Club cells form lung adenocarcinomas and maintain the alveoli of adult mice

Magda Spella[1*], Ioannis Lilis[1], Mario A.A. Pepe[2], Yuanyuan Chen[3], Maria Armaka[4], Anne-Sophie Lamort[2], Dimitra E. Zazara[1], Fani Roumelioti[4], Malamati Vreka[1,2], Nikolaos I. Kanellakis[1], Darcy E. Wagner[2], Anastasios D. Giannou[1], Vasileios Armenis[1], Kristina A.M. Arendt[2], Laura V. Klotz[2], Dimitrios Toumpanakis[5], Vassiliki Karavana[5], Spyros G. Zakynthinos[5], Ioanna Giopanou[1], Antonia Marazioti[1], Vassilis Aidinis[4], Rocio Sotillo[3,6], and Georgios T. Stathopoulos[1,2,6*].

[1] Laboratory for Molecular Respiratory Carcinogenesis, Department of Physiology, Faculty of Medicine, University of Patras, 26504 Rio, Greece

[2] Comprehensive Pneumology Center (CPC) and Institute for Lung Biology and Disease (iLBD), University Hospital, Ludwig-Maximilians University and Helmholtz Center Munich, Member of the German Center for Lung Research (DZL), 81377 Munich, Germany

[3] Division of Molecular Thoracic Oncology, German Cancer Research Center (DKFZ), Translational Lung Research Center (TLRC), Member of the German Center for Lung Research (DZL), 69120 Heidelberg, Germany

[4] Institute of Immunology, Biomedical Sciences Research Center "Alexander Fleming", Vari 16672, Greece

[5] 1st Department of Critical Care Medicine and Pulmonary Services, Evangelismos Hospital, School of Medicine, National and Kapodistrian University of Athens, 10675 Athens, Greece

[6] Equal senior co-authors

[*] Co-corresponding authors

Corresponding authors: Magda Spella, PhD and Georgios T. Stathopoulos, MD PhD; Laboratory for Molecular Respiratory Carcinogenesis, Department of Physiology, Faculty of Medicine, University of Patras; Basic Biomedical Sciences Building, 2nd Floor, Room B40, 1 Asklepiou Str., University Campus, 26504 Rio, Greece; Phone: +30-2610-969154; Fax: +30-2610-969176; E-mail: magsp@upatras.grand gstathop@upatras.gr; and Lung Carcinogenesis Group, Comprehensive Pneumology Center (CPC) and Institute for Lung Biology and Disease (iLBD), Max-Lebsche-Platz 31, 81377 Munich, Germany; Phone: +49 (89) 3187 1194; Fax: +49 (89) 3187 4661; E-mail: stathopoulos@helmholtz-muenchen.de.

ABSTRACT

Lung cancer and chronic lung diseases impose major disease burdens worldwide and are caused by inhaled noxious agents including tobacco smoke. The cellular origins of environmental-induced lung tumors and of the dysfunctional airway and alveolar epithelial turnover observed with chronic lung diseases are unknown. To address this, we combined mouse models of genetic labeling and ablation of airway (club) and alveolar cells with exposure to environmental noxious and carcinogenic agents. Club cells are shown to survive *KRAS* mutations and to form lung tumors after tobacco carcinogen exposure. Increasing numbers of club cells are found in the alveoli with aging and after lung injury, but go undetected since they express alveolar proteins. Ablation of club cells prevents chemical lung tumors and causes alveolar destruction in adult mice. Hence club cells are important in alveolar maintenance and carcinogenesis and may be a therapeutic target against premalignancy and chronic lung disease.

Word count, abstract: 148.

Keywords: Lung adenocarcinoma; Tobacco chemical; Urethane; Club cell; CCSP; SFTPC.

ABBREVIATIONS LIST AND MASTER LEGEND

AEC, airway epithelial cells; AMΦ, alveolar macrophages; ANOVA, analysis of variance; ATII, alveolar type II cells; BAL, bronchoalveolar lavage; BASC, bronchoalveolar stem cells; BHT, butylated hydroxytoluene; BMDM, bone-marrow-derived macrophages; C57BL/6 mice, mouse strain inherently resistant to chemical carcinogens; CCSP, Clara cell secretory protein; CCSP.CRE mice, mouse strain in which CRE expression is driven by the *Scgb1a1* promoter; CRE, causes recombination; ddPCR, digital droplet PCR; DTA mice, genetic suicide mouse strain that expresses Diphtheria toxin upon CRE-mediated recombination; EC, ethyl carbamate, urethane; FOXJ1, forkhead box J1; FVB mice, mouse strain inherently susceptible to chemical carcinogens; GFP, green fluorescent protein; GSEA, gene set enrichment analysis; *KRAS*, Kirsten rat sarcoma viral oncogene homologue; KRT5, keratin 5; LUAD, Lung adenocarcinoma; LYZ2, lysozyme 2; LYZ2.CRE mice, mouse strain in which CRE expression is driven by the *Lyz2* promoter; MCA, 3-methylcholanthrene; μCT, micro-computed tomography; *n*, sample size; NES.CRE mice, mouse strain in which CRE expression is driven by the *Nestin* neural promoter; *P*, probability; PCNA, proliferating cell nuclear antigen; LUC mice, mouse strain that reports for CRE-mediated recombination via firefly (Photinus pyralis) luciferase expression; SD, standard deviation; SFTPC, surfactant protein C; SFTPC.CRE mice, mouse strain in which CRE expression is driven by the *Sftpc* promoter; SOX2.CRE mice, mouse strain in which CRE expression is driven by the *Sox2* promoter; TOMATO, red fluorescent TdTomato fluorophore; TOMATO (mT/mG) mice, mouse strain that reports for CRE-mediated recombination via a switch from TOMATO to GFP fluorophore expression; TUBA1A, acetylated tubulin; VAV.CRE mice, mouse strain in which CRE expression is driven by the *Vav1* panhematopoietic promoter.

INTRODUCTION

Chronic lung diseases present tremendous health burdens attributed to dysfunctional alveolar repair [1-3]. Lung adenocarcinoma (LUAD), the leading cancer killer worldwide, is mainly caused by chemical carcinogens of tobacco smoke that induce mutations of the Kirsten rat sarcoma viral oncogene homologue (*KRAS*) in yet unidentified pulmonary cells [4-8]. The discovery of the cellular lineages and the transcriptional programs that underlie lung regeneration and carcinogenesis is extremely important, since epithelial developmental pathways are intimately related with oncogenic signaling to jointly regulate stemness and drug resistance [9, 10]. To this end, lineage-specific genes encoding epithelial proteins that support the physiological functions of the lungs were recently shown to suffer non-coding insertions and deletions in LUAD, lending further support to the longstanding notion that epithelial cells that express lung-restricted proteins are the cellular sources of LUAD [11].

However, these cells of origin of LUAD remain only partially charted. Previous pulmonary lineage tracing studies that utilized noxious insults and ectopic expression of oncogenes in the respiratory epithelium incriminated both airway and alveolar cells as progenitors of newly formed alveoli and/or LUAD in adult mice [12-18]. To this end, airway epithelial cells (AEC) line the bronchi and include ciliated, basal, goblet, and Clara or club cells; alveolar type II cells (ATII) and alveolar macrophages (AMΦ) are distributed across the distal lung parenchyma; and bronchoalveolar stem cells (BASC) with dual AEC/ATII properties are located at the bronchoalveolar junctions. Established markers currently used to label these pulmonary lineages include acetylated tubulin (TUBA1A) for ciliated cells, keratin 5 (KRT5) for basal cells, forkhead box J1 (FOXJ1) for goblet cells, Clara cell secretory protein (CCSP) for club cells, surfactant protein C (SFTPC) and lysozyme 2 (LYZ2) for ATII cells, and LYZ2 for AMΦ, are summarized in Figure 1A and Figure 1-figure supplement 1, and are extensively studied in [18, 19]. However, existing mouse models for lineage tracing feature

incomplete and/or promiscuous lung cell labeling, i.e. cellular markings fail to identify all cells of a target lineage (false negative marking) or wrongfully identify other cells outside of the target lineage (false positive marking) [12-18]. In addition, all studies that attempted to address the cellular origins of LUAD to date employed overexpression of oncogenes such as *KRAS*^{G12D} in the lungs, to conclude that ATII cells or BASC are the most probable culprits of the disease [13-18]. However, it was recently shown that oncogenic *KRAS*^{G12D}-driven mouse lung tumors do not imitate the mutational landscape of human LUAD as closely as tobacco carcinogen-induced LUAD do [7, 8, 20].

Here we aimed at identifying the cell lineage(s) that give rise to human-relevant tobacco carcinogen-triggered LUAD in mice and that regenerate adult murine alveoli after injury. For this, we combined mouse models of genetic labeling and ablation of airway and alveolar epithelial cells with noxious and tumorigenic insults to the adult lung. To achieve this, we adapted multi-hit chemical carcinogen exposure protocols to the murine C57BL/6 strain that is resistant to chemical tumor induction [21-23], and corroborated the findings with the FVB strain that is susceptible to single-hit carcinogenesis [20, 23, 24]. We show that aging, toxic, and carcinogen insults to the adult mouse lung cause expansion of airway-marked cells to the alveolar parenchyma, where they express the alveolar marker SFTPC and facilitate alveolar repair and carcinogenesis. In addition, we report how airway cells preferentially sustain chemical-induced *KRAS* mutations leading to LUAD that are spatially linked with neighboring bronchi. Moreover, genetic ablation of airway cells is shown to hinder alveolar maintenance and carcinogenesis in mice, indicating a central role for these cells in alveolar regeneration and LUAD triggered in response to environmental challenges.

RESULTS

Accurate genetic labeling of the airway lineage.

To evaluate the contribution of different epithelial lung cell lineages to chemical-induced LUAD, we crossed a CRE-reporter strain that switches somatic cells from membranous tdTomato (mT; hereafter TOMATO) to membranous GFP (mG; hereafter GFP) fluorescence upon CRE-mediated recombination (mT/mG; hereafter TOMATO mice) [25] to six different CRE-driver strains on the C57BL/6 background [18, 26-30]. This permitted the permanent genetic GFP-labeling of different lung cell lineages (mouse strains are listed in Figure 1A and Figure 1-figure supplement 2, in Materials and methods, and in Master Legend). Double heterozygote offspring at six postnatal weeks (i.e., after mouse lung development is complete; [12, 18]) were examined for GFP-labeling (results are shown in Figure 1A, Figure 1- figure supplements 3, 4, and in Figure 1-figure supplement 4- source data 1). This approach labeled permanently all AEC of GFP;CCSP.CRE mice, some AEC and all ATII of GFP;SFTPC.CRE mice, some ATII and all AMΦ of GFP;LYZ2.CRE mice, and various other cells in the remaining intercrosses (Figure 1A, Figure 1-figure supplements 3-5, and Figure 1-figure supplement 5-source data 1). Co-localization of GFP-labeling with lineage protein markers (listed in Figure 1A and Figure 1-figure supplement 1) revealed that genetic GFP-labeling in GFP;CCSP.CRE mice marked all airway epithelial cells including club and ciliated cells, in GFP;SFTPC.CRE mice most airway and all alveolar epithelial type II cells, and in GFP;LYZ2.CRE mice some alveolar epithelial type II cells and all alveolar macrophages (Figure 1B, Figure 1-figure supplements 6-8, Figure 1-figure supplement 7-source data 1, and Figure 1-figure supplement 8-source data 1).These findings show precise airway epithelial lineage labeling in GFP;CCSP.CRE mice and non-specific airway/alveolar/myeloid lineage labeling in GFP;SFTPC.CRE and GFP;LYZ2.CRE mice.

Airway cells in chemical-induced lung adenocarcinoma.

We next triggered LUAD in GFP;CCSP.CRE, GFP;SFTPC.CRE, and GFP;LYZ2.CRE mice on the C57BL/6 background using repetitive exposures to the tobacco carcinogens urethane (ethyl carbamate, EC; stand-alone mutagen and tumor promoter) [20, 21, 23, 24] or 3-methylcholanthrene followed by butylated hydroxytoluene (MCA/BHT; a two-hit mutagen/tumor promoter regimen) [22] (Figure 1C, Figure 1-figure supplements 9, 10, and Figure 1-figure supplement 10-source data 1). In both models, preneoplastic (airway epithelial hyperplasias and atypical alveolar hyperplasias) and neoplastic (adenoma and LUAD) lesions classified according to established guidelines [31] were located both in the airways and the alveolar regions. However, established lung tumors were most frequently located near or inside the airways (Figure 1C and Figure 1-figure supplement 11). All hyperplasias and tumors of GFP;SFTPC.CRE and some of GFP;LYZ2.CRE mice were GFP-labeled, but this was not informative, since baseline marking of GFP;SFTPC.CRE and GFP;LYZ2.CRE mice were non-specific. Interestingly, all hyperplasias and tumors of GFP;CCSP.CRE mice contained GFP-labeled airway cells that did not express the club cell marker CCSP anymore, but had acquired expression of the alveolar epithelial markers SFTPC with or without LYZ2 (Figure 1D, Figure 1-figure supplements 12-15, and Figure 1-figure supplement 13-source data 1). Identical results were recapitulated using single urethane hits to GFP;CCSP.CRE, GFP;SFTPC.CRE, and GFP;LYZ2.CRE mice backcrossed > F12 to the susceptible FVB strain, which result in human LUAD-like mutations including *Kras*^{Q61R} [20, 24, 32] (Figure 1D and Figure 1-figure supplements 16-19). Collectively, these data support that airway cells contribute to chemical-induced LUAD, shifting from airway to alveolar marker expression during carcinogenesis.

Airway cells sustain *Kras*^{Q61R} mutations and give rise to juxtabronchial tumors.

We next used digital droplet PCR (ddPCR) to determine the lung lineages that suffer *Kras*^{Q61R} driver mutations at early time-points after single urethane hits [20, 24, 32]. For this,

GFP;CCSP.CRE and GFP;LYZ2.CRE mice backcrossed > F12 to the susceptible FVB strain received urethane and duplexed ddPCR designed to single-copy-co-amplify *Kras* and *Rosa*^{mT} was performed one and two weeks later. Interestingly, GFP-labeled cells of both mouse strains had *Kras*^{Q61R} mutations at one week post-urethane, but *Kras*^{Q61R} mutations selectively persisted in GFP-labeled airway cells in the lungs of GFP;CCSP.CRE mice at two weeks (Figure 2A, Figure 2-figure supplement 1, and Figure 2-source data 1). In addition, three-dimensional reconstruction of tumor-bearing lungs of FVB mice at 6 months post-urethane using high-resolution micro-computed tomography (μCT) revealed that most lung tumors were spatially linked with the airways, in accord with pathology results (Figures 2B, 2C, and Figure 2-source data 2). These results support the involvement of airway cells in chemical-induced lung adenocarcinoma formation in mice.

Alveolar dissemination of airway-labeled cells during carcinogenesis.

Since airborne carcinogens act globally on the respiratory field [33], we examined non-neoplastic alveolar areas of carcinogen-treated GFP;CCSP.CRE mice, to discover markedly increased numbers of GFP-labeled cells in the alveoli of carcinogen-treated mice compared with saline-treated or naïve controls (Figure 3A, Figure 3-figure supplements 1, 2, and Figure 3-figure supplement 2-source data 1). Immunostaining revealed that juxtabronchial GFP-labeled cells still expressed CCSP, but lost CCSP and acquired SFTPC expression when located in alveoli and tumors (Figure 3B and Figure 3-figure supplements 3, 4). The expansion of airway cells after urethane exposure was also documented using bioluminescent imaging of double heterozygote offspring of CCSP.CRE intercrosses with Luciferase-expressing (LUC) mice [34], a strain emitting light specifically from airway epithelia (Figure 3-figure supplement 5, and Figure 3-figure supplement 5-source data 1). In addition, co-staining of human LUAD [35] for the alveolar marker SFTPC and the airway markers CCSP and KRT5 showed co-localization of SFTPC with KRT5 but not with CCSP (Figure 3C and

Figure 3-figure supplement 6). These results suggest that airway epithelial cells expand to alveolar regions during field cancerization by tobacco carcinogens, a process involving either direct alveolar cell recycling by airway epithelial cells or transient CCSP expression by alveolar cells during carcinogenesis. Moreover, that human and murine LUAD carry airway imprints although their location and protein expression suggests an alveolar origin [18, 36-39].

Airway cells in the aging and injured adult alveolus.

We next examined the kinetics of lineage-labeled cells during aging, injury, and repair. While the number of GFP-labeled cells in the alveoli of aging GFP;SFTPC.CRE and GFP;LYZ2.CRE mice was stable, GFP-labeled airway cells in the alveoli of aging GFP;CCSP.CRE mice progressively increased and expressed SFTPC protein (Figures 4A, 4B and Figure 4-source data 1). Bleomycin treatment, which depletes alveolar type II cells [40], accelerated the accumulation of GFP-labeled airway cells in the alveoli and in urethane-triggered LUAD (Figures 4C, 4D, Figure 4-figure supplements 1, 2, Figure 4-source data 2, and Figure 4-figure supplement 2-source data 1). GFP-labeled airway cells expressing the alveolar marker SFTPC also increased in the alveoli of GFP;CCSP.CRE mice exposed to perinatal hyperoxia that damages forming alveoli [41], and in the alveoli of GFP;CCSP.CRE mice treated with naphthalene that kills airway epithelial cells [39, 41], but were not identified within the airways of naphthalene-treated GFP;CCSP.CRE mice; these appeared to be repopulated by GFP-labeled airway cells that express the club cell marker CCSP (Figures 4E-4H, Figure 4-figure supplements 3, 4, Figure 4-source data 3, 4, and Figure 4-figure supplement 4-source data 1). In line with the latter finding, no GFP-labeled alveolar cells were identified in the airways of GFP;LYZ2.CRE mice recovering from naphthalene-induced injury (Figures 4G, 4H). Taken together, the data indicate that airway-originated cells repopulate both the airways and the alveoli during aging and recovery from injury, while

alveolar cells do not reconstitute the airways, in line with previous findings [18, 41]. The observed alveolar spread of airway-labeled cells was explained by either peripheral migration of airway cells or transient CCSP expression by regenerating alveolar cells.

Airway cells maintain alveoli and foster tumors.

To further examine the role of airway and alveolar cells in alveolar homeostasis and lung carcinogenesis, we ablated them by crossing CCSP.CRE, SFTPC.CRE, and LYZ2.CRE mice to mice expressing Diphtheria toxin in somatic cells upon CRE-mediated recombination (DTA mice) [42]. Triple transgenic GFP;DRIVER.CRE;DTA intercrosses were also generated to evaluate ablation efficiency. As expected, SFTPC.CRE;DTA and GFP;SFTPC.CRE;DTA mice were fetal lethal (no double or triple heterozygote offspring was obtained by $n > 3$ intercrosses, > 10 litters, and > 60 off-springs for each genotype; $P < 0.0001$, Fischer's exact test). However, all other ablated mice survived till adulthood. Airway epithelial ablation was complete in GFP;CCSP.CRE;DTA mice, while some GFP-labeled alveolar macrophages persisted in GFP;LYZ2.CRE;DTA mice, presumably freshly recruited monocytes initiating LYZ2 expression. Immunostaining revealed that the denuded airway epithelium of 12-week-old GFP;CCSP.CRE;DTA mice contained few flat CCSP+SFTPC+LYZ2+ immunoreactive cells, while the apparently intact alveolar spaces of GFP;LYZ2.CRE;DTA mice harbored only some CCSP-SFTPC-LYZ2+ immunoreactive alveolar macrophages (Figure 5A, Figure 5-figure supplements 1, 2, and Figure 5-figure supplement 2-source data 1). Remarkably, morphometric and functional analyses of 12-week-old DTA control, CCSP.CRE;DTA, and LYZ2.CRE;DTA mice showed that LYZ2.CRE;DTA mice displayed normal airway caliper and mean linear intercept (measures of airway and alveolar structure), normal number of CD45+CD11b+ myeloid cells in bronchoalveolar lavage (BAL; measure of airspace inflammation), and normal airways resistance and static compliance (measures of airway and alveolar function) compared with

DTA controls. However, CCSP.CRE;DTA mice displayed widened airway and alveolar dimensions with inflammatory interalveolar septal destruction evident by increased mean linear intercept, CD45+CD11b+ cells in BAL, and static compliance (Figures 5B, 5C and Figure 5-source data 1), mimicking human chronic obstructive pulmonary disease [1]. Finally, we exposed control and ablated mice to ten consecutive weekly urethane exposures. All mice survived six months into carcinogen treatment, and CCSP.CRE;DTA and LYZ2.CRE;DTA mice were equally protected from LUAD development compared with controls (Figures 5D, 5E, and Figure 5-source data 2). Taken together, these results show that the CCSP+ airway lineage maintains postnatal alveolar structure and function, and, together with the LYZ2+ alveolar lineage, are required for lung adenocarcinoma development.

Airway epithelial signatures in experimental and human lung adenocarcinoma.

We subsequently examined the transcriptomes of cell lines isolated from urethane-induced LUAD [32] and of murine lungs with those of murine AEC isolated from tracheal explants, of murine ATII cells [43], and of murine bone-marrow-derived macrophages (BMDM). The AEC transcriptome was specifically enriched in LUAD cells compared with whole lungs (Figures 6A, 6B, Figure 6-figure supplement 1, and Figure 6-source data 1). LUAD cell lines lost expression of epithelial markers compared with their native lungs, but displayed up-regulated expression of LUAD markers (i.e., *Krt18* and *Krt20*), of epidermal growth factor receptor ligands (*Areg* and *Ereg*), and of the *Myc* oncogene (Figure 6-figure supplements 2-4, and Figure 6-figure supplement 2-source data 1). Similar analyses of the transcriptomes of human LUAD and corresponding healthy lungs [44], and of primary human AEC, ATII, and AMΦ [45-47] also disclosed that the AEC transcriptome was significantly enriched in LUAD compared with healthy lungs (Figures 6C, 6D and Figure 6-source data 2). Gene set enrichment analyses (GSEA) showed that the mouse AEC transcriptome predominated over ATII/BMDM transcriptomes in LUAD cells (Figure 6E, Figure 6-figure supplement 5, and

Figure 6-source data 3). In addition, the human AEC transcriptome was enriched equally with ATII/AMΦ transcriptomes in human LUAD compared with healthy lungs (Figure 6F, Figure 6-figure supplement 6, and Figure 6-source data 4). These results showed the presence of an anticipated alveolar and an unexpected airway epithelial transcriptomic signature in tobacco carcinogen-induced LUAD of mice and men. The more pronounced results in mice were plausible by the early nature of the human surgical specimens examined compared with our murine cell lines that present advanced metastatic tumor cells.

DISCUSSION

We characterized the dynamics of respiratory epithelial cells in the postnatal mouse lung during aging and after challenge with noxious and carcinogenic insults. The contributions of airway cells to chemical-induced lung adenocarcinoma are described for the first time (Figure7A). Although the peripheral location and molecular phenotype of murine and human lung adenocarcinoma (i.e., the expression of the alveolar epithelial marker SFTPC) suggest an alveolar origin, we show here that both airway and alveolar cells are found in environmental-induced lung adenocarcinoma and that, in fact, airway cells may play a more prominent role during the initial steps of carcinogenesis. Furthermore, airway cells are implicated in postnatal alveolar maintenance during aging and recovery from injury. Our analyses facilitate insights into the dynamics of epithelial lineages in the postnatal lung (Figure7B) and indicate that airway cells are essential for the sustained structural and functional integrity of adult alveoli. Finally, mouse and human lung adenocarcinomas are shown to bare transcriptome markings of highly enriched airway signatures, rendering our findings plausible in both experimental and human lung adenocarcinoma.

This study addresses the cellular and molecular signatures of chemical-induced lung adenocarcinoma. Lung tumors induced in two different mouse strains by two different

chemical regimens contained in tobacco smoke are shown to contain airway epithelial markings. This is important because human lung adenocarcinoma is inflicted by chronic exposure to tobacco smoke and other environmental exposures [6-8, 20-22, 48, 49]. As such, the mutation profile of the human disease is more closely paralleled by chemical-induced murine lung tumors compared with lung cancers triggered by transgenic expression of *Kras*^{G12C} or *Kras*^{G12D} in the respiratory epithelium [20]. Although the latter transgenic tumors have been extensively studied [13-18], chemical-induced lung adenocarcinomas have not been investigated. In all mouse models we studied, all tumors contained the airway genetic marking, in contrast with the LYZ2 alveolar genetic marking which was dispensable for lung adenocarcinoma development. Our observations support the multi-stage field concept of chemical carcinogenesis [33], according to which tumor-initiated cells undergo multiple steps of genomic evolution and phenotypic appearance that include an obligatory airway-like stage. In fact, the prevalence of a different *Kras* mutation in urethane-induced tumors (*Kras*^{Q61R}) compared to *KRAS*$^{G12C/D}$ mutations in the transgenic mouse models has led to the suggestion that chemical carcinogens introduce *KRAS* mutations in a different population of tumor-initiating cells than mouse models of genetic *KRAS* activation [20]. Our findings of airway epithelial cells being more sensitive than alveolar type II cells to *Kras*^{Q61R} mutations during the initial steps of urethane-induced carcinogenesis further supports this notion and render airway cells an attractive novel target for premalignancy.

The consistent finding of CCSP genetic markings (indicative of airway epithelial origin) together with SFTPC and LYZ2 protein expression (indicative of alveolar epithelial phenotype) in chemical-triggered lung adenocarcinomas and their precursor lesions implies three different scenarios for lung adenocarcinoma formation: i) airway epithelial cells colonize the distal lung during carcinogenesis thereby activating obligate (SFTPC+) and dispensable (LYZ2+) alveolar transcriptomes; ii) alveolar cells transit through an obligate

CCSP+ with or without a dispensable LYZ2+ stage during the process; or iii) lung adenocarcinoma arises from multipotent progenitors that express multiple epithelial markers, such as those found during pulmonary embryogenesis, in human lung adenocarcinoma, and in other chronic lung diseases [18, 43, 50]. However, in our view, the propensity of airway cells to survive *KRAS* mutations during early carcinogenesis, the close airway-proximity of lung tumors revealed by μCT and histology, as well as the fact that CCSP-labeled cells did not express the CCSP marker anymore, support a bronchial origin of these tumors. This view is in line with recent evidence for tobacco smoke-induced epigenetic changes that sensitize human airway epithelial cells to a single *KRAS* mutation [51]. Along these lines, the split genetic markings of chemical-induced lung adenocarcinomas of GFP;LYZ2.CRE mice indicates that LYZ2-labeled alveolar cells are dispensable for environmental lung adenocarcinoma, as opposed to what was previously shown for genetically-triggered lung adenocarcinoma [18].

Our approach focused on the integral assessment of changes in lung epithelial kinetics and transcriptome signatures during aging, injury, and carcinogenesis. The perpetual cell labeling approach we adopted was preferred over pulsed lineage tracing models because of the unprecedented accuracy of our CCSP.CRE strain in exclusively and completely labeling airway epithelial cells at the conclusion of development, allowing tracking of subsequent changes in adulthood. The identification of transcriptional programs that are activated during lung repair and carcinogenesis are of great importance for lung biology and are likely to lead to therapeutic innovations [52]. To this end, insertions and deletions in lineage-restricted genes were recently shown to occur in human lung adenocarcinoma [11]. Moreover, integrin β_3 and TANK-binding kinase 1 partner with oncogenic *KRAS* signaling to mediate cancer stemness and drug resistance [9, 10]. Along these lines, our findings of the involvement of airway epithelial cells in lung maintenance, repair, and carcinogenesis imply that at least

some of these cells present lung stem cells with regenerative and malignant potential and thus marked therapeutic targets. This was evident in our hands by the facts that airway epithelial cells could maintain adult injured alveoli and sustain *KRAS* mutations induced by urethane.

In conclusion, airway cells contribute to alveolar maintenance and lung carcinogenesis in response to environmental challenges. Since defective epithelial repair underlies the pathogenesis of chronic lung diseases and since abundantly transcribed genes are central to the mutational processes that cause cancer, this finding is of potential therapeutic importance for chronic pulmonary diseases and lung cancer.

