Critical role for a high-plasticity cell state in lung cancer

https://doi.org/10.1038/s41586-025-09985-x

Received: 10 October 2024

Accepted: 1 December 2025

Jason E. Chan[1,2,10], Chun-Hao Pan[1,10], Jonathan Rub[1,3], Gary Guzman[1], Klavdija Krause[1], Emma Brown[1], Zeda Zhang[1], Hannah Styers[1], Griffin Hartmann[1], Zhuxuan Li[4], Xueqian Zhuang[1], Scott W. Lowe[1,5], Doron Betel[6,7], Yan Yan[1,8,9]✉ & Tuomas Tammela[1]✉

Plasticity—the ability of cells to undergo phenotypic transitions—drives cancer progression and therapy resistance[1–3]. Recent studies have suggested that plasticity in solid tumours is concentrated in a minority subset of cancer cells[4–6], yet functional studies examining this high-plasticity cell state (HPCS) in situ are lacking. Here we develop mouse models enabling the detection, longitudinal lineage tracing and ablation of the HPCS in autochthonous lung tumours in vivo. Lineage tracing reveals that the HPCS cells possess a high capacity for cell state transitions, giving rise to both early neoplastic (differentiated) and progressed lung cancer cell states in situ. Longitudinal lineage tracing using secreted luciferases reveals that HPCS-derived cells have a high capacity for growth compared with bulk cancer cells or another cancer cell state with features of differentiated lung epithelium. Ablation of HPCS cells in early neoplasias abrogates benign-to-malignant transition, whereas ablation in established tumours by suicide gene or chimeric antigen receptor (CAR) T cells robustly reduces tumour burden. We further demonstrate that the HPCS gives rise to therapy-resistant cell states, whereas HPCS ablation suppresses resistance to chemotherapy and oncoprotein-targeted therapy. Notably, an HPCS-like state is ubiquitous in regenerating epithelia and in carcinomas of multiple other tissues, revealing a convergence of plasticity programs. Our work establishes the HPCS as a critical hub enabling reciprocal transitions between cancer cell states. Targeting the HPCS in lung cancer and in other carcinomas may suppress cancer progression and eradicate treatment resistance.

Plasticity promotes cancer progression by enabling malignant cells to acquire states with a high capacity to proliferate, metastasize and adapt to stress[1,2]. Moreover, plasticity promotes resistance to chemotherapy[4,7] and oncoprotein-targeted therapy[8–10] in multiple cancer types, allowing cancer cells to acquire new cell states adapted to withstand therapeutic pressure through non-genetic mechanisms[1,2]. Thus, plasticity remains one of the most fundamental problems in cancer biology and one of the foremost challenges in clinical cancer management today. Yet, it is not apparent, conceptually or practically, how targeting cancer plasticity would be best achieved—that is, whether the focus should be on subsets of cells or rather on specific molecular programs.

The application of single-cell genomics and associated computational approaches over the past decade has enabled the unsupervised mapping of malignant cell states at a considerable scale and resolution[11,12]. Although limited to static snapshots, single-cell mRNA sequencing (scRNA-seq) studies have identified candidate transitional states in multiple solid tumour types[4,11–16]. Collectively, these studies suggest that plasticity is concentrated in specific subsets of cancer cells, warranting their functional analysis in time-dynamic experiments. As plasticity in cancer is fundamentally a temporal problem, new experimental strategies and model systems enabling the elucidation of the dynamic nature of plastic transitions are needed.

Lung adenocarcinoma (LUAD) is a prototype of a common, lethal and therapy-resistant solid tumour[17,18]. Important insights into the biology of human LUAD have emerged from the use of genetically engineered mouse models (GEMMs). In the most commonly used *KP* LUAD GEMMs, viral expression of Cre or Flp recombinase in lung epithelial cells leads to somatic activation of oncogenic *Kras*^{G12D} and inactivation of the tumour suppressor p53 (*Trp53*)[19–21]. The *KP* model incorporates de novo LUAD development in the relevant tissue site, recapitulating key molecular and histopathological features of the human disease, including responses to chemotherapy and KRAS-targeted therapy[9,22].

We recently used scRNA-seq to construct a map of LUAD evolution from the alveolar type 2 (AT2) cell of origin to advanced adenocarcinoma

---

[1]Cancer Biology and Genetics Program, Sloan Kettering Institute, Memorial Sloan Kettering Cancer Center, New York, NY, USA. [2]Division of Solid Tumor Oncology, Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, NY, USA. [3]Tri-I Program in Computational Biology & Medicine, Weill Cornell Medicine, New York, NY, USA. [4]BCMB Allied Program, Weill Cornell Graduate School of Medical Science, Weill Cornell Medicine, New York, NY, USA. [5]Howard Hughes Medical Institute, Chevy Chase, MD, USA. [6]Division of Hematology & Medical Oncology, Department of Medicine, Weill Cornell Medicine, New York, NY, USA. [7]Institute for Computational Biomedicine, Weill Cornell Medicine, New York, NY, USA. [8]College of Biomedicine and Health and College of Life Science and Technology, Huazhong Agricultural University, Wuhan, China. [9]Hubei Hongshan Laboratory, National Key Laboratory of Agricultural Microbiology, Huazhong Agricultural University, Wuhan, China. [10]These authors contributed equally: Jason E. Chan, Chun-Hao Pan. ✉e-mail: yanyan@mail.hzau.edu.cn; tammelat@mskcc.org

in the *KP* model[4]. This analysis identified a previously unknown cancer cell state, which emerges in early stages of lung tumorigenesis and is maintained in LUAD tumours throughout cancer progression. Notably, this cell state has transcriptomic features that are substantially different from AT2 cells, other adult tissue stem cells or cancer stem cells (CSCs)—notably, the concurrent expression of gene programs associated with a wide range of cellular identities[4]. Computational modelling of cancer cell state differentiation trajectories implicated this cell state as a key transition point in LUAD progression[4]. Given these attributes, we nominated this cellular subset as a candidate HPCS. Important recent work identified a cell state analogous to the HPCS in human early-stage lung neoplasias[5], underscoring the clinical relevance of the *KP* model for studying the HPCS. Together, these findings suggest that plasticity in lung cancer is concentrated in the HPCS subset, motivating functional interrogation of this state and its potential causal role in distinct steps of LUAD progression and in the context of therapy-associated transitions to drug-resistant cell states.

Reporters for HPCS tracing and ablation

We developed two reporter systems enabling us to visualize, isolate, lineage-trace and ablate the HPCS cells. We identified *Slc4a11* as the most specific gene marking the HPCS in autochthonous mouse *KP* LUAD tumours[4] (Fig. 1a,b, Extended Data Fig. 1a and Supplementary Table 1). *Slc4a11* encodes a sodium-coupled hydroxyl transporter that is highly expressed in corneal endothelial cells, inner ear sensory epithelia and renal medullary tubules in adults[23]. We knocked in a cDNA reporter cassette comprising an mScarlet bright red monomeric fluorescent protein, the tamoxifen-activatable Cre recombinase (*creERT2*)[24] and the human diphtheria toxin (DT) receptor (DTR)[25] suicide gene (hereafter, the *MCD* cassette) into the *Slc4a11* locus in *KPfrt* (*Kras*$^{frt\text{-}stop\text{-}frt(FSF)\text{-}G12D/+}$; *Trp53*$^{frt/frt}$)[26] mouse embryonic stem (mES) cells and generated mES-cell-derived chimeras (Extended Data Fig. 1b–d).

We also developed a lineage-tracing system marking cells undergoing Flp and/or Cre recombinase activity, enabling longitudinal monitoring of the growth potential of the HPCS and other cell states in vivo. Our system comprises a Flp-inducible *Gaussia princeps* luciferase (G-Luc)[27] linked to eGFP[28] (hereafter, *GG*) and Cre-inducible *Cypridina hilgendorfii* luciferase (C-Luc)[29] linked to TagBFP[30] (hereafter, *CB*) (Fig. 1c). Both luciferases are naturally secreted, enabling their detection from a small volume of plasma using specific substrates[31]. This enables non-terminal longitudinal sampling in the same animal to investigate the long-term growth potential of lineage-traced cells in vivo. The *GGCB* cassette was knocked into the ubiquitously active *Hipp11* locus[32] in *KPfrt* mES cells, followed by generation of mES-cell-derived chimeras (Extended Data Fig. 1e–g). We validated Flp-dependent activation of GG and the Cre-inducible switch of *GG* to *CB* in an ex vivo AT2 cell transformation assay[33] (Extended Data Fig. 1h–k), in subcutaneous transplants (Extended Data Fig. 2a–c) and in autochthonous *KPfrt* lung tumours (Extended Data Fig. 2d–g). Taken together, these results demonstrate that the *Hipp11*$^{FSF\text{-}GGCB/+}$ allele is a specific and sensitive reporter of Flp and Cre recombinase activity that can be used to longitudinally track growth of tumours initiated by Flp recombinase activity and growth potential of specific cell states marked by Cre recombinase activity.

In *KPfrt;Hipp11*$^{GGCB/+}$*;Slc4a11*$^{MCD/+}$ mice, intratracheal delivery of viral Flp induces *GG*$^+$ *KP* tumours in which the *Slc4a11*$^+$ HPCS cells are marked by mScarlet (*MCD*) expression (Fig. 1c,d). The average proportion of mScarlet$^+$ cells in autochthonous lung tumours at 15–16 weeks post-tumour initiation (PTI) was 17.0 ± 4.3% (mean ± s.e.m.) of the total GFP$^+$ pool of cancer cells (Fig. 1e (top)), closely mirroring the proportion of HPCS cells detected by scRNA-seq (13.1 ± 1.6%)[4] (Fig. 1a,b). The majority (78.4 ± 1.7%) of the mScarlet$^+$ cells expressed integrin α2 (Fig. 1e (bottom) and Extended Data Fig. 2h), another sensitive but less-specific marker of the HPCS than *Slc4a11*[4] (Fig. 1b and Supplementary Table 1). Spatial analysis in tissue sections revealed that mScarlet$^+$ cells comprise a subset of *KP* LUAD tumours in situ, localizing as small clusters or single cells throughout the lung tumours (Fig. 1f,g). Robust enrichment of the HPCS gene expression signature[4] and *Slc4a11* expression was observed in isolated mScarlet$^+$ cells (hereafter, HPCS cells) when compared to mScarlet$^-$ cells (Extended Data Fig. 2i). Furthermore, scRNA-seq analysis of GFP$^+$ cancer cells revealed concordant expression of the *Slc4a11*^{MCD} allele (marked by DTR) with *Slc4a11* and *Itga2* endogenous transcripts within cancer cells (Extended Data Fig. 2j). In summary, these results establish the *Slc4a11*^{MCD} allele as a faithful reporter of the HPCS in situ.

The HPCS is a cell state transition hub

Lung cancer progression is defined by loss of lung epithelial identity, marked by downregulation of the lung epithelial master regulator NKX2.1 and acquisition of programs associated with the embryonic foregut, marked by induction of the embryonal transcriptional regulator HMGA2[34,35]. We found that the density of the malignant cells in the HPCS increases with histopathological grade (Fig. 1h), which was associated with a reduced density of NKX2.1$^+$ cells and an increase in HMGA2$^+$ cells[34,35] (Extended Data Fig. 3). We found that *Nkx2-1* expression is associated with early neoplastic cell states with an alveolar or lung endoderm identity (Extended Data Fig. 4a (top)), whereas the HPCS exhibits reduced *Nkx2-1* expression and an induction of *Hmga2* (Extended Data Fig. 4a (bottom) and 4b–e). A computational time-series analysis of our original longitudinal scRNA-seq data spanning early neoplasias and advanced adenocarcinomas[4] corroborated that the *Nkx2-1*-to-*Hmga2* switch occurs in the HPCS (marked by *Slc4a11* and *Plaur*) (Extended Data Fig. 4f,g). Furthermore, this analysis suggests two long-term trajectories of lung cancer evolution emanating from the HPCS—one that maintains *Hmga2* expression and eventually undergoes epithelial–mesenchymal transition and another that reacquires *Nkx2-1* expression, loses *Hmga2* and ends up in a proximal ciliated lung epithelial-like state (Extended Data Fig. 4f). These data suggest that the HPCS functions as a key transition between the alveolar states in early neoplasias and the cell states that emerge later in lung cancer progression.

We examined HPCS transition potential (plasticity) in situ by lineage tracing in autochthonous *KPfrt;Rosa26*$^{mTmG/+}$*;Slc4a11*$^{MCD/+}$ early lung neoplasias and in established LUADs at 6 or 12 weeks PTI, respectively, with a single pulse of tamoxifen (Fig. 2a,b). In *Rosa26*$^{mTmG/+}$ mice, all cells are marked by tdTomato fluorescence, which can be switched to GFP fluorescence with Cre activation[36]. We observed faithful labelling of the HPCS cells at 3 days after labelling (Extended Data Fig. 5a,b), the time required for washout of tamoxifen and its active metabolite 4-OHT. After 14 days of tracing, we found that the percentage of cells actively in the HPCS (mScarlet$^+$) of the total pool of traced (GFP$^+$) cells decreased precipitously compared with at the 3-day baseline in both the 6–8 and 12–14 week tracing windows (Extended Data Fig. 5a,b), indicating that the majority of the traced cells exited the HPCS and acquired new fates during the 14-day trace. Notably, the fraction of HPCS cells over total cancer cells (mScarlet$^+$EPCAM$^+$) remained stable over time, but the fraction of traced cells over total EPCAM$^+$ cells—enriched for malignant cells—in microdissected tumours expanded over time, indicative of HPCS clonal expansion (Extended Data Fig. 5c–f). scRNA-seq analysis of sorted traced (GFP$^+$) cells in both early (6-week) and advanced (12-week) tumours showed that >93% and >67% of the traced cells occupied the HPCS, respectively, at the 3-day post-tamoxifen baseline (Fig. 2c–f and Extended Data Fig. 5g,h). These data indicate that the *Slc4a11*$^{MCD/+}$ reporter system enables lineage tracing of the HPCS at high specificity, although transitions out of the HPCS may occur more rapidly in adenocarcinomas in the 12–14 week tracing window. A low rate of labelling outside the HPCS is also possible.

Notably, scRNA-seq analysis of traced cells revealed that the HPCS gives rise to all cancer cell states observed in both adenomas (6–8 weeks) and adenocarcinomas (12–14 weeks) (Fig. 2c–f and Extended Data Fig. 5g,h). Consistent with this finding, a large fraction of the

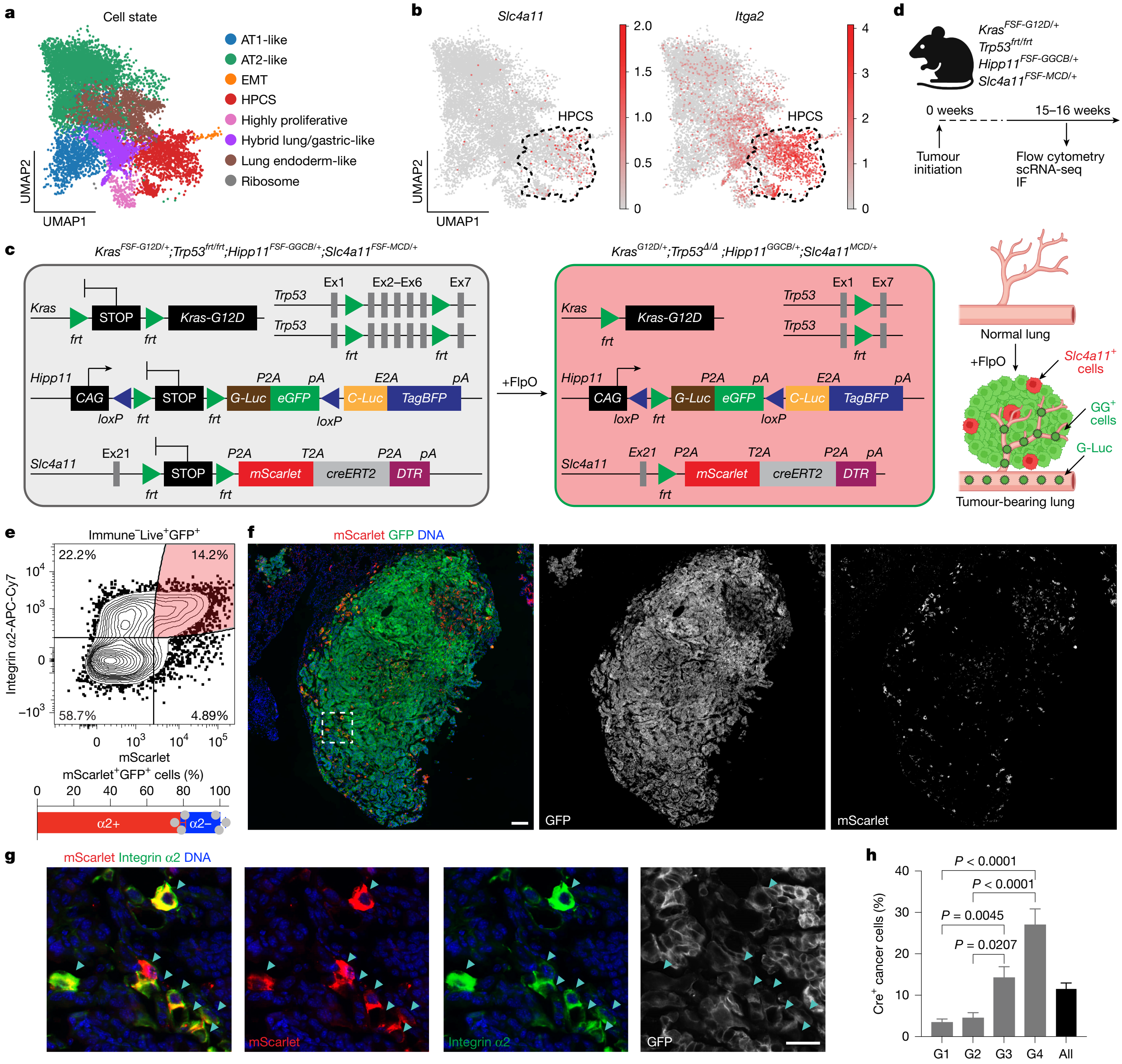


Fig. 1 | The *Slc4a11*$^{MCD/+}$ reporter system marks the HPCS in vivo. **a**, Single-cell transcriptomics from six *KP* mice at 15–16 weeks PTI identifies eight transcriptionally distinct LUAD cell states. **b**, The distribution of *Slc4a11* (left) and *Itga2* (right) gene expression. **c**, Schematic of *Kras*$^{FSF\text{-}G12D/+}$;*Trp53*$^{frt/frt}$;*Hipp11*$^{FSF\text{-}GGCB/+}$;*Slc4a11*$^{FSF\text{-}MCD/+}$ alleles before and after FlpO-mediated recombination (left). Right, FlpO lung delivery induces eGFP and G-Luc (GG$^+$, green). *Slc4a11*$^+$ cells express mScarlet, CreER^{T2} and DTR (MCD, red). **d**, Experimental schematic. Tumours were induced by intratracheally delivered PGK-FlpO lentivirus. **e**, Flow cytometry analysis highlighting co-expression of mScarlet and integrin α2 in CD45$^-$CD31$^-$CD11b$^-$CD11c$^-$F4/80$^-$TER-119$^-$Helix NP NIR$^-$ (Live$^+$)GFP$^+$ tumour cells (top). Bottom, the fraction of mScarlet$^+$GFP$^+$ cells positive for integrin α2. $n = 3$ mice. **f**, Immunofluorescence (IF) analysis of mScarlet$^+$ cells (red) within a GFP$^+$ LUAD tumour (green). Scale bar, 100 μm. **g**, Magnified view of the boxed area in **f**. The arrowheads indicate mScarlet and integrin α2 colocalization in GFP$^+$ cells. Scale bar, 20 μm. **h**, The percentage of Cre$^+$ (HPCS) cancer cells across histopathological grades in autochthonous *KPfrt;Slc4a11*$^{MCD/+}$ lung tumours. $n = 28$ tumours per grade, 5 mice. Statistical analysis was performed using one-way analysis of variance (ANOVA) with Dunnett's T3 multiple comparisons test. Data are mean ± s.e.m. The diagrams in **c** and **d** were created using BioRender. Tammela, T. (2025) https://BioRender.com/0lgfrw5.

HPCS-derived cells acquired expression of the differentiated lung epithelial markers *Nkx2-1* (encoding NKX2.1), *Sftpc* (encoding SPC) and *Hopx* and lost expression of *Hmga2* in both adenomas and adenocarcinomas at 14 days after lineage tracing, compared with at the 3-day baseline (Extended Data Fig. 5i–l). Taken together, these data indicate that the HPCS functions as a hub for cell state transitions between the early neoplastic, alveolar-like cell states and the progressed, endoderm-like states. Notably, these transitions occur through the HPCS in both forward and reverse directions along the axis of cancer progression (Extended Data Fig. 4f,g), towards both the alveolar and endoderm-like states, even in advanced adenocarcinomas. This acquisition of cellular diversity by the HPCS-derived cells in both adenomas and adenocarcinomas translated into robust, statistically significant increases in phenotypic volume, a quantitative measure of the diversity

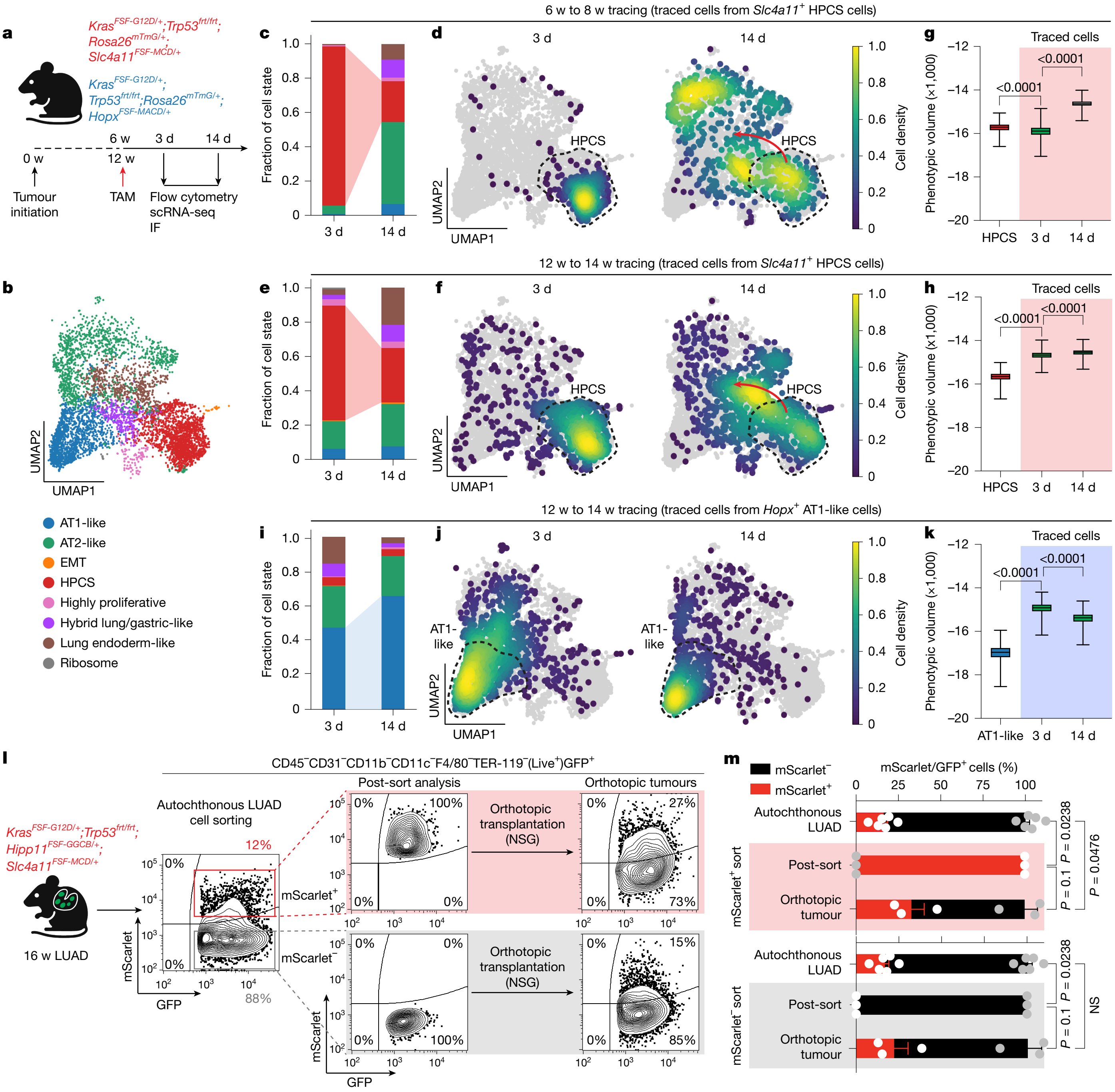


Fig. 2 | The HPCS is a hub for cell state transitions in vivo. a, The experimental design to lineage trace the HPCS (*Slc4a11*$^{FSF\text{-}MCD/+}$) or AT1-like (*Hopx*$^{FSF\text{-}MACD/+}$) cancer cells in autochthonous lung tumours. Note the switch between baseline (tdTomato$^+$) and lineage-traced (GFP$^+$) fluorescence in the *Rosa26*$^{mTmG/+}$ allele. **b**, scRNA-seq cell states in traced CD45$^-$CD31$^-$CD11b$^-$CD11c$^-$F4/80$^-$TER-119$^-$ Helix NP NIR$^-$(Live$^+$)GFP$^+$ LUAD cells collected at 6–8 or 12–14 weeks PTI. **c**,**d**, Transcriptomes of HPCS-derived cells traced at 6 weeks PTI collected at 3 days or 14 days after tracing. **c**, Cell state identities of HPCS-derived traced cells, coloured as in **b**. **d**, The density of traced cells. The HPCS is outlined. The red arrow indicates transdifferentiation. **e**,**f**, Transcriptomes of HPCS-derived cells traced at 12 weeks PTI collected at 3 days or 14 days after tracing. **e**, Cell state identities of HPCS-derived traced cells, coloured as in **b**. **f**, The density of traced cells. HPCS is outlined. The red arrow indicates transdifferentiation. **g**, Phenotypic volume calculated by bootstrap resampling (n = 1,000; 100 cells per iteration from pooled cells of 5 (3 days) or 4 mice (14 days, HPCS). P values were calculated using one-way ANOVA. **h**, Phenotypic volume as in **g** (pooled cells from 6 (3 days) or 3 (14 days, HPCS) mice. P values were calculated using one-way ANOVA. **i**–**k**, *Hopx*$^+$ AT1-like-cell-derived lineages traced at 12 weeks PTI and collected 3 days or 14 days after lineage tracing. **i**, The cell state identities of AT1-like-derived traced cells, coloured as in **b**. **j**, The density of traced cells. The AT1-like state is outlined. **k**, The phenotypic volume estimated as in **g**. Pooled cells from two mice per condition. P values were calculated using one-way ANOVA. **l**, Flow cytometry of primary LUAD cells collected from autochthonous lung tumours at 16 weeks PTI (left), post-sorted populations (middle) and orthotopically transplanted tumours (right). **m**, The proportion of mScarlet$^+$ and mScarlet$^-$ cells within GFP$^+$ populations from **l**. Autochthonous LUAD: n = 3 mice; post-sort: n = 3 samples; orthotopic tumour transplants: n = 3 mice. P values were calculated using Mann–Whitney U-tests. Data are mean ± s.e.m. P values for phenotypic volumes were calculated using one-way ANOVA with the Holm–Šídák test. The box plots show the median (centre line), 25th–75th percentiles (box limits) and minimum–maximum values (whiskers). The diagrams in **a** and **l** were created using BioRender. Tammela, T. (2025) https://BioRender.com/0lgfrw5.

of cellular phenotypes within cell populations[37] (Fig. 2g,h). Notably, the phenotypic volumes of the traced cells at 14 days after tamoxifen were significantly higher than the subset of HPCS cells alone or at the 3-day labelling baseline (Fig. 2g,h), suggesting that some traced cells exit the HPCS already during the 3-day tracing period, especially in the advanced tumours in the 12–14 week tracing window. These findings implicate the HPCS as a central driver of malignant cell state diversity in lung cancer.

