

Institute of Lung Biology and disease (ILBD) Comprehensive Pneumology Center (CPC)  
Helmholtz Zentrum München



Dissertation  
zum Erwerb des Doctor of Philosophy (Ph.D.)  
an der Medizinischen Fakultät der  
Ludwig-Maximilians-Universität zu München

***Mapping carcinogen-specific patterns of mutation acquisition  
in the lungs***

vorgelegt von: Mario Pepe

.....  
aus: Nocera Inferiore, Salerno, Italy

.....  
Jahr: 2021

Mit Genehmigung der Medizinischen Fakultät der  
Ludwig-Maximilians-Universität zu München

**First evaluator:** *Prof. Dr. Silke Meiners*

**Second evaluator:** *Dr. Georgios Stathopoulos*

**Third evaluator:** Prof. Dr. Alexander Dietrich

**Fourth evaluator:** Prof. Dr. Jürgen Behr

**Dean:** **Prof. Dr. med. dent. Reinhard Hickel**

date of the defense:

\_\_\_\_\_19/07/2021\_\_\_\_\_

## Affidavit

Mario Pepe

Hereby I declare,

That the submitted thesis entitled "Mapping carcinogen-specific patterns of mutation acquisition in the lungs" is my own work. I have only used the sources indicated and have not made unauthorized use of services of a third party. Where the work of others has been quoted or reproduced, the source is always given.

I further declare that the submitted thesis or parts thereof have not been presented as part of an examination degree to any other university.

Erding, 08/08/2021

Place, date

Mario Pepe

Signature of doctoral candidate

## Abbreviations

1	LUAD	Lung adenocarcinoma
2	NGS	Next generation sequencing
3	qPCR	Quantitative real-time PCR
4	RNAseq	Next generation RNA sequencing
5	NSCLC	Non-small cell lung carcinoma
6	EC	Urethane (also called ethyl carbamate)
7	DEN	Diethyl nitrosamine (also called N,N-Diethyl nitrous amide)
8.	<i>TP53</i>	Tumor protein p53
9.	<i>MLL2</i>	Histone-Lysine N-methyltransferase
10.	<i>CDKN2A</i>	Cyclin Dependent Kinase Inhibitor 2A
11.	<i>KEAP1</i>	Kelch like ECH associated protein 1
12.	<i>PTEN</i>	Phosphatase and tensin homolog
13	<i>KRAS</i>	KRAS proto-oncogene, GTPase
14	<i>EGFR</i>	Epidermal growth factor receptor
15	<i>ALK</i>	ALK receptor tyrosine kinase
16	BASC	Broncho-alveolar stem cell
17	AT	Alveolar type cell
18	PNEC	Pulmonary neuroendocrine cell
19	<i>Kras</i>	Kirsten rat sarcoma viral oncogene homolog
20	<i>Trp53</i>	Transformation related protein 53
21	GSEA	Gene set enrichment analysis
22	CRE	P1 bacteriophage CRE recombinase
23	GFP	green fluorescent protein
24	<i>CCSP</i>	Club cell secretory protein

25	<i>LYZ2</i>	Lysozyme C-2
26	<i>SFTPC</i>	Surfactant Protein C
27	ddPCR	Digital droplet PCR
28	dNTP	Deoxynucleotide
29	ddPCR <sup>Adv</sup>	advanced ddPCR assay

## List of publications

### Publication I:

Nikolaos I. Kanellakis\*, Anastasios D. Giannou\*, **Mario A. A. Pepe\***, Theodora Agalioti, Dimitra E. Zazara, Ioanna Giopanou, Ioannis Psallidas, Magda Spella, Antonia Marazioti, Kristina A. M. Arendt, Anne Sophie Lamort, Spyridon Champeris Tsaniras, Stavros Taraviras, Helen Papadaki, Ioannis Lilis† and Georgios T. Stathopoulos†. “Tobacco chemical-induced mouse lung adenocarcinoma cell lines pin the prolactin orthologue proliferin as a lung tumour promoter”. *Carcinogenesis*. 2019 Nov 25;40(11):1352-1362.

\* equal first author

† senior author

### Publication II:

Magda Spella, Ioannis Lilis, **Mario A. A. Pepe**, Yuanyuan Chen, Maria Armaka, Anne Sophie Lamort, Dimitra E. Zazara, Fani Roumelioti, Malamati Vreka, Nikolaos I. Kanellakis, Darcy E. Wagner, Anastasios D. Giannou, Vasileios Armenis, Kristina A. M. Arendt, Laura V. Klotz, Dimitrios Toumpanakis, Vassiliki Karavana, Spyros G. Zakyntinos, Ioanna Giopanou, Antonia Marazioti, Vassilis Aidinis, Rocio Sotillo, Georgios T. Stathopoulos†. “Club cells form lung adenocarcinomas and maintain the alveoli of adult mice”. *Elife*. 2019 May 29;8. pii: e45571.

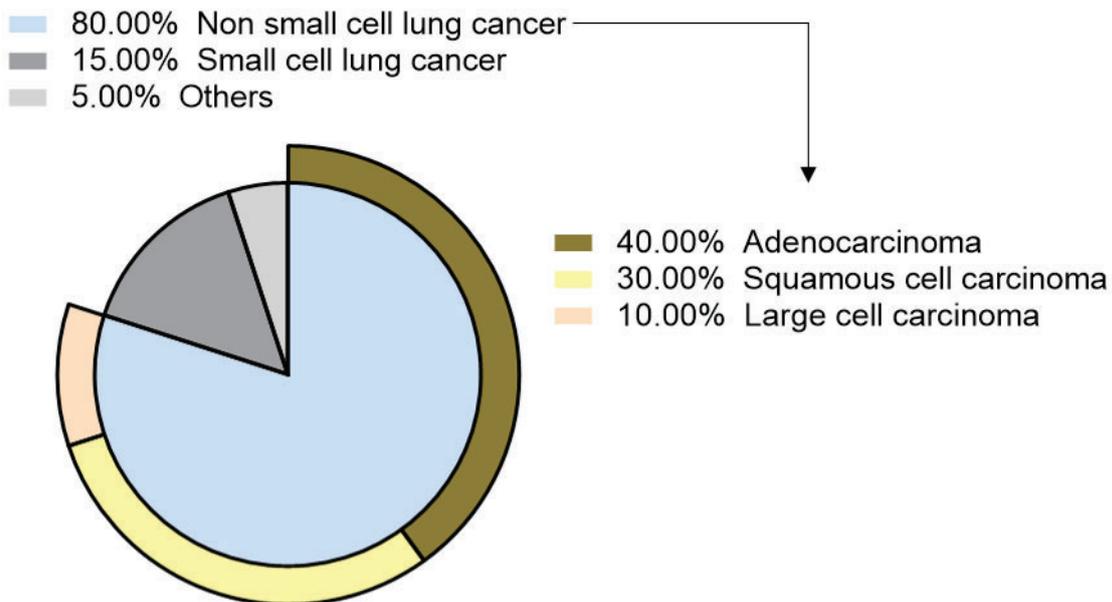
† senior author

## Table of Contents

Mapping carcinogen-specific patterns of mutation acquisition in the lungs .....	1
Affidavit .....	3
Abbreviations .....	4
List of publications .....	6
Table of Contents.....	7
Introduction .....	8
1.1. Tobacco's carcinogens mutational signature in LUAD .....	10
1.2. LUAD cell lineage of origin.....	11
1.3. Aim .....	13
2. Publication I: Tobacco chemical-induced mouse lung adenocarcinoma cell lines pin the prolactin orthologue proliferin as a lung tumour promoter.....	14
2.1. Summary .....	14
2.2. Contribution .....	15
2.3. Publication .....	17
3. Publication II: Club cells form lung adenocarcinomas and maintain the alveoli of adult mice .....	46
3.1. Summary .....	46
3.2. Contribution .....	47
3.3. Publication .....	48
4. Conclusion .....	134
5. References .....	136
Appendix A .....	139
Performing the ddPCR on a single genome copy .....	140
Identifying a ROI-REF couple with one single TaqMan® probe .....	141
Quantifying two different fragments emitting the same fluorescence .....	142
Spatial and longitudinal quantification of <i>Kras</i> <sup>Q61R</sup> mutation .....	143

## Introduction

Lung cancer presents the deadliest killer among malignant tumours both in men and women with over 1.7 million deaths per year. In 2030, the incidence of this dreadful disease will exceed 3.7 million, according to Globocan 2018 projections<sup>1</sup>. Lung cancer is classified in two major categories: small cell lung cancer (SCLC) which accounts for 15% of new cases, and non-small cell lung cancer (NSCLC) accounting for another 80% (Figure 1)<sup>1</sup>. In this last category of lung cancers, different subtypes are identified as follows: lung adenocarcinoma (LUAD), squamous cell carcinoma (SCC), and large cell lung carcinoma (LCLC)<sup>2, 3</sup>. These histological subtypes display different cellular appearances, arise in different locations in the lungs, are caused by different main risk factors, and display divergent mutation profiles<sup>2-7</sup>.



*Figure 1: Histologic classification of lung cancer. The data were collected from Bray et al. "Global cancer statistics". CA Cancer J Clin 2018. The plots were generated with Prism Graphpad 8<sup>th</sup>.*

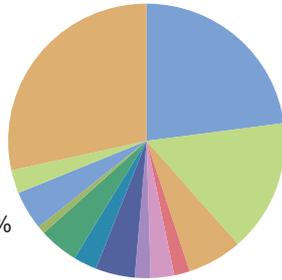
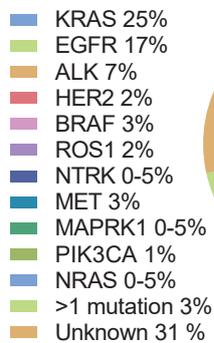
In the past, over 10% of NSCLC were assigned to the LCLC subtype. Currently, using a deeper classification of other histological subtypes, only 2% are diagnosed as LCLC<sup>4</sup>. The name of this category reflects the absence of a squamous or glandular cytology and the lack of differentiation of the

cells involved. This histological subtype can be found in almost every location in the lung<sup>2</sup>.

Previously SCC presented the most common subtype of NSCLC<sup>5</sup>. It generally localizes in the central part of the lungs, in proximity to the large airways<sup>2, 3</sup>. From the molecular point of view, SCC contains inactivating mutations in various tumor suppressor genes such as tumor protein p53 (*TP53*, mutated in 81% of cases), histone-Lysine N-Methyltransferase (*MLL2*, 20%), cyclin dependent kinase Inhibitor 2A (*CDKN2A* 15%), kelch like ECH associated protein 1 (*KEAP1* 12%), and phosphatase and tensin homolog (*PTEN* 8%) (Figure 2). In addition, recurrent gain-of-function mutations have been found in oncogenes such as phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha (*PIK3CA*, 16%) and neurofibromin 1 (*NF1*, 15%)<sup>4</sup>. The histology of SCC is characterized by flat and elongated cells that form multiple layers and feature partial keratinisation and follicle formation similar to the skin<sup>2, 3</sup>. The major risk factor is smoking (direct or second-hand smoke), while secondary (contributing) risk factors are thought to be asbestos, heredity, and radon<sup>3, 5</sup>.

Currently, LUAD is the most prominent histological subtype of lung cancer, accounting for over 40% of all cases of NSCLC<sup>1, 5</sup>. It tends to develop in the smaller airways, in proximity of the alveolar spaces<sup>3</sup>. The World Health Organization classifies LUAD in various histopathologic growth patterns (lepidic, acinar, papillary, micropapillary, solid, invasive mucinous, colloid, fetal, and enteric)<sup>3</sup>. LUAD histology is characterized by a glandular shape and presents features of mucus secreting cells. Major genomic alterations are present in the *KRAS* proto-oncogene GTPase (*KRAS*, 25-30%), epidermal growth factor receptor (*EGFR* 15-17%), and *ALK* receptor tyrosine kinase (*ALK*, 4-7%) driver oncogenes. Regarding the onco suppressor side, *TP53* (41%) is the single most mutated gene in LUAD, followed by *KEAP1*, (14%) (Figure 2)<sup>7, 8</sup>. LUAD risk factors are heredity factors<sup>9</sup>, as well as occupational exposure to agents such as silica, asbestos, radon, heavy metals, radiation, and diesel fumes<sup>2, 3</sup>. Nevertheless, the main etiologic agent remains tobacco smoke<sup>9-11</sup>.