353 # MATERIALS AND METHODS

354 ## Key Resources Table

Reagent type (species) or resource	Designation	Source or reference	Identifiers	Additional information
strain, strain background (*Mus musculus*)	C57BL/6	Jackson Laboratory	Stock #: 000664; RRID:IMSR_JAX:000664	
strain, strain background (*M. musculus*)	FVB	Jackson Laboratory	Stock #: 001800; RRID:IMSR_JAX:001800	
genetic reagent (*M. musculus*)	TOMATO	Jackson Laboratory	Stock #: 007676; RRID:IMSR_JAX:007676	PMID: 17868096
genetic reagent (*M. musculus*)	LUC	Jackson Laboratory	Stock #: 005125; RRID:IMSR_JAX:005125	PMID: 14717328
genetic reagent (*M. musculus*)	DTA	Jackson Laboratory	Stock #: 009669; RRID:IMSR_JAX:009669	PMID: 18354198
genetic reagent (*M. musculus*)	LYZ2.Cre	Jackson Laboratory	Stock #: 004781; RRID:IMSR_JAX:004781	PMID: 10621974
genetic reagent (*M. musculus*)	SOX2.Cre	Jackson Laboratory	Stock #: 008454; RRID:IMSR_JAX:008454	PMID: 14516668
genetic reagent (*M. musculus*)	VAV.Cre	Jackson Laboratory	Stock #: 008610; RRID:IMSR_JAX:008610	PMID: 9427694
genetic reagent (*M. musculus*)	NES.Cre	Jackson Laboratory	Stock #: 003771; RRID:IMSR_JAX:003771	PMID: 10471508
genetic reagent (*M. musculus*)	CCSP.Cre	European Mouse Mutant Archive	Stock #: EM:04965; RRID:IMSR_M231009	PMID: 22744859
genetic reagent (*M. musculus*)	SFTPC.Cre	Mouse Genome Informatics	RRID:MGI:3574949	PMID: 15716345

cell line (*M. musculus*)	LUAD cells	PMID: 30828726		Derived from urethane models
biological sample (*Homo sapiens*)	Lung adenocarcinomas	PMID: 26147201		Archival samples of patients with LUAD
antibody	rabbit poyclonal anti-PCNA	Abcam	Cat. #: ab2426; RRID:AB_303062	IHC (1:3000)
antibody	rabbit monoclonal anti-LYZ2	Abcam	Cat. #: ab108508; RRID:AB_10861277	IF (1:50)
antibody	rabbit polyclonal anti-KRT5	Abcam	Cat. #: ab53121; RRID:AB_869889	IF (1:200)
antibody	rabbit polyclonal anti-SFTPC	Santa Cruz Biotechnology	Cat. #: sc-13979; RRID:AB_2185502	IF (1:200)
antibody	rabbit polyclonal anti-CCSP	Santa Cruz Biotechnology	Cat. #: sc-25555; RRID:AB_2269914	IF (1:200)
antibody	goat polyclonal anti-CCSP	Santa Cruz Biotechnology	Cat. #: sc-9772; RRID:AB_2238819	IF (1:1000)
antibody	mouse monoclonal anti-acetylated α-tubulin	Sigma-Aldrich	Cat. #: T7451; RRID:AB_609894	IF (1:2000)
antibody	rabbit polyclonal anti-SFTPC	Merck-Millipore	Cat. #: AB3786; RRID:AB_91588	IF (1:500)
antibody	mouse monoclonal anti-KRT5 MA5-17057,	Thermo Fisher Scientific	Cat. #: MA5-17057; RRID:AB_2538529	IF (1:200)
antibody	mouse monoclonal anti-CD45 FITC conjugated	eBioscience	Cat. #: 11-0451-85; RRID:AB_465051	FC (0,05 μg)
antibody	mouse monoclonal anti-CD11b PE conjugated	eBioscience	Cat. #: 12-0112-82; RRID:AB_2734869	FC (0,05 μg)
antibody	donkey polyclonal anti-rabbit Alexa Fluor® 488	Molecular Probes	Cat. #: A21206; RRID:AB_141708	IF (1:500)
antibody	donkey polyclonal anti-goat Alexa Fluor® 568	Molecular Probes	Cat. #: A11057; RRID:AB_142581	IF (1:500)
antibody	donkey polyclonal anti-rabbit Alexa Fluor® 647	Molecular Probes	Cat. #: A31573; RRID:AB_2536183	IF (1:500)

antibody	donkey polyclonal anti-mouse Alexa Fluor® 647	Molecular Probes	Cat. #: A31571; RRID:AB_162542	IF (1:500)
antibody	donkey polyclonal anti-mouse Alexa Fluor® 568	Abcam	Cat. #: ab175700	IF (1:500)
sequence-based reagent	Digital droplet PCR primers	This paper	$Kras^{Q61R}$ mutation detection	Forward: ATCTGACGTGCTTTGCCTGT, Reverse: CCCTCCCCAGTTCTCATGTA
sequence-based reagent	Digital droplet PCR probe	This paper	$Kras^{Q61R}$ mutation detection	sequence: GACACAGCAGGTCAAGAGGAGTACA
sequence-based reagent	Digital droplet PCR primers and probe	Bio-Rad Laboratories	Registration #: dCNS685684912	Tomato allele detection
sequence-based reagent	Quantitative PCR	This paper	*Scgb1a1* gene	Forward: ATCACTGTGGTCATGCTGTCC, Reverse: GCTTCAGGGATGCCACATAAC
sequence-based reagent	Quantitative PCR	This paper	*Sftpc* gene	Forward: TCGTTGTCGTGGTGATTGTAG, Reverse: TCGTTGTCGTGGTGATTGTAG
sequence-based reagent	Quantitative PCR	This paper	*Gusb* gene	Forward: TTACTTTAAGACGCTGATCACC, Reverse: ACCTCCAAATGCCCATAGTC
commercial assay or kit	GenElute Mammalian Genomic DNA Minipreps Kit	Sigma-Aldrich	Cat. #: G1N70	
commercial assay	RNeasy Mini Kit	Qiagen	Cat. #: 74106	

or kit				
commercial assay or kit	SYBR FAST qPCR Kit	Kapa Biosystems	Cat. #: KK4600	
commercial assay or kit	MycoAlert Mycoplasma Detection Kit	LONZA	Cat. #: LT07-318	
chemical compound, drug	Urethane, ethyl carbamate (EC)	Sigma-Aldrich	Cat. #: U2500	1 g/Kg
chemical compound, drug	3-methylcholanthrene (MCA)	Sigma-Aldrich	Cat. #: 442388	15 mg/Kg
chemical compound, drug	Butylated hydroxytoluene (BHT)	Sigma-Aldrich	Cat. #: W218405	200 mg/Kg
chemical compound, drug	Naphthalene	Sigma-Aldrich	Cat. #: 84679	250 mg/Kg
chemical compound, drug	Bleomycin A2	Calbiochem	Cat. #: 203401	0.08 units
software, algorithm	Transcriptome Analysis Console Software	https://www.thermofisher.com/tw/zt/home/life-science/microarray-analysis/microarray-analysis-instruments-software-services/microarray-analysis-software/affymetrix-transcriptome-analysis-console-software.html	RRID:SCR_016519	
software, algorithm	FlowJo software	TreeStar	RRID:SCR_008520	
software, algorithm	FloMax Software	Partec	RRID:SCR_014437	
software, algorithm	Broad Institute pre-ranked GSEA module software	http://software.broadinstitute.org/gsea/index.jsp		PMID: 16199517
software, algorithm	NRECON software	Bruker		
software, algorithm	CT analysis (Ctan) software	Bruker		
software, algorithm	CTVox software	Bruker		

software, algorithm	QuantaSoft	Bio-Rad Laboratories (http://www.bio-rad.com/en-gr/sku/1864011-quantasoft-software-regulatory-edition?ID=1864011)		
software, algorithm	G*power	http://www.gpower.hhu.de/	RRID:SCR_013726	PMID: 17695343
software, algorithm	GraphPad Prism	http://www.graphpad.com/	RRID:SCR_002798	Version 8
software, algorithm	Fiji	http://fiji.sc	RRID:SCR_002285	PMID: 22743772
software, algorithm	Living Image software	Perkin-Elmer (http://www.perkinelmer.com/catalog/category/id/living%20image%20software)	RRID:SCR_014247	Version 4.2
other	Microarray data	This paper	Gene Expression Omnibus (GEO) accession ID: GSE94981	LUAD cells, bone marrow derived macrophages (BMDM), and tracheal AEC cells
other	Microarray data	Gene Expression Omnibus (GEO)	Accession ID: GSE82154; GSE55459; GSE46749; GSE18816; GSE43458	*M. musculus* ATII cells; *H. sapiens* AEC cells; *H. sapiens* ATII cells; *H. sapiens* AMΦ; *H. sapiens* non-smokers lung and LUAD
other	GeneChip Mouse Gene 2.0 ST array; GeneChip Human Gene 1.0 ST array	Thermo Fisher Scientific	Cat. #: 902119; Cat. #: 901085	
other	Hoechst33258 nuclear dye	Sigma-Aldrich	Cat. #: 14530	1:5000
other	D-Luciferin potassium salt	Gold Biotechnology	Cat. #: LUCK-100	1 mg
other	Trizol	Thermo Fisher Scientific	Cat. #: 15596026	

Key Resources Table

All raw data used to generate the main Figures and Figure Supplements are provided as *.xlsx Source Data files.

Study approval

All mice were bred at the Center for Animal Models of Disease of the University of Patras. Experiments were designed and approved *a priori* by the Veterinary Administration of the Prefecture of Western Greece (approval numbers 3741/16.11.2010, 60291/3035/19.03.2012, and 118018/578/30.04.2014) and were conducted according to Directive 2010/63/EU (http://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1486710385917&uri=CELEX:32010L0063). Male and female experimental mice were sex-, weight (20-25 g)-, and age (6-12 week)-matched. *n*= 588 experimental and *n* = 165 breeder mice were used for this report. Sample size was calculated using power analysis on G*power. Experiments were randomized across different cages and mouse lungs were always examined by two blinded researchers. Sample numbers are included in the figures and figure legends. Archival tissue samples of patients with LUAD [35] that underwent surgical resection with curative intent between 2001 and 2008 at the University Hospital of Patras were retrospectively enrolled. The observational protocol for these studies adhered to the Helsinki Declaration and was approved by the Ethics Committee of the University Hospital of Patras, and all patients gave written informed consent.

Reagents

Urethane, ethyl carbamate, EC, CAS# 51-79-6; 3-methylcholanthrene, 3-methyl-1,2-dyhydrobenzo[j]aceanthrylene, MCA, CAS# 56-49-5; butylated hydroxytoluene, 2,6-Di-tert-butyl-4-methylphenol, BHT, CAS# 128-37-0; naphthalene, CAS# 91-20-3, and Hoechst33258 nuclear dye (CAS# 23491-45-4), were from Sigma-Aldrich (St. Louis, MO). Bleomycin A2, ((3-{[(2'-{(5S,8S,9S,10R,13S)-15-{6-amino-2- [(1S)-3-amino-1-{[(2S)-2,3-diamino-3-oxopropyl]amino}-3-oxopropyl] -5-methylpyrimidin-4-yl}-13-[{[(2R,3S,4S,5S,6S)-3-{[(2R,3S,4S,5R,6R)-4-(carbamoyloxy)-3,5-dihydroxy-6-(hydroxymethyl) tetrahydro-2H-pyran-2-yl]oxy} -4,5-dihydroxy-6-(hydroxymethyl)

tetrahydro-2H-pyran-2-yl]oxy} (1H-imidazol-5-yl)methyl]-9-hydroxy-5-[(1R)-1-hydroxyethyl]-8,10-dimethyl-4,7,12,15-tetraoxo-3,6,11,14-tetraazapentadec-1-yl}-2,4'-bi-1,3-thiazol-4-yl)carbonyl]amino}propyl)(dimethyl)sulfonium; CAS #9041-93-4, was from Calbiochem (Darmstadt, Germany). D-Luciferin potassium salt, (4S)-2-(6-hydroxy-1,3-benzothiazol-2-yl)-4,5-dihydrothiazole-4-carboxylic acid, CAS #2591-17-5, was from Gold Biotechnology (St. Louis, MO).

Experimental mice

C57BL/6J (C57BL/6; #000664), FVB/NJ (FVB; #001800), *B6.129(Cg)-Gt(ROSA)26Sor*$^{tm4(ACTB\text{-}tdTomato,\text{-}EGFP)Luo}$*/J* [mT/mG; TOMATO; #007676; [25]], *FVB.129S6(B6)-Gt(ROSA)26Sor*$^{tm1(Luc)Kael}$*/J* [LUC; #005125; [34]], *B6.129P2-Gt(ROSA)26Sor*$^{tm1(DTA)Lky}$*/J* [DTA; #009669; [42]], *B6.129P2-Lyz2*$^{tm1(cre)Ifo}$*/J* [LYZ2.CRE; #004781; [18]], *B6.Cg-Tg(Sox2-cre)1Amc/J* [SOX2.CRE; #008454; [28]], *B6.Cg-Tg(Vav1-icre)A2Kio/J* [VAV.CRE; #008610; [29]], and *B6.Cg-Tg(Nes-cre)1Kln/J* [NES.CRE; #003771; [30]] mice were from Jackson Laboratories (Bar Harbor, MN). *B6;CBA-Tg(Scgb1a1-cre)1Vart/Flmg* (CCSP.CRE; European Mouse Mutant Archive #EM:04965) mice are described elsewhere [26] and *Tg(Sftpc-cre)1Blh* (SFTPC.CRE; Mouse Genome Informatics #MGI:3574949) mice were donated by their founder [27]. Mice were bred > F12 to the FVB background at the University of Patras Center for Animal Models of Disease.

Mouse models of lung adenocarcinoma

Six-week-old mice on the C57BL/6 background received ten consecutive weekly intraperitoneal urethane injections (1 g/Kg in 100 μL saline) and were sacrificed 6-7 months after the first injection, or four consecutive weekly intraperitoneal MCA (15 mg/Kg in 100 μL saline) followed by eight consecutive weekly intraperitoneal BHT injections (200 mg/Kg in 100 μL corn oil) and were sacrificed 6-7 months after the first injection. Six-week-old

mice on the FVB background received one intraperitoneal urethane injection (1 g/Kg in 100 μL saline) and were sacrificed 6-7 months later [20-24].

Mouse models of lung injury

Six-week-old mice (C57BL/6 background) received intratracheal bleomycin A2 (0.08 units in 50 μL saline) or intraperitoneal naphthalene (250 mg/Kg in 100 μL corn oil) [40, 41]. In addition, preterm mothers of the C57BL/6 background and their offspring were exposed to room air (21% oxygen; control) or 98% oxygen for two days before and four days after birth [41, 53]. Oxygen levels were continuously monitored. The gas stream was humidified to 40–70% by a deionized water-jacketed Nafion membrane tubing and delivered through a 0.22 μm filter before passage into a sealed Lexan polycarbonate chamber measuring 40 x 25 x 25 cm and accommodating 25 L gas at a flow rate of 5 L/min, resulting in complete gas exchange every 5 min. Mothers were cycled between litters on 21% and 98% oxygen every 24 hours to prevent oxygen toxicity and to control for nutritional support of the pups. After perinatal hyperoxia, mice remained at room air till sacrificed at eight weeks of age.

Urethane-induced lung adenocarcinoma cell lines

Lung tumors were dissected from surrounding healthy lung parenchyma under sterile conditions, minced into 1-mm pieces, and cultured at 37^0 C in 5% CO2-95% air using Dulbecco's Modified Eagle Medium (DMEM), 10% FBS, 2 mM L-glutamine, 1 mM pyruvate, 100 U/mL penicillin, and 100 U/mL streptomycin. All cell lines were immortal and indefinitely phenotypically stable over > 18 months and/or 60 passages, and were tumorigenic and metastatic in C57BL/6 mice [32]. Cell lines were cultured in DMEM supplemented with 10% FBS and 100 IU/mL penicillin/streptomycin and were maintained in humidified incubators at 37 °C with 95% air–5% CO_2. Cell lines were authenticated annually

using the short tandem repeat method and were tested negative for *Mycoplasma Spp.* biannually by MycoAlert Mycoplasma Detection Kit (LONZA; Verviers, Belgium).

Human lung adenocarcinomas

Ten archival formalin-fixed, paraffin-embedded tissue samples of patients with LUAD that underwent surgical resection with curative intent between 2001 and 2008 at the University Hospital of Patras were retrospectively enrolled [35]. The observational protocol for these studies adhered to the Helsinki Declaration and was approved by the Ethics Committee of the University Hospital of Patras, and all patients gave written informed consent.

Micro-computed tomography

Urethane or saline treated FVB mice were sacrificed six months post urethane/saline injection. Lungs were inflated and fixed with 10% neutral buffered formalin overnight. They were then dehydrated and chemically dried for μCT scanning using a method kindly provided by Jeroen Hostens (Bruker; Kontich, Belgium). Briefly, a gradient ethanol dehydration protocol (from 70-100%) was applied, followed by 2 hour incubation in Hexamethyldisilazane (HMDS; Sigma, St. Louis, MO) and 2 hour air-drying. The dehydrated lungs were then scanned in a Bruker SkyScan 1172 scanner at 41kV without filtration and with 5.94 μm voxel resolution (exposure: 440 ms). The X-ray projections were obtained at 0.35° intervals with a scanning angular rotation of 180° and two frames were averaged for each rotation under a mean of 10 frames per random movement. 3D reconstructions were performed using NRECON software (Bruker). Regions of interest for the whole lung and peripheral lung tissue were defined in the CT analysis software (CTan; Bruker), thresholds applied to detect tissue from background, and a 3D volume rendering of the lungs were performed using the CTVox software (Bruker).

Structural assessments in murine lungs

Mouse lungs were recoded (blinded) by laboratory members not participating in these studies and were always examined by two independent blinded participants of this study. The results obtained by each investigator were compared, and lungs were re-evaluated if deviant by > 20%. Lungs and lung tumors were initially inspected macroscopically under a Stemi DV4 stereoscope equipped with a micrometric scale incorporated into one eyepiece and an AxiocamERc 5s camera (Zeiss, Jena, Germany) in trans-illumination mode, allowing for visualization of both superficial and deeply-located lung tumors [23, 24]. Tumor location was charted and diameter (δ) was measured. Tumor number (multiplicity) per mouse was counted and mean tumor diameter per mouse was calculated as the average of individual diameters of all tumors found in a given mouse lung. Individual tumor volume was calculated as $\pi\delta^3/6$. Mean tumor volume per mouse was calculated as the average of individual volumes of all tumors found in a given mouse lung, and total lung tumor burden per mouse as their sum. Following macroscopic mapping of lung and lung tumor morphology, lungs of fluorescent reporter mice were imaged on a Leica MZ16F fluorescent stereomicroscope equipped with GFP and RFP filters and a DFC 300FX camera (Leica Microsystems, Heidelberg, Germany) in order to determine their macroscopic fluorescent pattern. Lung volume was measured by saline immersion, and lungs were embedded in paraffin, randomly sampled by cutting 5 μm-thick lung sections (n = 10/lung), mounted on glass slides, and stained with hematoxylin and eosin for morphometry and histologic typing of lung tumors. For this, a digital grid of 100 intersections of vertical lines (points) was superimposed on multiple digital images of all lung sections from lung tissue of a given mouse using Fiji academic freeware (https://fiji.sc/). Total lung tumor burden was determined by point counting of the ratio of the area occupied by neoplastic lesions versus total lung area and by extrapolating the average ratio per mouse to total lung volume [54]. The results of this stereologic approach were compared with the macroscopic method, and were scrutinized if deviant by > 20%. To evaluate alveolar

structure and size, we calculated mean linear intercept using randomly sampled hematoxylin and eosin-stained lung sections, as described elsewhere [54]. For this, a digital grid of twenty random horizontal lines was superimposed on multiple digital images of all lung sections from lung tissue of a given mouse using Fiji. Mean linear intercept was calculated by counting the intercepts of interalveolar septae with the lines and the formula: Σ{2 x (length of line/ number of intercepts)}/total number of lines. All quantifications were done by counting at least five random non-overlapping fields of view of at least ten sections per lung.

Histology and molecular phenotyping

For histology, lungs were inflated to 20 cmH_2O pressure that provides for a lung volume equivalent to the resting volume of the lungs (a.k.a. functional residual capacity in humans) and enables precise histologic observations on airway and alveolar structure avoiding false interpretations resulting from the study of compressed or over-inflated lungs [54]. Subsequently, lungs were fixed with 10% formalin overnight and were embedded in paraffin. Five-µm-thick paraffin sections were then counterstained with hematoxylin and eosin (Sigma, St. Louis, MO) and mounted with Entellan New (Merck Millipore, Darmstadt, Germany). For immunofluorescence, lungs were inflated with a 2:1 mixture of 4% paraformaldehyde:Tissue-Tek (Sakura, Tokyo, Japan), fixed in 4% paraformaldehyde overnight at 4°C, cryoprotected with 30% sucrose, embedded in Tissue-Tek and stored at -80°C. Ten-µm cryosections were then post-fixed in 4% paraformaldehyde for 10 min, treated with 0.3% Triton X-100 for 5 min, and incubated in blocking solution containing 10% fetal bovine serum (FBS), 3% bovine serum albumin (BSA), 0.1% polyoxyethylene (20) sorbitanmonolaurate (Tween 20) in 1x phosphate-buffered saline (PBS) for 1 hour. Following labeling with the indicated primary antibodies overnight at 4°C, sections were incubated with fluorescent secondary antibodies, counterstained with Hoechst 33258 and mounted with Mowiol 4-88 (Calbiochem, Darmstadt, Germany). The following primary antibodies were

used: rabbit anti-proliferating cell nuclear antigen (PCNA, 1:3000 dilution, ab2426, Abcam, London, UK), rabbit anti-LYZ2 (1:50 dilution, ab108508, Abcam), rabbit anti-KRT5 (1:200 dilution, ab53121, Abcam), rabbit anti-SFTPC (1:200 dilution, sc-13979, Santa Cruz, Dallas, TX), rabbit anti-CCSP (1:200 dilution, sc-25555, Santa Cruz), goat anti-CCSP (1:1000 dilution, sc-9772, Santa Cruz), mouse anti-acetylated α-tubulin (1:2000 dilution, T7451, Sigma-Aldrich, St. Lewis, MO), rabbit anti-SFTPC (1:500 dilution, AB3786, Merck-Millipore, Burlington, MA), and mouse anti-KRT5 (1:200 dilution, MA5-17057, Thermo Fisher Scientific, Waltham, MA). Alexa Fluor donkey anti-rabbit 488 (A21206, Thermo Fisher Scientific), Alexa Fluor donkey anti-mouse 568 (ab175700, Abcam), Alexa Fluor donkey anti-goat 568 (A11057, Thermo Fisher Scientific), Alexa Fluor donkey anti-rabbit 647 (A31573, Thermo Fisher Scientific), and Alexa Fluor donkey anti-mouse 647 (A31571, Thermo Fisher Scientific) secondary antibodies were used at 1:500 dilution. For isotype control, the primary antibody was omitted. Bright-field images were captured with an AxioLab.A1 microscope connected to an AxioCamERc 5s camera (Zeiss, Jena, Germany) whereas fluorescent microscopy was carried out either on an Axio Observer D1 inverted fluorescent microscope (Zeiss, Jena, Germany) or a TCS SP5 confocal microscope (Leica Microsystems, Wetzlar, Germany) with 20x, 40x and 63x lenses. Digital images were processed with Fiji. All quantifications of cellular populations were obtained by counting at least five random non-overlapping bronchial-, alveolar-, hyperplasia-, or tumor- containing fields of view per section.

Pulmonary function testing

Following anesthesia induced by intraperitoneal ketamine (100 mg/Kg) and xylazine (10 mL/Kg) and tracheostomy, mice were mechanically ventilated by a Flexivent rodent ventilator (Scireq, Montreal, Ontario, Canada). The whole procedure, described elsewhere [55], lasted 15 min. After a 3-min run-in period of ventilation with 21% oxygen, a tidal

volume of 10 mL/Kg, a respiratory rate of 150 breaths/min, and a positive end-expiratory pressure of 3 cmH_2O, paralysis was induced using 8 mg/Kg intraperitoneal succinyl choline, and total respiratory system impedance was obtained by applying an 8-sec-long pseudorandom frequency oscillation (0.5-19.75 Hz) to the airway opening. Thirty seconds prior to initiation of measurements, lung volume history was once controlled by a 6-sec-long inflation to 30 cm H_2O pressure. Measurements were repeated thrice at 60 sec intervals and were averaged. Data were fit into the constant phase model in order to fractionate total respiratory input impedance into airways resistance (Raw) and tissue damping and elastance coefficients. To obtain pressure-volume (PV) curves, the respiratory system was incrementally inflated and deflated to 40 mL/Kg total volume at seven steps each and airway pressures were recorded on each volume change. The slope of the linear portion of expiratory PV curves, which represents static compliance (Cst), a measure of airspace function, was calculated manually. Operators were blinded to animal genotype.

Digital droplet (dd)PCR

TOMATO, GFP;CCSP.CRE, and GFP;LYZ2.CRE mice (FVB strain) received one intraperitoneal injection of urethane (1 g/Kg) and lungs were then harvested one and two weeks post-urethane, homogenized, and subjected to DNA extraction and purification using GenElute Mammalian Genomic DNA Miniprep Kit (Sigma-Aldrich, St. Louis, MO). DNA concentration and quality were assessed using a Nanodrop 1000 spectrophotometer (Thermo Fisher Scientific,Waltham, MA). DNA concentration was converted to number of diploid copies according to the formula: DNA (ng/µL) / weight of mouse diploid genome (3.9 pg). Digital droplet PCR protocol and analysis was performed as described previously using reagents, equipment and software from BioRad Laboratories Inc. (Hercules, CA) [56]. In brief, 20000 genome copies were used. Samples were normalized internally according to the number of accepted droplets and inter-sample normalization was performed according to the

formula [x-min(x)]/[max(x)-min(x)],where x represents the actual, min(x) the minimum, and max(x) the maximum number of accepted droplets. The data were reported as % positive/accepted droplets. Sequences of *Kras*^{Q61R} primers and probe were: *Kras*^{Q61R} forward: ATCTGACGTGCTTTGCCTGT, *Kras*^{Q61R} reverse: CCCTCCCCAGTTCTCATGTA, and *Kras*^{Q61R} probe: GACACAGCAGGTCAAGAGGAGTACA. The *Rosa*^{mT} assay is registered as dCNS685684912 (Bio-Rad) with MIQE context: seq1:195-315:+ CCAGTTCATGTACGGCTCCAAGGCGTACGTGAAGCACCCCGCCGACATCCCCGAT TACAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTC GAGGACGGCGGTCT. Primers and fluorescently labeled probes were combined in a mixture containing 18 µM forward and reverse primers and 5 µM labeled probes (20x primer/Taqman probe mix). Reactions were assembled to contain 12.5 µL 2x ddPCR mix no-UTP, 1.25 µL 20x *Kras*^{Q61R}primer/Taqman probe Mix, 1.25 µL 20x *Rosa*^{mT} custom primer/Taqman probe Mix and 10 µL DNA diluted in nuclease-free water. The ddPCR protocol included a first denaturation step at 95°C for 10 min followed by 40 cycles of denaturation at 95°C for 30 sec and 40 cycles of annealing at 62.5°C for 60 sec, and was performed in a BioRad T100 Thermal cycler. Results were analyzed with a BioRad QX100 droplet reader using the QuantaSoft software. The amplitude gathering thresholds of positive droplets were set at 3500 for the *Rosa*^{mT} and at 10000 for the *Kras*^{Q61R} probe, according to the manufacturer's instructions.

Bronchoalveolar lavage (BAL)

BAL was performed using three sequential aliquots of 1000 µL sterile ice-cold phosphate-buffered saline (PBS). Fluid was combined and centrifuged at 260 g for 10 min to separate cells from supernatant. The cell pellet was resuspended in 1 ml PBS containing 2% fetal bovine serum, and the total cell count was determined using a grid hemocytometer according to the Neubauer method. Cell differentials were obtained by counting 400 cells on May-

Grünwald-Giemsa-stained cytocentrifugal specimens. Total BAL cell numbers were calculated by multiplying the percentage of each cell type by total BAL cell number [23, 24].

Bioluminescence imaging

LUC;CCSP.CRE mice, bioluminescent reporters of CCSP-labeled cell mass, received one intraperitoneal injection of saline (100 μL saline) or urethane (1g/Kg in 100 μL saline) and were serially imaged before treatment start, and at 150 and 210 days into treatment. Imaging was done on a Xenogen Lumina II (Perkin-Elmer, Waltham, MA) 5-20 min after delivery of 1 mg D-Luciferin sodium in 100 μL of sterile water to the retro-orbital vein, and data were analyzed using Living Image v.4.2 (Perkin-Elmer, Waltham, MA) [23, 24].

qPCR and microarrays

Triplicate cultures of 10^6 LUAD cells, BMDM (obtained by 1-week bone marrow incubation with 100 ng/mL M-CSF), and tracheal AEC (obtained by 1-week incubation of stripped mouse tracheal epithelium in DMEM) were subjected to RNA extraction using Trizol (Thermo Fisher) followed by column purification and DNA removal (Qiagen, Hilden, Germany). Whole lungs were homogenized in Trizol followed by the same procedure. Pooled RNA (5 μg) was quality tested (ABI 2000 Bioanalyzer; Agilent Technologies, Sta. Clara, CA), labeled, and hybridized to GeneChip Mouse Gene 2.0 ST arrays (Affymetrix, Sta. Clara, CA). All data were deposited at GEO (http://www.ncbi.nlm.nih.gov/geo/; Accession ID: GSE94981) and were analyzed on the Affymetrix Expression and Transcriptome Analysis Consoles together with previously reported [43-47] murine ATII and human AEC, ATII, AMΦ, non-smokers lung, and LUAD microarray data (Accession IDs: GSE82154, GSE55459, GSE46749, GSE18816, GSE43458). qPCR was performed using first strand synthesis with specific primers (*Scgb1a1*: ATCACTGTGGTCATGCTGTCC and GCTTCAGGGATGCCACATAAC; *Sftpc*: TCGTTGTCGTGGTGATTGTAG and

AGGTAGCGATGGTGTCTGCT; *Gusb*: TTACTTTAAGACGCTGATCACC and ACCTCCAAATGCCCATAGTC) and SYBR FAST qPCR Kit (Kapa Biosystems, Wilmington, MA) in a StepOne cycler (Applied Biosystems, Carlsbad, CA). Ct values from triplicate reactions were analyzed with the $2^{-\Delta CT}$ method relative to *Gusb*.

Flow cytometry

BAL cells were suspended in 50 µL PBS with 2% FBS and 0.1% NaN_3, were stained with anti-CD45 (#11-0451-85; eBioscience; Santa Clara, CA) and anti-CD11b (#12-0112-82; eBioscience; Santa Clara, CA) primary antibodies for 20 min in the dark at 0.5 µL antibody per million cells, and were analyzed on a CyFlowML cytometer with a sorter module using FloMax Software (Partec, Darmstadt, Germany) or FlowJo software (TreeStar, Ashland, OR), as described previously [32]. Perfused lungs were digested in RPMI-1640 medium containing collagenase XI (0.7 mg/mL; Sigma, St. Louis, MO) and type IV bovine pancreatic DNase (30 µg/mL; Sigma, St. Louis, MO) to obtain single-cell suspensions. After treatment with red blood cell lysis buffer (BioLegend; San Diego, CA), single-cell suspensions were analyzed on a LSR II flow cytometer (BD Bioscience, San Diego, CA), and data were examined with FlowJo. Dead cells were excluded using 4,6-diamidino-2-phenylindole (DAPI; Sigma, St. Louis, MO).

Microarray and gene set enrichment analyses (GSEA)

GSEA was performed with the Broad Institute pre-ranked GSEA module software (http://software.broadinstitute.org/gsea/index.jsp) [57]. In detail, genes significantly expressed (log2 normalized expression > 8) in murine tracheal airway cells, ATII cells [43], and BMDM were cross-examined against the murine lung and chemical-induced LUAD cell line transcriptomes. In addition, previously reported human AEC, ATII, and AMΦ cellular

signatures [45-47] were cross-examined against the previously described transcriptomes of human normal lung tissue from never-smokers and of LUAD [44].

Statistical analysis

Sample size was calculated using power analysis on G*power (http://www.gpower.hhu.de/), assuming $\alpha = 0.05$, $\beta = 0.05$, and effect size $d = 1.5$ [58]. No data were excluded from analyses. Animals were allocated to treatments by alternation and transgenic animals were enrolled case-control-wise. Data were collected by at least two blinded investigators from samples coded by non-blinded investigators. All data were normally distributed by Kolmogorov-Smirnov test, are given as mean ± SD, and sample size (n) always refers to biological and not technical replicates. Differences in frequency were examined by Fischer's exact and χ^2 tests and in means by t-test or one-way ANOVA with Bonferroni post-tests. Changes over time and interaction between two variables were examined by two-way ANOVA with Bonferroni post-tests. All probability (P) values are two-tailed and were considered significant when $P<0.05$. All analyses and plots were done on Prism v8.0 (GraphPad, La Jolla, CA).

ACKNOWLEDGMENTS

The authors thank the University of Patras Centre for Animal Models of Disease and Advanced Light Microscopy Facility for experimental support. MA and FR also wish to thank the InfrafrontierGR infrastructure (co-financed by the ERDF and NSRF 2007-2013) for supporting the μCT facility.

COMPETING INTERESTS

The authors declare no competing interests.

DATA AVAILABILITY

All raw data produced in this study are provided as *.xlsx source data supplements. The microarray data produced by this study were deposited at GEO (http://www.ncbi.nlm.nih.gov/geo/; Accession ID: GSE94981). Previously reported murine ATII and human AEC, ATII, AMΦ, non-smokers lung, and LUAD microarray data are available at GEO using Accession IDs GSE82154, GSE55459, GSE46749, GSE18816, and GSE43458).

FIGURES & FIGURE SUPPLEMENTS

This dataset contains 7 Figures, 1 Table, and 38 Figure Supplements. In detail, Figure 1 is linked with 19 Figure Supplements, Figure 2 is linked with 1 Figure Supplement, Figure 3 is linked with 6 Figure Supplements, Figure 4 is linked with 4 Figure Supplements, Figure 5 is linked with 2 Figure Supplements, and Figure 6 is linked with 6 Figure Supplements,

REFERENCES

1. Barnes PJ, Burney PG, Silverman EK, Celli BR, Vestbo J, Wedzicha JA, Wouters EF. Chronic obstructive pulmonary disease. Nat Rev Dis Primers. 2015 Dec 3;1:15076. doi: 10.1038/nrdp.2015.76.