To compare the plasticity of the HPCS to another defined, differentiated cancer cell state, we generated a reporter allele by knocking in an FSF-mScarlet-AkaLuc-CreER-DTR (hereafter, *FSF-MACD*) cassette into the *Hopx* locus in *KPfrt* mES cells (Extended Data Fig. 6a–c). *Hopx* marks AT1 cells in the normal lung and is enriched in the AT1-like cell state in LUAD tumours, although we note that its specificity for the AT1-like state is lower (82.2%) than that of *Slc4a11* for the HPCS (99.7%), as *Hopx* is also expressed in a subset of AT2-like cancer cells (Extended Data Fig. 6d–j and Supplementary Tables 2 and 3). In autochthonous *KPfrt;Hipp11*$^{FSF\text{-}GGCB/+}$*;Hopx*$^{FSF\text{-}MACD/+}$ LUAD tumours, cells expressing *Hopx* can be tracked by in vivo imaging of the super-bright luciferase AkaLuc[38] (Extended Data Fig. 6d,e), in addition to mScarlet fluorescence, which strongly coincides with *Hopx* mRNA and HOPX protein as well as the AT1-like state transcriptomic signature (Extended Data Fig. 6f–j), enabling faithful tracking of the AT1-like cell state in situ in LUAD tumours.

To measure the plasticity of the AT1-like cells, we performed similar lineage-tracing studies to those for the HPCS in established autochthonous *KPfrt;Rosa26*$^{mTmG/+}$*;Hopx*$^{MACD/+}$ adenocarcinomas (Fig. 2a and Extended Data Fig. 6k). Even though the fraction of mScarlet$^+$ cells remained constant in both the traced fraction and the total pool of cancer cells in the *KPfrt;Rosa26*$^{mTmG/+}$*;Hopx*$^{MACD/+}$ adenocarcinomas (Extended Data Fig. 6l,m), the fraction of cells deriving from the AT1-like cell statein the total pool of cancer cells declined over time (Extended Data Fig. 6n), suggesting clonal contraction. These results are markedly different from the HPCS tracing experiments, which showed rapid exit of the reporter-positive cells from the traced fraction and clonal expansion in adenocarcinomas (Extended Data Fig. 5a,b,d,f). scRNA-seq analysis of the lineage-traced AT1-like cells 3 days after a single pulse of tamoxifen (baseline) showed that 47.2 ± 5.46% (mean ± s.e.m.) of the traced cells reside in the AT1-like cell state, indicating relatively specific labelling of the AT1-like state, although, as expected, some labelling of AT2-like cells was also observed (Fig. 2i,j and Extended Data Fig. 6o,p). In marked contrast to the HPCS tracing experiments, the traced AT1-like cells remained in the AT1-like state and the phenotypic volume of the traced cells in fact contracted over the 12–14 week tracing window (Fig. 2i–k and Extended Data Fig. 6o,p). These results are consistent with a model in which high plasticity is concentrated in the HPCS, which gives rise to cancer cell states with fixed phenotypes and low plasticity, such as the AT1-like state.

Non-HPCS cells can acquire the HPCS

Given that the traced *Slc4a11*$^+$ cells rapidly exit the HPCS, yet the relative proportion of HPCS cells in tumours remains stable, we hypothesized that non-HPCS cells may be able to acquire the HPCS. To test this hypothesis, we transplanted primary non-HPCS (mScarlet$^-$GFP$^+$) cells from autochthonous *KPfrt;Hipp11*$^{GGCB/+}$*;Slc4a11*$^{MCD/+}$ lung tumours orthotopically into recipient mouse lungs through intratracheal instillation; pure HPCS (mScarlet$^+$GFP$^+$) cells were transplanted as a control (Fig. 2l–m). Consistent with a high growth potential (Extended Data Fig. 5d,f), the HPCS cells were around ninefold more efficient at forming tumours than the non-HPCS cells (Extended Data Fig. 6q). Notably, tumours arising from the non-HPCS population showed a similar fraction of mScarlet$^+$ cells to the autochthonous parent tumours (Fig. 2l,m), indicating that non-HPCS cells can acquire the HPCS. Similar results were obtained in subcutaneous serial allograft experiments (Extended Data Fig. 6r–t). Collectively, these data demonstrate that the HPCS is not a stem-like state that can arise only through self-renewal of pre-existing HPCS cells but, instead, dynamically interconverts between non-HPCS states.

HPCS drives tumour growth and progression

To examine the growth potential of the HPCS in situ, we used the *KPfrt;Hipp11*$^{FSF\text{-}GGCB/+}$*;Slc4a11*$^{FSF\text{-}MCD/+}$ model. As controls, we used *KPfrt;Hipp11*$^{FSF\text{-}GGCB/+}$*;Rosa26*$^{FSF\text{-}creERT2/+}$ mice to randomly trace a fraction of cancer cells—using a low dose of tamoxifen, as described previously[39]—and the *KPfrt;Hipp11*$^{FSF\text{-}GGCB/+}$*;Hopx*$^{FSF\text{-}MACD/+}$ mice to trace the AT1-like cell state (Fig. 3a,b). Longitudinal lineage tracing through secreted luciferases in established LUAD at 14–16 weeks after tumour initiation revealed that cancer cells derived from the HPCS have significantly higher growth potential and capacity for clonal expansion compared with derivatives of randomly labelled cells (Fig. 3c). By contrast, derivatives of the differentiated *Hopx*$^+$ AT1-like cell state have low growth potential (Fig. 3c), suggesting that the AT1-like cell state is outcompeted in LUAD tumours.

Despite the high growth potential of the HPCS, we found that the HPCS is highly quiescent (Fig. 3d and Extended Data Fig. 7a,b). In marked contrast, lineage-traced LUAD cells that had exited the HPCS (mScarlet$^-$GFP$^+$) showed high proliferative capacity as early as 3 days after tracing (mean ± s.e.m., 9.1 ± 2.7% Ki-67$^+$), which further increased by 14 days (15.1 ± 1.8% Ki-67$^+$) (Fig. 3e). The proliferation rate in the non-HPCS derivatives—that is, the non-traced cells (GFP$^-$PanCK$^+$)—remained constant over the 2-week lineage-tracing window (9.0 ± 1.1% (3 days) versus 10.2 ± 1.0% (14 days); Fig. 3e). EdU/BrdU dual pulsing[39] did not reveal a significant difference in S-phase duration between the HPCS derivatives and the non-HPCS derivatives (Extended Data Fig. 7c–f), nor did tamoxifen significantly confound our assessment of HPCS cell proliferation (Extended Data Fig. 7g). These data indicate that, while the HPCS is largely quiescent, a subset of its derivatives rapidly enter the cell cycle after exit from the HPCS, driving clonal expansion of the HPCS-derived cancer cells and tumour growth in situ.

Given that the HPCS emerges early during tumorigenesis and gives rise to advanced cell states, we hypothesized that it has a critical role in early tumour progression. To address this hypothesis, we ablated the HPCS in early autochthonous *KPfrt;Hipp11*$^{GGCB/+}$*;Slc4a11*$^{MCD/+}$ lung neoplasias by continuous systemic administration of DT starting at 1 week PTI and continuing to 7 weeks PTI (Fig. 3f,g). Consistent with the HPCS emerging after the initiation of early neoplasias, we did not detect a difference in the number of neoplastic nodules in the saline control versus DT groups (Extended Data Fig. 8a). However, we observed a substantial reduction in the overall neoplastic burden, which was largely driven by the absence of large (>0.15 mm^2) neoplastic lesions in the DT group (Fig. 3h and Extended Data Fig. 8b,c). Notably, ablation of the HPCS blunted progression of grade 1 atypical adenomatous hyperplasias (AAHs) to grade 2 adenomas (Extended Data Fig. 8d) and suppressed the emergence of cell states with high proliferation rates (Fig. 3i) or expression of HNF4α (Fig. 3j and Extended Data Fig. 8e), a marker of progression to a hybrid cell state with lung and gastric-like features that emerges after the HPCS in LUAD progression[4,40,41] (Extended Data Fig. 8f–h). These results indicate that the HPCS is not required for hyperplasia initiation but is critical for the progression of these neoplasias, giving rise to cell states with embryonal gene expression programs and/or a high mitotic index, which define more advanced cancers (Fig. 3k).

To test the role of the HPCS in the maintenance of established adenocarcinomas, we ablated the HPCS cells in autochthonous *KPfrt;Hipp11*$^{GGCB/+}$*;Slc4a11*$^{MCD/+}$ lung tumours at 14 weeks PTI through systemic administration of DT (Extended Data Fig. 8i,j). This intervention led to a rapid and pronounced ablation of mScarlet$^+$ cells and a decrease in tumour burden at 7 days (Extended Data Fig. 8k–m). ScRNA-seq analysis of residual GFP$^+$ cancer cells confirmed efficient

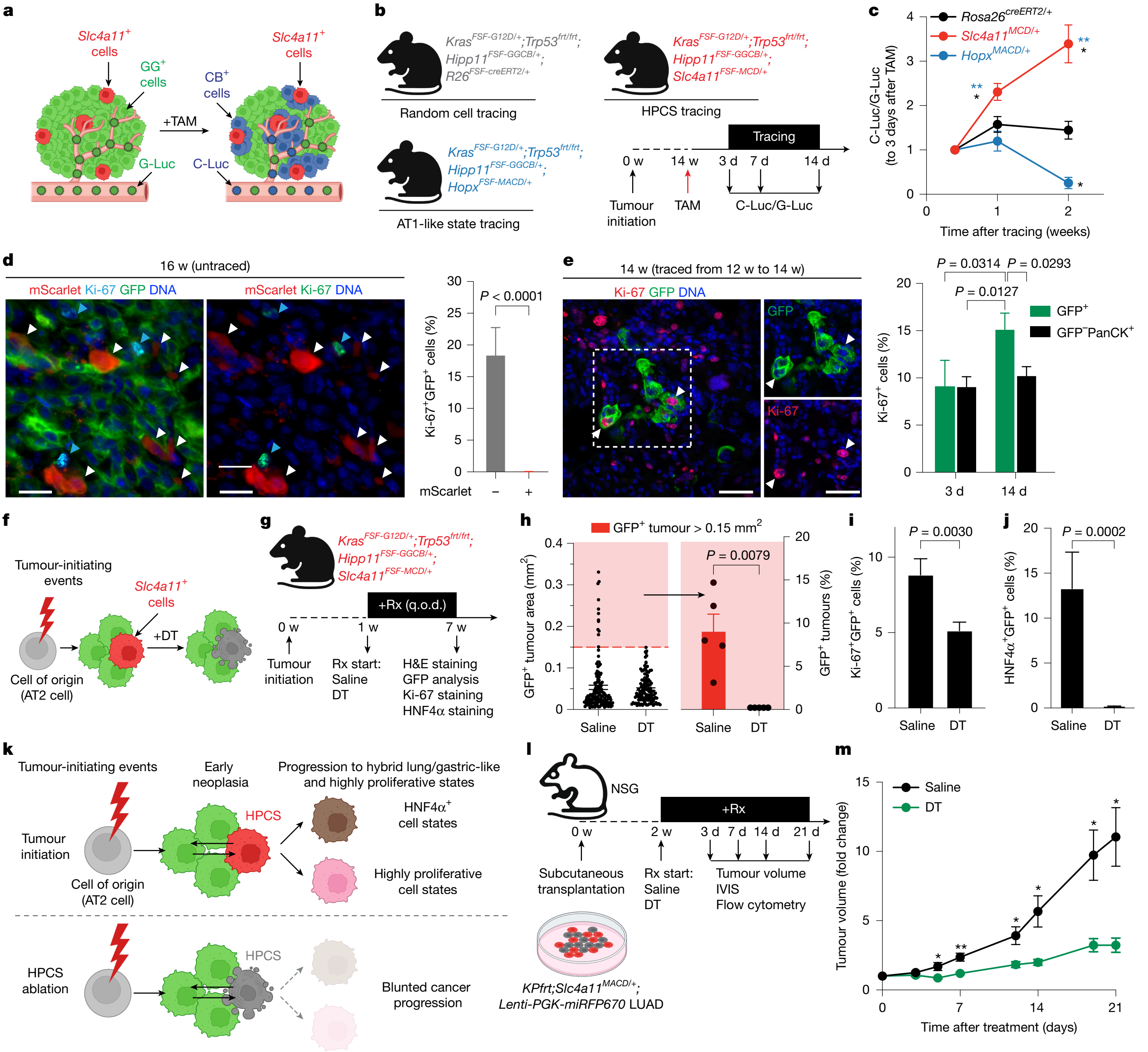


Fig. 3 | The HPCS drives tumour growth. **a**, *Hipp11*$^{GGCB/+}$;*Slc4a11*$^{MCD/+}$ LUAD before (left) and after (right) tamoxifen (TAM) administration. Tamoxifen switches GG to CB in *Slc4a11*$^{+}$ LUAD cells. **b**, Lineage-tracing comparison of random (*Rosa26*$^{creERT2/+}$), HPCS-derived (*Slc4a11*$^{MCD/+}$) and AT1-like cell state-derived (*Hopx*$^{MACD/+}$) LUAD cells in autochthonous lung tumours. **c**, The relative C-Luc/G-Luc ratio from blood normalized to day 3: HPCS (*Slc4a11*$^{MCD/+}$, $n = 5$ mice), random cell (*Rosa26*$^{creERT2/+}$, $n = 4$ mice) and AT1-like cell (*Hopx*$^{MACD/+}$, $n = 4$ mice) tracing. *P* values were calculated using two-way ANOVA with Tukey's test. PanCK, pan-cytokeratin. **d**, IF (left) and quantification (right) of Ki-67 in *KPfrt;Hipp11*$^{GGCB/+}$;*Slc4a11*$^{MCD/+}$ LUAD at 16 weeks PTI. mScarlet (white arrowheads) and Ki-67 (blue arrowheads) are mutually exclusive. Scale bars, 20 µm. $n = 49$ tumours, 3 mice. *P* values were calculated using Mann–Whitney *U*-tests. **e**, IF (left) and quantification (right) of Ki-67 cells in *KPfrt;Rosa26*$^{mTmG/+}$;*Slc4a11*$^{MCD/+}$ LUAD traced from 12 to 14 week PTI. Scale bars, 20 µm. Ki-67^{+}GFP^{+} cells: $n = 17$ tumours (3 days), 30 tumours (14 days); GFP^{-}PanCK^{+} cells: $n = 18$ tumours (3 days), 27 tumours (14 days). $n = 3$ mice per group. *P* values were calculated using two-way ANOVA with Tukey's test. **f**,**g**, Ablation of *Slc4a11*$^{+}$ HPCS cells in early neoplastic lesions. Mice received saline or DT (25 µg per kg, every other day (q.o.d.)). Schematics of ablation (**f**) and the treatment schedule (**g**) are shown. Rx, treatment. **h**, GFP^{+} tumour area (left). Saline: $n = 147$ tumours, 5 mice; DT: $n = 128$ tumours, 5 mice. Right, GFP^{+} tumours >0.15 mm^{2} in cross-sectional area. $n = 5$ mice per group. *P* values were calculated using Mann–Whitney *U*-tests. **i**, Ki-67^{+} cancer cells. $n = 21$ tumours, 3 mice per group. *P* values were calculated using Mann–Whitney *U*-tests. **j**, HNF4α^{+} cancer cells. $n = 12$ tumours, 3 mice per group. *P* values were calculated using Mann–Whitney *U*-tests. **k**, The effect of *Slc4a11*$^{+}$ HPCS cell ablation on subsequent cell states. **l**, Schematic of *Slc4a11*$^{+}$ HPCS ablation in subcutaneous allografts. Mice received saline or DT (25 µg per kg, every other day). IVIS, bioluminescence in vivo imaging system. **m**, Allograft volume after saline ($n = 16$ tumours, 8 mice) or DT ($n = 30$ tumours, 15 mice) administration. *P* values were calculated using Welch's *t*-tests. Data are mean ± s.e.m. The diagrams in **a**, **b**, **f**, **g**, **k** and **l** were created using BioRender. Tammela, T. (2025) https://BioRender.com/0lgfrw5.

ablation of *DTR*$^{+}$ cells, although we still observed *Slc4a11*$^{+}$ cells in residual tumours (Extended Data Fig. 8n). This suggested incomplete recombination of the FSF cassette in the *Slc4a11*$^{MCD/+}$ reporter allele in the residual tumours, enabling cancer cells to escape DT-mediated cytoablation. As such, our results probably underestimate the importance of the HPCS in the maintenance of LUAD tumours. To overcome this limitation, we generated a clonal *KP;Slc4a11*$^{MACD/+}$ LUAD cell line with complete excision of the FSF cassette (Extended Data Fig. 9a–c). As in the autochthonous tumours, the *Slc4a11*$^{MACD/+}$ reporter faithfully marks the HPCS in subcutaneous allografts in this model (Extended Data Fig. 9d–f), enabling lineage tracing (Extended Data Fig. 9g–i) and ablation of the HPCS cells in vivo (Fig. 3l and Extended Data Fig. 9j–l). Ablation of the HPCS produced a substantial anti-tumour response (Fig. 3m). Notably, ablation of the HPCS led to a much more robust reduction in cell state heterogeneity than the ablation the AT1-like cells (Extended Data Fig. 9m,n). By contrast, prolonged cytoablation of the AT1-like cell state in subcutaneous *KPfrt;Hopx*$^{MACD/+}$ allografts had no significant effect on tumour growth (Extended Data Fig. 10a–c), consistent with our results on longitudinal lineage tracing of the AT1-like state (Fig. 3c). Collectively, these results indicate that the HPCS is critical for the progression and growth of early lung neoplasias and established adenocarcinomas.

Drug resistance originates from the HPCS

We applied our longitudinal dual-luciferase lineage tracing system to test whether the HPCS gives rise to therapy-resistant states and minimal residual disease (MRD) in autochthonous *KPfrt;Hipp11*$^{GGCB/+}$*;Slc4a11*$^{MCD/+}$ lung tumours subjected to models of two standard-of-care therapies: cisplatin chemotherapy[22] and allele-specific KRAS(G12D) oncoprotein-targeted therapy (MRTX1133)[9,42,43] (Fig. 4a). Relative to saline, MRTX1133 markedly reduced non-HPCS-derived (G-Luc^{+}) cells, whereas cisplatin caused only a modest, non-significant decrease (Extended Data Fig. 10d). By contrast, HPCS-derived C-Luc^{+} cells survived and continued to grow under cisplatin, and initially expanded and then persisted under MRTX1133 (Extended Data Fig. 10e). Fluorescence-based lineage tracing in *KPfrt;Rosa26*$^{mTmG/+}$*;Slc4a11*$^{MCD/+}$ tumours corroborated these findings (Fig. 4a and Extended Data Fig. 10f). Thus, therapeutic responses primarily eliminate non-HPCS-derived cells, whereas HPCS-derived cells become strongly enriched in MRD (Fig. 4b and Extended Data Fig. 10g). Although the HPCS shows no relative growth advantage under MRTX1133 versus saline (Fig. 4b), the non-HPCS pool regresses with MRTX1133 (Extended Data Fig. 10d), while HPCS-derived cells continue to expand, albeit more slowly than under saline (Extended Data Fig. 10e).

We next investigated the fate of the HPCS under therapeutic pressure using scRNA-seq analysis of lineage-traced HPCS-derived cells in *KPfrt;Rosa26*$^{mTmG/+}$*;Slc4a11*$^{MCD/+}$ tumours after 3 weeks of therapy (Fig. 4a). Both cisplatin and KRAS(G12D) inhibition reduced the phenotypic volume of traced cells (Fig. 4c), indicating that therapy directs differentiation of the HPCS into fewer states relative to unperturbed tumours. Cisplatin promoted AT2-like and hybrid lung/gastric-like fates, whereas KRAS inhibition drove AT1-like differentiation, consistent with the AT1-like state's intrinsic resistance to KRAS inhibition, which we have previously reported[9] (Fig. 4d and Extended Data Fig. 10h). Both therapies decreased the fraction of cells persisting in the HPCS within the traced and total pools (Fig. 4d and Extended Data Fig. 10i,j), indicating that the HPCS itself is not drug resistant but it gives rise to MRD composed of resistant states. Combining DT-mediated HPCS ablation with cisplatin or MRTX1133 produced near-complete tumour eradication (Fig. 4e–g), whereas ablation of *Hopx*$^{+}$ AT1-like cells alone or with cisplatin did not yield therapeutic benefit (Extended Data Fig. 10k,l). Thus, the HPCS is a key transitional node that seeds therapy-resistant states and MRD.

We found strong enrichment of *Plaur* (encoding uPAR) in the HPCS (Fig. 4h and Extended Data Figs. 2j, 3, 4g and 11a). We delivered uPAR-directed CAR T cells (hereafter, uPAR CAR T cells), originally developed to eliminate senescent cells[44], systemically to mice with *KP;Rosa26*$^{tdTomato/+}$ lung tumours, which eradicated both uPAR^{+} and integrin α2^{+} cancer cells (Extended Data Fig. 11b–f), indicating effective HPCS cytoablation. Adoptive transfer of uPAR CAR T cells into mice with *KP;Slc4a11*$^{MACD/+}$ LUAD allografts produced a strong anti-tumour response with loss of *Slc4a11*-AkaLuc signal and delayed growth (Fig. 4i–k). Together, these findings show that the HPCS is a critical and therapeutically targetable cancer cell state that drives tumour progression, heterogeneity and therapy resistance.

HPCS recurs in carcinomas and epithelia

Given its essentiality in LUAD, we hypothesized that the HPCS may reflect a critical cancer cell state across multiple cancers. We confirmed presence of the HPCS signature in colorectal cancer (CRC), head and neck cancer (HNC), LUAD, ovarian cancer (OC), pancreatic ductal adenocarcinoma (PDAC), prostate adenocarcinoma (PCa) and cutaneous squamous cell carcinoma (SCC) single-cell datasets (Fig. 5a). We next correlated it with candidate plastic states linked to tumour progression and drug resistance and found that the HPCS scores significantly correlated with the core high-relapse cell state (coreHRC)[45] or oncofetal (OF)[46] program critical for CRC drug resistance and metastatic recurrence; with the basal state associated with poor prognosis in PDAC[47]; and with a plastic PCa cell state associated with transition to drug-resistant neuroendocrine cancer[16](Extended Data Fig. 11g). These correlations imply that the HPCS reflects a conserved cell state across carcinomas.

To further evaluate its broader relevance, we compared the HPCS signature with recurrent archetype states shared across human cancers independently of lineage[48,49]. Notably, the HPCS signature showed the highest correlation with the stress archetype across mouse and human carcinomas (Extended Data Figs. 11h and 12a). In human tumours, *SLC4A11*, *PLAUR*, the HPCS signature and the stress program likewise correlated with the plastic subsets previously explored (Fig. 5a,b). Moreover, *Slc4a11* and *Plaur* expression correlated with the previously defined KRT8^{+} alveolar intermediate cell state, a predicted LUAD precursor state in mouse and human tumours[5] (Extended Data Fig. 12b), although *Krt8* expression is broader and not restricted to the HPCS (Extended Data Fig. 12c). The convergence of these programs suggests functional similarity across carcinoma types.

As the HPCS drives cell state expansion and tumour growth, we hypothesized that it may represent a malignant correlate of an epithelial injury-induced transition state[3]. In the lung, this state arises in a subset of AT2 cells and is required for regeneration[50–52]. The HPCS signature, *Slc4a11* and *Plaur* strongly overlapped with this state in a scRNA-seq dataset of alveolar epithelial cells after lung injury[50] (Fig. 5c and Extended Data Fig. 12d). Consistent with this, *Slc4a11* was induced in AT2 cells in situ in hyperoxia-injured but not in uninjured mouse lung (Fig. 5d). *Slc4a11*, *Plaur*, the HPCS signature and the stress signature were also enriched in injury-associated regenerative epithelial states in pancreas, intestine and skin (Fig. 5c and Extended Data Fig. 12e). Thus, the HPCS is likely to reflect co-option of a latent regenerative program normally activated during epithelial injury.

Discussion

Using advanced multifunctional genetically engineered autochthonous lung cancer models, we demonstrate that a malignant HPCS is critical for cancer progression, maintenance and resistance to standard-of-care therapies. Functionally, the HPCS exhibits broad cell state transition capacity, including generating highly proliferative states, and therefore shares features with both adult stem cells and CSCs, including robust growth and differentiation potential[53]. However, in contrast to CSCs, the HPCS is acquired by neoplastic cells early in tumorigenesis and later by other cancer cell states in established adenocarcinomas. Instead of resembling homeostatic stem cells or CSCs, the HPCS is molecularly and functionally similar to a transient, plastic regenerative state induced in differentiated AT2

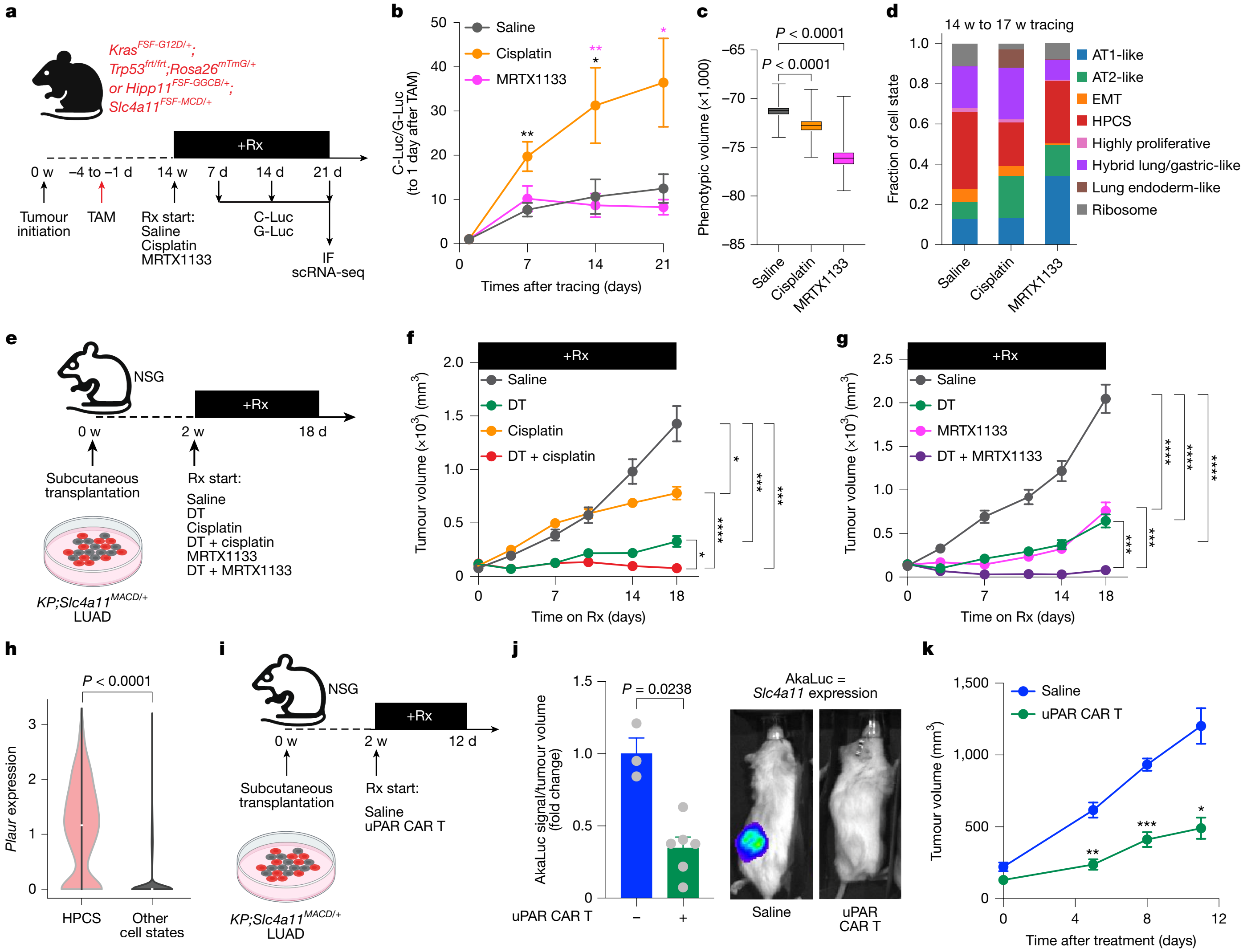


Fig. 4 | Drug-resistant cell states originate from the HPCS. a, Experiments assessing the effects of chemotherapy versus KRAS(G12D)-targeted therapy on HPCS-derived LUAD states. Mice received one or two doses of tamoxifen (200 mg per kg) before saline, cisplatin (1.5 mg per kg, every 3 days) or MRTX1133 (30 mg per kg, twice per day) for 3 weeks. G-Luc and C-Luc activity was recorded 1 day after tamoxifen and every 7 days thereafter. Traced CD45$^-$CD31$^-$CD11b$^-$CD11c$^-$F4/80$^-$TER-119$^-$Helix NP NIR$^-$(Live$^+$)GFP$^+$ LUAD cells from autochthonous *KPfrt;Rosa26*$^{mTmG/+}$*;Slc4a11*$^{MCD/+}$ tumours were isolated for scRNA-seq after 21 days of therapy. **b**, C-Luc/G-Luc ratios at the indicated timepoints (normalized to day 1) in the saline ($n = 5$), cisplatin ($n = 7$) or MRTX1133 ($n = 8$) treatment groups. *P* values were calculated using Welch's *t*-tests. **c**, The phenotypic volume of traced cells after 21 days of therapy, calculated as in Fig. 2g. $n = 4$ (saline), 6 (cisplatin) or 7 (MRTX1133). *P* values were calculated using one-way ANOVA with Dunnett's multiple-comparisons test. The box plots show the median (centre line), 25th–75th percentiles (box limits) and the minimum–maximum values (whiskers). **d**, The cell state composition of lineage-traced cells. **e**, Experimental design testing *Slc4a11*$^+$ HPCS LUAD cell ablation by DT (25 µg per kg, every other day) combined with saline, cisplatin chemotherapy (1.5 mg per kg, every 3 days) or KRAS(G12D)-targeted therapy (MRTX1133, 30 mg per kg, twice per day) in subcutaneous *KP;Slc4a11*$^{MACD/+}$ LUAD reporter allografts. **f**,**g**, Allograft volumes under the indicated therapies (cisplatin (**f**) and MRTX1133 (**g**) with DT). $n = 10$ tumours per group, 5 mice. *P* values were calculated using Mann–Whitney *U*-tests. **h**, Violin plot of *Plaur* gene expression in the HPCS versus other cell states. *P* values were calculated using Mann–Whitney *U*-tests. **i**, The design for intratumoural human-derived CAR T cell therapy targeting mouse uPAR on subcutaneous *KP;Slc4a11*$^{MACD/+}$ LUAD reporter allografts. **j**, *Slc4a11*-AkaLuc signal normalized to tumour volume measured 8 days after intratumoural injection. $n = 3$ (saline) or 6 (uPAR CAR T cells) mice; *P* values were calculated using Mann–Whitney *U*-tests. **k**, Allograft volume under saline ($n = 3$ mice) or uPAR CAR T cells ($n = 6$ mice). *P* values were calculated using Mann–Whitney *U*-tests. Data are mean ± s.e.m. The diagrams in **a**, **e** and **i** were created using BioRender. Tammela, T. (2025) https://BioRender.com/0lgfrw5.

cells by injury-associated paracrine signals[50–52], consistent with similar acquired plasticity in diverse epithelial regeneration contexts[3,54,55]. Future work will identify specific niche-derived inducers of the HPCS in vivo and use reductionist systems to induce the HPCS in non-plastic cancer cells.

