Title

Alveolar type I cells can give rise to KRAS-induced lung adenocarcinoma

Journal

Cell Reports, 42(12)

ISSN

2639-1856

Authors

Yang, Minxiao

Shen, Hua

Flodby, Per

et al.

Publication Date

2023-12-01

DOI

10.1016/j.celrep.2023.113286

Peer reviewed

HHS Public Access

Author manuscript

Cell Rep. Author manuscript; available in PMC 2024 February 05.

Published in final edited form as:

Cell Rep. 2023 December 26; 42(12): 113286. doi:10.1016/j.celrep.2023.113286.

Alveolar type I cells can give rise to KRAS-induced lung adenocarcinoma

Minxiao Yang[1,2,3,4,5,11], **Hua Shen**[6,11], **Per Flodby**[6], **Michael D. Koss**[4,7], **Rania Bassiouni**[1,4], **Yixin Liu**[5,6], **Tea Jashashvili**[8], **Aaron Neely**[1,5], **Ezuka Ogbolu**[1], **Jonathan Castillo**[1,2,3,4,5], **Theresa Ryan Stueve**[2,3,4], **Daniel J. Mullen**[2,3,4], **Amy L. Ryan**[9], **John Carpten**[1,4], **Alessandra Castaldi**[10], **W. Dean Wallace**[4,7], **Beiyun Zhou**[4,5,6], **Zea Borok**[10], **Crystal N. Marconett**[1,2,3,4,5,12,*]

[1]Department of Translational Genomics, University of Southern California, Los Angeles, CA 90089, USA

[2]Department of Surgery, University of Southern California, Los Angeles, CA 90089, USA

[3]Department of Biochemistry and Molecular Medicine, University of Southern California, Los Angeles, CA 90089, USA

[4]Norris Comprehensive Cancer Center, University of Southern California, Los Angeles, CA 90089, USA

[5]Hastings Center for Pulmonary Research, University of Southern California, Los Angeles, CA 90089, USA

[6]Division of Pulmonary, Critical Care and Sleep Medicine, Department of Medicine, University of Southern California, Los Angeles, CA 90089, USA

[7]Department of Pathology, Keck School of Medicine, University of Southern California, Los Angeles, CA 90089, USA

[8]Department of Integrative Anatomical Sciences, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA

[9]Department of Anatomy and Cell Biology, Carver College of Medicine, University of Iowa, Iowa, IA 52242, USA

[10]Division of Pulmonary, Critical Care and Sleep Medicine, Department of Medicine, University of California, San Diego, La Jolla, CA 92093, USA

*Correspondence: crystal.marconett@med.usc.edu.

AUTHOR CONTRIBUTIONS

M.Y., A.L.R., A.C., J.C., B.Z., Z.B., and C.N.M. supervised the work, designed all the experiments, and interpreted the data. H.S., P.F., B.Z., and Z.B. generated transgenic mouse model, *Gramd2*-CreERT2. M.Y. maintained mouse strains, crossed the indicated mouse models with the $KRAS^{G12D}$ model, and established AT1 cell-derived LUAD models. M.Y., A.N., and E.O. performed survival analysis. T.J. supervised and performed micro-CT scanning and biostatistics study. M.D.K., W.D.W., Y.L., and M.Y. supervised and performed the histology studies. C.N.M., T.R.S., D.J.M., and R.B. performed the bioinformatic studies. All the authors contributed to the experiments and manuscript preparation. M.Y., Z.B., and C.N.M. wrote the manuscript.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.celrep.2023.113286.

DECLARATION OF INTERESTS

The authors declare no competing interests.

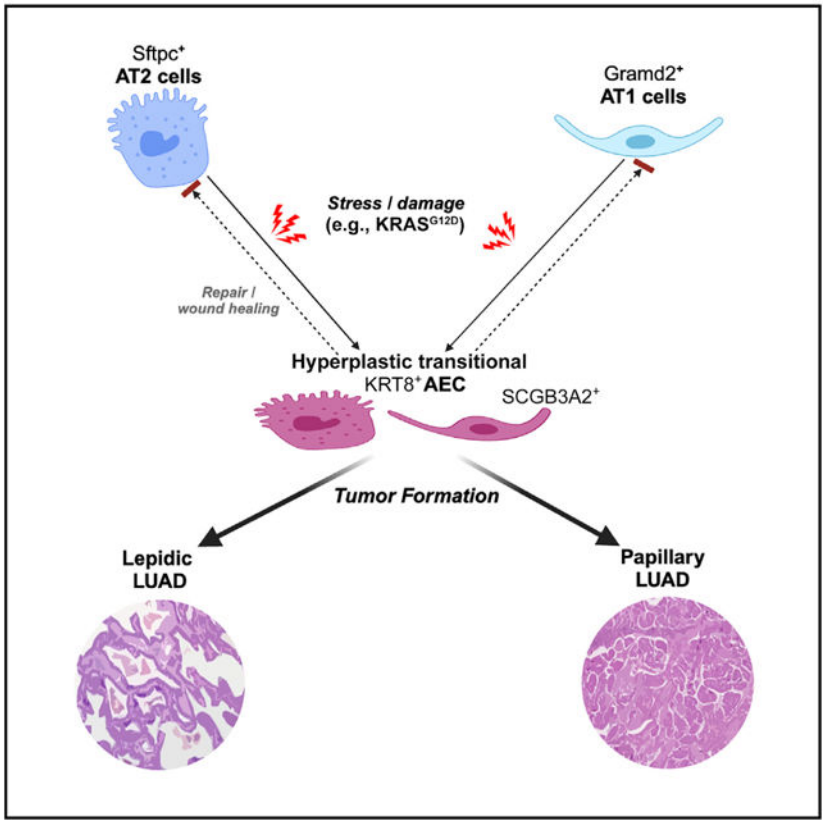
[11]These authors contributed equally

[12]Lead contact

SUMMARY

Lung adenocarcinoma (LUAD) is the most prevalent subtype of lung cancer and presents clinically with a high degree of biological heterogeneity and distinct clinical outcomes. The current paradigm of LUAD etiology posits alveolar epithelial type II (AT2) cells as the primary cell of origin, while the role of AT1 cells in LUAD oncogenesis remains unknown. Here, we examine oncogenic transformation in mouse Gram-domain containing 2 (*Gramd2*)$^+$ AT1 cells via oncogenic KRAS^{G12D}. Activation of KRAS^{G12D} in AT1 cells induces multifocal LUAD, primarily of papillary histology. Furthermore, KRT8$^+$ intermediate cell states were observed in both AT2- and AT1-derived LUAD, but SCGB3A2$^+$, another intermediate cell marker, was primarily associated with AT1 cells, suggesting different mechanisms of tumor evolution. Collectively, our study reveals that *Gramd2*$^+$ AT1 cells can serve as a cell of origin for LUAD and suggests that distinct subtypes of LUAD based on cell of origin be considered in the development of therapeutics.

Graphical Abstract



In brief

Yang et al. show that alveolar epithelial type I (AT1) cells can serve as a cell of origin for lung adenocarcinoma (LUAD). In addition, cell of origin can be associated with known properties of LUAD, including histologic subtypes and transcriptomic signatures, suggesting different mechanisms in tumor evolution.

INTRODUCTION

Lung cancer is the leading cause of cancer-related death in the United States and worldwide.[1] Originating in the parenchyma, lung adenocarcinoma (LUAD) is the most common subtype of lung cancer.[2] LUAD encompasses several pathologic subtypes, including solid, lepidic, papillary/micropapillary, and acinar, each with differing expectations of patient survival, and presents with a wide spectrum of genetic, epigenetic, and pathologic variation.[3] Several studies have linked epidermal growth factor receptor (EGFR) status to improved survival outcomes,[4] whereas KRAS mutations often occur concurrently with TP53 mutations and result in higher-grade LUAD with poorer survival outcomes.[5] However, the biology underlying these disparate observations is largely unknown.

LUAD occurrence is localized to the distal lung, where the predominant epithelial cell types are surfactant-producing alveolar epithelial type II (AT2) cells and large, delicate alveolar epithelial type I (AT1) cells.[6] AT1 cells cover over 95% of the alveolar epithelial surface and are largely responsible for facilitating gas exchange.[7] AT1 cells have been thought to be terminally differentiated[8] and consequently unable to proliferate.[9] Due to the delicate nature of AT1 cells, they are susceptible to injury. Several mouse lineage studies have shown that AT2, and more recently club cells, can facilitate replenishment of the AT1 cell population by proliferating and further differentiating to an AT1 cell identity.[10,11] The regenerative and self-renewal capacity of AT2 cells helped to establish this cell type as the primary cell of origin for LUAD.[2,12,13] In contrast, AT1 cells have yet to be thoroughly explored as a potential cell of origin for LUAD in large part due to a lack of highly cell-specific markers.

In addition to their distinct morphology, AT1 cells are distinguished by expression of a combination of markers, including aquaporin 5 (AQP5), podoplanin (PDPN), homeodomain-only protein homeobox (HOPX), G protein-coupled receptor class C group 5 (GPRC5A), and advanced glycosylation end-product-specific receptor (human: AGER, mouse: Rage), among others.[14] While each of these cell markers has demonstrated utility in identifying AT1 cells, most have modest RNA expression in other lung cell types. This confounds interpretation of previous attempts to activate oncogenic drivers, specifically in AT1 cells. Specific examples of this include HOPX, which has been used previously to demonstrate AT1 to AT2 cell differentiation, and the formation of lung nodules.[15] However, *Hopx* mRNA has been observed in AT2 cells in multiple studies,[14,16,17] raising the possibility that the observed results are due to RNA expression of study-specific drivers in AT2 cells.

Many mouse models have been utilized for the generation of LUAD lesions in mice. Prior studies have focused on leveraging *Sftpc*-CreERT2 lineage-traced mice[18,19] as a method of inducing LUAD from AT2 cells,[12] and, potentially from dual-positive *Sftpc*$^{+}$ *Scgb1a1*$^{+}$

bronchioalveolar stem cells (BASCs).[19,20] Classically, EGFR-driven LUAD mouse models have been generated using the *Scgb1a1/CCSP* promoter to drive expression[21]; however, these studies often require additional mutations and/or deletions to the EGFR axis not typically seen in primary LUAD to force lesion formation, suggesting that this mode of induction may not accurately reflect the cellular origins of LUAD in humans. Additional studies used the EGFR oncogenic driver in combination with the co-occurring mutations. For example, KEAP1 driven by *Gprc5a* has shown the ability to form histologically defined LUAD.[19,22] However, expression of Gprc5a in subsets of distal basal, as well as BASC populations in mice, call into question which specific cell type or types was able to give rise to observed LUAD lesions.[19] Therefore, we set out to determine if AT1 cells can also serve as a cell of origin for LUAD by employing a novel transgenic mouse model in which oncogenic $KRAS^{G12D}$ is activated following tamoxifen-inducible Cre recombination selectively in *Gramd2*$^{+}$ AT1 cells.

RESULTS

$KRAS^{G12D}$ activation in AT1 cells induces multifocal LUAD lesions *in vivo*

Novel insight into the cellular composition of LUAD has recently come to light through the single-cell characterization of human LUAD.[23] Interestingly, this study identified several LUAD patient tumors with distinct AT1 cell signatures. To functionally test AT1 cell contribution to LUAD, we activated the $KRAS^{G12D}$ oncogenic driver mutation specifically in AT1 cells. KRAS mutations contribute to ~30% of LUAD cases, with the most common G12D variant recapitulated in a well-established mouse model that can give rise to LUAD when activated in AT2 cells.[12,24] Our laboratory has previously invested in the identification of line-age-restricted gene expression patterns for AT1 cell markers within distal lung and subsequently identified *GRAMD2* as highly specific for AT1 cells.[14] Additional single-cell RNA sequencing studies have also observed significant enrichment for *Gramd2* in AT1 cells.[17] To leverage the endogenous regulatory regions that confer AT1 cell specificity to *Gramd2*, we inserted the tamoxifen-inducible CreERT2 transgene for genetic recombination into the 3′ end of the endogenous *Gramd2* gene to take advantage of the preexisting regulatory structure that confers AT1 cell specificity to *Gramd2* (Figure S1A). *Gramd2*-CreERT2:mTmG mice induced with tamoxifen demonstrated membranous expression of green fluorescent protein (GFP) in many, but not all, $AQP5^{+}$ AT1 cells (Figures S1B–S1D).[25] In contrast, GFP was not expressed in $SFTPC^{+}$ AT2 cells (Figures S1E–S1G), making it a valuable model for testing if AT1 cells can contribute to LUAD *in vivo*.

Gramd2-CreERT2 mice were then crossed with KRAS-LSL-G12D mice to create *Gramd2*:$KRAS^{G12D}$ mice alongside *Sftpc*:$KRAS^{G12D}$ mice, which are known to induce LUAD in AT2 cells. These transgenic models underwent intraperitoneal (IP) injection with optimized tamoxifen (TAM) concentrations, as previously described[26] and hereafter named *Gramd2*:$KRAS^{G12D}$ or *Sftpc*:$KRAS^{G12D}$. Harvested lungs were characterized via silicification followed by whole lung micro-computed tomography (micro-CT) scanning as performed previously[27] to determine if lesions could form upon $KRAS^{G12D}$ activation in AT1 cells.

Post-tamoxifen treatment, *Gramd2*:KRAS^{G12D} mice formed multifocal lesions throughout the whole lung (Figure 1A), as did the positive control *Sftpc*:KRAS^{G12D} mice. Corn oil vehicle controls for *Gramd2*:KRAS^{G12D} showed no appreciable lesions (Figure S2), indicating that Cre-mediated recombination did not occur unless specifically activated using tamoxifen.

To quantify the incidence and distribution of lesion formation, we performed a double-blind analysis where lesion multiplicity was counted, volumes calculated, and distance to the nearest bronchiole branchpoint measured. We observed approximately half the number of lesions in *Gramd2*:KRAS^{G12D} lungs (n = 189, 199, and 202 respectively) as compared with *Sftpc*:KRAS^{G12D} lungs (n = 439, 241, and 361, respectively) ($p < 0.001$), which may be due to several considerations including but not limited to optimization of Cre activity or pathogenetic properties between cell types. The mean volume of individual lesions varied depending on cell of origin as well, with *Gramd2*:KRAS^{G12D} lesions (median size $1.5e^{-2}$ mm^3, mean volume of $4.3e^{-2}$ mm^3 being significantly larger than those in the *Sftpc*:KRAS^{G12D} mice (mean tumor volume $1.0e^{-2}$ mm^3, median tumor volume of $2.2e^{-2}$ mm^3) model ($p < 0.001$) (Figure 1B). The median distance between each individual lesion and the nearest bronchiole also differed significantly between *Gramd2*:KRAS^{G12D} (0.7337 mm) and *Sftpc*:KRAS^{G12D} (0.6557 mm) mice ($p < 0.001$) (Figure 1C). One region in one *Sftpc*:KRAS^{G12D} lung was ~400x outside the range of volumes observed for all other lesions ($9.5e^{-2}$mm^3), and we were unable to determine if this represented one or multiple lesions due to limits in scanning resolution. To understand how cell of origin affected survival, three independent rounds of *Sftpc*:KRAS^{G12D} and *Gramd2*:KRAS^{G12D} mice were split into tamoxifen treatment and corn oil control groups and their overall survival measured. We observed significant decreases in survival compared with corn oil controls in both *Sftpc*:KRAS^{G12D} (n = 18) and *Gramd2*:KRAS^{G12D} (n = 21) mice (Figure 1D). Taken together, our results indicated that activation of KRAS^{G12D} in *Gramd2*$^+$ AT1 cells resulted in lesions detectable by micro-CT. However, whether detected lesions were consistent with known properties of lung cancer, including histologic patterning and subtypes, remained unclear.