2. Lozano R, Naghavi M, Foreman K, Lim S, Shibuya K, Aboyans V, Abraham J, Adair T, Aggarwal R, Ahn SY, Alvarado M, Anderson HR, Anderson LM, Andrews KG, AtkinsonC, Baddour LM, Barker-Collo S, Bartels DH, Bell ML, Benjamin EJ, Bennett D,Bhalla K, Bikbov B, Bin Abdulhak A, Birbeck G, Blyth F, Bolliger I, Boufous S,Bucello C, Burch M, Burney P, Carapetis J, Chen H, Chou D, Chugh SS, Coffeng LE, Colan SD, Colquhoun S, Colson KE, Condon J, Connor MD, Cooper LT, Corriere M,Cortinovis M, de Vaccaro KC, Couser W, Cowie BC, Criqui MH, Cross M, DabhadkarKC, Dahodwala N, De Leo D, Degenhardt L, Delossantos A, Denenberg J, Des Jarlais DC, Dharmaratne SD, Dorsey ER, Driscoll T, Duber H, Ebel B, Erwin PJ, EspindolaP, Ezzati M, Feigin V, Flaxman AD, Forouzanfar MH, Fowkes FG, Franklin R, FransenM, Freeman MK, Gabriel SE, Gakidou E, Gaspari F, Gillum RF, Gonzalez-Medina D,Halasa YA, Haring D, Harrison JE, Havmoeller R, Hay RJ, Hoen B, Hotez PJ, Hoy D, Jacobsen KH, James SL, Jasrasaria R, Jayaraman S, Johns N, Karthikeyan G,Kassebaum N, Keren A, Khoo JP, Knowlton LM, Kobusingye O, Koranteng A,Krishnamurthi R, Lipnick M, Lipshultz SE, Ohno SL, Mabweijano J, MacIntyre MF,Mallinger L, March L, Marks GB, Marks R, Matsumori A, Matzopoulos R, Mayosi BM,McAnulty JH, McDermott MM, McGrath J, Mensah GA, Merriman TR, Michaud C, MillerM, Miller TR, Mock C, Mocumbi AO, Mokdad AA, Moran A, Mulholland K, Nair MN,Naldi L, Narayan KM, Nasseri K, Norman P, O'Donnell M, Omer SB, Ortblad K,Osborne R, Ozgediz D, Pahari B, Pandian JD, Rivero AP, Padilla RP, Perez-Ruiz F, Perico N, Phillips D, Pierce K, Pope CA 3rd, Porrini E, Pourmalek F, Raju M,Ranganathan D, Rehm JT, Rein DB, Remuzzi G, Rivara FP, Roberts T, De León FR,Rosenfeld LC, Rushton L, Sacco RL, Salomon JA, Sampson U, Sanman E, Schwebel DC, Segui-Gomez M, Shepard DS, Singh D, Singleton J, Sliwa K, Smith E, Steer A,Taylor JA, Thomas B, Tleyjeh IM, Towbin JA, Truelsen T, Undurraga EA,Venketasubramanian N, Vijayakumar L, Vos T, Wagner GR, Wang M, Wang W, Watt K,Weinstock MA, Weintraub R, Wilkinson JD, Woolf AD, Wulf S, Yeh PH, Yip P,Zabetian A, Zheng ZJ, Lopez AD, Murray CJ, AlMazroa MA, Memish ZA. Global andregional mortality from 235 causes of death for 20 age groups in 1990 and 2010: asystematic analysis for the Global Burden of Disease Study 2010. Lancet. 2012 Dec15;380(9859):2095-128. doi: 10.1016/S0140-6736(12)61728-0.

3. Spella M, Lilis I, Stathopoulos GT. Shared epithelial pathways to lung repair and disease. Eur Respir Rev. 2017 Jun 28;26(144). pii: 170048. doi: 10.1183/16000617.0048-2017.

4. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. CA Cancer J Clin. 2015 Mar;65(2):87-108. doi: 10.3322/caac.21262.

5. Forbes SA, Bindal N, Bamford S, Cole C, Kok CY, Beare D, Jia M, Shepherd R, Leung K, Menzies A, Teague JW, Campbell PJ, Stratton MR, Futreal PA. COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. Nucleic Acids Res. 2011 Jan;39(Database issue):D945-50. doi: 10.1093/nar/gkq929.

6. Hecht SS. Tobacco smoke carcinogens and lung cancer. J Natl Cancer Inst. 1999 Jul 21;91(14):1194-210.doi:10.1093/jnci/91.14.1194.

7. Campbell JD, Alexandrov A, Kim J, Wala J, Berger AH, Pedamallu CS, Shukla SA, Guo G, Brooks AN, Murray BA, Imielinski M, Hu X, Ling S, Akbani R, Rosenberg M,Cibulskis C, Ramachandran A, Collisson EA, Kwiatkowski DJ, Lawrence MS, WeinsteinJN, Verhaak RG, Wu CJ, Hammerman PS, Cherniack AD, Getz G; Cancer Genome AtlasResearch Network, Artyomov MN, Schreiber R, Govindan R, Meyerson M. Distinctpatterns of somatic genome alterations in lung adenocarcinomas and squamous cell carcinomas. Nat Genet. 2016 Jun;48(6):607-16. doi: 10.1038/ng.3564.

8. Cancer Genome Atlas Research Network. Comprehensive molecular profiling of lung adenocarcinoma. Nature. 2014 Jul 31;511(7511):543-50. doi: 10.1038/nature13385.

9. Barbie DA, Tamayo P, Boehm JS, Kim SY, Moody SE, Dunn IF, Schinzel AC, Sandy P, Meylan E, Scholl C, Fröhling S, Chan EM, Sos ML, Michel K, Mermel C, Silver SJ, Weir BA, Reiling JH, Sheng Q, Gupta PB, Wadlow RC, Le H, Hoersch S, Wittner BS, Ramaswamy S, Livingston DM, Sabatini DM, Meyerson M, Thomas RK, Lander ES, Mesirov JP, Root DE, Gilliland DG, Jacks T, Hahn WC. Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. Nature. 2009 Nov 5;462(7269):108-12. doi: 10.1038/nature08460.

10. Seguin L, Kato S, Franovic A, Camargo MF, Lesperance J, Elliott KC, Yebra M, Mielgo A, Lowy AM, Husain H, Cascone T, Diao L, Wang J, Wistuba II, Heymach JV, Lippman SM, Desgrosellier JS, Anand S, Weis SM, Cheresh DA. An integrin β_3-KRAS-RalB complex drives tumour stemness and resistance to EGFR inhibition. Nat Cell Biol. 2014 May;16(5):457-68. doi: 10.1038/ncb2953.

11. Imielinski M, Guo G, Meyerson M. Insertions and Deletions Target Lineage-Defining Genes in Human Cancers. Cell. 2017 Jan 26;168(3):460-472.e14. doi: 10.1016/j.cell.2016.12.025.

12. Zuo W, Zhang T, Wu DZ, Guan SP, Liew AA, Yamamoto Y, Wang X, Lim SJ, Vincent M, Lessard M, Crum CP, Xian W, McKeon F. p63(+)Krt5(+) distal airway stem cells are essential for lung regeneration. Nature. 2015 Jan 29;517(7536):616-20. doi: 10.1038/nature13903.

13. Kim CF, Jackson EL, Woolfenden AE, Lawrence S, Babar I, Vogel S, Crowley D, Bronson RT, Jacks T. Identification of bronchioalveolar stem cells in normal lung and lung cancer. Cell. 2005 Jun 17;121(6):823-35.doi:10.1016/j.cell.2005.03.032.

14. Cho HC, Lai CY, Shao LE, Yu J. Identification of tumorigenic cells in Kras(G12D)-induced lung adenocarcinoma. Cancer Res. 2011 Dec 1;71(23):7250-8. doi: 10.1158/0008-5472.CAN-11-0903.

15. Xu X, Rock JR, Lu Y, Futtner C, Schwab B, Guinney J, Hogan BL, Onaitis MW. Evidence for type II cells as cells of origin of K-Ras-induced distal lung adenocarcinoma. Proc Natl Acad Sci U S A. 2012 Mar 27;109(13):4910-5. doi: 10.1073/pnas.1112499109.

16. Sutherland KD, Song JY, Kwon MC, Proost N, Zevenhoven J, Berns A. Multiple cells-of-origin of mutant K-Ras-induced mouse lung adenocarcinoma. Proc Natl Acad Sci U S A. 2014 Apr 1;111(13):4952-7. doi: 10.1073/pnas.1319963111.

17. Mainardi S, Mijimolle N, Francoz S, Vicente-Dueñas C, Sánchez-García I, Barbacid M. Identification of cancer initiating cells in K-Ras driven lung adenocarcinoma. Proc Natl Acad Sci U S A. 2014 Jan 7;111(1):255-60. doi: 10.1073/pnas.1320383110.

18. Desai TJ, Brownfield DG, Krasnow MA. Alveolar progenitor and stem cells in lung development, renewal and cancer. Nature. 2014 Mar 13;507(7491):190-4. doi: 10.1038/nature12930.

19. Treutlein B, Brownfield DG, Wu AR, Neff NF, Mantalas GL, Espinoza FH, Desai TJ, Krasnow MA, Quake SR. Reconstructing lineage hierarchies of the distal lung epithelium using single-cell RNA-seq. Nature. 2014 May 15;509(7500):371-5. doi: 10.1038/nature13173.

20. Westcott PM, Halliwill KD, To MD, Rashid M, Rust AG, Keane TM, Delrosario R, Jen KY, Gurley KE, Kemp CJ, Fredlund E, Quigley DA, Adams DJ, Balmain A. The mutational landscapes of genetic and chemical models of Kras-driven lung cancer. Nature. 2015 Jan 22;517(7535):489-92. doi: 10.1038/nature13898.

21. Miller YE, Dwyer-Nield LD, Keith RL, Le M, Franklin WA, Malkinson AM. Induction of a high incidence of lung tumors in C57BL/6 mice with multiple ethyl carbamate injections. Cancer Lett. 2003 Aug 20;198(2):139-44.doi: 10.1016/S0304-3835(03)00309-4.

22. Malkinson AM, Koski KM, Evans WA, Festing MF. Butylated hydroxytoluene exposure is necessary to induce lung tumors in BALB mice treated with 3-methylcholanthrene. Cancer Res. 1997 Jul 15;57(14):2832-4. PMID: 9230183.

23. Stathopoulos GT, Sherrill TP, Cheng DS, Scoggins RM, Han W, Polosukhin VV, Connelly L, Yull FE, Fingleton B, Blackwell TS. Epithelial NF-kappaB activation promotes urethane-induced lung carcinogenesis. Proc Natl Acad Sci U S A. 2007 Nov 20;104(47):18514-9. doi: 10.1073/pnas.0705316104.

24. Vreka M, Lilis I, Papageorgopoulou M, Giotopoulou GA, Lianou M, Giopanou I,Kanellakis NI, Spella M, Agalioti T, Armenis V, Goldmann T, Marwitz S, Yull FE,Blackwell TS, Pasparakis M, Marazioti A, Stathopoulos GT. IκB Kinase α IsRequired for Development and Progression of KRAS-Mutant Lung Adenocarcinoma.Cancer Res. 2018 Jun 1;78(11):2939-2951. doi: 10.1158/0008-5472.CAN-17-1944.

25. Muzumdar MD, Tasic B, Miyamichi K, Li L, Luo L. A global double-fluorescent Cre reporter mouse. Genesis. 2007 Sep;45(9):593-605.doi: 10.1002/dvg.20335.

26. Oikonomou N, Mouratis MA, Tzouvelekis A, Kaffe E, Valavanis C, Vilaras G, Karameris A, Prestwich GD, Bouros D, Aidinis V. Pulmonary autotaxin expression contributes to the pathogenesis of pulmonary fibrosis. Am J Respir Cell Mol Biol. 2012 Nov;47(5):566-74. doi: 10.1165/rcmb.2012-0004OC.

27. Okubo T, Knoepfler PS, Eisenman RN, Hogan BL. Nmyc plays an essential role during lung development as a dosage-sensitive regulator of progenitor cell proliferation and differentiation. Development. 2005 Mar;132(6):1363-74.doi: 10.1242/dev.01678.

28. Hayashi S, Lewis P, Pevny L, McMahon AP. Efficient gene modulation in mouse epiblast using a Sox2Cre transgenic mouse strain. Mech Dev. 2002 Dec;119 Suppl 1:S97-S101.doi: 10.1016/S0925-4773(03)00099-6.

29. Ogilvy S, Elefanty AG, Visvader J, Bath ML, Harris AW, Adams JM. Transcriptional regulation of vav, a gene expressed throughout the hematopoietic compartment. Blood. 1998 Jan 15;91(2):419-30. PMID: 9427694.

30. Tronche F, Kellendonk C, Kretz O, Gass P, Anlag K, Orban PC, Bock R, Klein R, Schütz G. Disruption of the glucocorticoid receptor gene in the nervous system results in reduced anxiety. Nat Genet. 1999 Sep;23(1):99-103.doi: 10.1038/12703.

31. Nikitin AY, Alcaraz A, Anver MR, Bronson RT, Cardiff RD, Dixon D, Fraire AE, Gabrielson EW, Gunning WT, Haines DC, Kaufman MH, Linnoila RI, Maronpot RR, Rabson AS, Reddick RL, Rehm S, Rozengurt N, Schuller HM, Shmidt EN, Travis WD, Ward JM, Jacks T. Classification of proliferative pulmonary lesions of the mouse: recommendations of the mouse models of human cancers consortium. Cancer Res. 2004 Apr 1;64(7):2307-16.doi: 10.1158/0008-5472.CAN-03-3376.

32. Kanellakis NI, Giannou AD, Pepe MA, Agalioti T, Zazara DE, Giopanou I, Psallidas I, Spella M, Marazioti A, Arendt KAM, Lamort AS, Tsaniras SC, Taraviras S, Papadaki H, Lilis I, Stathopoulos GT. Tobacco chemical-induced mouse lung adenocarcinoma cell lines pin the prolactin orthologue proliferin as a lung tumour promoter. Carcinogenesis. 2019 Mar 4. pii: bgz047. doi: 10.1093/carcin/bgz047. [Epub ahead of print].

33. Franklin WA, Gazdar AF, Haney J, Wistuba II, La Rosa FG, Kennedy T, Ritchey DM, Miller YE. Widely dispersed p53 mutation in respiratory epithelium. A novel mechanism for field carcinogenesis. J Clin Invest. 1997 Oct 15;100(8):2133-7. doi: 10.1172/JCI119748.

34. Safran M, Kim WY, Kung AL, Horner JW, DePinho RA, Kaelin WG Jr. Mouse reporter strain for noninvasive bioluminescent imaging of cells that have undergone Cre-mediated recombination. Mol Imaging. 2003 Oct;2(4):297-302. PubMed PMID: 14717328.

35. Giopanou I, Lilis I, Papaleonidopoulos V, Marazioti A, Spella M, Vreka M, Papadaki H, Stathopoulos GT. Comprehensive Evaluation of Nuclear Factor-κB Expression Patterns in Non-Small Cell Lung Cancer. PLoS One. 2015 Jul 6;10(7):e0132527. doi: 10.1371/journal.pone.0132527.

36. National Lung Screening Trial Research Team, Aberle DR, Adams AM, Berg CD, Black WC, Clapp JD, Fagerstrom RM, Gareen IF, Gatsonis C, Marcus PM, Sicks JD. Reduced lung-cancer mortality with low-dose computed tomographic screening. N Engl J Med. 2011 Aug 4;365(5):395-409. doi: 10.1056/NEJMoa1102873.

37. Mason RJ, Kalina M, Nielsen LD, Malkinson AM, Shannon JM. Surfactant protein C expression in urethane-induced murine pulmonary tumors. Am J Pathol. 2000 Jan;156(1):175-82.doi: 10.1016/S0002-9440(10)64717-7.

38. Lindskog C, Fagerberg L, Hallström B, Edlund K, Hellwig B, Rahnenführer J, Kampf C, Uhlén M, Pontén F, Micke P. The lung-specific proteome defined by integration of transcriptomics and antibody-based profiling. FASEB J. 2014 Dec;28(12):5184-96. doi: 10.1096/fj.14-254862.

39. Sutherland KD, Berns A. Cell of origin of lung cancer. Mol Oncol. 2010 Oct;4(5):397-403. doi: 10.1016/j.molonc.2010.05.002.

40. Lawson WE, Polosukhin VV, Zoia O, Stathopoulos GT, Han W, Plieth D, Loyd JE, Neilson EG, Blackwell TS. Characterization of fibroblast-specific protein 1 in pulmonary fibrosis. Am J Respir Crit Care Med. 2005 Apr 15;171(8):899-907. 10.1164/rccm.200311-1535OC.

41. Rawlins EL, Okubo T, Xue Y, Brass DM, Auten RL, Hasegawa H, Wang F, Hogan BL. The role of Scgb1a1+ Clara cells in the long-term maintenance and repair of lung airway, but not alveolar, epithelium. Cell Stem Cell. 2009 Jun 5;4(6):525-34. doi: 10.1016/j.stem.2009.04.002.

42. Voehringer D, Liang HE, Locksley RM. Homeostasis and effector function of lymphopenia-induced "memory-like" T cells in constitutively T cell-depleted mice. J Immunol. 2008 Apr 1;180(7):4742-53.doi: 10.4049/jimmunol.180.7.4742.

43. Frank DB, Peng T, Zepp JA, Snitow M, Vincent TL, Penkala IJ, Cui Z, Herriges MJ, Morley MP, Zhou S, Lu MM, Morrisey EE. Emergence of a Wave of Wnt Signaling that Regulates Lung Alveologenesis by Controlling Epithelial Self-Renewal and Differentiation. Cell Rep. 2016 Nov 22;17(9):2312-2325. doi: 10.1016/j.celrep.2016.11.001.

44. Kabbout M, Garcia MM, Fujimoto J, Liu DD, Woods D, Chow CW, Mendoza G, Momin AA, James BP, Solis L, Behrens C, Lee JJ, Wistuba II, Kadara H. ETS2 mediated tumor suppressive function and MET oncogene inhibition in human non-small cell lung cancer. Clin Cancer Res. 2013 Jul 1;19(13):3383-95. doi: 10.1158/1078-0432.CCR-13-0341.

45. Clark JG, Kim KH, Basom RS, Gharib SA. Plasticity of airway epithelial cell transcriptome in response to flagellin. PLoS One. 2015 Feb 10;10(2):e0115486. doi: 10.1371/journal.pone.0115486.

46. Dancer RC, Parekh D, Lax S, D'Souza V, Zheng S, Bassford CR, Park D, Bartis DG, Mahida R, Turner AM, Sapey E, Wei W, Naidu B, Stewart PM, Fraser WD, Christopher KB, Cooper MS, Gao F, Sansom DM, Martineau AR, Perkins GD, Thickett DR. Vitamin D deficiency contributes directly to the acute respiratory distress syndrome (ARDS). Thorax. 2015 Jul;70(7):617-24. doi: 10.1136/thoraxjnl-2014-206680.

47. Lee SM, Gardy JL, Cheung CY, Cheung TK, Hui KP, Ip NY, Guan Y, Hancock RE, Peiris JS. Systems-level comparison of host-responses elicited by avian H5N1 and seasonal H1N1 influenza viruses in primary human macrophages. PLoS One. 2009 Dec 14;4(12):e8072. doi: 10.1371/journal.pone.0008072.

48. Alexandrov LB, Ju YS, Haase K, Van Loo P, Martincorena I, Nik-Zainal S, Totoki Y, Fujimoto A, Nakagawa H, Shibata T, Campbell PJ, Vineis P, Phillips DH, Stratton MR.

Mutational signatures associated with tobacco smoking in human cancer. Science. 2016 Nov 4;354(6312):618-622.doi: 10.1126/science.aag0299.

49. Castelletti N, Kaiser JC, Simonetto C, Furukawa K, Küchenhoff H, Stathopoulos GT. Risk of lung adenocarcinoma from smoking and radiation arises in distinct molecular pathways. Carcinogenesis. 2019 Feb 19. pii: bgz036. doi: 10.1093/carcin/bgz036.

50. Xu Y, Mizuno T, Sridharan A, Du Y, Guo M, Tang J, Wikenheiser-Brokamp KA, Perl AT, Funari VA, Gokey JJ, Stripp BR, Whitsett JA. Single-cell RNA sequencing identifies diverse roles of epithelial cells in idiopathic pulmonary fibrosis. JCI Insight. 2016 Dec 8;1(20):e90558. doi: 10.1172/jci.insight.90558.

51. Vaz M, Hwang SY, Kagiampakis I, Phallen J, Patil A, O'Hagan HM, Murphy L, Zahnow CA, Gabrielson E, Velculescu VE, Easwaran HP, Baylin SB. Chronic Cigarette Smoke-Induced Epigenomic Changes Precede Sensitization of Bronchial Epithelial Cells to Single-Step Transformation by KRAS Mutations. Cancer Cell. 2017 Sep 11;32(3):360-376.e6. doi: 10.1016/j.ccell.2017.08.006.

52. Nagel R, Semenova EA, Berns A. Drugging the addict: non-oncogene addiction as a target for cancer therapy. EMBO Rep. 2016 Nov;17(11):1516-1531.doi: 10.15252/embr.201643030.

53. Yee M, Chess PR, McGrath-Morrow SA, Wang Z, Gelein R, Zhou R, Dean DA, Notter RH, O'Reilly MA. Neonatal oxygen adversely affects lung function in adult mice without altering surfactant composition or activity. Am J Physiol Lung Cell Mol Physiol. 2009 Oct;297(4):L641-9. doi: 10.1152/ajplung.00023.2009.

54. Hsia CC, Hyde DM, Ochs M, Weibel ER; ATS/ERS Joint Task Force on Quantitative Assessment of Lung Structure. An official research policy statement of the American Thoracic Society/European Respiratory Society: standards for quantitative assessment of lung structure. Am J Respir Crit Care Med. 2010 Feb 15;181(4):394-418. doi: 10.1164/rccm.200809-1522ST.

55. Manali ED, Moschos C, Triantafillidou C, Kotanidou A, Psallidas I, Karabela SP, Roussos C, Papiris S, Armaganidis A, Stathopoulos GT, Maniatis NA. Static and dynamic mechanics of the murine lung after intratracheal bleomycin. BMC Pulm Med. 2011 May 31;11:33. doi: 10.1186/1471-2466-11-33.

56. Mazaika E, Homsy J. Digital Droplet PCR: CNV Analysis and Other Applications. Curr Protoc Hum Genet. 2014 Jul 14;82:7.24.1-13. doi: 10.1002/0471142905.hg0724s82.

57. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. 2005 Oct 25;102(43):15545-50.

58. Faul F, Erdfelder E, Lang AG, Buchner A. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behav Res Methods. 2007 May;39(2):175-91.doi: 10.3758/BF03193146.

FIGURE LEGENDS

Figure 1. Airway cells in urethane-induced lung tumors. (A) Cartoon of the different lung epithelial lineages, their distribution in the airways (club, goblet, ciliated, and basal cells) and the alveoli (alveolar type I and II cells), their permanent fluorescent genetic labeling in the reporter mice used in this study (green color), and the protein markers used for their identification. See also Figure 1-figure supplements 1-5. **(B)** Lung sections from naïve 6-week-old GFP;CCSP.CRE mice (*n* = 22), in which all airway cells bear permanent genetic GFP+ (green arrows) and all other cells TOMATO+ (red arrows) labels, counterstained with nuclear Hoechst33258 dye (top) or immunostained for the club cell marker CCSP and the alveolar type II cell marker SFTPC (bottom). a, alveoli; b, bronchi; v, vein. See also Figure 1-figure supplements 6-8. **(C)** Proliferating cell nuclear antigen (PCNA; brown) & hematoxylin (blue)-stained (top) and CCSP (green) & Hoechst33258 (blue)-stained (bottom) lung tumor sections of urethane-treated C57BL/6 mice six months post-treatment (*n* = 5/group), depicting endobronchial lung adenocarcinomas (white arrows). See also Figure 1-figure supplements 9-11. **(D)** Lung sections of GFP;CCSP.CRE mice (*n*=10) at six months post-urethane treatment bearing hyperplasias and tumors (dashed outlines) immunostained for the club cell marker CCSP (left) and the alveolar type II cell marker SFTPC (right). Note the GFP-labeled lesions of airway origin that have lost CCSP and have acquired SFTPC immunoreactivity. See also Figure 1-figure supplements 12-19. CCSP, Clara cell secretory protein; TUBA1A, acetylated α-tubulin; SFTPC, surfactant protein C; LYZ2, lysozyme 2; FOXJ1, forkhead box J1; KRT5, keratin 5.

Figure 2. Airway cells sustain *Kras*^{Q61R} mutations inflicted by urethane and give rise to juxtabronchial lung adenocarcinomas. **(A)** DNA was extracted from the lungs of GFP;CCSP.CRE and GFP;LYZ2.CRE mice (FVB strain) one and two weeks post-urethane treatment (*n* = 5/group). Summary of duplexed digital droplet PCR (ddPCR) results using primers and probes specific for the *Rosa*^{mT} and the *Kras*^{WT} sequences. Note that all cell types equally suffer initial *Kras*^{Q61R} mutations, but only GFP-labeled cells of GFP;CCSP.CRE mice (i.e. airway cells) maintain the *Kras*^{Q61R} mutation after two weeks. See also Figure 2-figure supplement 1. Data are shown as violin plot. *P*, overall probability, two-way ANOVA. ***: $P < 0.001$ compared with all other groups, Bonferroni post-tests. **(B)** Representative high-resolution micro-computed tomography (μCT) lung sections (top) and three-dimensional reconstructions (bottom) from urethane-treated FVB mice six months after treatment (*n* = 10). Note lung tumors attached to (green arrows) or contained within (blue arrows) the airways, as well as lung tumors with no obvious link to a bronchus (red arrows). **(C)** Summary of results from μCT (data from Figure 2B) and pathology (data from Figure 1C) shown as violin plot. *P*, probability, two-way ANOVA.*, ***, and ****: $P < 0.05$, $P < 0.001$, and $P < 0.0001$, respectively, compared with airway-attached tumors, Bonferroni post-tests. Shown are also Spearman's correlation coefficient (ρ) and probability (*P*) for correlation of μCT and pathology results.

Figure 3.Expansion of airway cells in the tumor-initiated lung. **(A)** Non-neoplastic alveolar regions from lung sections of saline-, urethane (ethyl carbamate, EC)-, and 3-methyl-1,2-dyhydrobenzo[j]aceanthrylene/butylated hydroxytoluene (MCA/BHT)-treated GFP;CCSP.CRE mice at six months into treatment (*n* = 8 mice/group). Note the few GFP-labeled cells of saline-treated mice and their increased numbers in carcinogen-treated mice (arrows). See also Figure 3-figure supplements 1, 2. **(B)** Juxtabronchial region from lung section of urethane-treated GFP;CCSP.CRE mouse at six months into treatment (*n* = 22) stained for the alveolar type II cell marker SFTPC. Arrows and legend indicate different phenotypes of extrabronchial GFP-labeled cells. See also Figure 3-figure supplements 3-5. **(C)** Merged high-power image of SFTPC and KRT5 co-staining of human lung adenocarcinoma (*n* = 10) shows significant co-localization of the two markers in a subset of tumor cells (arrows). See also Figure 3-figure supplement 6. CCSP, Clara cell secretory protein; SFTPC, surfactant protein C; KRT5, keratin 5.

Figure 4. Airway cells in alveolar repair. (A) Non-neoplastic alveolar regions from lung sections of aging GFP;CCSP.CRE mice (bottom right section is also SFTPC-immunostained) show increasing numbers of alveolar GFP-labeled cells with age (arrows). Green arrows: genetically GFP-labeled, SFTPC-immunoreactive airway cells in alveolus of 15-month-old GFP;CCSP.CRE mouse. Color-coded boxes indicate time-windows of experiments in (B-D). **(B)** Data summary (*n* = 5 mice/time-point) from (A) shown as violin plot. *P*, probability, one-way ANOVA. ns, ***, and ****: $P > 0.05$, $P < 0.001$, and $P < 0.0001$, respectively, for comparison with time-point zero by Bonferroni post-tests. **(C)** SFTPC-immunostained lung sections of GFP;CCSP.CRE mice show accelerated increase of alveolar GFP-labeled SFTPC-immunoreactive airway cells after bleomycin treatment (arrows). **(D)** Data summary from (C) shown as violin plots (*n* = 4 mice/time-point). *P*, probabilities, one-way ANOVA. ns, *, **, ***, and ****: $P > 0.05$, $P < 0.05$, $P < 0.01$, $P < 0.001$, and $P < 0.0001$, respectively, for comparison with day zero by Bonferroni post-tests. (**E)** SFTPC-stained lung sections of GFP;CCSP.CRE mice at two months after perinatal exposure to 98% O_2 show enlarged alveoli (evident by increased mean linear intercept) enriched in GFP-labeled SFTPC-immunoreactive airway cells (arrows) compared with 21% O_2. See also Figure 4-figure supplements 1, 2. **(F)** Data summary from (E) shown as violin plots (*n* = 6 mice/group). *P*, probabilities, t-test. **(G)** Lung sections (top) of GFP;CCSP.CRE mice (*n* = 5 mice/group) show enrichment of alveoli in GFP-labeled cells post-naphthalene treatment (arrows). Lung sections (bottom) of GFP;LYZ2.CRE mice (*n* = 5 mice/group) at six weeks post-naphthalene show no bronchial (b) GFP-labeled cells. See also Figure 4-figure supplements 3, 4. **(H)** Data summary from (G) shown as violin plot (*n* = 5 mice/time-point). *P*, probability, two-way ANOVA. ns and ****: $P > 0.05$ and $P < 0.0001$, respectively, for comparison with corn oil by Bonferroni post-tests. CCSP, Clara cell secretory protein; SFTPC, surfactant protein C; LYZ2, lysozyme 2.

Figure 5. Airway cell-ablated mice display alveolar destruction and are protected from carcinogenesis. **(A)** Lineage marker-immunostained lung sections of 12-week-old GFP;CCSP.CRE;DTA and GFP;LYZ2.CRE;DTA mice (*n* = 6/group) show increased bronchial and alveolar size and flat CCSP+SFTPC+LYZ2+ cells in the airways of GFP;CCSP.CRE;DTA mice (green arrows), and CCSP-SFTPC-LYZ2+ alveolar macrophages in the airspaces of GFP;LYZ2.CRE;DTA mice (blue arrows). See also Figure 5-figure supplements 1, 2. **(B)** Hematoxylin and eosin-stained lung sections (*n* = 6/group) from 12-week-old DTA (controls), CCSP.CRE;DTA (airway epithelial suicide model), and LYZ2.CRE;DTA (alveolar epithelial suicide model) mice. **(C)** Data summaries of mean linear intercept, bronchoalveolar lavage (BAL) myeloid cells, pressure-volume curves, airway resistance, and static compliance (*n* = 6-10/group) from 12-week-old DTA, CCSP.CRE;DTA, and LYZ2.CRE;DTA mice shown as violin plots. *P*, probabilities, one-way ANOVA. ns, **, and ***: $P > 0.05$, $P < 0.01$, and $P < 0.001$, respectively, for the indicated comparisons, Bonferroni post-tests. **(D)** Lung photographs of control, CCSP.CRE;DTA, and LYZ2.CRE;DTA mice at six months into treatment with urethane started at six weeks of age. **(E)** Incidence table and data summaries of lung tumors from (D) (violin plots; *n* is given in table). *P*, probabilities, χ^2-test (table) and one-way ANOVA (graphs). ns, *, **, and ***: $P > 0.05$, $P < 0.05$, $P < 0.01$, and $P < 0.001$, respectively, for the indicated comparisons, Fischer's exact tests (table) or Bonferroni post-tests (graphs). a, alveoli; b, bronchi; ps, pleural space; v, vessel. CCSP, Clara cell secretory protein; SFTPC, surfactant protein C; LYZ2, lysozyme 2.

Figure 6. Airway and alveolar signatures in murine and human lung adenocarcinoma (LUAD). (A, B) RNA of mouse urethane-induced LUAD cell lines, lungs obtained pre- and one week post-urethane treatment , airway epithelial cells (AEC), alveolar type II cells (ATII), and bone marrow-derived macrophages (BMDM) was examined by Affymetrix Mouse Gene ST2.0 microarrays (n = 4/group). **(A)** Heat map of genes significantly differentially expressed (overall ANOVA and FDR $P < 10^{-6}$) shows accurate hierarchical clustering. **(B)** Expression of the 30 top-represented transcripts of AEC, ATII, and BMDM in lungs and LUAD cells. See also Figure 6-figure supplements 1-4. **(C, D)** RNA of human LUAD (n = 40), never-smoker lung tissue (n = 30), primary AEC (n = 5), primary ATII (n = 4), and alveolar macrophages (AMΦ; n = 9) was analyzed by Affymetrix Human Gene ST1.0 microarrays. (**C)** Heat map of genes significantly differentially expressed (ΔGE > 5-fold) between LUAD and lung (ANOVA and FDR $P < 10^{-3}$) shows accurate hierarchical clustering. **(D)** Mean expression levels of the 30 top-represented transcripts of human AEC, ATII, and AMΦ in lungs and LUAD. **(E, F)** Gene set enrichment analyses, including normalized enrichment scores (NES), of mouse (E) and human (F) AEC, ATII, and BMDM/AMΦ signatures (defined as the top 1% expressed genes overall or exclusive to the cell type; n = 2) in mouse and human LUAD transcriptomes shows significant enrichment of the AEC (but not the ATII and BMDM/AMΦ) signature compared with lung (nominal $P < 0.0001$ for all, family-wise error rates FWER < 0.01). Gene symbols indicate the top 3 lagging genes from each signature and shows loss of *Scgb1a1* (encoding CCSP) by LUAD. See also Figure 6-figure supplements 5, 6. Data are given as violin plots. P, two-way ANOVA probabilities. ns, *, **, and ***: $P > 0.05$, $P < 0.05$, $P < 0.01$, and $P < 0.001$ for the indicated comparisons by Bonferroni post-tests. ANOVA, analysis of variance; FDR, false discovery rate.

