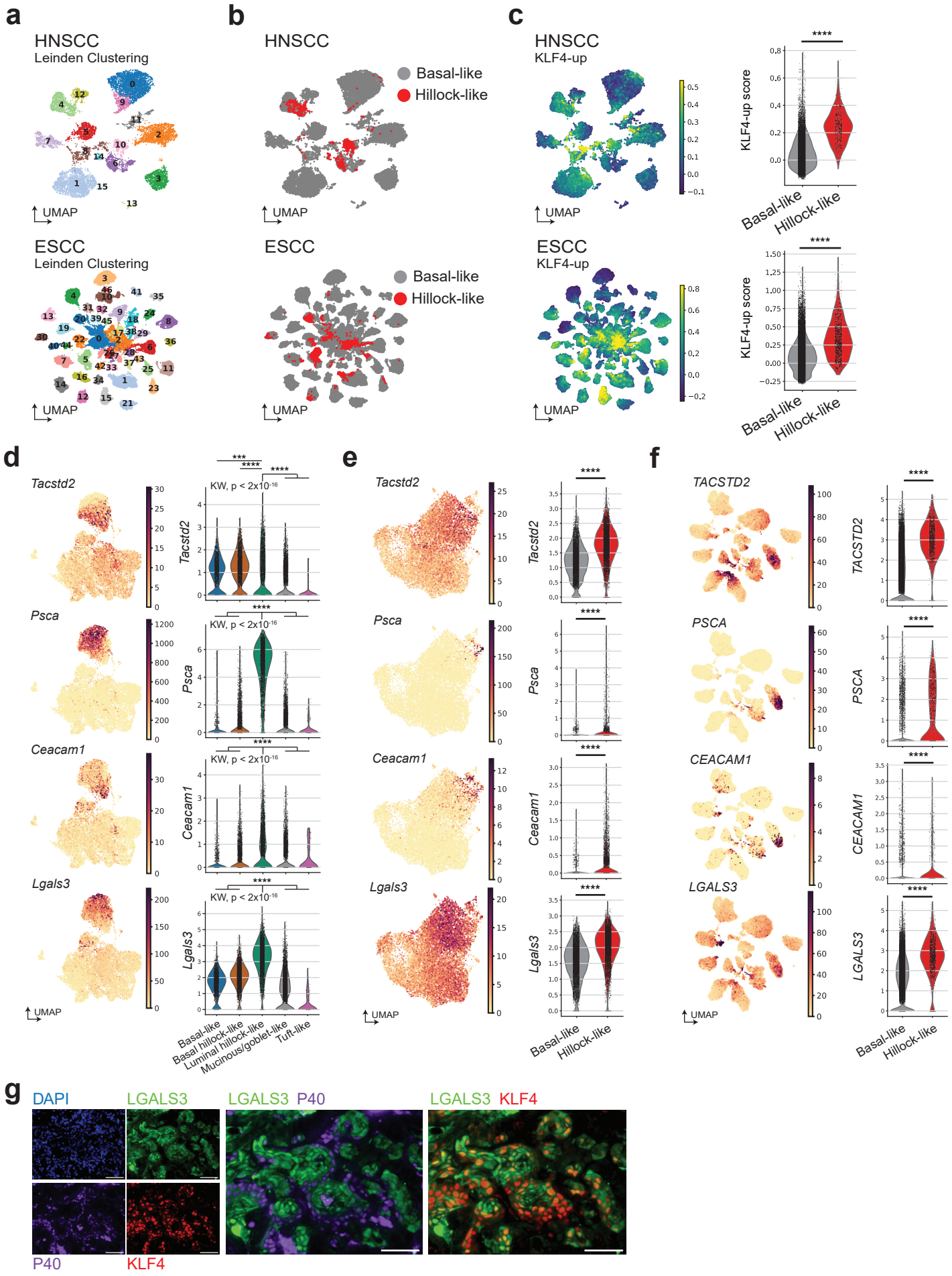

a
H520
sgNTC
sgKLF4-2
sgKLF4-3
KLF4
KLF4 overexposure
KRT13
KRT13 overexposure
HSP90
b
H226
TRE-GFP
TRE-KLF4
Doxy (µg/mL)
GFP
KLF4
KRT13
KRT13 overexposure
HSP90
c
H226
Relative luminescence (normalized to -Doxy)
Doxy (µg/mL)
TRE-GFP
TRE-KLF4
d
BCi-NS1.1
sgNTC
sgKLF4-2
sgKLF4-3
KLF4
KLF4 overexposure
KRT13
HSP90
e
BEAS-2B
TRE-GFP
TRE-KLF4
250ng/mL Doxy (48hrs)
GFP
KLF4
KRT13
P21
HSP90
f
BEAS-2B
Relative Confluency normalized to t=0
Hours
TRE-GFP
TRE-KLF4
0 ng/mL Doxy
250 ng/mL Doxy
1000 ng/mL Doxy
g
BEAS-2B TRE-KLF4
log2(fold change) +Doxy/-Doxy
KRT13
DEG index
h
SNL tumour organoids sgKlf4i-1
log2(fold change) sgKlf4i-1/sgNTCi
Krt13
DEG index
i
anti-KLF4
IgG
Krt13
GKRE: TAGAAAAGAAGGG
chr11:100,121,971-100,121,983
j
SNL primary tumour
KLF4-up
UMAP
KW, p < 2x10-16
KLF4-up score
Basal-like
Basal hillock-like
Luminal hillock-like
Mucinous
Tuft-like
k
Human squamous
KLF4-up
UMAP
KLF4-up score
Basal-like
Hillock-like
l
SNL tumour organoid
KLF4-up
UMAP
KLF4-up score
Basal-like
Hillock-like
m
Relative confluency (normalized to sgNTC)
Cisplatin [µM]
sgNTCi
sgKlf4i-1
sgKlf4i-2
n
Relative confluency (normalized to sgNTC)
H2O2 [µM]
sgNTCi
sgKlf4i-1
sgKlf4i-2
o
Relative confluency (normalized to vehicle)
1mM NAC
2.5µM Cis
sgNTCi
sgKlf4i-1
sgKlf4i-2

Figure 7. KLF4 promotes a KRT13$^+$ hillock-like state and resistance to cisplatin. a) Immunoblot for indicated proteins in the human LUSC cell line H520 expressing non-targeting control (*sgNTC*) or *sgKLF4*. **b)** Immunoblot for indicated proteins in the human H226 LUSC cells with *TRE-GFP* or *-KLF4* overexpression constructs at the indicated dose or without (-) doxycycline (doxy) exposure for 48 hrs. **c)** Celltiter-Glo (Promega) luminescence normalized to no doxy (-) after 5 days of growth in H226 cells treated with the indicated dose of doxy every 48 hrs. Two-way ANOVA, comparing within genotype only. **d)** Immunoblot for indicated proteins in human BCi-NS1.1 cell lines expressing non-targeting control (*sgNTC*) or *sgKLF4*. **e)** 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. **f)** 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 (Doxy). One-way ANOVA. **g)** Waterfall plot of differentially expressed genes from RNA-sequencing of TRE-KLF4 BEAS-2B cells treated +/- doxy as in **7f**. **h)** Waterfall plot of differentially expressed genes from RNA-sequencing of SNL tumour organoids comparing *sgKlf4i-1* to *sgNTC* from **6h-i**. **i)** IGV genome viewer of the *Krt13* locus for KLF4 CUT&RUN reads in SNL tumour organoids. **j-l)** KLF4-up score applied to SNL tumour cells (**j**), human squamous tumour cells (**k**), and organoid cells (**l**) projected onto UMAP space from **3a, 4d,** and **5h** with score expression violin plots by cell state assignment. Kruskal-Wallis followed by Dunn's test with a Bonferonni correction; all comparisons $p < 2x10^{-16}$ (**j**), Mann-Whitney U test (**k,l**); ****$p \leq 0.0001$. (see **Methods** and **Supplementary Table 13**). **m)** Organoid growth after treatment with cisplatin determined by GFP+ confluency using an Incucyte live-cell imaging system (Sartorius) normalized to vehicle treated organoids followed by normalization to *sgNTCi* after 5 days. Two-way ANOVA, multiple comparisons to *sgNTCi* within each dose. **n)** Organoid growth after treatment with hydrogen peroxide (H_2O_2) determined by GFP+ confluency using an Incucyte live-cell imaging system (Sartorius) normalized to vehicle treated organoids followed by normalization to *sgNTCi* after 5 days. Two-way ANOVA, multiple comparisons to *sgNTCi* within each dose. **o)** Organoid growth after treatment with cisplatin and N-acetylcystine (NAC) determined by GFP+ confluency using an Incucyte live-cell imaging system (Sartorius) normalized to vehicle treated organoids after 5 days. Vehicle treatment represented by dashed line at relative growth 1 for each cell line. Two-way ANOVA, multiple comparisons to vehicle within each genotype. **For all panels,** error bars represent mean +/- SD; *$p \leq 0.05$; **$p \leq 0.01$;***$p \leq 0.001$; ****$p \leq 0.0001$ or exact values are shown. DAPI (blue) marks nuclei. HSP90 serves as loading control for immunoblot.