## LUAD



## SCC

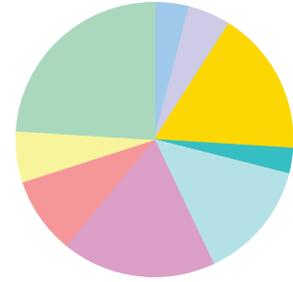
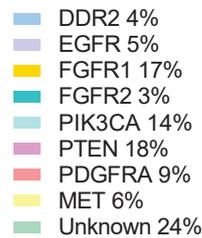


Figure 2: Driver genes mutations in LUAD and SCC. Data are collected from the cancer genome atlas (<https://portal.gdc.cancer.gov/>) and catalogue of somatic mutation in cancer (<https://cancer.sanger.ac.uk/cosmic>). The gene names are reported as HGNC nomenclature. <1 mutation= mutation in more than one driver gene. The plots were generated with Prism Graphpad 8<sup>th</sup>.

### 1.1. Tobacco's carcinogens mutational signature in LUAD

The advent of deep sequencing and transcriptome profiling opened new possibilities to further characterize and describe LUAD. In the past 10 years a plethora of sequencing studies populated the scientific literature, using as starting material the cancer patient specimens. According to the concept of carcinogen-induced mutational signatures<sup>7-10</sup>, every carcinogen tends to leave a specific imprint on normal human cells. This effect persists throughout the carcinogenic process and can be detected in fully malignant tumors. According to this, it is possible retrospectively decipher what a carcinogen caused, namely a carcinogen signature<sup>7, 8, 15</sup>. Therefore, the definition of such carcinogen signatures has primarily relied on correlation between clinically documented exposures (i.e. smoking) and sequencing results from various human cancers<sup>7-10, 13, 14</sup>. For the tobacco smoke, where the composition consists of a complex mixture of chemical compounds<sup>11</sup>, the signature derived it is the result of the addiction and combination of all carcinogens induced mutation present in this complex mixture. Indeed, the different specific carcinogen-induced alterations contribute with different patterns and weights at the whole smoking signature<sup>9, 16</sup>. Due to the mixture of carcinogens present in the tobacco smoke, and the accumulation of

several mutational events during the tumorigenesis process, it is not possible to dissect the contribution of every single carcinogen from a fully malignant smoking-associated tumor<sup>7-10</sup>. Furthermore, the effect of a single carcinogen is not only the genomic alteration that the carcinogen caused, but also the downstream repercussion that this genomic alteration will induce, and the feedbacks that would derive from this event<sup>17</sup>.

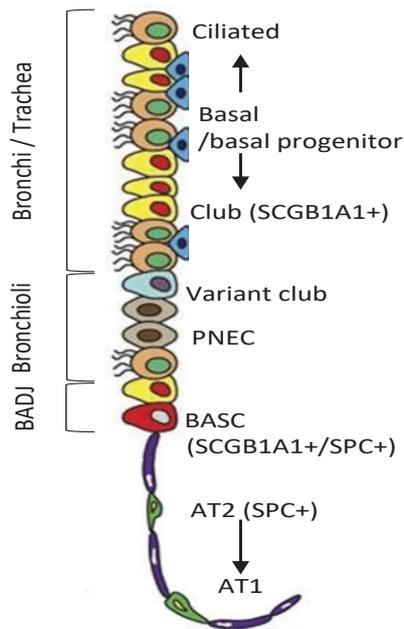
In consequence, a gap emerged between cutting-edge technology and the possibility of finely mapping the carcinogen-induced molecular alterations in the lungs. Indeed, for the tobacco smoke induced alterations, the necessity arose to develop different animal models of carcinogen induced LUAD to dissect the carcinogenic effect of the tobacco smoke in the different alteration processes related to the different carcinogens. In this matter, only one study (Westcott et al, Nature 2015) attempted to define a specific carcinogen-induced signature and it was limited to only two carcinogens that could mimic the mutational pattern observed in humans<sup>17</sup>. The remarkable results obtained in this study are hampered by the different stromal and immune-infiltrating cells contaminating the sequencing samples<sup>18, 19</sup>.

Therefore, it seems of essential importance to obtain the appropriate starting material for defining a signature induced by a single carcinogen. In this optic, deriving a pure cancer cell would be a good starting point. Despite that, only a carcinogen signature is not enough to precisely map the pattern of mutation acquisition in the lung. It remains uncertain in which pulmonary cell lineage the LUAD tumorigenesis initiates.

## 1.2. LUAD cell lineage of origin

Different cell lineages contribute to constitute the lungs airway epithelium. In the lower respiratory tract, we can recognize the respiratory epithelium and the respiratory zone<sup>21</sup>. The respiratory epithelium lines trachea, bronchi and bronchioles, and is composed from Goblet cells with the function of secreting mucus to protect the membrane. The ciliated cells mainly dedicate to the mucus clearance mechanism<sup>21</sup>. Among the ciliated cells we also find the basal cells, representing the stem cells, or the progenitors of the airway epithelium. At the level of terminal bronchioles, we also observe the club cells; able specifically to secrete different kinds of surfactants, and characterized by their glandular form<sup>22</sup>. In the respiratory zone, at the level

of the alveoli, we find alveolar macrophages and pneumocytes. In pneumocytes, we recognize alveolar type 1 (AT1) and alveolar type 2 (AT2) subtypes<sup>21</sup>. AT1 cells are devoted to the gas exchange, their shape is typically squamous with a thin wall and a flat shape. The AT2 cell are cuboidal and their function is related to the production of lung surfactant. Pulmonary neuroendocrine cell (PNEC) and variant club cell complete the lung histology<sup>22</sup> (Figure 3).



*Figure 3: Cells of the lungs: schematic representation of different cell lineages composing the respiratory epithelium. In brackets are reported the different specific marker per cell lineage. Broncho-alveolar stem cell (BASC); Broncho-alveolar ductal junct. The figure was readapted from Sutherland KD., Berns A. "Cell of origin of lung cancer". Molecular Oncology 4, 397-403 (2010).*

In patients, LUAD stains positive for markers of AT2 cell and club cells with antibodies detecting surfactant protein C (SPC) and secretoglobin family 1A member 1 (SCGB1A1)<sup>23</sup>. These findings gave rise to the hypothesis that AT2 cells were the LUAD cells of origin. To elaborate on this hypothesis, most scientists deployed different transgenic mouse models mainly based on the point mutations in KRAS codon 12/13<sup>24-26</sup>. These mutations, that represent the most common in KRAS driven LUAD in human<sup>27</sup>, lead to the oncogenic form of KRAS. With these animal models, using different CRE drivers, it is possible to induce the expression of the oncogenic KRAS in different cell lineages<sup>24-26, 28</sup>. Having as target the AT2, the first studies focused their attention on this cell lineage and Broncho-alveolar stem cells (BASC) that stained positive to AT2 and club cell markers<sup>25</sup>. Another study had as investigation target the club cells using the same approach to try to resolve the implication of this lineage in the LUAD tumorigenesis<sup>28</sup>. Both studies concluded that KRAS mutation in different cell lineages can start the

tumorigenesis process , but full malignancies arose exclusively from AT2 cells and other cell lineages gave rise only to hyperplasia<sup>25, 28</sup>. An uncharted matter in this approach is to detect the smoked induced LUAD cell of origin. This cannot be achieved using a genetic model, where all the series of events related to smoking and tumorigenesis are not recapitulated and neither possibly recaptured<sup>7, 8, 12, 17</sup>.

In summary, the state of the art in smoking-induced LUAD consists of an unknown initiating cell lineage, a mutational profile accounting as key mutation *KRAS* and *TP53* genes, and a location in the lung at the level of the small airways in proximity of the alveolar space. In this context our group established a specific and pivotal way to induce LUAD in mice<sup>30</sup>. We also used a combination of genetically induced mouse models, and/or different triggering carcinogens to recapitulate early event of thoracic malignancy<sup>30, 31</sup>. On behalf of that, we are in the position of combining different experimental approaches to tackle the identification of the smoked induced LUAD cells of origin and define specific carcinogen induced signatures. In this broad topic, I focused on a method to map early events of alterations in the lung, after the carcinogen insult, and profiled the transcriptional pattern of pure cancer cell lines derived from different carcinogens.

### 1.3. Aim

The aim of this dissertation is to combine different approaches for mapping alterations induced by tobacco smoke carcinogens during LUAD tumorigenesis. In our research group, we achieved to induce LUAD with different carcinogens in susceptible mouse strains. Using those different models of carcinogen-induced LUAD, this thesis aims to:

- a) Map the gene expression profiles of carcinogen induced LUAD cell lines established using injection of two different tobacco carcinogens into two differentially susceptible mouse strains.
- b) Identify the cellular origin of LUAD using mouse models of respiratory epithelial cell fate.

## 2. Publication I: Tobacco chemical-induced mouse lung adenocarcinoma cell lines pin the prolactin orthologue proliferin as a lung tumour promoter

### 2.1. Summary

LUAD represents the most common histological subtype of and the most frequent cause of death from non-small cell lung carcinoma (NSCLC), accounting for over 600,000 deaths per year worldwide<sup>1</sup>. As is also true for other histological subtypes of NSCLC, smoking represents the main cause of LUAD. The molecular profile of this disease is completely different in smokers compared with never smokers, carrying over ten times the amount of mutations as well as a unique transcriptome<sup>7, 13, 33</sup>. The need for faithful models of smoking-induced LUAD that fully recapitulate the biological behavior and the molecular profile of LUAD is unmet.

Hence, mouse models of tobacco carcinogen-triggered LUAD were selected for our purposes, and we induced LUAD in two different strains of inbred mice: FVB and Balb/c, using two different carcinogens of tobacco smoke: urethane and diethyl nitrosamine, applied according to established protocols. All mice developed large tumors that, after histological confirmation, were used to establish truly malignant carcinogen-induced LUAD cell lines ( $n = 7$ ). Independent from which carcinogen was used, all cell lines were harboring mutations in *Kras* and *Trp53* genes. They are mouse homologs of two of the most relevant altered human genes in smokers' LUAD (*KRAS* and *TP53*)<sup>7</sup>. All cell lines were immortals, phenotypically stables and indefinitely proliferatives *in vitro*. Those cell lines were able to cause, through subcutaneous, intravenous, and intrapleural injection, primary solid tumors at the injection site and lung metastases in syngeneic mice, which were uniformly lethal.

Gene expression profiling of our carcinogen induced LUAD cell lines, in comparison with naive mouse lungs and various other cell types (alveolar epithelial cells, immune cells, and other cancer cell lines), revealed a unique gene expression signature. This transcriptomic fingerprint featured statistically significant alterations in key pathways for cancer development, such as DNA replication, purine metabolism, and mitotic checkpoint regulation. Interestingly, the carcinogen-induced LUAD cell line signature

sufficed to precisely dichotomize normal from cancer samples from the Biomarker-integrated Approaches of Targeted Therapy for Lung Cancer Elimination (BATTLE) study<sup>35</sup>. Moreover, gene-set enrichment analysis disclosed a strong positive enrichment of the LUAD cell line transcriptomic signature among smokers' patients and a significant negative enrichment among never smokers' from the BATTLE dataset.

In conclusion, our carcinogen-induced cell lines are able to closely mimic smokers' LUAD in terms of biological behavior and molecular profile, representing a new tool to study the pathobiology of this disease.

## 2.2. Contribution

My personal contribution to this publication consisted of defining the transcriptome profile of the different carcinogen-induced LUAD cell lines, revealing the unique transcriptomic signature of those cell lines and cross-examining those results with a publicly available dataset of human LUAD transcriptomes. For this purpose, we performed gene expression profiling using Affymetrix Mouse Gene 2.0 ST Arrays and analyzed the samples listed below using dedicated software (Transcriptome analysis console 4, TAC4; Thermo Fisher Scientific, Waltham, MA):

- Tobacco carcinogen-induced LUAD cell lines ( $n = 6$ )
- Lung from naive syngeneic mice ( $n = 4$ )
- Mouse type II alveolar epithelial cells ( $n = 5$ )
- Mouse airway epithelial cells ( $n = 4$ )
- Mouse mast cells ( $n = 4$ )
- Mouse macrophages ( $n = 4$ )
- Other murine cancer cell lines ( $n = 4$ )

The results of the above analyses are visible in figure 4A of the publication.

After gene expression profiling, I performed pathway and gene-set enrichment analyses (GSEA) (figures 4E and 4F). Pathway analysis was performed using the WikiPathway function of TAC4. GSEA was performed using GSEA4 (Broad institute; available at <https://www.gsea-msigdb.org/gsea/index.jsp>). To cross-validate the carcinogen-induced LUAD cell line signature in human LUAD transcriptomes, I derived from our mouse

gene list the relative human orthologs using OrthoDB, and other relevant webtools. This “humanized” gene list was then used for two different purposes:

- As a metric of distance to compute hierarchical clustering between human LUAD transcriptomes and matched normal lung tissues.
- To build a custom gene-set and analyze its enrichment in transcriptome profiles of smokers’ and never smokers’ LUAD.