Figure 7. Proposed role of airway-marked cells in murine lung maintenance and adenocarcinoma. **(A)** Our evidence supports the existence of distinct developmental ancestries for airway epithelial (AEC) and alveolar type II (ATII) cells, notwithstanding their common descent from an early (possibly *Sftpc*+) lung epithelial progenitor. The developmental airway lineage (*Scgb1a1*+*Sftpc*±; green) gives rise to all types of airway cells, including club, ciliated, goblet, basal, and other cells, while the developmental ATII lineage (*Sftpc*+*Lyz2*±; red) gives rise to ATII cells before birth. These lineages appear to be segregated in the growing unaffected lung of the mouse till the age of six weeks, which roughly corresponds to a human age of six years, where cellular proliferation in the human lungs ceases. Thereafter, and likely due to the continuous exposure of the lungs to inhaled noxious agents, gradual expansion of *Scgb1a1*+*Sftpc*± marked cells ensues. Upon lung injury, this process is accelerated. Similarly, during carcinogenesis caused by chemical tobacco smoke carcinogens, *Scgb1a1*+*Sftpc*± marked cells expand and are ubiquitously present in peripheral lung adenocarcinomas. **(B)** Proposed neonatal proportions and postnatal dynamics of pulmonary epithelial cells during adulthood. Estimated proportions of lineage-marked cells at birth, based on flow cytometry and co-localization of proteinaceous and genetic cell marking. Lung lineages appear to be segregated in the growing lung till the age of full lung development (six weeks in mice and 6-8 years in humans) or till lung injury ensues. Schematic of proposed postnatal redistribution of marked cells in the adult lung. Upon injury, during multi-stage field carcinogenesis, or even during unchallenged aging, *Scgb1a1*+ marked cells appear in the distal alveolar regions, thereby maintaining lung structure and function. Bubble size indicates relative marked cell abundance. CCSP, Clara cell secretory protein; FOXJ1, forkhead box J1; KRT5, keratin 5; LYZ2, lysozyme 2; SFTPC, surfactant protein C; TUB1A1, acetylated α-tubulin.

LEGENDS TO FIGURE SUPPLEMENTS

FIGURE 1 - FIGURE SUPPLEMENTS

Figure 1 - Figure Supplement 1. Table of pulmonary lineage markers and key abbreviations used in this study. TUBA1A, Tubulin alpha 1a or acetylated tubulin; KRT5, Keratin 5; FOXJ1, Forkhead box J1; CCSP, Secretoglobin, family 1A, member 1 (uteroglobin) or Clara cell secretory protein or Clara cell 10 KDa protein; SFTPC, Surfactant protein C; LYZ2, Lysozyme 2; AEC, airway epithelial cells; BASC, bronchoalveolar stem cells; ATII, alveolar type II cells or type II pneumocytes; AMΦ, alveolar macrophages.

Figure 1 - Figure Supplement 2. Genetic labeling of pulmonary lineages in eleven mouse strains and intercrosses: summary of results. CRE, causes recombination; TOMATO, tdTomato; GFP, green fluorescent protein; CCSP, Clara cell secretory protein; SFTPC, surfactant protein C; LYZ2, lysozyme 2; SOX2, sex determining region Y (SRY)-box 2; VAV, Vav Guanine Nucleotide Exchange Factor 1; NES, nestin; JAX, Jackson Laboratories; EMMA, European Mutant Mouse Archive; MGI, Mouse Genome Informatics; AEC, airway epithelial cells; BASC, bronchoalveolar stem cells; ATII, alveolar type II cells or type II pneumocytes; AMΦ, alveolar macrophages; BM, bone marrow (myeloid) cells. Symbols indicate: - (white), no genetic labeling; + (magenta), complete genetic labeling; ± (blue), partial genetic labeling.

Figure 1 - Figure Supplement 3. Genetic labeling of pulmonary lineages in seven lineage reporter strains on the C57BL/6 background: representative images. Representative photographs (top row) and green epifluorescence images (second row) of whole lungs, as well as fluorescent microscopic images of lung sections for nuclear Hoechst33258 stain (third row), endogenous TOMATO (fourth row), endogenous GFP (fifth row), and merged images

(bottom row) of genetically marked mice on the C57BL/6 background employed in these studies (described in detail in Figure 1 - Figure Supplement 2) at six postnatal weeks (*n* = 5/mouse strain). b, bronchi; a, alveoli; ps, pleural space.

Figure 1 - Figure Supplement 4. Genetic labeling of pulmonary lineages in seven lineage reporter strains on the C57BL/6 background: data summary. XY plot of GFP-labeled airway versus alveolar cells from *n* = 5 mice/mouse strain. Arrows denote the three lineage-reporter strains selected for further study including GFP;CCSP.CRE (green), GFP;LYZ2.CRE (blue), and GFP;SFTPC.CRE (olive) mice. Data are given as mean ± SD.

Figure 1 - Figure Supplement 5. Flow cytometric quantification of lineage-labeled cells in three lineage reporter strains on the C57BL/6 background. Schematic representation of genetic lineage labeling of GFP;CCSP.CRE, GFP;SFTPC.CRE, and GFP;LYZ2.CRE mice (left), flow cytometric gating strategy to quantify GFP+ and TOMATO+ cells (middle), and violin plot from *n* = 5, 3, and 6 mice/strain (right). Numbers are mean ± SD. *P*, overall probability, two-way ANOVA.

Figure 1 - Figure Supplement 6. Genetic lineage labels of protein-marked cells in three lineage reporter strains on the C57BL/6 background: representative images. Representative merged fluorescent microscopic images from lineage marker-stained lung sections of 6-week-old lineage-labeled mice (*n* = 5/group). Arrows indicate cells expressing the respective marker protein with (green) or without (red) genetic lineage-labeling. CCSP, Clara cell secretory protein; TUBA1A, acetylated tubulin; SFTPC, surfactant protein C; LYZ2, lysozyme 2; b, bronchi; a, alveoli.

Figure 1 - Figure Supplement 7. Genetic lineage labels of protein-marked cells in seven lineage reporter strains on the C57BL/6 background: data summary. XY plot of ratios of

genetic GFP-labeled to protein marker CCSP and SFTPC-immunoreactive cells (*n* = 5/group). Arrows denote the three lineage-reporter strains selected for further study including GFP;CCSP.CRE (green), GFP;LYZ2.CRE (blue), and GFP;SFTPC.CRE (olive) mice. Data are given as mean ± SD.

Figure 1 - Figure Supplement 8. Protein markings of lineage-labeled cells in three lineage reporter strains on the C57BL/6 background: data summary. Quantification of protein marker expression of genetic-labeled cells of GFP;CCSP.CRE, GFP;LYZ2.CRE, and GFP;SFTPC.CRE mice (*n* = 6/strain) for Clara cell secretory protein (CCSP), surfactant protein C (SFTPC), and lysozyme 2 (LYZ2). Data are given as violin plots. *P*, overall probability, two-way ANOVA.

Figure 1 - Figure Supplement 9. Two carcinogen regimens for reproducible lung tumor induction in naturally resistant C57BL/6 mice. Top: schematic of multi-hit urethane administration tailored to yield 90% tumor incidence in C57BL/6 mice: ten weekly intraperitoneal injections of 1 g/Kg urethane (ethyl carbamate, EC; grey arrows) are initiated at six weeks after birth (pink arrow) and lungs are examined six months after the first urethane injection (black arrow). Bottom: 3-methyl-1,2-dyhydrobenzo[j]aceanthrylene (MCA)/butylated hydroxytoluene (BHT) regimen tailored to yield 90% tumor incidence in C57BL/6 mice. Four weekly intraperitoneal injections of 15 mg/Kg MCA (red arrows) initiated at six weeks after birth (pink arrow) are followed by eight weekly intraperitoneal injections of 200 mg/Kg BHT (blue arrows) and lung examination at six months after first MCA dose (black arrow).

Figure 1 - Figure Supplement 10. Lung tumors induced in C57BL/6 mice by two carcinogen regimens. Eighty-four C57BL/6 mice received ten weekly intraperitoneal injections of 1 g/Kg urethane (ethyl carbamate, EC) initiated at six weeks of age and lungs

were examined six months after the first urethane injection (black font and symbols). Twenty C57BL/6 mice received four weekly intraperitoneal injections of 15 mg/Kg 3-methyl-1,2-dyhydrobenzo[j]aceanthrylene (MCA) followed by eight weekly intraperitoneal injections of 200 mg/Kg butylated hydroxytoluene (BHT) and lungs were examined six months after the first MCA dose (grey font and symbols). Table shows tumor incidence and graph shows tumor number versus mean tumor diameter. Each small circle represents one mouse and each large circle with error bar the means for each carcinogen regimen.

Figure 1 - Figure Supplement 11. Airway links of urethane-induced lung adenocarcinomas. Proliferating cell nuclear antigen (PCNA)-stained lung sections of urethane-treated C57BL/6 mice at six months post-treatment start. Arrows: airway hyperplasias (grey) and lung adenocarcinomas (black) arising within a bronchus (left) and apparently in an alveolar region but adjacent to a bronchus (right).

Figure 1 - Figure Supplement 12. Genetic labeling of urethane-induced lung adenocarcinomas in four lineage reporter strains on the C57BL/6 background: representative images. Representative photographs (top row) and green epifluorescence images (second row), as well as merged fluorescent microscopic images of lung sections for nuclear Hoechst33258 stain, endogenous TOMATO, and endogenous GFP (bottom three rows), of tumor-bearing lungs from genetically marked mice employed in these studies (described in detail in Figure 1 -figure supplement 2) at six months after initiation of ten weekly intraperitoneal urethane injections (*n* = 30, 22, 18, and 20/strain, respectively). b, bronchi. Top two rows: arrows indicate lung tumors. Bottom three rows: white arrows indicate GFP-labeled cells in apparently non-affected alveolar areas of GFP;CCSP.CRE mice; green arrow indicates rare GFP+ cell in non-affected central airway of GFP;LYZ2.CRE mouse. Note the absence of GFP-labeling of lung tumors in TOMATO mice, the complete

GFP-labeling in GFP;CCSP.CRE and GFP;SFTPC.CRE mice, and the partial GFP-labeling in GFP;LYZ2.CRE mice.

Figure 1 - Figure Supplement 13. Genetic labeling of urethane-induced lung adenocarcinomas in four lineage reporter strains on the C57BL/6 background: data summary. XY plot of percentage of GFP-labeled tumors/lung versus GFP-labeled tumor cells/tumor averaged per lung in strains from Figure 1 -figure supplement 12 (*n* = 30, 22, 18, and 20/group, respectively). Data are given as mean ± SD.

Figure 1 - Figure Supplement 14. Genetic labeling of MCA/BHT-induced lung adenocarcinomas in two lineage reporter strains on the C57BL/6 background: representative images. Single-channel (endogenous TOMATO and GFP labeling and Hoechst 33258 nuclear stain) and merged images of lung hyperplasias and tumors (dashed outlines) of TOMATO and GFP;CCSP.CRE mice at six months after initiation of treatment with four weekly intraperitoneal injections of 15 mg/Kg 3-methyl-1,2-dyhydrobenzo[j]aceanthrylene (MCA) followed by eight weekly intraperitoneal injections of 200 mg/Kg butylated hydroxytoluene (BHT) (*n* = 8/group). Note the absence of GFP-labeling in lesions of TOMATO mice and the GFP-labeled lesions of GFP;CCSP.CRE mice.

Figure 1 - Figure Supplement 15. Protein marker expression of urethane-induced lung adenocarcinomas in three lineage-labeled mouse strains on the C57BL/6 background: representative images. Lineage marker protein-stained lung adenocarcinomas (dashed outlines) from genetically marked mice (*n* = 10/group). Note the genetic GFP-labeled tumor cells of GFP;CCSP.CRE mice that have lost CCSP and have acquired SFTPC with or without LYZ2 protein marker expression. CCSP, Clara cell secretory protein; TUBA1A, acetylated α-tubulin; SFTPC, surfactant protein C; LYZ2, lysozyme 2.

Figure 1 - Figure Supplement 16. Genetic lineage labels of protein-marked cells in three lineage reporter strains on the FVB background: representative images. Representative merged fluorescent microscopic images from lineage marker-stained lung sections of 6-week-old lineage reporter mice (*n* = 5/group). Arrows indicate cells expressing the respective marker protein with (green) or without (red) genetic lineage-labeling. CCSP, Clara cell secretory protein; SFTPC, surfactant protein C; LYZ2, lysozyme 2; b, bronchi; a, alveoli.

Figure 1 - Figure Supplement 17. A single-hit mouse model for urethane-induced lung adenocarcinoma induction in naturally susceptible FVB mice. Schematic of single-hit urethane administration tailored to yield 100% tumor incidence in FVB mice: one intraperitoneal injection of 1 g/Kg urethane (ethyl carbamate, EC; grey arrow) is delivered at six weeks after birth (pink arrow) and lungs are examined six months later (black arrow).

Figure 1 - Figure Supplement 18. High-throughput epifluorescent detection of genetic labeling of urethane-induced lung adenocarcinomas in four lineage reporter strains on the FVB background: representative images. Representative photographs (top) and green (middle) and red (bottom) epifluorescence images of tumor-bearing lungs from genetically lineage-marked FVB mice at six months after a single intraperitoneal urethane injection ($n \geq$ 8/strain). Arrows indicate all (white), GFP-labeled (green), and TOMATO-labeled (red) lung tumors.

Figure 1 - Figure Supplement 19. Genetic labeling of urethane-induced lung adenocarcinomas in three lineage reporter strains on the FVB background: representative images. Representative merged fluorescent microscopic images of lineage marker protein-stained lung tumors (dashed outlines) from genetically marked mice (FVB background) at six months after a single intraperitoneal urethane injection ($n \geq$ 10/strain). Note the genetic GFP-labeled tumor cells of GFP;CCSP.CRE mice that have lost CCSP and

have acquired SFTPC with or without LYZ2 protein marker expression. CCSP, Clara cell secretory protein; SFTPC, surfactant protein C; LYZ2, lysozyme 2.

FIGURE 2 - FIGURE SUPPLEMENTS

Figure 2 - Figure Supplement 1. Airway cells sustain *Kras*^{Q61R} mutations inflicted by urethane. DNA was extracted from the lungs of GFP;CCSP.CRE and GFP;LYZ2.CRE mice (FVB strain) one and two weeks post-urethane treatment (*n* = 5/group). Representative gating strategy of digital droplet PCR (ddPCR) using primers and probes specific for the *Rosa*^{mT} and the *Kras*^{WT} sequences. Dashed outlines indicated GFP+*Kras*^{Q61R}+ droplet gates.

FIGURE 3 - FIGURE SUPPLEMENTS

Figure 3 - Figure Supplement 1. Airway-labeled cells in the alveoli of carcinogen-exposed C57BL/6 mice: representative images. Single-channel microscopy images (endogenous TOMATO and GFP labeling with Hoechst 33258 nuclear stain) of non-neoplastic alveolar regions of GFP;CCSP.CRE mice treated as in Figure 3A.

Figure 3 - Figure Supplement 2. Airway-labeled cells in the alveoli of carcinogen-exposed C57BL/6 mice: data summary. Data summary (shown as violin plot) from GFP;CCSP.CRE mice treated as in Figure 3A (*n* = 10/group). *P*, overall probability, one-way ANOVA. ns and **: $P > 0.05$ and $P < 0.01$ for the indicated comparisons, Bonferroni post-tests.

Figure 3 - Figure Supplement 3. Airway-labeled cells in the alveoli of carcinogen-exposed mice express SFTPC. Single-channel images of non-neoplastic distal lung regions of urethane-treated GFP;CCSP.CRE mice at six months into treatment (*n* = 22), stained for the lung cell markers Clara cell secretory protein (CCSP), acetylated α-tubulin (TUBA1A),

and surfactant protein C (SFTPC). Note the genetic GFP-labeled tumor cells that have lost CCSP and have acquired SFTPC protein marker expression.

Figure 3 - Figure Supplement 4. Airway-labeled cells in environmental-induced lung tumors express SFTPC. Juxtabronchial regions, alveolar hyperplasias, and tumors (dashed lines) of lungs from urethane-treated GFP;CCSP.CRE mice at six months into treatment (*n* = 22) stained for lineage marker proteins Clara cell secretory protein (CCSP), acetylated α-tubulin (TUBA1A), and surfactant protein C (SFTPC). Arrows and legend indicate different phenotypes of extrabronchial GFP-labeled cells. a, alveoli; b, bronchi.

Figure 3 - Figure Supplement 5. *In vivo* bioluminescent detection of the airway lineage in the lungs of saline- and carcinogen-treated mice. Representative merged bioluminescence/photographic images (left) and data summary (right) of LUC;CCSP.CRE mice (FVB background) before and seven months after saline (one intraperitoneal injection of 100 µL; *n* = 6) or urethane (one intraperitoneal injection of 1 g/Kg in 100 µL saline; *n* = 5) treatment. Note that in this model light is emitted exclusively by genetically CCSP-labeled cells over the lungs. Note also the signal decrease in saline- and increase in urethane-treated mice. Data are given as mean ± SD. *P*, overall probability, two-way ANOVA. ***: $P < 0.001$ for comparison with saline, Bonferroni post-test.

Figure 3 - Figure Supplement 6. Human lung adenocarcinomas co-express airway and alveolar markers. Co-staining of human lung adenocarcinomas for SFTPC and either CCSP (A; *n* = 10) or KRT5 (B; *n* = 10) shows absence of CCSP expression and significant co-localization of SFTPC and KRT5 in a subset of tumor cells. CCSP, Clara cell secretory protein; KRT5, keratin 5; SFTPC, surfactant protein C.

FIGURE 4 - FIGURE SUPPLEMENTS

Figure 4 - Figure Supplement 1. Alveolar type II cell ablation using bleomycin pre-treatment increases airway-labeled cells in urethane-induced lung tumors: representative images. Representative epifluorescence (top) and merged fluorescent microscopy (bottom) images of tumor-bearing lungs and lung tumors of six-week-old GFP;CCSP.CRE mice that received intratracheal saline or 0.08 units bleomycin (*n* = 6/group), were allowed to recover for one month, and subsequently received ten weekly intraperitoneal injections of 1 g/Kg urethane to be sacrificed six months after the first urethane injection. Arrows and dashed outlines indicate lung tumors.

Figure 4 - Figure Supplement 2. Alveolar type II cell ablation using bleomycin pre-treatment increases airway-labeled cells in urethane-induced lung tumors: data summary. Violin plot of GFP-labeled tumors/mouse (*n* = 6 mice/group) and GFP-labeled cells/tumor (*n* = 12 tumors/group; *n* = 2 tumors/mouse were examined) from experiment described in Figure 4 -figure supplement 1. Note the enrichment of lung adenocarcinomas in GFP-labeled cells in response to bleomycin, which depletes resident alveolar type II cells. *P*, overall probability, two-way ANOVA. **: $P < 0.01$ for comparison with saline, Bonferroni post-test.

Figure 4 - Figure Supplement 3. Airway epithelial cell ablation using naphthalene is restored by airway-labeled cells: representative images. Representative fluorescent microscopic images of lungs of GFP;CCSP.CRE mice at different time-points after intraperitoneal injection of 250 mg/Kg naphthalene given at six weeks of age. Shown are merges of Hoechst 33258-stain, endogenous TOMATO- and GFP-labeling, and immunostains for surfactant protein C (SFTPC, left) or Clara cell secretory protein (CCSP, right). Arrows denote naphthalene-induced airway epithelial gaps that are restored by GFP-labeled airway cells that express CCSP, but not SFTPC protein.

Figure 4 - Figure Supplement 4. Airway epithelial cell ablation by naphthalene: data summary. Violin plot of percentage of GFP-labeled airway cells from experiment described in Figure 4 -figure supplement 3 (n = 6 mice/time-point). P, overall probability, one-way ANOVA. ***: P< 0.001 for the comparison with day zero, Bonferroni post-test.

FIGURE 5 - FIGURE SUPPLEMENTS

Figure 5 - Figure Supplement 1. Triple transgenic mouse models for validation of genetic pulmonary lineage ablation: representative images. Representative lung sections of 12-week-old GFP;CCSP.CRE, GFP;LYZ2.CRE, GFP;CCSP.CRE;DTA, and GFP;LYZ2.CRE;DTA mice (n = 6/group). Shown are merges of Hoechst 33258-stained endogenous TOMATO- and GFP-labeling. Note increased bronchial (b) and alveolar (a) size, complete airway epithelial denudement, and prominent distortion of bronchial and alveolar structure of GFP;CCSP.CRE;DTA mice compared with other strains, mimicking chronic obstructive pulmonary disease. Note also the presence of some GFP-labeled alveolar macrophages in GFP;LYZ2.CRE;DTA mice (arrows). a, alveoli; b, bronchi.

Figure 5 - Figure Supplement 2. Triple transgenic mouse models for validation of genetic pulmonary lineage ablation: data summary. Violin plot of GFP-labeling of lung sections of 12-week-old mice from Figure 5 -figure supplement 1 (n = 6/group). Note the complete ablation of airway cells in GFP;CCSP.CRE mice and the persistence of some GFP-labeled alveolar macrophages in GFP;LYZ2.CRE;DTA mice. Measurements were from at least five non-overlapping tumor, airway, or alveolar fields/lung. P, overall probability, two-way ANOVA. ns and ****: P> 0.05 and P< 0.0001, respectively, for the indicated comparisons by Bonferroni post-tests.

FIGURE 6 - FIGURE SUPPLEMENTS

Figure 6 - Figure Supplement 1. Lineage-specific gene expression in mouse lung adenocarcinoma cell lines induced by urethane compared with mouse lungs. RNA of mouse urethane-induced lung adenocarcinoma (LUAD) cell lines, lungs obtained pre- and one week post-urethane treatment, and airway epithelial cells (AEC), alveolar type II cells (ATII), and bone marrow-derived macrophages (BMDM) was examined by Affymetrix Mouse Gene ST2.0 microarrays (*n* = 4/group). Shown is the number of genes out of the 30 top-represented transcripts of AEC, ATII, and BMDM within the top-2000-expressed genes of lungs and LUAD cells.

Figure 6 - Figure Supplement 2. Loss of lineage marker expression in mouse lung adenocarcinoma cell lines induced by urethane. Mean expression levels of selected transcripts, including lineage markers and markers of histologic subtype in lung adenocarcinoma (LUAD) cell lines compared with lungs pre- and one week post-urethane treatment (A and B, microarrays from Figure 6 -figure supplement 1, *n* = 2/group; C, qPCR, *n* = 3/group). AD, adenocarcinoma; SQ, squamous cell carcinoma; SC, small cell carcinoma. *P*, overall probability, two-way ANOVA. ****: $P < 0.0001$ for the highlighted genes compared with lungs (red, significantly down-regulated; green, significantly up-regulated).

Figure 6 - Figure Supplement 3. Loss of lineage marker expression in mouse lung adenocarcinoma cell lines induced by urethane compared with mouse lungs: heat maps. 528 genes differentially expressed between six different lung adenocarcinoma cell lines cultured from urethane-induced lung tumors and six benign respiratory mouse samples, including lungs of saline- and urethane-treated mice obtained at one week post-treatment, as well as primary mouse tracheal epithelial cells using the cut-offs indicated. Whole heat map (left) showing the accurate hierarchical clustering of the samples according to differentially expressed genes, as well as the top over- and under-represented genes (right). Note the

universal loss of expression of lineage markers by lung adenocarcinoma cells (genes in red font). ANOVA, analysis of variance; FDR, false discovery rate.

Figure 6 - Figure Supplement 4. Loss of lineage marker expression in mouse lung adenocarcinoma cell lines induced by urethane compared with mouse lungs: volcano plot. Shown are selected top over- and under-represented genes (arrows) from microarrays from Figure 6 -figure supplement 2.

Figure 6 - Figure Supplement 5. Mouse gene set enrichment analyses. Shown are gene set enrichment analyses of airway epithelial cell (AEC), alveolar type II cell (ATII), and bone marrow-derived macrophage (BMDM) transcriptome signatures in mouse lungs (top) and urethane-induced lung adenocarcinoma (LUAD) cell lines (bottom) transcriptomes. The data were used to design Figure 6E.

Figure 6 - Figure Supplement 6. Human gene set enrichment analyses. Affymetrix Human Gene ST1.0 microarrays hybridized with RNA of human lung adenocarcinomas (LUAD; n = 40), never-smoker lung tissues (n = 30), primary airway epithelial cells (AEC; n = 5), primary alveolar type II cells (ATII; n = 4), and alveolar macrophages (AMΦ; n = 9) were cross-examined. Shown are gene set enrichment analyses of AEC, ATII, and AMΦ signatures in lung (top) and LUAD (bottom) transcriptomes. The data were used to design Figure 6F.

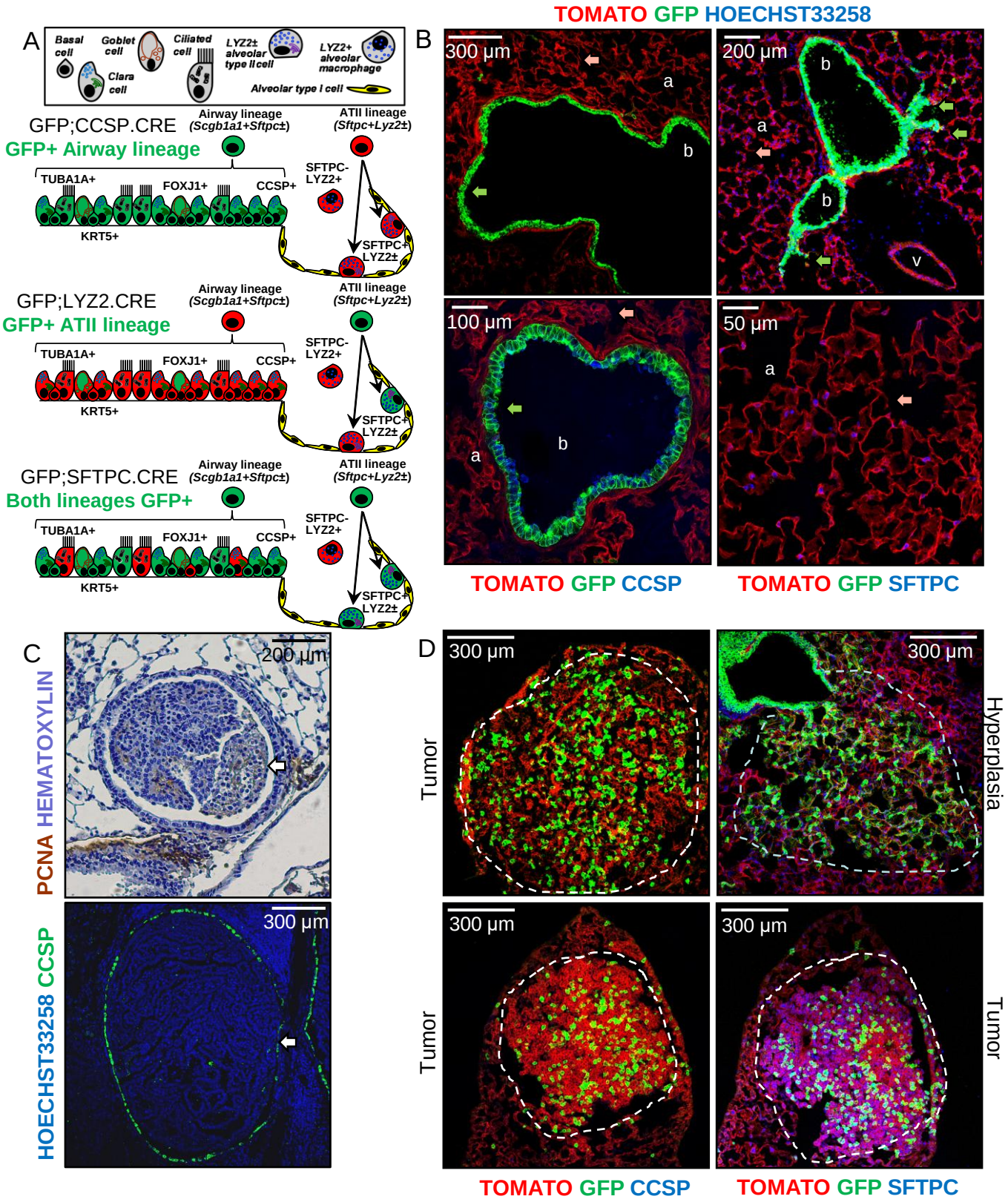

A
Basal cell
Goblet cell
Ciliated cell
LYZ2± alveolar type II cell
LYZ2+ alveolar macrophage
Clara cell
Alveolar type I cell
GFP;CCSP.CRE
GFP+ Airway lineage
Airway lineage (Scgb1a1+Sftpc±)
ATII lineage (Sftpc+Lyz2±)
TUBA1A+
FOXJ1+
CCSP+
KRT5+
SFTPC- LYZ2+
SFTPC+ LYZ2±
GFP;LYZ2.CRE
GFP+ ATII lineage
GFP;SFTPC.CRE
Both lineages GFP+
B
TOMATO GFP HOECHST33258
300 µm
200 µm
100 µm
50 µm
a
b
v
TOMATO GFP CCSP
TOMATO GFP SFTPC
C
PCNA HEMATOXYLIN
HOECHST33258 CCSP
200 µm
300 µm
D
Tumor
Hyperplasia
300 µm
TOMATO GFP CCSP
TOMATO GFP SFTPC