See also Extended Data Fig. 7 and 8.

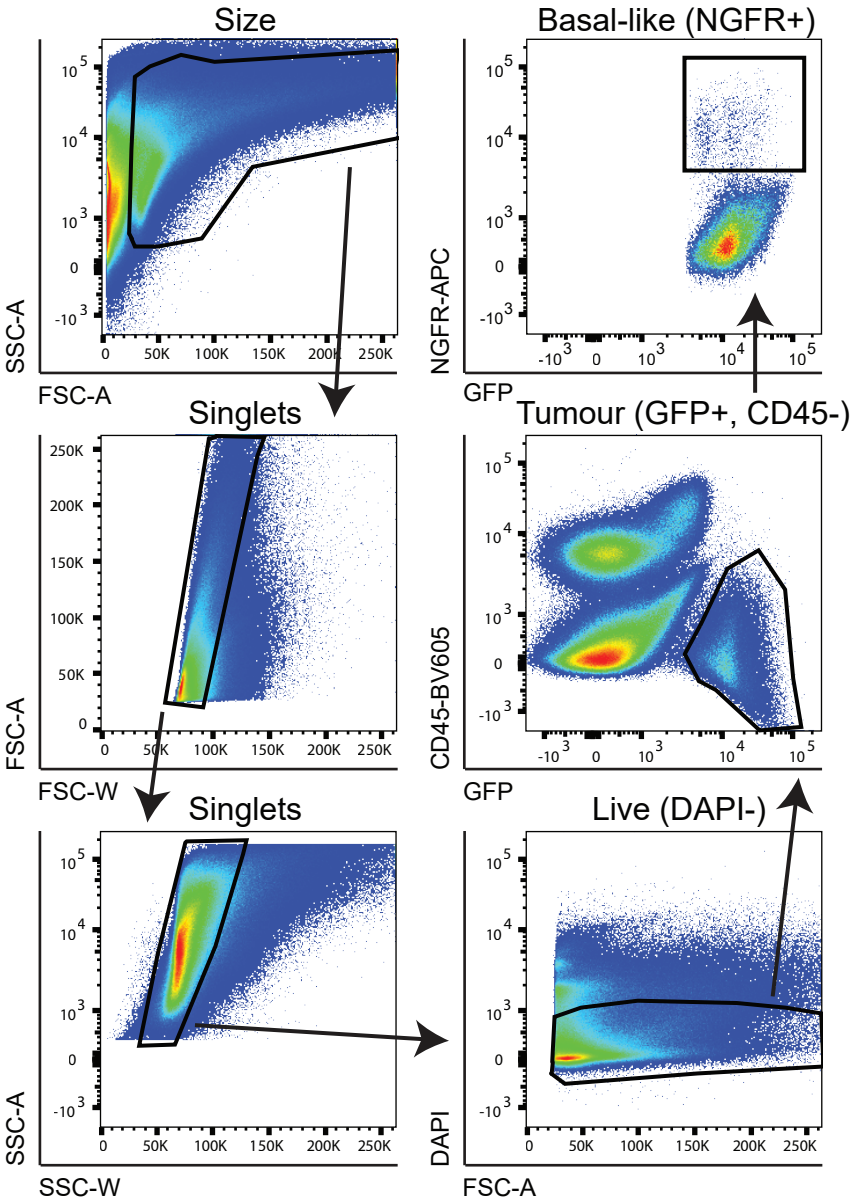
Figure 8


a
HNSCC
Leinden Clustering
UMAP
ESCC
Leinden Clustering
UMAP
b
HNSCC
Basal-like
Hillock-like
UMAP
ESCC
Basal-like
Hillock-like
UMAP
c
HNSCC
KLF4-up
UMAP
KLF4-up score
Basal-like
Hillock-like
ESCC
KLF4-up
UMAP
KLF4-up score
Basal-like
Hillock-like
d
Tacstd2
Psca
Ceacam1
Lgals3
UMAP
KW, p < 2x10-16
Basal-like
Basal hillock-like
Luminal hillock-like
Mucinous/goblet-like
Tuft-like
e
Tacstd2
Psca
Ceacam1
Lgals3
UMAP
Basal-like
Hillock-like
f
TACSTD2
PSCA
CEACAM1
LGALS3
UMAP
Basal-like
Hillock-like
g
DAPI
LGALS3
LGALS3 P40
LGALS3 KLF4
P40
KLF4

1726 **Figure 8. A conserved hillock-like state is present across squamous tumours and is**
1727 **enriched for potential therapeutic targets. a)** UMAP of scRNA-seq from head & neck tumour
1728 cells from 15 patients[95] (Kurten et al) or from esophageal tumours from 60 patients[96] (Zhang et
1729 al) labeled by Leiden clustering. **b)** Cell type classification of head & neck SCC or esophageal
1730 SCC single cell data based on hillock score expression from cells in **8a**. **c)** UMAP of KLF4-up
1731 score applied to head & neck and esophageal SCC cells from **8a** with score expression in violin
1732 plots grouped by cell state assignment. Mann-Whitney U test; ****$p \leq 0.0001$. **d-f)** Gene
1733 expression in SNL tumour cells (**d**), organoid cells (**e**), and human squamous tumour cells (**f**)
1734 projected onto UMAP space from **3a, 4d,** and **5h** with score expression violin plots by cell state
1735 assignment. Kruskal-Wallis followed by Dunn's test with a Bonferonni correction; all comparisons
1736 $p < 2x10^{-16}$ (**d**), Mann-Whitney U test (**e,f**); ****$p \leq 0.0001$. **g)** Representative immunofluorescence
1737 (IF) co-staining for LGALS3 (green), KLF4 (red), and P40 (purple) in SNL squamous tumours.
1738 Representative of n = 2 Ad-CMV-SNL mice. Scale bars, 50 μm.
1739 **See also Extended Data Fig. 9 and 10.**
1740