The HPCS persists and proportionally expands with advancing histopathological grade after its early emergence. As its physiological correlate is induced by epithelial injury, regenerative niche signals may similarly induce the HPCS in a subset of cancer cells, facilitating cell state transitions. In this model, the HPCS serves as a transition hub rather than sitting atop a hierarchy, in contrast to the CSC model (Fig. 5e), thereby providing a central mechanism for maintaining intratumoural heterogeneity. Conversely, the AT1-like state exhibits low plasticity and growth potential, representing the opposite end of a malignant plasticity spectrum[4]. Other malignant states may also harbour plasticity; for example, the hybrid lung/gastric state shows co-accessibility at pulmonary and gastric genes[40,41], suggesting the potential for lung- or gastric-like differentiation. Similar lineage-tracing and ablation studies are needed to functionally interrogate these states in situ.

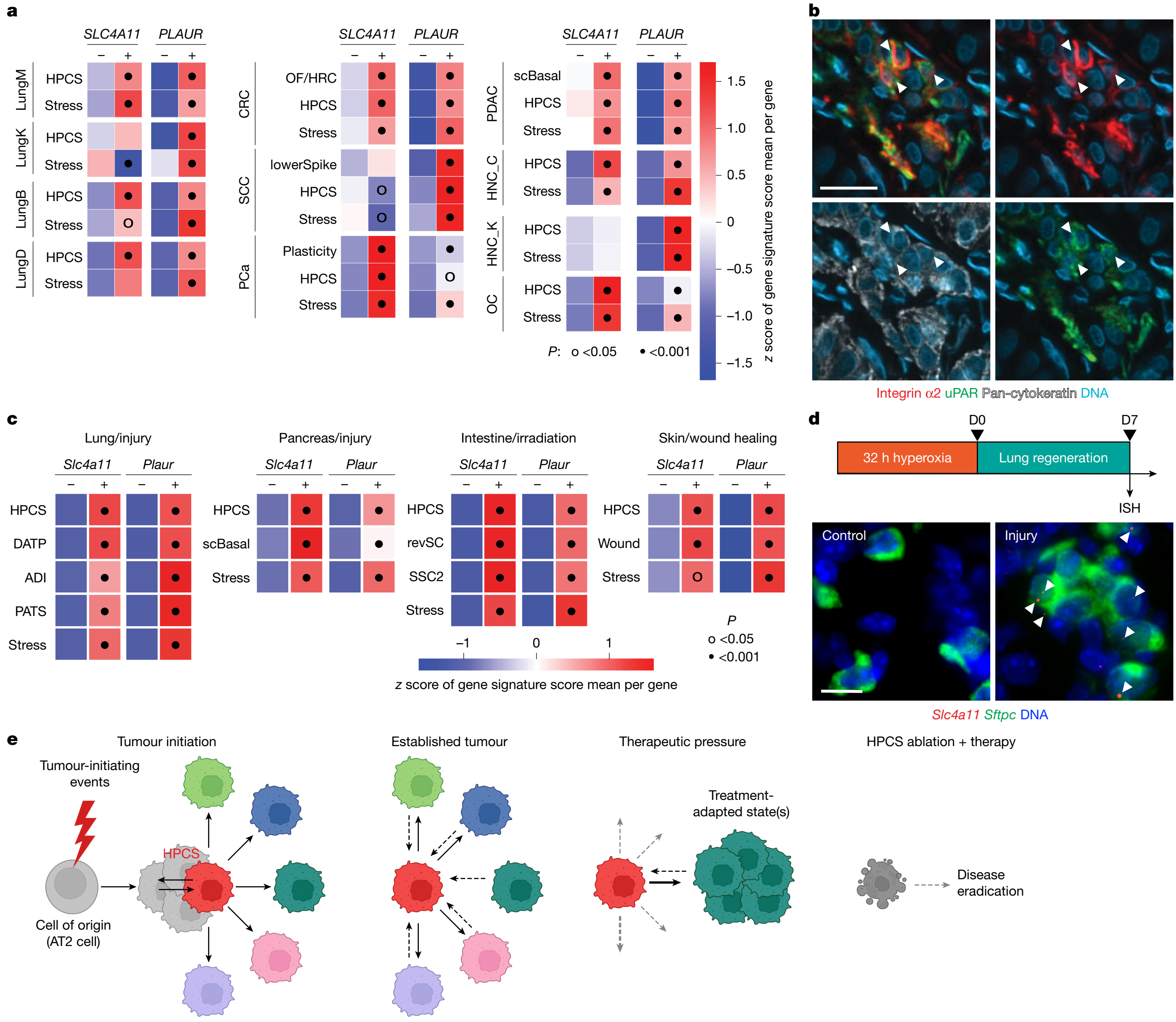


Fig. 5 | An HPCS-like state marked by *Slc4a11* and *Plaur* is ubiquitously present in multiple carcinomas and regenerating epithelia. a, The mean gene signature score in *SLC4A11*- or *PLAUR*-negative and -positive cells scaled by a program-specific *z* score across human scRNA-seq datasets (Supplementary Table 5). For each dataset gene pair, significance is depicted in the positive cell column (*t*-test). Data are from four human LUAD datasets (LungM[5], LungK[57], LungB[58], LungD[33]) and plastic cell states (OF/HRC[45,46], lowerSpike[15], plasticity[16]) in other cancers (CRC, SCC, PCa, PDAC, head and neck cancers (HNC_C/HNC_K) and ovarian cancer (OC)). **b**, uPAR (green), integrin α2 (red) and pan-cytokeratin (white) in human LUAD. The arrowheads indicate uPAR^{+}integrin α2^{+} double-positive cells. Scale bar, 25 µm. **c**, The mean gene signature score in *Slc4a11*- or *Plaur*-negative and -positive cells scaled by a program-specific *z* score in mouse scRNA-seq datasets of epithelial tissue regeneration, including lung (damage-associated transient progenitor (DATP)[50]; KRT8^{+} alveolar differentiation intermediate (ADI)[52]; pre-AT1 transitional cell state (PATS))[51], pancreas (scBasal cell state)[47], intestine (revival stem cell (revSC)[59]; singly profiled stem cell cluster 2 (SSC2))[60] and skin (wound)[61]. Significance is depicted in the positive cell column, as determined using *t*-tests. **d**, *Slc4a11* (red) and *Sftpc* (green) mRNA in normal lungs versus 7 days after hyperoxia-induced lung injury. Scale bar, 10 µm. **e**, Schematic of key findings (see the main text). All *P* values are unadjusted. Reference datasets and gene signatures are provided in Supplementary Table 5. The diagram in **e** was created using BioRender. Tammela, T. (2025) https://BioRender.com/0lgfrw5.

We demonstrate that the HPCS is a stepping stone for expanding intratumoural heterogeneity during early tumorigenesis and is critical for LUAD progression beyond small nodules. On the basis of this finding, the HPCS enables transition of premalignant lesions—which are abundant in normal, healthy tissues—to lethal malignancy. Thus, targeting the HPCS in early neoplasias may enable early cancer interception. The HPCS also remains central in established tumours, as its ablation induces tumour regression and growth suppression. First, HPCS-derived cells possess significantly higher growth potential than average LUAD cells, and therefore ablating the HPCS removes a major growth driver. Second, because non-HPCS cells can acquire the HPCS, continuous cytoablation eliminates cells as they convert, contributing to anti-tumour activity. This conversion rate may accelerate as HPCS niches are vacated by the cytoablation, consistent with previous findings in stem-like colon cancer cell ablation[56]. Third, the HPCS may exert non-cell-autonomous effects, such as immunosuppression or secretion of paracrine factors that support tumour growth; notably, multiple genes that encode growth factors and cytokines enriched in the HPCS (for example, *Tnf*, *Il23*, *Vegf*, *Cxcl2*, *Wnt7a*, *Wnt7b*, *Wnt10a*, *Areg*, *Ereg* and *Tgfa*)[4] have established or probable roles in cancer progression. Understanding how the HPCS contributes to the composition and organization of the tumour ecosystem is a key future direction.

Perhaps the most intriguing explanation for the anti-tumour effect is that eliminating the HPCS disrupts the bridge enabling cancer cells to transition and adapt during tumour evolution, thereby stalling growth. In support of this model, early HPCS ablation prevented emergence of states needed for hyperplasia-to-adenoma or adenoma-to-adenocarcinoma transitions. This adaptability model also aligns with the strong synergy of HPCS ablation with standard LUAD therapies: removing the HPCS prevents treatment-sensitive states from transitioning into resistant states—such as the AT1-like state resistant to KRAS inhibition[9]—leading to robust therapeutic efficacy.

Our findings identify the HPCS as a therapeutic entry point for targeting plasticity-enabled resistance, challenging the paradigm that oncologic therapy inevitably yields a drug-tolerant MRD. We find that the HPCS is a major MRD source, particularly after chemotherapy, positioning it as a prime target for limiting or eradicating MRD (Fig. 5e). uPAR-directed CAR T cell therapy targeting the HPCS in LUAD and potentially other carcinomas is one translational avenue; however, as uPAR is enriched but not universal within the HPCS, antigen escape and relapse remain possible, implying that eliminating this state may require multi-antigen targeting. The similarity of the HPCS to the conserved stress-signature-associated state across diverse human tumours—marked by *SLC4A11* and *PLAUR*—suggests that plasticity is concentrated within the stress-signature-associated state, offering broad therapeutic potential. Importantly, in healthy tissues, the HPCS program emerges only during epithelial regeneration and is absent in homeostasis, implying that cytoablation or targeting of HPCS drivers may be safe in patients. Taken together, our findings show that targeting the HPCS is a promising strategy for suppressing tumour progression and eradicating therapy resistance across carcinomas.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-025-09985-x.

1. Perez-Gonzalez, A., Bevant, K. & Blanpain, C. Cancer cell plasticity during tumor progression, metastasis and response to therapy. *Nat. Cancer* **4**, 1063–1082 (2023).
2. Quintanal-Villalonga, A. et al. Lineage plasticity in cancer: a shared pathway of therapeutic resistance. *Nat. Rev. Clin. Oncol.* **17**, 360–371 (2020).
3. Torborg, S. R., Li, Z., Chan, J. E. & Tammela, T. Cellular and molecular mechanisms of plasticity in cancer. *Trends Cancer* **8**, 735–746 (2022).
4. Marjanovic, N. D. et al. Emergence of a high-plasticity cell state during lung cancer evolution. *Cancer Cell* **38**, 229–246 (2020).
5. Han, G. et al. An atlas of epithelial cell states and plasticity in lung adenocarcinoma. *Nature* **627**, 656–663 (2024).
6. Patel, A. S. & Yanai, I. A developmental constraint model of cancer cell states and tumor heterogeneity. *Cell* **187**, 2907–2918 (2024).
7. Cazet, A. S. et al. Targeting stromal remodeling and cancer stem cell plasticity overcomes chemoresistance in triple negative breast cancer. *Nat. Commun.* **9**, 2897 (2018).
8. Boumahdi, S. & de Sauvage, F. J. The great escape: tumour cell plasticity in resistance to targeted therapy. *Nat. Rev. Drug Discov.* **19**, 39–56 (2020).
9. Li, Z. et al. Alveolar differentiation drives resistance to KRAS inhibition in lung adenocarcinoma. *Cancer Discov.* **14**, 308–325 (2024).
10. Tong, X. et al. Adeno-to-squamous transition drives resistance to KRAS inhibition in *LKB1* mutant lung cancer. *Cancer Cell* **42**, 413–428 (2024).
11. Rozenblatt-Rosen, O. et al. The Human Tumor Atlas Network: charting tumor transitions across space and time at single-cell resolution. *Cell* **181**, 236–249 (2020).
12. Tirosh, I. & Suva, M. L. Cancer cell states: lessons from ten years of single-cell RNA-sequencing of human tumors. *Cancer Cell* **42**, 1497–1506 (2024).
13. Patel, A. P. et al. Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science* **344**, 1396–1401 (2014).
14. Tirosh, I. et al. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* **352**, 189–196 (2016).
15. Taylor, M. A. et al. Stem-cell states converge in multistage cutaneous squamous cell carcinoma development. *Science* **384**, eadi7453 (2024).
16. Chan, J. M. et al. Lineage plasticity in prostate cancer depends on JAK/STAT inflammatory signaling. *Science* **377**, 1180–1191 (2022).
17. Kratzer, T. B. et al. Lung cancer statistics, 2023. *Cancer* **130**, 1330–1348 (2024).
18. Siegel, R. L., Kratzer, T. B., Giaquinto, A. N., Sung, H. & Jemal, A. Cancer statistics, 2025. *CA Cancer J. Clin.* **75**, 10–45 (2025).
19. Jackson, E. L. et al. Analysis of lung tumor initiation and progression using conditional expression of oncogenic K-ras. *Genes Dev.* **15**, 3243–3248 (2001).
20. Jackson, E. L. et al. The differential effects of mutant p53 alleles on advanced murine lung cancer. *Cancer Res.* **65**, 10280–10288 (2005).
21. Young, N. P., Crowley, D. & Jacks, T. Uncoupling cancer mutations reveals critical timing of p53 loss in sarcomagenesis. *Cancer Res.* **71**, 4040–4047 (2011).
22. Oliver, T. G. et al. Chronic cisplatin treatment promotes enhanced damage repair and tumor progression in a mouse model of lung cancer. *Genes Dev.* **24**, 837–852 (2010).
23. Groger, N. et al. SLC4A11 prevents osmotic imbalance leading to corneal endothelial dystrophy, deafness, and polyuria. *J. Biol. Chem.* **285**, 14467–14474 (2010).
24. Leone, D. P. et al. Tamoxifen-inducible glia-specific Cre mice for somatic mutagenesis in oligodendrocytes and Schwann cells. *Mol. Cell. Neurosci.* **22**, 430–440 (2003).
25. Buch, T. et al. A Cre-inducible diphtheria toxin receptor mediates cell lineage ablation after toxin administration. *Nat. Methods* **2**, 419–426 (2005).
26. Lee, C. L. et al. Generation of primary tumors with Flp recombinase in FRT-flanked p53 mice. *Dis. Model. Mech.* **5**, 397–402 (2012).
27. Tannous, B. A. Gaussia luciferase reporter assay for monitoring biological processes in culture and in vivo. *Nat. Protoc.* **4**, 582–591 (2009).
28. Cormack, B. P., Valdivia, R. H. & Falkow, S. FACS-optimized mutants of the green fluorescent protein (GFP). *Gene* **173**, 33–38 (1996).
29. Nakajima, Y., Kobayashi, K., Yamagishi, K., Enomoto, T. & Ohmiya, Y. cDNA cloning and characterization of a secreted luciferase from the luminous Japanese ostracod, *Cypridina noctiluca*. *Biosci. Biotechnol. Biochem.* **68**, 565–570 (2004).
30. Subach, O. M., Cranfill, P. J., Davidson, M. W. & Verkhusha, V. V. An enhanced monomeric blue fluorescent protein with the high chemical stability of the chromophore. *PLoS ONE* **6**, e28674 (2011).
31. Charles, J. P. et al. Monitoring the dynamics of clonal tumour evolution in vivo using secreted luciferases. *Nat. Commun.* **5**, 3981 (2014).
32. Hippenmeyer, S. et al. Genetic mosaic dissection of Lis1 and Ndel1 in neuronal migration. *Neuron* **68**, 695–709 (2010).
33. Dost, A. F. M. et al. Organoids model transcriptional hallmarks of oncogenic KRAS activation in lung epithelial progenitor cells. *Cell Stem Cell* **27**, 663–678 (2020).
34. Snyder, E. L. et al. Nkx2-1 represses a latent gastric differentiation program in lung adenocarcinoma. *Mol. Cell* **50**, 185–199 (2013).
35. Winslow, M. M. et al. Suppression of lung adenocarcinoma progression by Nkx2-1. *Nature* **473**, 101–104 (2011).
36. Muzumdar, M. D., Tasic, B., Miyamichi, K., Li, L. & Luo, L. A global double-fluorescent Cre reporter mouse. *Genesis* **45**, 593–605 (2007).
37. Azizi, E. et al. Single-cell map of diverse immune phenotypes in the breast tumor microenvironment. *Cell* **174**, 1293–1308 (2018).
38. Iwano, S. et al. Single-cell bioluminescence imaging of deep tissue in freely moving animals. *Science* **359**, 935–939 (2018).
39. Driessens, G., Beck, B., Caauwe, A., Simons, B. D. & Blanpain, C. Defining the mode of tumour growth by clonal analysis. *Nature* **488**, 527–530 (2012).
40. Feng, A., Yermalovich, A. & Meyerson, M. Lineage rewiring in lung adenocarcinoma via HNF4α and NKX2-1 dynamics. *Genes Dev.* https://doi.org/10.1101/gad.353142.125 (2025).
41. Fort, G. et al. Opposing lineage specifiers induce a protumor hybrid identity state in lung adenocarcinoma. *Genes Dev.* https://doi.org/10.1101/gad.352742.125 (2025).
42. Wang, X. et al. Identification of MRTX1133, a noncovalent, potent, and selective KRAS(G12D) Inhibitor. *J. Med. Chem.* **65**, 3123–3133 (2022).
43. Hallin, J. et al. Anti-tumor efficacy of a potent and selective non-covalent KRAS(G12D) inhibitor. *Nat. Med.* **28**, 2171–2182 (2022).
44. Amor, C. et al. Senolytic CAR T cells reverse senescence-associated pathologies. *Nature* **583**, 127–132 (2020).
45. Canellas-Socias, A. et al. Metastatic recurrence in colorectal cancer arises from residual EMP1$^+$ cells. *Nature* **611**, 603–613 (2022).
46. Han, T. et al. Lineage reversion drives WNT independence in intestinal cancer. *Cancer Discov.* **10**, 1590–1609 (2020).
47. Raghavan, S. et al. Microenvironment drives cell state, plasticity, and drug response in pancreatic cancer. *Cell* **184**, 6119–6137 (2021).
48. Barkley, D. et al. Cancer cell states recur across tumor types and form specific interactions with the tumor microenvironment. *Nat. Genet.* **54**, 1192–1201 (2022).
49. Gavish, A. et al. Hallmarks of transcriptional intratumour heterogeneity across a thousand tumours. *Nature* **618**, 598–606 (2023).
50. Choi, J. et al. Inflammatory signals induce AT2 cell-derived damage-associated transient progenitors that mediate alveolar regeneration. *Cell Stem Cell* **27**, 366–382 (2020).
51. Kobayashi, Y. et al. Persistence of a regeneration-associated, transitional alveolar epithelial cell state in pulmonary fibrosis. *Nat. Cell Biol.* **22**, 934–946 (2020).
52. Strunz, M. et al. Alveolar regeneration through a Krt8$^+$ transitional stem cell state that persists in human lung fibrosis. *Nat. Commun.* **11**, 3559 (2020).
53. Batlle, E. & Clevers, H. Cancer stem cells revisited. *Nat. Med.* **23**, 1124–1134 (2017).
54. Ge, Y. et al. Stem cell lineage infidelity drives wound repair and cancer. *Cell* **169**, 636–650 (2017).
55. Gola, A. & Fuchs, E. Environmental control of lineage plasticity and stem cell memory. *Curr. Opin. Cell Biol.* **69**, 88–95 (2021).

56. Shimokawa, M. et al. Visualization and targeting of LGR5$^+$ human colon cancer stem cells. *Nature* **545**, 187–192 (2017).
57. Kim, N. et al. Single-cell RNA sequencing demonstrates the molecular and cellular reprogramming of metastatic lung adenocarcinoma. *Nat. Commun.* **11**, 2285 (2020).
58. Bischoff, P. et al. Single-cell RNA sequencing reveals distinct tumor microenvironmental patterns in lung adenocarcinoma. *Oncogene* **40**, 6748–6758 (2021).
59. Ayyaz, A. et al. Single-cell transcriptomes of the regenerating intestine reveal a revival stem cell. *Nature* **569**, 121–125 (2019).
60. Morral, C. et al. p53 promotes revival stem cells in the regenerating intestine after severe radiation injury. *Nat. Commun.* **15**, 3018 (2024).
61. Haensel, D. et al. Defining epidermal basal cell states during skin homeostasis and wound healing using single-cell transcriptomics. *Cell Rep.* **30**, 3932–3947 (2020).

Methods

Mice

Previously published genetically engineered mouse strains were used in this study: *Kras*$^{LSL\text{-}G12D/+}$ (ref. 19), *Trp53*$^{flox/flox}$ (ref. 62), *Kras*$^{FSF\text{-}G12D/+}$ (ref. 21), *Trp53*$^{frt/frt}$ (ref. 26), *Rosa26*$^{mTmG/+}$ (ref. 36); and *Hipp11*$^{FSF\text{-}GGCB}$, *Hipp11*$^{FSF\text{-}BG}$, *Slc4a11*$^{FSF\text{-}MCD}$ and *Hopx*$^{FSF\text{-}MACD}$ reporters were generated in this study as described in detail below. All mice bearing autochthonous *KP* lung tumours were maintained in a C57BL/6 × Sv129 mixed background. *NOD.Cg-Prkdc*^{scid}*;Il2rg*^{tm1Wjl}*/SzJ*(NSG)[63] mice (The Jackson Laboratory, 005557) were used as recipients in all allotransplant studies. All mice were monitored by the investigators and veterinary staff at the Research Animal Resource Center at Memorial Sloan Kettering Cancer Center (MSKCC) and housed under a 12 h–12 h light–dark cycle at 20–25 °C and 30–70% humidity with food and water provided ad libitum.

Autochthonous and transplantation models of lung cancer

Autochthonous LUAD tumours were induced in *Kras*$^{LSL\text{-}G12D/+}$*;Trp53*$^{flox/flox}$ or *Kras*$^{FSF\text{-}G12D/+}$*;Trp53*$^{frt/frt}$ (*KPfrt*) mice with 1×10^8–1×10^9 plaque-forming units (PFU) of AdSPC-Cre, AdSPC-FlpO (Iowa Viral Vector Core), or lentiviral FlpO at 3×10^5 or 6×10^5 transforming units, as previously described[64], in mice that were aged between 8 and 12 weeks. Immunocompromised NSG mice were used as recipients for either subcutaneous, orthotopic or intravenous transplantation of *KP* LUAD cell line allografts. For subcutaneous transplantation, cells were resuspended in S-MEM (Gibco, 11380-037) and mixed with Matrigel (Thermo Fisher Scientific, CB-40230C) at a 1:1 ratio. Then, 250,000 cells were implanted subcutaneously into both flanks of NSG mice. For orthotopic transplantation, sorted cells were resuspended in PBS (Gibco, 10010-023) and intratracheally administrated to NSG mice. For intravenous transplantation, 200,000 cells were resuspended in S-MEM and injected into NSG mice through the tail vein. All cell lines were continuously monitored for mycoplasma contamination. Approximately equal numbers of male and female mice were included in all experimental groups in all mouse experiments. Mice were treated in accordance with all relevant institutional and national guidelines and regulations, and mice were euthanized by CO_2 asphyxiation, followed by intracardiac perfusion with S-MEM to clear tissues of blood when appropriate. A complete list of mice along with age, sex and age of tumour used in experiments is available (Supplementary Table 4). All animal studies were approved by the MSKCC Institutional Animal Care and Use Committee (protocol 17-11-008). Sample sizes were determined based on our previous experience with similar models rather than statistical methods. We found this sufficient to detect biologically meaningful differences while minimizing animal use; experiments were randomized when feasible. Blinding was not possible as treatment effects on tumour volume were readily distinguishable between groups. Tumour burden limit was defined as a single tumour >2 cm in diameter, tumour volume >10% of body mass or multiple tumours with a cumulative volume >3,000 mm^3. These limits were not exceeded in any of our experiments.

Generation of donor vectors for embryonic stem cell targeting

For the generation of the Slc4a11-FSF-MCD donor vector, homology arms of around 1,200 bp in length 5′ and 3′ to the end of *Slc4a11* exon 21 (Extended Data Fig. 1b) were amplified from genomic DNA of C57BL/6 mES cells using high-fidelity PCR (NEB, M0494). A homology-directed repair template donor vector was constructed by flanking the frt-bGlobinpA-(PGK-Hygromycin-pA)i-frt-P2A-mScarlet-T2A-CreER T2-P2A-DTR-WPRE-bGHpA cassette with the 5′ and 3′ homology arms and cloned into the pUC19 plasmid backbone (Takara Bio, 638949) using Gibson assembly (NEB, E2611).

For the generation of the Hipp11-FSF-GGCB donor vector, homology arms of around 5,000 bp in length 5′ and 3′ to the safe harbour of *Hipp11* intergenic region (positioned between the *Eif4enif1* and *Drg1* genes; Extended Data Fig. 1e) were amplified from genomic DNA of C57BL/6 mES cells using high-fidelity PCR. A homology-directed repair template donor vector was constructed by flanking the CAG-loxP-frt-Neomycin-PGKpA-SV40pA-frt-G-Luc-P2A-meGFP-bGlobinpA-loxP-C-Luc-E2A-TagBFP-3xFlag-WPRE-bGHpA (GGCB) cassette with the 5′ and 3′ homology arms and cloned into the pUC19 plasmid backbone using Gibson assembly.

For the generation of Hopx-FSF-MACD donor vector, homology arms of around 1,500 bp in length 5′ and 3′ to the end of *Hopx* exon 3 (Extended Data Fig. 6a) were amplified from genomic DNA of C57BL/6 mES cells using high-fidelity PCR. A homology-directed repair template donor vector was constructed by flanking the frt-bGlobinpA-(PGK-Hygromycin-pA) i-frt-P2A-mScarlet-AkaLuc-T2A-CreERT2-P2A-DTR-WPRE-bGHpA cassette with the 5′ and 3′ homology arms and cloned into the pUC19 plasmid backbone using Gibson assembly.