Gramd2$^+$ AT1 cells give rise to LUAD with predominantly papillary histology

To determine the histologic properties of the lesions we observed in *Gramd2*:KRAS^{G12D} lungs, matching formalin-fixed paraffin-embedded (FFPE) sections were stained with hematoxylin and eosin (H&E) and underwent double-blind histologic evaluation alongside *Sftpc*:KRAS^{G12D} positive controls. Additional negative controls evaluated included wild-type (WT) C57BL/6J, *Sftpc*-CreERT2 without KRAS^{G12D}, *Gramd2*-CreERT2 without KRAS^{G12D}, and KRAS-LSL-G12D mice without Cre drivers. No hyperplastic foci were detected in H&E sections in negative control sections (Figure 2A). AT2 cell-derived *Sftpc*:KRAS^{G12D} lesions were classified solely as lepidic adenocarcinoma, whereas *Gramd2*:KRAS^{G12D} resulted in a mixture of lepidic and papillary adenocarcinoma (LA and PA), with papillary features being the predominant histologic subtype (Figure 2B). PA is characterized by fibrovascular cores and unappreciable alveolar lining that is replaced by malignant cuboidal and/or columnar cells,[28] whereas LA is characterized as non-mucinous adenocarcinoma that shows predominantly lepidic growth pattern but lacks

architectural complexity and lymphovascular or perineural invasion.[28,29] We also observed PA extending into the bronchial mucosa in *Gramd2*:KRAS^{G12D} lungs, hereafter termed bronchial infiltrative adenocarcinoma (BIA, Figures 2B; Figure S3). These BIAs did not form dense nodules and therefore may have not been detectable using densitometric tracing in the micro-CT images described above. The identified BIA lesions had classical PA histology within the bronchial lumen but failed to penetrate the lung epithelium lining the respiratory airways. This has been observed in multiple mouse models of LUAD[22] but is not a hallmark of human LUAD.

A section from each lung was evaluated to quantify differences in histology observed between KRAS^{G12D}-induced tumors based on cell of origin. The number of lesions of each histologic type was recorded from three different mice per genotype. This was done to avoid duplicate sampling errors that may occur from multiple sections of the same lung. We observed that LUAD of AT2 cell origin was significantly enriched for lepidic histology ($p = 0.005$). In contrast, LUAD of AT1 cell origin was significantly enriched for papillary histology ($p = 0.006$) compared with AT2-cell-derived tumors (Figure 2C). This was intriguing, as there are known survival differences between these histologic subtypes of human LUAD, with lepidic having significantly better overall survival outcomes.[30]

To confirm that the adenocarcinomas we observed in *Gramd2*:KRAS^{G12D} mice were of lung origin, we performed immunohistochemical (IHC) staining for standard markers of human LUAD that are currently used for classifying human tumors in clinical diagnosis. Nuclear homeobox protein NKX2–1 (also known as TTF-1) expression is a known characteristic of LUAD.[31,32] We observed positive nuclear NKX2–1 staining within the *Gramd2*:KRAS^{G12D} (Figure 2D) as well as control *Sftpc*:KRAS^{G12D} (Figure S4) tumors, indicating that the tumors were likely primary LUAD lesions and not metastatic from other tissues. However, NKX2–1 can also be indicative of carcinoma of thyroid origin. We therefore stained for thyroglobulin (TGB), which is present in thyroid carcinoma but absent in LUAD, to exclude a thyroid origin for these tumors. Tumors from *Gramd2*:KRAS^{G12D} and *Sftpc*:KRAS^{G12D} were negative for TGB staining (Figures 2D and S4). Combined with NKX2–1^{+} staining, these findings confirm that these adenocarcinomas are likely primary tumors of lung origin. We also performed Napsin A (NAPSA) staining to determine if *Gramd2*$^{+}$ tumors were enriched for the NAPSA PA marker.[33] We observed that *Gramd2*:KRAS^{G12D} papillary LUAD derived from AT1 cells were positive for NAPSA staining (Figure 2D), whereas AT2-cell-derived *Sftpc*:KRAS^{G12D} lepidic adenocarcinomas were not (Figure S4). Additional staining for cytokeratin deposition (clinical marker AE1/AE3),[32] CD68^{+} immune cell infiltration, and GATA3 were also performed to determine consistency between human LUAD staining patterns and what was observed in *Gramd2*:KRAS^{G12D} mice tumors (Figure S4).

Spatial transcriptomic profiling reveals distinct cell-of-origin-specific transcriptomic presentation tied to differential histology

To understand the mechanism(s) by which activation of KRAS^{G12D} leads to differential histologic manifestations of LUAD, we undertook spatial transcriptomic profiling using the 10X Genomics Visium platform. Spatial transcriptomics was chosen over other

technologies such as single-cell RNA sequencing to allow integration of histologic presentation and spatial distribution with transcriptomic signatures. This was especially important for *Gramd2*:KRAS^{G12D} lungs, as multiple LUAD histologic subtypes were observed, including BIA, PA, and LA, and the use of spatial technology allowed us to track their distinct histologic features throughout downstream analysis. Three biological replicates of *Gramd2*:KRAS^{G12D}- and *Sftpc*:KRAS^{G12D}-derived lung sections underwent transcriptomic profiling that included pathologist-verified LUAD lesions, alongside *Sftpc*-CreERT2 and *Gramd2*-CreERT2 controls (Figure S5). H&E-stained images generated by the Visium pipeline confirmed histologic characterization of these pathologist-defined LUAD lesions (Figure 3A) and that the *Gramd2*: KRAS^{G12D} samples included LUADs of multiple histologic subtypes (Figure 3B). Expression data from all individual lung samples were then clustered together to evaluate transcriptomic similarities and differences between *Gramd2*:KRAS^{G12D}- and *Sftpc*:KRAS^{G12D}-derived LUAD. We observed 24 distinct integrated clusters (ICs), so named because each individual 55-μm spot on the spatial array contains between 1 and 10 cells (Figure 3C). Normal alveolar tissue ICs (primarily IC1, red) were rarely observed in either *Gramd2*:KRAS^{G12D} or *Sftpc*:KRAS^{G12D} samples. Histologic analysis revealed that distinct ICs were associated with pathologist-defined LUAD histologic subtypes, with IC12 associated with PA, IC14 associated with BIA, and IC17 associated with LA. The distribution of these among samples showed an enrichment for PA and BIA in *Gramd2*:KRAS^{G12D} and enrichment for LA in *Sftpc*:KRAS^{G12D} samples (Figure 3D).

Uniform manifold approximation and projection (UMAP) was used for dimensional reduction and visualization of IC similarities and differences (Figure 3E). We observed a high degree of concordance between control *Gramd2*-CreERT2 and *Sftpc*-CreERT2, composed predominantly of IC1 and IC18 (Figure 3F). There was also a high degree of concordance between LUAD samples generated from the same cell of origin (Figure 3G), except for IC3 and IC10, which were enriched solely in *Gramd2*:KRAS^{G12D} replicate lung 3. This demonstrated that spatial transcriptomic profiling could discern distinct transcriptomic signatures associated with diverse histologic presentation.

Expression analysis reveals distinct transcriptomic signatures within AT1-derived LUAD

We then sought to determine if the histologically defined LUAD lesions recapitulated known gene expression signatures associated with human LUAD. To do this, we applied ESTIMATE, a tool developed on pan-cancer signatures from The Cancer Genome Atlas (TCGA) to estimate tumor content from expression data (Figure 4A).[34] We observed that ESTIMATE-computed array spots with high tumor purity were enriched in histologically defined LUAD clusters, but not completely concordant, as many spots included in histologically defined LUAD clusters did not meet the tumor purity threshold of ≥0.84 that were mutually exclusive from regions of immune (Figure S6) and stromal (Figure S7) enrichment. The 0.84 threshold was chosen based on the tumor purity level that eliminated >99% of array spots from control samples (Figure 4B). The four ICs identified previously were also the most highly enriched for elevated tumor purity via ESTIMATE (Figure 4B, starred ICs). UMAP projection demonstrated that spots with high tumor purity segregated into three regions (Figure 4C), corresponding to BIA, LA, and PA histopathology. An intermediate, “mixed” IC between LA and PA with high tumor purity

was also detected (IC20, pink). Closer examination of these high tumor purity regions demonstrated that subsets of multiple clusters were also included in the PA histologic subtype using ESTIMATE tumor purity thresholds (IC4, IC7, IC8, IC12); however, most array spots aligned to PA fell within IC12 (Figure 4D).

To determine how histologic presentation translated to transcriptomic states in LUAD, we undertook differential expression analysis among those ICs that were composed of a majority of array spots that had high tumor purity levels, which included IC12 (PA-associated), IC14 (BIA-associated), IC17 (LA-associated), and IC20 (LUAD with mixed histology). These ICs also showed elevated proliferation levels compared with other ICs, further supporting their role in carcinogenesis (Figure S8). Supervised clustering of the top 300 differentially expressed genes in each cluster revealed enriched expression patterns (Figure 4F). Among these were several known lung-cell-type-specific markers including the previously described AT1 cell markers *Sema3b*, *Spock2*, *Sema3g*, and *Rtkn2*.[14,35,36] Of note, IC14 with distinct BIA histology was enriched for several *Cpaf* genes, including but not limited to *Foxj1*, *Dnah5*, *Cpaf65*, *Cpaf43*, and *Cpaf100*, all of which have distinct functions in the formation and assembly of ciliated structures,[37] indicating that the BIA phenotype not observed in human LUAD is likely derived from low-level *Gramd2* expression in ciliated cells. This was further confirmed by immunofluorescent staining of BIA lesions with FOXJ1, a known ciliated cell marker (Figure S9). BIA did not appear to arise from club cells, another epithelial cell type resident in the distal bronchiolar epithelium, as SCGB1A1 staining in BIA was variegated (Figure S3) and not exclusively associated with tumor nuclei.

To further determine how known carcinogenic genes varied within each LUAD-associated IC, we performed gene set enrichment analysis (GSEA) using Protein Analysis Through Evolutionary Relationships (PANTHER) on IC12, IC14, IC17, and IC20.[38,39] GSEA showed that several pathways with distinct functions in carcinogenesis were significantly enriched, including proliferation, dysregulated cellular signaling, disruptions in cell-cell contact and extracellular matrix deposition (Figure 4G). Of note, PANTHER analysis confirmed that IC14 with BIA histology was enriched for cilia formation, motility, and axoneme assembly, suggesting that there is a small subset of ciliated cells present in the lower bronchiolar space with *Gramd2* expression that was not detected by previous methods. Separate from BIA in IC14, expression of key cancer-associated genes was compared between *Gramd2*:KRAS^{G12D} and *Sftpc*:KRAS^{G12D}. Several known cancer drivers in LUAD were enriched in *Gramd2*:KRAS^{G12D}, including *Egr1*,[40] a known downstream target of receptor tyrosine kinase signaling, *Cyp2s1*,[41] a member of the cytochrome p450 family, and *Hes6*, a known component of Notch signaling.[42]

Cellular heterogeneity and marker staining varies between AT2- and AT1-derived LUAD

To further interrogate alveolar epithelial cell composition within *Gramd2*:KRAS^{G12D}- and *Sftpc*:KRAS^{G12D}-derived LUADs, we examined the distribution of cell-type-specific markers within the LUAD-associated ICs (Figure 5A). Interestingly, with the exception of IC14 derived from multiciliated cells, the distribution of AT2- and AT1-associated cell markers showed no significant expression differences per cluster. This was also observed

when stratifying expression based on *Sftpc*:KRAS^{G12D} vs. *Gramd2*:KRAS^{G12D} cell of origin (Figure 5B). It was anticipated that *Gramd2*:KRAS^{G12D}- and *Sftpc*:KRAS^{G12D}-derived LUAD would have mutually exclusive cell type marker expression, with *Gramd2*-driven LUAD enriched for AT1 cell markers, and *Sftpc*-driven LUAD enriched for AT2 cell markers. Instead, we observed that both *Sftpc*- and *Gramd2*-driven LUAD were enriched for both cell-type-specific markers. This was further confirmed by observing the correlation between a panel of AT2 (red) and AT1 (green) cell-type-specific markers, including *Sftpc* vs. *Gramd2* (Figure 5C), *Abca3* vs. *Ager* (Figure 5D), and *Slc34a2* vs. *Aqp5* (Figure 5E).

As the cell markers listed above are known to have distinct cell-type-specific expression, we sought to determine if the mixed signature was the result of carcinogenic processes dysregulating gene expression, aberrant activation of KRAS^{G12D} oncogene in a cell type other than the one expected from Cre recombinase off target effects, or the result of multiple cells being contained within each spatial transcription array spot by performing immunofluorescence staining for AT1 cell marker AQP5 and AT2 cell marker SFTPC on both *Gramd2*:KRAS^{G12D}- and *Sftpc*:KRAS^{G12D}-derived LUAD (Figure 5F). Immunofluorescence staining confirmed that SFTPC and AQP5 were expressed in LUAD derived from both *Sftpc*:KRAS^{G12D} and *Gramd2*:KRAS^{G12D}. Interestingly, the staining patterns for AT1 and AT2 cells varied between LUAD with differing cells of origin, indicating that the LUAD lesions were composed of multiple cell types containing distinct AT2- and AT1-derived cells. Intracellular staining patterns for SFTPC were similar in both *Sftpc*:KRAS^{G12D}- and *Gramd2*:KRAS^{G12D}-derived LUAD, with SFTPC staining surrounding the periphery of cuboidal cells reminiscent of AT2 cell architecture. In contrast, AQP5 staining patterns varied dramatically in the induced adenocarcinomas depending on cell of origin. In *Sftpc*:KRAS^{G12D} tumors, AQP5 staining was long and planar, reminiscent of normal AT1 cell architecture and occurred at the boundaries of the SFTPC^{+} hyperproliferative regions. In contrast, AQP5 staining in *Gramd2*:KRAS^{G12D}-derived LUAD was markedly rounded and dispersed throughout the tumor. These results suggest that co-occurrence of AT2 and AT1 cell markers within array spots was due to different cell types present within each spot. Our results also suggest that the contribution to differential histology observed between *Sftpc*:KRAS^{G12D}- and *Gramd2*:KRAS^{G12D}-derived LUAD may be in part aided by the disruption and redistribution of AT1 cell architecture.

AT2- and AT1-derived LUAD both exhibit KRT8^{+} intermediate cell states

Keratin 8 (KRT8^{+}) intermediate cell states have described roles in multiple lung diseases and have been implicated in cellular response to stress.[11,43] To determine if KRT8^{+} intermediate cell states may form during LUAD tumorigenesis, and if there were differences between *Sftpc*:KRAS^{G12D} and *Gramd2*:K-RAS^{G12D}, we examined the distribution of *Krt8* expression across spatial transcriptomic profiles (Figure 6A). While LUAD of BIA histology (IC14) had one of the highest levels of *Krt8* expression (Figure 6B), elevated *Krt8* was observed within LA-associated IC17 and PA-associated IC14, as indicated by asterisks. *Krt8* was elevated in both cells of origin within all LUAD-associated ICs (Figure 6C). We therefore sought to investigate the relationship between KRT8 and histologic presentation in *Gramd2*- and *Sftpc*-driven LUAD.