Extended Data Fig. 1

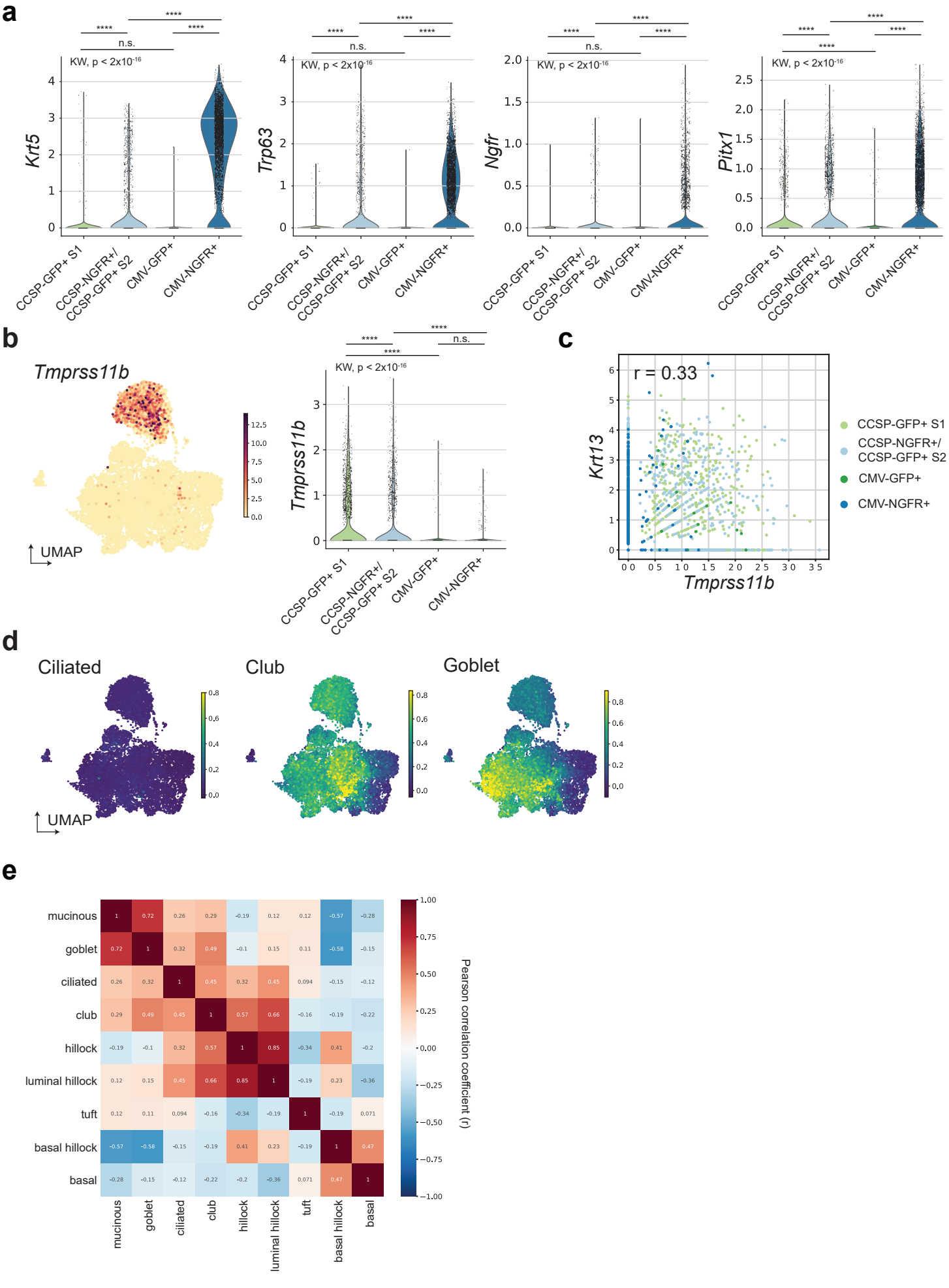
a



1741 **Extended Data Fig. 1. NGFR$^+$ tumour cells sorted from SNL tumours for scRNA-seq, related**
1742 **to Fig. 2. a)** Representative flow cytometry plots for gating strategy of DAPI$^-$CD45$^-$GFP$^+$NGFR$^+$
1743 SNL tumour cells. SSC: side scatter, FSC: forward scatter, A: Area, W: Width. Arrows in upper
1744 left graph flow down to the right, and up back to the upper right panel, indicating progressive
1745 gating during the flow sort.

1746

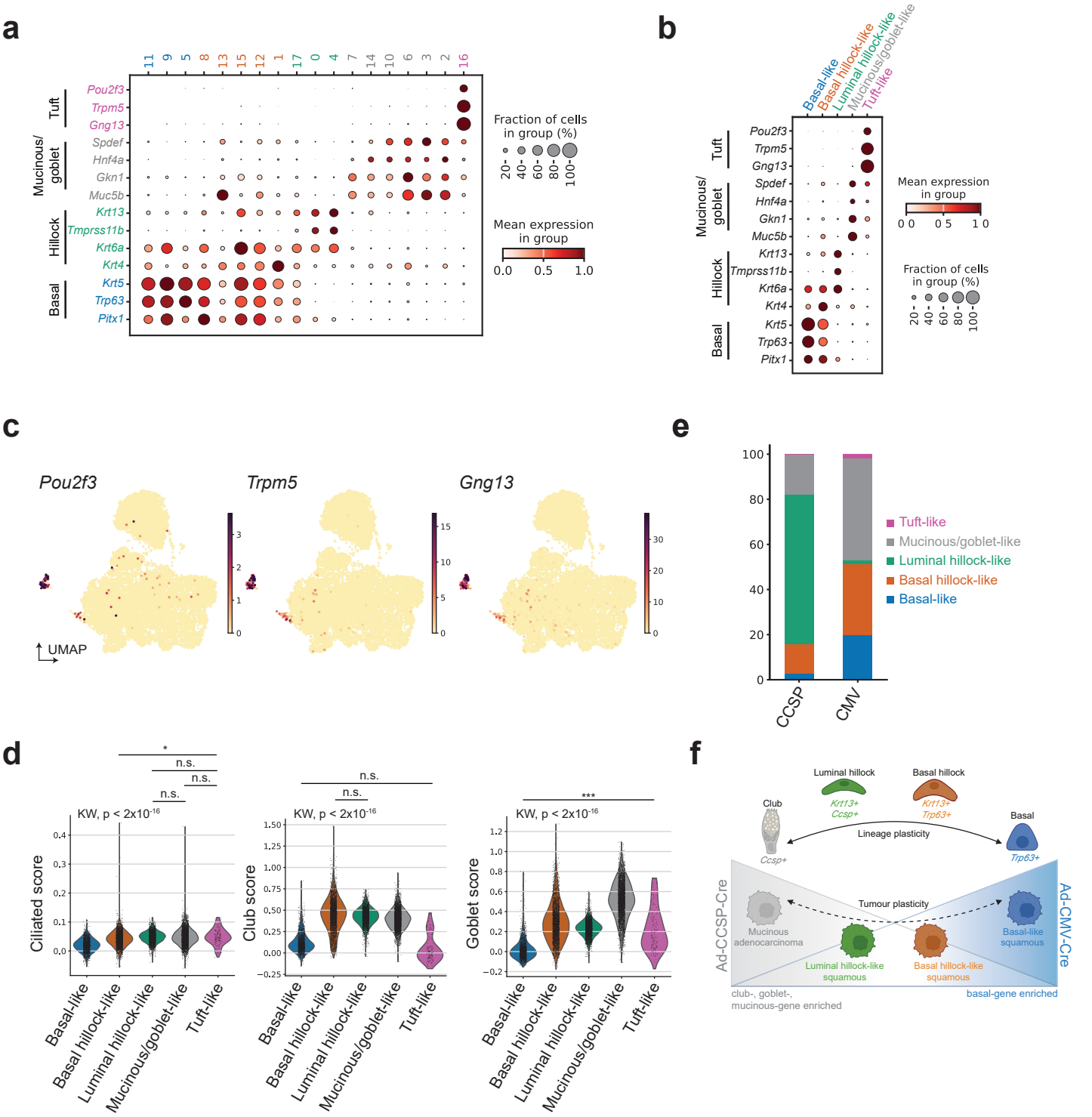
Extended Data Fig. 2



1747 **Extended Data Fig. 2. *Tmprss11b* correlates with *Krt13 expression*, related to Fig. 3. a)**
1748 Violin plots for basal marker genes by sample. Kruskal-Wallis followed by Dunn's test with
1749 Bonferroni correction, not all comparisons shown. **b)** scRNA-seq expression of *Tmprss11b*
1750 projected on UMAP space from cells in **3a** (left) and violin plot for *Tmprss11b* expression by
1751 sample (right). Kruskal-Wallis followed by Dunn's test with Bonferroni correction, not all
1752 comparisons shown. **c)** Correlation between *Krt13* and *Tmprss11b* expression in cells from **3a**
1753 with the Pearson correlation coefficient (r) shown. **d)** Gene scores applied to SNL tumour cells
1754 projected onto UMAP space from **3a** (see **Methods** and **Supplementary Table 2**). **e)** Correlation
1755 analysis between gene scores in SNL tumour single cell data in **3a** with the Pearson correlation
1756 coefficient (r) shown. **For all panels,** ****$p \leq 0.0001$.