The results are reported in the publication in figures 6E and 6F

## 2.3. Publication

## ORIGINAL ARTICLE

# Tobacco chemical-induced mouse lung adenocarcinoma cell lines pin the prolactin orthologue proliferin as a lung tumour promoter

Nikolaos I. Kanellakis<sup>1,†,\*</sup>, Anastasios D. Giannou<sup>1,†</sup>, Mario A.A. Pepe<sup>2,†</sup>, Theodora Agalioti<sup>1</sup>, Dimitra E. Zazara<sup>1</sup>, Ioanna Giopanou<sup>1</sup>, Ioannis Psallidas<sup>1</sup>, Magda Spella<sup>1</sup>, Antonia Marazioti<sup>1</sup>, Kristina A.M. Arendt<sup>2</sup>, Anne Sophie Lamort<sup>2</sup>, Spyridon Champeris Tsaniras<sup>3</sup>, Stavros Taraviras<sup>3</sup>, Helen Papadaki<sup>4</sup>, Ioannis Lilis<sup>1,†</sup> and Georgios T. Stathopoulos<sup>1,2,\*†</sup>

<sup>1</sup>Laboratory for Molecular Respiratory Carcinogenesis, Department of Physiology, Faculty of Medicine, University of Patras, Rio, Achaia 26504, Greece, <sup>2</sup>Lung Carcinogenesis Group, Comprehensive Pneumology Center (CPC) and Institute for Lung Biology and Disease (iLBD), Ludwig-Maximilian University and Helmholtz Center Munich, Member of the German Center for Lung Research (DZL), Munich, Bavaria 81377, Germany, <sup>3</sup>Stem Cell Biology Laboratory, Department of Physiology, Faculty of Medicine and <sup>4</sup>Department of Anatomy, Faculty of Medicine, University of Patras, Rio, Achaia 26504, Greece

\*To whom correspondence should be addressed. Laboratory for Molecular Respiratory Carcinogenesis, Department of Physiology, Faculty of Medicine, University of Patras, Biomedical Sciences Research Building, 2nd floor, Room B40, 1 Asklepiou Str., University Campus, 26504 Rio, Greece. Tel: +30 2610 969154; Fax: +30 2610 969176 Email: [gstathop@upatras.gr](mailto:gstathop@upatras.gr)

<sup>†</sup>These authors contributed equally to this work.

## Abstract

Lung adenocarcinoma (LADC) is the leading cause of cancer death worldwide. Nevertheless, syngeneic mouse models of the disease are sparse, and cell lines suitable for transplantable and immunocompetent mouse models of LADC remain unmet needs. We established multiple mouse LADC cell lines by repeatedly exposing two mouse strains (*FVB*, *Balb/c*) to the tobacco carcinogens urethane or diethylnitrosamine and by culturing out the resulting lung tumours for prolonged periods of time. Characterization of the resulting cell lines ( $n = 7$ ) showed that they were immortal and phenotypically stable *in vitro*, and oncogenic, metastatic and lethal *in vivo*. The primary tumours that gave rise to the cell lines, as well as secondary tumours generated by transplantation of the cell lines, displayed typical LADC features, such as glandular architecture and mucin and thyroid transcription factor 1 expression. Moreover, these cells exhibited marked molecular similarity with human smokers' LADC, including carcinogen-specific *Kras* point mutations (*Kras*<sup>Q61R</sup> in urethane- and *Kras*<sup>Q61H</sup> in diethylnitrosamine-triggered cell lines) and *Trp53* deletions and displayed stemness features. Interestingly, all cell lines overexpressed proliferin, a murine prolactin orthologue, which functioned as a lung tumour promoter. Furthermore, prolactin was overexpressed and portended poor prognosis in human LADC. In conclusion, we report the first LADC cell lines derived from mice exposed to tobacco carcinogens. These cells closely resemble human LADC and provide a valuable tool for the functional investigation of the pathobiology of the disease.

## Introduction

Lung cancer is the leading cause of cancer death worldwide accounting for 1.6 million deaths in 2012, including 270 000 in

the European Union and 170 000 in the USA, with lung adenocarcinoma (LADC) accounting for half of the cases (1,2). Lung

Received: August 16, 2018; Revised: February 13, 2019; Accepted: February 27, 2019

© The Author(s) 2019. Published by Oxford University Press. All rights reserved. For Permissions, please email: [journals.permissions@oup.com](mailto:journals.permissions@oup.com).

**Abbreviation**

EC	ethyl carbamate
LADC	lung adenocarcinoma
PRL	proliferin, prolactin
RT-PCR	reverse transcription PCR

cancer is mainly caused by chemical carcinogens of tobacco smoke (3–5). Smoking-induced carcinomas including LADC bear thousands of mutations per genome, including gain-of-function point substitutions in critical codons of the KRAS proto-oncogene and deletion or loss-of-function point substitutions of the tumour suppressor TP53, encountered respectively in 20–40% and 80–90% of LADC (6–9). There is an unmet need for relevant mouse models of smoking-induced carcinomas such as LADC, which cannot be fully recapitulated by genetic models (10). Such vehicles would aid in the identification of new lung cancer genes, in the distinction of true cancer drivers from passenger events and in the development of new therapies.

To more comprehensively understand lung cancer initiation, evolution and signalling, faithful mouse models of the disease are invaluable (11). Although elaborate genetic mouse models of lung cancer are available, they do not fully recapitulate smoke-induced carcinogenesis, because they are based on a single or a few transgenes that are turned on artificially (10,12,13) and tumours often regress after transgenes are turned-off (14). Importantly, evidence suggests that mouse models of tobacco carcinogen-induced LADC are closely related to the human disease (10). In specific, urethane promotes LADC development through induction of KRAS and other oncogene mutations that are also found in human LADC (10,15). However, existing tobacco carcinogen-triggered mouse lung tumour models are not really thought to be malignant, despite that these lung tumours feature high similarities to human lung cancer (16,17). Therefore, truly malignant and transplantable mouse tobacco carcinogen-derived lung cancer cell lines do not exist. Such cells are invaluable as they could be used in immunocompetent mice to faithfully recapitulate human lung cancer on a background of full tumour–host interactions (13).

Proliferins (PRLs), also known as mitogen regulated proteins, are four murine glycoprotein orthologues of human prolactin (18–22). PRLs are highly expressed in the murine placenta during embryogenesis, as well as in highly proliferative adult mouse tissues such as skin hair follicles and small intestinal crypts, where they function to drive cellular proliferation, angiogenesis and wound healing (18,23,24). PRL expression levels have been correlated with fibrosarcoma progression in mice (24); however, their role in LADC remains unknown.

Here, we report the establishment of a battery of murine cell lines derived from LADC of two inbred mouse strains following exposure to two different tobacco carcinogens, and we show that they are malignant. These cell lines, faithful mouse models of human smoking-induced LADC, unveiled an unexpected LADC-promoting role for PRL in mice and for its orthologue prolactin in humans.

**Materials and methods****Ethics approval**

Experiments were carefully designed and approved a priori by the Veterinary Administration of the Prefecture of Western Greece (approval protocol 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/LexUriServ/LexUriServ.do?uri=OJ:L:2010:276:0033:0079:EN:PDF>).

**Murine cell lines used and authentication method**

The murine cancer cell lines used were Lewis lung carcinoma, B16F10 skin melanoma and PANO2 pancreatic adenocarcinoma (all from the National Cancer Institute Tumour Depository, Frederick, MD), as well as MC38 colon adenocarcinoma cells (obtained from Dr. Timothy Blackwell, Vanderbilt University, Nashville, TN) and AE17 pleural mesothelioma cells (obtained from Dr. Timothy Blackwell, Vanderbilt University). All cell lines have been described previously in detail (25). NIH 3T3 cells were from the American Type Culture Collection (Manassas, VA). Cells were cultured at 37°C in 5% CO<sub>2</sub>-95% air using Dulbecco's modified Eagle's medium 10% foetal bovine serum, 2 mM L-glutamine, 1 mM pyruvate, 100 U/ml penicillin and 100 mg/ml streptomycin and were tested biannually for identity by short tandem repeats and *Mycoplasma* spp. by PCR.

**Derivation of mouse LADC cell lines**

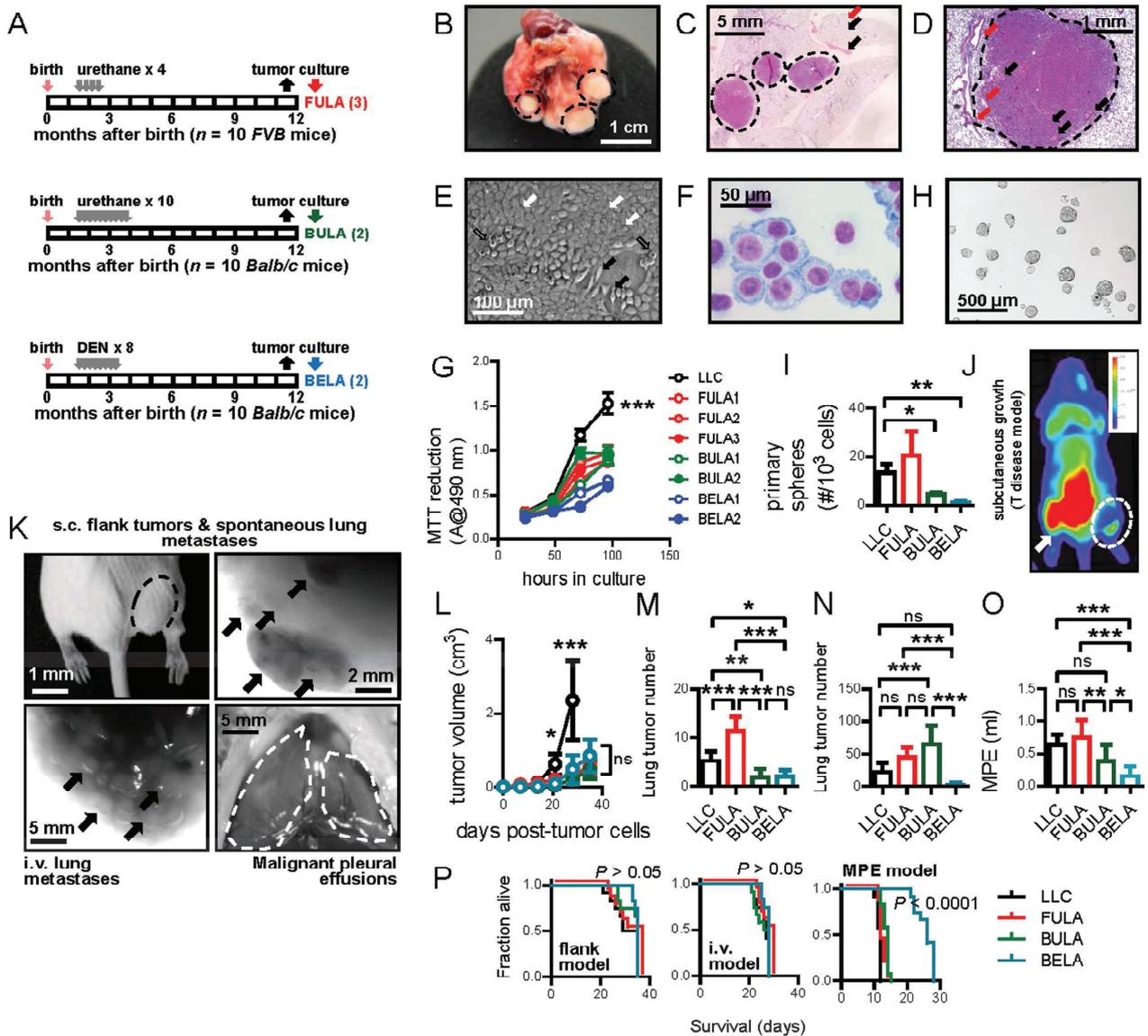
Ten months after first carcinogen [urethane, ethyl carbamate (EC), CH<sub>3</sub>CH<sub>2</sub>OCNH<sub>2</sub>, CAS #51-79-6; diethylnitrosamine, N,N-diethylnitrosamine, DEN, C<sub>4</sub>H<sub>10</sub>N<sub>2</sub>O, CAS # 55-18-5] exposure, mice were killed, lung tumours were dissected from surrounding healthy lung parenchyma under sterile conditions, were halved, one-half was processed for histology, and the other half was chopped into 1 mm pieces and seeded to cell culture dishes. Cells were cultured under standard conditions outlined in the [Supplementary data](#), available at [Carcinogenesis Online](#). When adenocarcinoma was diagnosed for a given tumour, its corresponding culture was passaged *in vitro* over a period of 18 months and 60 passages, whichever occurred first. All mouse LADC cell lines were deposited at the Laboratory for Molecular Respiratory Carcinogenesis cell line facility (<http://www.lmrc.upatras.gr>) and are available on request (lmrc@upatras.gr).