To confirm our transcriptomic analysis results that *Krt8* was elevated in both *Gramd2*- and *Sftpc*-driven LUAD, we evaluated KRT8 staining patterns in the $KRAS^{G12D}$-initiated mouse tumors by performing immunofluorescence on both *Gramd2*:$KRAS^{G12D}$ and *Sftpc*:$KRAS^{G12D}$ tumors. Staining patterns in control C57BL/6J mice were consistent with previous literature[11] (Figure S10), in that the alveolar epithelium expressed little to no $KRT8^{+}$ staining under normal physiologic conditions (Figure 6D). Also consistent with previous literature was the presence of $KRT8^{+}$ cells within the bronchial tree (Figure S11). We then examined how the distribution of KRT8 varied between *Sftpc*:$KRAS^{G12D}$-derived lepidic LUAD and *Gramd2*:$KRAS^{G12D}$-derived lepidic LUAD and observed that KRT8 staining was elevated in both tumor types and surrounded tumor nuclei (Figures 6E and 6F).

We then sought to determine if the tumor nuclei surrounded by $KRT8^{+}$ signal showed cell-type-specific marker expression consistent with the cell in which $KRAS^{G12D}$ had been activated. Co-staining of KRT8 and SFTPC in *Sftpc*:$KRAS^{G12D}$ lepidic LUAD showed that tumor nuclei were surrounded by SFTPC expression, as anticipated (Figure 6E). Unfortunately, the *GRAMD2* antibody was incompatible with KRT8, therefore AQP5 was used as a surrogate AT1 cell marker to interrogate AT1 cell distribution in *Sftpc*:$KRAS^{G12D}$ lepidic LUAD. AQP5 was also present in *Sftpc*:$KRAS^{G12D}$ LUAD as observed previously (Figure 5), and costained with KRT8 (Figure 6F). *Gramd2*:$KRAS^{G12D}$ papillary LUAD also showed distinct overlap between AQP5 and KRT8 staining. Bronchiolar infiltrative LUAD observed in *Gramd2*:$KRAS^{G12D}$ mice also showed distinct expression of KRT8 surrounding tumor nuclei, which were also positive for SFTPC (Figures 6G; Figure S3). Our results indicate that KRT8 up-regulation was present in LUAD tumors from both alveolar epithelial cells of origin and may reflect the carcinogenic stress response in distal lung.

KRT8 expression levels are associated with poor outcomes in human LUAD.[47] Expression of *KRT8* was elevated in human LUAD from TCGA-LUAD dataset (Figure 6H), and we also observed that elevated *KRT8* expression was associated with poorer overall survival in human LUAD (Figure 6I) in Kaplan-Meier plot (KMplot).[44,45] This was intriguing, as *KRT8* has been described as a transitional cell state during the maladaptive repair phase following fibrosis and may play a role in the carcinogenic process. We then evaluated whether *KRT8* levels are induced during differentiation of purified human AT2 cells in 2D culture into AT1-like cells, as previously described.[46] We observed that *KRT8* was elevated during human AEC differentiation (Figure 6J), further supporting that tumors may arise from a $KRT8^{+}$ intermediate AEC cell stress state in both AT2 and AT1 cells.

In sum, we observed KRT8 expression in lesions derived from both models. We therefore hypothesize that either AT1 or AT2 cells, when exposed to stress conditions, such as activation of $KRAS^{G12D}$, would adopt a defensive transitional cell state to attempt to resolve the injury. This $KRT8^{+}$ cell state has been previously observed following bleomycin injury to resolve into reformation of the normal alveolar epithelium,[11] and is unable to do so when accumulating mutations that give rise to cancer.

SCGB3A2 is associated specifically with AT1-derived LUAD

To further characterize the intermediate cell state observed in *Gramd2*:$KRAS^{G12D}$-initiated LUAD, we examined the profile of an additional marker of lung epithelial transitional

cell states, secretoglobin 3a2 (SCGB3A2), which has been shown to generate multiple distal lung lineages in response to cellular damage and/or stress.[48,49] SCGB3A2 was also associated with KRT8$^+$ expression in response to alveolar stress and injury resolution.[11] *Scgb3a2* was elevated in LUAD-associated ICs (Figure 7A), but the expression of *Scgb3a2* was elevated in *Gramd2*:KRAS^{G12D} relative to expression in *Sftpc*:KRAS^{G12D} LUAD array spots. While *Scgb3a2* was elevated in BIA, it was also elevated in the other LUAD-associated ICs (Figure 7B), indicating that LUAD may pass through a *Scgb3a2*$^+$ cellular state during carcinogenesis.

Staining for SCGB3A2 (red) alongside KRT8 (green) cell markers was negative in alveolar tissue of WT (Figure 7C) and control *Sftpc*-CreERT2 and *Gramd2*-CreERT2 controls (Figure 7D). In contrast, SCGB3A2 staining overlapped with KRT8 staining patterns specifically in *Gramd2*:KRAS^{G12D} LUAD lesions but was mutually exclusive in *Sftpc*:KRAS^{G12D} lesions (Figure 7E). This was confirmed with higher-magnification imaging, showing distinct pattens of SCGB3A2 mutual exclusivity with KRT8 in *Sftpc*:KRAS^{G12D} and overlap in *Gramd2*:KRAS^{G12D} LUAD lesions, suggesting AT1 cells are capable of transitioning through SCGB3A2$^+$ intermediate cell states while AT2 cells cannot. This in turn suggests that transitional cell states that occur during LUAD oncogenesis may vary between cells of origin.

DISCUSSION

Results of this study suggest that *Gramd2*$^+$ AT1 cells can give rise to histologically verified and transcriptomically distinct LUAD, and therefore that *Gramd2*$^+$ AT1 cells can serve as a cell origin of LUAD. Furthermore, different cells of origin can give rise to LUAD with different histologic patterns. *Gramd2*$^+$ AT1 cells gave rise primarily to papillary adenocarcinoma, whereas *Sftpc*$^+$ AT2 cells gave rise solely to lepidic adenocarcinoma. Lepidic adenocarcinoma is known to have improved overall survival outcome compared with all other types of LUAD (~92% 5-year survival rate), and PA also has improved outcomes (~70% 5-year) compared with acinar and solid LUAD histologies (<15% 5-year). Our results suggest that other mature epithelial cell types resident in the distal lung may also contribute to LUAD that manifests in different histologic patterns not observed in our model systems. It has been previously suggested that the BASC resident at the bronchoalveolar junction (BADJ) can also give rise to LUAD,[20] although a thorough characterization of histologic subtypes has to our knowledge not yet been performed. It is also possible that the distal respiratory bronchiolar epithelium, which is present in humans but not in mice, may harbor unique cell populations that can give rise to LUAD.

Gramd2 was initially identified as an AT1 cell marker in 2-dimensional cultures of AT2 cells differentiating into AT1-like cells whose expression was conserved across human, mouse, and rat.[14] This initial study saw induction of *GRAMD2* after 4 days of differentiation *in vitro*, when other known AT1 cell markers were also expressed, such as AQP5, PDPN, and AGER (among others). However, by day 4 in culture there was an obvious induction of the intermediate cell marker KRT8 in this cell population, which could indicate that *GRAMD2*, as has been observed for HOPX,[50] is also induced in the AEC-intermediate cell state, which is present at low levels in the adult lung,[50] as well as in the fully differentiated

"mature" AT1 cell. *Igfbp2* is currently considered the most lineage-restricted marker for mature AT1 cells,[51] therefore it would be interesting to see if KRAS^{G12D} activation within the *Igfbp2*-restricted AT1 cell population can give rise to papillary LUAD as well.

Our results also indicate that AT1 and AT2 cell lineages may pass through a KRT8 intermediate cell state during carcinogenesis. *Krt8*$^{+}$ cells are stress-induced intermediate cells that can arise from AT2 and the more proximally located club cells that subsequently differentiate into AT1 cells and repopulate the lung epithelial lining.[11] To our knowledge, this study is the first demonstration that LUAD may progress through this intermediate state. However, unlike bleomycin-induced fibrosis, the cells do not appear able to resolve the initial stressor and maintain the hallmarks of this intermediate population. *Gramd2*$^{+}$ AT1 derived LUAD colocalizes expression between KRT8 and SCGB3A2, whereas SFTPC^{+} AT2- derived LUAD did not. Elevated expression of KRT8 is known to predict poor patient outcome for LUAD and is associated with epithelial-to-mesenchymal transition,[47] further implicating this intermediate transitional cell state in the etiology of LUAD. Although this study established a role for KRT8 and SCGB3A2 in LUAD, we did not longitudinally track LUAD evolution from cell of origin, and therefore results presented on intermediate cell states are inferred. Therefore, the precise connection between alveolar cell origin, cellular plasticity, and intermediate cell states during tumorigenesis will require further investigation. Lineage tracing studies could be used to answer the timing and trajectory of intermediate cell states contributing to *Gramd2*$^{+}$ AT1-LUAD.

Another unexpected finding in the *Gramd2*$^{+}$ AT1 derived LUAD in mouse model was the development of bronchiolar infiltrative adenocarcinoma (BIA). This has been observed previously in *Gprc5a*$^{+}$ cell-specific induction of EGFR oncogenic activity[22] and has largely been characterized as a species-specific phenomenon, as this does not occur within human LUAD. These species-specific effects may be due to the lack of distal respiratory epithelium and the interactions between cells of origin and these regions. Additionally, in human LUAD, KRAS^{G12D} mutations are significantly associated with concurrent loss of TP53 function (TP53-LOH). In transgenic mouse models combining KRAS mutations with TP53-LOH increases the multiplicity, size, and degree of dedifferentiation observed in tumors.[24] It is possible that a combination of KRAS^{G12D} with TP53 LOH in *Gramd2*$^{+}$ AT1 cells may result in higher-grade tumors, reflecting known histologic patterns observed in human LUAD with mixed histology.

Based on our spatial transcriptomic analysis, the BIA phenotype was derived from a ciliated cell population, which was histologically, spatially, and transcriptomically distinct from the alveolar papillary lesions observed in the same *Gramd2*$^{+}$-AT1-derived LUAD model. Interestingly, *Sftpc*:KRAS^{G12D} also presented with lesions of BIA histology in our transcriptomic analysis, but to a lesser extent, suggesting that *Sftpc*-CreERT2 may be activated in a similar subset of ciliated cells. Our micro-CT scanning results demonstrated that nodule formation in both *Sftpc*:KRAS^{G12D} and *Gramd2*:KRAS^{G12D} was restricted to alveolar regions and equidistant from terminal bronchi. This would argue that either micro-CT was not able to detect the BIA phenotype as it did not grow as distinct nodules, or that any nodules formed from BIA were equidistant from the terminal bronchi and derived from a population of ciliated cells in near proximity to the alveolar space. In addition,

our spatial transcriptomic analysis shows that BIA was also highly enriched for cilia formation, axoneme assembly, and other pathways with known involvement in multiciliated cell function. Taken together, it appears that the BIA lesions, derived from ciliated epithelial cells and not consistent with any known histology in human LUAD, do not serve as a cell of origin in human LUAD and may be a source of the dissention in the field as to whether mouse models accurately reflect LUAD histology of human lesions.

In summary, our results demonstrate that AT1 cells can serve as a cell of origin for LUAD. Additionally, our results indicate that different cells originating in the distal lung give rise to LUAD of different histologic and molecular phenotypes. This work is consistent with recent observations that *Hopx*$^{+}$ AT1 cells have proliferative capacity.[15] Importantly, the same mutation, KRAS^{G12D}, caused different cellular signaling disruptions depending on the cell of origin in which the oncogene was activated, which could lead to the diverse phenotypic presentations observed. Our work suggests that LUAD is not one single cancer type that arises from AT2 and, to a lesser extent club cells, but instead is a collection of cancers that occur within spatial proximity in the distal lung, each with distinct morphologic, phenotypic, and transcriptomic differences that derive from the cell of origin of LUAD.

Limitations of the study

This study is performed largely on mouse lung samples; therefore, it remains to be determined if human patient samples exhibit similar findings. In addition, data presented in this study are reflective of fully formed tumors. It remains to be determined how cell of origin plasticity influences characteristics of the resultant tumor. Our findings indicating the AT1 cell population can serve as a cell of origin for LUAD are based on the *Gramd2*-CreERT2 mouse model; however, another other Cre driver model, *Hopx*-CreER, has recently been shown to induce AT1-derived LUAD, but with a different histologic presentation than observed in this study.[52] Further research will be necessary to determine if specificity of individual mouse models is the source of this difference. Our results do not infer applications in patient treatment, including clinical diagnosis, prognosis, and intervention.

STAR★METHODS

RESOURCE AVAILABILITY

Lead contact—Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact: Crystal N. Marconett (cmarcone@usc.edu). Institutional and funding agency requirements for resource and reagent sharing will be followed.

Materials availability—All materials are commercially available. *Gramd2*-CreERT2 mice are available by Material Transfer Agreement from Dr. Zea Borok, MD (zborok@health.ucsd.edu) with a standard Materials Transfer Agreement.

Data and code availability

- Spatial transcriptomic RNA-sequencing data were uploaded to NCBI Gene Expression Omnibus (GEO: GSE215858).

- TCGA data are available for download through the Gene (GDC) portal (https://portal.gdc.cancer.gov/).
- Code used in R-based analyses is available as Data S1.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

All animal protocols are reviewed semiannually and annually and monitored continuously by an Institutional Animal Care and Use Committee (IACUC). All mice procedures were performed in accordance with protocols approved by the University of Southern California (USC) IACUC and followed the National Institutes of Health (NIH) Guide for the Care and Use of Laboratory Animals. All mice were housed in groups of ≤5 animals per cage in the USC's Animal Resources Center and maintained in accordance with the US National Institutes of Health Guide for Care and Use of Laboratory Animals. The facility has received full AAALAC accreditation. The facility was kept under regular lighting conditions (12 h light/dark cycle) with a regular feeding and cage-cleaning schedule. Room temperature was maintained at 21 ± 2°C. Mice were all maintained on a C57BL/6J genetic background (backcrossed every three generations), and all genotypes were determined by polymerase chain reaction (PCR). 2–4 month old male mice were used for experiments.

Gramd2*-CreERT2 mice—Gramd2*-CreERT2 mice were produced by Applied Stem Cell, Inc. (Milpitas, CA) and bred and maintained on a C57BL/6J background. The details of mouse strain construction are described in method details.

Sftpc*-CreERT2 mice—Sftpc*-CreERT2 mice were a kind gift of Harold Chapman at the University of California, San Francisco.

KRAS-LSL-G12D mice—KRAS-LSL-G12D mice (B6.129S4-Krastm4Tyj/J; Jackson Laboratory strain #:008179) and their wild-type littermates were bred and maintained on a C57BL/6J background.

***Gramd2*:KRAS^{G12D} and Sftpc:KRAS^{G12D} mice—**We generated double transgenic *Sftpc*:KRAS^{G12D} mice by crossing KRAS-LSL-G12D heterozygous mice with *Sftpc*-CreERT2 transgenic mouse line. The *Sftpc*-CreERT2 mouse line has been used previously to track AT2 cell fate.[12] *Gramd2*-CreERT2 mice were crossed to heterozygous KRAS-LSL-G12D mice to generate *Gramd2*:KRAS^{G12D} transgenic line. *Gramd2*-CreERT2 and *Sftpc*-CreERT2 mice were in a B6/129 mixed strain background (129 N1), while KRAS-LSL-G12D mice are in a B6/129S4 mixed strain background (C57BL/6J N10).