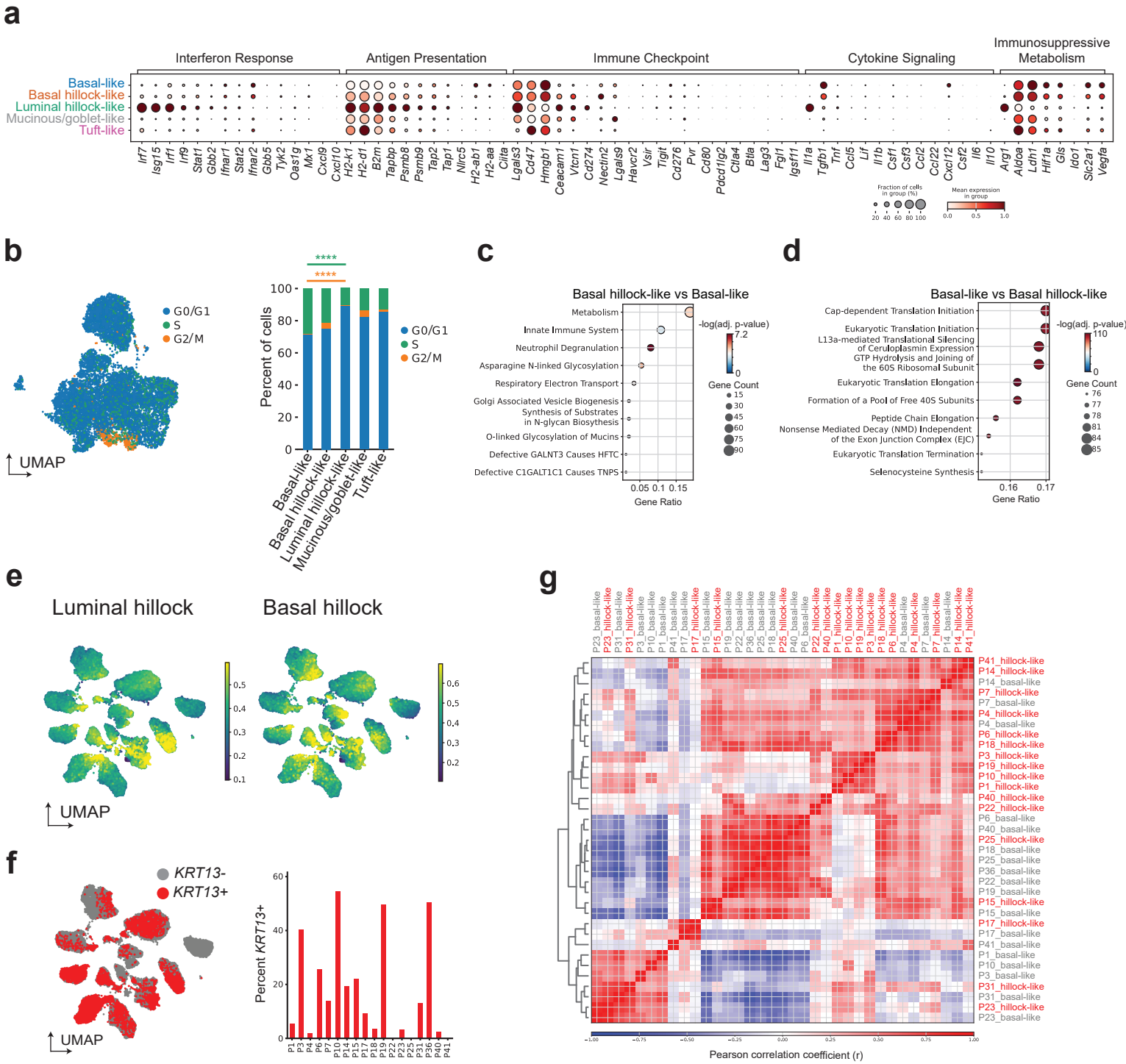
1757

Extended Data Fig. 3



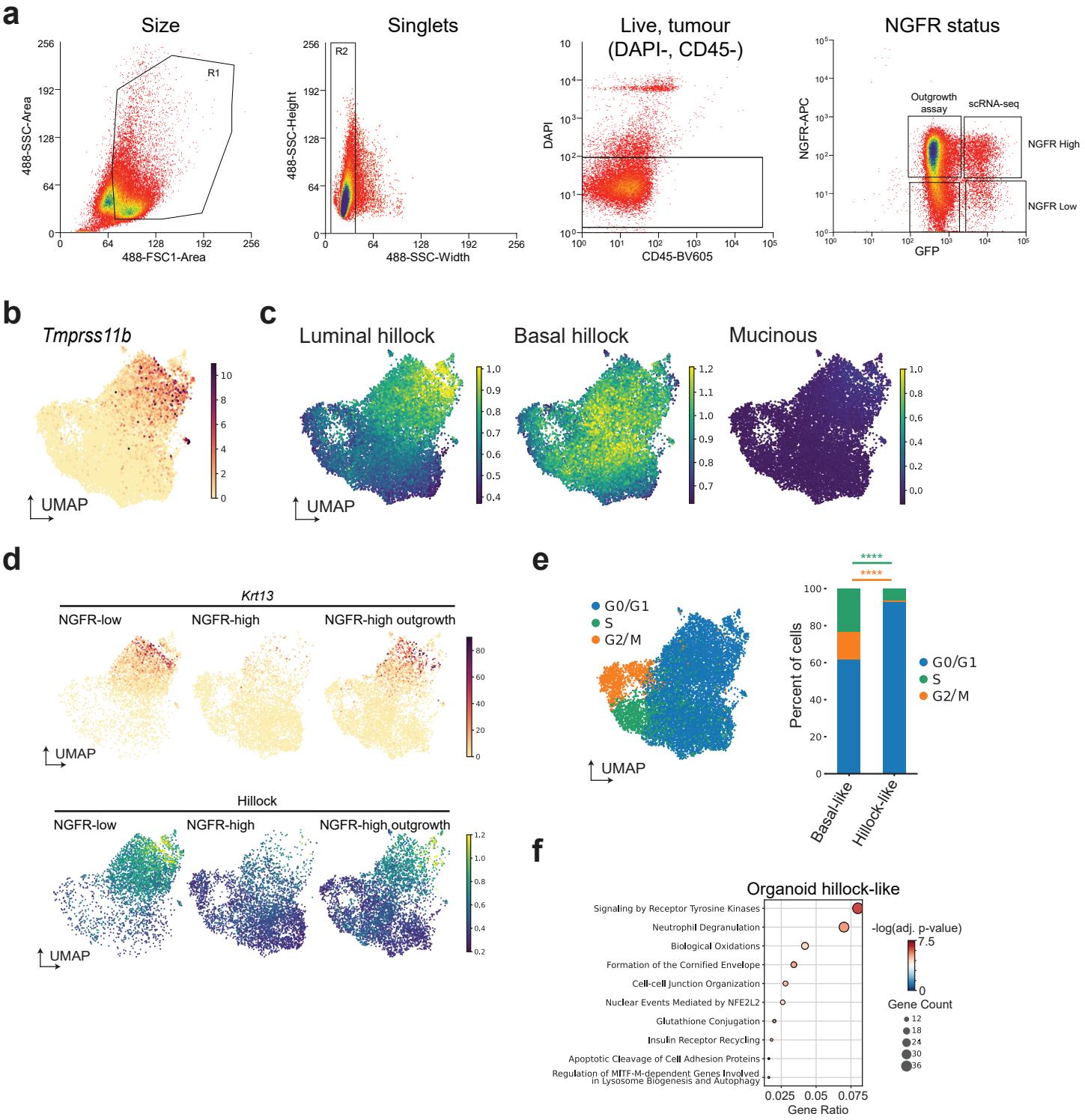
1758 **Extended Data Fig. 3. Defining cell states by transcriptional profiles in SNL tumours**,
1759 **related to Fig. 3. a)** Scanpy dot plot of scRNA-seq gene expression by Leiden clustering. Text
1760 coloring of Leiden clusters indicates cell type assignment, reflected in b. **b)** Scanpy dot plot of
1761 scRNA-seq gene expression by assigned cell type. For **a)** and **b)**, Dot size indicates percent of
1762 cells expressing each transcript and the color indicates average expression per group, as
1763 indicated in figure. **c)** scRNA-seq expression of tuft marker genes projected on UMAP from cells
1764 in **3a**. **d)** Violin plots of gene score expression by cell type assignment. Kruskal-Wallis followed
1765 by Dunn's test with Bonferroni correction, all comparison $p < 0.0001$ unless otherwise shown. **e)**
1766 Stacked bar plot showing percent of cells occupying each cell state in Ad-CCSP- versus Ad-CMV-
1767 initiated tumour samples from scRNA-seq data in **3h**. **f)** Schematic depicting potential lineage
1768 plasticity in normal cell types and SNL tumor cell states. Created with Biorender.