Figure 1

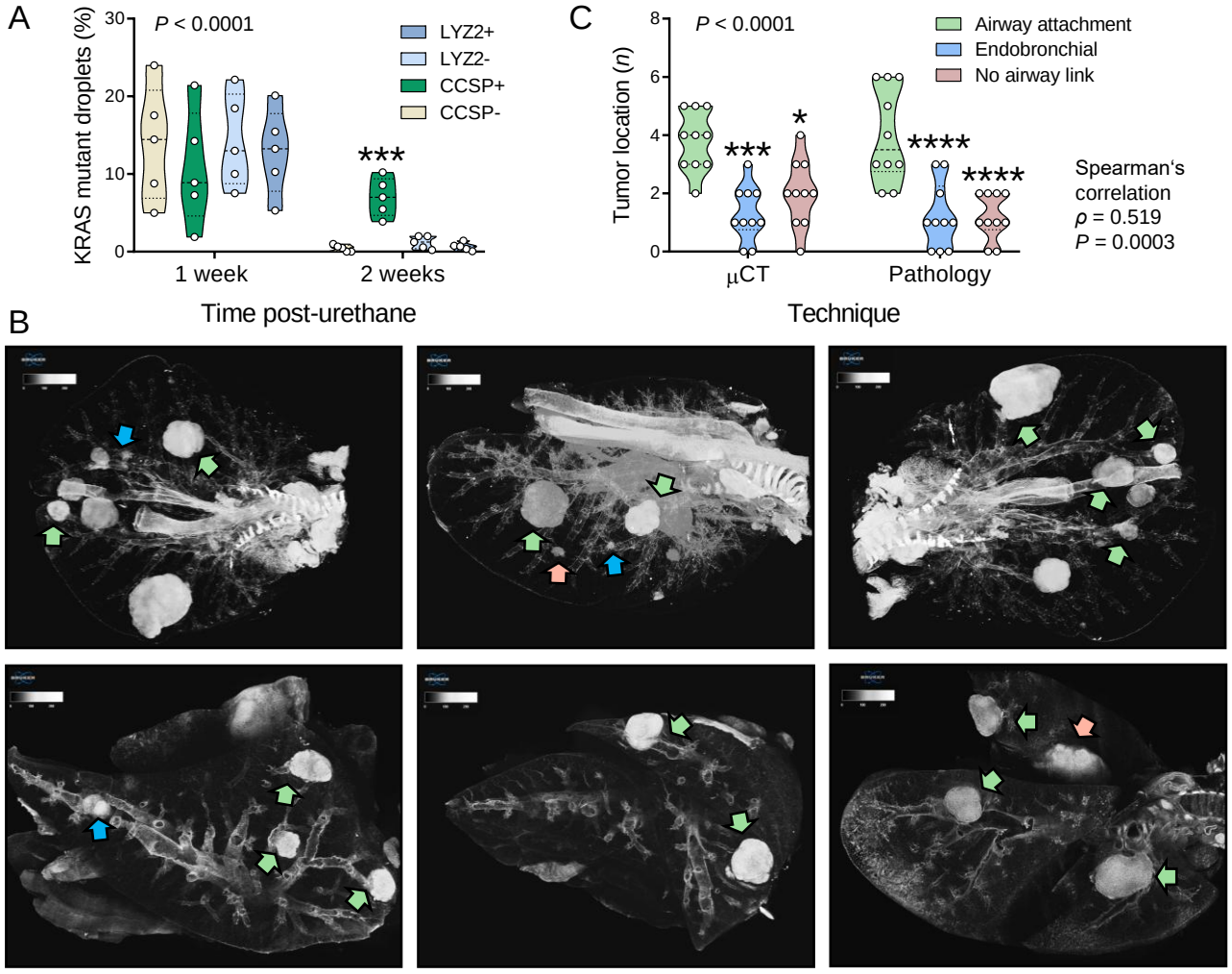

A
KRAS mutant droplets (%)
P < 0.0001
LYZ2+
LYZ2-
CCSP+
CCSP-
***
1 week
2 weeks
Time post-urethane
C
Tumor location (n)
P < 0.0001
Airway attachment
Endobronchial
No airway link
***
*
****
****
Spearman's correlation
ρ = 0.519
P = 0.0003
μCT
Pathology
Technique
B

Figure 2

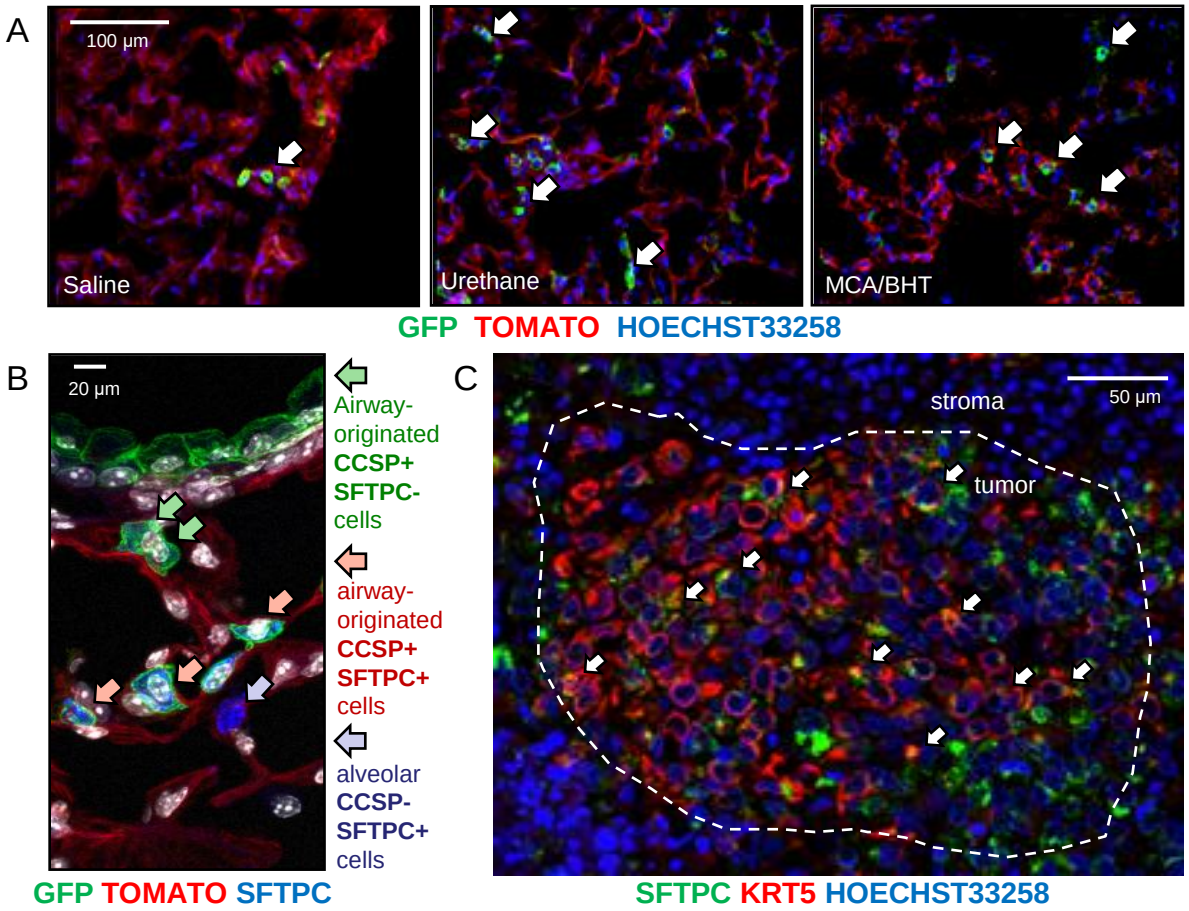

A
100 µm
Saline
Urethane
MCA/BHT
GFP TOMATO HOECHST33258
B
20 µm
Airway-originated CCSP+ SFTPC- cells
airway-originated CCSP+ SFTPC+ cells
alveolar CCSP- SFTPC+ cells
GFP TOMATO SFTPC
C
stroma
50 µm
tumor
SFTPC KRT5 HOECHST33258

Figure 3

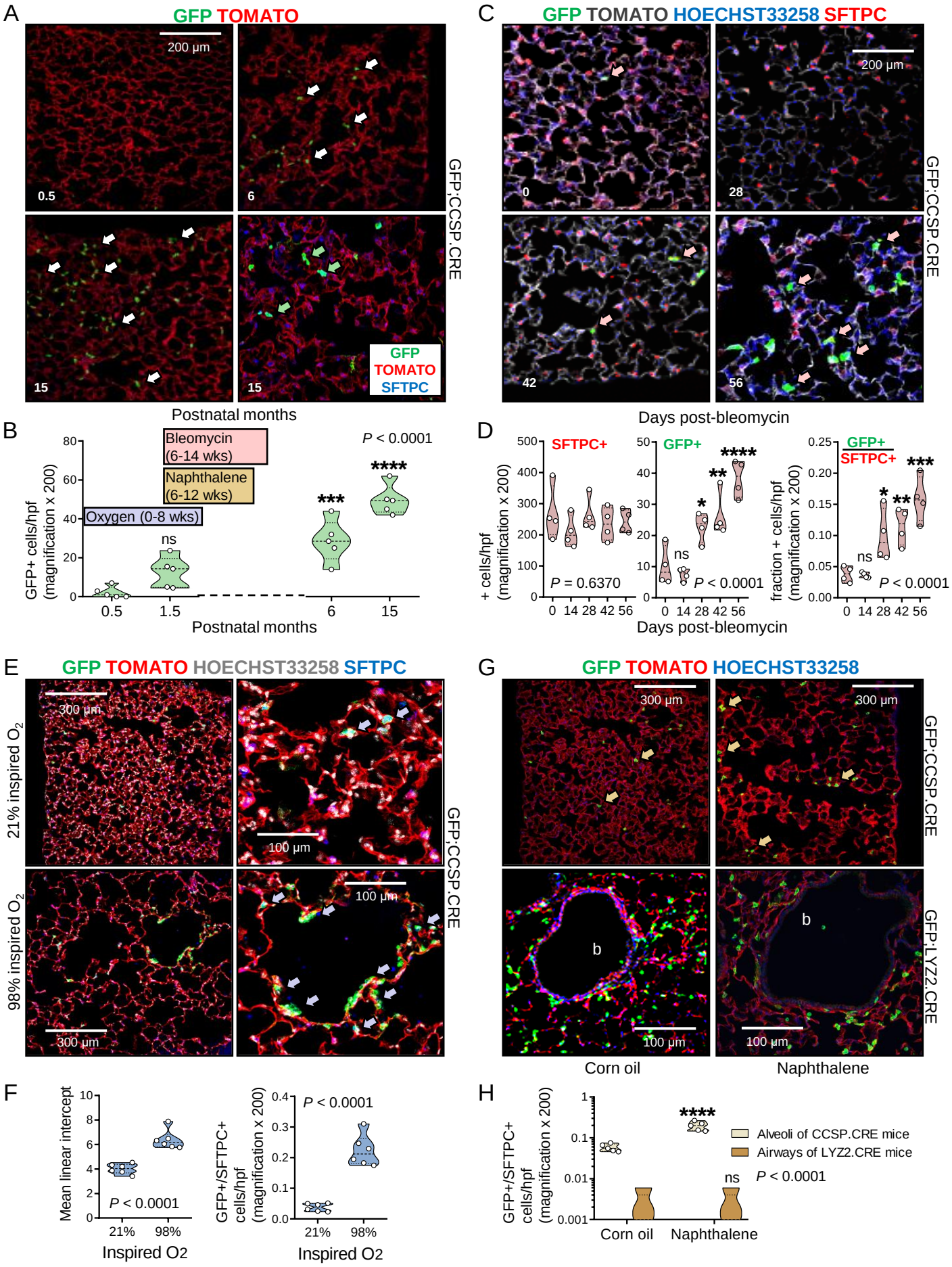

A
GFP TOMATO
200 µm
0.5
6
15
15
GFP
TOMATO
SFTPC
GFP;CCSP.CRE
Postnatal months
B
Bleomycin (6-14 wks)
Naphthalene (6-12 wks)
Oxygen (0-8 wks)
P < 0.0001
ns
***
****
GFP+ cells/hpf (magnification x 200)
Postnatal months
C
GFP TOMATO HOECHST33258 SFTPC
200 µm
0
28
42
56
GFP;CCSP.CRE
Days post-bleomycin
D
SFTPC+
GFP+
GFP+ SFTPC+
P = 0.6370
P < 0.0001
ns
*
**
***
****
+ cells/hpf (magnification x 200)
fraction + cells/hpf (magnification x 200)
Days post-bleomycin
E
GFP TOMATO HOECHST33258 SFTPC
21% inspired O2
98% inspired O2
300 µm
100 µm
GFP;CCSP.CRE
F
Mean linear intercept
P < 0.0001
21%
98%
Inspired O2
GFP+/SFTPC+ cells/hpf (magnification x 200)
P < 0.0001
Inspired O2
G
GFP TOMATO HOECHST33258
300 µm
100 µm
b
GFP;CCSP.CRE
GFP;LYZ2.CRE
Corn oil
Naphthalene
H
GFP+/SFTPC+ cells/hpf (magnification x 200)
****
ns
Alveoli of CCSP.CRE mice
Airways of LYZ2.CRE mice
P < 0.0001
Corn oil
Naphthalene

Figure 4

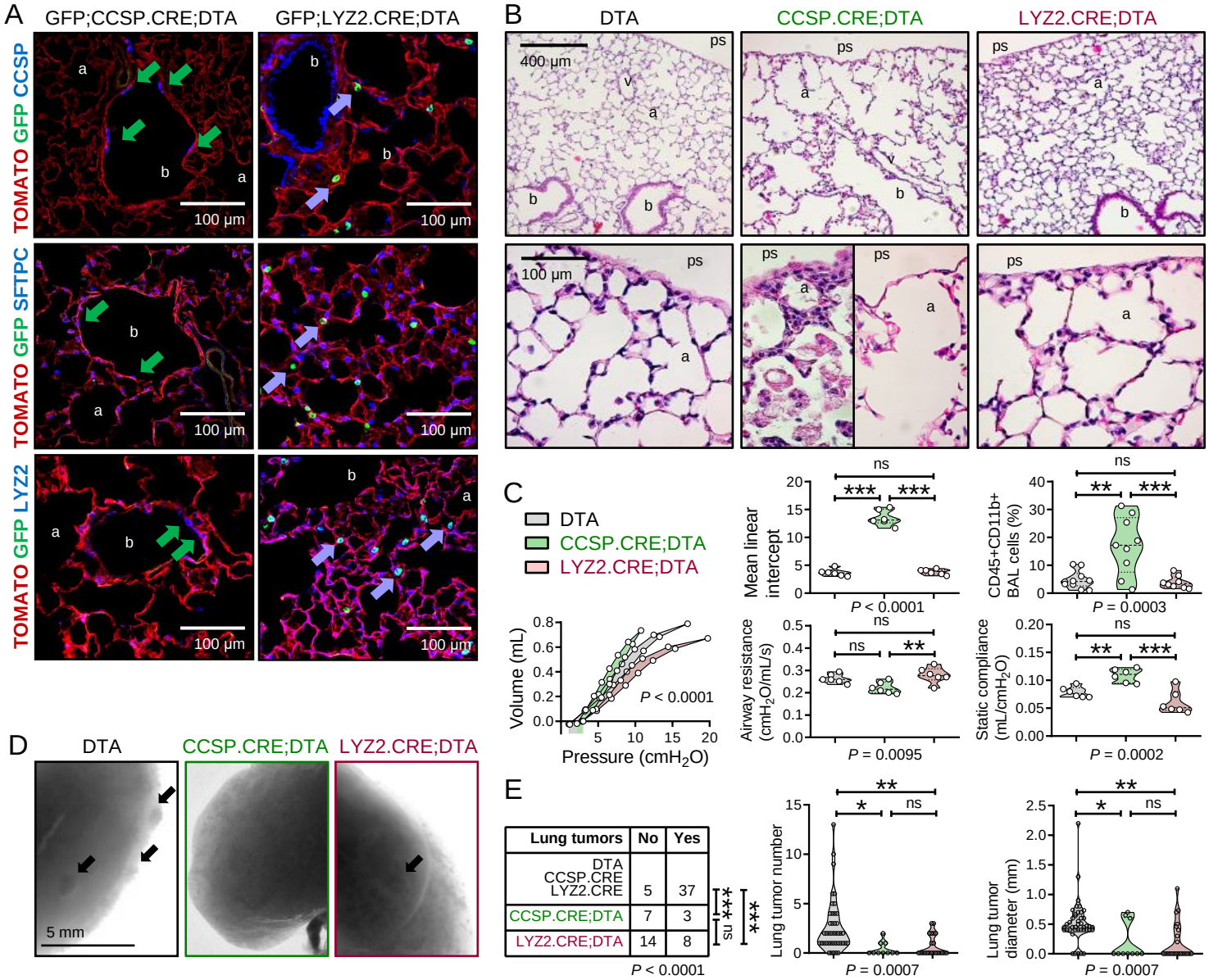

A
GFP;CCSP.CRE;DTA
GFP;LYZ2.CRE;DTA
TOMATO GFP CCSP
TOMATO GFP SFTPC
TOMATO GFP LYZ2
100 µm
B
DTA
CCSP.CRE;DTA
LYZ2.CRE;DTA
400 µm
100 µm
ps
C
DTA
CCSP.CRE;DTA
LYZ2.CRE;DTA
Volume (mL)
Pressure (cmH2O)
P < 0.0001
Mean linear intercept
P < 0.0001
CD45+CD11b+ BAL cells (%)
P = 0.0003
Airway resistance (cmH2O/mL/s)
P = 0.0095
Static compliance (mL/cmH2O)
P = 0.0002
ns
D
DTA
CCSP.CRE;DTA
LYZ2.CRE;DTA
5 mm
E
Lung tumors
No
Yes
DTA CCSP.CRE LYZ2.CRE
5
37
CCSP.CRE;DTA
7
3
LYZ2.CRE;DTA
14
8
P < 0.0001
Lung tumor number
P = 0.0007
Lung tumor diameter (mm)
P = 0.0007

Figure 5

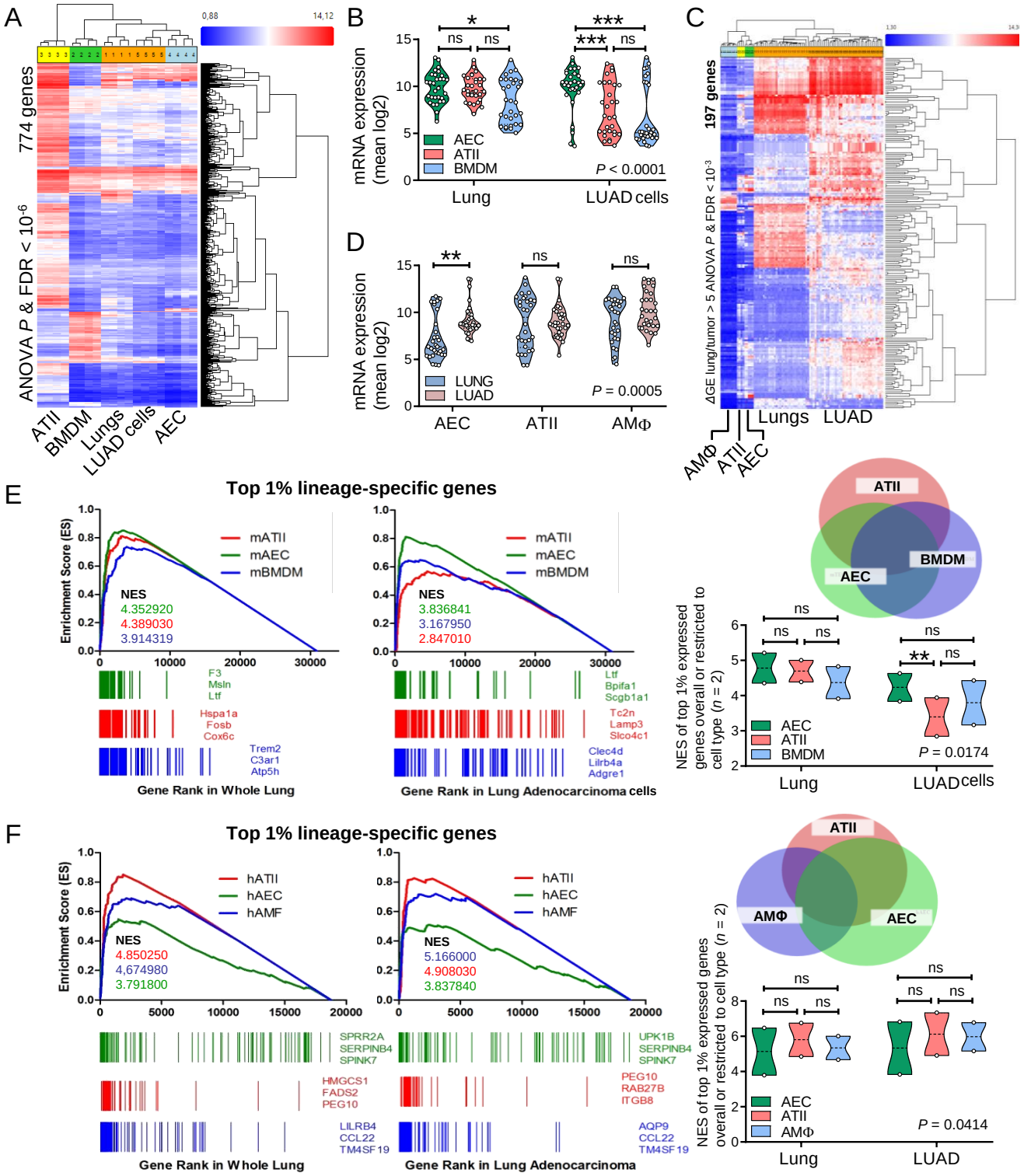

A
774 genes
ANOVA P & FDR < 10-6
ATII
BMDM
Lungs
LUAD cells
AEC
B
mRNA expression (mean log2)
AEC
ATII
BMDM
P < 0.0001
Lung
LUAD cells
C
197 genes
ΔGE lung/tumor > 5 ANOVA P & FDR < 10-3
Lungs
LUAD
AMΦ
ATII
AEC
D
mRNA expression (mean log2)
LUNG
LUAD
P = 0.0005
AEC
ATII
AMΦ
E
Top 1% lineage-specific genes
Enrichment Score (ES)
mATII
mAEC
mBMDM
NES
4.352920
4.389030
3.914319
3.836841
3.167950
2.847010
F3
Msln
Ltf
Hspa1a
Fosb
Cox6c
Trem2
C3ar1
Atp5h
Ltf
Bpifa1
Scgb1a1
Tc2n
Lamp3
Slco4c1
Clec4d
Lilrb4a
Adgre1
Gene Rank in Whole Lung
Gene Rank in Lung Adenocarcinoma cells
ATII
BMDM
AEC
NES of top 1% expressed genes overall or restricted to cell type (n = 2)
AEC
ATII
BMDM
P = 0.0174
Lung
LUAD cells
F
Top 1% lineage-specific genes
Enrichment Score (ES)
hATII
hAEC
hAMF
NES
4.850250
4,674980
3.791800
5.166000
4.908030
3.837840
SPRR2A
SERPINB4
SPINK7
HMGCS1
FADS2
PEG10
LILRB4
CCL22
TM4SF19
UPK1B
SERPINB4
SPINK7
PEG10
RAB27B
ITGB8
AQP9
CCL22
TM4SF19
Gene Rank in Whole Lung
Gene Rank in Lung Adenocarcinoma
ATII
AMΦ
AEC
NES of top 1% expressed genes overall or restricted to cell type (n = 2)
AEC
ATII
AMΦ
P = 0.0414
Lung
LUAD

Figure 6

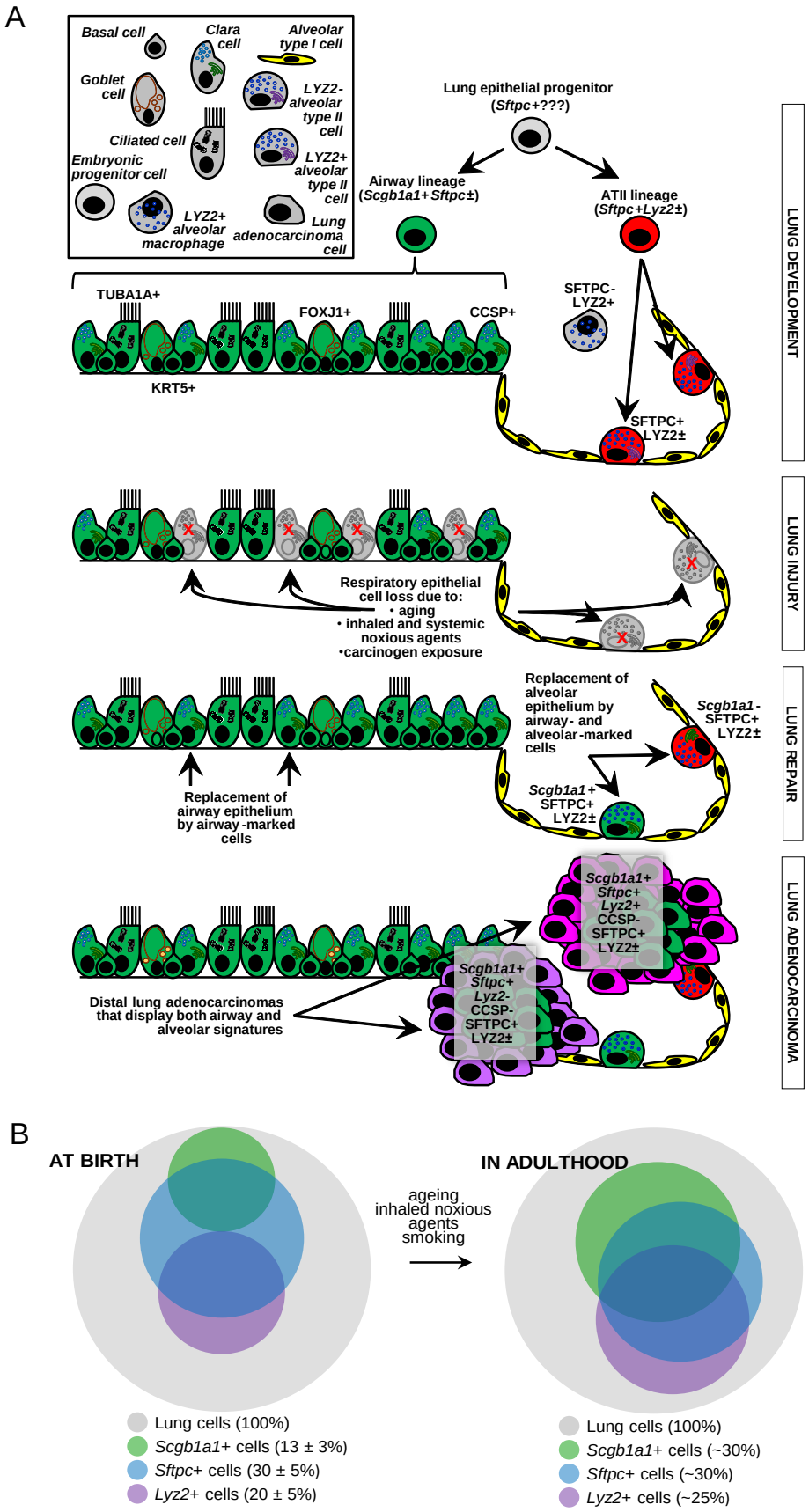

A
Basal cell
Clara cell
Alveolar type I cell
Goblet cell
LYZ2- alveolar type II cell
Ciliated cell
LYZ2+ alveolar type II cell
Embryonic progenitor cell
LYZ2+ alveolar macrophage
Lung adenocarcinoma cell
Lung epithelial progenitor (Sftpc+???)
Airway lineage (Scgb1a1+ Sftpc±)
ATII lineage (Sftpc+Lyz2±)
TUBA1A+
FOXJ1+
CCSP+
KRT5+
SFTPC- LYZ2+
SFTPC+ LYZ2±
LUNG DEVELOPMENT
Respiratory epithelial cell loss due to:
• aging
• inhaled and systemic noxious agents
•carcinogen exposure
LUNG INJURY
Replacement of alveolar epithelium by airway- and alveolar-marked cells
Scgb1a1- SFTPC+ LYZ2±
Scgb1a1+ SFTPC+ LYZ2±
Replacement of airway epithelium by airway-marked cells
LUNG REPAIR
Scgb1a1+ Sftpc+ Lyz2+ CCSP- SFTPC+ LYZ2±
Scgb1a1+ Sftpc+ Lyz2- CCSP- SFTPC+ LYZ2±
Distal lung adenocarcinomas that display both airway and alveolar signatures
LUNG ADENOCARCINOMA
B
AT BIRTH
IN ADULTHOOD
ageing
inhaled noxious agents
smoking
Lung cells (100%)
Scgb1a1+ cells (13 ± 3%)
Sftpc+ cells (30 ± 5%)
Lyz2+ cells (20 ± 5%)
Lung cells (100%)
Scgb1a1+ cells (~30%)
Sftpc+ cells (~30%)
Lyz2+ cells (~25%)

Figure 7

Marker Protein	Official Name, Aliases	Target Lineage	Coding Gene	
			Human	Mouse
TUBA1A	Tubulin, alpha 1a, acetylated tubulin	Ciliated AEC	*TUBA1A*	*Tuba1a*
KRT5	Keratin 5	Basal AEC	*KRT5*	*Krt5*
FOXJ1	Forkhead box J1	Goblet AEC	*FOXJ1*	*Foxj1*
CCSP	Secretoglobin, family 1A, member 1 (uteroglobin), Clara cell secretory protein, Clara cell 10 KDa protein	Club AEC, BASC	*SCGB1A1*	*Scgb1a1*
SFTPC	Surfactant protein C	ATII, BASC	*SFTPC*	*Sftpc*
LYZ2	Lysozyme 2	ATII, AMΦ	*LYZ2*	*Lyz2*

Strain Category		CRE reporter	CRE reporter x CRE driver intercrosses					
Short Name		TOMATO	GFP; CCSP. Cre	GFP; SFTPC. Cre	GFP; LYZ2. Cre	GFP; SOX2. Cre	GFP; VAV. Cre	GFP; NES. Cre
Full Name		B6.129(Cg)-Gt(ROSA)26Sortm4(ACTB-tdTomato,-EGFP)Luo/J	B6;CBA-Tg(Scgb1a1-cre)1Vart/Flmg	Tg(Sftpc-cre)1Blh	B6.129P2-Lyz2tm1(cre)Ifo/J	B6.Cg-Tg(Sox2-cre)1Amc/J	B6.Cg-Tg(Vav1-icre)A2Kio/J	B6.Cg-Tg(Nes-cre)1Kln/J
Reference		[25]	[26]	[27]	[18]	[28]	[29]	[30]
ID		JAX # 007676	EMMA # EM:04965	MGI # 3574949	JAX # 004781	JAX # 008454	JAX # 008610	JAX # 003771
Background tested		C57BL/6 FVB	C57BL/6 FVB	C57BL/6 FVB	C57BL/6 FVB	C57BL/6	C57BL/6	C57BL/6
Cells Labeled	AEC	-	+	±	-	+	-	±
	ATII	-	-	+	±	+	-	-
	AMΦ	-	-	-	+	+	+	-
	BASC	-	+	+	-	+	-	-
	BM	-	-	-	±	+	+	-