METHOD DETAILS

Generation of *Gramd2*:CreERT2 transgenic mouse model—The *Gramd2*-CreERT2 knockin mouse line was produced by Applied Stem Cell, Inc. (Milpitas, CA). A CRISPR/Cas-assisted gene targeting approach was used in mouse embryonic stem (ES) cells to knockin CreERT2 and eGFP into the endogenous *Gramd2* locus. This was accomplished

by using a “traditional” targeting vector, including flanking homologous sequences of *Gramd2* and sequences encoding CreERT2, eGFP and a NeoR positive selection marker in between. NeoR was flanked by FRT sites to enable its removal by Flp/FRT recombination in a later step. The knockin was designed so that the endogenous stop codon of *Gramd2* was disrupted and the last coding triplet in *Gramd2* was immediately followed by an E2A sequence, CreERT2, and a T2A sequence followed by eGFP-Stop-polyA, all in-frame. A nuclear localization signal (NLS) was added to the 3′ end of eGFP. This strategy, using E2A and T2A sequences, should enable expression of separate *GRAMD2*, CreERT2 and nGFP proteins from the knockin allele. Homologous recombination in ES cells (derived from a 129 mouse sub-strain) was assisted by simultaneous electroporation of the targeting vector together with Cas9 and a guide RNA (gRNA) that directed Cas9 to a site immediately upstream of the stop codon of *Gramd2*. ESC clones were screened for the presence of the targeted gene by PCR and ploidy was tested by chromosome count analysis. ESC clones with greater than 65% euploidy qualified for microinjection. Selected clones were transferred into pseudo pregnant mice.

Chimeric males in the C57 BL6J transmitting the knockin allele through the germline were crossed to FLPer (knockin of FLPe variant of the Saccharomyces cerevisiae FLP1 recombinase) females (Jackson Laboratories, Bar Harbor, ME; Stock #016226) to remove the PGK-NeoR selection marker. Neither Western blot nor immunofluorescence staining detected expression of GFP protein, indicating inability to obtain significant co-expression of nGFP with CreERT2, despite detectable GFP mRNA, perhaps due to defective translation. For detection of the lineage label, *Gramd2*:CreERT2-nGFP mice were subsequently crossed to mTmG mice (Jackson Laboratories, Bar Harbor, ME; Stock #007576) also on the 129S6/SvEvTac background. *Gramd2*:CreERT2-nGFP and mTmG double heterozygous mice (*Gramd2*-CreERT2:mTmG) were used for evaluation of cell-specific GFP expression.

Mouse genotyping

All mouse procedures underwent ethical review by the Institutional Animal Care and Use Committee (IACUC) at USC prior to experimentation (IACUC Protocol #20963). Pups were weaned before day 21 after birth. For each pup, 0.1 cm of tail was cut and stored in a 1.5 mL microcentrifuge tube. 100 μL DirectPCR (#102-T, Viagen) with 10 μL proteinase K (50 μg/mL, Qiagen, DNeasy Blood & Tissue Kit, #69504) was then added to each tube and incubated with the tail overnight at 60°C. To inactivate proteinase K, a 2-h incubation at 92°C was performed the next day, and crude DNA was ready for use. A concentration of 50 ng/ul was used as a template for subsequent genotyping PCR reactions. PCR was performed using GoTaq G2 Hot Start Polymerase (Promega, Lot: 0000362382). For *Gramd2* gene, reactions consisted of 1 μL DNA, 5 μL buffer, 2 μL MgCl2, 0.5 μL dNTP, 1 μL common forward primer, mutant and wild type reverse primer, 0.12 μL polymerase, and nuclease-free water to a total volume of 20 μL. For the *KRAS* gene, the PCR recipe is similar except using 3 μL of DNA template. For *Sftpc* gene, the PCR recipe is similar except for two sets of primers: one for the mutant *Sftpc* allele and one for the wild type *Sftpc* allele. For *Gramd2* and *KRAS* genes, the reaction was incubated for 2 min at 94°C, 35 cycles, including 20 s at 94°C, 30 s at 58°C and 90 s at 72°C, and 10 min at 72°C using the C1000 Touch Thermal

Cycler (Bio-Rad, Los Angeles, CA). For *Sftpc* gene, the reaction was incubated for 3 min at 94°C, 35 cycles, including 30 s at 94°C, 30 s at 58°C and 40 s at 72°C, and 2 min at 72°C using the C1000 Touch Thermal Cycler (Bio-Rad, Los Angeles, CA). Primer sequences are listed in the Supplemental Methods.

Gel electrophoresis

PCR products were loaded into a 2.0% agarose gel, which was prepared using LE Agarose (VWR, LIFE SCIENCE, CAS: 9012-36-6) and nuclease-free water. The PowerPac Basic Power Supply (BIO-RAD, PowerPac Basic) was used to perform gel electrophoresis at 80V for 1 h. Gels were imaged on a Molecular Imager ChemiDox XRS+ (BIO-RAD, Los Angeles, CA).

Tamoxifen administration

Tamoxifen (Sigma-Aldrich, CAS: 10540-29-1) powder was dissolved in corn oil (Sigma-Aldrich, CAS: 8001-30-7) to a final concentration of 50 mg/mL. To ensure that tamoxifen was fully dissolved in corn oil, the solid-liquid mixture was gently shaken and incubated in a shaking incubator (SHEL LAB, model: SSI5R; cat: #33–804R) at 65°C for 2 h. Six-week-old mice were injected with tamoxifen via intraperitoneal injection. For *Gramd2*:KRAS^{G12D}, *Gramd2*:CreERT2, KRAS-LSL-G12D and WT mice strains, a stock concentration of 50 mg/kg tamoxifen with a final dosage of 200 mg/kg per mouse was introduced via intraperitoneal (IP) injection three times on alternating days. For *Sftpc*:KRAS^{G12D} and *Sftpc*:CreERT2 mice strains, the working concentration of 25 mg/mL tamoxifen was administered by IP injection twice on alternating days to a final dosage of 100 mg/kg.

Dissection and processing of lung samples

At 14 weeks post-tamoxifen injection, mice were euthanized by injecting 100 μL Euthasol (ANADA #200–071, Virbac) via intraperitoneal injection. The lungs were then perfused with phosphate buffered saline (PBS, 21–031-CV, CORNING) and inflated with 4% paraformaldehyde (PFA, CAS: 30520-89-4, Sigma-Aldrich) at a pressure of 25 cm water. After fixation, the lungs were transferred into a 50 mL Falcon tube and fixed in 4% PFA overnight. Fixed lungs were then washed three times with 15 mL sterile PBS and stored in 70% ethanol for future use.

Embedding and sectioning

Embedding and sectioning were accomplished with assistance of the USC Translational Research Core in the USC School of Pharmacy. Briefly, lung samples were processed in an automatic tissue processor (Thermo Scientific Spin Tissue Processor Microm STP 120) following a standard gradient of dehydration (70%, 80%, 95% and 100% of ethanol), clearing (Clear-Rite 3) and paraffin infiltration. After embedding, samples were sectioned at 5 μm using a rotary microtome (Thermo Fisher Microm HM310 Rotary Microtome) and affixed to clean slides.

Silicification & micro-CT scanning

After fixation with 4% PFA using insufflation, the lungs of *Gramd2*:KRAS^{G12D} and *Sftpc*:KRAS^{G12D} mice were gently rinsed with PBS, and then with 50% ethanol, before placing in 70% ethanol. This rinse was repeated three times. Next, lung samples were dehydrated using an ethanol gradient at room temperature, in 70%, 80%, and 90% ethanol solution each for 2 h, and then in 100% ethanol, and left overnight. After lung samples were dehydrated using ethanol, we used the low-surface-tension solvent hexamethyldisilazane (HMDS) (SHBG4111V, Sigma-Aldrich) to incubate samples. For infiltration, the lung samples were left in the chemical hood for 1–2 h until they were thoroughly dried and solidified as previously described[54].

The dried and solidified lungs underwent micro-CT scanning at 6–8 μm voxel resolution at the USC Molecular Imaging Center in the Department of Radiology at Keck School of Medicine, USC using a Phoenix nanotom M micro-CT scanner system. The scans were performed using the following parameters: 60kVp, 200 μA using 1440 projections along a 360-degree rotation at one frame per second rate. Raw image data were reconstructed into 16-bit DICOM images. Visualization of X-ray image data, 3D surface renderings of lesions inside of lungs, and quantification of lesion volumes and distances between the lesions and terminal bronchioles were performed using VGSTUDIO MAX 3.3.2.170119 64 bit (© Copyright 1997–2019 by Volume Graphics GmbH).

The micro-CT images and 3D surface renderings were used to segment the multifocal lesions throughout the lung, calculate each lesion's mean and median volume (mm^3), and collect the x, y, and z coordinates of the center point for each lesion. We also collected the x, y, and z coordinates of landmarks placed in each terminal bronchiole just before the alveolar ducts. Then all possible distances between each lesion and landmark of each terminal bronchiole were calculated, and minimal distance was chosen as the distance between the lesion and terminal bronchiole. This calculation was repeated until we found all possible minimal distances between the lesion and terminal bronchiole. One lesion (<0.01%) in *Sftpc*:KRAS^{G12D} was several standard deviations outside the rest of the data cohort and excluded from additional analyses.

Kaplan-Meier survival analysis

Mice (Mus musculus) was selected as the model animal to generate lung adenocarcinomas which served as samples for survival analysis studies. The KRAS-LSL-G12D gene was activated in the mice (*Gramd2*:KRAS^{G12D} and *Sftpc*:KRAS^{G12D}) via tamoxifen injection, and mice in the same genotype that serve as vehicle controls were injected with corn oil. After sample size calculation (under p value of 5% (p = 0.05) and a power of 80%), at least 20 mice in each oncogenic group and 3 males and 3 females in vehicle control group were used in this analysis. The health status of all mice post injection was evaluated daily according to PASTER Animal Assessment Metric.[55] Euthanasia was performed and recorded immediately when the mice presented with critical illness that is below the threshold indicated in PASTER Metric. According to the survival duration, the Kaplan-Meier plot is depicted by using KM plotter (KMplot.com[44,45]) and the p

value evaluating significance of survival difference is calculated by using DATAtab Online Statistics Calculator.

H&E staining

H&E staining was accomplished with assistance of the USC Immunohistochemistry laboratory, Department of Pathology, Keck School of Medicine, USC. The slides were deparaffinized and stained in Hematoxylin and Eosin using an automated stainer (Varistain Gemini ES Automated Slide Stainer). Samples underwent blinding prior to pathology review.

Immunohistochemistry (IHC) staining

IHC staining was accomplished with assistance of the USC Immunohistochemistry laboratory in the Department of Pathology at the Keck School of Medicine, USC. First, sections were baked at 60°C for 1–2 h and then left to cool for 15 min. After cooling, the following steps were processed in the BOND-III Fully Automated IHC/ISH Staining System (Leica Biosystems). Sections were first deparaffinized and then underwent antigen retrieval by incubating with EDTA-based epitope retrieval solution (BOND Epitope Retrieval Solution 2, Cas #AR9640, Leica Biosystems) for 20 min. After antigen retrieval, slides were washed with distilled water 3 times for 2 min each. Next, processed sections were incubated with Multi-Cytokeratin (AE1/AE3) (10 mg/mL, CAS #PA0909, Leica-biosystems), Anti-CD68 Monoclonal Antibody, Unconjugated, Clone 514H12 (37 mg/L, CAS #PA0286, Leica-biosystems), Anti-GATA-3 Monoclonal Antibody, Clone L50–823 (1:20, CAS #CM405B, BIOCARE), Anti-Napsin A Monoclonal Antibody, Clone TMU-Ad02 (1:100, CAS #CM388A, BIOCARE), Anti-thyroglobulin Monoclonal Antibody, Clone 2H11 + 6E1 (1:250, CAS #340M-15, Cell Marque), and TTF-1 (1:500, Cell Marque, Cas #343M-96) Antibody for 15 min. After primary antibody incubation sections were washed with 0.2% Tween diluted in PBS for 2 min for three times at 25°C. Next, sections were incubated with post primary block and polymer (BOND IHC Polymer Detection Kit (DS9800), Leica Biosystems). Each incubation lasted for 8 min. After that, sections were blocked with peroxide for 5 min, then staining was performed with DAB Chromogen (BOND IHC Polymer Detection Kit (DS9800), Leica Biosystems) for 10 min. In addition, sections were stained with Mayer's Hematoxylin (American Master Tech). The processed sections were then taken out of the Leica Bond-III Auto-stainer and were dehydrated by dipping in 95% isopropyl alcohol for 1 min, 100% isopropyl alcohol for 1 min, and xylene for 1 min. Dehydrated sections were mounted and covered with glass coverslips.

Immunofluorescence (IF) staining

Lungs were inflation-fixed with 4% paraformaldehyde (PFA) in PBS at 25 cm H2O pressure at 4°C overnight. Lungs were isolated, dehydrated in ethanol and embedded in paraffin. After de-paraffinization was performed on FFPE lung section slides, high pH (9.0) antigen retrieval was performed using antigen unmasking solution (#H3301, Vector Laboratories, Burlingame, CA). Sections (5 μm) were washed and permeabilized with PBS containing 0.1% Triton X-100. Sections were blocked with CAS block (#008120, Life Technologies) at room temperature for 30 min to 1 h, then incubated with primary antibodies at 4°C overnight. Primary antibodies were as follows: GFP (1:200–500, #13970, Abcam,

Eugene, OR #A6455, Invitrogen), AQP5 (1:200, #Aqp-005, Alomone Labs, Limerick, PA), Polyclonal Anti-*GRAMD2* Antibody (1:100, #HPA029435, Atlas Antibodies), proSFTPC (aka proSPC,1:200–500, #WRAB-SPC-9337, Seven Hills, Cincinnati, OH), Rat-anti-Keratin, type II/Cytokeratin 8 (KRT8) antibody (1:100, DSHB, CAS #ab531826), Mouse UGRP1/SCGB3A2 Antibody (1:300, #: AF3465, R&D Systems), and FOXJ1 Monoclonal Antibody (1:300, # 14-9965-82, Thermo Fisher). Mouse monoclonal FOXJ1 antibody was biotinylated by using Biotinylation Kit (ab201795, Abcam), and antibody signal of KRT8 was amplified using biotinylated anti-rat antibody (1:300, 0.5 mg, FisherScientific Vector Laboratories, CAS #NC9016344), followed by staining with Streptavidin, Alexa Fluor 488 conjugate (1: 300, Invitrogen, CAS #S11223). Antibody signal of AQP5, SFTPC, *GRAMD2* or SCGB3A2 was visualized by staining with IgG (H + L) Highly Cross-Adsorbed Goat anti-Rabbit, Alexa Fluor 647 (1:300, ThermoFisher, CAS #A-21245). Slides were mounted with media containing 4′,6-diamidino-2-phenylindole (DAPI) or propidium idodide (PI). Confocal (1 μm optical sections, Leica TCS SP8 confocal) images were captured at the same settings for all sections within an experimental group.