**Availability of data**

Microarray data are publicly available at Gene Expression Omnibus (GEO) DataSets (<https://www.ncbi.nlm.nih.gov/gds/>) using accession IDs GSE94981 (LADC, lungs, airway epithelial cells, mast cells and macrophages), GSE82154 (alveolar epithelial type II cells cells), GSE58188 (other cancer cells including Lewis lung carcinoma, MC38, AE17, B16F10 and PANO2 cells) and GSE43458 (BATTLE trial).

**Results****Novel mouse LADC cell lines generated by exposure of inbred mouse strains to tobacco carcinogens**

To develop murine LADC cell lines, we repeatedly exposed FVB and *Balb/c* mice to the tobacco carcinogens urethane (EC) and diethylnitrosamine [N-nitrosodiethylamine (DEN)]. For this, mice received repetitive intraperitoneal EC (1 g/kg) or DEN (200 mg/kg) injections and were observed for prolonged periods of time for true LADC to develop (Figure 1A). Indeed, mice developed large tumours that were harvested under sterile conditions and bisected, one-half of the tumours were always used for histological examination and the other half minced for long-term culture under standard conditions (Figure 1B). Histology revealed that some tumours were LADC showing abundant mitoses, invasion of adjacent lung structures and necrosis (Figures 1C and D). Cells from these tumours were cultured for a period of over 18 months and/or 60 passages, such that only truly malignant cells survived. This simple method has been shown not to introduce new artificial mutations that are not present pre-culture (26). The resulting LADC cell lines ( $n = 7$ ) were named XYLA# with X signifying the mouse strain (F, FVB; B, *Balb/c*), Y the carcinogen used (U, EC; E, DEN), LA lung adenocarcinoma, and # their serial number by derivation date. All cell lines were immortal, phenotypically stable and indefinitely proliferative *in vitro* where they displayed spindle shapes and anoikis (Figures 1E and F). In addition, all cell lines exhibited nuclear atypia and stemness as they were able to form tumour-spheres *in vitro*, a capacity unique to stem and cancer



cells (Figures 1G–I). Remarkably, on subcutaneous delivery of 1 million cells/mouse to syngeneic mice, LADC cell lines were able to form primary solid tumours at the injection site, as well as spontaneous pulmonary metastases (Figures 1J–L). Intravenous delivery of 250 000 LADC cells to syngeneic mice caused lung metastases as well, and intrapleural injection of 150 000 LADC cells to syngeneic mice triggered malignant pleural effusions (Figures 1M–O). All LADC cell lines were uniformly lethal regardless of injection route, confirming their malignant nature (Figure 1P). The primary tumours that gave rise to the cell lines, as well as secondary tumours generated by transplantation of the cell lines, all displayed typical LADC features, such as glandular architecture and mucin and thyroid transcription factor 1 expression (Figure 2). These results firmly support that the cell lines derived from lung tumours of tobacco carcinogen-treated mice are true LADC cells that can recapitulate the metastatic patterns of human LADC.

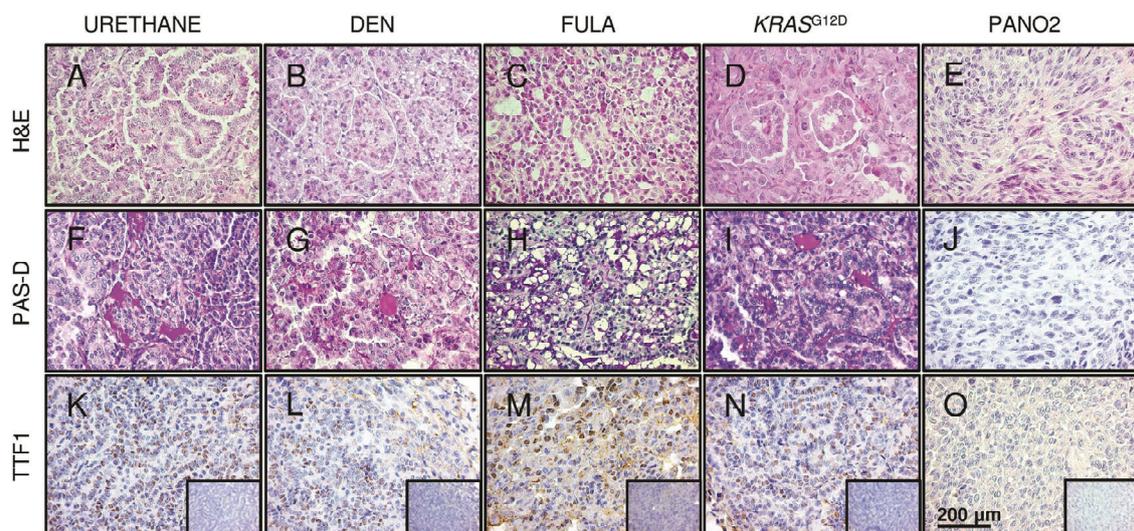
### Tobacco chemical-induced mouse LADC cell lines harbour *Kras* mutations and *Trp53* loss

Activating *KRAS* mutations and loss or mutation of *TRP53* are common in human LADC of smokers (7,27,28). We hence sought to determine whether our LADC cell lines were similar to human LADC in terms of *Kras* and *Trp53* status (Figure 3). Reverse transcription PCR (RT-PCR) followed by direct complementary DNA sequencing of *Kras* (target) and *Nras* (control) transcripts revealed the presence of heterozygous *Kras*<sup>Q61R</sup> mutations in all EC-induced cell lines and heterozygous *Kras*<sup>Q61H</sup> mutations in all DEN-induced cell lines, but no *Nras* mutations (Figures 3A and B; Supplementary Figure 1, available at *Carcinogenesis* Online). Interestingly, all *Kras*<sup>Q61R</sup>-mutant cell lines generated using EC also expressed a nonsense-mediated decay transcript, in addition to the mutant and wild-type (<sup>WT</sup>) transcripts (Figure 3A; Supplementary Figure 1, available at *Carcinogenesis* Online), a mechanism thought to prevent the expression of mutant proteins (29). To assess the *Trp53* status

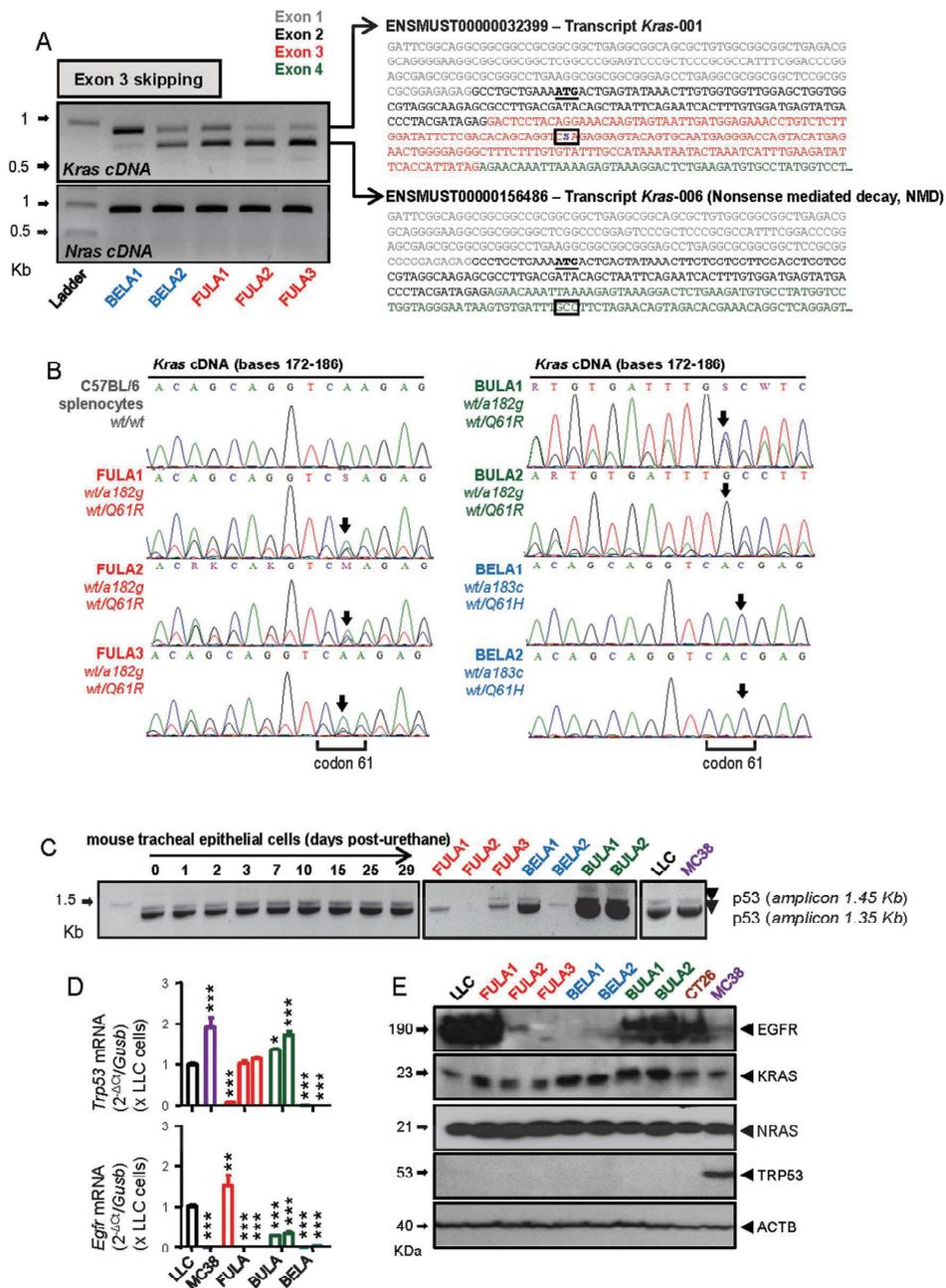
of our carcinogen-induced LADC cell lines, they were cross-examined with mouse tracheal epithelial cells obtained over a time course post-EC exposure and with two non-carcinogen-derived mouse cancer cell lines with defined *Kras* and *Trp53* status: Lewis lung carcinoma cells with mutant *Kras*<sup>G12C</sup> and *Trp53*<sup>WT</sup> and MC38 colon adenocarcinoma cells with mutant *Kras*<sup>G13R</sup> and mutant *Trp53*<sup>R178P</sup> (25,30). RT-PCR and quantitative real-time PCR (qPCR) showed different patterns of mono- or bi-allelic *Trp53* loss (Figure 3C and D), whereas western immunoblots and immunocytochemistry did not detect *Trp53* mutations (Figure 3E; Supplementary Figure 2, available at *Carcinogenesis* Online). Diverse EGFR expression patterns were determined via qPCR and western immunoblots that did not correlate with carcinogen or mouse strain used (Figure 3D and E), whereas Sanger sequencing yielded *Egfr*<sup>WT</sup> in all cell lines. Notably, these results show that our murine tobacco carcinogen-triggered LADC cell lines bear *Kras*<sup>MUT</sup> alleles and exhibit patterns of *Trp53* loss that resemble the human LADC of smokers. Moreover and in accord with a comprehensive genomic screen of carcinogen-induced LADC (10), the data indicate that each tobacco carcinogen inflicts a defined *KRAS* point mutation (single nucleotide variation): EC causes *Kras*<sup>Q61R</sup> and DEN *Kras*<sup>Q61H</sup> mutation. Finally, the data identify for the first time mutant *KRAS*-associated nonsense-mediated decay in our EC-generated cell lines with *Kras*<sup>Q61R</sup> mutations, which together with codon bias can explain the notorious absence of mutant *KRAS* reads in RNA-sequencing studies (29,31).

### Tobacco chemical-induced mouse LADC cell lines overexpress stemness and cancer genes

Global gene expression analysis of our LADC cells in comparison with total lung RNA from naive mice and various other cell types [Gene Expression Omnibus (GEO) DataSets accession IDs GSE94981 for LADC cell lines, lungs, airway epithelial cells, mast cells and macrophages; GSE82154 for alveolar epithelial type II cells and GSE58188 for other cancer



**Figure 2.** Tobacco carcinogen-induced mouse lung carcinomas are classified as adenocarcinomas. Primary urethane (A, F, K) and DEN (B, G, L)-induced LADC tumours that gave rise to the LADC cell lines, secondary subcutaneous tumours generated by transplantation of FULA cells into syngeneic mice (C, H, M), *KRAS*<sup>G12D</sup>-driven LADC (Ref. (45); D, I, N), as well as subcutaneous tumours of pancreatic adenocarcinoma PANO2 cells (Ref. (25); E, J, O), were stained for haematoxylin and eosin (H&E), Periodic acid-Schiff–diastase (PAS-D), and thyroid transcription factor 1 (TTF1). (A–E) H&E-stained representative tumour sections. Note the typical glandular–solenooid structure of LADC tumours (A–D) and the solid form of PANO2 tumours (E). (F–J) PAS-D stain for visualization of mucin. Note the adenocarcinoma–distinctive positive mucin staining of all LADC (F–I) and the negative PAS-D staining of PANO2 tumours (J). (K–O) Immunostaining for TTF1 (NKX2-1). Note the LADC–distinctive nuclear immunoreactivity of all LADC (K–N) and the negative results from PANO2 tumours (O).