Validation of the *Hipp11*^{GGCB} reporter

To validate the functionality of the GGCB cassette, we performed ex vivo transformation of AT2 cells isolated from a *KPfrt;Hipp11*$^{FSF\text{-}GGCB/+}$ chimeric mouse using lentiviral vectors encoding either codon-optimized Flp recombinase (*flpO*) alone or *flpO* linked to *creERT2*. In these experiments, FlpO activates oncogenic KRAS(G12D), deletes *Trp53*, and initiates expression of the GG cassette, whereas subsequent activation of CreER^{T2} with 4-hydroxytamoxifen (4-OHT) results in a switch from GG to CB (Extended Data Fig. 1h–k). G-Luc activity was increased 13 days after transformation in all conditions (Extended Data Fig. 1i (top)), whereas C-Luc activity was observed only after 4-OHT stimulation in organoids transduced with the vector encoding both *flpO* and *creERT2* (Extended Data Fig. 1i (bottom)). Moreover, we performed flow cytometry and fluorescence imaging analyses on organoids under these four conditions. We found that eGFP was expressed at the baseline following *flpO* and a switch to TagBFP was observed only after 4-OHT exposure in the organoids transduced with both *flpO* and *creERT2* (Extended Data Fig. 1j,k). Similar results were obtained in subcutaneous transplants (Extended Data Fig. 2a–c) and autochthonous lung tumours (Extended Data Fig. 2d–g) in vivo, both by detection of G-Luc and C-Luc from repeated blood samples and by fluorescence imaging of tumours at the end point.

Embryonic stem cell targeting, genotyping and chimera generation

A Kras$^{FSF\text{-}G12D/+}$*;Trp53*$^{frt/frt}$ (*KPfrt*) mES cell line in the C57BL/6J background was generated by crossing a hormone-primed C57BL/6J *Trp53*$^{frt/frt}$ female mouse with a *Kras*$^{FSF\text{-}G12D/+}$*;Trp53*$^{frt/frt}$ male mouse. At 3.5 days after coitum, blastocysts were flushed out from the pregnant uterus, isolated and cultured on a mouse embryonic fibroblast (MEF) feeder layer. Individual ES cell lines were genotyped by PCR detection of *Kras*$^{FSF\text{-}G12D/+}$, *Trp53*$^{frt/frt}$ and *Zfy* (Y-chromosome-specific).

For the generation of *Slc4a11*$^{FSF\text{-}MCD/+}$, *Hopx*$^{FSF\text{-}MACD/+}$ and *Hipp11*$^{FSF\text{-}GGCB/+}$ knock-in mES cells, donor vectors (Slc4a11-FSF-MCD, Hopx-FSF-MACD or Hipp11-FSF-GGCB, respectively) and ribonucleoprotein (RNP) complex containing HiFi Cas9 nuclease (IDT, 1081061) and crRNA–tracrRNA duplex (IDT) were co-transfected into the *KPfrt* mES cell line by electroporation (Lonza, 4D Nucleofector). A list of the sequences of crRNAs is provided in Supplementary Table 4.

KPfrt mES cells were thawed 2 days before targeting and the media were changed 1 day and 2 h before electroporation. Before electroporation, sequence-specific crRNA and universal tracrRNA were resuspended in IDTE buffer (IDT) at a concentration of 200 µM and the crRNA–tracrRNA duplex was then formed (final concentration, 44 µM) by combining an equimolar concentration of crRNA and tracrRNA and annealing at 95 °C for 5 min (followed by cooling down to room temperature at ramp rate of 0.1 °C s^{-1}). RNP complexes were formed by combining 22 pmol of crRNA–tracrRNA duplex and 22 pmol HiFi Cas9 nuclease and incubating at room temperature for 20 min. For each electroporation, 500,000 mES cells depleted of MEFs, 1 µl donor vector (3 µg µl^{-1}), 1 μl RNP complex, 2 μl electroporation enhancer (10 μM, IDT), 16.4 μl Nucleofector P3 primary cell solution and 3.6 μl Nucleofector Supplement 1 were combined and loaded into electroporation cuvette. The ES cells were then plated onto the top of feeder MEFs and, 48 h later, the ES cells were selected with either hygromycin (Slc4a11-FSF-MCD and Hopx-FSF-MACD, 150 μg ml^{-1}) or G418 (Hipp11-FSF-GGCB, 400 μg ml^{-1}) for 1 week. Resistant clones were manually picked, expanded and validated by genotyping using the primers listed in Supplementary Table 4.

Generation of genetically engineered reporter mouse strains

Chimeric F$_0$ mice were obtained by injecting genotype-verified mES cells into host embryos at the eight-cell stage and genotyped at 2 weeks of age. F$_0$ mice were crossed into the *Kras*$^{FSF\text{-}G12D/+}$*;Trp53*$^{frt/frt}$ background to generate mice appropriate for the given experiments.

Generation of LUAD reporter and lineage-tracing cell lines

For the generation of the *Slc4a11*$^{MACD/+}$*;KP* LUAD reporter cell line, a *KP* LUAD cell line derived from a mouse bearing autochthonous *Kras*$^{LSL\text{-}G12D/+}$*;Trp53*$^{flox/flox}$ tumours at 24 weeks PTI was generated. The *KP* LUAD cells were then co-transfected with the Slc4a11-FSF-MACD donor vector together with the U6-sgSlc4a11-EFS-Cas9 vector expressing guide RNA (ACATATGGGGAGGTATGAGC) targeting the last exon of *Slc4a11* at a 1:1 ratio using Lipofectamine 3000 (Thermo Fisher Scientific, L3000015). The transfected cells were selected with hygromycin (Sigma-Aldrich, 400053) at a concentration of 150 μg ml^{-1} for 2 weeks and single-cell-derived drug-resistant clones were manually picked for expansion and genotyping with the following primers (5′ KI_F1 and 5′ KI_R1; 3′ KI_F1 and 3′ KI_R1). The single-cell-derived clones were transduced with AdCMVFlpO (Iowa Viral Vector Core) at a multiplicity of infection (MOI) of 500 to remove the frt-bGlobinpA-(PGK-Hygromycin-pA)i-frt STOP cassette. Excision of the STOP cassette was confirmed by genotyping spanning the left homology arm and mScarlet (recombined; 5′ KI_F1 and 5′ KI_R2) and by flow cytometry analysis detecting mScarlet fluorescence (Extended Data Fig. 9c,d). An additional *Slc4a11-MACD;KP* LUAD reporter cell line was generated through ex vivo transformation of an AT2 organoid culture intermediate, as described previously[9] (see below). The lentiviral lineage tracing vectors (Lenti-EFS-Flex-TagBFP-PGK-eGFP or PGK-Gluc-miRFP670-EFS-lox-BFP-lox)[9] (Extended Data Fig. 9g–i) were transduced into the *Slc4a11-MACD;KP* LUAD reporter cell lines and sorted using fluorescence-activated cell sorting (FACS) based on eGFP or miRFP670 fluorescence. *KPfrt;Hopx*$^{MACD/+}$ LUAD reporter cell lines were generated as previously described[9] or derived from autochthonous *Kras*$^{FSF\text{-}G12D/+}$*;Trp53*$^{frt/frt}$*;Hopx*$^{MACD/+}$*;Hipp11*$^{BG/+}$ tumours at 20 weeks PTI, in which the *Hipp11*$^{BG/+}$ allele enables switching from baseline (TagBFP^{+}) to lineage-traced (GFP^{+}) fluorescence. All cell lines tested negative for mycoplasma contamination.

Dissociation of LUADs and lung tissue

For isolation of normal AT2 cells and autochthonous LUAD cells, mice were euthanized at the indicated timepoints after tumour induction and were perfused with sterile S-MEM (Gibco, 11380-037) through the right ventricle of the heart. Dissected lungs or microdissected tumours were dissociated with a mixture of dispase II (Corning, 354235, 0.6 U ml^{-1}), collagenase type IV (Thermo Fisher Scientific, 17104019, 167 U ml^{-1}) and DNase I (StemCell Technologies, 07469, 10 U ml^{-1}) in S-MEM solution at 37 °C as previously described[4] for 1 h. The dissociated cells were filtered using a 100 μm filter and centrifuged at 1,500 rpm for 10 min at 4 °C. The supernatant was removed by aspiration and red blood cell lysis was performed using BD Pharm Lyse (BD Biosciences, 555899) for 1 min on ice. Cells were then washed with sterile medium containing 2% heat-inactivated FBS (Hyclone, SH30910.03), passed through a 40 μm filter and pelleted at 300*g* for 5 min at 4 °C. The supernatant was removed, and live cells were purified using the Akadeum Dead Cell Removal Microbubble kit according to the manufacturer's instructions (Akadeum Life Sciences, 11510-211). Cells were resuspended in FACS buffer (2% heat-inactivated FBS in PBS) and counted for use in FACS, as described below.

Flow cytometry analysis and FACS

Cells were prepared as described above, and Fc block (BD Biosciences, 553142) was added on ice for 10 min before staining with the appropriate antibody panel (Supplementary Table 4). After 20 min of staining on ice, cells were washed twice with FACS buffer and pelleted by a 5 min spin (300*g* at 4 °C). The cell pellets were resuspended in PBS with 2% heat-inactivated FBS containing DAPI (Sigma-Aldrich, D9542, 1 μg ml^{-1}) or Helix NP NIR (BioLegend, 425301, 5 nM) to identify dead cells. Cell sorting was performed at the Flow Cytometry Core Facility at Sloan Kettering Institute/MSKCC, using a BD FACS Aria Sorter. Cells were sorted using the 4-way purity mode. Cancer cells were sorted as (CD45/CD31/CD11b/CD11c/F4/80/TER-119)$^{-}$/Helix NP NIR^{-} or DAPI^{-} (live); the specific fluorescence-positive cell populations are indicated in each experiment.

Alveolar organoid culture and ex vivo transformation protocol

FACS-purified AT2 cells (gated as MHCII^{+}EPCAM^{+}SCA1^{-}podoplanin^{-}lineage^{-}(CD45/CD31/CD11b/CD11c/F4/80/Ter-119)$^{-}$DAPI^{-}) from *KPfrt;Hipp11*$^{GGCB/+}$ chimeras were transduced by lentivirus (Lenti-PGK-FlpO or Lenti-PGK-FlpO-P2A-creERT2) at a MOI of 10 by spinfection (600*g*, 37 °C, 30 min). Then, 4,000 transduced AT2 cells were resuspended with 50,000 primary pulmonary endothelial cells isolated from 4-week-old *Rosa26*$^{mTmG/+}$ mice by FACS (CD31^{+}CD45^{-}DAPI^{-}) in 50 μl alveolar organoid culture medium (Ham's F-12 (Thermo Fisher Scientific, 11765047), 10% FBS (Hyclone, SH30910.03), 1% GlutaMAX (Thermo Fisher Scientific, 35050061), 1% penicillin–streptomycin (Thermo Fisher Scientific, 15070063), 1% ITS (Millipore Sigma, I3126) and 1% HEPES (Thermo Fisher Scientific, 15630080)). The resuspension was then mixed with 50 μl Matrigel (Thermo Fisher Scientific, CB-40230C) and placed into cell culture inserts (Thermo Fisher Scientific, 08-770). Alveolar organoid culture medium (500 μl) was added to the reservoir (Thermo Fisher Scientific, 353504) outside the insert and replaced every 3 days. Primary organoids were digested at day 7 for secondary organoid culture with 5 U ml^{-1} dispase (Corning, 354235) for 1 h at 37 °C and replating without endothelial cells to select for transformed tumour spheres. 4-OHT (Sigma-Aldrich, H6278) was added at a concentration of 1 μM at day 10. The supernatants were collected every 3 days for G-Luc and C-Luc measurement starting at day 1. Organoids were imaged using the EVOS M5000 microscope and dissociated for flow cytometry analysis at day 16 (6 days after 4-OHT).

IF and immunohistochemistry

Mice were euthanized by CO_2 asphyxiation followed by systemic perfusion with S-MEM (Gibco, 11380-037) or PBS (Gibco, 10010-023) to clear lungs of blood. Tissues were fixed in 10% neutral-buffered formalin (Sigma-Aldrich, HT501128) for 24–48 h at 4 °C and either embedded in paraffin or dehydrated using 30% sucrose for 16–24 h before embedding in OCT compound (Thermo Fisher Scientific, 23-730-571) at −80 °C.

IF imaging was performed on 5 μm formalin-fixed paraffin-embedded (FFPE) sections or 7 μm cryosections. FFPE sections were deparaffinized and heat-induced antigen retrieval was performed using EDTA antigen retrieval buffer (Sigma-Aldrich, E1161). For cryosections, the slides were air-dried for 1 h at room temperature and fixed in acetone at −20 °C for 10 min. The sections were blocked in donkey immunomix (0.2% BSA (Sigma-Aldrich, 810533), 5% donkey serum (Thermo Fisher Scientific, 31874), 0.3% Triton X-100 (Thermo Fisher Scientific, BP151-100) in PBS (Gibco, 10010-023)) at room temperature for 30 min. Incubation of primary antibodies against GFP (Abcam, ab13970), integrin α2 (Abcam, ab181548), Ki-67 (Thermo Fisher Scientific, 14-5698-82), pan-cytokeratin (Agilent Dako, M351529-2), uPAR (R&D Systems, AF807), Flag (Sigma-Aldrich, F1804), NKX2.1 (Abcam, ab76013), HMGA2 (Cell Signaling, 8179), SPC (Sigma-Aldrich, AB3786), HOPX (Santa Cruz

Biotechnology, sc-398703), Cre recombinase (Cell Signaling, 15036) and HNF4α (Cell Signaling, 3113) diluted in donkey immunomix was performed at 4 °C overnight. AlexaFluor-conjugated secondary antibodies raised in donkey were used for signal detection (Invitrogen, A31573, A78948, A10037, A78947). Sections were counterstained with 1 μg ml^{-1} DAPI (Sigma-Aldrich, D9542) for 10 min and mounted with coverslips using Mowiol mounting reagent (EMD Millipore, 475904). Mounted slides were imaged using the Zeiss Axio Imager Z2 and ZEN 2.3 software or digitally scanned using Mirax Midi-Scanner (Carl Zeiss AG). Image analysis was performed using Fiji software.

Haematoxylin and eosin staining was performed using a standard protocol and tumour grades were assigned using an AI-based Aiforia software (NSCLC_v25 algorithm, Aiforia Technologies), as described previously[65]. For immunohistochemistry, tissue sections were incubated at 58 °C for 1 h and loaded onto the Leica Bond RX and dewaxed at 72 °C, followed by antigen retrieval using EDTA-based ER2 solution (Leica, AR9640) at 100 °C for 20 min. Primary antibodies against Cre recombinase (Cell Signaling Technology, 15036), uPAR (R&D Systems, AF543), NKX2-1 (Abcam, ab76013) or HMGA2 (Cell Signaling Technology, 8179) were incubated at room temperature for 1 h, followed by an 8-min incubation with the Leica Bond Polymer anti-rabbit or anti-goat HRP reagent (Polymer Refine Detection Kit, Leica, DS9800). The mixed DAB reagent (Polymer Refine Detection Kit) was applied for 10 min, followed by haematoxylin counterstaining (Refine Detection Kit) for 10 min. After staining, the slides were rinsed in water, dehydrated through a graded ethanol series (70%, 90%, 100%), cleared three times in HistoClear II (National Diagnostics, HS-202) and mounted with Permount (Thermo Fisher Scientific, SP15). Image analysis was performed using Fiji software.

Catalogue numbers and dilutions of all antibodies are provided in Supplementary Table 4.

In vivo EdU/BrdU dual labelling and imaging

After sequential in vivo incorporation of EdU (20 mg per kg, intraperitoneal (i.p.), 16 h before euthanasia) and BrdU (100 mg per kg, i.p., 4 h before euthanasia), lung tissues were collected, fixed in formalin (Sigma-Aldrich, HT501128), embedded in paraffin, sectioned and mounted onto slides using standard FFPE procedures. Tissue sections were deparaffinized and rehydrated, followed by EdU staining using the Click-iT Plus EdU Cell Proliferation Kit (Invitrogen, C10637). Next, automated multiplex IF was conducted with the Leica Bond BX staining system. The sections were treated with EDTA-based epitope retrieval ER2 solution (Leica, AR9640) for 20 min at 100 °C. Primary antibodies against BrdU (Roche, 1170376), GFP (Abcam, ab13970) and Cre (BioLegend, 908001) were used. Secondary antibodies were incubated followed by nuclear counterstaining with DAPI (Sigma-Aldrich, 5 μg ml^{-1}). The slides were mounted using Mowiol mounting reagent (Calbiochem) before imaging. Image analysis was performed using Fiji software.

In situ hybridization

mRNA in situ hybridization was performed on FFPE tissues using the manual Advanced Cell Diagnostics RNAscope 2.5 HD Reagent Kit (322350) or the RNAscope Multiplex Fluorescent Reagent Kit v2 (323100) according to the manufacturer's instructions. Antigen retrieval times and protease digestion times were 15 and 20 min for mouse LUAD tissues, respectively. Probes for *Slc4a11* and *Sftpc*, as well as Opal dyes from Akoya Biosciences and their dilutions for use with multiplex fluorescence in situ hybridization are listed in Supplementary Table 4.

AkaLuc in vivo bioluminescence imaging

NSG mice bearing subcutaneous transplants of *KP;Slc4a11-MACD* or *KPfrt;Hopx-MACD* reporter cells and mice bearing autochthonous *KPfrt;;Hipp11*$^{GGCB/+}$*;Hopx*$^{MACD/+}$ LUAD tumours were i.p. injected with 100 μl of 30 mM AkaLumine-HCl substrate (Sigma-Aldrich, 808350) resuspended in PBS and imaged on the IVIS Lumina II system (PerkinElmer).

Plasma sampling and G-Luc/C-Luc measurements

Whole venous blood was collected by puncturing the submandibular vein, followed by the collection of 100 μl of blood into capillary blood collector vials (Thermo Fisher Scientific, 02-675-185). Plasma was separated by centrifugation at >8,000*g* for 10 min at 4 °C. For G-Luc measurement, plasma or cell culture supernatant was diluted 1:10 in PBS and 200 μM *Gaussia* luciferase substrate coelenterazine-h (NanoLight, 3011) was added. For C-Luc measurement, plasma or cell culture supernatant was diluted 1:100 in PBS and 0.617 μM *Cypridina* luciferase substrate vargulin (NanoLight, 305) was added. Luminescence was immediately measured on the BioTek Cytation 1 (Agilent) system at room temperature. This approach enables measurement of HPCS or AT1-like state growth potential (C-Luc) compared with the bulk of the tumour (G-Luc) by longitudinal measurements in the blood.

Administration of DT, MRTX1133 and cisplatin

DT (Sigma-Aldrich, D0564) was first dissolved in sterile water and diluted in sterile saline for i.p. treatment (at 50 μg per kg daily or 25 μg per kg every other day for long-term treatment and combination treatment studies). Cisplatin (West-Ward, NDC, 0143-9504-01) was dosed at 1.5 mg per kg i.p. every 3 days. MRTX1133 (a gift from Mirati Pharmaceuticals) in captisol (HY-17031, MedChem Express) at 30 mg per kg, twice per day was given i.p. as previously described[9].

Lineage tracing of *Slc4a11*$^{+}$, *Hopx*$^{+}$ and random *Rosa26*$^{creERT2/+}$ cell state cells

Mice bearing autochthonous lung tumours were administered one or two doses of tamoxifen (200 mg per kg by oral gavage) at the indicated timepoints. For scRNA-seq analysis of *Hopx*$^{+}$ cells and for randomly labelling *Rosa26*$^{CreERT2/+}$ cancer cells, one dose of tamoxifen (20 mg per kg by oral gavage) was provided. The baseline measurement at 3 days was chosen to account for the conversion of tamoxifen to its active metabolite 4-OHT, recombination of the lineage-traced cells and elimination of residual 4-OHT. Tamoxifen was dissolved in corn oil at 20 mg ml^{-1} or 2 mg ml^{-1} at 60 °C for 1 h, as described previously[9].

Generation of lentivirus

HEK293FreeStyle (HEK293FS) cells were transfected with lentiviral transfer plasmids and the second-generation lentiviral packaging plasmid psPAX2 (Addgene, 12260) and the envelope plasmid pMD2.G (Addgene, 12259) using either the TransIT-LT1 kit (Mirus Bio, MIR 6000) or Lipofectamine 2000 Transfection Reagent (Thermo Fisher Scientific, 11668500). At 24 h after transfection, the medium was discarded and replaced with fresh complete medium. Viral medium was collected and filtered through 0.45-μm PES filters (Cytiva, 6780-2504) at 48 h and 72 h after transfection. All viral media collected were concentrated using an ultracentrifuge with rotor speed set at >130,000*g* for 2 h at 4 °C. The supernatant was discarded into bleach and viral pellets were allowed to solubilize overnight at 4 °C. Concentrated virus was gently mixed and aliquoted. The aliquots were immediately placed on dry ice and stored at −80 °C. A fibroblast reporter GreenGo cell line expressing GFP after Flp-mediated recombination was used to titre lentivirus, as described previously[66].

Generation of uPAR CAR T cells

Both mouse SFG γ-retroviral m.uPAR-m28z and human SFG γ-retroviral m.uPAR-h28z plasmids were previously described[44]. In the human m.uPAR-h28z CAR, the anti-mouse uPAR scFV is preceded by a human CD8a signal peptide and followed by a CD28 hinge-transmembrane-intracellular domain, a CD3z intracellular signalling domain and is linked to a P2A sequence to simultaneously express truncated LNGFR. In the mouse m.uPAR-m28z CAR, the anti-mouse uPAR scFV is preceded by a mouse CD8a signal peptide and followed by the MYC-tag sequence, mouse CD28 transmembrane and mouse CD3z intracellular domain[67]. Plasmids encoding the SFG γ-retroviral vectors were used to transfect gpg29 fibroblasts (H29) to generate VSV-G pseudotyped retroviral supernatants, which were used to construct stable retrovirus-producing cell lines as previously described[67,68]. To isolate human T cells from peripheral blood, buffy coats from anonymous healthy donors were purchased from the New York Blood Center. Peripheral blood mononuclear cells were isolated by Ficoll-based density-gradient centrifugation. T cells were purified using the human Pan T cell isolation kit (Miltenyi Biotec, 130-096-535), stimulated with CD3/CD28 T cell activator Dynabeads (Invitrogen, 11131D) as described previously[69], and cultured in X-VIVO 15 (Lonza, BEBP04-744Q) supplemented with 5% human serum (Gemini Bio-Products, 100-110-100), 5 ng ml^{-1} interleukin-7 and 5 ng ml^{-1} IL-15 (PeproTech, 200-07-10UG and 200-15-10UG, respectively). T cells were counted using an automated cell counter Vi-CELL BLU (Beckman). Then, 48 h after initiating T cell activation, T cells were transduced with retroviral supernatants by centrifugation on RetroNectin-coated plates (Takara, T110B). Transduction efficiencies were determined 4 days later using flow cytometry and CAR T cells were adoptively transferred into mice or used for in vitro experiments. All blood samples were handled according to the required ethical and safety procedures. To isolate mouse T cells from peripheral blood, SV-129 and C57BL/6 mixed background mice were euthanized and the spleens collected. After tissue dissection and red blood cell lysis, primary mouse T cells were purified using the mouse Pan T cell Isolation Kit (Miltenyi Biotec, 130-095-130). Purified T cells were cultured in RPMI-1640 (Thermo Fisher Scientific, 11-875-119) supplemented with 10% FBS (GeminiBio, 900-108), 10 mM HEPES (Thermo Fisher Scientific, 15-630-080), 2 mM L-glutamine (Thermo Fisher Scientific, A2916801), MEM non-essential amino acids 1× (Thermo Fisher Scientific, 11140050), 55 μM β-mercaptoethanol (Thermo Fisher Scientific, A2916801), 1 mM sodium pyruvate (Thermo Fisher Scientific, 11360070), 100 IU ml^{-1} recombinant human IL-2 (Proleukin; Novartis) and mouse anti-CD3/28 Dynabeads (Gibco, 11453D) at a bead:cell ratio of 1:2. T cells were spinoculated with retroviral supernatant collected from Phoenix-ECO cells 24 h after initial T cell activation as described[67,70] and used for functional analysis 3–4 days later.

Administration of uPAR CAR T cells

For uPAR CAR T transfusion into autochthonous *KP* tumour bearing mice, i.p. cyclophosphamide (200 mg per kg, Long Grove Pharmaceutical; NDC, 81298-8114-1) was administered to control and uPAR CAR T treatment groups. Then, 16 h later, a total of 2×10^6 mouse-derived uPAR CAR T cells per mouse was administered through i.p. injection in uPAR-CAR-T-cell-treated mice. Mice were monitored daily and collected 7 days after uPAR CAR T transfusion as indicated. For human CAR T transfusion into NSG allografts, between 5×10^6 and 7.5×10^6 human-derived CAR T cells targeting mouse uPAR were administered intratumourally per mouse.

Hyperoxia lung injury

Alveolar injury was induced in 14-week-old C57BL/6 mice by a 32 h exposure to >85% O$_2$ in a hyperoxia chamber (BioSpherix), with FiO$_2$ concentration maintained at a constant flow of around 3 lO$_2$ per min and monitored by an in-line oxygen analyser. Mice were euthanized on day 7 after the 32 h exposure, followed by collection of lungs for histological analysis.

Processing of cells for droplet-based scRNA-seq

Single-cell suspensions from LUAD tumours were prepared and stained as above. The samples were multiplexed using the TotalSeq B cell hashing protocol[71] (BioLegend; Supplementary Table 4). Live sorted cells were collected by flow cytometry, washed once with PBS containing 1% BSA and resuspended to a final concentration of 700–1,300 cells per μl of PBS + 1% BSA and processed by droplet-based scRNA-seq as below.

scRNA-seq

Single-cell suspensions were stained with Trypan blue, and the Countess II Automated Cell Counter (Thermo Fisher Scientific) was used to assess both the cell number and viability. After quality control, the samples were loaded onto Next GEM Chip G (14143, 15123, 15342, 15488, 15600, 15601, 15771) or GEM-X Single Cell Chip (16235, 16318, 16562, 16686, 17402, 17483, 17543, 17721) (10x Genomics PN 1000690 and 2000060) and GEM generation, cDNA synthesis, cDNA amplification and library preparation of around 40,000–50,000 cells proceeded using the Chromium Next GEM Single Cell 3′ Kit v3.1 or GEM-X Single Cell 3′ Kit v4 (10x Genomics, 1000268 and 1000691) according to the manufacturer's protocol. cDNA amplification included 11–12 cycles, and 78–863 ng of the material was used to prepare sequencing libraries with 8–14 cycles of PCR. Indexed libraries were pooled and sequenced on the NovaSeq 6000 (14143, 16235, 16562) or X (15123, 15342, 15488, 15600, 15601, 15771, 16235, 16318, 16562, 16686, 17402, 17483, 17543, 17721) system in a PE28/88 run using the NovaSeq 6000 S4 (200 Cycles) or X 10B (100 Cycles) or 25B (300 cycles) Reagent Kit (Illumina). An average of 38,000 paired reads was generated per cell.

Cell surface protein feature barcode analysis

Amplification products generated using the methods described above included both cDNA and feature barcodes tagged with cell barcodes and unique molecular identifiers. Smaller feature barcode fragments were separated from longer amplified cDNA using a 0.6× cleanup using aMPure XP beads (Beckman Coulter, A63882). Libraries were constructed using the 3′ Feature Barcode Kit (10x Genomics, 1000276) according to the manufacturer's protocol with 10–12 cycles of PCR. Indexed libraries were pooled and sequenced on the NovaSeq 6000 (14143_B,16562_B) or X (15123_B, 15342_B, 15488_B, 15600_B, 15601_B, 15771_B, 16235_B, 16318_B, 16562_B, 16686_B, 17402_B, 17543_B) system in a PE28/88 run using the NovaSeq 6000 S4 (200 Cycles) or X 10B (100 cycles) or 25B (300 cycles) Reagent Kit (Illumina). An average of 359 million paired reads was generated per sample.

Computational analyses

Jupyter notebooks executing the analysis workflow and figure generation are available on GitHub (https://github.com/dbetel/HPCS_LUAD). All generated sequencing data and count matrices are available at the NCBI Gene Expression Omnibus under accession number GSE277777.

Processing of scRNA-seq data

FASTQ files of scRNA-seq data generated on the 10x Chromium or ChromiumX platform were processed using the standard CellRanger pipeline (version ≥6.2)[72]. Reads were aligned to a custom GRCm38/mm10 reference genomes including the *eGFP*, *TagBFP, tdTomato, Gluc, Cluc, Akaluc, mScarlet, creERT2* and *DTR* transgenes. Cell-gene count matrices were analysed using a combination of published packages and custom scripts centred around the SCANPY/AnnData ecosystem[73]. scRNA-seq datasets from different mouse models and primary patient samples were analysed separately using similar workflows.

scRNA-seq data were compiled into a combined count matrix. In general, cells with less than 300 unique molecular identifiers (UMIs), more than 10–20% mitochondrial UMIs and low complexity based on the number of detected genes versus number of UMIs were removed as indicated in the available code. Where applicable, doublets were filtered by modelling the TotalSeq B hash count distribution as a Bayesian Gaussian mixture model with variational inference[74]. The same method was used to demultiplex the sample into individual hashes. UMI counts were normalized using the default CPM normalization. In the case of non-hashed transplant samples, the R package scDblFinder[75] was used to detect doublets, which were then removed before further analysis.