Extended Data Fig. 4



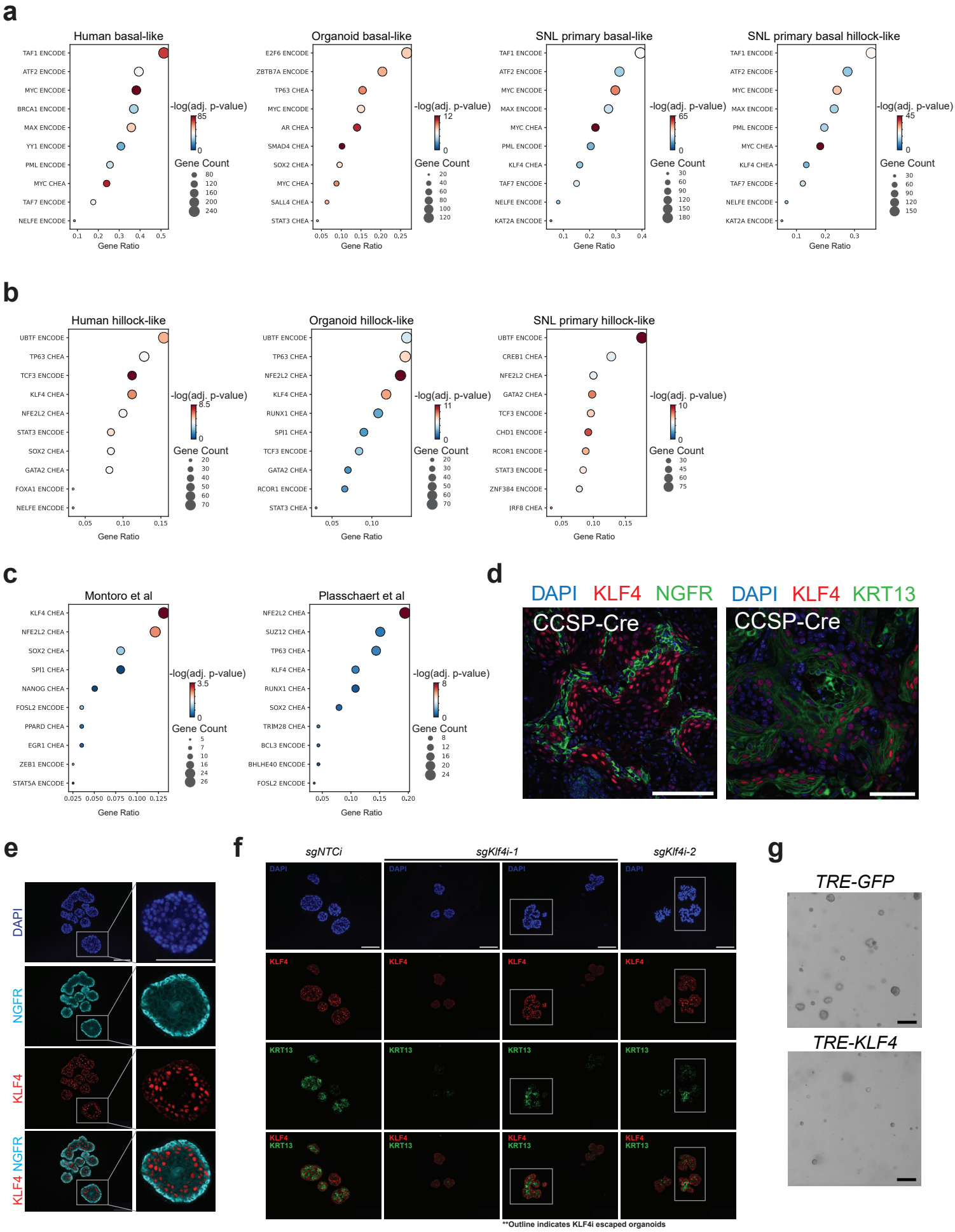
1769 **Extended Data Fig. 4. The hillock state in squamous tumours may be immunomodulatory**
1770 **and is relatively non-proliferative**, **related to Fig. 4. a)** Scanpy dot plot of scRNA-seq gene
1771 expression of SNL tumours depicted in **Fig. 3a,** by Leiden clustering. Dot size indicates percent
1772 of cells expressing each transcript and the color indicates the average expression per cluster
1773 indicated in figure. Text coloring indicates cell type assignment. **b)** Cell cycle assignments applied
1774 to cells in UMAP space as in **3a** (see **Methods**) and stacked bar plot showing percent of cells
1775 occupying each cell cycle phase by cell state assignment as in **3h**. Statistics shown compare
1776 G2/M (orange) or S-phase (green) gene scores between assigned cell states. Kruskal-Wallis test
1777 followed by Dunn's test, not all comparisons shown. **c-d)** Gene set enrichment analysis comparing
1778 basal-like and basal hillock-like SNL tumour cells from **3h**. The ENRICHR gene set used was
1779 'Reactome Pathways 2024'. **e)** Luminal hillock and basal hillock scores applied to human lung
1780 squamous cell carcinoma tumours in UMAP space as in **4d** (see **Methods** and **Supplementary**
1781 **Tables 4** and **5**). **f)** UMAP showing assignment of *KRT13* state from **4d** (left) and *KRT13*
1782 assignment per patient tumour (n=18) as a percentage of all tumour cells per sample (right). **g)**
1783 Correlation analysis of the whole trasncriptome of patient tumours divided by cell state from single
1784 cell data in **4d, j** with the Pearson correlation coefficient (r) shown. Red labels indicate hillock-like
1785 cells from each patient; grey labels indicate basal-like cells from each patient.
1786 **See also Supplementary Table 8 and 9.**
1787

Extended Data Fig. 5



Extended Data Fig. 5. NGFR$^+$ cells give rise to KRT13$^+$ hillock-like cells, related to Fig. 5. a) Gating strategy for NGFR-high and low cells from dissociated SNL tumour organoids cells in **5f-l**. GFP/NGFR-high cells were used for the "NGFR-high" scRNA-seq sample and replated for the "NGFR-high outgrowth" scRNA-seq. A 50:50 mixture of GFP-high:GFP-low NGFR-low cells were used for the "NGFR-low" scRNA-seq sample. GFP low samples were used for organoid outgrowth assays. **b)** *Tmprss11b* expression projected onto UMAP space from **5h**. **c)** Luminal hillock, basal hillock, and mucinous scores applied to cells projected onto UMAP space from **5h** (see **Methods** and **Supplementary Tables 3-5**).**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**. Statistics shown compare G2/M (orange) or S-phase (green) gene scores between assigned cell states. Mann-Whiteney U-Test. ****$p \leq 0.0001$. **f)** Gene set enrichment analysis comparing organoid cells in **5l** by assigned cell type. The ENRICHR gene set used was 'Reactome Pathways 2024'.
See also Supplementary Table 11.