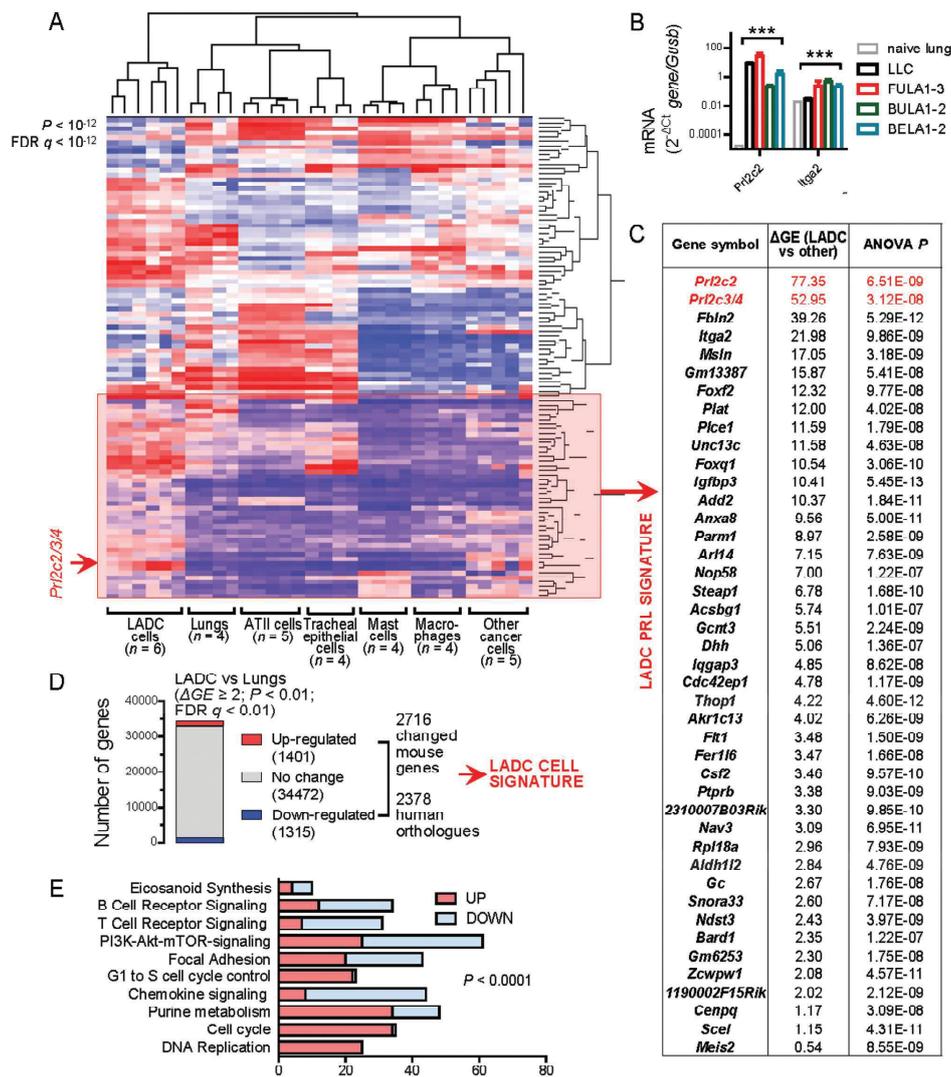


Downloaded from https://academic.oup.com/carcin/article-abstract/40/11/1352/5368948 by GSF Zentralbibliothek user on 16 March 2020

**Figure 3.** Tobacco carcinogen-induced mouse lung adenocarcinoma cell lines bear codon 61 *Kras* mutations and exhibit loss of *Trp53*. (A) *Kras* and *Nras* messenger RNA (mRNA) expression by RT-PCR of select chemical-induced lung adenocarcinoma cell lines and sequences of the amplicons cut and extracted from the gels together with their matching Ensembl annotations. Note the shorter nonsense-mediated decay (NMD) transcript. (B) Complementary DNA Sanger sequencing traces of splenocytes of WT C57BL/6 mouse and of the chemical-induced lung adenocarcinoma cell lines reported here. Note the heterozygous *Kras*<sup>Q61R</sup> and *Kras*<sup>Q61H</sup> single nucleotide variants (arrows) in all urethane- and DEN-induced cell lines, respectively. Note also the superimposition of WT, mutant and NMD *Kras* traces in urethane-induced cell lines. No *Egfr* and *Nras* mutations were detected. (C) *Trp53* mRNA expression by RT-PCR of mouse tracheal epithelial cells cultured from the lungs of urethane-exposed mice at various time-points post-injection, of select chemical-induced lung adenocarcinoma cell lines and of Lewis lung carcinoma (LLC; *Kras*<sup>G12C</sup>, *Trp53*<sup>WT</sup>) and MC38 colon adenocarcinoma (*Kras*<sup>G12S</sup>, *Trp53*<sup>R178P</sup>) cells. (D) *Trp53* and *Egfr* mRNA expression by qPCR of LLC, MC38 and chemical-induced LADC cell lines relative to Gusb. Data are presented as mean ± SD (n = 3/group). \*P < 0.05, \*\*P < 0.01 and \*\*\*P < 0.001, for comparison with LLC cells by one-way analysis of variance with Bonferroni post-tests. (E) EGFR, KRAS, NRAS, TRP53 and ACTB protein expression of LLC, chemical-induced LADC and CT26 (*Kras*<sup>WT</sup>, *Trp53*<sup>WT</sup>) and MC38 colon adenocarcinoma cells by western immunoblot. Note the absence of detectable labile TRP53<sup>WT</sup> expression in all but MC38 cells that bear mutant *Trp53*<sup>R178P</sup> that results in abnormally stable but non-functional TRP53 protein. The immunoblot has been cropped. LADC cell line naming convention XYLA# denotes X for mouse strain (F, FVB; B, Balb/c), Y for carcinogen used (U, EC; E, DEN), LA for lung adenocarcinoma and # for serial number by derivation date.

cells; freely available at <https://www.ncbi.nlm.nih.gov/gds/>] identified a distinct transcriptomic pattern of LADC cells, including differential expression of a 43 gene-signature that

comprised several cancer and stemness genes, such as *Itga2* (32) and PRL transcripts *Prl2c2/Prl2c3/Prl2c4*. Gene expression was validated by qPCR (Figure 4A–C). Similar gene expression

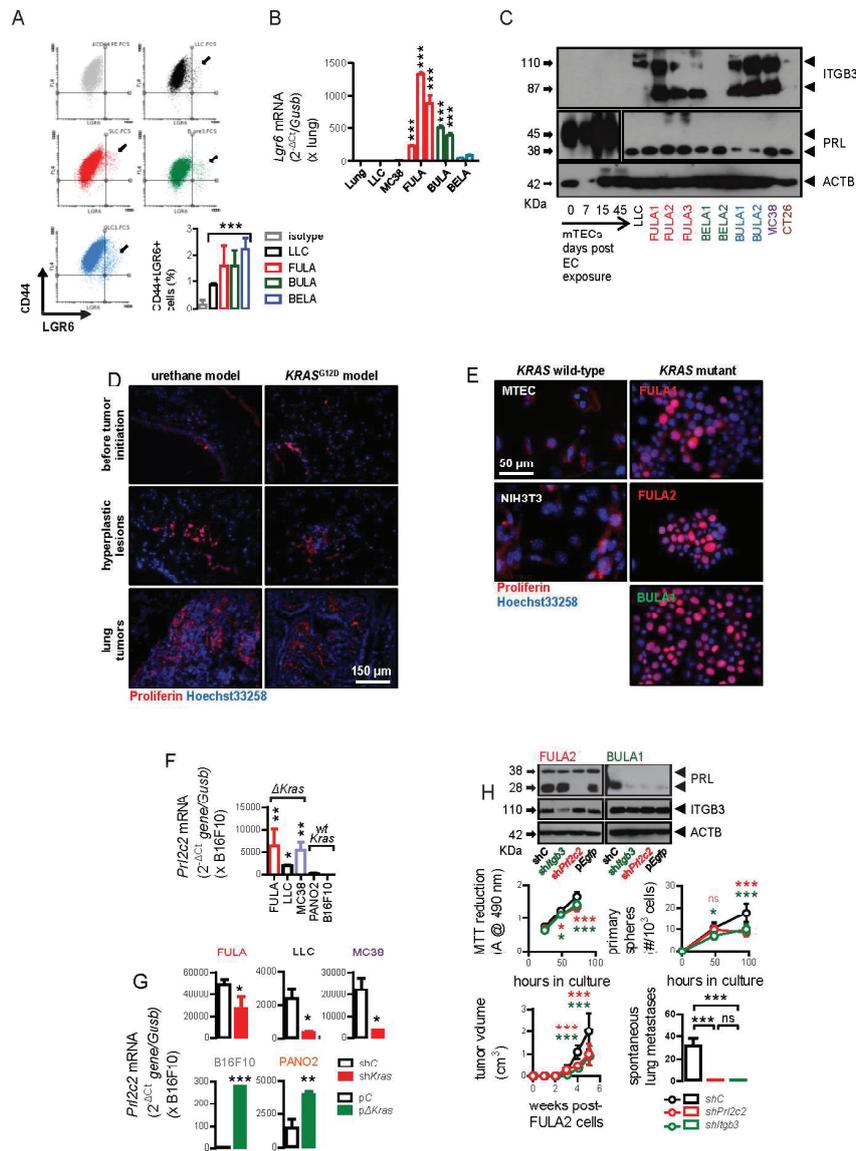


**Figure 4.** Comparative transcriptome profiling of carcinogen-induced mouse lung adenocarcinoma cell lines identifies focal overexpression of PRLs. (A) Unsupervised hierarchical clustering of global transcriptomes of chemical-induced LADC cell lines, total mouse lung RNA, alveolar type 2 cells, airway epithelial cells, bone marrow-derived mast cells and macrophages and other cancer cell lines by microarray [GEO DataSets accession IDs GSE94981 for LADC cell lines, lungs, tracheal epithelial cells, mast cells and macrophages; GSE82154 for alveolar epithelial type 2 cells (ATII cells) and GSE58188 for other cancer cell lines; freely available at <https://www.ncbi.nlm.nih.gov/gds/>]. Cut-off used was statistical significance by analysis of variance (ANOVA) ( $P$ ) and false discovery rate  $q < 10^{-12}$ . (B) PRL transcript (*Pr2c2*, *Pr2c3* and *Pr2c4*) and *Itga2* expression of mouse LADC cells and naive lungs relative to *Gusb* by qPCR. Data are presented as mean  $\pm$  SD. \*\*\* $P < 0.001$  for all comparisons with naive lungs by one-way ANOVA with Bonferroni post-tests. (C) The 43 transcripts differentially expressed in LADC cells versus all other groups and clustering together with PRL transcripts (A, red box) comprising the LADC PRL signature.  $\Delta GE$ , differential gene expression; ANOVA, analysis of variance;  $P$ , probability. (D, E) Summary of the murine genes differentially expressed in LADC cells and naive lungs using the cut-offs shown and of their human orthologues comprising the LADC cell line signature (D) and pathway analysis thereof (E).

analyses comparing LADC cell lines only to mouse lungs revealed a broader transcriptomic signature of the LADC cells, which indicated significant perturbation of pathways significant for cancer cells, such as DNA replication, cell cycle and purine metabolism pathways (Figure 4D and E). To examine the stemness of LADC cells, we determined RNA and protein levels of the lung and cancer stemness markers *Lgr6* and *Itgb3* (33,34). Remarkably, LADC cells displayed significant expression levels of both genes (Figure 5A–C; Supplementary Figures 3 and 4, available at *Carcinogenesis* Online), indicating a prominent cancer stemness potency in line with their tumour sphere-forming capacity.