To identify highly variable features, variance-stabilizing transformation and dimensionality reduction were performed on normalized, $\log_2$-transformed count data using principal component analysis. The resulting dimensionality-reduced count matrices were used as an input for uniform manifold approximation and projection embedding and unsupervised clustering using the Leiden algorithm[76].

Cell state classifications

To compare data generated from the droplet based 10x Chromium platforms with our previous work[4], we first identified the common genes between each new 10x dataset and our previously published single-cell dataset, which was generated using the SmartSeq2 method[77]. For each 10x dataset, we trained a multiclass logistic regression model using the scikit-learn LogisticRegression class with options multi_class='multinomial' and solver='lbfgs' using our original cluster labels (hereafter, cell state identities) and gene counts from our previous work[4], using only the genes in common between both datasets. We then used this model to classify the cells generated in our 10x dataset. To assign cell state identities to each Leiden cluster generated from our SCANPY pipeline above, we took a pluralistic voting approach in which the cell states that were the most represented in a Leiden cluster were used as that cluster's cell state identity. A rare exception to this were those Leiden clusters where a significant proportion of cells were identified as highly proliferative and were subsequently shown to have high *Mki67* expression. These Leiden clusters were assigned a highly proliferative cell state identity. Cell state assignments can be found in the source code available at the GitHub repository above. Notably, based on recent work identifying a hybrid lung/gastric-like cell state expressing *Nkx2-1* and *Hnf4α*[40,41], we reannotated cells classified as clusters 8 and 10 in our original work[4] to comprise the hybrid lung/gastric-like cell state.

Marker evaluation for classification of cell states

The presence of marker transcripts was determined on processed cell counts as transcript levels greater than the minimum value detected by scRNA-seq (that is, non-zero counts). Cell state classifications were determined as above, with the modification that putative transcript markers were removed as input factors from all processing that could bias cell state scores and assignments towards the HPCS. We calculated true-positive (TP), false-positive (FP), true-negative (TN) and false-negative (FN) metrics and calculated marker sensitivity, specificity, positive predictive values, and negative predictive value as below:

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}}$$

$$\text{Positive predictive value} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

$$\text{Negative predictive value} = \frac{\text{TN}}{\text{TN} + \text{FN}}$$

For similar calculations using the SmartSeq2 LUAD data[4] from which the signatures were originally derived, we reprocessed raw count data using the pipeline described above up to the Leiden clustering step. We then calculated gene scores using the built-in SCANPY score_genes function using the top 100 genes from the previously determined HPCS gene signature[4] (Supplementary Table 5), again with the marker genes removed to prevent bias. We labelled clusters enriched for the HPCS gene signature as HPCS clusters and calculated the sensitivity, specificity, positive predictive value and negative predictive values as described above for the marker gene of interest. The results for all sensitivity, specificity, positive predictive value and negative predictive values for marker genes are provided in Supplementary Tables 1–3. We estimate the sensitivity of the HPCS to be 9.26% and 65.4% based on the 10x droplet-based and SmartSeq2 scRNA-seq data, respectively, while the specificity is 99.7% and 84.3%. Although we acknowledge that we cannot completely rule out the possibility that a minor proportion of non-HPCS cells are traced in our experiments, based on the specificity of the reporter system as calculated above, this is likely to be an insignificant minority of cells.

Phenotypic volume calculations

Phenotypic volumes, quantitative measures capturing the diversity of cellular phenotypes in cell populations, were performed on highly variable genes, as previously described[37]. Distributions of phenotypic volumes were calculated by sampling 100 cells randomly with replacement from the cells of interest and calculating the phenotypic volume for 1,000 replicates per cell population. Statistical significance was determined using either *t*-tests or ANOVA to compare the distributions of phenotypic volumes.

Gene signature score calculation and correlation

Gene signatures were compiled from a variety of sources listed in Supplementary Table 5 and were used to calculate scores using the SCANPY score_genes function. Scatter plots and Pearson correlations were generated by independently calculating each cell's gene signature score and HPCS score and then comparing the distribution of scores. Lines of best fit and r^2 values were calculated using scipy.stats.linregress and statistical significance for Pearson correlations was determined using an exact distribution with the built-in scipy.stats.pearsonr function. Gene signature scores were used to compare cells with or without *Slc4a11* expression as indicated.

External data analysis

All gene signatures and analysed datasets from previously published works are listed in Supplementary Table 5.

Time-series analysis

SmartSeq2 scRNA-seq LUAD data (GEO: GSE152607)[4] were downloaded from the NCBI GEO. scRNA-seq data from wild-type AT2 cells, *KP* adenoma and *KP* adenocarcinoma tumour data from the 2, 12, 20 and 30 week post-tumour induction timepoints were used with Moscot's[78] TemporalProblem to determine intertimepoint couplings through optimal transport, and the RealTimeKernel.from_moscot() function was used to convert the intertimepoint couplings into a CellRank[79] transition matrix. This matrix was used with a generalized perron-cluster cluster analysis estimator to identify terminal macrostates. Gene expression trajectories to the terminal macrostates were plotted with the Cellrank gene_trends function using a generalized additive model and the time component was determined by Palantir[80] pseudotime. The relative estimated start and end of the HPCS was modelled in Palantir pseudotime by plotting the trend of a calculated HPCS score and comparing it to known marker genes using the above process. The calculated HPCS score was defined using the top 100 genes of the HPCS gene signature (Supplementary Table 5) with the SCANPY score_genes function.

Software versions

SCANPY (v. ≥1.9), pingouin (v.0.5.4), gseapy (v.1.1.1), numpy (v. ≥1.26), scipy (v. ≥1.12), scikit-learn (v. ≥1.13), leidenalg (v.0.10.2), matplotlib (v.3.8.4), Cellrank (v.2.0.7), Palantir (v.1.4.1), R (v.4.3.3), FIJI/ImageJ (v. >1.54) and GraphPad (v. >9.0) were used.

Statistics and reproducibility

Statistical analyses were performed using Student's *t*-tests, Welch's tests, Mann–Whitney *U*-tests, Kruskal–Wallis tests, Holm–Šídák tests, one-way ANOVA or two-way ANOVA, as appropriate. All statistical analyses performed were in a two-sided manner. Statistical significance for the figures is indicated as raw P values or with asterisks. $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$, $^{****}P < 0.0001$; NS, not significant. Representative micrographs from experiments were performed at least three times independently with similar results. For Fig. 5f, hyperoxia injury was performed at three different timepoints for one mouse each, with *Slc4a11*-related expression peaking at 7 days. All of the experiments were performed as biological replicates.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

All generated sequencing data and count matrices are available at the NCBI Gene Expression Omnibus under accession code GSE277777. A full list of gene signatures and external datasets used in the analysis of this study is available in Supplementary Table 5. Source data are provided with this paper.

Code availability

Jupyter notebooks executing the analysis workflow and figure generation are available on GitHub (https://github.com/dbetel/HPCS_LUAD). A Zenodo repository is also available[81] (https://doi.org/10.5281/zenodo.17662770).

62. Marino, S., Vooijs, M., van Der Gulden, H., Jonkers, J. & Berns, A. Induction of medulloblastomas in p53-null mutant mice by somatic inactivation of Rb in the external granular layer cells of the cerebellum. *Genes Dev.* **14**, 994–1004 (2000).
63. Ishikawa, F. et al. Development of functional human blood and immune systems in NOD/SCID/IL2 receptor gamma chain^{null} mice. *Blood* **106**, 1565–1573 (2005).
64. Tammela, T. et al. A Wnt-producing niche drives proliferative potential and progression in lung adenocarcinoma. *Nature* **545**, 355–359 (2017).
65. LaFave, L. M. et al. Epigenomic state transitions characterize tumor progression in mouse lung adenocarcinoma. *Cancer Cell* **38**, 212–228 (2020).
66. Sanchez-Rivera, F. J. et al. Rapid modelling of cooperating genetic events in cancer through somatic genome editing. *Nature* **516**, 428–431 (2014).
67. Kuhn, N. F. et al. CD40 ligand-modified chimeric antigen receptor T cells enhance antitumor function by eliciting an endogenous antitumor response. *Cancer Cell* **35**, 473–488 (2019).
68. Brentjens, R. J. et al. Eradication of systemic B-cell tumors by genetically targeted human T lymphocytes co-stimulated by CD80 and interleukin-15. *Nat. Med.* **9**, 279–286 (2003).
69. Feucht, J. et al. Calibration of CAR activation potential directs alternative T cell fates and therapeutic potency. *Nat. Med.* **25**, 82–88 (2019).
70. Davila, M. L., Kloss, C. C., Gunset, G. & Sadelain, M. CD19 CAR-targeted T cells induce long-term remission and B cell aplasia in an immunocompetent mouse model of B cell acute lymphoblastic leukemia. *PLoS ONE* **8**, e61338 (2013).
71. Stoeckius, M. et al. Cell hashing with barcoded antibodies enables multiplexing and doublet detection for single cell genomics. *Genome Biol.* **19**, 224 (2018).
72. Zheng, G. X. et al. Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* **8**, 14049 (2017).
73. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15 (2018).
74. Rub, J. et al. Deep-learning tool ScVital enables species-agnostic integration of cancer cell states. *Cancer Res.* https://doi.org/10.1158/0008-5472.CAN-24-4889 (2025).
75. Germain, P. L., Lun, A., Garcia Meixide, C., Macnair, W. & Robinson, M. D. Doublet identification in single-cell sequencing data using scDblFinder. *F1000Res* **10**, 979 (2021).
76. Traag, V. A., Waltman, L. & van Eck, N. J. From Louvain to Leiden: guaranteeing well-connected communities. *Sci. Rep.* **9**, 5233 (2019).
77. Picelli, S. et al. Full-length RNA-seq from single cells using Smart-seq2. *Nat. Protoc.* **9**, 171–181 (2014).
78. Klein, D. et al. Mapping cells through time and space with moscot. *Nature* **638**, 1065–1075 (2025).
79. Weiler, P., Lange, M., Klein, M., Pe'er, D. & Theis, F. CellRank 2: unified fate mapping in multiview single-cell data. *Nat. Methods* **21**, 1196–1205 (2024).
80. Setty, M. et al. Characterization of cell fate probabilities in single-cell data with Palantir. *Nat. Biotechnol.* **37**, 451–460 (2019).
81. Chan, J. E. et al. Code for 'Critical role for a high-plasticity cell state in lung cancer'. *Zenodo* https://doi.org/10.5281/zenodo.17662770 (2025).
82. Liu, Z. et al. Systematic comparison of 2A peptides for cloning multi-genes in a polycistronic vector. *Sci. Rep.* **7**, 2193 (2017).
83. Moon, K. R. et al. Visualizing structure and transitions in high-dimensional biological data. *Nat. Biotechnol.* **37**, 1482–1492 (2019).

Acknowledgements We thank the members of the Tammela laboratory for discussions; J. Teixeira for supporting and motivating this work; L. Bombardelli for advice and suggestions regarding the G-Luc and C-Luc secreted luciferases; K. Manova, W. Kang, M. Tipping and the members of the Molecular Cytology Core for histology support; E. Chan and E. Rosiek for help with image analysis and quantification; E. de Stanchina and the Antitumor Assessment Core Facility for support with drug administration and tumour transplant experiments; R. Gardner, M. Kweens and A. Longhini for FACS support; N. Mohibullah and the members of the Integrated Genomics Operation for next-generation sequencing support; H. Alcorn and O. Grbovic-Huezo for laboratory management; C. Sotomayer for help with testing code; M. Blum, B. Christensen, S. Ding, J. Guthrie, A. Hudson, A. Moore, E. Rivas-Hernandez and C. Sussman for help with experiments and mouse genotyping; and J. Christensen and J. Hallin for providing MRTX1133. This work was supported by NIH K08-CA267072, the NIH Loan Repayment Program, and the Linn Fund for Sarcoma Research (to J.E.C.); American Cancer Society award – ACS Fairfield County CT Research Council (PF-25-1422234-01-PFCBI, https://doi.org/10.53354/ACS.PF-25-1422234-01-PFCBI.pc.gr.230344; to C.-H.P.); a PhD fellowship from Boehringer Ingelheim Fond (to K.K.); a Damon Runyon postdoctoral fellowship (2467-22), a postdoctoral fellowship from the American Federation for Aging Research, and the St. Louis Ovarian Cancer Awareness Research Grant for Ovarian Cancer from the Foundation for Women's Cancer (to Z.Z.); New York Stem Cell Science NYSTEM training award (C32559GG), the Center for Stem Cell Biology at MSKCC and support from the Druckenmiller Center for Lung Cancer Research at MSKCC (to X.Z.); NIH R01-AG054720 (to D.B.); NIA R01-AG065396, support from the MSK Technology Development Fund (TDF) (FP00009954), the Mark Foundation Grant for Cancer Research, HHMI and the Geoffrey Beene Cancer Research Center (to S.W.L); the Sigrid Juselius Foundation, National Natural Science Foundation 82373443, Fundamental Research Funds for Central University 2662025SYPY005, Huazhong Agricultural University Pilot Project Fund, the Open Project funded by Key Laboratory of Carcinogenesis and Translational Research, Ministry of Education (2025 Open Project-4) (to Y.Y.); NIH R01-CA270116, NIH R01-CA293718, an AACR Next Generation Transformative Award, and Josie Robertson and Rita Allen Scholarships (to T.T.); and the NIH/NCI Cancer Center Support Grant P30-CA08748 (to MSKCC). We acknowledge the use of the Antitumor Assessment, Integrated Genomics Operation, Flow Cytometry, Molecular Cytology Core, and Single Cell Analytics and Innovation Lab Facilities at the Sloan Kettering Institute, funded by CCSG P30-CA08748, Cycle for Survival, and the Marie-Josée and Henry R. Kravis Center for Molecular Oncology. The schematics were created using BioRender (https://BioRender.com/0lgfrw5).

Author contributions J.E.C., C.-H.P., Y.Y. and T.T. conceived and designed the study. J.E.C. and T.T. wrote the manuscript. C.-H.P., Y.Y. and D.B. contributed to the writing of the manuscript. J.E.C., C.-H.P., Y.Y., D.B. and T.T. interpreted the data. J.E.C., C.-H.P., Y.Y., K.K., E.B., Z.Z., H.S., G.G., G.H., Z.L. and X.Z. performed experiments and data analysis. Z.Z. and S.W.L. contributed CAR T cells and associated methodology. J.E.C., J.R. and D.B. performed computational analysis. J.E.C. and D.B. supervised computational analysis and interpreted data. S.W.L. and T.T. supervised experimental work. S.W.L. and T.T. obtained funding. All of the authors approved the final manuscript.

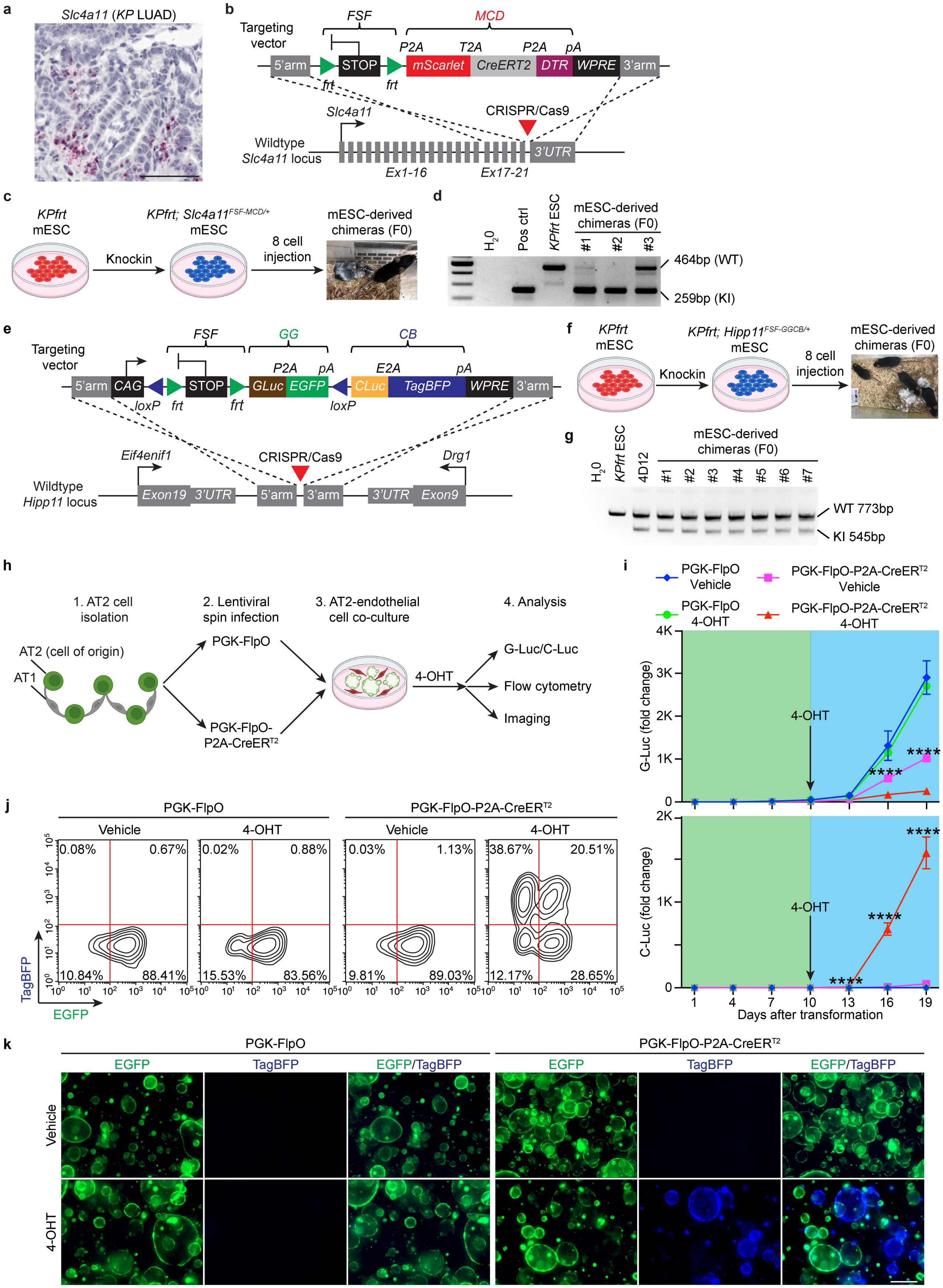
Competing interests T.T. is a scientific advisor with equity interests in Lime Therapeutics. His spouse is an employee of and has equity in Recursion Pharmaceuticals. The Tammela laboratory receives funding from Ono Pharma related to targeting the HPCS, although this funding did not directly support this work. S.W.L. is a consultant and holds equity in Blueprint Medicines, ORIC Pharmaceuticals, Mirimus, PMV Pharmaceuticals, Faeth Therapeutics and Senecea Therapeutics, and is a consultant for Fate Therapeutics. S.W.L. has equity in a joint venture developed by MSKCC and a cell therapy company to develop senolytic cell-based therapies for non-cancer indications. The company has licensed MSKCC IP, including hUPAR binders. The Mark Foundation provided the Endeavor Award to Scott Lowe for "Harnessing Senescence Biology for Immune Oncology". The other authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41586-025-09985-x.

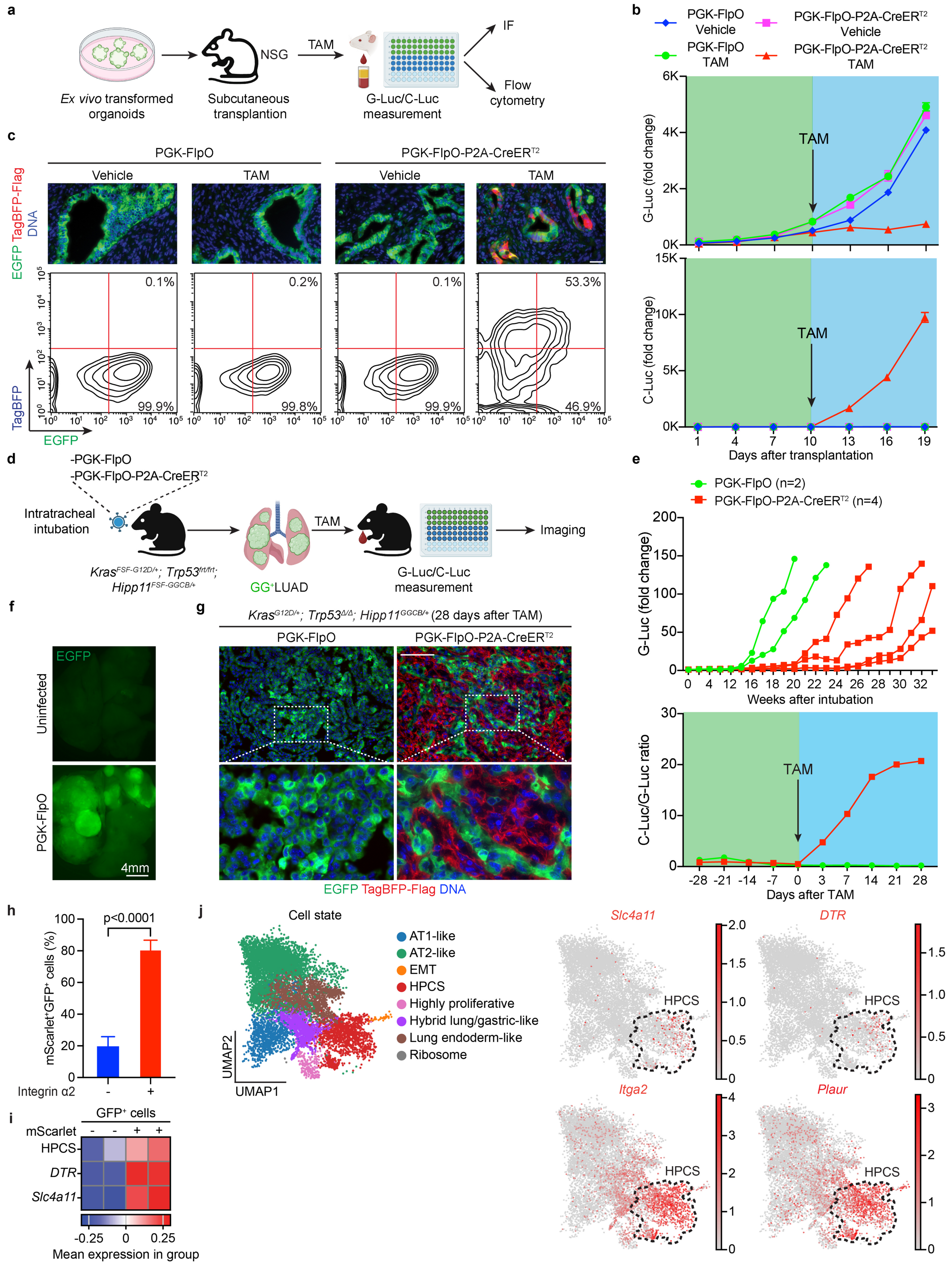
Correspondence and requests for materials should be addressed to Yan Yan or Tuomas Tammela.

Peer review information *Nature* thanks Aaron Hata and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.



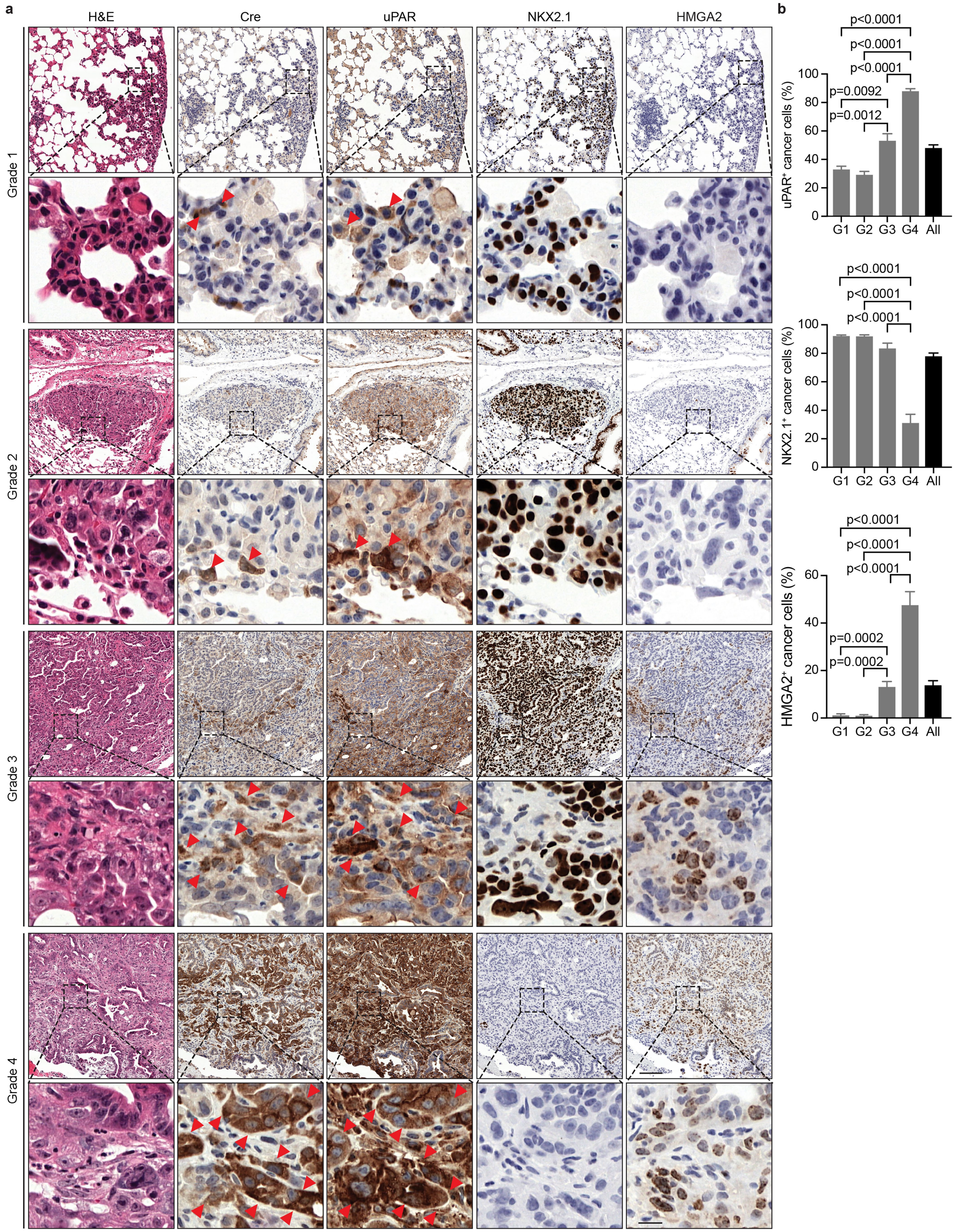
Extended Data Fig. 1 | See next page for caption.