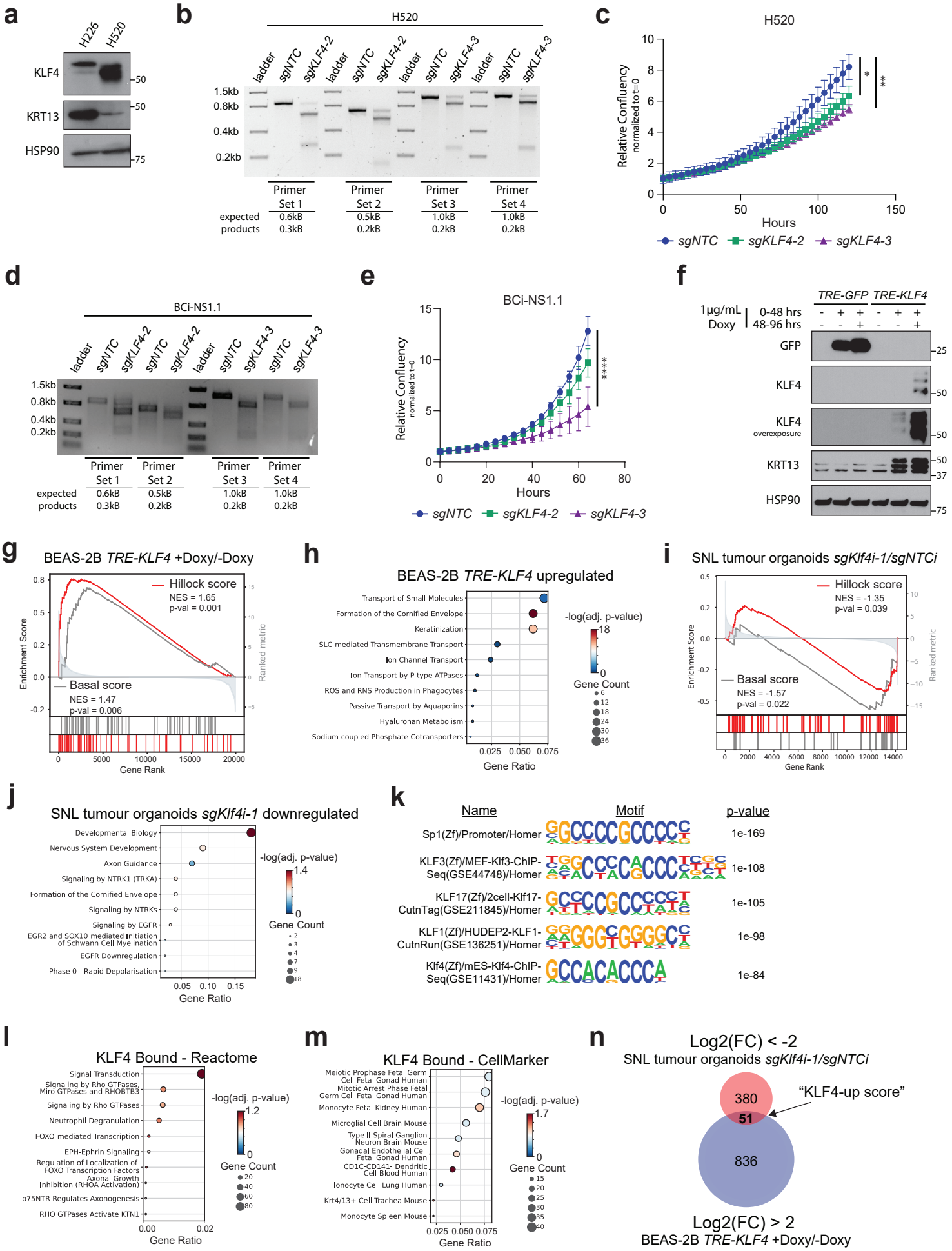
Extended Data Fig. 6



1804 **Extended Data Fig. 6. Identifying candidate regulators of the hillock cell state and KLF4 as**
1805 **a regulator of KRT13, related to Fig. 6. a)** ENRICHR analysis for predicted regulators of the
1806 basal-like state in human basal-like cells, organoid basal-like cells, SNL tumour basal-like and
1807 basal hillock-like cells using 'ENCODE and ChEA Consensus TFs from ChIP-X'. **b)** ENRICHR
1808 analysis for predicted regulators of the hillock-like state in human hillock-like cells, organoid
1809 hillock-like cells, and SNL tumour luminal hillock-like cells using 'ENCODE and ChEA Consensus
1810 TFs from ChIP-X'. **c)** ENRICHR analysis for predicted regulators of the hillock state in normal lung
1811 tissue from the indicated publications (Montoro et al, Nature, 2018; Plasschaert et al, Nature,
1812 2018) using 'ENCODE and ChEA Consensus TFs from ChIP-X' (see **Methods - Differential**
1813 **gene expression analyses** and **Supplementary Table 11**). **d)** IF staining for NGFR (green) and
1814 KLF4 (red) or KRT13 (green) and KLF4 (red) in squamous tumours from Ad-CCSP-initiated SNL
1815 tumours. Representative of n = 3 mice. Scale bars, 100 μm. **e)** Representative IF co-staining for
1816 KLF4 (red) and NGFR (cyan) in SNL tumour organoids. Scale bars, 100 μm. Representative of n
1817 = 2 experiments at with least 4 different 20x fields per experiment. **f)** Representative IF co-staining
1818 for KRT13 (green) and KLF4 (red) in SNL tumour organoids with KLF4 knockdown using
1819 CRISPRi. Scale bars, 100 μm. Images are uncropped coinciding to representative images shown
1820 in **6h**. White boxes indicate *sgKlf4i* expressing organoids with escaped KLF4 expression. DAPI
1821 (blue) marks nuclei. **g)** Representative brightfield images of SNL tumour organoids infected with
1822 TRE-GFP or -KLF4 treated with 1μg/mL doxycycline every 48 hrs for 7 days. Scale bars, 200 μm.