### The prolactin orthologue PRL drives the *in vitro* and *in vivo* growth of tobacco chemical-induced mouse LADC cell lines

Because microarray analyses identified PRL transcripts to be the most abundantly and specifically overexpressed by LADC cell lines compared with other samples (Figure 4B; and Supplementary Figure 5, available at *Carcinogenesis* Online), we validated the microarray (Figure 4C; Supplementary Figure 6, available at *Carcinogenesis* Online) and sought to functionally investigate its role in LADC development and evolution. Interestingly, PRL was overexpressed in experimental murine



**Figure 5.** Carcinogen-induced mouse lung adenocarcinoma cell lines overexpress lung and cancer stemness markers. Proliferin drives lung adenocarcinoma growth in vitro and in vivo. (A) Representative dotplots and data summary of flow cytometry of lung cells and LLC and chemical-induced LADC cell lines for the cancer stem cell marker CD44 and the lung stem cell marker LGR6 identified significant proportions of CD44+LGR6+ cells (arrows). Data are presented as mean ± SD (n = 3/group). \*\*\* P < 0.001 for comparison with lung cells by one-way analysis of variance (ANOVA) with Bonferroni post-tests. (B) *Lgr6* messenger RNA expression by qPCR of lungs and LLC, MC38 and chemical-induced LADC cell lines relative to *Gusb*. Data are presented as mean ± SD (n = 3/group). \*\*\* P < 0.001 for comparison with lungs by one-way ANOVA with Bonferroni post-tests. (C) ITGB3, PRL and ACTB protein expression of mouse tracheal epithelial cells (mTECs) cultured from the lungs of urethane-exposed mice at various time-points post-injection, of select chemical-induced LADC cell lines and of LLC, MC38 colon adenocarcinoma and CT26 colon adenocarcinoma cells by western immunoblot. Immunoblot has been cropped. (D) Immunoreactivity of murine lungs from the urethane and KRAS<sup>G12D</sup> LADC models for proliferin (red colour) before tumour initiation (top), at early stages of tumour progression (middle) and when harbouring LADC (bottom). Blue colour indicates nuclear Hoechst33258 counterstaining. Note the increased nuclear PRL expression in LADCs. (E) Immunoreactivity of benign mouse tracheal epithelial cells (mTECs) and NIH 3T3 fibroblasts and of select chemical-induced LADC cells for proliferin (red colour). Blue colour indicates nuclear Hoechst33258 counterstaining. Note the increased nuclear PRL expression in LADCs. (F) *Prl2c2* messenger RNA expression by qPCR of different mouse cancer cell lines with (FULA, LLC, MC38) and without (PANO2, B16F10) *Kras* mutations (25) relative to *Gusb*. Data are presented as mean ± SD (n = 3/group). \*\*\* P < 0.001 for comparison with B16F10 cells by one-way ANOVA with Bonferroni post-tests. (G) *Prl2c2* gene expression of parental and *Kras*-modulated (red: sh*Kras*-expressing; green: pΔ*Kras*2B-expressing) cancer cell lines (25) relative to *Gusb* by qPCR shows that *Prl2c2* expression is KRAS-driven. Data are presented as mean ± SD (n = 3/group). \*P < 0.05, \*\*P < 0.01 and \*\*\*P < 0.001, for comparison with parental cells by Student's t-test. (H) FULA1 and BULA2 cells were stably transfected with target-specific short hairpin (shRNA) against *Prl2c2* or *Itgb3* or random shRNA pools (shC). Shown are representative results of PRL, ITGB3 and ACTB protein expression by western immunoblot, 3-(4,5-dimethylthiazole-2-yl)-2,5-diphenyl tetrazolium bromide assay and tumour sphere formation capacity in vitro, as well as primary tumour growth and lung metastasis formation in vivo on injection of 1 million cells to syngeneic FVB or Balb/c mice. Data presented are mean ± SD (n = 3–6/group) obtained from FULA2 cells, but identical results were obtained using BULA1 cells. \*P < 0.05 and \*\*\*P < 0.001, for comparison of the colour-coded silenced cells with control-transfected cells by one or two-way ANOVA with Bonferroni post-tests.

hyperplastic lesions and LADC from both urethane and KRAS<sup>G12D</sup> models (Figure 5D). Furthermore, PRL expression was found to be mutant KRAS-associated and -driven, with KRAS<sup>MUT</sup> cells

displaying exclusive nuclear PRL immunoreactivity, in stark contrast with the cytoplasmic signal of KRAS<sup>WT</sup> cells (Figure 5E). Silencing of *Kras* expression in KRAS<sup>MUT</sup> cells resulted in

decreased PRL expression levels whereas plasmid-mediated overexpression of  $\Delta$ Kras2B<sup>G12C</sup> transcript in KRAS<sup>WT</sup> cells induced the levels of PRL (Figure 5F and G). Importantly, short hairpin RNA-mediated silencing of the most abundant PRL transcript *Prl2c2* side by side with the important cancer stemness transcript *Itgb3* led to comparable and significant decreases in *in vitro* cell proliferation and tumour-sphere formation, as well as in *in vivo* subcutaneous tumour growth rates and spontaneous metastatic capacity to the lungs (Figure 5H), indicating that PRL is an important lung tumour promoter as was shown previously for integrin  $\beta$ 3 (33).

### Prolactin is overexpressed in human LADC and portends poor survival

We next assessed a potential role for the human PRL orthologue prolactin (also abbreviated PRL) in human LADC. In a sample set (GEO DataSets accession ID: GSE43458) of 30 normal lung tissues from never smokers, 40 LADC from never smokers and 40 LADC from smokers from the BATTLE trial (35), a 77 gene set representing the PRL signalling pathway [Wiki Prolactin Signaling Pathway, Homo Sapiens; <http://www.wikipathways.org/index.php/Pathway:WP2037> (36)] could accurately cluster normal samples from cancer tissues (Figure 6A and B). Gene expression analysis revealed that PRL messenger RNA was significantly overexpressed in LADC tissues compared with normal lung tissues from the BATTLE study (Figure 6C). Moreover, PRL immunoreactivity was stronger in LADC tissues from our centre (37) compared with surrounding non-cancerous lung tissues (Figure 6D). Interestingly, a 2378 gene transcriptomic signature of our LADC cell lines identified earlier (Figure 4D) managed to accurately cluster normal from cancer samples of the BATTLE study (Figure 6E). Furthermore, gene set enrichment analysis (38); <http://software.broadinstitute.org/gsea/index.jsp>) of the transcriptomic signature and the BATTLE study dataset samples revealed highly positive enrichment of the transcriptomic signature in LADC samples from smokers but negative enrichment in LADC samples from never smokers (Figure 6F).

Importantly, patients with lung cancer from the Kaplan-Meier Plotter database (<http://kmplot.com/analysis/index.php?p=service&default=true>) (39); with high PRL expression displayed significantly shorter survival compared with patients with lower expression levels. Interestingly, the dismal survival effect of PRL was restricted to female patients with LADC (Figure 7A–E). Remarkably, multivariate Cox regression analyses revealed that together with increasing tumour stage, high PRL expression is an independent negative prognosticator of overall survival in patients with lung cancer, independent of sex, smoking history and histology (Figure 7F and G).

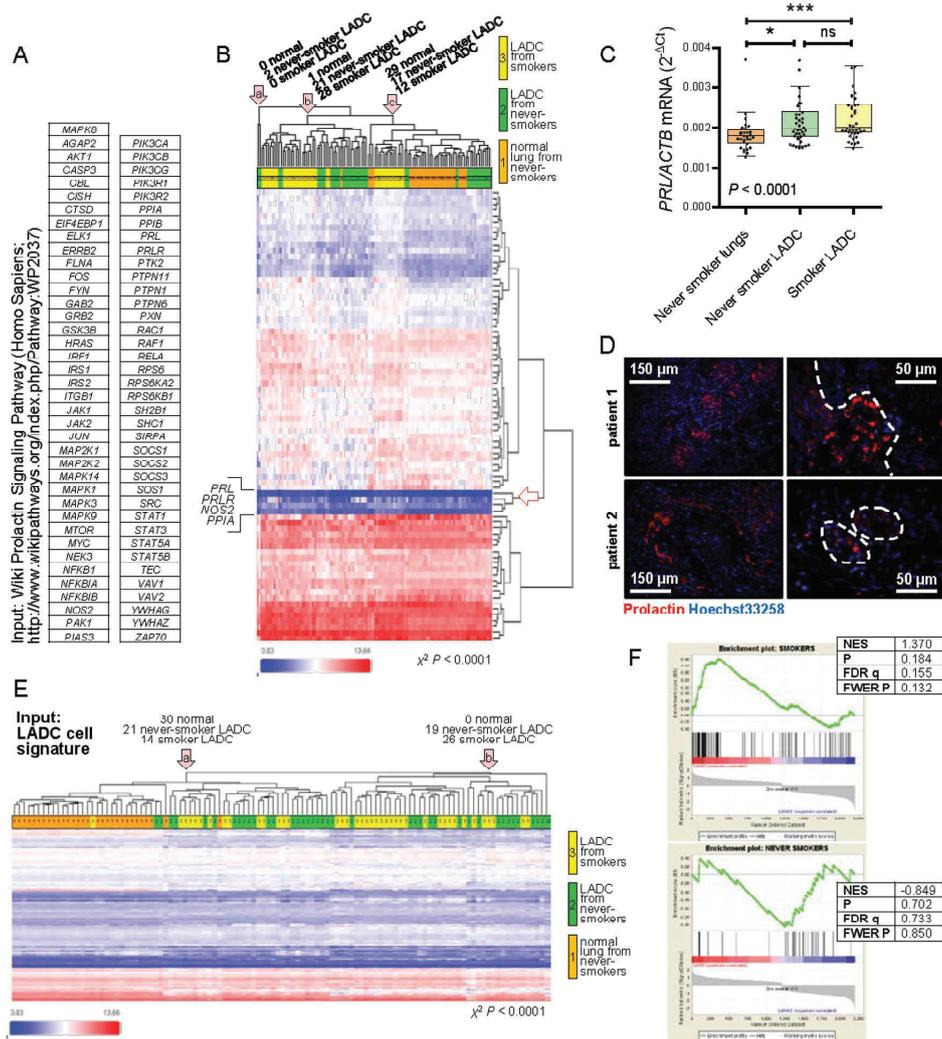
### Discussion

Here, we report the first-ever derivation of multiple true LADC cell lines obtained from lung tumours generated in mice by exposure to tobacco carcinogens. We comprehensively characterized their properties *in vitro* and *in vivo* and clearly show that they present true adenocarcinoma cell lines that (i) display cancer stem cell properties; (ii) grow and metastasize in the lungs and pleural space of syngeneic mice similar to the human disease; and (iii) carry expression and mutation profiles that resemble human LADC of smokers. Importantly, a universal signature of these cell lines across different tobacco carcinogens and mouse strains used to generate them is shown to be present in human LADC, rendering them relevant tools for research on this disease. This signature revealed that these tobacco carcinogen-inflicted

LADC cell lines depend on PRL for sustained growth and metastasis. Moreover, the human PRL orthologue prolactin was overexpressed in human LADC and was linked with poor survival. Hence, our murine LADC cell lines prove for the first time that tobacco chemical-induced lung tumours in mice are indeed malignant, address the unmet need for faithful mouse models of smoking-induced human LADC in syngeneic immunocompetent mice and present exciting new tools for the discovery of novel drivers and treatments of the human disease in the future.

This is the first study designed and implemented to develop transplantable mouse models of human tobacco carcinogen-induced LADC. Although lung cancer is the leading cause of cancer death worldwide (1,2), LADC is its most common histological subtype with increasing incidence, and tobacco smoking is the main cause of the disease (3–5), tools for research are still sparse. Although cell lines and transplantable models have spearheaded lung cancer research and discovery, only a handful of murine cell lines for syngeneic transplantable models exist, complemented by a multitude of human cell lines for xenograft models in immunocompromised mice that lack an adaptive immune system (16,40). This shortcoming has been overcome by the development of transgenic mouse models that recapitulate salient features of human LADC (41,42). However, genetic LADC models are not metastatic (43), display copy number alterations rather than the heavy load of single nucleotide variants found in human LADC of smokers (6) and in chemical-induced LADC of mice (10) and often display histological appearances not reminiscent of the human disease (40). Although chemical LADC models were long discovered and widely used, they were neglected in the era of transgenic models, thought to present adenomas rather than carcinomas. However, the strengths of chemical models, including their high mutation load, the predominance of single nucleotide variants and their interaction between carcinogen, exposure protocol and host genetic background render them lucrative (10,13,16). To this end, we used tobacco chemicals to identify a cardinal role for nuclear factor- $\kappa$ B signalling in LADC (44,45) and another group discovered important mechanisms of genomic context- and organ-specific KRAS-driven carcinogenesis using chemical models (46,47). Our work provides for the first time mouse models of LADC that combine the strengths of transplantable and chemical models: our LADC cell lines are readily transplantable in syngeneic mice, metastasize like human LADC and carry *Kras/Trp53* lesions and gene expression profiles that resemble the human disease.