Extended Data Fig. 1 | Construction of the *Slc4a11*$^{FSF-MCD/+}$ and *Hipp11*$^{FSF-GGCB/+}$ reporter alleles and validation of the *Hipp11*$^{FSF-GGCB/+}$ allele in vitro. a, *Slc4a11* mRNA in *KP* LUAD at 20 weeks post-tumour initiation (PTI). Scale bar: 50 µm. **b**, Components of the *Slc4a11*$^{FSF-MCD/+}$ reporter and targeting strategy by CRISPR/Cas9 mediated homology dependent repair (HDR). To restrict the expression of the *MCD* cassette to *Slc4a11*$^{+}$ cancer cells, we introduced a frt-stop-frt (*FSF*) cassette[21] upstream of the *MCD*. Upon removal of the *FSF* cassette, *MCD* is linked to the last coding exon of *Slc4a11* by a 2 A peptide sequence[82]. This approach preserves activity of the endogenous gene and links the *MCD* reporter tightly to native gene-regulatory elements, increasing the fidelity of the reporter. The strategy for generating genetically engineered *Slc4a11*$^{MCD/+}$ knock-in (KI) reporter enabling lineage tracing and ablation of *Slc4a11*$^{+}$ HPCS in *KP* LUAD was to knock in the *frt-stop-frt-P2A-mScarlet-T2A-CreERT2-P2A-DTR-WPRE* (*FSF-MCD*) reporter construct in frame into the stop codon of exon 21 of *Slc4a11* gene in the presence of CRISPR/Cas9 gene editing. **c**, Experimental outline to generate *Kras*$^{FSF-G12D/+}$*; Trp53*$^{frt/frt}$*; Slc4a11*$^{FSF-MCD/+}$ reporter mouse. The *FSF-MCD* reporter construct was knocked into a mouse embryonic stem cell (mESC) harbouring the KRAS(G12D) (*Kras*$^{frt-stop-frt(FSF)-G12D/+}$)[21] and p53 loss-of-function (*Trp53*$^{frt/frt}$)[26] alterations (***KPfrt***) and the correctly targeted mESC clones were microinjected into 8-cell stage embryos to obtain mESC-derived chimeras. **d**, Gel electrophoresis of the PCR products from a correctly targeted mESC clone (pos ctrl), parental *KPfrt* mESC, or mESC-derived chimeras using primer pairs detecting either wild-type (WT, 464 bp, *top*) or knock-in (KI, 259 bp, *bottom*) alleles, respectively. **e**, Vector map of *Hipp11*$^{FSF-GGCB/+}$ targeting vector and strategy for the generation of *Hipp11*$^{FSF-GGCB/+}$ reporter by CRISPR/Cas9 mediated dependent HDR. *CAG-LoxP-frt-stop-frt-G-Luc-P2A-EGFP-pA-LoxP-C-Luc-E2A-TagBFP-pA-WPRE* (*FSF-GGCB*) reporter construct was knocked into the *Hipp11* intergenic region safe harbour (positioned between *Eif4enif1* and *Drg1* genes) by CRISPR/Cas9 gene editing. FlpO-mediated recombination removes the *FSF* cassette and activates the *G-Luc-P2A-EGFP (GG)* element, whereas Cre-mediated recombination removes the *loxP-stop-loxP* (*LSL*) cassette and activates the *C-Luc-E2A-TagBFP* (*CB*) element. G-Luc: *Gaussia* luciferase; C-Luc: *Cypridia* luciferase. **f**, Experimental outline to generate the *Kras*$^{FSF-G12D/+}$*; Trp53*$^{frt/frt}$*; Hipp11*$^{FSF-GGCB/+}$ reporter mouse. *CAG-FSF-GGCB* reporter construct described in panel (**e**) was knocked into *KPfrt* mESCs, the correctly targeted mESC clones were microinjected into 8-cell stage embryos to obtain mESC-derived chimeras. **g**, Gel electrophoresis of the PCR products from either a parental mESC (*KPfrt* mESC), a correctly targeted mESC clone (4D12), or mESC-derived chimeras using primer pairs detecting either wild-type (WT, 773 bp, *top*) or knock-in (KI, 545 bp, *bottom*) alleles. **h**, Experimental outline to validate the *Hipp11*$^{FSF-GGCB}$ reporter in an alveolar type 2 (AT2) cell ex vivo transformation assay. AT2 cells isolated from the chimeras were transduced with lentivirus expressing either FlpO or FlpO-P2A-CreERT2 to remove the *FSF* cassette and activate oncogenic *Kras*$^{G12D/+}$ and delete *Trp53*, resulting in the generation of *Kras*$^{G12D/+}$ mutant and *Trp53* deficient LUAD organoids. The transformed organoids were exposed to 4-OHT and analysed with the indicated approaches. **i**, Longitudinal monitoring of G-Luc (*top*) and C-Luc (*bottom*) activities depicted as fold change over day 0 from ex vivo transformed LUAD organoids as in (**h**). Organoids were exposed to 4-OHT (1 µM) at day 10. Media was refreshed every 3 days before supernatant was collected for luciferase activity measurement. n = 6 for each group. Error bars are SEM. Two-way ANOVA with Dunnett's multiple comparison test. **j**, Flow cytometry analysis of EGFP vs. TagBFP expression from PGK-FlpO (*left*) or PGK-FlpO-P2A-CreERT2 (*right*) transformed organoids 6 days after treatment with either vehicle control or 4-OHT. TagBFP is only induced in organoids harbouring CreER^{T2} and exposed to 4-OHT. **k**, Fluorescent images of EGFP and TagBFP from LUAD organoids transformed with indicated lentivirus 6 days after treatment with either vehicle control or 4-OHT. Consistent with (**j**), TagBFP is only present in organoids harbouring CreER^{T2} and exposed to 4-OHT. Scale bar: 200 µm. The diagrams in **c**, **f** and **h** were created using BioRender. Tammela, T. (2025) https://BioRender.com/0lgfrw5.



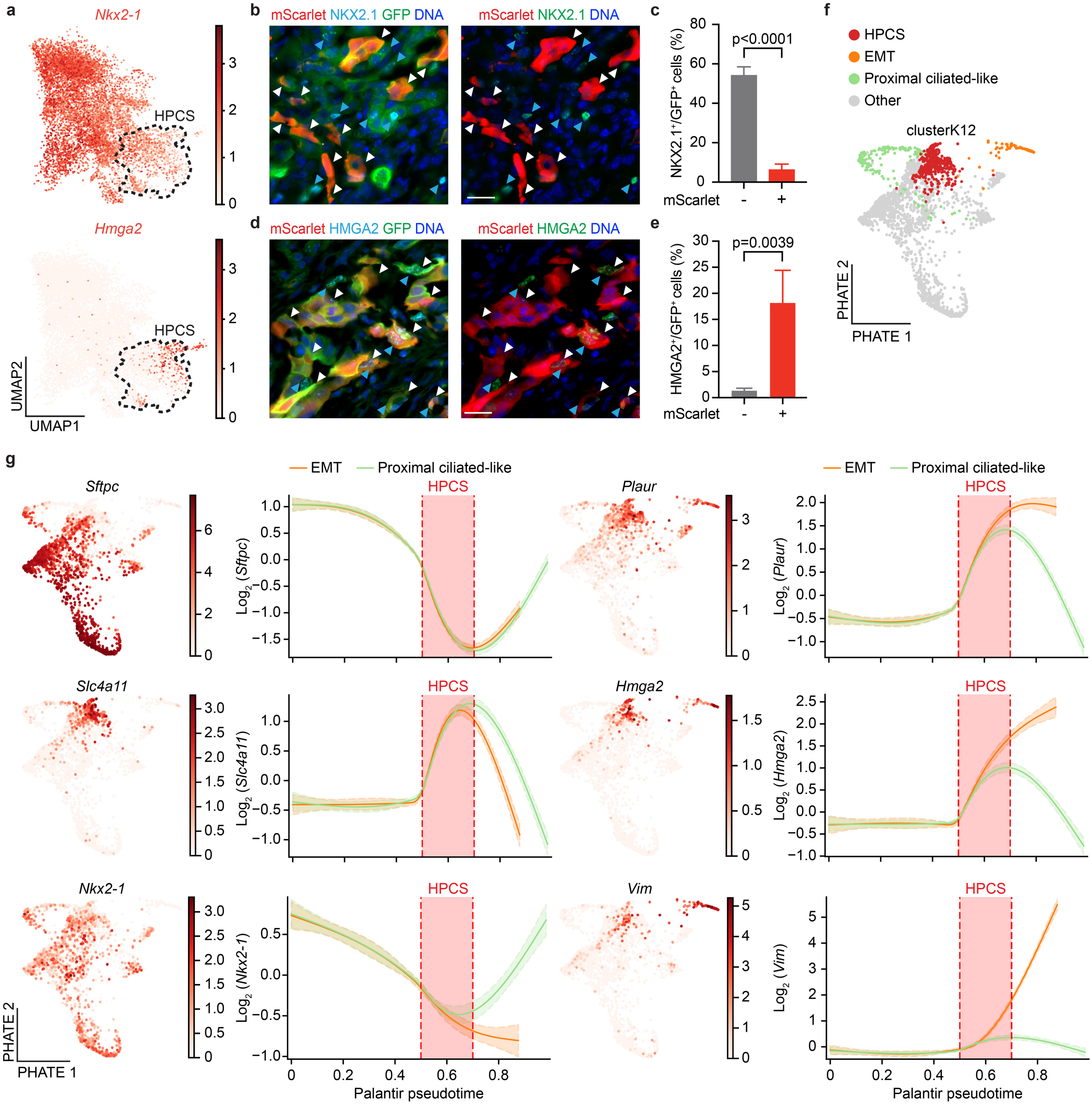
Extended Data Fig. 2 | See next page for caption.

Extended Data Fig. 2 | Validation of the *Hipp11*$^{GGCB/+}$ and *Slc4a11*$^{MCD/+}$ reporter alleles in vivo. a, Experimental outline to validate the *Hipp11*$^{FSF-GGCB/+}$ reporter in subcutaneous LUAD allografts derived from ex vivo transformed AT2 cell organoids. Tumour organoids were transplanted into the subcutaneous flank of NSG mice and administered with tamoxifen (TAM, 200 mg/kg) at day 10 post-transplantation. Mice were cheek bled every 3 days for G-Luc and C-Luc measurements. Tumours were harvested at day 19 for flow cytometry analysis and IF staining. **b**, Longitudinal monitoring of G-Luc (*top*) and C-Luc (*bottom*) activity (shown as fold changes relative to day 1) at the indicated time points from the indicated groups of tumour-bearing mice. n = 2. **c**, IF images (*top*) and flow cytometry plots (*bottom*) for the analysis of EGFP and TagBFP expression from transplanted tumours at indicated conditions. TagBFP IF was performed using anti-Flag antibody detecting the TagBFP-3xFlag fusion protein. Scale bar: 20 µm. **d**, Experimental outline to validate the *Hipp11*$^{FSF-GGCB/+}$ reporter in autochthonous LUAD tumours. *Kras*$^{FSF-G12D/+}$*; Trp53*$^{frt/frt}$*; Hipp11*$^{FSF-GGCB/+}$ mice were intubated with lentivirus expressing either PGK-FlpO or PGK-FlpO-P2A-CreER^{T2} to initiate LUAD and were cheek bled every other week to monitor tumour development. Mice from both groups were treated with a single dose of TAM (200 mg/kg) when the G-Luc activity had increased 10-fold compared to non-infected control mice. Mice were cheek bled every 3 days after TAM treatment to assess the C-Luc/G-Luc ratio. **e**, Longitudinal monitoring of G-Luc activity (*top*) or C-Luc/G-Luc ratios (*bottom*) from mice intubated with lentivirus expressing either PGK-FlpO or PGK-FlpO-P2A-CreER^{T2}. G-Luc and C-Luc activities were measured and shown as fold change over non-infected control mice. n = 2 for PGK-FlpO; n = 4 for PGK-FlpO-P2A-CreER^{T2}. **f**, Dissection microscope images depicting EGFP signal in untransduced (*top*) vs PGK-FlpO (*bottom*) transduced *KP*; *Hipp11*$^{GGCB/+}$ mice. **g**, IF staining of tumours from (**f**) stained with antibodies detecting EGFP (green) and Flag (TagBFP-Flag, red). Scale bar: 200 µm. Error bars are SEM. **h**, Percentage of mScarlet^{+}/GFP^{+} cells staining for integrin α2 measured by IF imaging. n = 19 tumours, 3 mice. Welch's t-test. Error bars are SEM. **i**, Heatmap of HPCS program, *Slc4a11*^{MCD} reporter, and endogenous *Slc4a11* expression in mScarlet^{-}/GFP^{+} and mScarlet^{+}/GFP^{+} LUAD cells. **j**, Single-cell LUAD transcriptomes in autochthonous *KPfrt; Hipp11*$^{FSF-GGCB/+}$*; Slc4a11*$^{MCD/+}$ lung tumours at 15-16 weeks PTI. Eight annotated LUAD cell states that molecularly define LUAD tumours are annotated. Distribution of *Slc4a11* (*upper left*), *DTR* (*upper right*), *Itga2* (*lower left*), and *Plaur* (*lower right*) gene expression. The diagrams in **a** and **d** were created using BioRender. Tammela, T. (2025) https://BioRender.com/0lgfrw5.

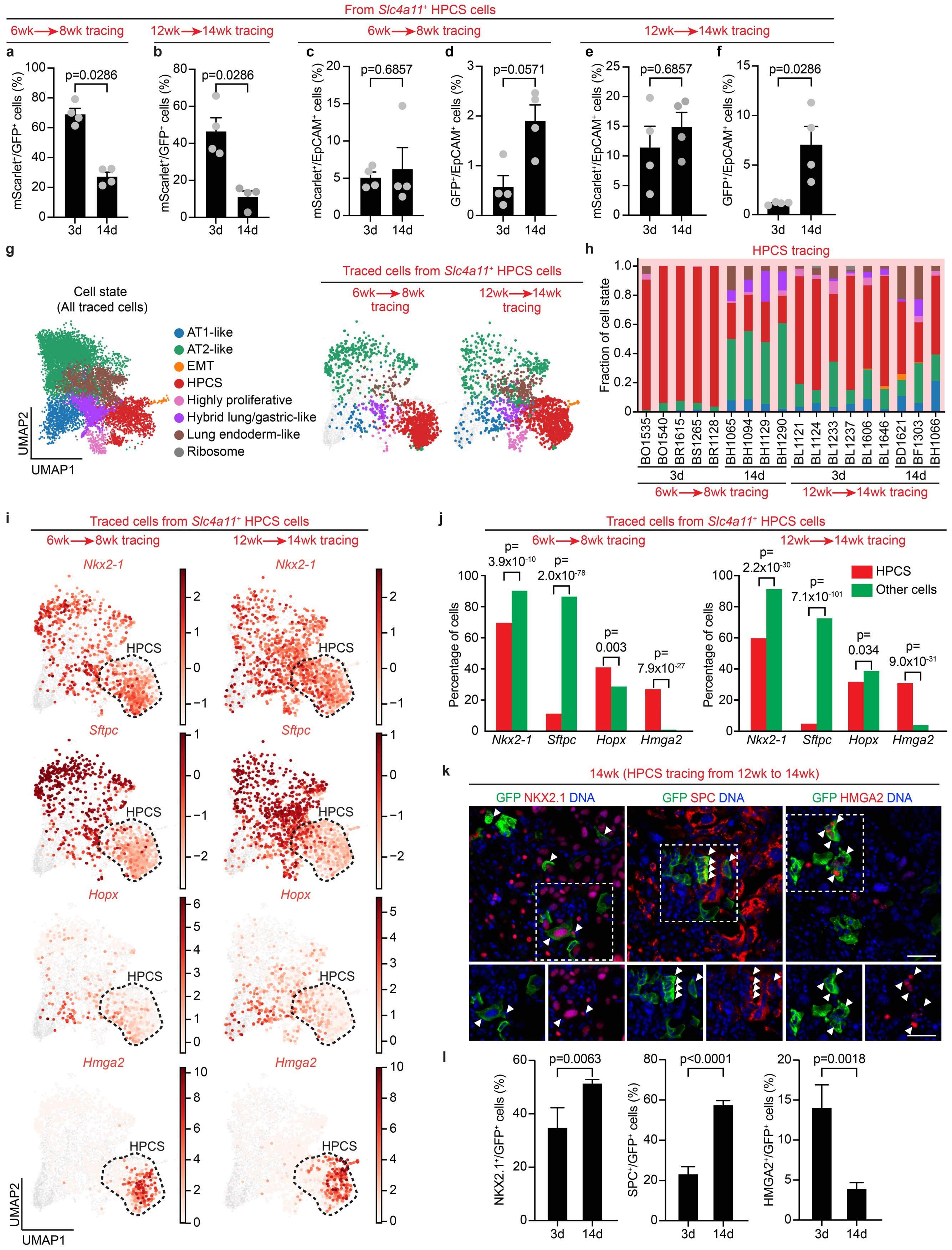


Extended Data Fig. 3 | See next page for caption.

Extended Data Fig. 3 | HPCS, uPAR, NKX2.1, and HMGA2 across tumours of distinct histopathological grades. a, Representative H&E and immunohistochemistry (IHC) for Cre (to detect the Slc4a11-MCD reporter), uPAR, NKX2.1 and HMGA2 on serial sections from autochthonous *KPfrt; Hipp11*$^{GGCB/+}$*; Slc4a11*$^{FSF\text{-}MCD/+}$ LUAD tumours at 7 and 16 weeks PTI. Histopathological grades were assigned using Aiforia artificial intelligence (AI)-based image analysis software. Scale bar: 100 µm (low magnification) and 10 µm (high magnification). **b**, uPAR^{+} (*top*), NKX2.1^{+} (*middle*), HMGA2^{+} (*bottom*) tumour cell percentages across histopathological grades. uPAR: n = 30 tumours per grade, 4 mice; NKX2.1: 28 tumours per grade, 4 mice; HMGA2: 31 tumours per grade, 4 mice. One-way ANOVA with Dunnett's T3 multiple comparisons test. Error bars are SEM.

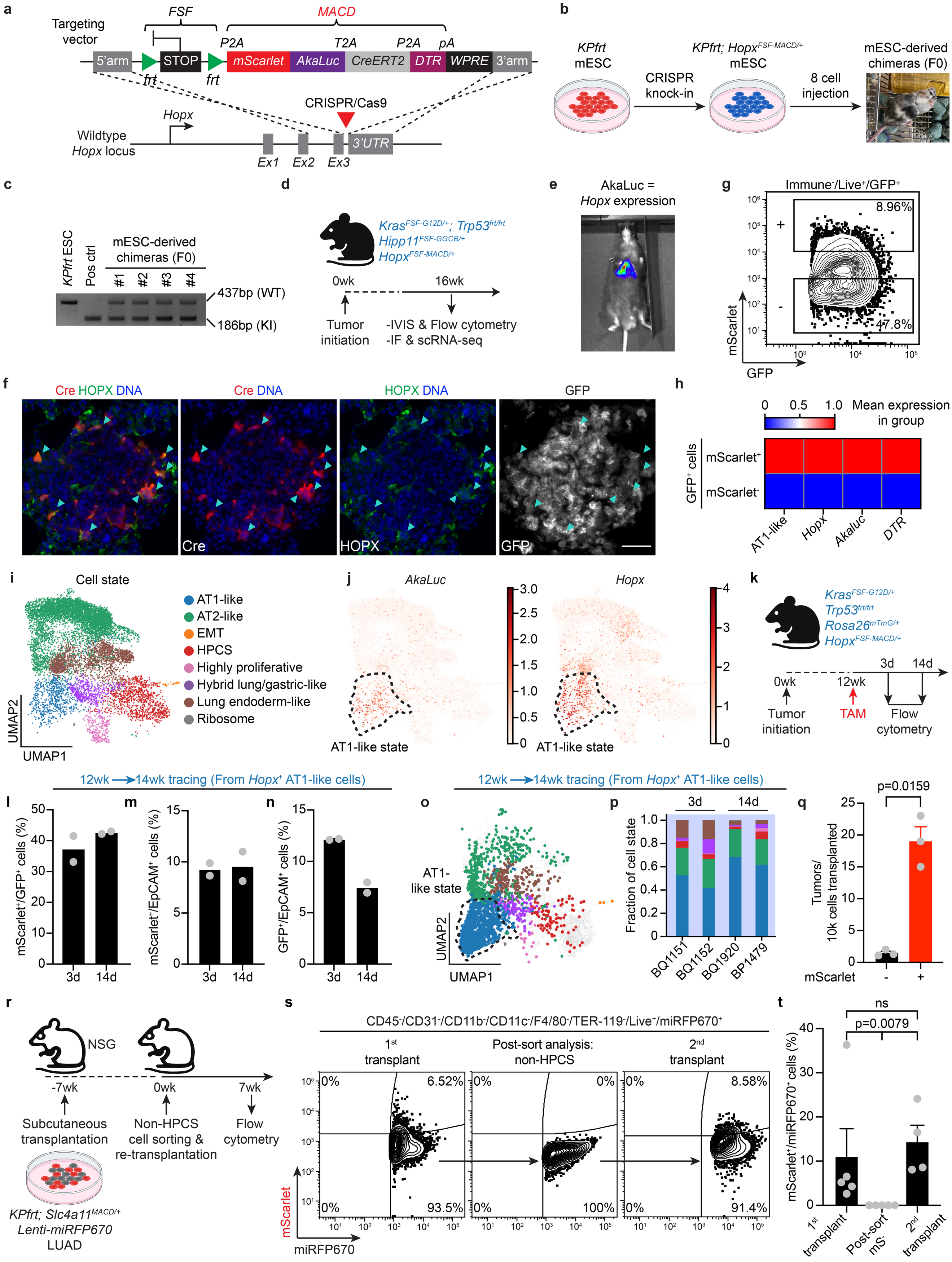


Extended Data Fig. 4 | Association of HPCS with lung cancer progression markers. **a**, *Nkx2-1* (*top*) and *Hmga2* (*bottom*) expression in LUAD cells from (Fig. 1a). **b**, IF showing mScarlet (white arrowheads) and NKX2.1 (blue arrowheads) mutual exclusivity in GFP+ LUAD cells in *KPfrt*; *Hipp11*$^{GGCB/+}$; *Slc4a11*$^{MCD/+}$ lung tumours at 16 weeks PTI. Scale bar: 20 µm. **c**, Percentage of NKX2.1+ cells in mScarlet−/GFP+ vs. mScarlet+/GFP+ LUAD cell subsets (n = 53 tumours, 5 mice; Mann-Whitney U test). **d**, IF showing mScarlet (white arrowheads) and HMGA2 (blue arrowheads) co-expression in subsets of GFP+ LUAD cells in *KPfrt*; *Hipp11*$^{GGCB/+}$; *Slc4a11*$^{MCD/+}$ lung tumours at 16 weeks PTI. Scale bar: 20 µm. **e**, Percentage of HMGA2+ cells in mScarlet−/GFP+ vs. mScarlet+/GFP+ subsets (n = 21 tumours, 5 mice; Mann-Whitney U test). **f**, Predicted terminal cell states (EMT, Proximal ciliated-like) and the HPCS projected on the LUAD progression PHATE[83] map from Marjanovic*, Hofree*, Chan* et al.[4] **g**, Expression of the indicated genes plotted as PHATE maps (*left*) and gene expression trajectories for the indicated marker genes plotted over Palantir[80] *pseudotime*, with cell state trajectories predicted by CellRank[79], and estimated time spent in HPCS cell state shaded in red (see *'Time Series Analyses'* in Methods) (*right*).



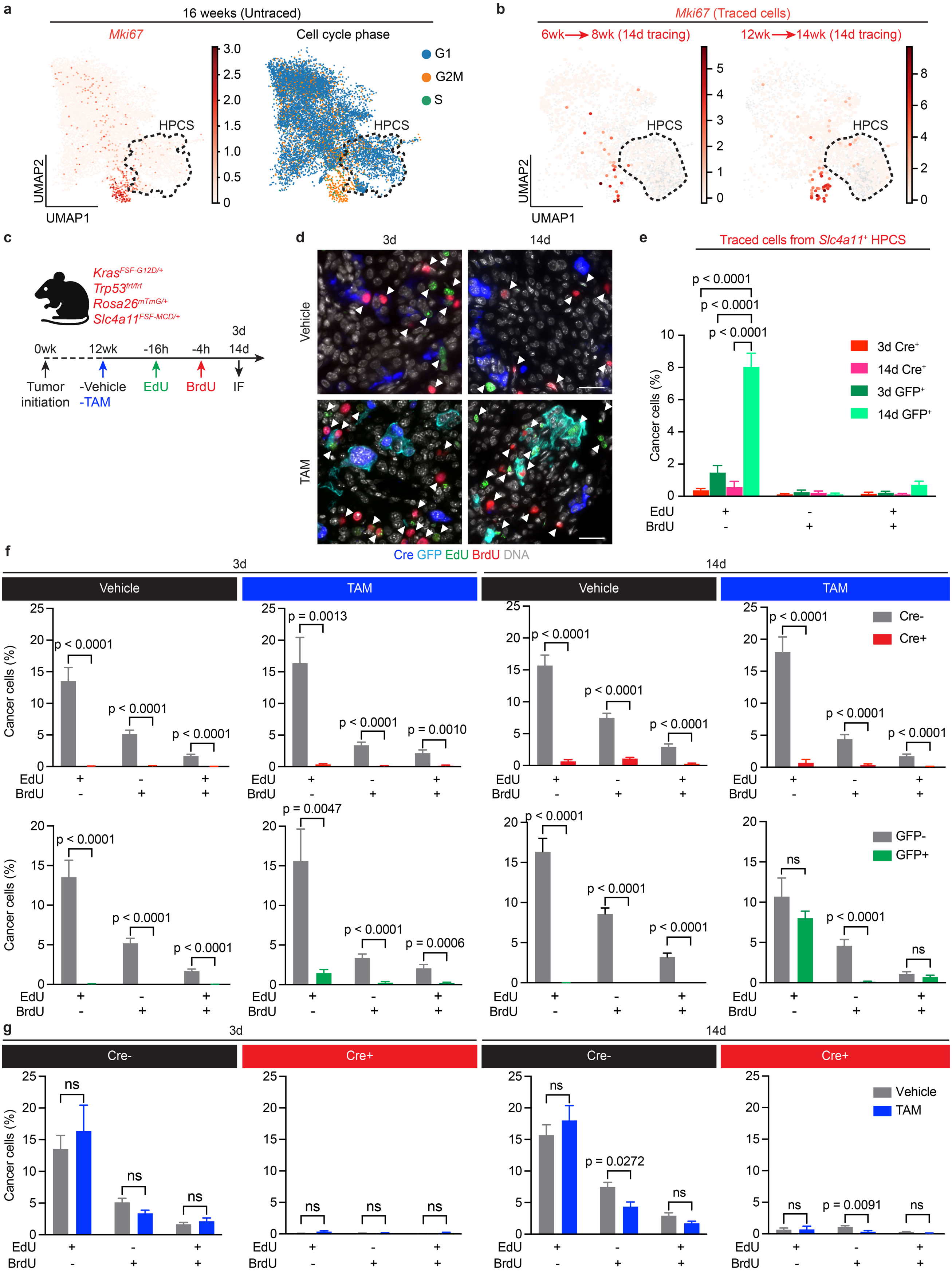
Extended Data Fig. 5 | See next page for caption.

Extended Data Fig. 5 | Flow cytometry, scRNA-seq data analyses of HPCS lineage tracing, and immunofluorescence imaging examining traced HPCS cells and progression markers. a-b, Quantification of HPCS in lineage-traced cells after tracing (n = 4 mice per time point; Mann-Whitney U test). **c-d**, mScarlet$^+$/EpCAM$^+$ (HPCS/all cancer cells, **c**) and GFP$^+$/EpCAM$^+$ (Traced cells/all cancer cells, **d**) at 3 days (3 d) or 14 days (14 d) of lineage tracing (6 to 8 weeks PTI). n = 4 mice per time point. Mann-Whitney U test. **e-f**, mScarlet$^+$/EpCAM$^+$ (HPCS/all cancer cells, **e**) and GFP$^+$/EpCAM$^+$ (Traced cells/all cancer cells, **f**) at 3 days (3 d) or 14 days (14 d) of lineage tracing (12 to 14 week). n = 4 mice per time point. Mann-Whitney U test. **g**, Cell states in combined tracing experiments from Fig. 2a. scRNA-seq data from HPCS lineage-traced cells are plotted at the indicated timepoints. **h**, Stacked bar graphs showing the distribution of cell states across individual mice from HPCS tracing experiments shown as stacked bar graphs as in Fig. 2c, e. Error bars are SEM. **i**, Distribution of *Nkx2-1*, *Sftpc*, *Hopx*, and *Hmga2* expression from traced LUAD cells at 8 weeks (6 to 8 week tracing, *left*) or 14 weeks (12 to 14 week tracing, *right*) PTI. **j**, Percentage of HPCS (red) or other cell states (green) expressing the indicated genes in traced LUAD cells at 8 weeks (6 to 8 week tracing, *left*) or 14 weeks (12 to 14 week tracing, *right*) PTI. Fisher's exact test. **k**, IF images showing co-staining of GFP (green) with either NKX2.1 (*left*, red), SPC (*middle*, red), or HMGA2 (*right*, red). Boxed areas are shown as individual insets below their respective images. Scale bar: 20 µm. **l**, Percentage of NKX2.1$^+$ (*left*), SPC$^+$ (*middle*) or HMGA2$^+$ (*right*) cells in GFP$^+$ LUAD cells quantified from IF co-staining of LUAD tissues harvested at 14 weeks PTI after 3 (3 d) and 14 (14 d) days post-tracing. NKX2.1: n = 21 tumours, 3 mice (3 d) and n = 111 tumours, 3 mice (14 d). SPC: n = 59 tumours, 3 mice (3 d) and n = 89 tumours, 3 mice (14 d). HMGA2: n = 84 tumours, 3 mice (3 d) and n = 35 tumours, 3 mice (14 d). Mann-Whitney U test. Error bars are SEM.