Visium 10X spatial transcriptomic profiling

Spatial transcriptomic profiling was performed at the Molecular Genomics Core, part of the Norris Comprehensive Cancer Center, using the manufacturer's instructions. Briefly, two biological replicates (one block for each) of *Gramd2*:KRAS^{G12D} and *Sftpc*:KRAS^{G12D} underwent initial 10 μM sectioning at the Tissue Pathology Core, part of the Norris Comprehensive Cancer Center. Sections were H&E stained (as above) and underwent pathology review. Regions of Interest (ROIs) then underwent sample preparation including test slide sample sequencing using the Visium Tissue Section Test Slide (PN-2000460, 10X Genomics, Dublin, CA, USA). RNA quality assessment of test slides was determined by first extracting RNA using the Qiagen RNeasy FFPE kit (#73504, Qiagen, Hilden, Germany) and measuring RNA concentration using a Qubit fluorometer (#Q33238, ThermoFisher Scientific, Waltham, MA, USA). The percentage of total fragments >200 nt in length was determined by running the samples using a 4200 Tapestation (#G2991AA, Agilent, Santa Clara, CA). Once samples were determined to be of high enough quality to continue, 5 μM FFPE sections were placed in a 42°C water bath and once rehydrated, adhered to the Visium Spatial Gene Expression Slide (PN-2000233, 10X Genomics). Subsequently, samples were dried at 42°C for 3 h in a desiccation chamber and underwent deparaffinization using Qiagen Deparaffinization Solution (Cat # 19093) at 60°C for 2 h. Subsequently H&E staining (#MHS16, #HT110116 Millipore Sigma, Burlington MA, USA) was performed according to Visium technical parameters (CG000409) and images were captured with a Zeiss Axioscan2 microscope using a 103 objective. Decrosslinking was performed according to 10X standard protocol (CG000407) and immediately hybridized to the Visium Mouse Transcriptome probe set V1.0, which contained 20,551 genes targeted by 20,873 probes. Post-probe extension, sequencing library construction was performed using unique sample indices using the Dual Index Kit TS, Set A (PN-1000251) for Illumina-compatible paired-end sequencing (2×100). Raw sequencing data were processed with the Space Ranger pipeline (10x Genomics) spaceranger mkfastq for demultiplexing of sequencing data. Spaceranger performed count alignment, unique molecular identifier counting, tissue detection, and alignment.

QUANTIFICATION AND STATISTICAL ANALYSIS

Sample sizes, statistical tests used, and p values for each figure and panel are indicated in the figure legends. Initial tumor identification studies were performed on a cohort of 5 mice per genotype. Follow-up studies were performed on groups with ≥3 mice per genotype. Statistical comparisons between two groups were made using a Student's two-tailed t test. Comparisons between multiple groups were performed using ANOVA with Bonferroni multiple testing correction. Three biological replicates of *Gramd2*:KRAS^{G12D} and *Sftpc*:KRAS^{G12D} underwent spatial transcriptomic profiling. Differential expression analysis in spatial data were made using Seurat with the FindMarkers functionality.[53]

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

The authors would like to thank Junji Watanabe, PhD, and the Translational Research Lab in the School of Pharmacy for assistance with tissue embedding and sectioning; Seth Ruffins, PhD, and the Optical Imaging Core at the Broad California Institute for Regenerative Medicine (CIRM) for assistance with immunofluorescence; Peter Conti, MD, PhD, and the Molecular Imaging Center for assistance with micro-CT; and David Wesley Craig, PhD, and the Molecular Genomics Core for spatial transcriptomic sequencing. Core usage at the Norris Comprehensive Cancer Center is supported by the National Cancer Institute (P30 CA014089). Funding for this study was provided by the National Cancer Institute (NCI R01 CA262258) to C.N.M., American Cancer Society (RSG-20-135-01) to C.N.M., R35 HL135747 to Z.B., the Wright Foundation, and the Departments of Surgery and of Translational Genomics, Keck School of Medicine, USC. Training activities for coauthors from diverse backgrounds were supported in part by the NCI-sponsored CaRE2 Florida-California Health Equity Center (U54 CA23396, U54 CA233444, and U54 CA233465).

INCLUSION AND DIVERSITY

One or more of the authors of this paper self-identifies as an underrepresented ethnic minority in science. One or more of the authors of this paper self-identifies as a member of the LGBTQIA+ community. One or more of the authors of this paper received support from a program designed to increase minority representation in science.

REFERENCES

1. Siegel RL, Miller KD, Wagle NS, and Jemal A (2023). Cancer statistics. CA A Cancer J. Clin 73, 17–48.
2. Ferone G, Lee MC, Sage J, and Berns A (2020). Cells of origin of lung cancers: lessons from mouse studies. Genes Dev. 34, 1017–1032. [PubMed: 32747478]
3. Tang ER, Schreiner AM, and Pua BB (2014). Advances in lung adenocarcinoma classification: a summary of the new international multidisciplinary classification system (IASLC/ATS/ERS). J. Thorac. Dis 6 (Suppl 5), S489–S501. [PubMed: 25349701]
4. Chen Z, Liu X, Zhao J, Yang H, and Teng X (2014). Correlation of EGFR mutation and histological subtype according to the IASLC/ATS/ERS classification of lung adenocarcinoma. Int. J. Clin. Exp. Pathol 7, 8039–8045. [PubMed: 25550848]
5. Shen M, Qi R, Ren J, Lv D, and Yang H (2021). Characterization with KRAS mutant is a critical determinant in immunotherapy and other multiple therapies for non-small cell lung cancer. Front. Oncol 11, 780655. [PubMed: 35070984]
6. Weibel ER (1971). The mystery of "non-nucleated plates" in the alveolar epithelium of the lung explained. Acta Anat. 78, 425–443. [PubMed: 4995841]

7. Weibel ER (2015). On the tricks alveolar epithelial cells play to make a good lung. Am. J. Respir. Crit. Care Med 191, 504–513. [PubMed: 25723823]
8. Brody JS, and Williams MC (1992). Pulmonary alveolar epithelial cell differentiation. Annu. Rev. Physiol 54, 351–371. [PubMed: 1562179]
9. Williams MC (2003). Alveolar type I cells: molecular phenotype and development. Annu. Rev. Physiol 65, 669–695. [PubMed: 12428023]
10. Aspal M, and Zemans RL (2020). Mechanisms of ATII-to-ATI cell differentiation during lung regeneration. Int. J. Mol. Sci 21, 3188. [PubMed: 32366033]
11. Strunz M, Simon LM, Ansari M, Kathiriya JJ, Angelidis I, Mayr CH, Tsidiridis G, Lange M, Mattner LF, Yee M, et al. (2020). Alveolar regeneration through a Krt8+ transitional stem cell state that persists in human lung fibrosis. Nat. Commun 11, 3559. [PubMed: 32678092]
12. Xu X, Rock JR, Lu Y, Futtner C, Schwab B, Guinney J, Hogan BLM, and Onaitis MW (2012). Evidence for type II cells as cells of origin of K-Ras-induced distal lung adenocarcinoma. Proc. Natl. Acad. Sci. USA 109, 4910–4915. [PubMed: 22411819]
13. Wang Z, Li Z, Zhou K, Wang C, Jiang L, Zhang L, Yang Y, Luo W, Qiao W, Wang G, et al. (2021). Deciphering cell lineage specification of human lung adenocarcinoma with single-cell RNA sequencing. Nat. Commun 12, 6500. [PubMed: 34764257]
14. Marconett CN, Zhou B, Sunohara M, Pouldar TM, Wang H, Liu Y, Rieger ME, Tran E, Flodby P, Siegmund KD, et al. (2017). Cross-species transcriptome profiling identifies new alveolar epithelial type i cell-specific genes. Am. J. Respir. Cell Mol. Biol 56, 310–321. [PubMed: 27749084]
15. Jain R, Barkauskas CE, Takeda N, Bowie EJ, Aghajanian H, Wang Q, Padmanabhan A, Manderfield LJ, Gupta M, Li D, et al. (2015). Plasticity of Hopx(+) type I alveolar cells to regenerate type II cells in the lung. Nat. Commun 6, 6727. [PubMed: 25865356]
16. Adams TS, Schupp JC, Poli S, Ayaub EA, Neumark N, Ahangari F, Chu SG, Raby BA, DeIuliis G, Januszyk M, et al. (2020). Single-cell RNA-seq reveals ectopic and aberrant lung-resident cell populations in idiopathic pulmonary fibrosis. Sci. Adv 6, eaba1983. [PubMed: 32832599]
17. Travaglini KJ, Nabhan AN, Penland L, Sinha R, Gillich A, Sit RV, Chang S, Conley SD, Mori Y, Seita J, et al. (2020). A molecular cell atlas of the human lung from single-cell RNA sequencing. Nature 587, 619–625. [PubMed: 33208946]
18. Rock JR, Barkauskas CE, Cronce MJ, Xue Y, Harris JR, Liang J, Noble PW, and Hogan BLM (2011). Multiple stromal populations contribute to pulmonary fibrosis without evidence for epithelial to mesenchymal transition. Proc. Natl. Acad. Sci. USA 108, E1475–E1483. [PubMed: 22123957]
19. Yin H, Jing B, Xu D, Guo W, Sun B, Zhang J, Liao Y, Song H, Wang T, Liu S, et al. (2022). Identification of active bronchioalveolar stem cells as the cell of origin in lung adenocarcinoma. Cancer Res. 82, 1025–1037. [PubMed: 35045987]
20. Kim CFB, Jackson EL, Woolfenden AE, Lawrence S, Babar I, Vogel S, Crowley D, Bronson RT, and Jacks T (2005). Identification of bronchioalveolar stem cells in normal lung and lung cancer. Cell 121, 823–835. [PubMed: 15960971]
21. Kwon MC, and Berns A (2013). Mouse models for lung cancer. Mol. Oncol 7, 165–177. [PubMed: 23481268]
22. Fujimoto J, Nunomura-Nakamura S, Liu Y, Lang W, McDowell T, Jakubek Y, Ezzeddine D, Kapere Ochieng J, Petersen J, Davies G, et al. (2017). Development of Kras mutant lung adenocarcinoma in mice with knockout of the airway lineage-specific gene Gprc5a. Int. J. Cancer 141, 1589–1599. [PubMed: 28653505]
23. Laughney AM, Hu J, Campbell NR, Bakhoum SF, Setty M, Lavallée VP, Xie Y, Masilionis I, Carr AJ, Kottapalli S, et al. (2020). Regenerative lineages and immune-mediated pruning in lung cancer metastasis. Nat. Med 26, 259–269. [PubMed: 32042191]
24. Sutherland KD, Song JY, Kwon MC, Proost N, Zevenhoven J, and Berns A (2014). Multiple cells-of-origin of mutant K-Ras-induced mouse lung adenocarcinoma. Proc. Natl. Acad. Sci. USA 111, 4952–4957. [PubMed: 24586047]

25. Shen H, Flodby P, Ji Y, Liu Y, Castaldi A, Zhou B, et al. (2021). Novel transgenic mouse strain *Gramd2C*^{reERt2} for investigation of alveolar epithelial type I cell plasticity and phenotype. Am. J. Respir. Crit. Care Med, A1037.
26. Gui YS, Wang L, Tian X, Feng R, Ma A, Cai B, Zhang H, and Xu KF (2012). SPC-Cre-ERT2 transgenic mouse for temporal gene deletion in alveolar epithelial cells. PLoS One 7, e46076. [PubMed: 23049940]
27. Zhou B, Flodby P, Luo J, Castillo DR, Liu Y, Yu FX, McConnell A, Varghese B, Li G, Chimge NO, et al. (2018). Claudin-18-mediated YAP activity regulates lung stem and progenitor cell homeostasis and tumorigenesis. J. Clin. Invest 128, 970–984. [PubMed: 29400695]
28. Travis WD, Brambilla E, Noguchi M, Nicholson AG, Geisinger KR, Yatabe Y, Beer DG, Powell CA, Riely GJ, Van Schil PE, et al. (2011). International association for the study of lung cancer/american thoracic society/european respiratory society international multidisciplinary classification of lung adenocarcinoma. J. Thorac. Oncol 6, 244–285. [PubMed: 21252716]
29. Chalela R, Curull V, Enríquez C, Pijuan L, Bellosillo B, and Gea J (2017). Lung adenocarcinoma: from molecular basis to genome-guided therapy and immunotherapy. J. Thorac. Dis 9, 2142–2158. [PubMed: 28840016]
30. Nguyen TT, Lee HS, Burt BM, Wu J, Zhang J, Amos CI, and Cheng C (2022). A lepidic gene signature predicts patient prognosis and sensitivity to immunotherapy in lung adenocarcinoma. Genome Med. 14, 5. [PubMed: 35016696]
31. Nakamura N, Miyagi E, Murata S.i., Kawaoi A, and Katoh R (2002). Expression of thyroid transcription factor-1 in normal and neoplastic lung tissues. Mod. Pathol 15, 1058–1067. [PubMed: 12379752]
32. Johansson L (2004). Histopathologic classification of lung cancer: Relevance of cytokeratin and TTF-1 immunophenotyping. Ann. Diagn. Pathol 8, 259–267. [PubMed: 15494931]
33. Chernock RD, El-Mofty SK, Becker N, and Lewis JS (2013). Napsin A expression in anaplastic, poorly differentiated, and micropapillary pattern thyroid carcinomas. Am. J. Surg. Pathol 37, 1215–1222. [PubMed: 23681073]
34. Yoshihara K, Shahmoradgoli M, Martínez E, Vegesna R, Kim H, Torres-Garcia W, Treviño V, Shen H, Laird PW, Levine DA, et al. (2013). Inferring tumour purity and stromal and immune cell admixture from expression data. Nat. Commun 4, 2612. [PubMed: 24113773]
35. Little DR, Gerner-Mauro KN, Flodby P, Crandall ED, Borok Z, Akiyama H, Kimura S, Ostrin EJ, and Chen J (2019). Transcriptional control of lung alveolar type 1 cell development and maintenance by NK homeobox 2–1. Proc. Natl. Acad. Sci. USA 116, 20545–20555. [PubMed: 31548395]
36. Flodby P, Kim YH, Beard LL, Gao D, Ji Y, Kage H, Liebler JM, Minoo P, Kim KJ, Borok Z, and Crandall ED (2016). Knockout mice reveal a major role for alveolar epithelial type i cells in alveolar fluid clearance. Am. J. Respir. Cell Mol. Biol 55, 395–406. [PubMed: 27064541]
37. Hua K, and Ferland RJ (2018). Primary cilia proteins: ciliary and extraciliary sites and functions. Cell. Mol. Life Sci 75, 1521–1540. [PubMed: 29305615]
38. Thomas PD, Ebert D, Muruganujan A, Mushayahama T, Albou LP, and Mi H (2022). PANTHER: Making genome-scale phylogenetics accessible to all. Protein Sci. 31, 8–22. [PubMed: 34717010]
39. Mi H, and Thomas P (2009). PANTHER pathway: an ontology-based pathway database coupled with data analysis tools. Methods Mol. Biol 563, 123–140. [PubMed: 19597783]
40. Wang B, Guo H, Yu H, Chen Y, Xu H, and Zhao G (2021). The role of the transcription factor EGR1 in cancer. Front. Oncol 11, 642547. [PubMed: 33842351]
41. Guo H, Zeng B, Wang L, Ge C, Zuo X, Li Y, Ding W, Deng L, Zhang J, Qian X, et al. (2021). Knockdown CYP2S1 inhibits lung cancer cells proliferation and migration. Cancer Biomarkers 32, 531–539. [PubMed: 34275895]
42. Krossa I, Strub T, Martel A, Nahon-Esteve S, Lassalle S, Hofman P, Baillif S, Ballotti R, and Bertolotto C (2022). Recent advances in understanding the role of HES6 in cancers. Theranostics 12, 4374–4385. [PubMed: 35673577]
43. Jiang P, Gil de Rubio R, Hrycaj SM, Gurczynski SJ, Riemondy KA, Moore BB, Omary MB, Ridge KM, and Zemans RL (2020). Ineffectual type 2-to-type 1 alveolar epithelial cell differentiation

in idiopathic pulmonary fibrosis: persistence of the KRT8. Am. J. Respir. Crit. Care Med 201, 1443–1447. [PubMed: 32073903]