In addition to new research tools, our findings also provide important conceptual advances. By applying Robert Koch's postulates, we prove for the first time beyond doubt that at least some chemical-induced lung tumours in mice are malignant adenocarcinomas. In addition, the isolation of the true tumour-initiating cells from these tumours will probably lead to the identification of new disease drivers and mechanisms, such as the PRL/prolactin pathway reported here. Future sequencing of these cells will hopefully yield yet unknown perturbed genes and pathways that go undetected by large scale molecular fingerprinting projects that examine heterotypic tumours (10,27). This can be appreciated by the average ~70-fold overexpression of PRL by our LADC cell lines relative to naive murine lungs, compared with the ~1.3- to 1.5-fold overexpression of prolactin in human LADC relative to naive human lungs (Figure 6C; Supplementary Figure 5, available at *Carcinogenesis* Online). We are currently fingerprinting our cell lines, aiming at the functional identification of the genomic imprints of different tobacco (and other) carcinogens on the murine DNA in the nucleotide, trinucleotide, gene, locus and chromosome levels, aiming to expand and

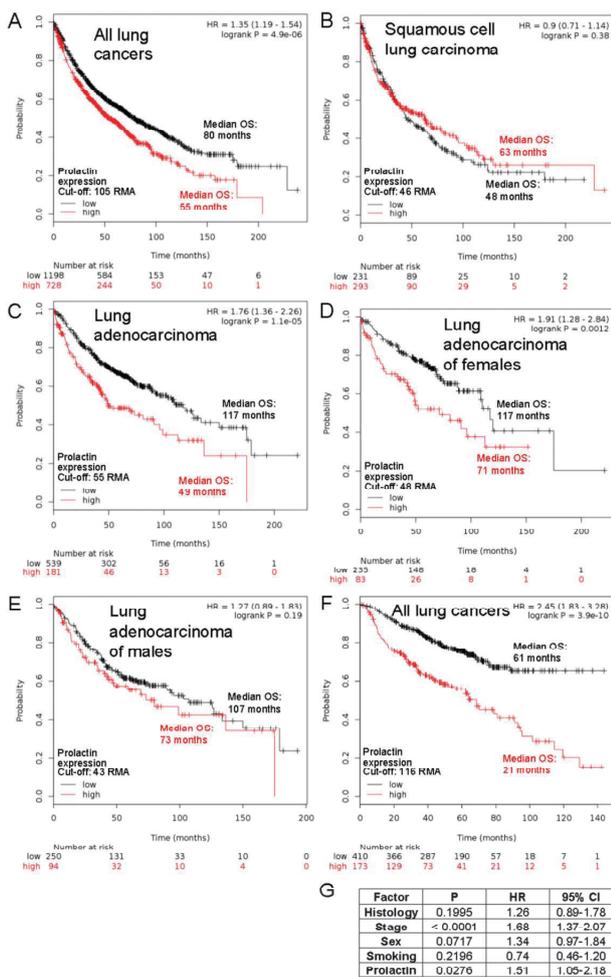


**Figure 6.** Prolactin is overexpressed in human lung adenocarcinoma. (A) Genes that comprise the Wiki Prolactin Signaling Pathway (Homo Sapiens; <http://www.wikipathways.org/index.php/Pathway/WP2037>) used as input for analyses of the BATTLE study in (B) (36). (B) Unsupervised hierarchical clustering of the BATTLE trial transcriptomic dataset including 30 normal lung tissues from never smokers, 40 LADC from never smokers and 40 LADC from smokers (GEO DataSets accession ID: GSE43458 (35)); by the Wiki Prolactin Signaling Pathway from (A) significantly distinguishes normal samples from LADC tissues ( $\chi^2$  and hypergeometric test  $P < 0.0001$ ). (C) Prolactin (PRL) transcript PRL normalized to  $\beta$ -actin (ACTB) transcript ACTB expression of patients from the BATTLE trial shows increased PRL expression in LADC compared with normal lung tissues. Data are presented as median with Tukey's whiskers (boxes: interquartile range; bars: 50% extreme quartiles) and raw data points (dots) ( $n = 30$ –40/group).  $P$  denotes overall one-way analysis of variance probability and \* and \*\*\* denote  $P < 0.05$  and  $P < 0.001$ , respectively, for comparisons indicated by Bonferroni post-tests. (D) Immunoreactivity of representative LADC tissues from our centre (37) for prolactin (red colour). Blue colour indicates nuclear Hoechst33258 counterstaining. Note the increased PRL expression in LADCs (dashed lines) compared with adjacent tissues. (E) Unsupervised hierarchical clustering of the BATTLE trial transcriptomic dataset by a 2378 gene transcriptomic signature of our LADC cell lines identified in this study (Figure 4D) significantly distinguishes normal samples from LADC tissues ( $\chi^2$  and hypergeometric test  $P < 0.0001$ ). (F) Gene set enrichment analysis (GSEA (38); <http://software.broadinstitute.org/gsea/index.jsp>) of a 2378 gene transcriptomic signature of our LADC cell lines identified in this study (Figure 4D) in smokers' and never-smokers' LADC from the BATTLE study. Note that the signature of our tobacco carcinogen-induced LADC cell lines was significantly positively enriched in LADC from smokers from the BATTLE trial, but negatively enriched in LADC from never smokers. NES, normalized enrichment score; P, nominal probability; FDR q, false discovery rate probability; FWER P, family-wise error rate probability. Note that FDR q and FWER  $P < 0.25$  are considered significant in GSEA.

validate the important findings of Alexandrov et al., who initiated the process of defining carcinogenic DNA imprints by clinical correlation across multiple human cancers (48,49). Taking into account that tumour initiating cells might be resistant to conventional chemotherapy regimens, LADC cells may also present optimal models to study drug response and test novel therapies (50).

The method to generate chemical-induced LADC cell lines reported here may also be useful for future research, as it can be applied to any genetically modified mouse strain, yielding a

powerful tool to study gene function in cancer, comprising the methods currently available for this, such as CRISPR/Cas9. For example, we have derived LADC cells from urethane-treated mice carrying conditionally deleted (floxed) *Trp53* alleles and have performed causes recombination-mediated recombination (i.e. *Trp53* deletion) *in vitro*, gaining important insights into the role of functional *Trp53* in osteopontin signalling (30). We are currently applying this technique to an array of reporter, knock-in/out and conditional mice, garnering important insights into LADC biology.



**Figure 7.** Prolactin expression is associated specifically with poor survival of female patients with lung adenocarcinoma. (A–E) Kaplan–Meier survival plots with univariate Cox regression hazard ratios (HR) and 95% confidence intervals (95% CI) of patients with lung cancer (A), squamous cell lung carcinoma (B), lung adenocarcinoma (C) and females (D) and males (E) with lung adenocarcinoma, stratified by prolactin messenger RNA (mRNA) expression as determined by microarray (probe ID: 205445\_at). Optimal cut-offs were determined by dichotomizing patient data by all possible percentiles of prolactin expression. Data were from the Kaplan–Meier Plotter database (<http://kmplot.com/analysis/index.php?p=service&default=true>) (39). Note the sex- and histology-specific impact of prolactin expression on survival. (F, G) Kaplan–Meier survival plot with multivariate Cox regression HR and 95% CI of patients with lung cancer stratified by prolactin mRNA expression (F). Analyses were done as earlier, this time using multivariate function and entering prolactin expression, histological subtype, stage, sex and smoking history as co-variables. Results of multivariate Cox regression analysis (G) show that together with increasing stage, high prolactin expression is an independent dismal predictor of survival in lung cancer. RMA, robust multi-array units; OS, overall survival; HR, hazard ratio of high versus low expressing patients; 95% CI, 95% confidence intervals; P, log-rank test or Cox regression probability values.

LADC cell lines were derived from two mouse strains (FVB and Balb/c) exposed repeatedly to the cigarette carcinogens urethane (EC) and diethylnitrosamine (DEN). This procedure simulates tobacco smoking in humans (10,16). Importantly, the culture of these cells does not introduce changes other than those induced by the carcinogenic process (26), a fact pending validation. To this end, transcriptomic analyses revealed

ubiquitous altered expression of a signature comprised 2716 genes in our LADC cell lines that included a 43 gene set tightly linked with PRL. PRL and its human counterpart prolactin were validated as potent LADC drivers using observational studies in murine and human LADC, as well as functional studies in mice. The discovery of the role of PRL/prolactin signalling in LADC underscores the value of our LADC cell lines as research vehicles and warrants further investigation of PRL/prolactin as candidate therapeutic targets.

In conclusion, the tobacco carcinogen-inflicted murine LADC cell lines reported and made available here are valuable tools for research and discovery and can be used in multifaceted ways for future identification of molecular signatures, driver genes and pathways and drugs against LADC. These cell lines made possible the identification of PRL/prolactin signalling as lung tumour promoter.

## Supplementary material

Supplementary data are available at *Carcinogenesis* online.

## Funding

European Research Council 2010 Starting Independent Investigator and 2015 Proof of Concept Grants (grant numbers 260524 and 679345 respectively to G.T.S.). I.G. is a recipient of a Greek State Scholarship Foundation (IKY) programme co-financed by the European Union (European Social Fund-ESF), by Greek national funds through an action entitled ‘Reinforcement of Postdoctoral Researchers’ (NSRF 2014–2020).

**Conflict of Interest Statement:** The authors declare no competing interests.

## References

- Torre, L.A. et al. (2016) Global cancer incidence and mortality rates and trends—an update. *Cancer Epidemiol. Biomarkers Prev.*, 25, 16–27.
- Torre, L.A. et al. (2015) Global cancer statistics, 2012. *CA. Cancer J. Clin.*, 65, 87–108.
- Alberg, A.J. et al. (2013) Epidemiology of lung cancer: diagnosis and management of lung cancer, 3rd ed: American College of Chest Physicians evidence-based clinical practice guidelines. *Chest*, 143, e1S–e29S.
- Sun, S. et al. (2007) Lung cancer in never smokers—a different disease. *Nat. Rev. Cancer*, 7, 778–790.
- Hecht, S.S. (1999) Tobacco smoke carcinogens and lung cancer. *J. Natl. Cancer Inst.*, 91, 1194–1210.
- Garraway, L.A. et al. (2013) Lessons from the cancer genome. *Cell*, 153, 17–37.
- Ding, L. et al. (2008) Somatic mutations affect key pathways in lung adenocarcinoma. *Nature*, 455, 1069–1075.
- Graziano, S.L. et al. (1999) Prognostic significance of K-ras codon 12 mutations in patients with resected stage I and II non-small-cell lung cancer. *J. Clin. Oncol.*, 17, 668–675.
- Nelson, M.A. et al. (1996) Detection of K-ras gene mutations in non-neoplastic lung tissue and lung cancers. *Cancer Lett.*, 103, 115–121.
- Westcott, P.M. et al. (2015) The mutational landscapes of genetic and chemical models of Kras-driven lung cancer. *Nature*, 517, 489–492.
- Ollila, S. et al. (2011) The tumor suppressor kinase LKB1: lessons from mouse models. *J. Mol. Cell Biol.*, 3, 330–340.
- Jackson, E.L. et al. (2001) Analysis of lung tumor initiation and progression using conditional expression of oncogenic K-ras. *Genes Dev.*, 15, 3243–3248.
- de Seranno, S. et al. (2010) Progress and applications of mouse models for human lung cancer. *Eur. Respir. J.*, 35, 426–443.
- Desai, T.J. et al. (2014) Alveolar progenitor and stem cells in lung development, renewal and cancer. *Nature*, 507, 190–194.