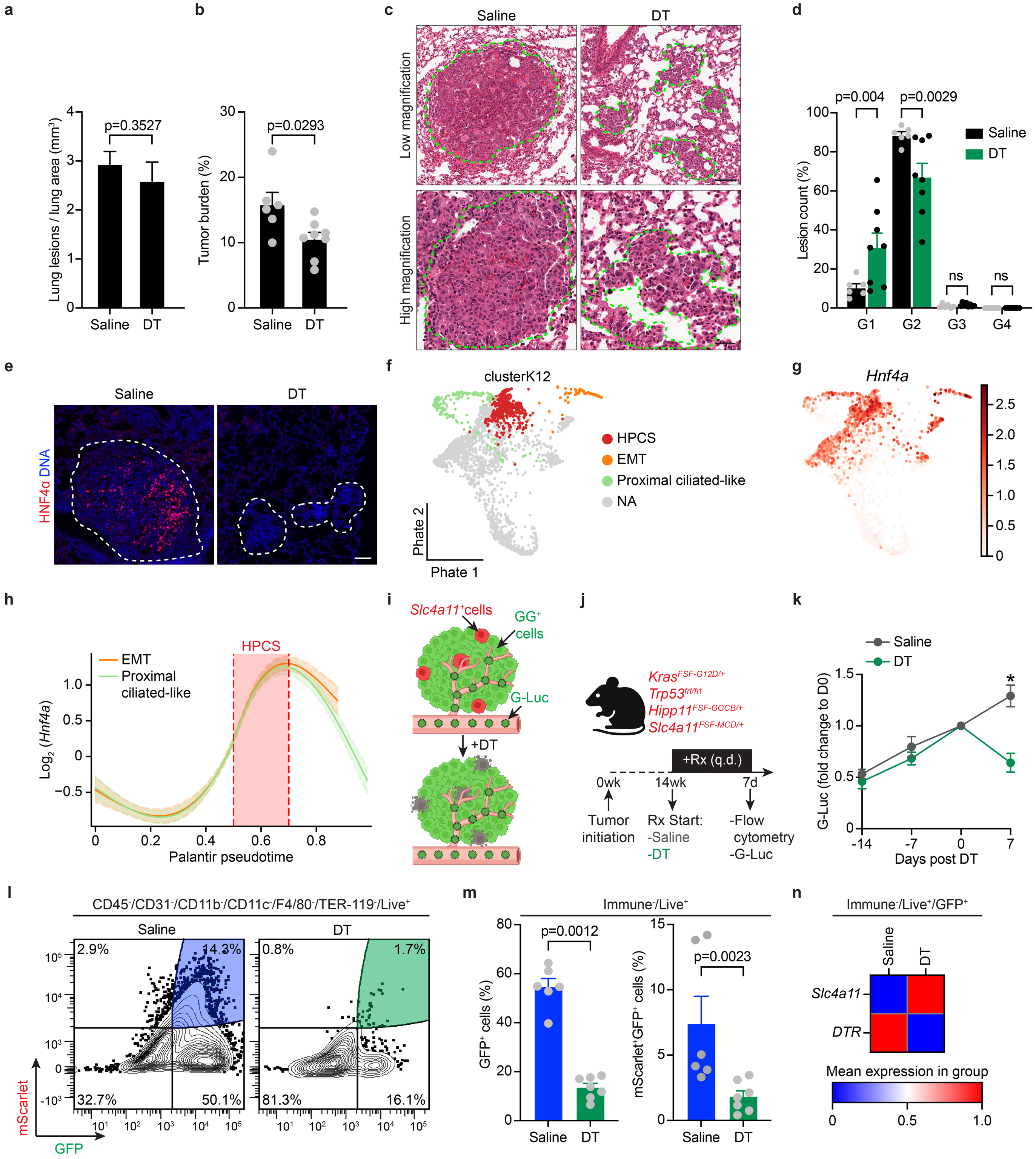
Extended Data Fig. 6 | See next page for caption.

Extended Data Fig. 6 | The *Hopx*$^{MACD/+}$ reporter allele and acquisition of the HPCS by non-HPCS cells. a, Map of *Hopx*$^{FSF\text{-}MACD/+}$ reporter and targeting strategy by CRISPR/Cas9 mediated HDR. Strategy for generating the genetically engineered *Hopx*$^{MACD/+}$ knock-in (KI) reporter enabling lineage-tracing and ablation of *Hopx*$^{+}$ AT1-like cells in *KP* LUAD: the *frt-stop-frt-P2A-mScarlet-T2A-Akaluc-P2A-CreERT2-P2A-DTR-WPRE* (*FSF-MACD*) reporter construct was knocked in frame into the stop codon of exon 3 of the *Hopx* gene by CRISPR/Cas9 gene editing. **b**, Experimental outline to generate *Kras*$^{FSF\text{-}G12D/+}$*; Trp53*$^{frt/frt}$*; Hopx*$^{FSF\text{-}MACD/+}$ reporter mouse. *FSF-MACD* reporter construct was knocked into *KPfrt* mESCs, and the correctly targeted mESC clones were microinjected into 8-cell stage embryos to obtain mESC-derived chimeras. **c**, Gel electrophoresis of the PCR products from the parental *KPfrt* mESCs, a correctly targeted mESC clone (positive control), or mESC-derived chimeras using primer pairs detecting either wild-type (WT, 437 bp, *top*) or knock-in (KI, 186 bp, *bottom*) alleles, respectively. **d**, Experimental design to validate the *Hopx*$^{FSF\text{-}MACD/+}$ reporter. Tumours were initiated by intratracheally delivered PGK-FlpO lentivirus in *KPfrt; Hipp11*$^{GGCB/+}$*; Hopx*$^{MACD/+}$ mice and LUAD tumours were harvested 16 weeks PTI for flow cytometry, IF staining, and scRNA-seq. **e**, Bioluminescence imaging of mice with autochthonous *KPfrt; Hipp11*$^{GGCB/+}$*; Hopx*$^{MACD/+}$ LUAD tumours. AkaLuc bioluminescence reports for *Hopx* expression. **f**, Representative IF images of mScarlet, HOPX, and GFP showing co-localization of mScarlet with HOPX (arrowheads) in GFP^{+} tumour cells. Scale bar: 50 µm. **g**, Representative flow cytometry plot showing the GFP^{+} tumour cells with mScarlet^{+} or mScarlet^{-} populations (highlighted regions) sorted from *KPfrt; Hipp11*$^{GGCB/+}$*; Hopx*$^{MACD/+}$ LUAD tumours 16 weeks PTI for scRNA-seq. **h**, Heatmap of scaled expression values for the AT1-like gene expression program, *Hopx*, or reporter cassette (as measured by *AkaLuc* and *DTR* expression) from sorted mScarlet^{+}/GFP^{+} or mScarlet^{-}/GFP^{+} LUAD cancer cells. n = 2 mice. **i**, LUAD cell states from the sorted cells shown in panel (**g**). **j**, Gene expression of *AkaLuc (left)* or *Hopx* (*right*) in GFP^{+} tumour cells sorted from autochthonous *KPfrt; Hipp11*$^{GGCB/+}$*; Hopx*$^{MACD/+}$ lung tumours. **k**, Experimental design to lineage-trace *Hopx*$^{+}$ AT1-like LUAD cells in *Kras*$^{FSF\text{-}G12D/+}$*; Trp53*$^{frt/frt}$*; Rosa26*$^{mTmG/+}$*; Hopx*$^{FSF\text{-}MACD/+}$ mice. At 12 weeks PTI, tumour-bearing mice were administered a single dose of tamoxifen (TAM; 20 mg/kg), which caused cells with high *Hopx*$^{MACD/+}$ expression to switch from tdTomato to GFP expression. Tumours were harvested at 3 days (3 d) and 14 days (14 d) for flow cytometry analysis. **l-n**, mScarlet^{+}/GFP^{+} (AT1-like/all traced cells, **l**), mScarlet^{+}/EpCAM^{+} (AT1-like/all cancer cells, **m**), and GFP^{+}/EpCAM^{+} (Traced cells/all cancer cells, **n**) cancer cells at 3 or 14 days of lineage tracing. n = 2 mice per time point. **o**, Analysis of the combined data from tracing experiments from Fig. 2a. and scRNA-seq data from AT1-like lineage-traced cells. **p**, Stacked bar graphs showing the distribution of cell states across individual mice from the AT1-like tracing experiments shown in (**o**) and stacked bar graphs in Fig. 2i. **q**, Tumour number per 10,000 (10k) cells transplanted. Transplanted cells were either mScarlet^{-} or mScarlet^{+} (Fig. 2l). n = 3 mice. Welch's t test. **r**, Experimental schema to examine the emergence of HPCS from non-HPCS derived transplantation. A *KPfrt; Slc4a11*$^{MACD/+}$ LUAD cell line expressing the miRFP670 reporter was subcutaneously transplanted into NSG mice (−7wk). mScarlet^{-}/miRFP670^{+} cells (non-HPCS) were sorted and subcutaneously transplanted into NSG mice (0wk). Tumours were harvested for flow cytometry analysis at 7 weeks post-transplantation. **s**, Representative flow cytometry plot of mScarlet and miRFP670 expression in tumours of the first transplantation (1st transplant), sorted non-HPCS cells (post-sort analysis), and second transplantation (2nd transplant). Cells were gated as CD45^{-}/CD31^{-}/CD11b^{-}/CD11c^{-}/F4/80^{-}/TER-119^{-}/DAPI^{-}(Live^{+})/miRFP670^{+}. **t**, Average percentage of mScarlet^{+} cells in miRFP670^{+} populations from the indicated groups. Post-sort mS^{-}: Post-sort analysis on non-HPCS cells as in (**s**). n = 4 mice (1st transplant), 5 samples (post-sort), 4 mice (2nd transplant). Mann-Whitney U test. Error bars are SEM. The diagrams in **b**, **d**, **k** and **r** were created using BioRender. Tammela, T. (2025) https://BioRender.com/0lgfrw5.



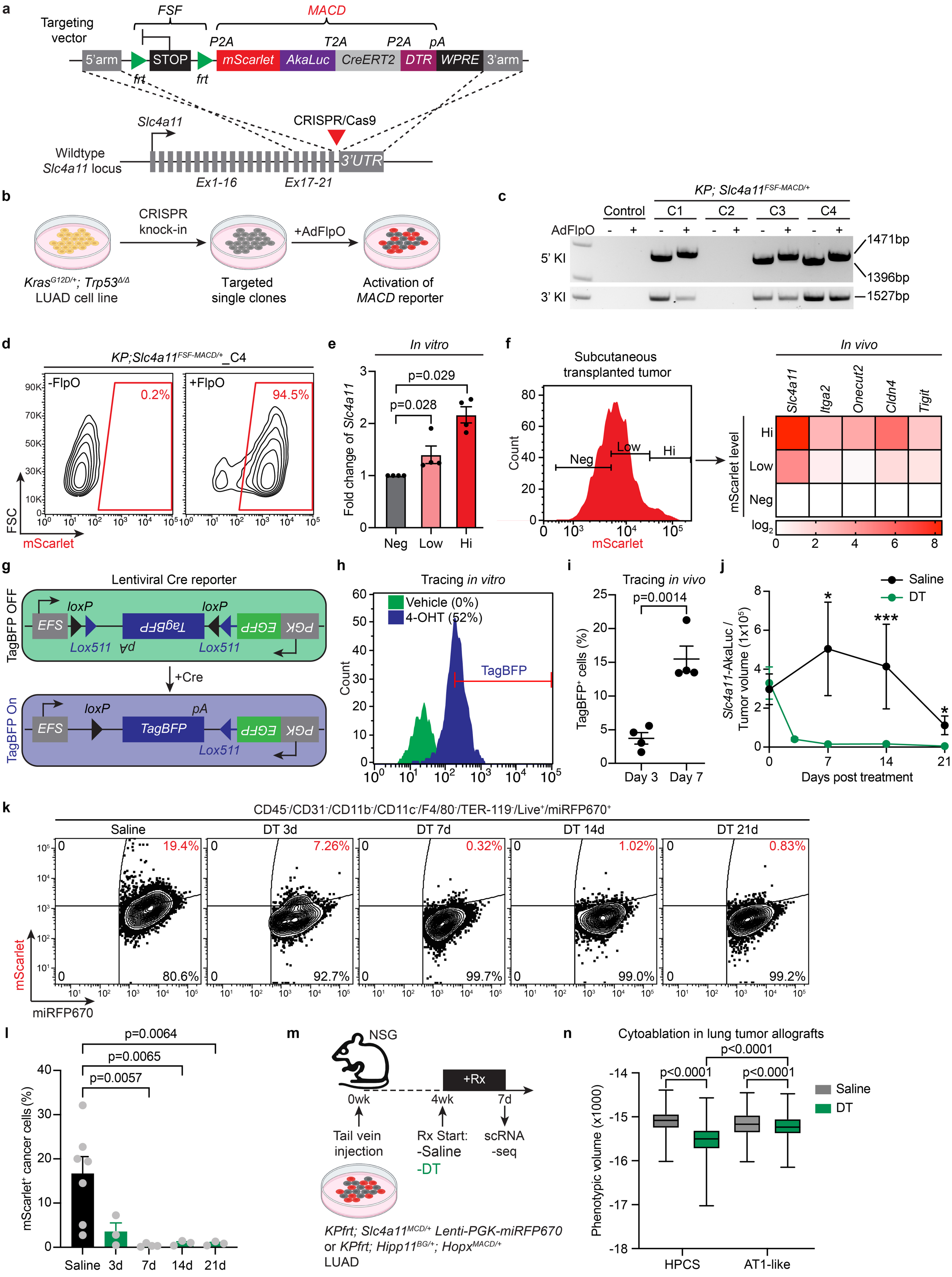
Extended Data Fig. 7 | See next page for caption.

Extended Data Fig. 7 | Analysis of proliferation in the HPCS and in HPCS-derived cells. a, *Mki67* expression (*left*) and the predicted cell cycle phase (*right*) in *KPfrt; Hipp11*$^{GGCB/+}$*; Slc4a11*$^{MCD/+}$ mice at 15-16 weeks PTI. **b**, *Mki67* expression in early-grade neoplasias and established adenocarcinomas, 14-days post-tamoxifen (TAM) labelling, in cells harvested at 8 weeks (*left*) and 14 weeks (*right*) PTI. **c**, Experimental design for EdU/BrdU dual labelling in *KPfrt; Rosa26*$^{mTmG/+}$*; Slc4a11*$^{MCD/+}$ mice. At 12 weeks PTI, tumour-bearing mice received a single dose of TAM (200 mg/kg), enabling lineage tracing of HPCS cells via a switch from tdTomato to GFP expression. EdU and BrdU were administered at 16 h and 4 h, respectively, before tumour harvest at 3 days (3 d) and 14 days (14 d) post-lineage tracing for IF analysis. **d**, Images showing IF staining of Cre (HPCS, blue), GFP (traced-HPCS, turquoise), EdU (green), BrdU (red), and DNA (DAPI, white) with (*bottom*) or without (*top*) TAM treatment at 3 d (*left*) or 14 d (*right*) post-lineage tracing. EdU$^+$ and BrdU$^+$ cells are pointed with arrowheads. Scale bar: 30 µm. **e**, Percentage of EdU$^+$/BrdU$^-$, EdU$^-$/BrdU$^+$, and EdU$^+$/BrdU$^+$ in Cre$^+$ (HPCS) or GFP$^+$ (lineage-traced) cells at 3 d or 14 d following lineage tracing. n = 30 tumours, 4 mice per group. Two-way ANOVA with Tukey's test. **f**, Percentage of EdU$^+$/BrdU$^-$, EdU$^-$/BrdU$^+$, and EdU$^+$/BrdU$^+$ comparing Cre$^+$ (HPCS) and Cre$^-$ (non-HPCS) cells (*top*) and GFP$^+$ (traced) and GFP$^-$ (non-traced) cells (*bottom*) at 3 d or 14 d following TAM or vehicle treatment. n = 30 tumours, 4 mice per group. Two-way ANOVA with Šídák's multiple comparisons test. **g**, Percentage of EdU$^+$/BrdU$^-$, EdU$^-$/BrdU$^+$, and EdU$^+$/BrdU$^+$ comparing vehicle and TAM groups between Cre$^-$ (non-HPCS) and Cre$^+$ (HPCS) cells at 3 d or 14 d following treatment. n = 30 tumours, 4 mice per group. Two-way ANOVA with Šídák's multiple comparisons test. Error bars are SEM. The diagram in **c** was created using BioRender. Tammela, T. (2025) https://BioRender.com/0lgfrw5.



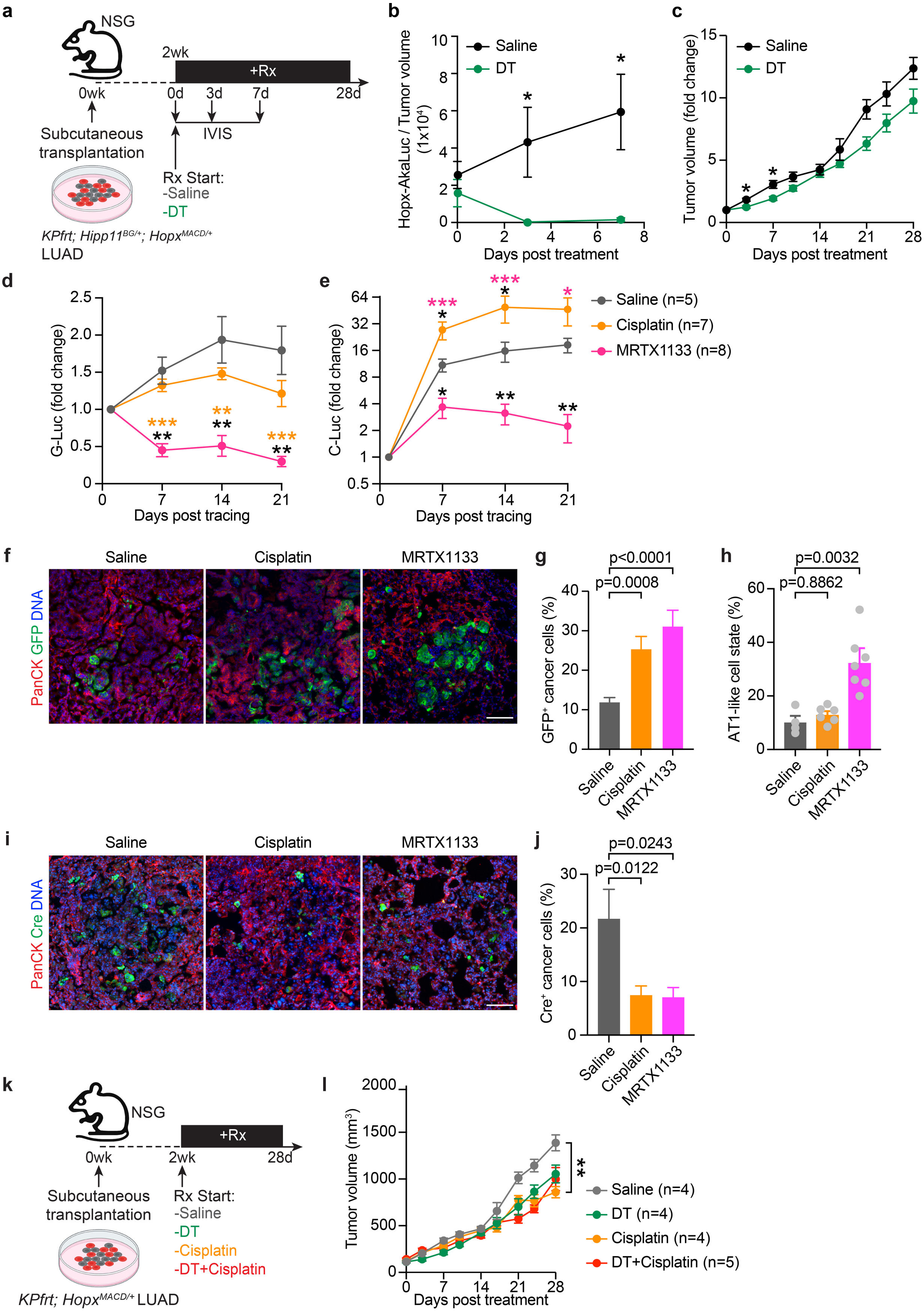
Extended Data Fig. 8 | See next page for caption.

Extended Data Fig. 8 | Histopathological and flow cytometry analyses of HPCS ablation in early neoplasias and in adenocarcinomas. a, Number of tumours normalized by lung area in saline vs. DT treated groups. n = 10 lung areas, 3 mice per group. Mann-Whitney U test. **b**, Tumour burden (total tumour area / total lung area) in saline (6 mice) vs. DT (8 mice) treated groups. Mann-Whitney U test. **c**, Representative H&E stained lung sections showing tumour boundaries (dashed lines) in saline control vs. HPCS-ablated (DT) mice. Scale bar: 100 µm (low magnification) and 40 µm (high magnification). **d**, Distribution of tumour numbers across tumour grades in saline vs. DT treated groups. n = 6 mice (saline) and 8 mice (DT). Two-way ANOVA with Šídák's multiple comparisons test. **e**, HNF4α (red) and DNA (blue) in lung sections in saline control vs. HPCS-ablated (DT) mice. Scale bar: 100 µm. **f**, Predicted end cell states (EMT, Proximal ciliated-like) and the HPCS projected on the LUAD progression PHATE[82] map from Marjanovic et.[4]. **g**, *Hnf4a* expression. **h**, Gene expression trajectories for *Hnf4a* plotted in *Palantir*[80] pseudotime space, with estimated time spent in the HPCS shaded in red. **i-j**, Experimental design to test the effect of ablating *Slc4a11*$^{+}$ HPCS on tumour growth in autochthonous *KPfrt; Hipp11*$^{GGCB/+}$*; Slc4a11*$^{MCD/+}$ LUAD-bearing mice as determined by longitudinal G-Luc activity monitoring. Experimental schema of LUAD bearing mice 14 weeks PTI were administered saline or DT (50 µg/kg, daily) and analysed at 7 days as indicated. **k**, G-Luc activity (secreted by GG^{+} tumour cells) shown as fold change at indicated time points normalized to day 0 post-DT treatment. n = 3 mice per group. Welch's t test. **l**, Representative flow cytometry plot of saline- and DT-treated (7 day) tumours. Shaded regions indicate gated populations analysed. **m**, Total tumour burden (GFP^{+} cells, *left*) or HPCS (mScarlet^{+}/GFP^{+} cells, *right*) as a percentage of live, non-immune cells [CD45^{-}/CD31^{-}/CD11b^{-}/CD11c^{-}/F480^{-}/TER-119^{-}/Helix NP NIR^{-}(Live)$^{+}$]. n = 5 (Saline) and 7 (DT) mice. Mann-Whitney U test. **n**, Heatmap of scaled mean gene expression for *DTR* (*Slc4a11*$^{MCD/+}$ allele) and native *Slc4a11* in saline vs. HPCS-ablated (DT) GFP^{+} cancer cells. Error bars are SEM. The diagrams in **i** and **j** were created using BioRender. Tammela, T. (2025) https://BioRender.com/0lgfrw5.



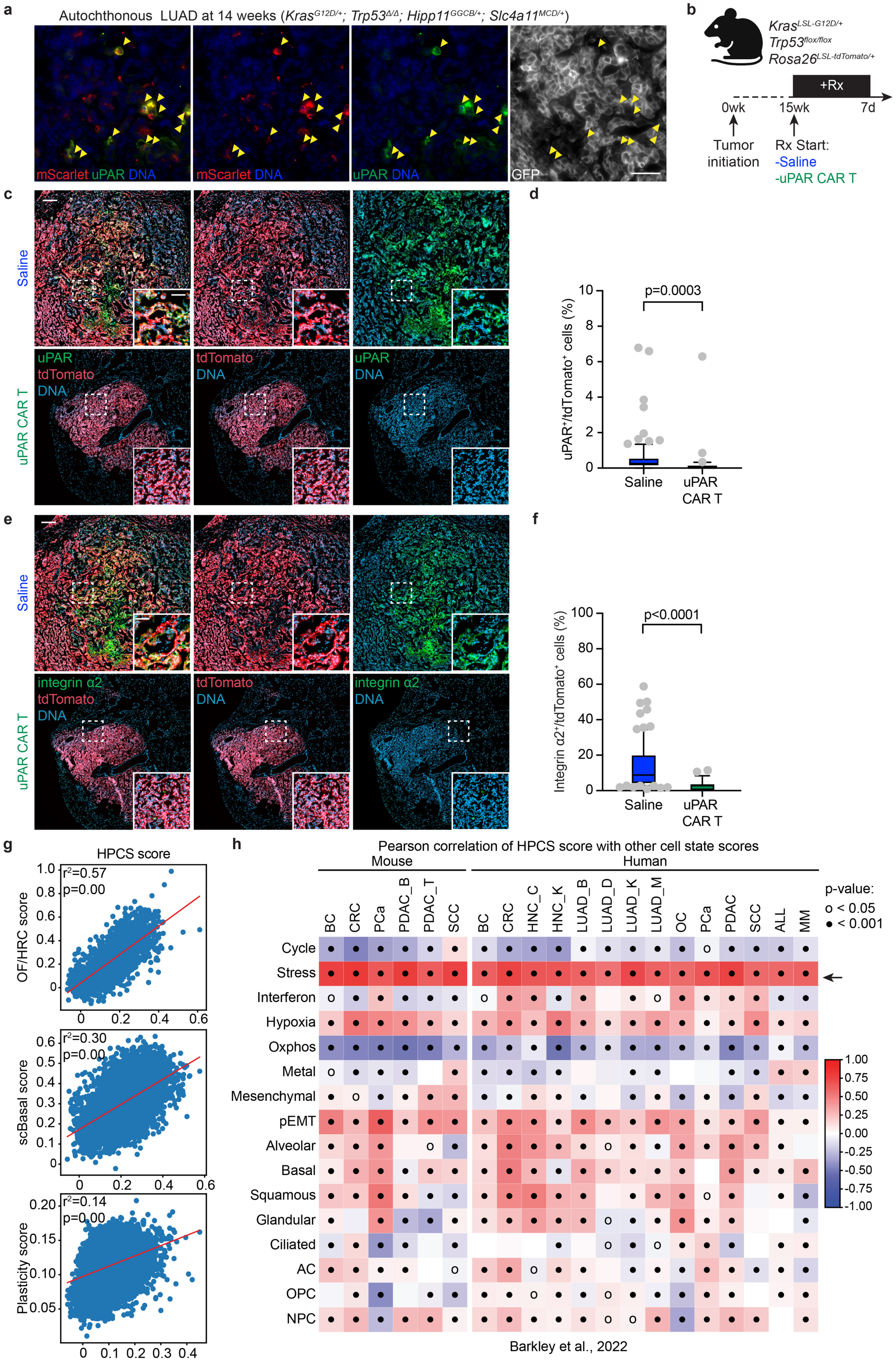
Extended Data Fig. 9 | See next page for caption.