1823

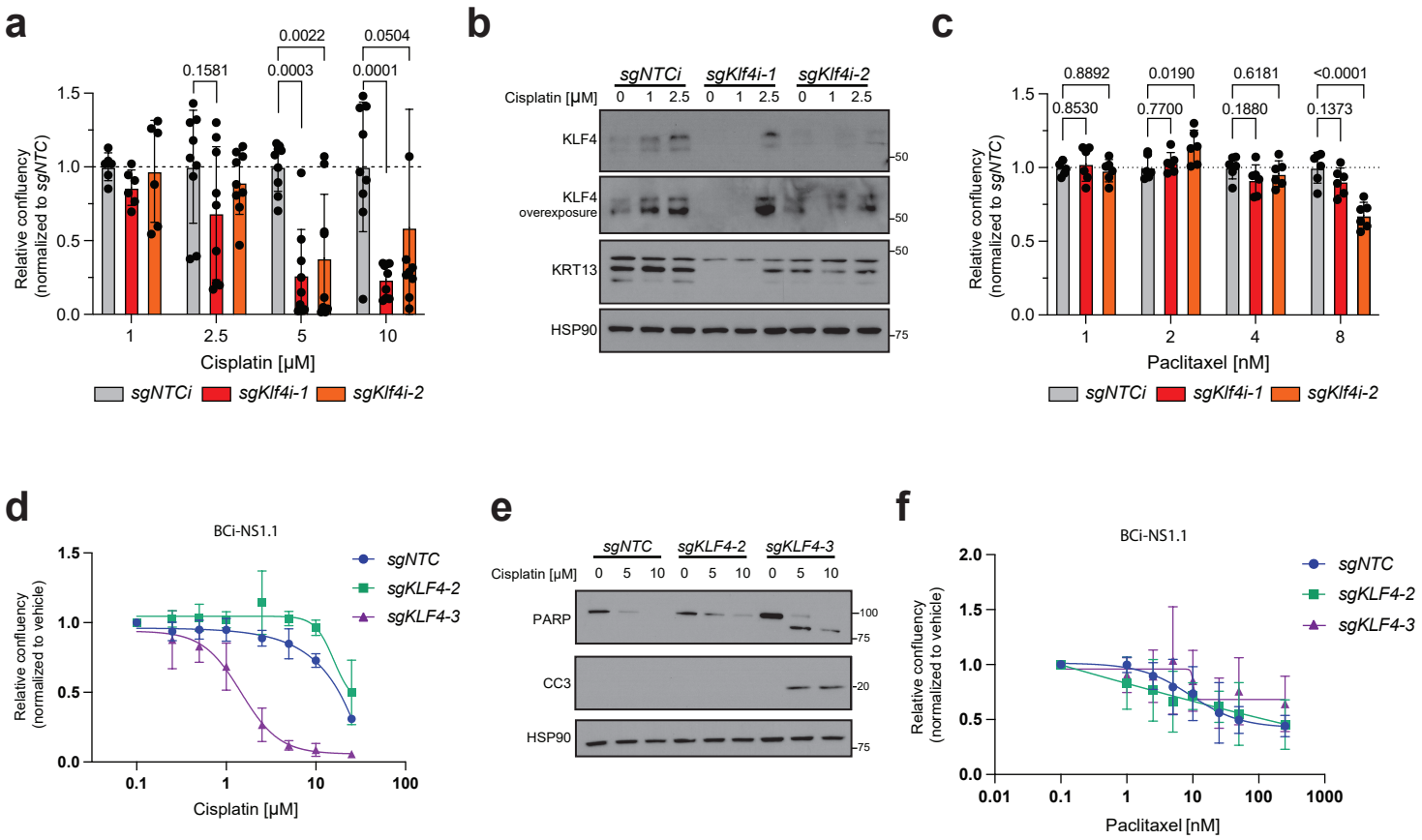
Extended Data Fig. 7



1824 **Extended Data Fig. 7. KLF4 regulates the KRT13^{+} hillock-like-state, related to Fig. 7. a)**
1825 Immunoblot for indicated proteins in the human LUSC cell lines H226 and H520. **b)** T7 surveyor
1826 assay for *KLF4* gene editing in human H520. Expected fragment sizes after T7 cutting of
1827 mismatched (edited) DNA are shown below. **c)** Cell growth determined by confluency normalized
1828 to initial plating density of H520 cells. One-way ANOVA at endpoint. **d)** T7 surveyor assay for
1829 *KLF4* gene editing in human BCi-NS1.1 cells. Expected fragment sizes after T7 cutting of
1830 mismatched (edited) DNA are shown below. **e)** Cell growth determined by confluency normalized
1831 to initial plating density of BCi-NS1.1 cells. One-way ANOVA. **f)** Immunoblot for indicated proteins
1832 in BEAS-2B cells infected with TRE-GFP or KLF4 treated with (+) or without (-) doxycycline (doxy)
1833 for the indicated times. HSP90 serves as loading control. **g)** GSEA for the general hillock score
1834 and basal score derived from RNA-sequencing of TRE-KLF4 BEAS-2B cells treated +/- doxy as
1835 in **7g**. **h)** Gene set enrichment analysis of the top 100 upregulated genes by TRE-KLF4 in BEAS-
1836 2B cells treated +/- doxy as in **7g**. The ENRICHR gene set used was 'Reactome Pathways 2024'.
1837 **i)** GSEA for the general hillock score and basal score derived from RNA-sequencing of SNL
1838 tumour organoids as in **7h**. **j)** Gene set enrichment analysis of the top 100 downregulated genes
1839 by *sgKlf4i-1* in tumour organoids as in **7h**. The ENRICHR gene set used was 'Reactome Pathways
1840 2024'. **k)** HOMER motif analysis of SEACR called peaks from KLF4 CUT&RUN in SNL tumour
1841 orgnaoids. **l,m)** ENRICHR analysis of the top 500 peaks annotated as protein coding promoters
1842 bound by KLF4 in SNL tumour organoids. The ENRICHR gene set used was 'Reactome
1843 Pathways 2024' and 'CellMarker 2024' respectively. **n)** Overlap of top genes downregulated by
1844 KLF4 knockdown in organoids and KLF4 overexpression in BEAS-2B. 51 overlapping genes used
1845 to generate the "KLF4-up" score.
1846 **See also Supplementary Table 13 and 14.**
1847