44. Győrffy B, Surowiak P, Budczies J, and Lánczky A (2013). Online survival analysis software to assess the prognostic value of biomarkers using transcriptomic data in non-small-cell lung cancer. PLoS One 8, e82241. [PubMed: 24367507]
45. Lánczky A, and Győrffy B (2021). Web-based survival analysis tool tailored for medical research (KMplot): development and implementation. J. Med. Internet Res 23, e27633. [PubMed: 34309564]
46. Marconett CN, Zhou B, Rieger ME, Selamat SA, Dubourd M, Fang X, Lynch SK, Stueve TR, Siegmund KD, Berman BP, et al. (2013). Integrated transcriptomic and epigenomic analysis of primary human lung epithelial cell differentiation. PLoS Genet. 9, e1003513. [PubMed: 23818859]
47. Wang W, He J, Lu H, Kong Q, and Lin S (2020). KRT8 and KRT19, associated with EMT, are hypomethylated and overexpressed in lung adenocarcinoma and link to unfavorable prognosis. Biosci. Rep 40.
48. McCauley KB, Alysandratos KD, Jacob A, Hawkins F, Caballero IS, Vedaie M, Yang W, Slovik KJ, Morley M, Carraro G, et al. (2018). Single-cell transcriptomic profiling of pluripotent stem cell-derived SCGB3A2+ airway epithelium. Stem Cell Rep. 10, 1579–1595.
49. Conchola AS, Frum T, Xiao Z, Hsu PP, Kaur K, Downey MS, Hein RFC, Miller AJ, Tsai YH, Wu A, et al. (2023). Regionally distinct progenitor cells in the lower airway give rise to neuroendocrine and multiciliated cells in the developing human lung. Proc. Natl. Acad. Sci. USA 120, e2210113120. [PubMed: 37279279]
50. Liebler JM, Marconett CN, Juul N, Wang H, Liu Y, Flodby P, Laird-Offringa IA, Minoo P, and Zhou B (2016). Combinations of differentiation markers distinguish subpopulations of alveolar epithelial cells in adult lung. Am. J. Physiol. Lung Cell Mol. Physiol 310, L114–L120. [PubMed: 26545903]
51. Wang Y, Tang Z, Huang H, Li J, Wang Z, Yu Y, Zhang C, Li J, Dai H, Wang F, et al. (2018). Pulmonary alveolar type I cell population consists of two distinct subtypes that differ in cell fate. Proc. Natl. Acad. Sci. USA 115, 2407–2412. [PubMed: 29463737]
52. Juul NH, Yoon JK, Martinez MC, Rishi N, Kazadaeva YI, Morri M, Neff NF, Trope WL, Shrager JB, Sinha R, and Desai TJ (2023). KRAS(G12D) drives lepidic adenocarcinoma through stem-cell reprogramming. Nature 619, 860–867. [PubMed: 37468622]
53. Hafemeister C, and Satija R (2019). Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. Genome Biol. 20, 296. [PubMed: 31870423]
54. Sarabia-Vallejos MA, Ayala-Jeria P, and Hurtado DE (2021). Three-dimensional whole-organ characterization of the regional alveolar morphology in normal murine lungs. Front. Physiol 12, 755468. [PubMed: 34955878]
55. Paster EV, Villines KA, and Hickman DL (2009). Endpoints for mouse abdominal tumor models: refinement of current criteria. Comp. Med 59, 234–241. [PubMed: 19619413]

Highlights

- AT1 cells can give rise to KRAS-driven LUAD using transgenic mouse models
- Alveolar cell of origin influences histologic presentation of LUAD
- $KRT8^+$ intermediate cell state marker is associated with both subtypes of LUAD
- $KRT8^+$ cells were $SCGB3A2^+$ only in AT1-derived LUAD

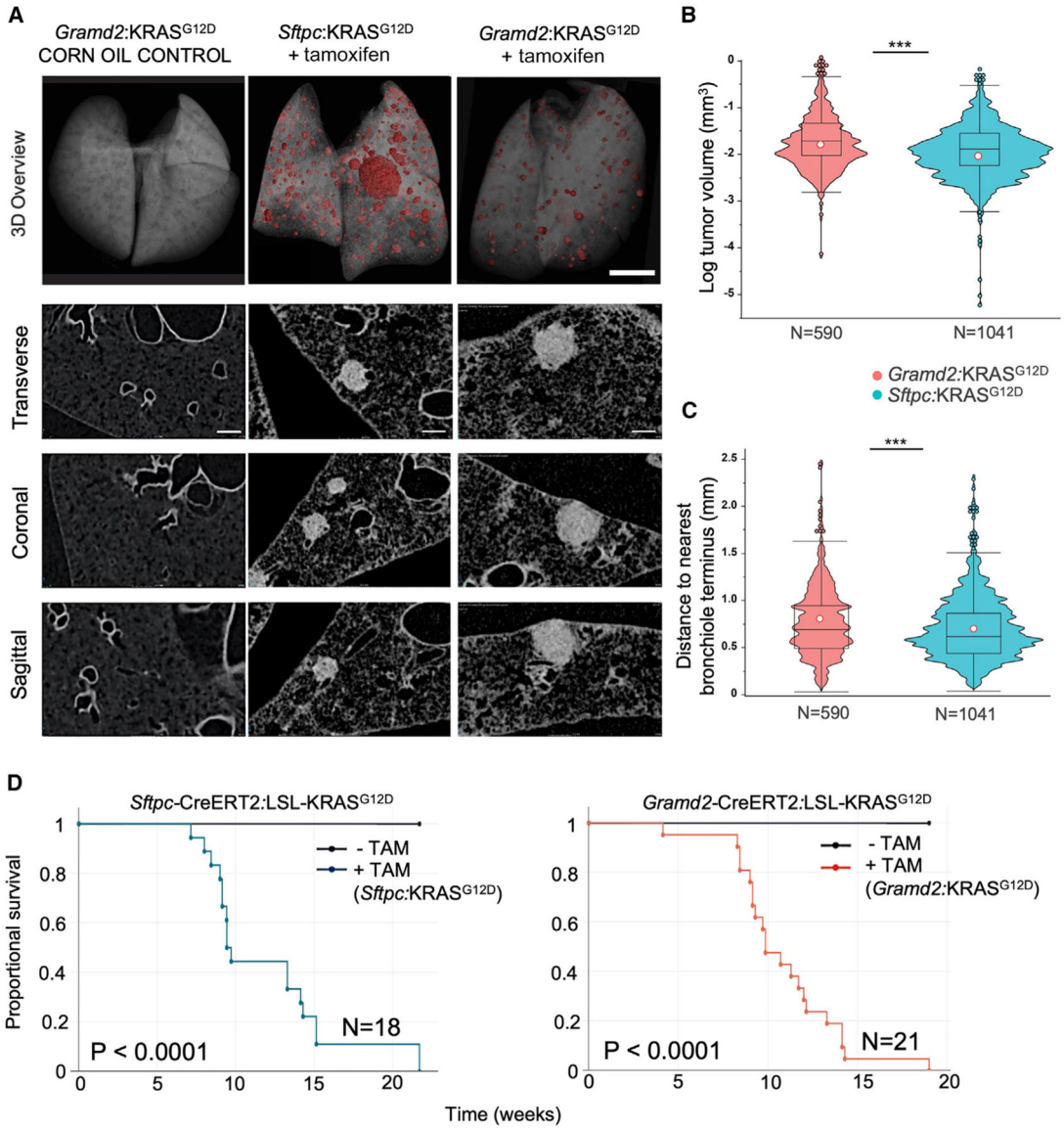


Figure 1. KRAS^{G12D} activation in AT1 cells induces multifocal LUAD lesions *in vivo*

(A) Micro-CT scanning was performed on a transgenic mouse lung representative of each genotype, *Gramd2*:KRAS^{G12D} (corn oil treated and tamoxifen treated) and *Sftpc*:KRAS^{G12D} (tamoxifen treated), at 14 weeks post tamoxifen induction, and the three different imaging planes, transverse, sagittal, and coronal, and three-dimensional (3D) overview are shown in order (scale bars in A 3D overview were 3 mm in each image, and 0.2 mm in transverse, coronal, and sagittal images). Experiment was performed in three different biological replicates per group.

(B) Violin plot using weighted area with −0.5 smoothness indicates each tumor's volume across three independent biological replicates, overlay with max, Q3, median, Q1, and min boxplots. Red circle indicates mean values across all three samples in the group. Unpaired two-tailed Student's t test was used to determine significance. *p < 0.05, **p < 0.01, ***p < 0.001.

(C) Violin plot using weighted area with −0.5 smoothness indicates the distance of each tumor to its nearest bronchiole across three independent biological replicates, overlaid with boxplots showing the quartile 3, median, and quartile 1 range. Red circles indicate mean values across all three samples in the group. All nodules identified from each representative lung are included. Unpaired two-tailed Student's t test was used to determine significance. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

(D) Survival analysis of *Sftpc*:$KRAS^{G12D}$ (left) and *Gramd2*:$KRAS^{G12D}$ (right) post-tamoxifen treatment. n = 18 for *Sftpc*:$KRAS^{G12D}$ mice; n = 21 for *Gramd2*:K-RAS^{G12D} mice. Log rank test was used to determine significance.

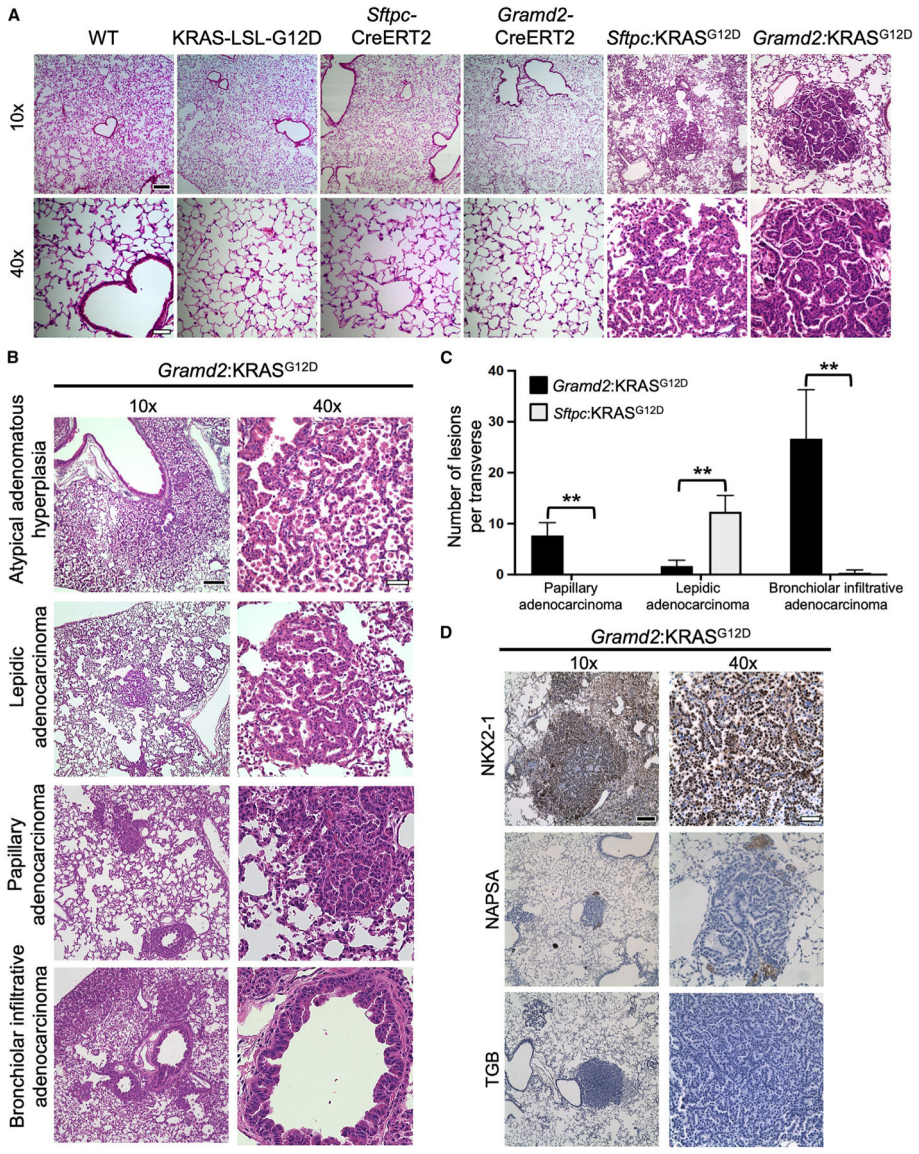


Figure 2. Gramd2^{+} AT1 cells give rise to LUAD with predominantly papillary histology

(A) Representative hematoxylin and eosin (H&E) staining of lungs 14 weeks after tamoxifen injection from control mice (C57BL/6J (WT), KRAS-LSL-G12D, *Sftpc*-CreERT2, *Gramd2*-CreERT2, as well as *Sftpc*:KRAS^{G12D} and *Gramd2*:KRAS^{G12D} mice. n = 3 for all genotypes.

(B) Variation in histologic lesions from *Gramd2*:KRAS^{G12D} mouse lungs, including atypical adenomatous hyperplasia, lepidic adenocarcinoma, papillary adenocarcinoma, and bronchial infiltrative adenocarcinoma. (Scale bars in A and B, black with white outline, 100 μm (10X); white with black outline, 25 μm (40X)). n = 3 for all genotypes.

(C) Quantification of histologic subtypes of lung adenocarcinomas from different mouse genotypes: *Gramd2*:KRAS^{G12D} (black bars; n = 3; 14 weeks) vs. *Sftpc*:KRAS^{G12D} (gray bars; n = 3; 14 weeks). The bar graphs represent the average number of different types of lesions per section per lung. x axis = subtypes of observed LUAD lesions; y axis = number

of lesions per transverse lung section. Data represent means ± SEM. Unpaired two-tailed Student's t test was used to determine significance. *p < 0.05, **p < 0.01, ***p < 0.001. (D) Immunohistochemistry (IHC) analysis of *Gramd2*:KRAS^{G12D} tumor sections for NKX2–1, Napsin A (NAPSA), and thyroglobulin (TGB). Scale bars: Black with white outline, 100 μm (10X); white with black outline, 25 μm (40X). Experiment was performed on three biological replicates.