Extended Data Fig. 9 | Construction, validation, and cytoablation of a *Slc4a11*$^{MACD/+}$ reporter cell line and cytoablation of HPCS or AT1-like cells in allotransplants. **a**, Vector map of *Slc4a11*$^{FSF\text{-}MACD/+}$ and targeting strategy of the reporter by CRISPR/Cas9 and homology-directed repair (HDR). *frt-stop-frt-P2A-mScarlet-T2A-Akaluc-P2A-CreERT2-P2A-DTR-WPRE* (*FSF-MACD*) reporter construct was knocked in frame into the *Slc4a11* locus in the presence of CRISPR/Cas9 mediated gene editing. Akaluc: Akaluciferase; P2A and T2A: short polypeptide cleavage site. **b**, Experimental design to generate *Slc4a11*$^{MACD/+}$ reporter cell line. A *Kras*$^{G12D/+}$*; Trp53*$^{-/-}$ (*KP*) LUAD cell line was transfected with the *Slc4a11-MACD* targeting vector in the presence of U6-sg*Slc4a11*-EFS-Cas9 and targeted single clones were transduced with adenoviral-CMV-FlpO (AdFlpO) to remove the *FSF* cassette and activate the *MACD* reporter. **c**, Gel electrophoresis of the PCR products from targeted clones before and after AdFlpO using primers detecting knock-in of the 5′ (5′ KI, *top*) and 3′ (3′ KI, *bottom*) arms, respectively. Successful removal of the *FSF* cassette resulted in an increase of band size in the PCR detection of the 5′ arm knock-in. **d**, Flow cytometry analysis of mScarlet expression in targeted single clone 4 (*Slc4a11-FSF-MACD*_C4) of *KP* LUAD before and after AdFlpO transduction. **e**, Relative expression of *Slc4a11* transcripts (shown as fold change) measured by qRT-PCR in mScarlet negative (Neg), low (Low) or high (Hi) sorted cell populations from *Slc4a11-FSF-MACD*_C4 *KP* LUAD cells transduced with AdFlpO. n = 4. Mann-Whitney U test. **f**, *Left*: Representative flow cytometry plot of mScarlet expression from dissociated subcutaneously transplanted *Slc4a11-FSF-MACD*_C4 *KP* LUAD tumour. Cells were sorted for negative, low, or high expression of mScarlet as indicated. *Right*: Heatmap showing the relative expression of indicated genes (columns) measured by qRT-PCR from the sorted cell populations (rows). n = 2 biological replicates, scale is log2 as indicated. **g**, Schematic of lentiviral Cre recombinase reporter construct for the testing of CreER^{T2} functionality from the *Slc4a11*$^{FSF\text{-}MACD/+}$ *KP* LUAD reporter. *Slc4a11*$^{FSF\text{-}MACD/+}$ *KP* LUAD cells infected with the lentiviral Cre reporter expresses GFP alone and upon exposure to 4-hydroxytamoxifen (4-OHT) expresses TagBFP in *Slc4a11*$^{+}$ cells. **h**, Flow cytometry analysis of TagBFP expression from *Slc4a11-FSF-MACD* *KP* LUAD cells transduced with the lentiviral Cre reporter treated with either vehicle or with 4-OHT (1 μM) for 72 h. **i**, Percentage of TagBFP^{+} tumour cells from dissociated subcutaneous transplanted *Slc4a11*$^{FSF\text{-}MACD/+}$ *KP* LUAD (with lentiviral Cre reporter) at day 3 or day 7 post-tamoxifen (TAM, 200 mg/kg) treatment. n = 4 mice. Unpaired t-test. **j**, *Slc4a11*-AkaLuc bioluminescence signal intensity normalized to tumour volume in saline- vs. DT-treated allograft tumours (n = 8 tumours each from 4 mice per group). Measurements were taken at the indicated time points. Welch's t test. **k**, A *KPfrt*; *Slc4a11*$^{MACD/+}$ LUAD cell line expressing the miRFP670 reporter was subcutaneously transplanted into NSG mice. Flow cytometry analysis of mScarlet versus miRFP670 expression in LUAD allografts following saline or DT for the indicated time. Cells were gated as CD45^{-}/CD31^{-}/CD11b^{-}/CD11c^{-}/F4/80^{-}/TER-119^{-}/DAPI^{-} (live^{+})/miRFP670^{+}. **l**, Percentage of mScarlet^{+} cells within the miRFP670^{+} LUAD cell population as gated in (**k**) in subcutaneous LUAD allografts treated with either saline or DT over the indicated time course. n = 7 (saline), 3 (3 d), 4 (7 d), 3 (14 d), and 3 (21 d) mice. One-way ANOVA with Dunnett's multiple comparisons test. **m**, Experimental design to test the effects of ablating *Slc4a11*$^{+}$ HPCS or *Hopx*$^{+}$ AT1-like cells on tumour heterogeneity in intravenously delivered lung transplants of *KPfrt; Slc4a11*$^{MCD/+}$ or *KPfrt; Hopx1*$^{MACD/+}$ LUAD reporter cells. Four weeks after transplantation, mice were administered saline or DT (25 μg/kg, q.o.d.). Tumours were harvested after 7 days on treatment and miRFP670^{+} cancer cells were sorted for scRNA-seq. **n**, Phenotypic volume calculated by bootstrap resampling (n = 1000; 100 cells sampled with replacement per iteration from pooled cells of 2 mice [Saline] or 3 mice [DT] in each cell state group). Box plots show median (centre), 25th–75th percentiles (box), and minimum–maximum (whiskers). One-way ANOVA with Holm-Šídák test. Error bars are SEM. The diagrams in **b** and **m** were created using BioRender. Tammela, T. (2025) https://BioRender.com/0lgfrw5.



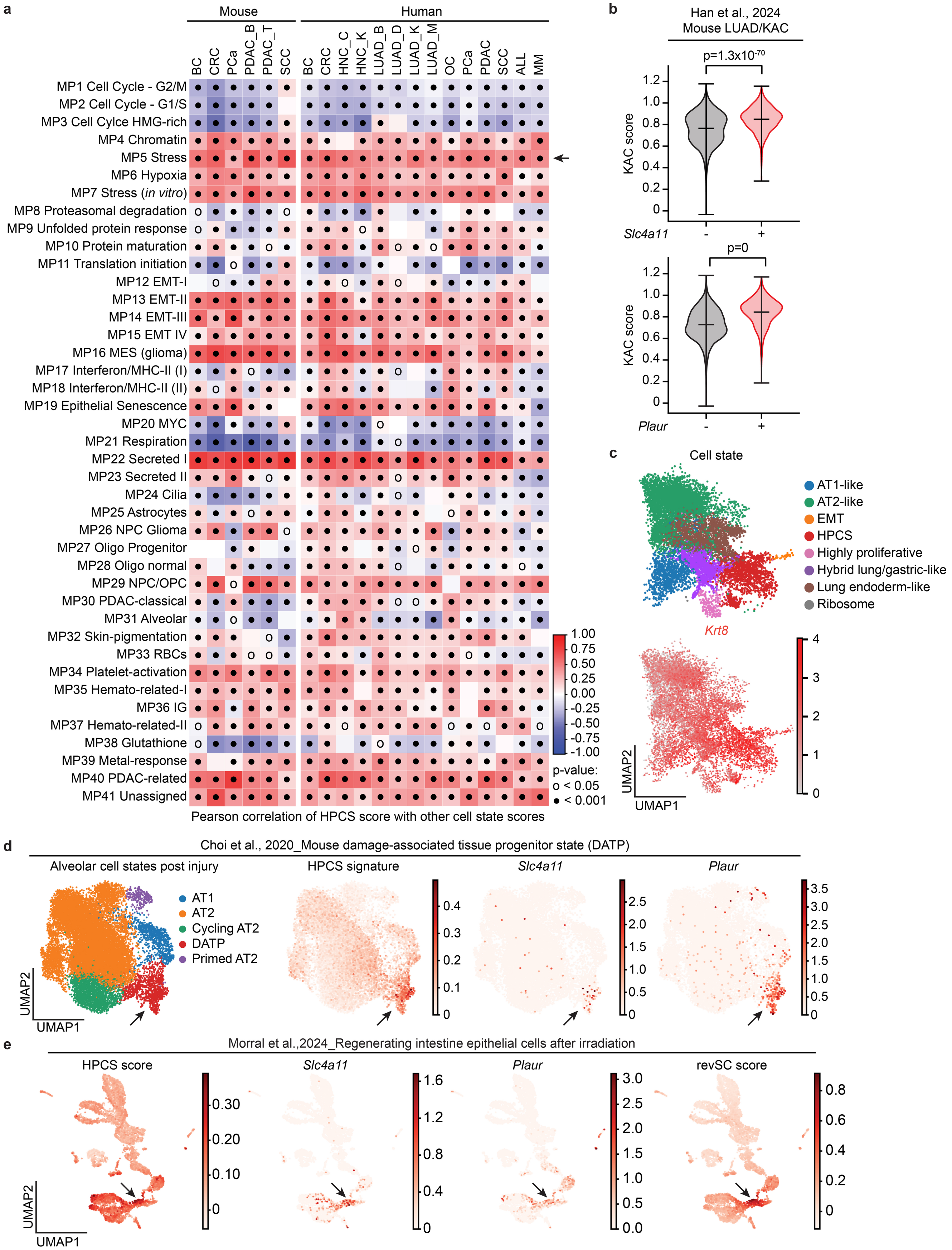
Extended Data Fig. 10 | See next page for caption.

Extended Data Fig. 10 | Combination of HPCS or AT1-like cell state ablation with lung cancer therapies. a-c, Data relating to Fig. 3l–m. **a**, Experimental design to test the effects of ablating *Hopx*$^{+}$ AT1-like cells on tumour growth in subcutaneous *KPfrt; Hopx*$^{MACD/+}$ LUAD reporter allografts. Two weeks after transplantation, allograft-bearing mice were administered saline or DT (25 µg/kg, q.o.d.). Tumour volume measurement and AkaLuc bioluminescence signal detection were performed at the indicated time point. IVIS: in vivo imaging system. **b**, *Hopx*-AkaLuc bioluminescence signal intensity normalized to tumour volume in saline-treated (n = 10 tumours, 5 mice) vs. DT-treated (n = 8 tumours, 4 mice) allografts. Welch's test. **c**, Tumour volume at indicated timepoints in allografts in mice administered either saline or DT. n = 8 tumours, 4 mice per group. Welch's test. **d-e**, Data relating to Fig. 4a, b. G-Luc (**d**) and C-Luc (**e**) activity at indicated time points (normalized to day 1) in saline (n = 5 mice), cisplatin (n = 7 mice), or MRTX1133 (n = 8 mice) treated groups. Note: Fold change of C-Luc under MRTX1133 therapy: 4x (7 d), 3x (14 d), and 2x (21 d). Welch's t-test. **f-j**, Data relating to Fig. 4c, d. **f**, Co-staining of PanCK (red), GFP (green, traced-HPCS) in *KPfrt; Rosa26*$^{mTmG/+}$*; Slc4a11*$^{MCD/+}$ LUAD tumours under indicated therapies for 3 weeks. Scalebar: 10 µm. **g**, Percentage of GFP^{+} cancer (panCK^{+}) cells under saline (n = 34 tumours, 3 mice), cisplatin (n = 27 tumours, 3 mice), or MRTX1133 (n = 20 tumours, 3 mice) treatments. One-way ANOVA with Tukey's multiple comparisons test. **h**, Percentage of HPCS-derived cells in the AT1-like cell state after 21 days of MRTX1133 therapy based on scRNA-seq analyses. n = 4 mice (saline), n = 6 mice (cisplatin), or n = 7 mice (MRTX1133). Kruskal-Wallis test. **i**, Co-staining of PanCK (red) and Cre (green, HPCS) in *KPfrt; Rosa26*$^{mTmG/+}$*; Slc4a11*$^{MCD/+}$ LUAD tumours under indicated therapies for 3 weeks. Scalebar: 10 µm. **j**, Percentage of Cre^{+} cancer cells under saline (n = 17 tumours, 3 mice), cisplatin (n = 21 tumours, 3 mice), or MRTX1133 (n = 13 tumours, 3 mice) therapy. One-way ANOVA with Tukey's multiple comparisons test. **k-l**, Data relating to Fig. 4e, f. **k**, Experimental design to test the effect of ablating *Hopx*$^{+}$ AT1-like LUAD cells by DT (25 µg/kg, q.o.d.) in the context of either saline control or cisplatin chemotherapy (1.5 mg/kg, every 3 days) in subcutaneous *KPfrt; Hopx*$^{MACD/+}$ LUAD reporter allografts. **l**, Volume of subcutaneous *KPfrt; Hopx*$^{MACD/+}$ LUAD allografts subjected to saline control, DT, or cisplatin, alone, or in combination, as in (**k**). Saline, DT, and cisplatin: n = 8 tumours, 4 mice. Combination: n = 10 tumours, 5 mice. Two-way ANOVA with Tukey's multiple comparisons test. Error bars are SEM. The diagrams in **a** and **k** were created using BioRender. Tammela, T. (2025) https://BioRender.com/0lgfrw5.



Extended Data Fig. 11 | See next page for caption.

Extended Data Fig. 11 | uPAR CAR-T cell-mediated HPCS eradication and analysis of a HPCS-like state across cancers. a, mScarlet, uPAR, and GFP immunofluorescence. Yellow arrowheads indicate co-localization of mScarlet with uPAR (yellow arrowheads) in GFP$^+$ tumour cells. Scale bar: 20 µm. **b**, Outline of the experimental design to evaluate uPAR CAR-T cells (2×10^6 cells/mouse) in autochthonous *KP*; *Rosa26*^{tdTomato}/+ LUAD with one dose of either saline or uPAR CAR-T cells. **c**, Representative IF images of uPAR and tdTomato in autochthonous *KP; Rosa26*$^{tdTomato/+}$ LUAD tumours treated with saline versus uPAR CAR T cells. Inset depicts boxed regions. Large image scalebar: 200 µm, inset scalebar: 60 µm. **d**, Quantification of uPAR expressing tdTomato$^+$ cancer cells in saline (n = 93 primary tumours from one mouse) vs. uPAR CAR T cell (n = 31 primary tumours from one mouse) treated mice shown as a boxplot; Kruskal-Wallis test. **e**, Representative IF images of integrin α2 and tdTomato in autochthonous *KP; Rosa26*$^{tdTomato/+}$ LUAD bearing mice treated with vehicle or uPAR CAR T cells. Inset depicts regions boxed with a dashed line. Large image scalebar: 200 µm, inset scalebar: 60 µm. **f**, Boxplot of integrin α2$^+$/tdTomato$^+$ tumour cell percentages in saline (n = 83 primary tumours from one mouse) vs. uPAR CAR T (n = 25 primary tumours from one mouse) treated mice; Kruskal-Wallis test. **g**, Pearson correlation between the HPCS signature score (*x*-axis) and plastic cell states (*y*-axis) from colorectal cancer (OF/HRC: oncofetal/core high relapse[45]) (*top*), pancreatic ductal adenocarcinoma (scBasal[47]) (*middle*), or castration-resistant prostate cancer adenocarcinoma (Plasticity[16]) (*bottom*). **h**, Heatmap of Pearson correlations calculated between the HPCS and recurrent pan-cancer cell states (AC: astrocyte-like, OPC: oligodendrocyte progenitor cell-like, NPC: neural progenitor cell-like)[48], across cancer type (BC: breast cancer, CRC: colorectal cancer, HNC: head and neck cancer, LUAD: lung adenocarcinoma, OC: ovarian cancer, PCa: prostate adenocarcinoma, PDAC: pancreatic ductal adenocarcinoma, SCC: squamous cell carcinoma of the skin, ALL: acute lymphoblastic leukaemia, MM: multiple myeloma). Arrow indicates the *Stress*-associated cell state from ref. 48. Full study list in Supplementary Table 5. All boxplots are depicted with the box (25th–75th percentiles), centre line (median), and whiskers (10th and 90th percentiles). The diagram in **b** was created using BioRender. Tammela, T. (2025) https://BioRender.com/0lgfrw5.



Extended Data Fig. 12 | See next page for caption.

Extended Data Fig. 12 | Analysis of a HPCS-like state across cancers and epithelial injury models. a, Heatmap of Pearson correlations calculated between the HPCS and recurrent pan-cancer cell states (RBC: red blood cells, IG: immunoglobulins)[49], divided by cancer type (BC: breast cancer, CRC: colorectal cancer, HNC: head and neck cancer, LUAD: lung adenocarcinoma, OC: ovarian cancer, PCa: prostate adenocarcinoma, PDAC: pancreatic adenocarcinoma, SCC: cutaneous squamous cell carcinoma, ALL: acute lymphoblastic leukaemia, MM: multiple myeloma). Arrows indicate correlations to the *Stress*-associated cell state. A full list of studies is listed in Supplementary Table 5. p values are unadjusted and calculated from the exact distribution of Pearson correlation coefficient. **b**, Violin plot of the *Krt8*$^+$ alveolar intermediate cell state (KAC)[5] score in *Slc4a11* (*top*) and *Plaur* (*bottom*) negative and positive cells. t-test. **c**, Expression of *Krt8*, a marker of the "KAC" subset of LUAD cells, in *KP* lung adenocarcinoma cells. Note broad, near-uniform expression of *Krt8* throughout all the LUAD cell states. **d**, Classification of cell states (*far left*), expression of the HPCS program (*middle left*), *Slc4a11* (*middle right*) or *Plaur* (*far right*) in scRNA-seq data obtained from injured and non-injured primary lung tissue. Arrows point to the damage associated transient progenitor (DATP) cell state[50]. **e**, Distribution of the HPCS program (*far left*), expression of *Slc4a11* (*middle left*) or *Plaur* (*middle right*) in scRNA-seq data obtained from injured and non-injured primary mouse small intestine tissue[60]. Arrows point to the damage associated revival stem cell (revSC) cell state.

nature portfolio

Corresponding author(s): Yan Yan and Tuomas Tammela

Last updated by author(s): Nov 14, 2025

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐	☒	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	FASTQ files of scRNA-seq data generated on the 10X Chromium or ChromiumX platform were processed using the standard CellRanger pipeline (version >=6.2). Reads were aligned to a custom GRCm38 / mm10 references including the EGFP, TagBFP, tdTomato, Gluc, Cluc, Akaluc, mScarlet, CreERT2, and DTR
Data analysis	Cell-gene count matrices were analyzed using a combination of published packages and custom scripts centered around the SCANPY / AnnData ecosystem. Single-cell RNA-sequencing datasets from different mouse models and primary patient samples were analyzed separately using similar workflows. Jupyter notebooks executing the analysis workflow and figure generation are available on Github at https://github.com/dbetel/HPCS_LUAD. Software used include SCANPY (v >=1.9), numpy (v >=1.26), scipy (v >=1.12), scikit-learn (v >=1.13), leidenalg (v 0.10.2), matplotlib (v 3.8.4), Cellrank (v2.0.7), Palantir (v1.4.1), R (v4.3.3). Information to recreate the conda environment used to run the analysis is available at the Github address listed above. Graphpad Prism Software (>v9.0) was used to analyze and create figures for the manuscript. Flow cytometry data was analyzed with FlowJo Software (v10.10.0). Image quantification was performed using FIJI/ImageJ (v1.54).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

All generated sequencing data and count matrices are available at the NCBI Gene-Expression Omnibus under accession record GSE277777. A list of previously published datasets analyzed in this manuscript is available in Supplementary Table 5.

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	All analysis on human material was performed on previously published data. References to the studies used in this work are available in Supplementary Table 5.
Reporting on race, ethnicity, or other socially relevant groupings	All analysis on human material was performed on previously published data. References to the studies used in this work are available in Supplementary Table 5.
Population characteristics	All analysis on human material was performed on previously published data. References to the studies used in this work are available in Supplementary Table 5.
Recruitment	All analysis on human material was performed on previously published data. References to the studies used in this work are available in Supplementary Table 5.
Ethics oversight	All analysis on human material was performed on previously published data. References to the studies used in this work are available in Supplementary Table 5.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No sample size calculation was performed for the given animal studies. For Slc4a11-MCD/+, Hopx-MACD/+, Rosa26-CreERT2, Rosa26-mTmG/+ and Hipp11-GGCB/+ studies, we were restricted by the number of mice with the correct genotype to be included in the experimental groups. For transplantation studies, initial numbers of mice for each group were n>5, which we felt should suffice to show a statistical difference if the effect size was robust and significantly large. For animal studies and other experiments, sample sizes were determined based on our prior experience with similar models, sufficient to detect biologically meaningful differences while minimizing animal use.
Data exclusions	We excluded analysis of the single cell data from the 7 day diphtheria toxin ablation experiment when we realized Slc4a11 expression persisted despite DT administration (Extended Data Figure 8m-n). We suspect this is related to incomplete recombination of the reporter allele and have indicated as much in the manuscript.
Replication	All biological replicates for single cell studies were included in our analysis using the TotalSeq-B cell hashing method to minimize batch effect. Luciferase assays were performed with at least three biological replicates. All in vivo studies included biological replicates, with the number of mice and tumors (where applicable) indicated. For single-cell and transplant in vivo studies, experiments were independently performed at least 3 times, with consistent results observed across replicates. For luciferase assays and other experiments, each experiment was independently repeated 3 times, and all attempts at replication yielded comparable results.
Randomization	All mice were randomly allocated to their experimental groups. Where possible, sex was matched. For genetic targeting and other experiments, allocation was not randomized because group assignment was inherent to the design. Key variables, including cell density and culture conditions, were carefully controlled to ensure consistency and avoid bias.

Blinding

Blinding was not possible given the robustness of the effects on treatment (vehicle vs chemotherapy or Kras inhibitors), where changes in tumor volume were readily apparent and easily distinguishable between groups.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☐	☒ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☐	☒ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used

A full list of antibodies used in the study are available in Supplementary Table 4.

Validation

Mouse IHC/IF Antibodies
Integrin α2 Rabbit mAb [EPR17338] Abcam ab181548, IP, WB, ICC/IF, Flow (Intracellular), IHC-P; KO validated, cited in 27 publications
RFP Rabbit pAb Rockland 600-401-379, WB, ELISA, IHC, IF, FC, EM, FISH, IP, cited in 1187 publications
GFP Chicken pAb Thermo Fisher Scientific A10262, WB, IHC, IHC-P, IHC-F, ICC/IF, Flow, IP, cited in 375 publications
NKX2-1 Rabbit mAb [EP1584Y] Abcam ab76013, IHC-P, Flow, WB, ICC/IF, cited in 175 publications
HMGA2 Rabbit mAb [D1A7] Cell Signaling 8179, WB, IHC, IF cited in 70 publications
Ki-67 Rat mAb [SolA15] Thermo Fisher Scientific 14-5698-82, WBC, IHC, IHC-P, IHC-F, ICC/IF, Flow, IP, cited in 378 publications
Flag Mouse mAb [M2] Sigma F1804, WB, IP, IHC, IF, ICC, cited in 9626 publications
Cre Recombinase Rabbit mAb [D7L7L] Cell Signaling 15036, WB, IHC, IF, Flow, cited in 84 publications
HOPX Mouse mAb [E1] Santa Cruz Biotechnology sc-398703, IP, WB, IHC-P, ELISA, Flow, cited in 84 publications
SPC Rabbit pAb Sigma-Aldrich AB3786, IHC, WB, cited in 367 publications
uPAR Goat pAb R&D Systems AF534, WB, Flow, IHC, CyTOF, cited in 16 publications
HNF4α Rabbit mAb [C11F12] Cell Signaling 3113, WB, IHC, IF, cited in 107 publications
Bromodeoxyuridinefrom mouse IgG1 Roche 11170376 [BMC9318], Flow, IHC https://www.scientificlabs.com/en/product/antibodies/11170376001#specification
GFP Chicken pAb Abcam ab13970, WB, ICC/IF, cited in over 3610 publications
Cre Recombinase Rabbit pAb 908001, WB, cited in 21 publications
pan Cytokeratin Mouse mAB [AE1/AE3], M351529-2 https://www.agilent.com/store/en_US/Prod-M351529-2/M351529-2

Human IHC/IF Antibodies
uPAR Goat pAb R&D Systems AF807, WB, IHC, IP, cited in 13 publications
Integrin α2 Rabbit mAb [EPR17338] Abcam ab181548, IP, WB, ICC/IF, Flow (Intracellular), IHC-P; KO validated, cited in 27 publications
pan keratin Mouse mAb [80] Abcam ab8068, Flow, WB, IHC-P, ICC/IF, ICC-Fr, cited in 41 publications

TotalSeq'"-B anti-mouse Hashtag Antibodies
https://www.biolegend.com/fr-ch/products/totalseq-b0301-anti-mouse-hashtag-1-antibody-17771
https://www.biolegend.com/fr-ch/products/totalseq-b0302-anti-mouse-hashtag-2-antibody-17772
https://www.biolegend.com/fr-ch/products/totalseq-b0303-anti-mouse-hashtag-3-antibody-17773
https://www.biolegend.com/fr-ch/products/totalseq-b0304-anti-mouse-hashtag-4-antibody-17774
https://www.biolegend.com/fr-ch/products/totalseq-b0305-anti-mouse-hashtag-5-antibody-17775
https://www.biolegend.com/fr-ch/products/totalseq-b0306-anti-mouse-hashtag-6-antibody-17776
https://www.biolegend.com/fr-ch/products/totalseq-b0307-anti-mouse-hashtag-7-antibody-17777
https://www.biolegend.com/fr-ch/products/totalseq-b0308-anti-mouse-hashtag-8-antibody-17778
https://www.biolegend.com/fr-ch/products/totalseq-b0309-anti-mouse-hashtag-9-antibody-17779
https://www.biolegend.com/fr-ch/products/totalseq-b0310-anti-mouse-hashtag-10-antibody-18225
https://www.biolegend.com/fr-ch/products/totalseq-b0311-anti-mouse-hashtag-11-antibody-22812
https://www.biolegend.com/fr-ch/products/totalseq-b0312-anti-mouse-hashtag-12-antibody-22803
https://www.biolegend.com/fr-ch/products/totalseq-b0313-anti-mouse-hashtag-13-antibody-22804
https://www.biolegend.com/fr-ch/products/totalseq-b0314-anti-mouse-hashtag-14-antibody-22805
https://www.biolegend.com/fr-ch/products/totalseq-b0315-anti-mouse-hashtag-15-antibody-22806

All primary antibodies were validated for the species used (mouse and human), either by the manufacturer or in published studies. Validation details, including manufacturer statements, relevant citations, and profiles from online antibody databases, are provided (catalog numbers listed above).

Eukaryotic cell lines

Policy information about cell lines and Sex and Gender in Research

Cell line source(s)	Cell lines were derived from KP female mice in the laboratory.
Authentication	Cell lines were generated in house and were genotyped as shown in Extended Data Figures 1d, 1g, 6c, and 9c.
Mycoplasma contamination	All cell lines used tested negative for Mycoplasma contamination.
Commonly misidentified lines (See ICLAC register)	No commonly misidentified cell lines were used in the study.

Animals and other research organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research, and Sex and Gender in Research

Laboratory animals	All immunocompetent mice used in this study are C57BL/6 x Sv129 mixed background. Immunodeficient mice were NOD.Cg-Prkdcscid Il2rgtm1Wjl/SzJ (aka NSG mice) bought mainly from The Jackson Laboratory. Mice used in the study are between 8-30 weeks old. A full list of immunocompetent mice used in this study are available in Supplementary Table 4.
Wild animals	No wild animals were used in this study.
Reporting on sex	Where possible, sex was matched during randomization. A full list of animals including age and sex are available in Supplementary Table 4.
Field-collected samples	This study did not involve samples collected from the field.
Ethics oversight	All animal studies were approved by the Memorial Sloan Kettering Cancer Center (MSKCC) Institutional Animal Care and Use Committee (protocol # 17-11-008).

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Plants

Seed stocks	*Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.*
Novel plant genotypes	*Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.*
Authentication	*Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosiacism, off-target gene editing) were examined.*

Flow Cytometry

Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	Dissociation of lung adenocarcinomas and lung tissue For isolation of normal AT2 cells and autochthonous LUAD cells, mice were euthanized at the indicated time points post tumor induction and were perfused with sterile S-MEM (Gibco, #11380-037) through the right ventricle of the heart. Dissected lungs or microdissected tumors were dissociated with a mixture of Dispase II (Corning, #354235, 0.6 U/ml), Collagenase Type IV (Thermo Fisher Scientific, #17104019; 0.167 U/ml), and DNase I (Stemcell Technologies, #07469; 10 U/ml) in S-MEM solution at 37 °C as previously described for 1 hour. The dissociated cells were filtered using a 100 µm filter and spun at 1500 rpm for 10 minutes at 4 °C. The supernatant was removed by aspiration and red blood cell lysis was performed

using BD Pharm Lyse (BD Biosciences, #555899) for 1 minute on ice. Cells were then washed with sterile media containing 2% heat inactivated (HI)-FBS (Hyclone, #SH30910.03), passed through a 40 µm filter, and pelleted at 300 g for 5 minutes at 4 °C. The supernatant was removed, and live cells were purified using the Akadeum Dead Cell Removal Microbubble kit as per manufacturer instructions (Akadeum Life Sciences, #11510-211). Cells were resuspended in Fluorescence-Activated Cell Sorting (FACS) buffer media (2% heat-inactivated FBS in PBS) and counted for use in FACS below.

Flow cytometry analysis and fluorescence-activated cell sorting (FACS)
Cells were prepared as above, and Fc block (BD Biosciences, #553142) was added on ice for 10 minutes prior to being stained with the appropriate antibody panel (Extended Data Table 4). After 20 minutes of staining on ice, cells were washed twice with FACS buffer and pelleted with a 5 minutes, 300 g spin at 4 °C. Cell pellets were resuspended in PBS with 2% HI-FBS containing DAPI (Sigma Aldrich, #D9542, 1 µg/ml) or Helix NP NIR (Biolegend, #425301, 5 nM) to identify dead cells. Cell sorting was performed at the Flow Cytometry Core Facility at Sloan Kettering Institute/MSKCC, using a BD FACS Aria Sorter. Cells were sorted using the '4-way purity' mode. Cancer cells were sorted as (CD45/CD31/CD11b/CD11c/F4/80/TER-119)-/Helix NP NIR- or DAPI- (live+) with specific fluorescent positive cell populations indicated in each experiment.

Instrument	BD FACS Aria
Software	BD FACSDiva (v9.0) was used to collect data during FACS. FlowJo (v10.10.0) was used to analyze the collected data.
Cell population abundance	Post-sort analysis was performed where applicable and indicated. When post-sort purity was not determined by flow analysis, the material was instead processed directly for single-cell sequencing. In general, >90% of the cells sequenced were as expected for each given flow panel.
Gating strategy	Cancer cells were sorted as (CD45/CD31/CD11b/CD11c/F4/80/TER-119)-/Helix NP NIR- or DAPI- (live+) with specific fluorescent positive cell populations indicated in each experiment. For Alveolar organoid culture and ex vivo transformation protocol, FACS-purified AT2 cells were sorted and gated as MHCII+/EpCAM+/Sca1-/podoplanin-/lineage-(CD45, CD31, CD11b, CD11c, F4/80 and Ter-119)-/DAPI-.

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.