Figure 1 - Figure Supplement 2

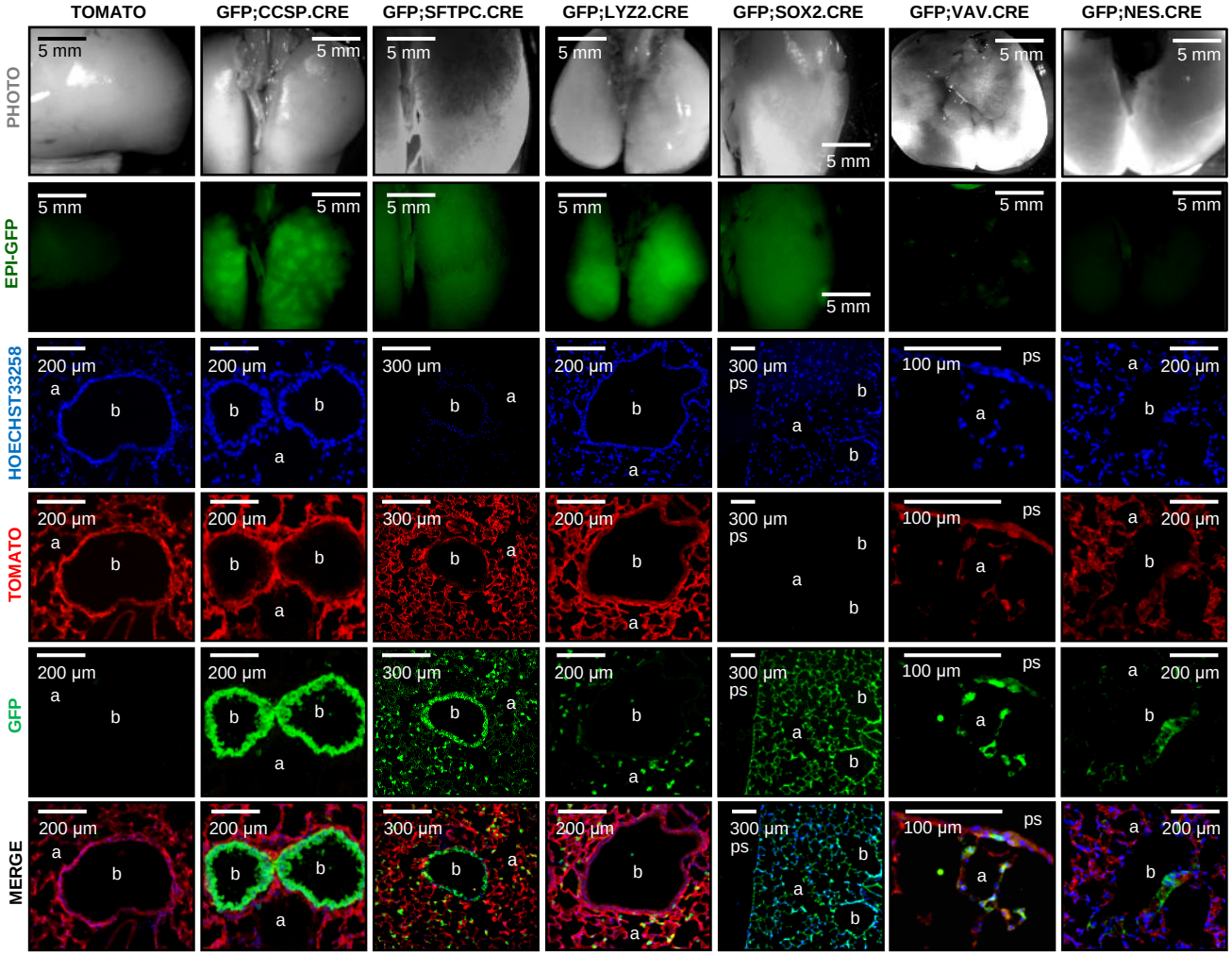


Figure 1 - Figure Supplement 3

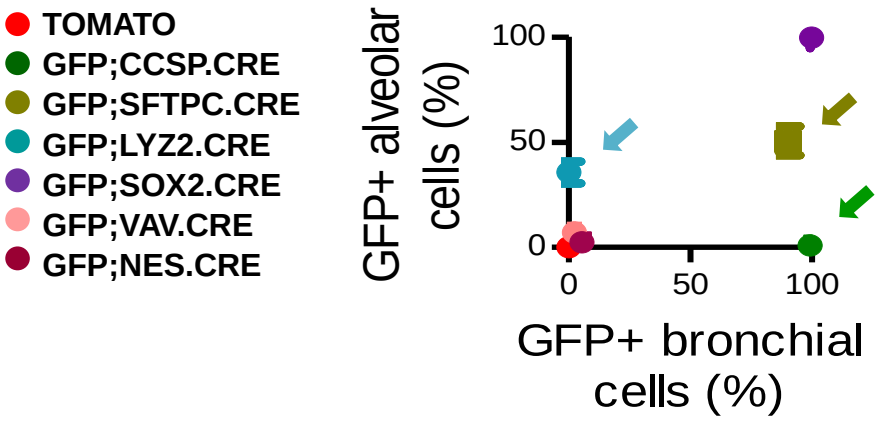


Figure 1 - Figure Supplement 4

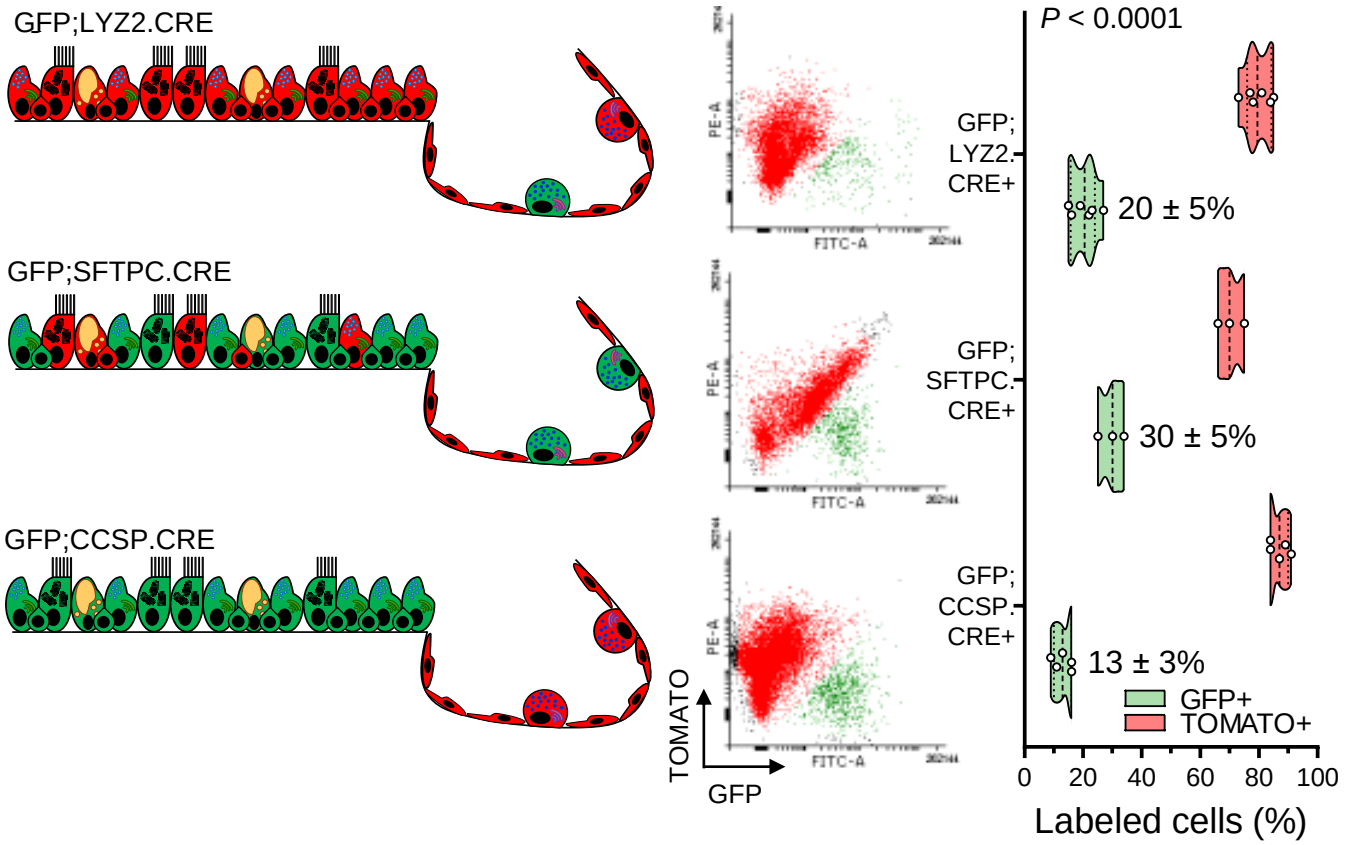


Figure 1 - Figure Supplement 5

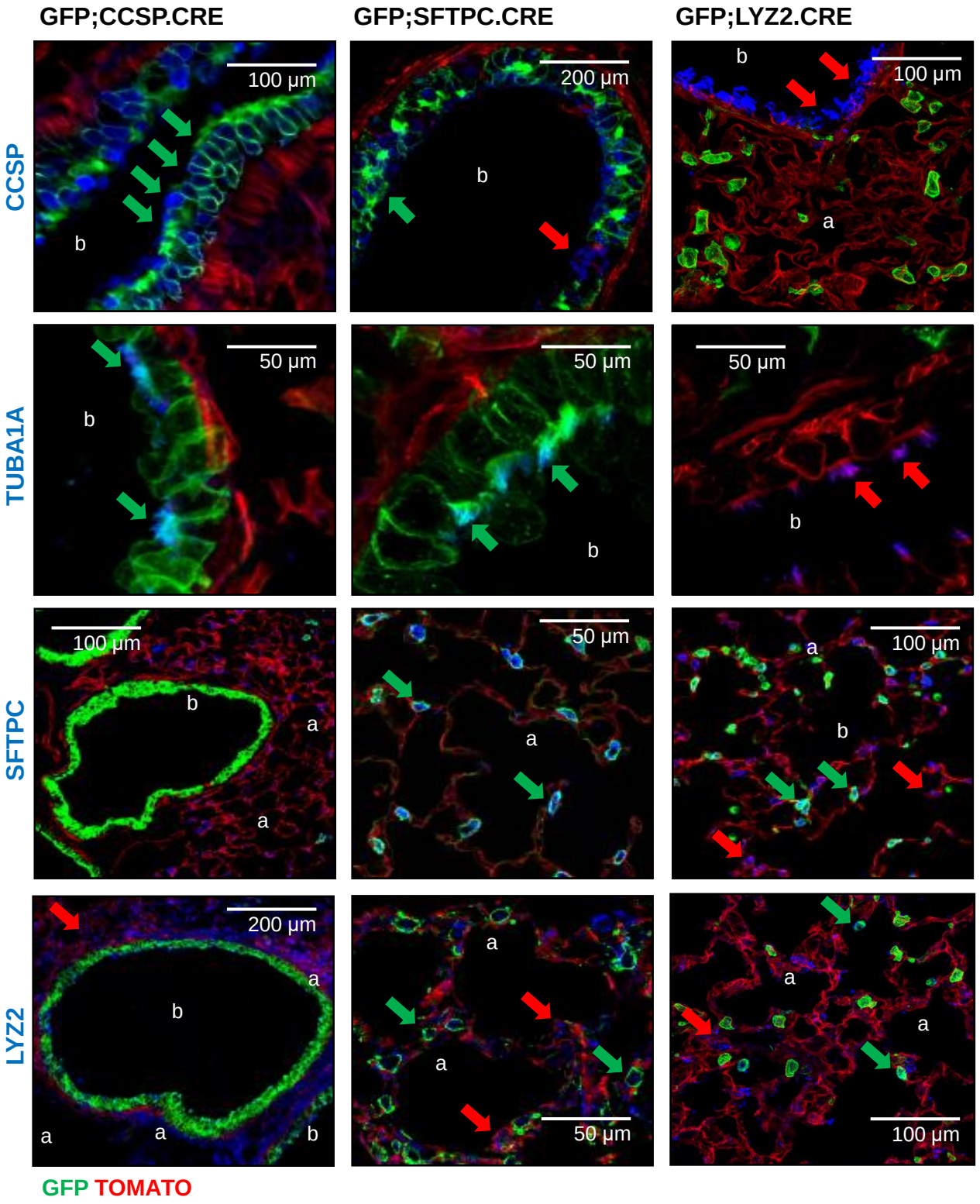


Figure 1 - Figure Supplement 6

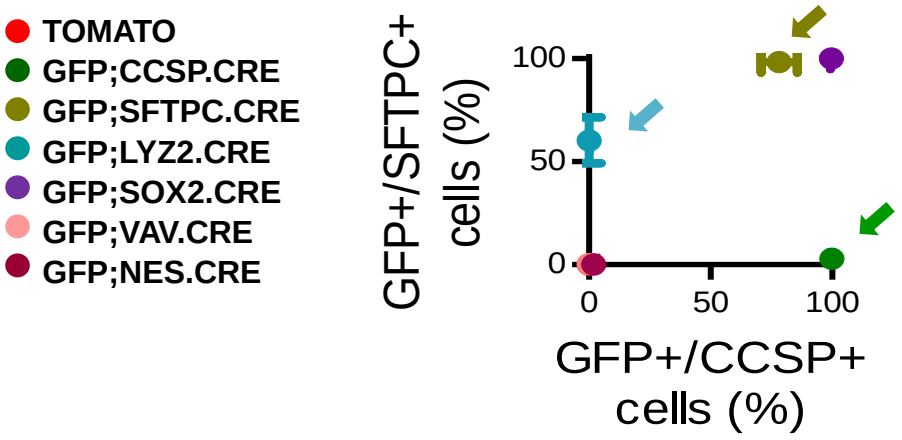


Figure 1 - Figure Supplement 7

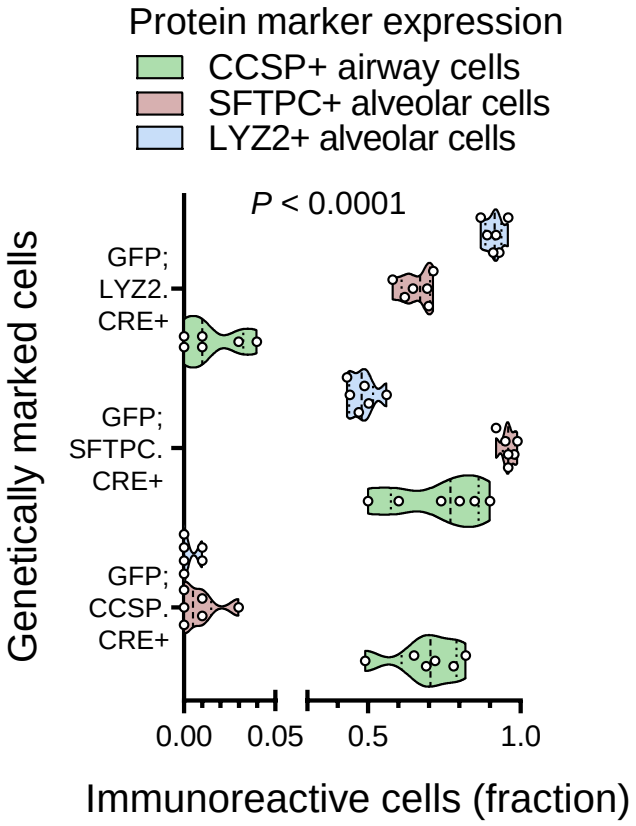


Figure 1 - Figure Supplement 8

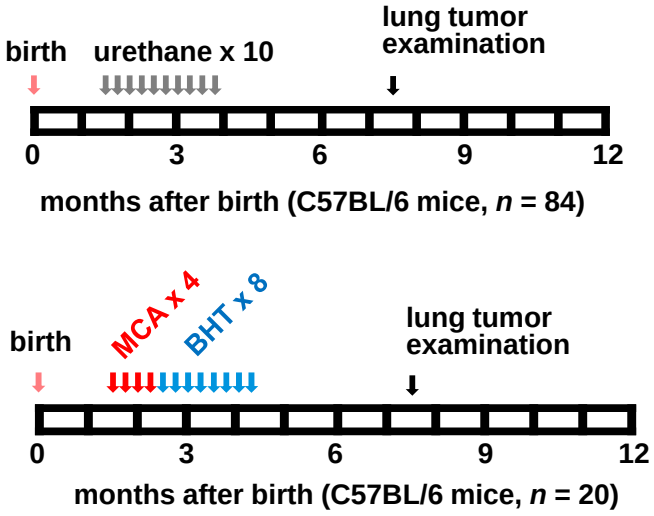


Figure 1 - Figure Supplement 9

	Lung Tumors	No	Yes	%
o	EC	9	75	89
o	MCA/BHT	2	18	90

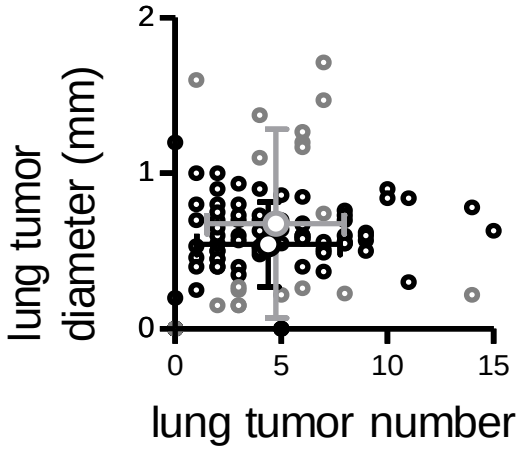


Figure 1 - Figure Supplement 10

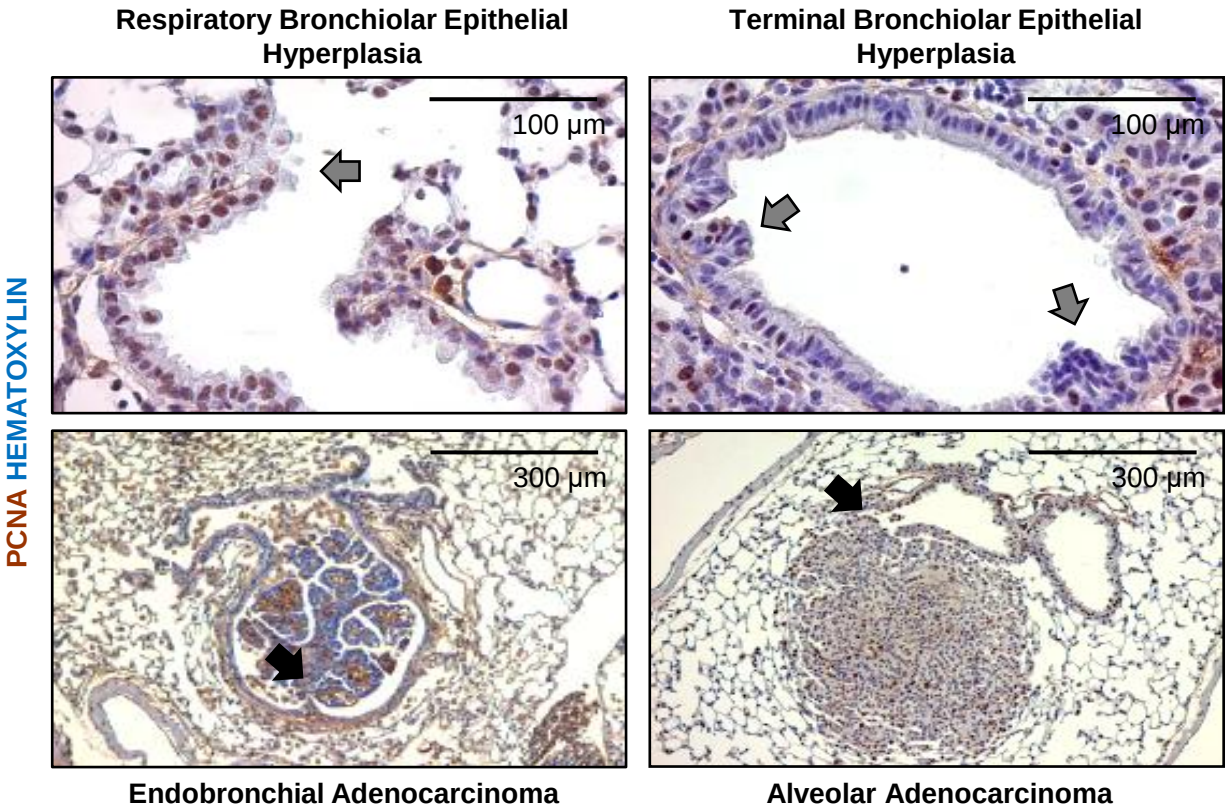


Figure 1 - Figure Supplement 11

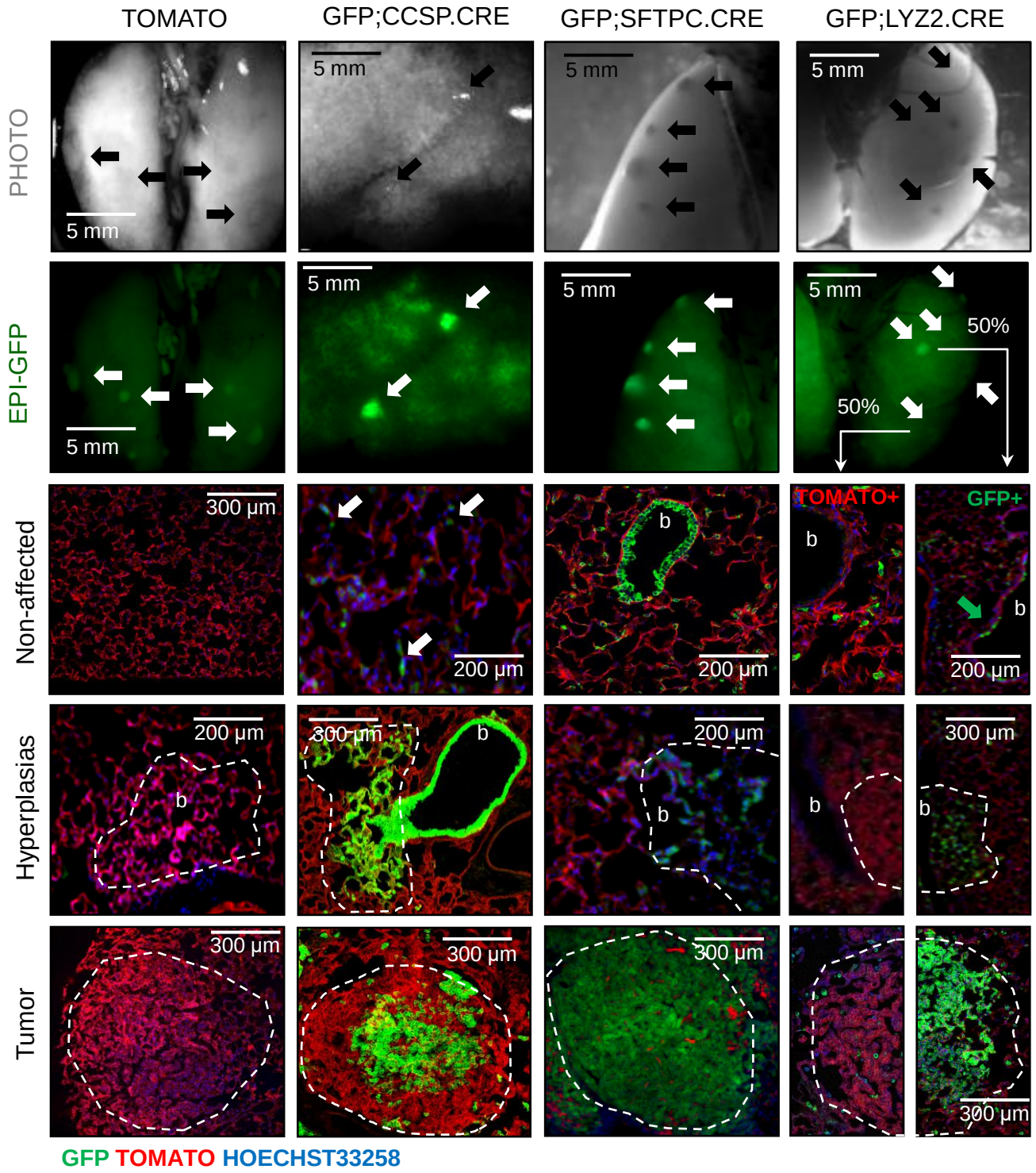


Figure 1 - Figure Supplement 12

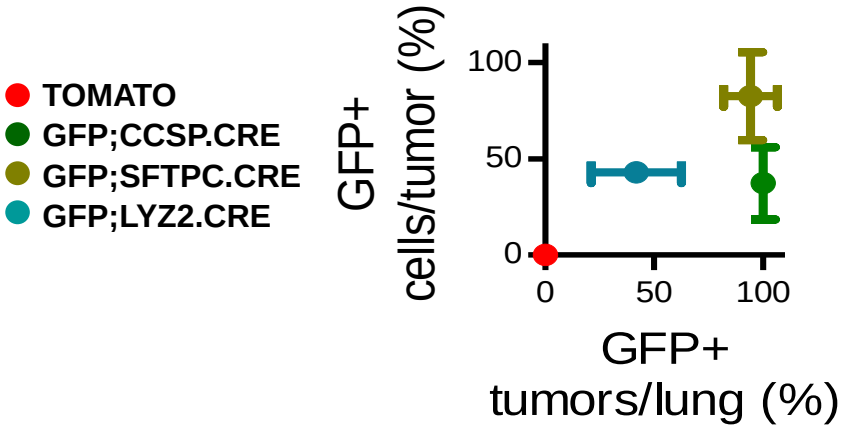


Figure 1 - Figure Supplement 13

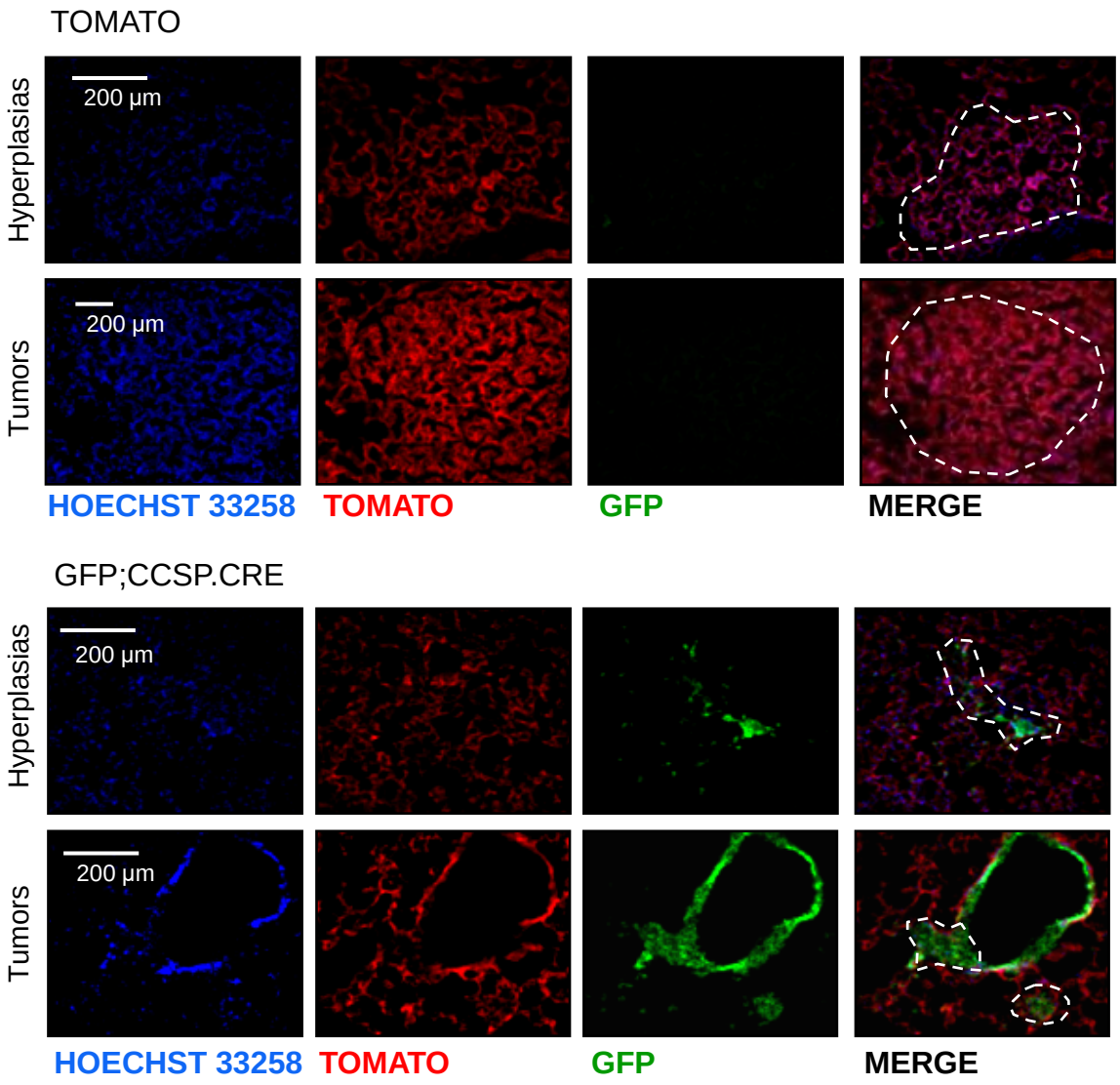