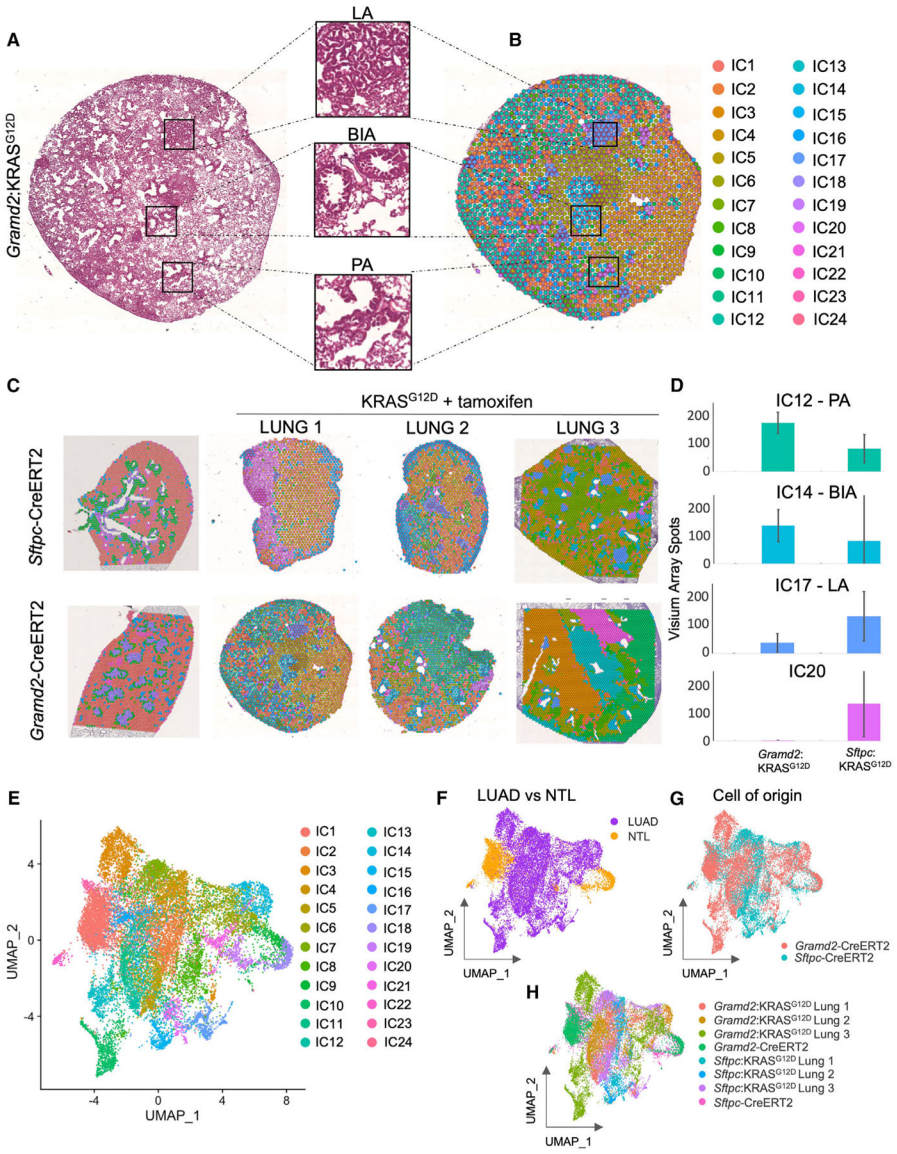


Figure 3. Spatial transcriptomic profiling reveals distinct cell-origin-specific molecular and phenotypic presentation

(A) H&E section generated as part of the Visium 10X spatial transcriptomic profiling performed on representative *Gramd2*:KRAS^{G12D} lung that contains multiple LUAD lesions.

(B) High magnification views of distinct histologic regions within *Gramd2*:KRAS^{G12D} lung sections.

(C) Spatial distribution of integrated clusters (ICs) across all six lung samples within the dataset. Colors indicate distinct ICs. Spatial transcriptomic sequencing was performed on three biological replicates from *Sftpc*:KRAS^{G12D} and *Gramd2*:KRAS^{G12D} mouse lungs.

(D) Barplots of average number of array spots per section. Average (n = 3 samples) and error bars (percent standard deviation) are shown.

(E) UMAP projection of spatial transcriptomics expression data colored by integrated cluster (IC).

(F) UMAP projection of Visium spatial transcriptomic dataset colored by tamoxifen-activated $KRAS^{G12D}$ sample status. Purple = LUAD (tamoxifen-activated $KRAS^{G12D}$), orange = non-tumor lung.
(G) UMAP projection of Visium spatial transcriptomic dataset colored by cell of origin. *Sftpc*:$KRAS^{G12D}$ (teal) and *Gramd2*:$KRAS^{G12D}$ (pink).
(H) UMAP projection of Visium spatial transcriptomic dataset colored by sample genotype and replicate.

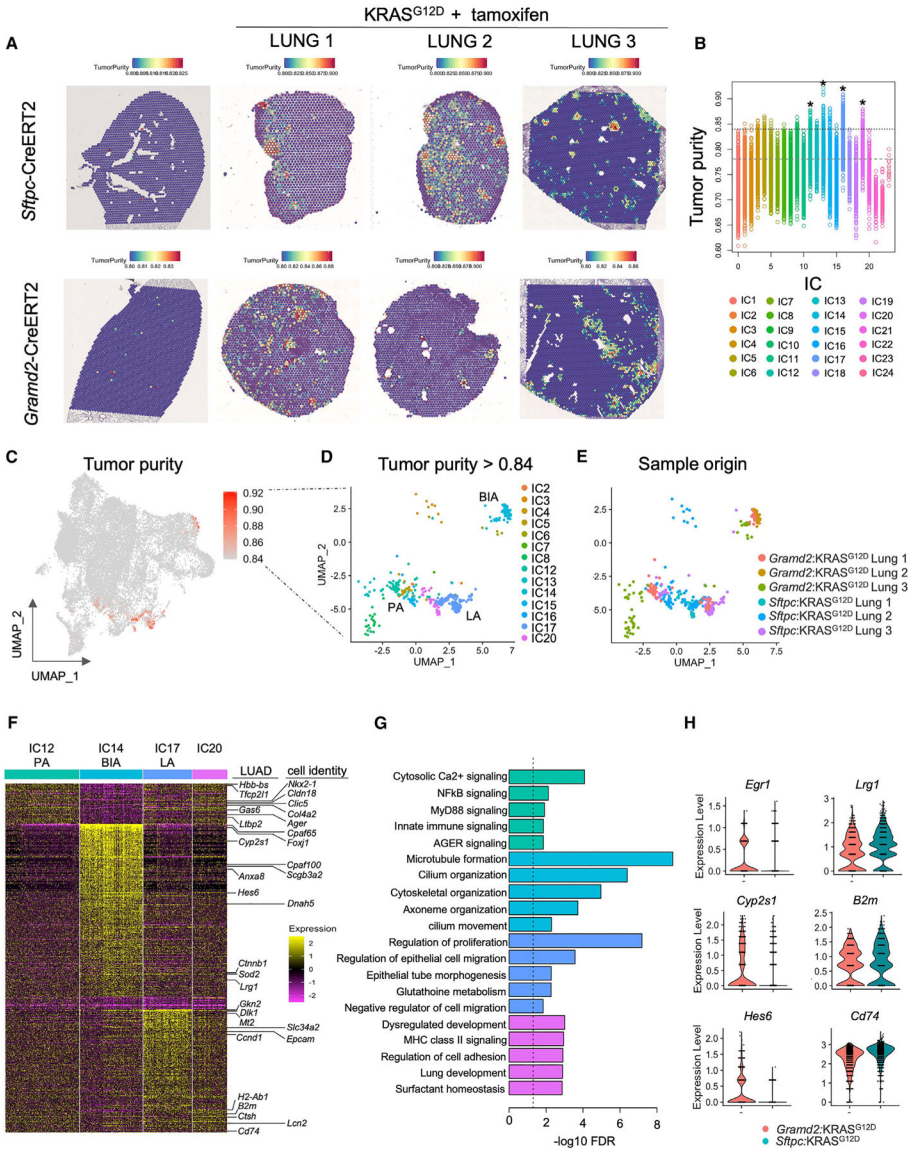


Figure 4. Expression analysis reveals distinct transcriptomic signatures within AT1-derived LUAD

(A) ESTIMATE was applied to quantitate tumor purity within samples; Red = high tumor purity, blue = low tumor purity.

(B) Correlation between tumor purity and cluster identity. Colors = Integrated Cluster ID. Asterisks indicate ICs where a large percentage of array spots correlated to pathologically defined LUAD lesions.

(C) UMAP projection of expression data from all samples colored by tumor purity; Red = high tumor purity, gray = low tumor purity.

(D) UMAP projection of a subset of spatial transcriptomic spots with tumor purity greater than or equal to 0.84. Colors = Integrated Cluster ID, Teal (IC14) = bronchiolar infiltrative adenocarcinoma (BIA) histology, purple (IC17) = lepidic adenocarcinoma, pink = lepidic adenocarcinoma with or without co-occurring micropapillary features, designated “mixed” histology, sea foam green (IC12) = papillary adenocarcinoma (PA) histology.

(E) UMAP projection of spots with tumor purity ≥0.84. Colors indicate sample origin.
(F) Heatmap of top 300 differentially expressed genes between ICs from spots with high tumor purity. Select genes relating to carcinogenesis or lung cell-type-specific markers are indicated. Yellow = highly expressed, purple = little to no expression. Integrated cluster (IC) color as indicated previously in (B).
(G) Bar plot of PANTHER gene set enrichment analysis (GSEA). The −log10 of false-discovery rate (FDR) corrected p values are indicated; dark green = highly significant FDR of Gene Ontology (GO) biologic processes category enrichment.
(H) Violin plots of top differentially expressed genes between *Sftpc*:$KRAS^{G12D}$ (teal) and *Gramd2*:$KRAS^{G12D}$ (pink) across LUAD ICs.

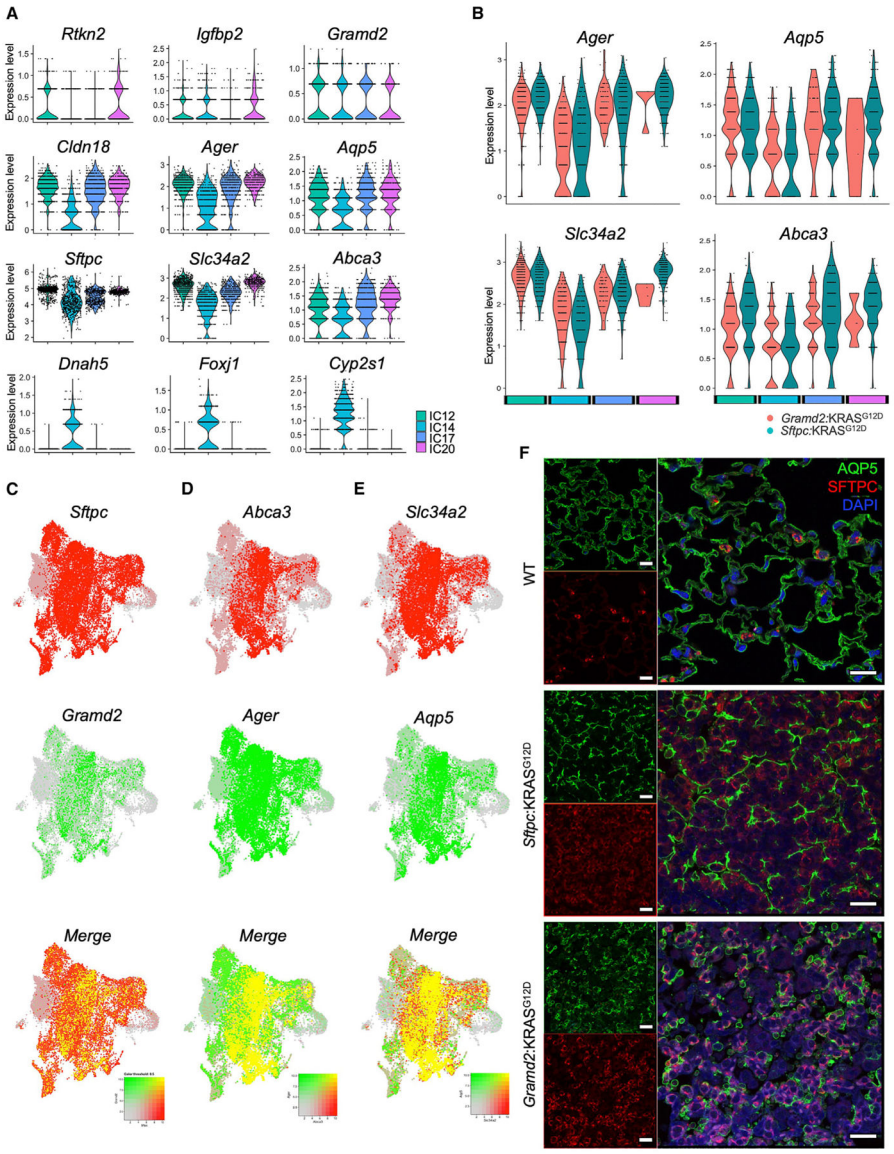


Figure 5. Cellular heterogeneity and marker staining varies between AT2- and AT1-derived LUAD

(A) Violin plots of the indicated AT1 and AT2 cell-specific markers in ICs with high tumor purity.

(B) Violin plots of the indicated cell-type-specific markers, split by cell of origin, teal = *Sftpc*:KRAS^{G12D}, pink = *Gramd2*:KRAS^{G12D}.

(C) UMAPs of correlation between cell-type-specific markers per array spot. *Sftpc* = red, *Gramd2* = green, co-expressed = yellow.

(D) UMAPs of correlation between cell-type-specific markers per array spot. *Abca3* = red, *Ager* = green, co-expressed = yellow.

(E) UMAPs of correlation between cell-type-specific markers per array spot. *Slc34a2* = red, *Aqp5* = green, co-expressed = yellow.

(F) Immunofluorescence imaging of colocalization between SFTPC (red) and AQP5 (green) within either wild-type C57BL/6J (WT), *Sftpc*:KRAS^{G12D} or *Gramd2*:KRAS^{G12D} tumors, n = 3 per genotype. Scale (white bar), 10 μm (63x).

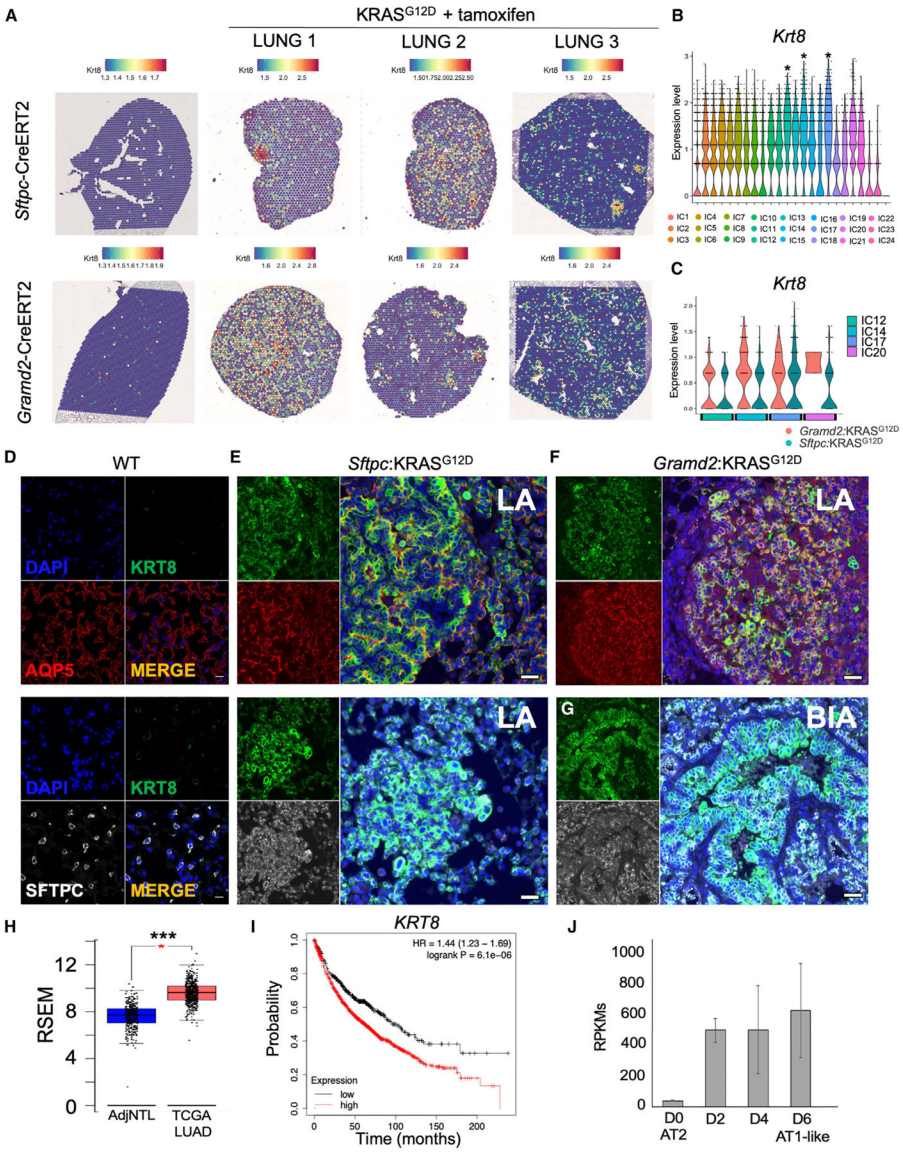


Figure 6. AT2- and AT1-derived LUAD both exhibit *Krt8*$^{+}$ intermediate cell states

(A) Spatial distribution of *Krt8* throughout the entire Visium spatial transcriptomic dataset. Red = high *Krt8* expression, blue = low *Krt8* expression.