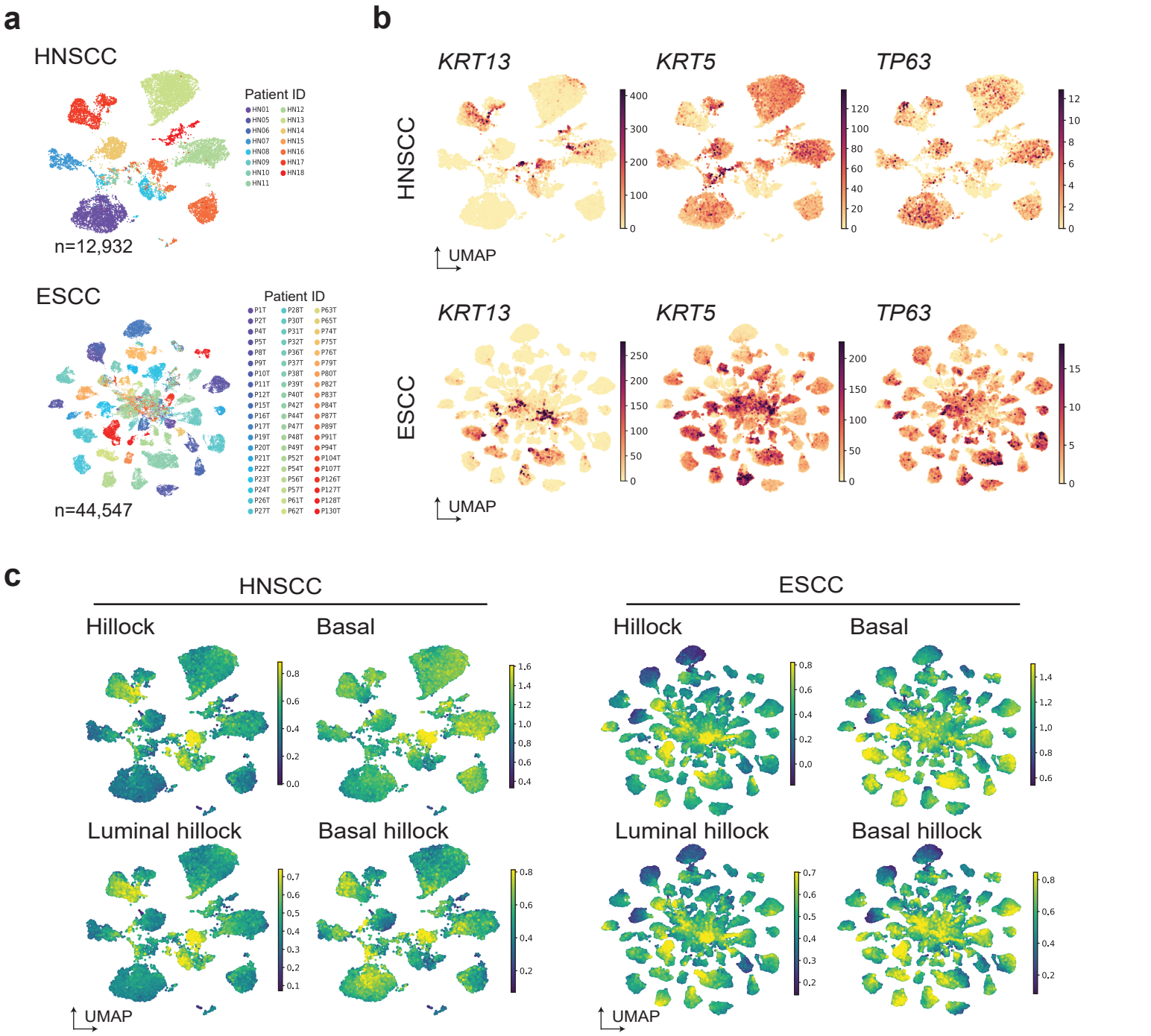
1848

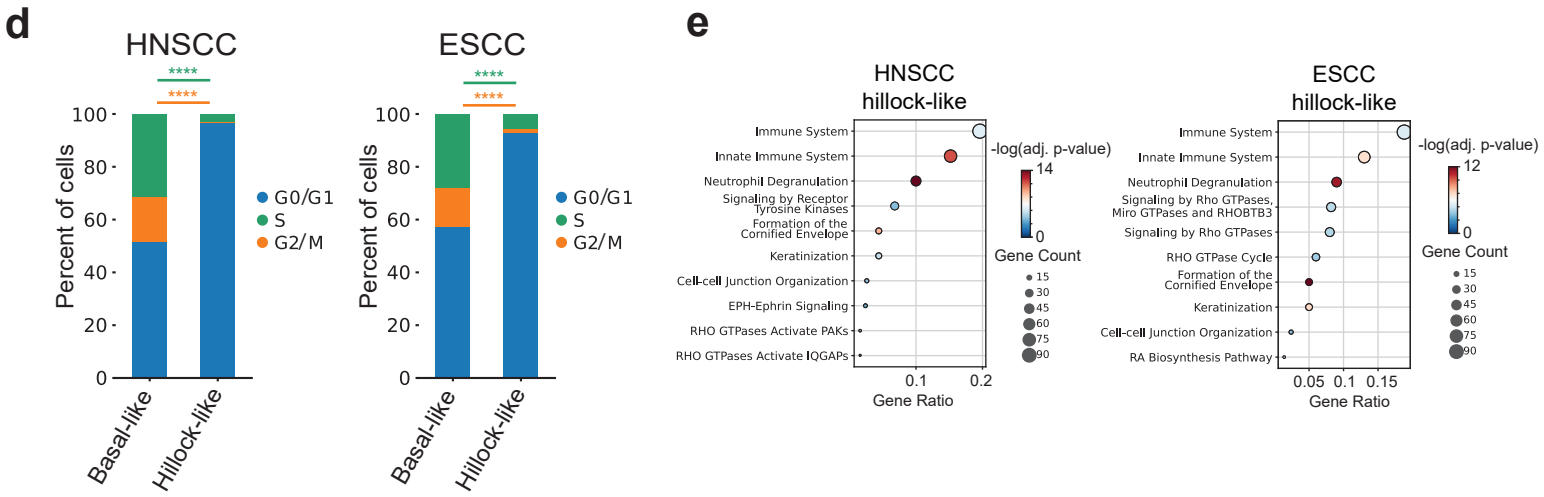
Extended Data Fig. 8



1849 **Extended Data Fig. 8. KLF4 regulates resistance to cisplatin but not paclitaxel, related to**
1850 **Fig. 7. a)** Organoid growth after treatment with cisplatin determined by GFP+ confluency using
1851 an Incucyte live-cell imaging system (Sartorius) normalized to vehicle treated organoids followed
1852 by normalization to *sgNTCi* after 10 days, cells were partially dissociated and related on day 7.
1853 Two-way ANOVA, multiple comparisons to *sgNTCi* within each dose. **b)** Immunoblot for indicated
1854 proteins in SNL tumour organoids after treatment with cisplatin from **Extended Data Fig. 8a**. **c)**
1855 Organoid growth after treatment with paclitaxel determined by GFP+ confluency using an Incucyte
1856 live-cell imaging system (Sartorius) normalized to vehicle treated organoids followed by
1857 normalization to *sgNTCi* after 5 days. Two-way ANOVA, multiple comparisons to *sgNTCi* within
1858 each dose. **d)** Human BCi-NS1.1 cell line growth after treatment with cisplatin for 48 hrs
1859 determined by confluency using an Incucyte live-cell imaging system (Sartorius) normalized to
1860 vehicle treated cells. **e)** Immunoblot for indicated proteins in human BCi-NS1.1 cell lines
1861 expressing non-targeting control (sgNTC) or sgKLF4 after 48 hours of treatment with the indicated
1862 doses of cisplatin. **f)** Human BCi-NS1.1 cell line growth after treatment with paclitaxel for 72 hrs
1863 determined by confluency using an Incucyte live-cell imaging system (Sartorius) normalized to
1864 vehicle treated cells. **For all panels,** error bars represent mean +/- SD; exact p-values are listed.
1865

Extended Data Fig. 9

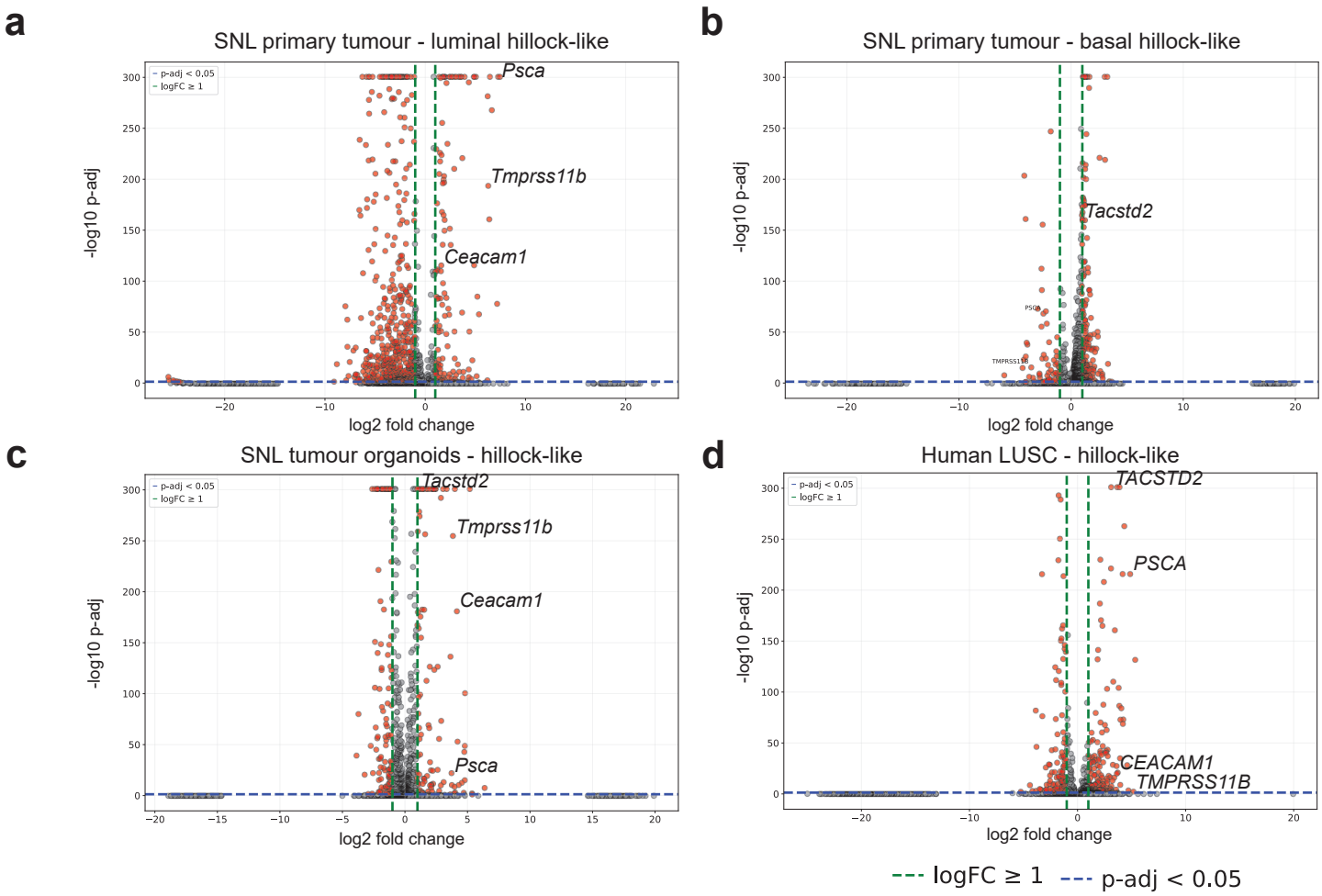




1866 **Extended Data Fig. 9. Hillock-like cells are present in squamous tumours of the esophagus**
1867 **and head & neck, related to Fig. 8. a)** UMAP of scRNA-seq from human head & neck tumour
1868 cells from 15 patients[95] or from esophageal tumours from 60 patients[96] labeled by patient (HN, or
1869 P) sample ID. **b)** scRNA-seq expression of squamous (*TP63, KRT5)* or hillock markers (*KRT13*)
1870 in UMAP space from cells in **8a**. **c)** Human hillock score, basal score, luminal hillock score, and
1871 basal hillock score (**Supplementary Tables 1-5**) applied to tumour cells as in **8a** with score
1872 projected in UMAP. **d)** Cell cycle assignments applied to cells from **8a** (see **Methods**) shown as
1873 a stacked bar plot with percent of cells occupying each cell cycle phase by cell state assignment
1874 as in **8H**. Statistics shown compare G2/M (orange) and S-phase (green) gene scores between
1875 assigned cell states. Mann-Whiteney U-Test. ****$p \leq 0.0001$. **e)** Gene set enrichment analysis
1876 comparing human tumour cells in **8b** by assigned cell type. The ENRICHR gene set used was
1877 'Reactome Pathways 2024'.

1878

Extended Data Fig. 10



1879 **Extended Data Fig. 10. Differentially expressed cell surface protein-encoding genes by cell**
1880 **state using scRNA-seq data, related to Fig. 8. a-d)** Volcano plots of DEGs by cell state filtered
1881 for predicted cell surface protein encoding genes.
1882 **See also Supplementary Table 15.**
1883

1884