Figure 1 - Figure Supplement 14

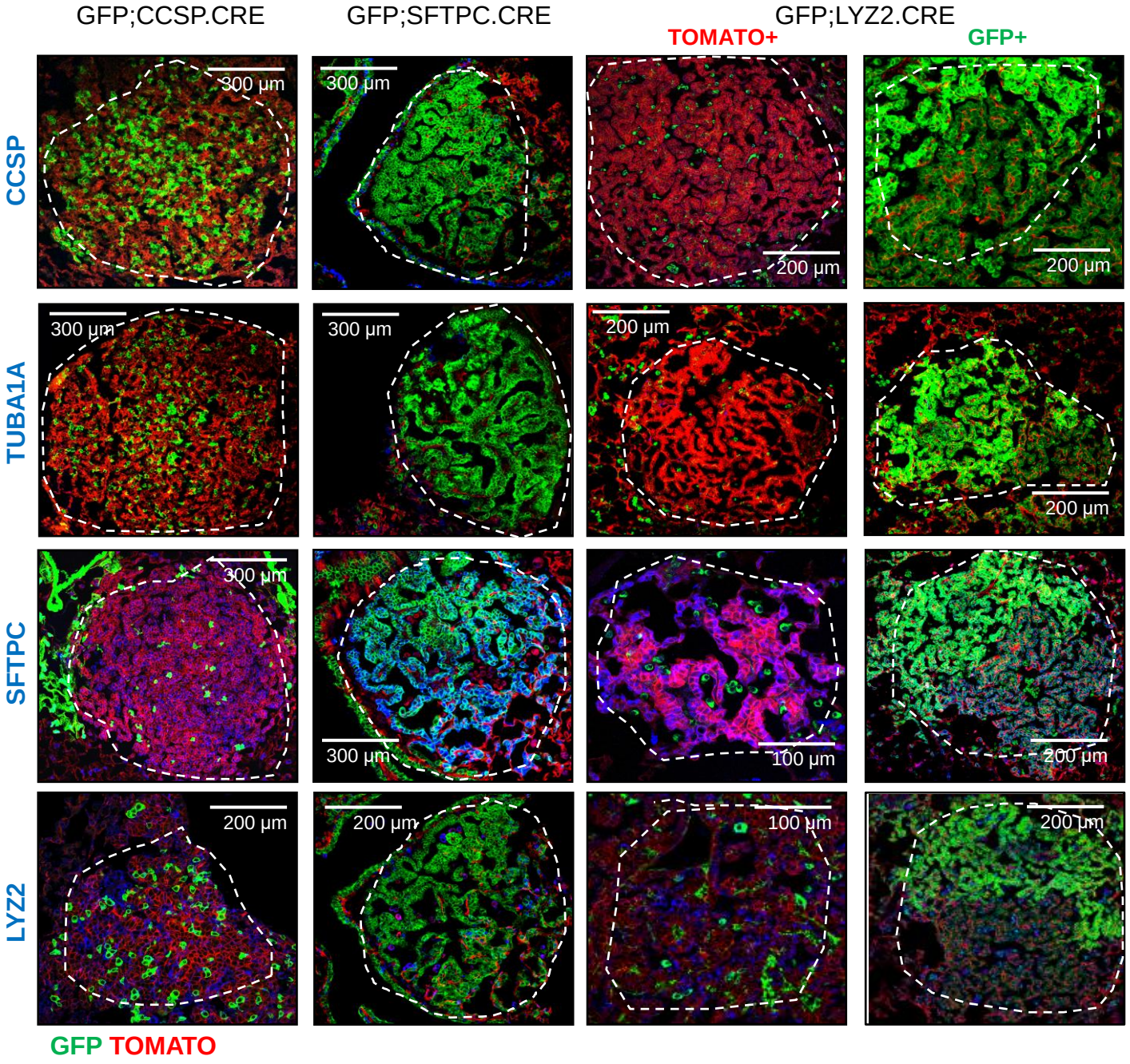


Figure 1 - Figure Supplement 15

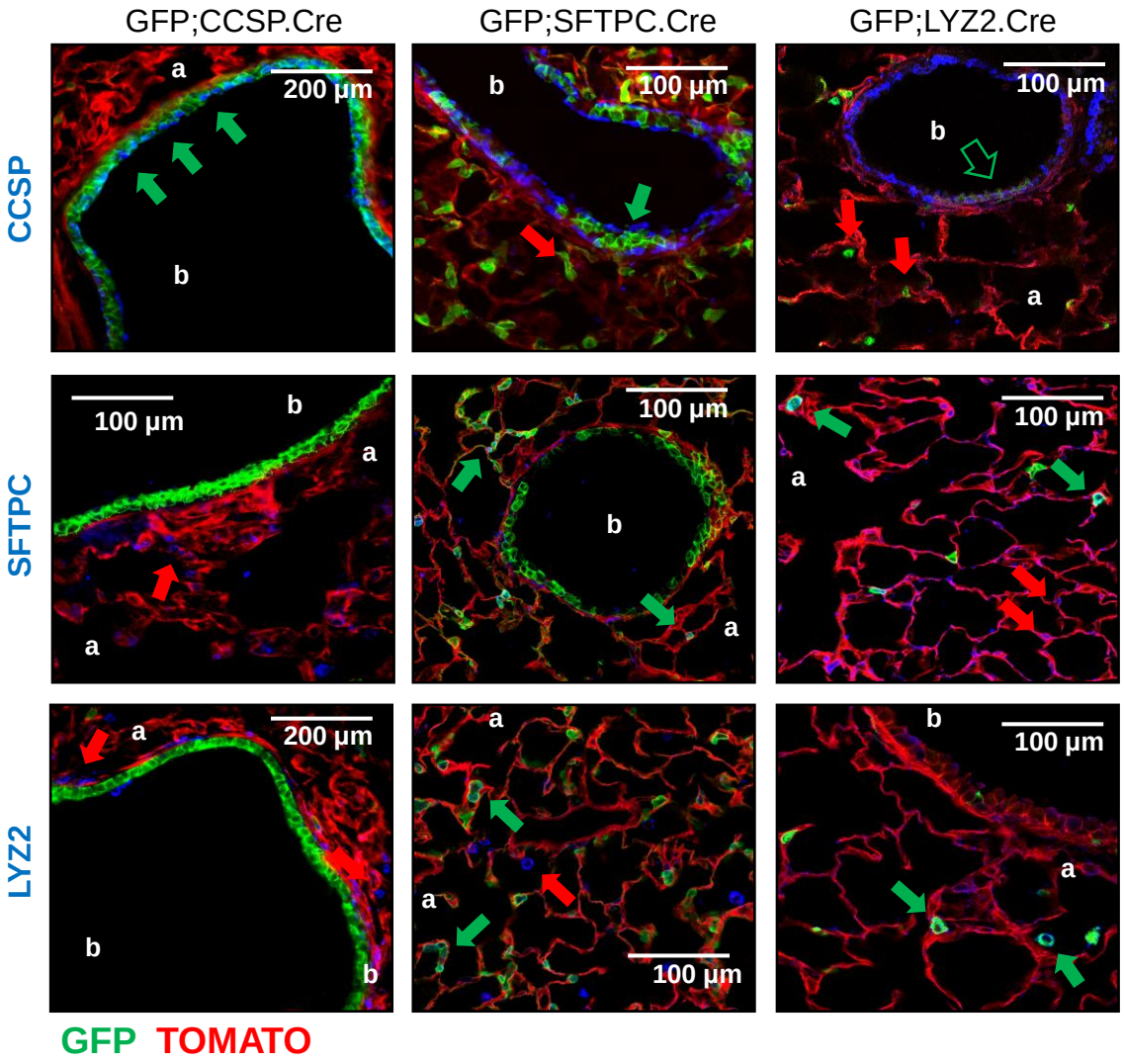


Figure 1 - Figure Supplement 16

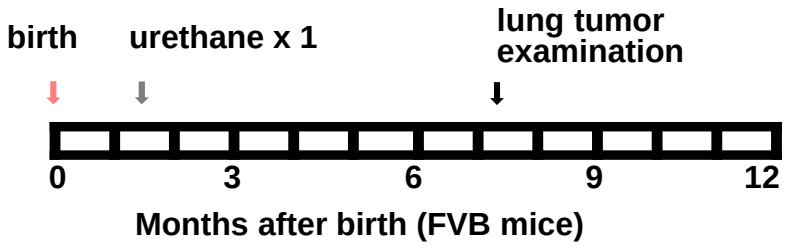


Figure 1 - Figure Supplement 17

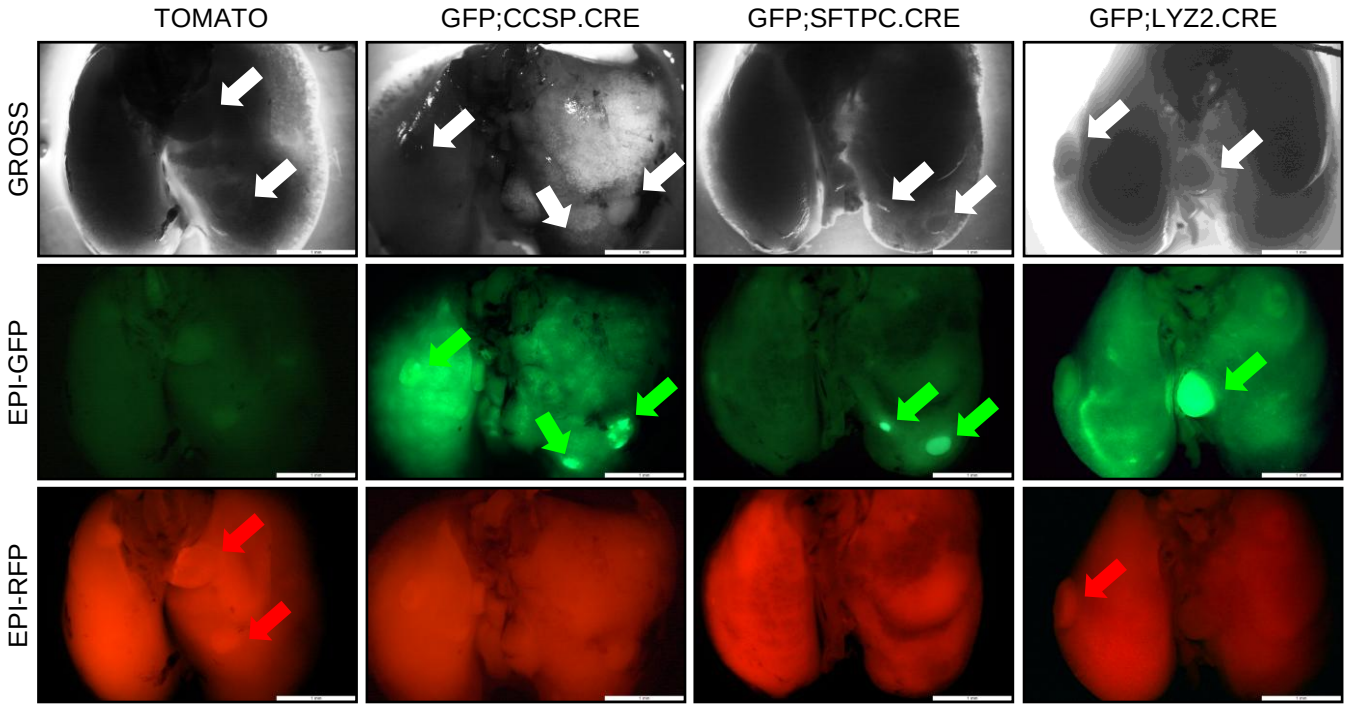


Figure 1 - Figure Supplement 18

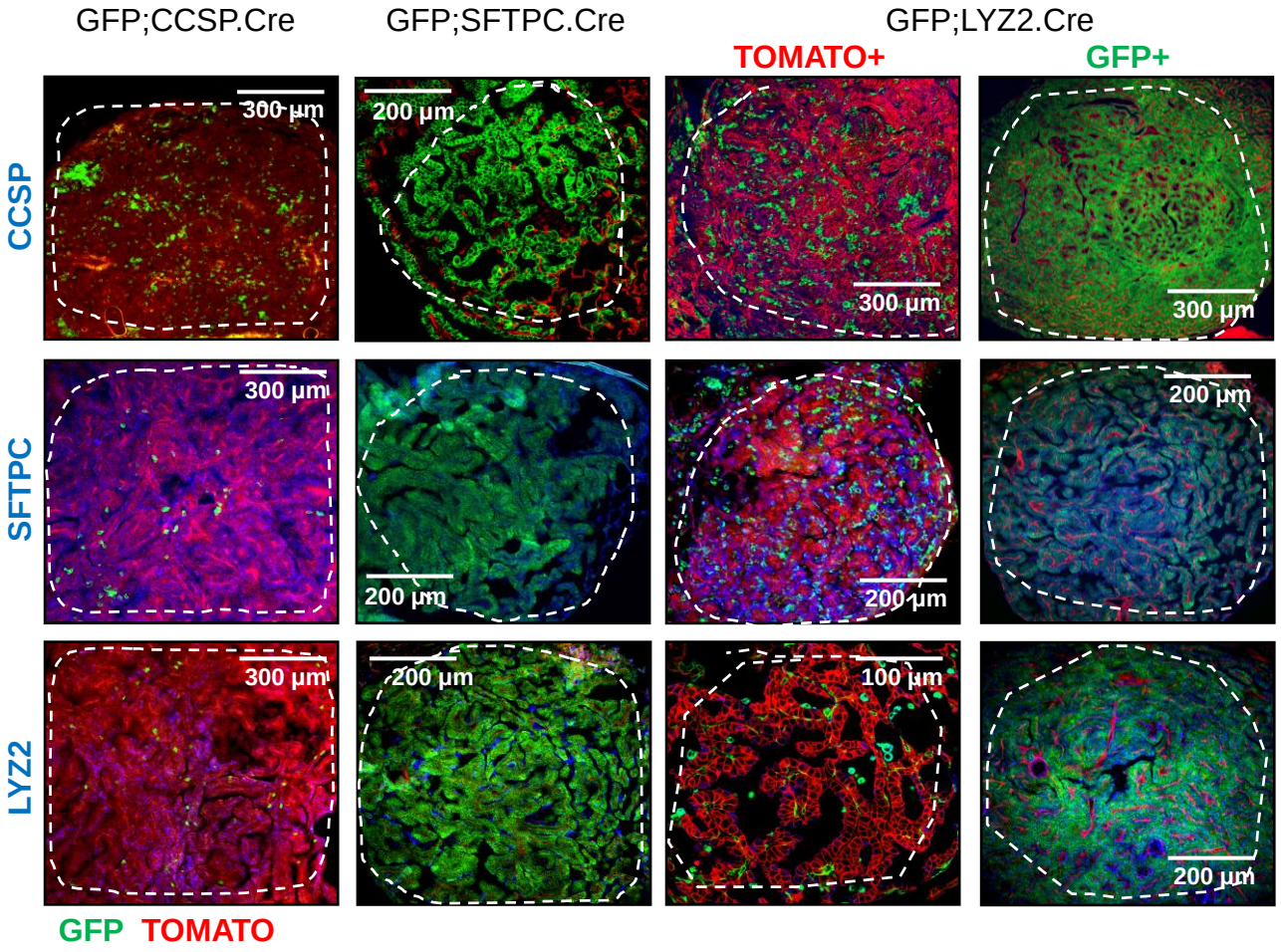


Figure 1 - Figure Supplement 19

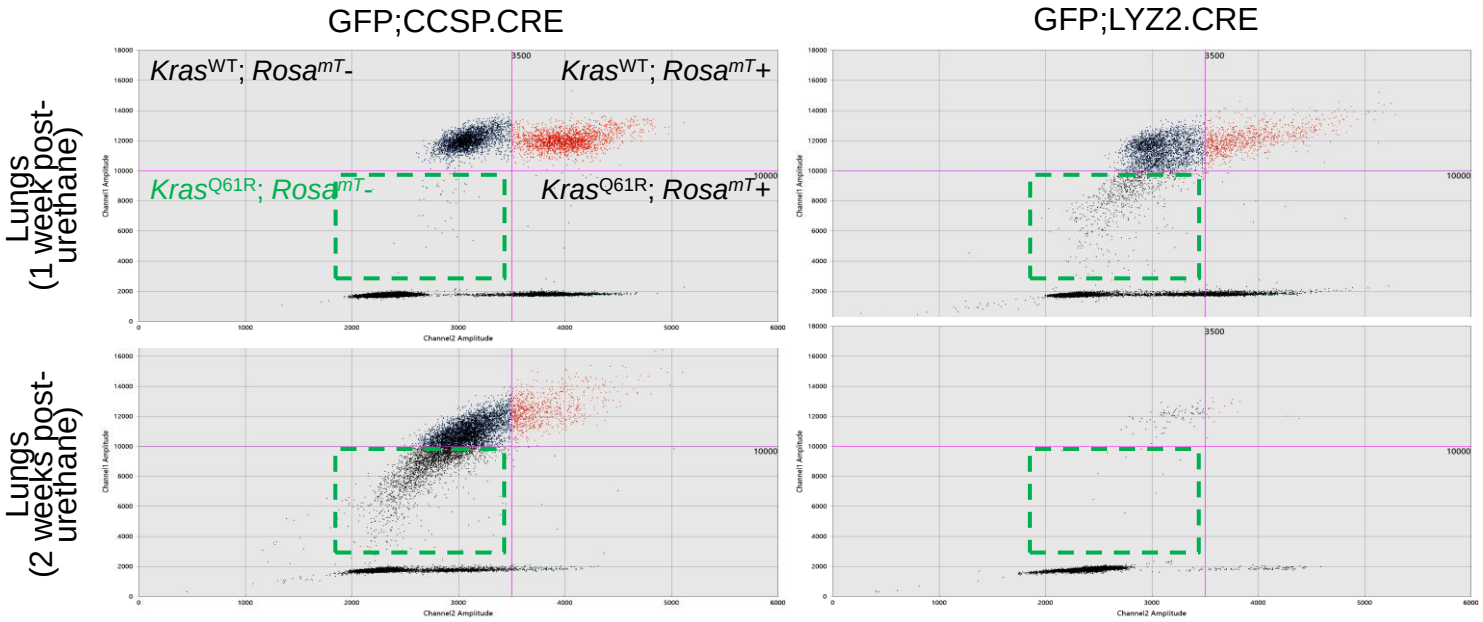


Figure 2 - Figure Supplement 1

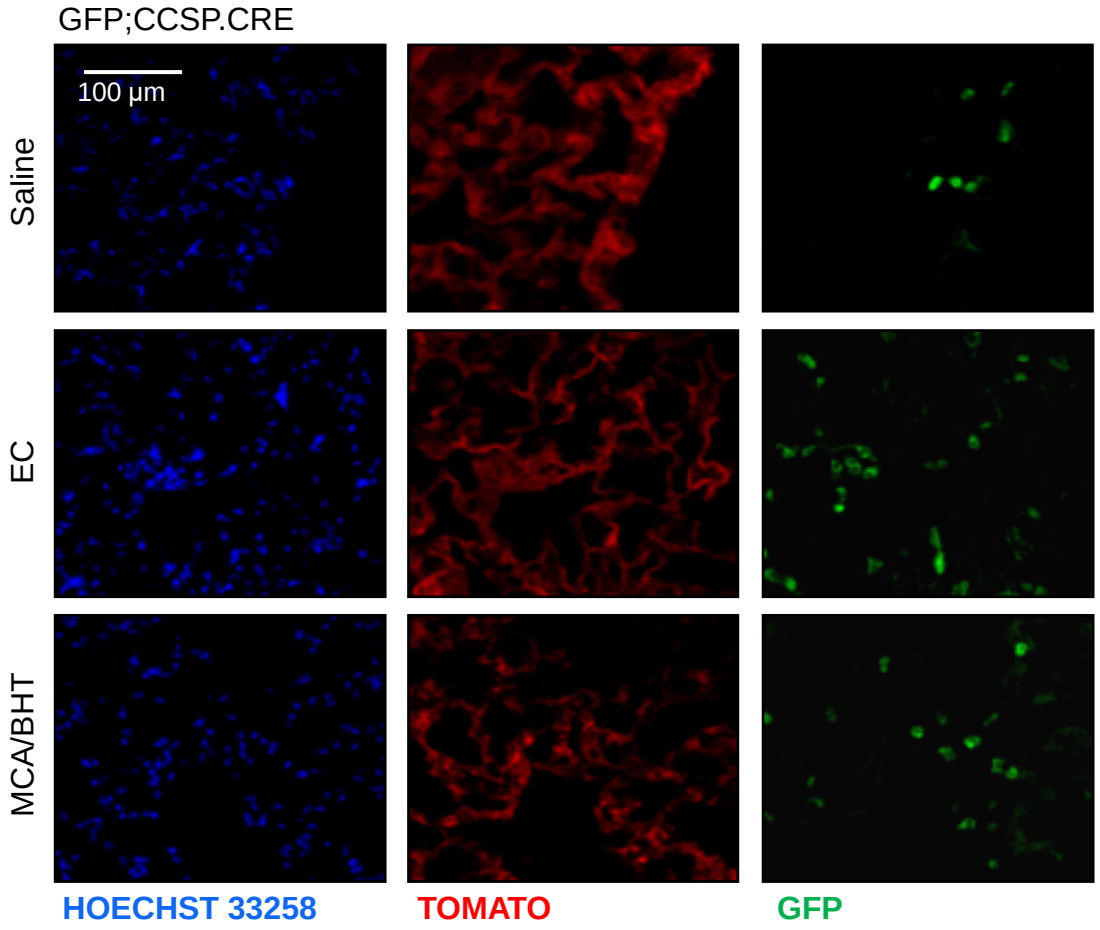


Figure 3 - Figure Supplement 1

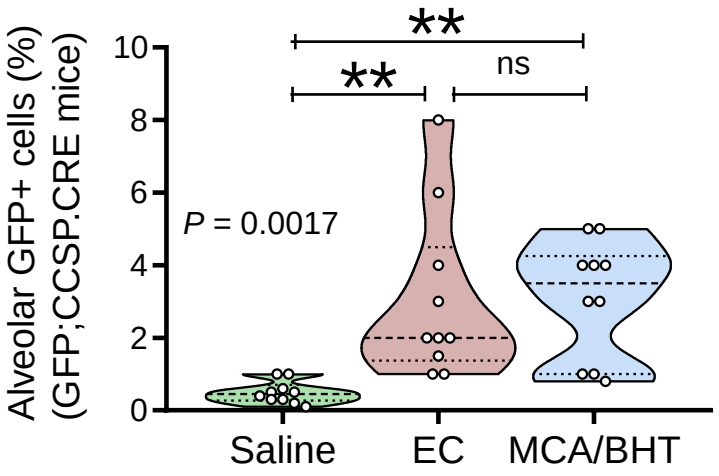


Figure 3 - Figure Supplement 2

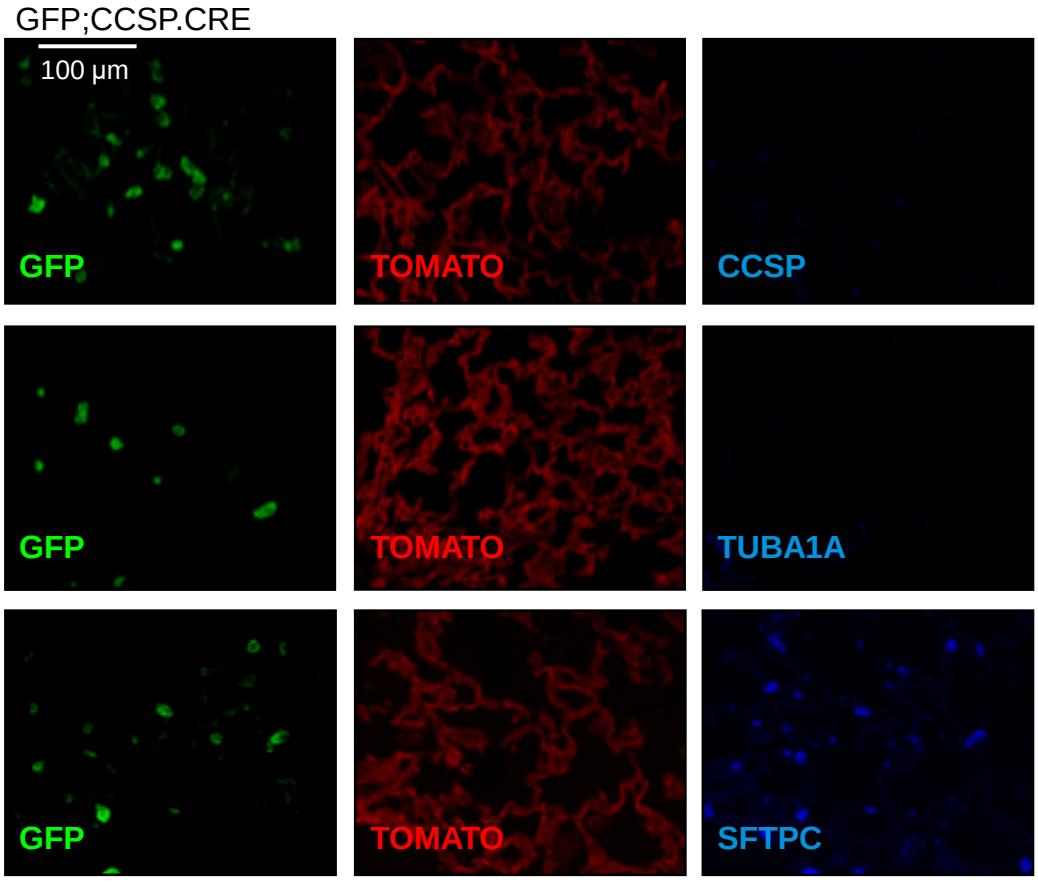


Figure 3 - Figure Supplement 3

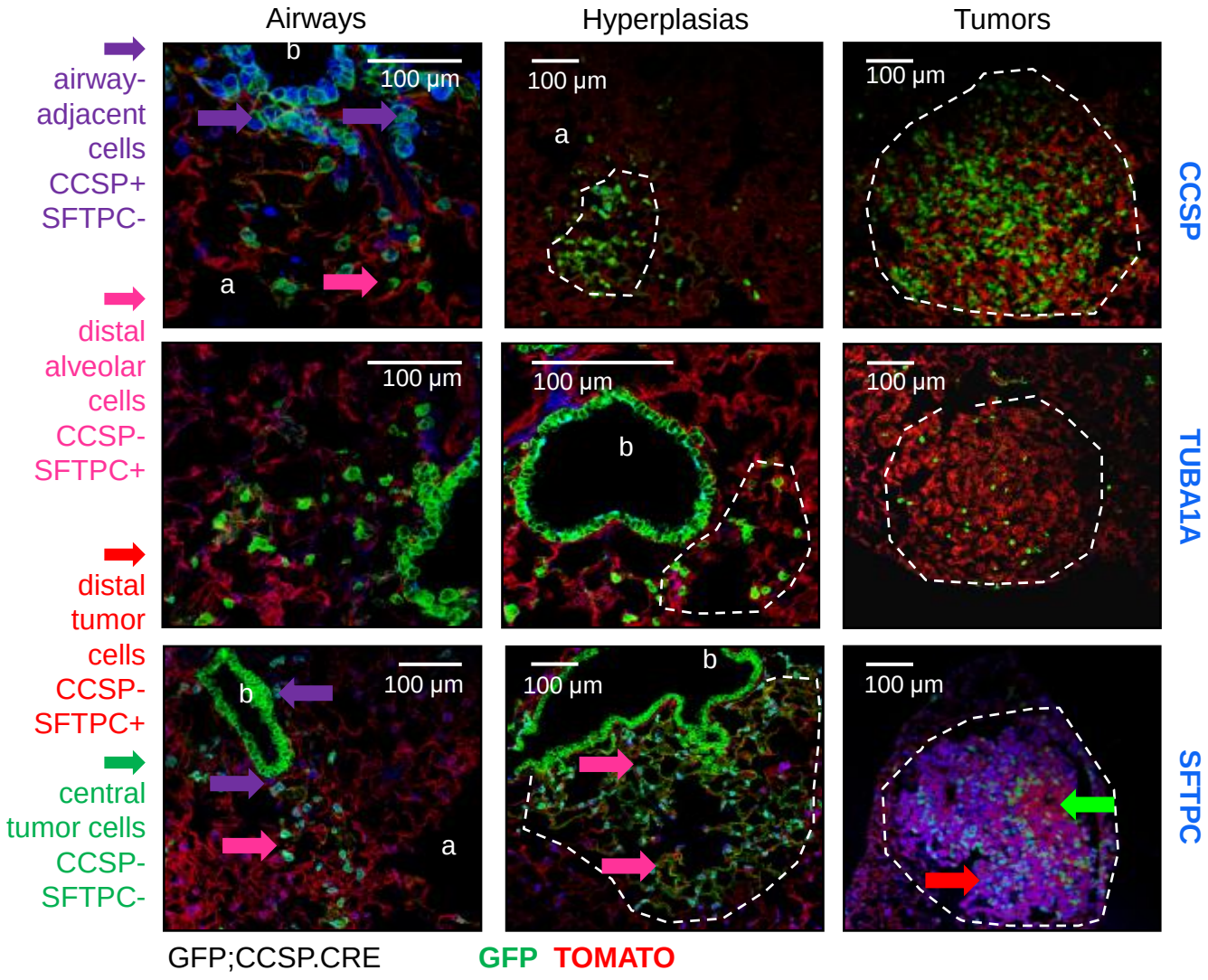


Figure 3 - Figure Supplement 4

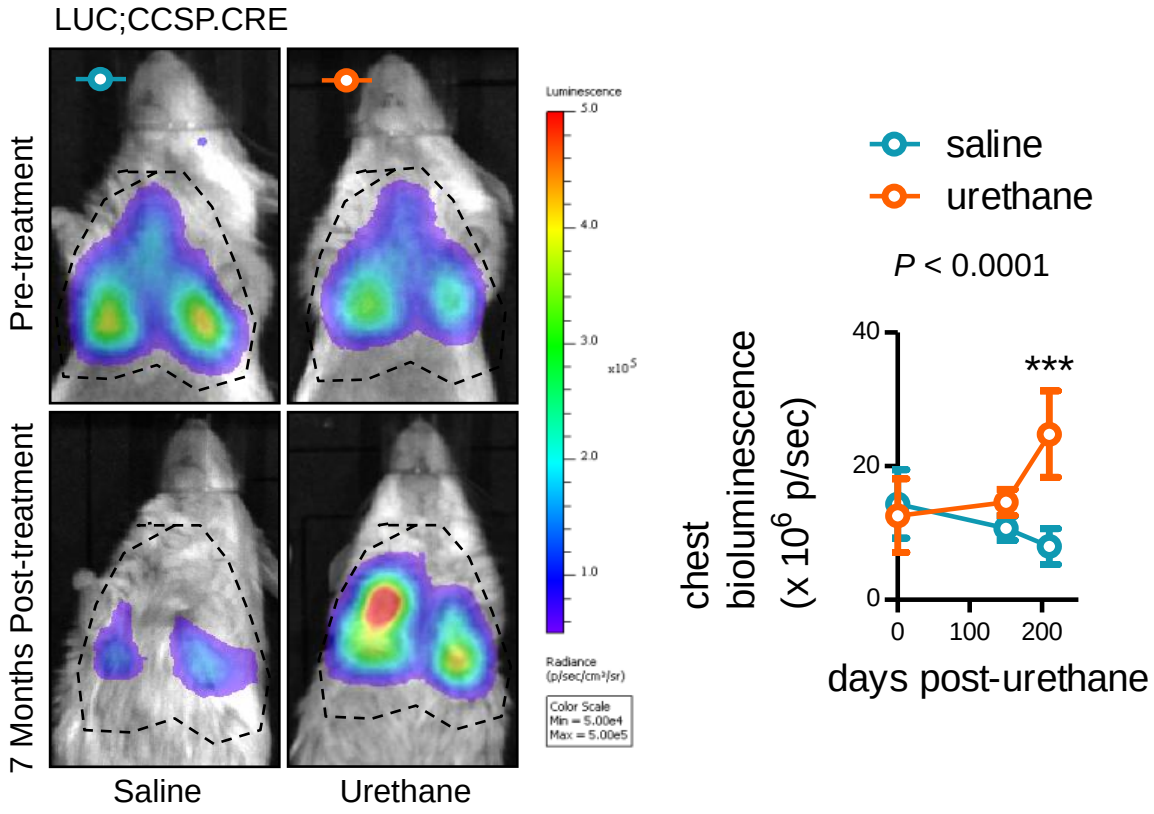


Figure 3 - Figure Supplement 5

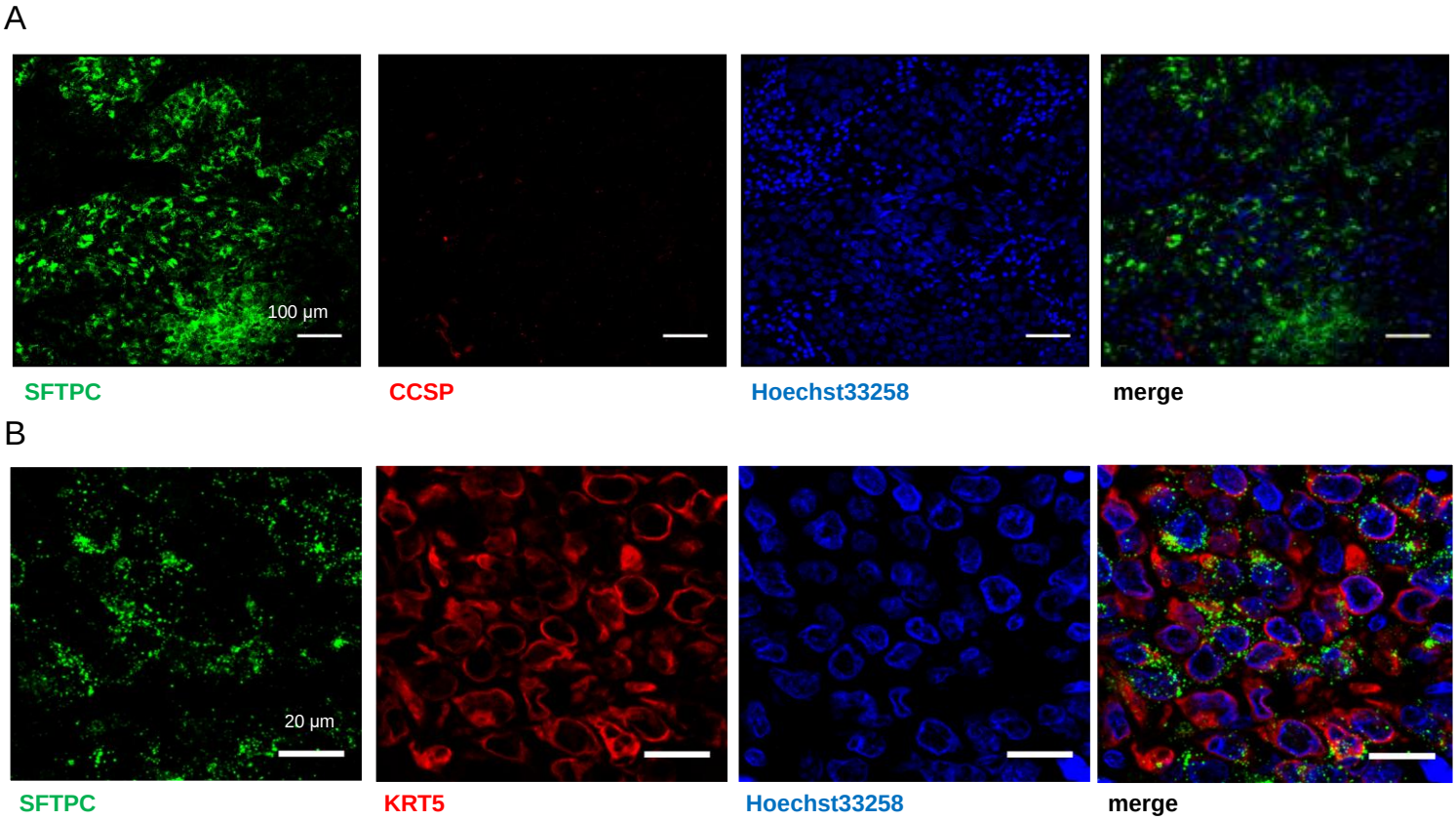


Figure 3 - Figure Supplement 6