(B) Violin plot of *Krt8* expression within individual ICs in all array spots (bottom). Colors indicate distinct ICs.

(C) Violin plot of *Krt8* expression within LUAD-associated ICs. Colors indicate distinct cell of origin. *Sftpc*:KRAS^{G12D} = teal, *Gramd2*:KRAS^{G12D} = pink.

(D) IF staining of KRT8 in control C57BL/6J mouse lungs. Red = AQP5, white = SFTPC, green = KRT8, blue = DAPI, orange = merged. SFTPC was originally captured in TRITC and was converted to grayscale in ImageJ.

(E) IF staining of KRT8 in *Sftpc*:KRAS^{G12D} LUAD with lepidic patterning (LA), colors are as in (D).

(F) IF staining of KRT8 in *Gramd2*:KRAS^{G12D} LUAD with lepidic patterning (LA), colors are as in (D).

(G) IF staining of KRT8 in *Gramd2*:KRAS^{G12D} LUAD with bronchiolar-infiltrative (BIA-LUAD) phenotype, colors are as in (D) (Scale bars in D, E, F and G, white scale bar in lower right corners, 20 μM [20X]; n = 3 per genotype).

(H) Expression of KRT8 in TCGA-LUAD. Blue = adjacent non-tumor lung (AdjNTL), red = LUAD. RSEM = RNA sequencing by expectation maximization. Unpaired two-tailed Student's t test for two variables was used to determine significance. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

(I) Overall survival of LUAD patients stratified by KRT8 expression. Plot generated in KMplot.[44,45] Black = low expression, red = high expression. Logrank p value of differences in survival between groups is shown.

(J) Expression of KRT8 in publicly available AEC 2D culture bulk RNA sequencing.[46] Expression measured in reads per kilobase of gene per millions mapped in sample (RPKMs). Error bars represent mean with standard deviation. (N = 3).

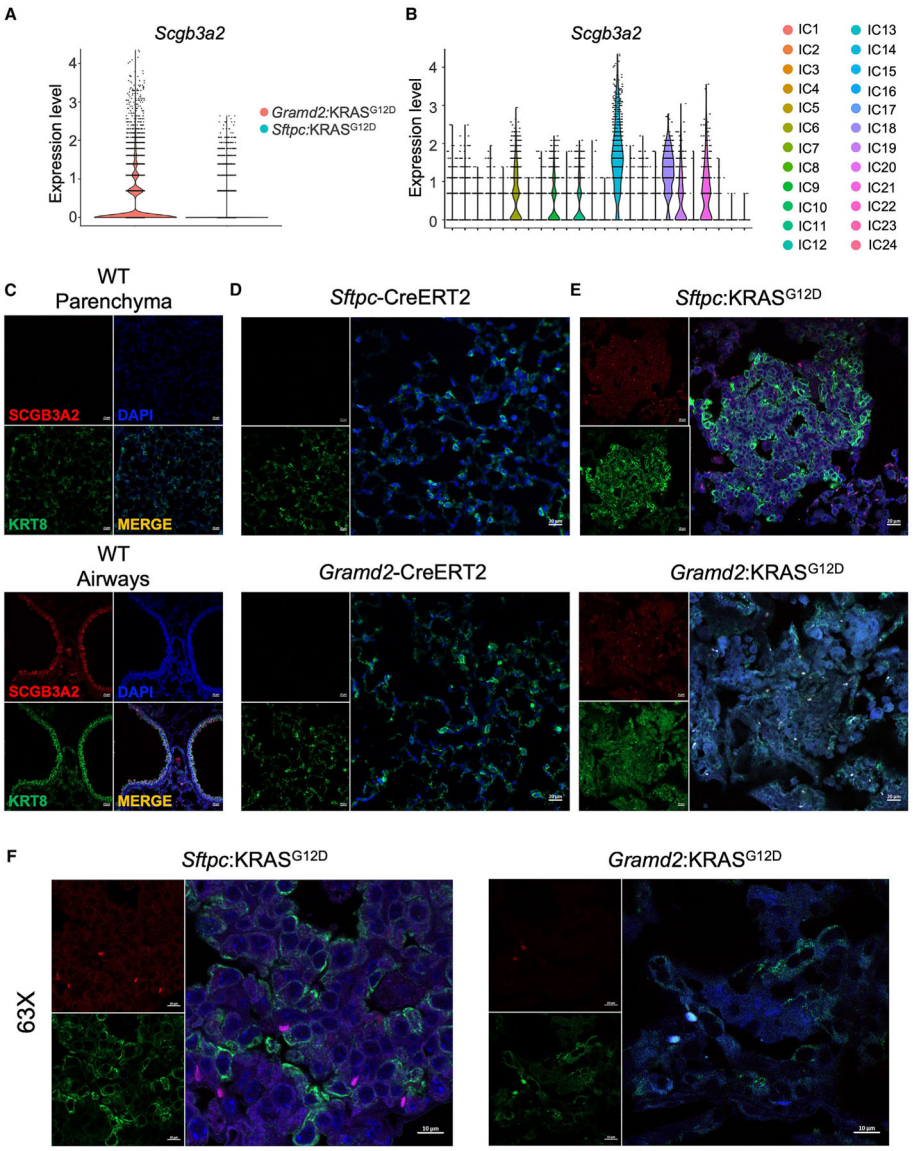


Figure 7. SCGB3A2 is associated specifically with AT1-derived LUAD

(A) Violin plot of *Scgb3a2* expression between *Gramd2*:KRAS^{G12D} and *Sftpc*:KRAS^{G12D} Visium spatial transcriptomic datasets. *Sftpc*:KRAS^{G12D} = teal, *Gramd2*:KRAS^{G12D} = pink.

(B) Violin plot of *Scgb3a2* expression within individual ICs in all array spots present in Visium spatial transcriptomic datasets. Colors indicate distinct ICs.

(C) IF staining of SCGB3A2 in wild-type (WT) control B57 J/L mice lungs in distal airways (parenchyma, top panel) and bronchioles (airway, bottom panel). Red = AQP5, white = SFTPC, green = KRT8, blue = DAPI, orange = merged. SFTPC was originally captured in TRITC and was converted to grayscale in ImageJ. White bar = 20 μm (40X); n = 3 per genotype.

(D) Staining as in (C) on *Sftpc*-CreERT2 (top), *Gramd2*-CreERT2 (bottom), and control lungs. White bar = 20 μm (40X); n = 3 per genotype.

(E) Staining as in (C) on *Sftpc*:KRAS^{G12D} (top) and *Gramd2*:KRAS^{G12D} (bottom) LUAD lesions. White bar = 20 μm (40X); n = 3 per genotype).
(F) Staining as in (C) on *Sftpc*:KRAS^{G12D} (left) and *Gramd2*:KRAS^{G12D} (right) LUAD lesions. Scale bars in (F), white bar = 10 μm (63X); n = 3 per genotype.

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Anti-mouse Multi-Cytokeratin Monoclonal Antibody, Unconjugated	Leica Biosystems	Cat# NCL-AE1/AE3; RRID:AB_564119
Anti-mouse CD68 Monoclonal Antibody, Unconjugated (Clone 514H12)	Leica Biosystems	Cat# PA0286; RRID:AB_10554758
Anti-mouse GATA-3 Monoclonal Antibody (Clone L50-823)	BIOCARE	Cat #CM405A; RRID:AB_10895444
Anti-mouse Napsin A Monoclonal Antibody (Clone TMU-Ad 02)	BIOCARE	Cat #CM388A; RRID:AB_10582498
Anti-Thyroglobulin Monoclonal Antibody, Unconjugated (Clone 2H11 &6E1)	Cell Marque	Cat# 340M-15; RRID:AB_1158894
Anti-TTF-1 Monoclonal Antibody, Unconjugated (Clone 8G7G3/1)	Cell Marque	Cat# 343M-96; RRID:AB_1158937
Anti-mouse Green Fluorescent Protein (GFP) Polyclonal Antibody, Unconjugated	Thermo Fisher Scientific	Cat# A-6455; RRID:AB_221570
Anti-mouse Aquaporin 5 Antibody	Alomone Labs	Cat# AQP-005; RRID:AB_2039736
Anti-mouse pro SP-C Polyclonal Antibody, Unconjugated	Seven Hills Bioreagents	Cat #WRAB-SPC-9337; RRID:AB_451721
Anti-mouse GRAMD2 Polyclonal Antibody	Atlas Antibodies	Cat# HPA029435; RRID:AB_10601811
Anti-mouse Keratin, type II; cytokeratin 8/18; EndoA Antibody - Brulet, P./Kemler, R.; Institut Pasteur	DSHB	Cat# TROMA-I; RRID:AB_531826
Anti-Mouse Ugrp1 Polyclonal Antibody, Unconjugated	R and D Systems	Cat# AF3465; RRID:AB_2183550
Anti-Mouse FOXJ1 Monoclonal Antibody eBioscience (Clone 2A5)	Thermo Fisher Scientific	Cat# 14-9965-82; RRID:AB_1548835
Anti-Rat IgG Antibody Biotinylated	Vector Laboratories	Cat #NC9016344 RRID:AB_2336206
Goat anti-Rabbit IgG (H + L) Highly Cross-Adsorbed Secondary Antibody, Alexa Fluor™ 647	Thermo Fisher Scientific	Cat# A-21245; RRID:AB_2535813
Chemicals, peptides, and recombinant proteins		
Streptavidin, Alexa Fluor™ 488 conjugate	Thermo Fisher	Cat #S11223
Hexamethyldisilazane (HMDS)	Sigma-Aldrich	Cat #SHBG4111V
Antigen unmasking solution	Vector Laboratories	Cat #H3301
Hematoxylin and eosin	Millipore Sigma	Cat #HT110116
CAS blocking buffer	Life Technologies	Cat #008120
Deparaffinization solution	Qiagen	Cat # 19093
Critical commercial assays		
Biotinylation Kit/Biotin Conjugation Kit (Fast, Type A)	Abcam	Cat #ab201795
BOND Polymer Refine Detection	Leica Biosystems	Cat #DS9800
Qiagen RNeasy FFPE kit	Qiagen	Cat #73504
Dual Index Kit TS, Set A	10X GENOMICS	Cat #PN-1000251
Click-iT™ EdU Cell Proliferation Kit	Invitrogen	Cat #C10337
Deposited data		
Data files for spatial transcriptomic data	This paper	GSE215858
Public RNA-seq data of human LUAD	TCGA	https://portal.gdc.cancer.gov/
Experimental models: Organisms/strains		

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Mouse: *Gramd2*:CreERT2	Applied Stem Cell, Inc.	N/A
Mouse: *Sffpc*:CreERT2	Harold Chapman, University of California, San Francisco	N/A
Mouse: B6.129S4-Krastm4Tyj/J	Jackson Laboratories	Strain #:008179
Mouse: Gt(ROSA)26Sortm4(ACTB-tdTomato,-EGFP)Luo/J	Jackson Laboratories	Strain #:007576
Oligonucleotides		
Gramd2-CreERT2 common forward: 5'-CTAGTCCTGTCCTCGTCCTATC-3'	This paper	N/A
Gramd2-CreERT2 mutant allele reverse: 5'-GGGAAACCATTTCCGGTTATTC-3'	This paper	N/A
Gramd2-CreERT2 WT allele reverse: 5'-CACATCCCAGCCTTCTCAAA-3'	This paper	N/A
Sffpc-CreERT2 WT allele forward: 5'-TGGTTCCGAGTCCGATTCTTC-3'	This paper	N/A
Sffpc-CreERT2 WT allele reverse: 5'-CCTTTTGCTCTGTTCCCCATTA-3'	This paper	N/A
Sffpc-CreERT2 mutant allele forward: 5'-TGAGGTTCGCAAGAACCTGATGGA-3'	This paper	N/A
Sffpc-CreERT2 mutant allele reverse: 5'-ACCAGCTTGCATGATCTCCGGTAT-3'	This paper	N/A
KRAS-LSL-G12D common reverse: 5'-CTGCATAGTACGCTATACCCTGT-3'	This paper	N/A
KRAS-LSL-G12D mutant forward: 5'-GCAGGTCGAGGGACCTAATA-3'	This paper	N/A
KRAS-LSL-G12D WT forward: 5'-TGTCTTTCCCCAGCACAGT-3'	This paper	N/A
mTmG common reverse: 5'-CTTTAAGCCTGCCCAGAA GA −3'	This paper	N/A
mTmG mutant allele forward: 5'- TAGAGCTTGCGGAACCCTTC −3'	This paper	N/A
mTmG WT allele forward: 5'- AGGGAGCTGCAGTGGAGTAG −3'	This paper	N/A
Software and algorithms		
VGSTUDIO MAX 3.3.2.170119 (64 bit)	Volume Graphics GmbH	https://www.volumegraphics.com/en/products/vgsm/whats-new-in-vgstudio-max-3-3-x.html
Space Ranger pipeline (10x Genomics)	10X GENOMICS	https://support.10xgenomics.com/spatial-gene-expression/software
R 4.0.5	R	https://www.r-project.org/
R Studio (v 1.4.1717)	RStudio	https://www.rstudio.com/
Seurat (v4.1.1)	Hafemeister et al.[53]	https://satijalab.org/seurat/
ESTIMATE (v1.0.13)	Yoshihara et al.[34]	https://bioinformatics.mdanderson.org/estimate/rpackage.html
IPA (v 01-20-04)	QIAGEN	https://digitalinsights.qiagen.com/products-overview/discovery-insights-portfolio/analysis-and-visualization/qiagen-ipa/
Other		
Zeiss LSM 800 Confocal Laser Scanning Microscope	ZEISS	N/A

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Phoenix Nanotom® M micro-CT scanner system	Baker Hughes	N/A
Applied Biosystems ProFlex PCR System	Thermo Fisher Scientific	N/A
ChemiDoc MP Imaging System	Bio-Rad	N/A