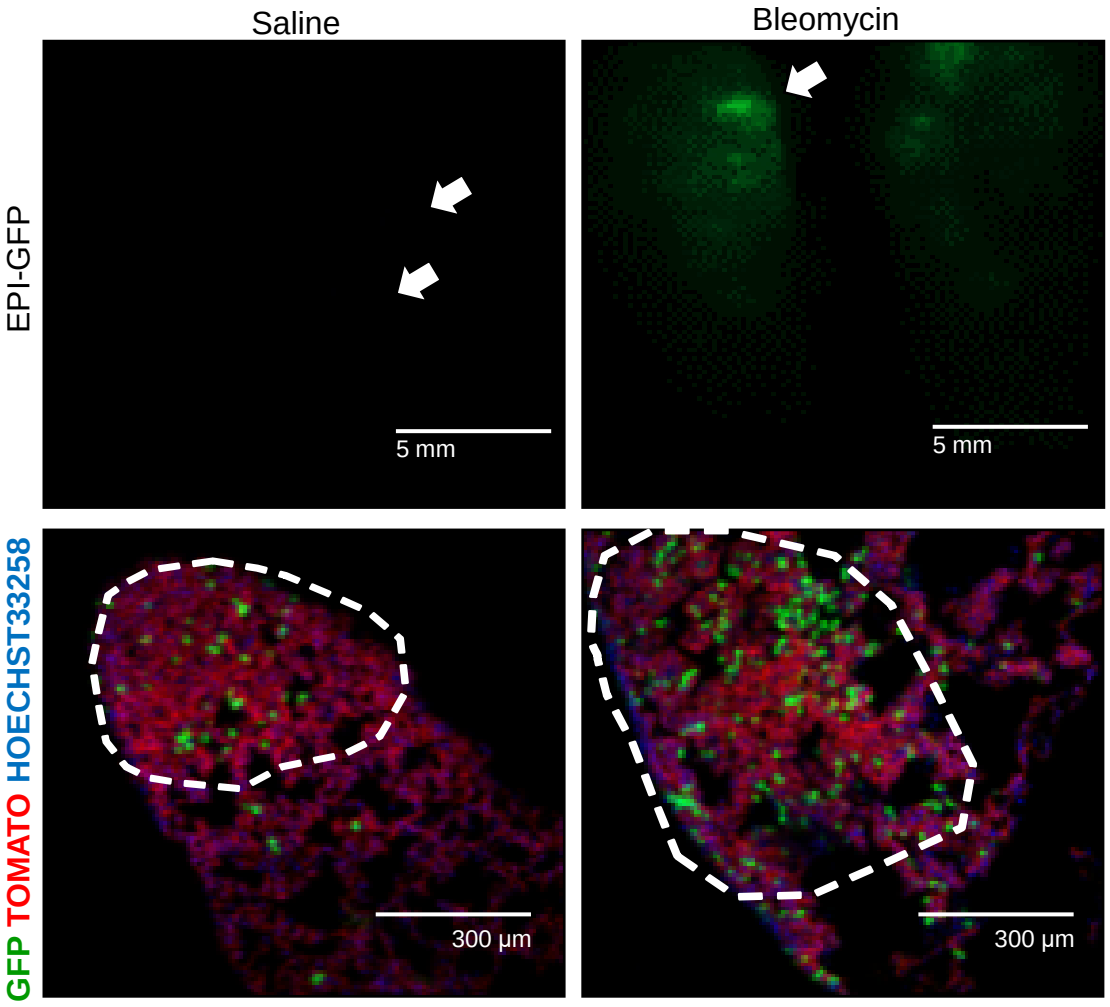


Figure 4 - Figure Supplement 1

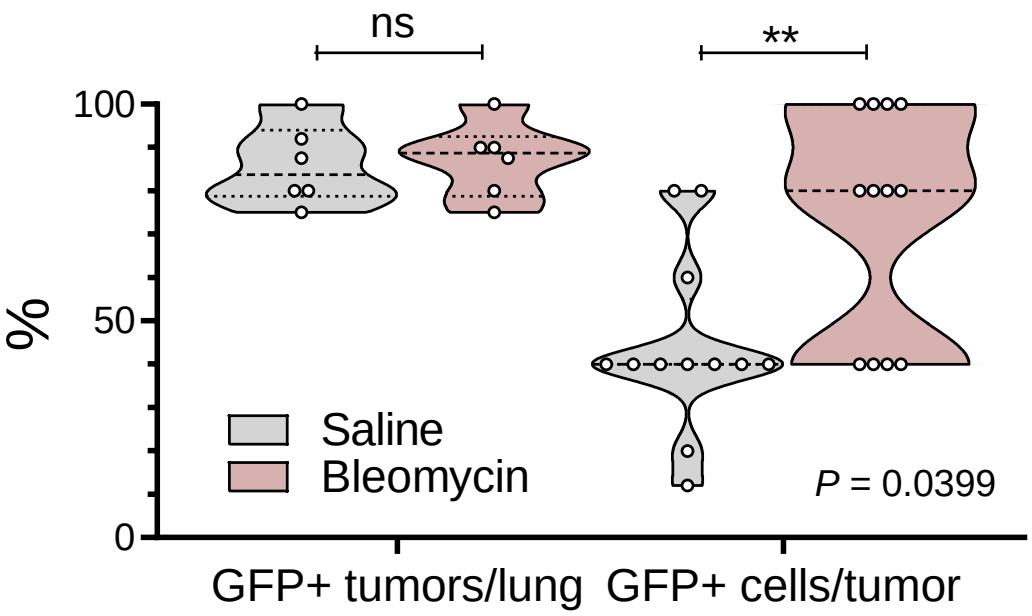


Figure 4 - Figure Supplement 2

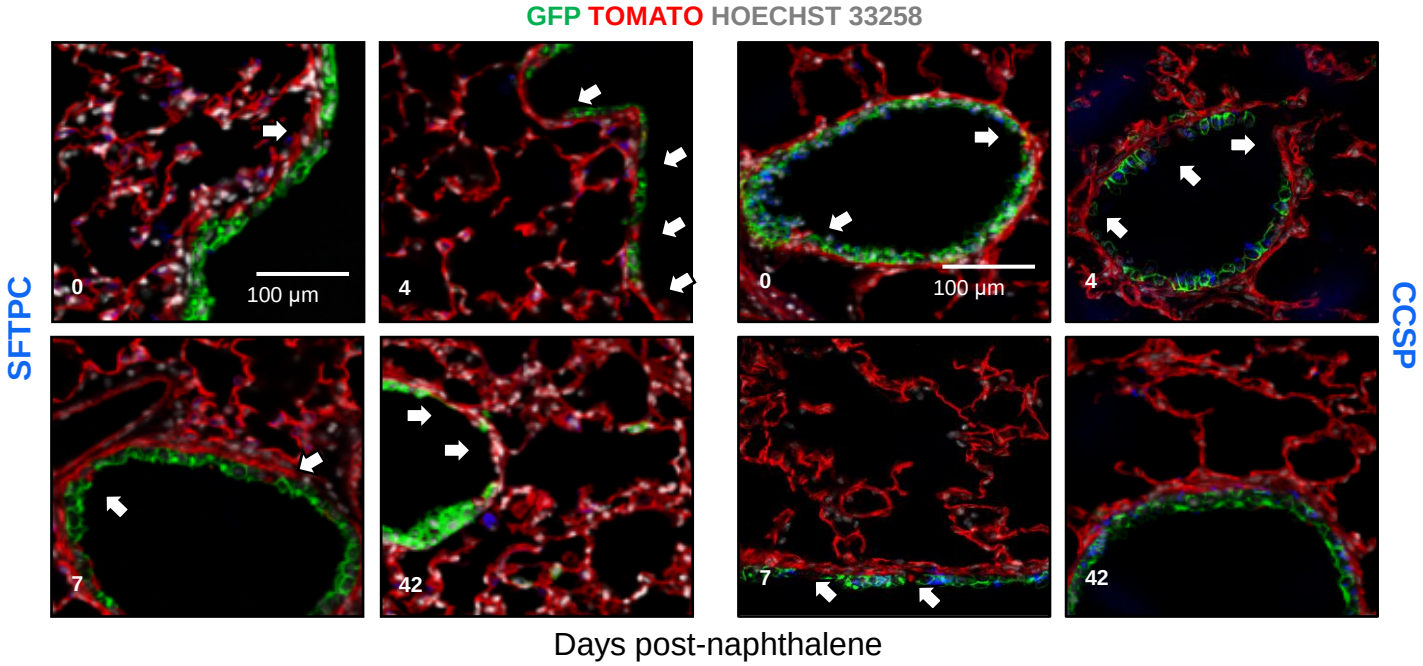


Figure 4 - Figure Supplement 3

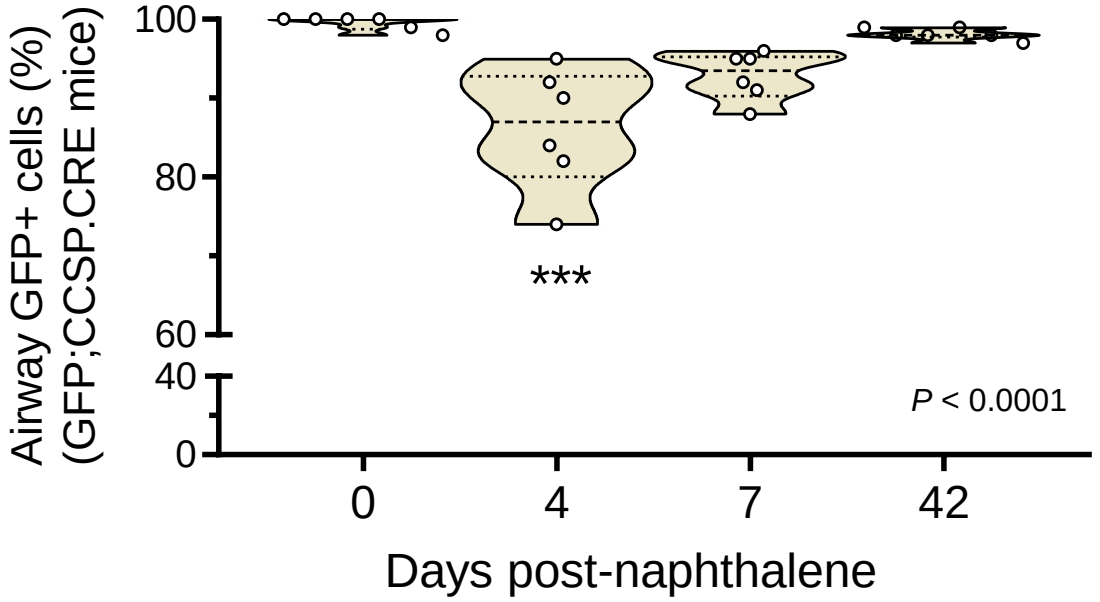
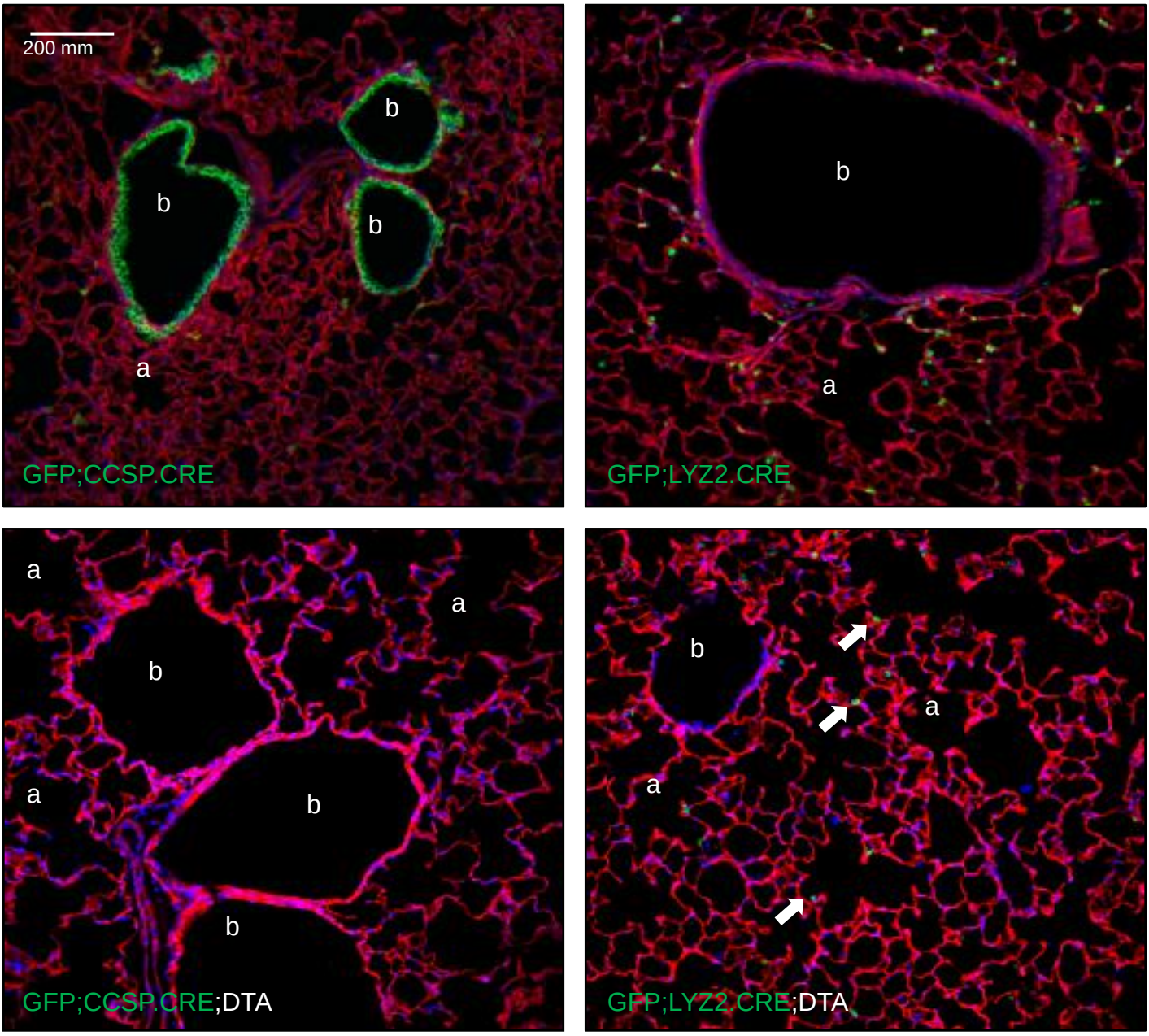


Figure 4 - Figure Supplement 4



HOECHST33258

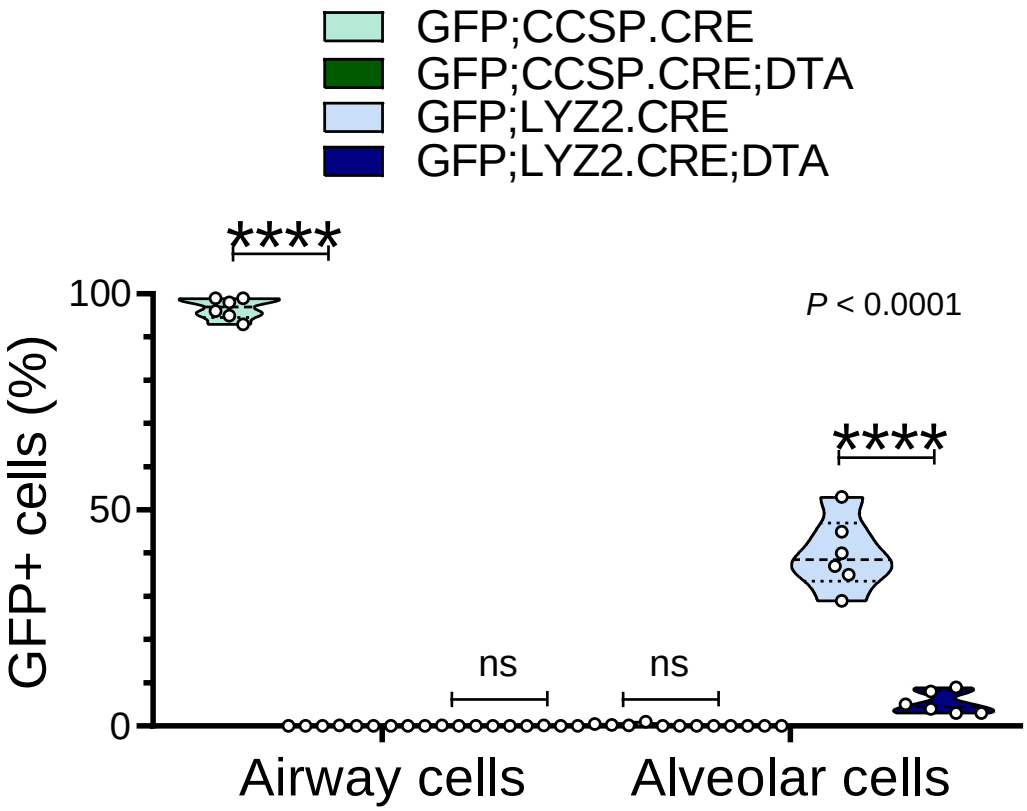


Figure 5 - Figure Supplement 2

		Signature genes	
		Lung	LUAD cells
Signature genes	AEC	23	25
	ATII	24	10
	BMDM	14	9

P = 0.1168

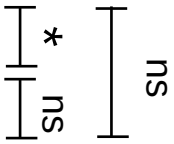


Figure 6 - Figure Supplement 1

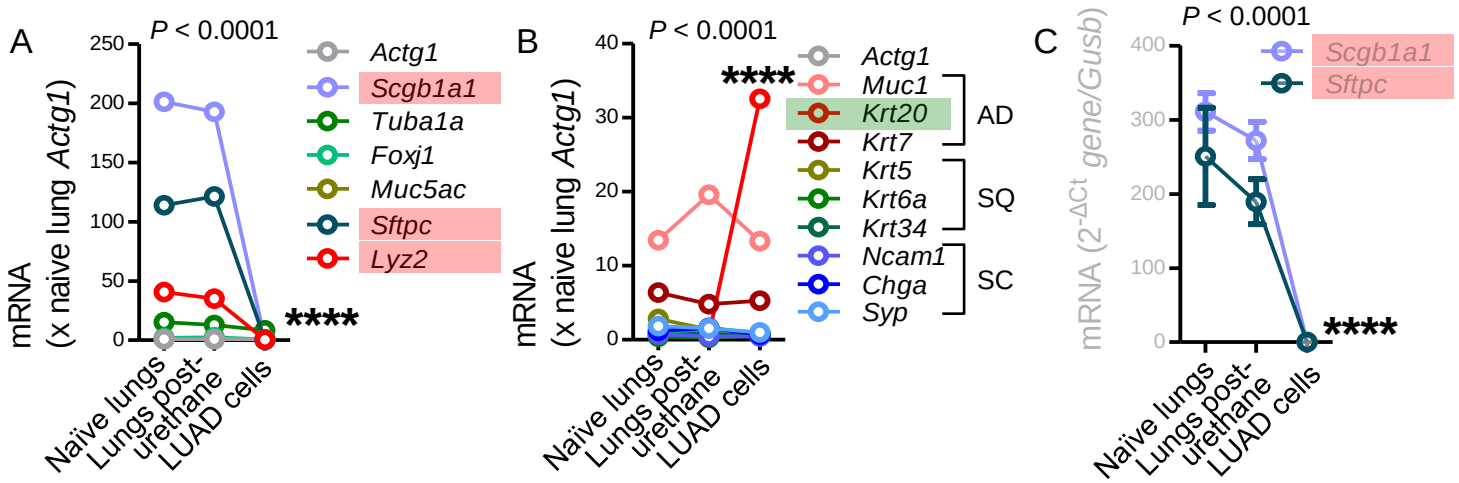


Figure 6 - Figure Supplement 2

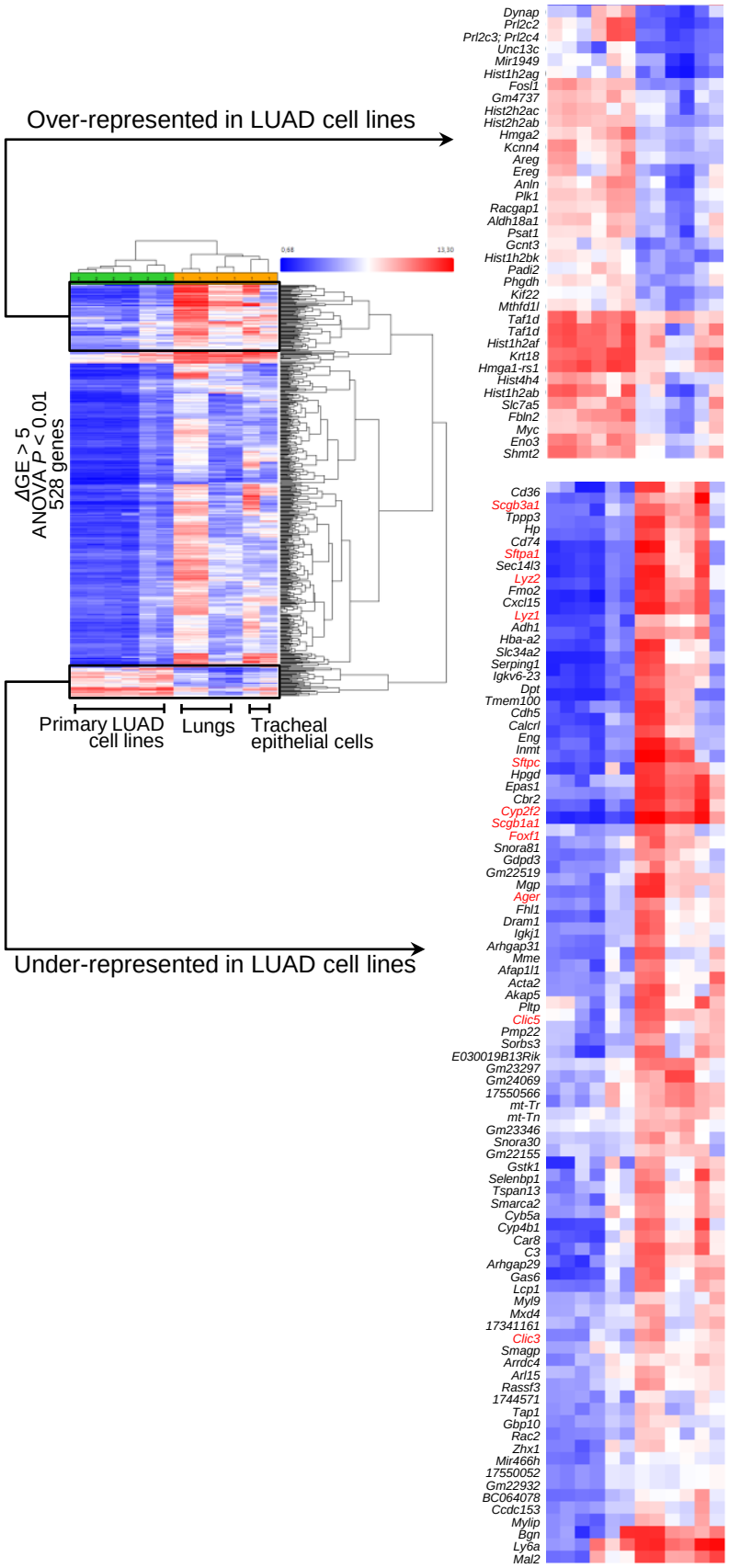


Figure 6 - Figure Supplement 3

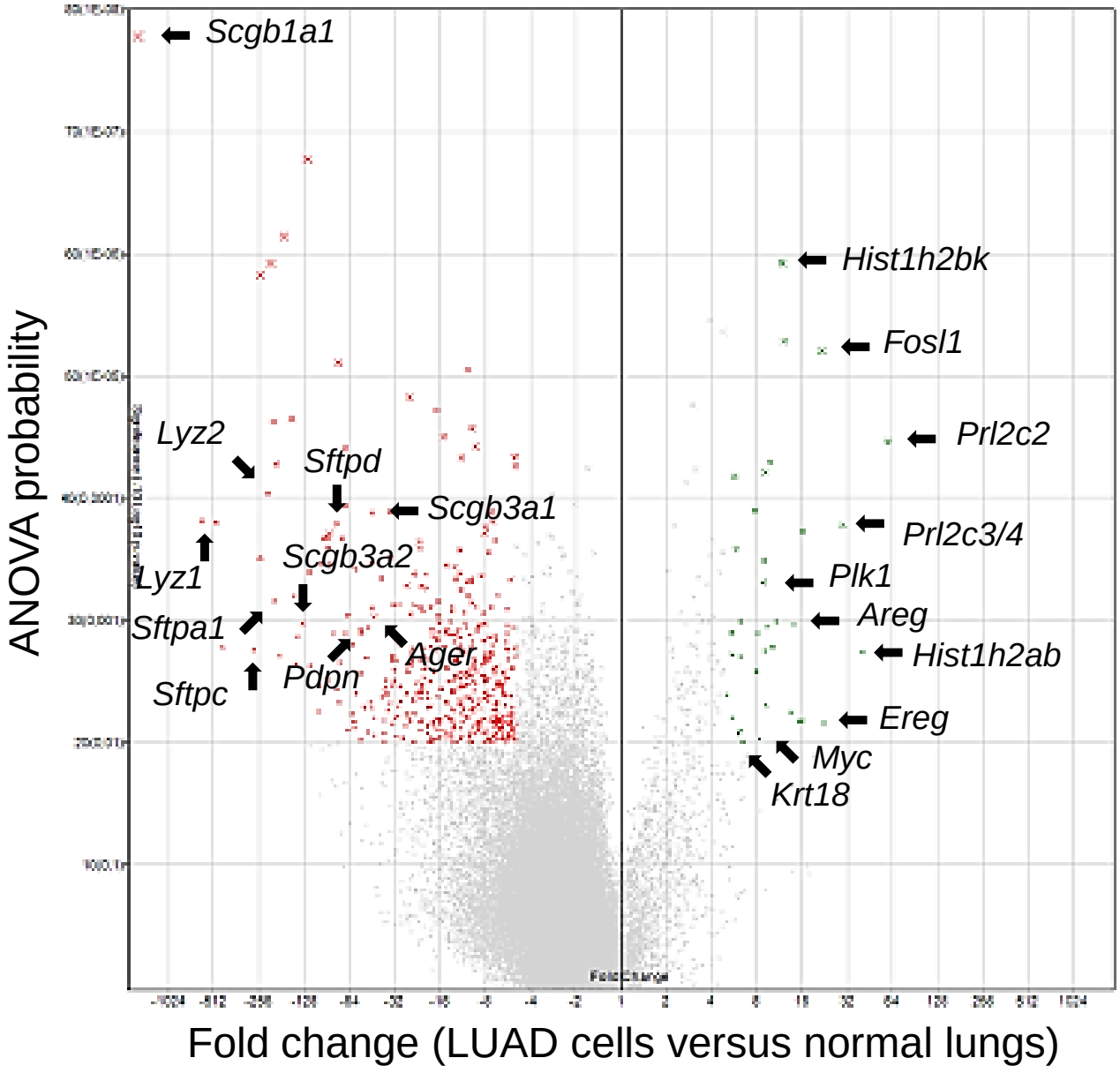


Figure 6 - Figure Supplement 4

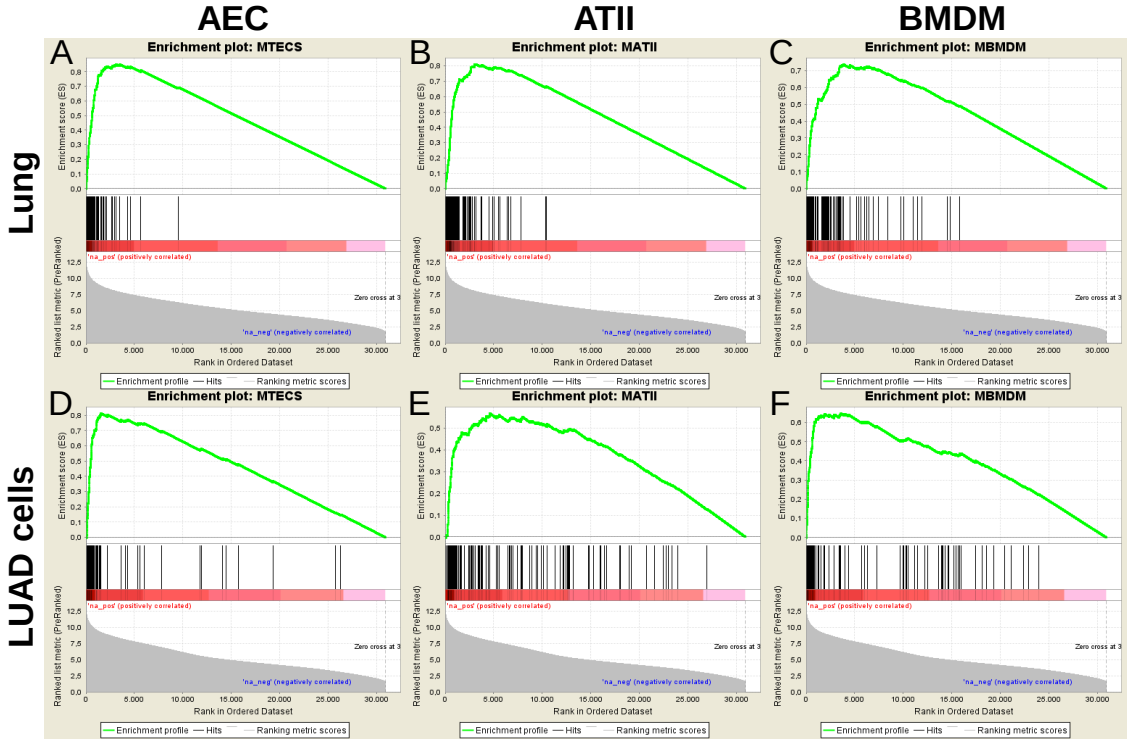


Figure 6 - Figure Supplement 5

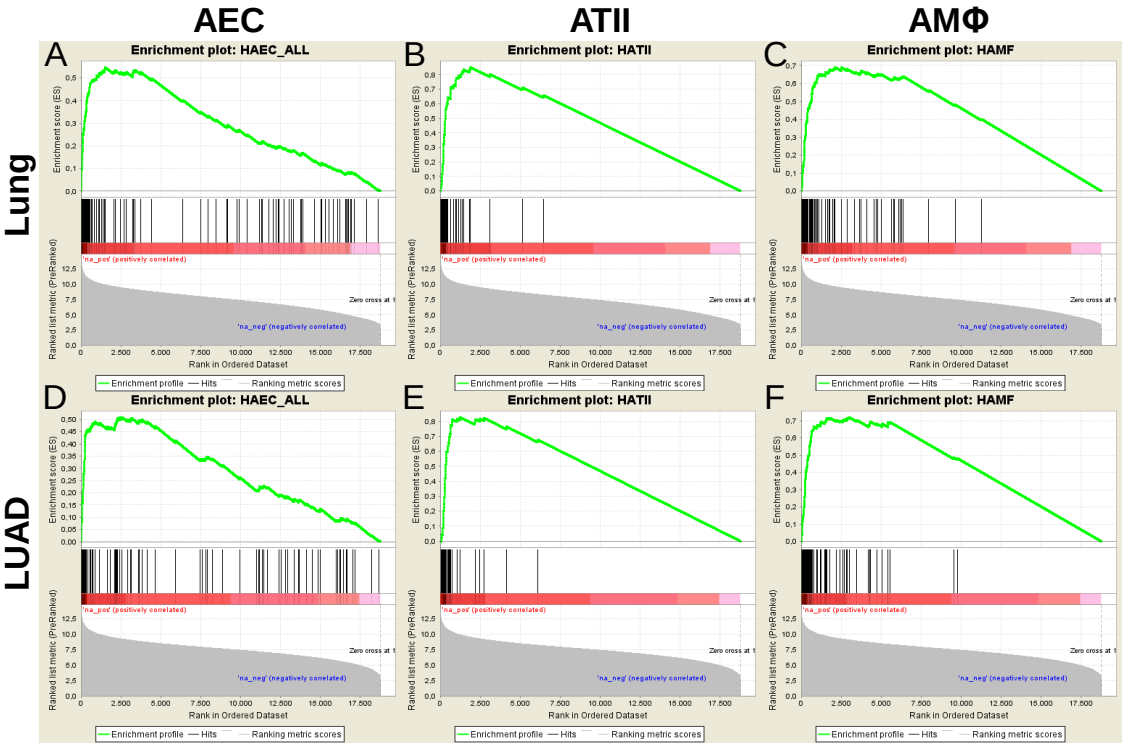


Figure 6 - Figure Supplement 6