15. You, M. et al. (1989) Activation of the Ki-ras protooncogene in spontaneously occurring and chemically induced lung tumors of the strain A mouse. *Proc. Natl. Acad. Sci. USA.*, 86, 3070–3074.
16. Meuwissen, R. et al. (2005) Mouse models for human lung cancer. *Genes Dev.*, 19, 643–664.
17. Tuveson, D.A. et al. (1999) Modeling human lung cancer in mice: similarities and shortcomings. *Oncogene*, 18, 5318–5324.
18. Fassett, J.T. et al. (2000) Mrp4, a new mitogen-regulated protein/proliferin gene; unique in this gene family for its expression in the adult mouse tail and ear. *Endocrinology*, 141, 1863–1871.
19. Nilsen-Hamilton, M. et al. (1987) Detection of proteins induced by growth regulators. *Methods Enzymol.*, 147, 427–444.
20. Nilsen-Hamilton, M. et al. (1987) Relationship between mitogen-regulated protein (MRP) and proliferin (PLF), a member of the prolactin/growth hormone family. *Gene*, 51, 163–170.
21. Wilder, E.L. et al. (1986) Expression of multiple proliferin genes in mouse cells. *Mol. Cell. Biol.*, 6, 3283–3286.
22. Fang, Y. et al. (1999) Signaling between the placenta and the uterus involving the mitogen-regulated protein/proliferins. *Endocrinology*, 140, 5239–5249.
23. Corbacho, A.M. et al. (2002) Roles of prolactin and related members of the prolactin/growth hormone/placental lactogen family in angiogenesis. *J. Endocrinol.*, 173, 219–238.
24. Toft, D.J. et al. (2001) Reactivation of proliferin gene expression is associated with increased angiogenesis in a cell culture model of fibrosarcoma tumor progression. *Proc. Natl. Acad. Sci. USA.*, 98, 13055–13059.
25. Agalioi, T. et al. (2017) Mutant KRAS promotes malignant pleural effusion formation. *Nat. Commun.*, 8, 15205.
26. Pauli, C. et al. (2017) Personalized *In Vitro* and *In Vivo* cancer models to guide precision medicine. *Cancer Discov.*, 7, 462–477.
27. Cancer Genome Atlas Research, N. (2014) Comprehensive molecular profiling of lung adenocarcinoma. *Nature*, 511, 543–550.
28. Wu, K. et al. (2015) Frequent alterations in cytoskeleton remodeling genes in primary and metastatic lung adenocarcinomas. *Nat. Commun.*, 6, 10131.
29. Chang, Y.F. et al. (2007) The nonsense-mediated decay RNA surveillance pathway. *Annu. Rev. Biochem.*, 76, 51–74.
30. Giopanou, I. et al. (2017) Tumor-derived osteopontin isoforms cooperate with TRP53 and CCL2 to promote lung metastasis. *Oncoimmunology*, 6, e1256528.
31. Lampson, B.L. et al. (2013) Rare codons regulate KRas oncogenesis. *Curr. Biol.*, 23, 70–75.
32. Su, Y.J. et al. (2015) Polarized cell migration induces cancer type-specific CD133/integrin/Src/Akt/GSK3 $\beta$ / $\beta$ -catenin signaling required for maintenance of cancer stem cell properties. *Oncotarget*, 6, 38029–38045.
33. Seguin, L. et al. (2014) An integrin  $\beta_3$ -KRAS-RalB complex drives tumour stemness and resistance to EGFR inhibition. *Nat. Cell Biol.*, 16, 457–468.
34. Oeztuerk-Winder, F. et al. (2012) Regulation of human lung alveolar multipotent cells by a novel p38 $\alpha$  MAPK/miR-17-92 axis. *EMBO J.*, 31, 3431–3441.
35. Kabbout, M. et al. (2013) ETS2 mediated tumor suppressive function and MET oncogene inhibition in human non-small cell lung cancer. *Clin. Cancer Res.*, 19, 3383–3395.
36. Kelder, T. et al. (2012) WikiPathways: building research communities on biological pathways. *Nucleic Acids Res.*, 40(Database issue), D1301–D1307.
37. Giopanou, I. et al. (2015) Comprehensive evaluation of nuclear factor- $\kappa$ B expression patterns in non-small cell lung cancer. *PLoS One*, 10, e0132527.
38. Subramanian, A. et al. (2007) GSEA-P: a desktop application for Gene Set Enrichment Analysis. *Bioinformatics*, 23, 3251–3253.
39. Györfy, B. et al. (2013) Online survival analysis software to assess the prognostic value of biomarkers using transcriptomic data in non-small-cell lung cancer. *PLoS One*, 8, e82241.
40. Gazdar, A.F. et al. (2016) Correction: “From mice to men and back: an assessment of preclinical model systems for the study of lung cancers”. *J. Thorac. Oncol.*, 11, e88–e89.
41. Kim, C.F. et al. (2005) Mouse models of human non-small-cell lung cancer: raising the bar. *Cold Spring Harb. Symp. Quant. Biol.*, 70, 241–250.
42. Kwak, I. et al. (2004) Genetically engineered mouse models for lung cancer. *Annu. Rev. Physiol.*, 66, 647–663.
43. Vanharanta, S. et al. (2013) Origins of metastatic traits. *Cancer Cell*, 24, 410–421.
44. Stathopoulos, G.T. et al. (2007) Epithelial NF- $\kappa$ B activation promotes urethane-induced lung carcinogenesis. *Proc. Natl. Acad. Sci. USA.*, 104, 18514–18519.
45. Vreka, M. et al. (2018) I $\kappa$ B Kinase  $\alpha$  is required for development and progression of KRAS-mutant lung adenocarcinoma. *Cancer Res.*, 78, 2939–2951.
46. To, M.D. et al. (2013) Interactions between wild-type and mutant Ras genes in lung and skin carcinogenesis. *Oncogene*, 32, 4028–4033.
47. To, M.D. et al. (2008) Kras regulatory elements and exon 4A determine mutation specificity in lung cancer. *Nat. Genet.*, 40, 1240–1244.
48. Alexandrov, L.B. et al. (2016) Mutational signatures associated with tobacco smoking in human cancer. *Science*, 354, 618–622.
49. Behjati, S. et al.; ICGC Prostate Group. (2016) Mutational signatures of ionizing radiation in second malignancies. *Nat. Commun.*, 7, 12605.
50. Zhou, B.B. et al. (2009) Tumour-initiating cells: challenges and opportunities for anticancer drug discovery. *Nat. Rev. Drug Discov.*, 8, 806–823.

Tobacco chemical-induced mouse lung  
adenocarcinoma cell lines pin the prolactin  
orthologue proliferin as a lung tumour promoter

Nikolaos I. Kanellakis, Anastasios D. Giannou, Mario A. Pepe, Theodora Agalioti,  
Dimitra E. Zazara, Ioanna Giopanou, Ioannis Psallidas, Magda Spella, Antonia  
Marazioti, Kristina A. M. Arendt, Anne-Sophie Lamort, Spyridon Champeris Tsaniras,  
Stavros Taraviras, Helen Papadaki, Ioannis Lilis, and Georgios T. Stathopoulos

**Supplementary Material**

## Supplementary Materials and methods

### Mouse models of endogenous lung adenocarcinoma

*FVB* (#001800), *Balb/c* (#001026), *FVB-Tg*(CAG-luc,-GFP)L2G85Chco/J (called CAG.Luc.eGFP; #008450; (1), B6.129S4-Krastm4Tyj/J (called *KRAS*<sup>G12D</sup>; #008179; (2), and *C57BL/6* (#000664) mice were obtained from Jackson Laboratories (Bar Harbor, MN) and were bred on the corresponding background at the University of Patras Center for Animal Models of Disease. Experimental mice were sex-, weight (20-25 g)-, and age (6-12 week)-matched.

For chemically induced lung carcinogenesis, *FVB*, and *Balb/c* mice received the tobacco carcinogens urethane (Sigma Aldrich, U2500) intraperitoneally (1g/Kg in 100  $\mu$ l phosphate-buffered saline) or diethylnitrosamine (200 mg/kg) (Sigma Aldrich, N0756) and were sacrificed after ten months (3,4). For mutant *KRAS*-driven lung tumorigenesis, *C57BL/6* mice heterozygous for the loxP-STOP-loxP.*KRAS*<sup>G12D</sup> transgene (*KRAS*<sup>G12D</sup> mice), which express mutant *KRAS* in any somatic cell upon CRE-mediated recombination, received  $5 \times 10^8$  intratracheal plaque-forming units of adenovirus encoding CRE recombinase (Ad-*Cre*; Baylor College of Medicine, Houston, TX) and were killed after four months (2).

### Mouse models of transplantable lung adenocarcinoma

For heterotopic lung adenocarcinoma (LADC) development, mice were anesthetized using isoflurane inhalation and received s.c. injections of 100  $\mu$ L phosphate-buffered saline (PBS) containing  $1 \times 10^6$  mouse cancer cells. Three vertical tumor dimensions ( $\delta_1$ ,  $\delta_2$ , and  $\delta_3$ ) were monitored longitudinally and tumor volume was calculated

using the formula  $\pi * \delta_1 * \delta_2 * \delta_3 / 6$  as described elsewhere (5-7). Mice were sacrificed after 3-4 weeks. For forced lung metastasis induction, mice were anesthetized using isoflurane inhalation and received i.v. injections of 100  $\mu$ L phosphate-buffered saline (PBS) containing  $0.25 \times 10^6$  murine cells. Mice were sacrificed after two weeks. Lung tumors were counted and sized using a Stemi DV4 stereoscope (Zeiss, Jena, Germany) in transillumination mode to visualize both superficial and deep lung metastases. For malignant pleural effusion precipitation, syngeneic mice received 150,000 murine cancer cells intrapleurally and mice were sacrificed after two weeks. Pleural effusions and tumors were evaluated as described elsewhere (5,6).

### **Isolation and culture of mouse cell types and lines**

Mouse airway epithelial cells were cultured from the stripped tracheal epithelia of eight-week-old *FVB* and *Balb/c* mice as described previously (8). Bone marrow-derived macrophages were generated after one-week culture of whole bone marrow cells flushed from the four long bones (two tibias and two femurs) of *FVB* mice with 100 ng/mL recombinant murine (rm) macrophage colony-stimulating factor, thereby passaging the adherent cells, as described elsewhere (9). Bone marrow-derived mast cells were generated after four-week culture of whole bone marrow cells flushed from the four long bones (two tibias and two femurs) of *FVB* mice with 100 ng/mL rm interleukin-3 with or without 100 ng/mL rm KIT-ligand, thereby passaging the non-adherent cells (7).

### **Cytology, cytometry, and histology**

For May-Grünwald Giemsa stain, cells were fixed with methanol for 2 min, were stained with May-Grünwald stain in 1 mM Na<sub>2</sub>HPO<sub>4</sub>, 2.5 mM KH<sub>2</sub>PO<sub>4</sub>, pH = 6.4 for 6 min, and subsequently with Giemsa stain in 2 mM Na<sub>2</sub>HPO<sub>4</sub>, 5 mM KH<sub>2</sub>PO<sub>4</sub>, pH = 6.4 for 40 min, were washed with H<sub>2</sub>O, and were dried. Slides were mounted with Entellan (Merck Millipore, Darmstadt, Germany), coverslipped, and analyzed. For flow cytometry and fluorescence-activated cell sorting (FACS), 10<sup>6</sup> cells suspended in 50 ml FACS buffer (PBS supplemented with 2% FBS and 0,1% NaN<sub>3</sub>) were stained with the indicated antibodies according to the manufacturer's instructions and recommended dilutions (Supplementary table ST1) for 20 min in the dark, were washed with FACS buffer from excess antibody, and were resuspended in 1ml FACS buffer for further analysis. For histology, murine lungs were inflated, fixed in 4% paraformaldehyde overnight, embedded in paraffin or OCT and were stored at room temperature or -80°C, respectively. Five-µm paraffin or 10-µm-cryosections were mounted on glass slides. Sections were labeled using the indicated antibodies (Supplementary table ST1), counterstained with Envision (Dako, Carpinteria, CA) or Hoechst 33258 (Sigma), and mounted with Entellan new (Merck Millipore) or Mowiol 4-88 (Calbiochem, Gibbstown, NJ). For isotype control, primary antibody was omitted. Immunoreactivity was quantified as described previously (10). For hematoxylin-eosin staining, slides were incubated with deionized H<sub>2</sub>O for 2 minutes at room temperature and transferred to hematoxylin solution (Papanicolaou's solution 1b hematoxylin solution S; Merck Millipore) for 30 seconds at room temperature (RT). Then slides were washed up with tap water, incubated for 1 second in 1% acid alcohol at RT, incubated with lithium solution for 4 seconds at RT and washed with tap water again. Subsequently slides were incubated in eosin solution (Eosin Y solution 0.5% alcoholic; Merck Millipore) for 1 minute at RT and

washed with tap water. PAS-D staining was done as described elsewhere (11). Finally slides were dehydrated, coverslipped, and analyzed. Bright-field and fluorescent microscopy were carried out on AxioLab.A1 (Zeiss), AxioObserver.D1 (Zeiss) or TCS SP5 (Leica) microscopes. Digital images were processed with Fiji academic software (12) .

### **Cell proliferation and tumor-sphere assays**

In vitro cancer cell proliferation was determined using the 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT) assay, as described previously (10). Tumor-sphere generation assay was performed as described elsewhere (13)

### **Bioluminescent imaging**

Mice were imaged on a Xenogen Lumina II after i.v. delivery of 1 mg D-luciferin (Gold Biotechnology). Data were analyzed on Living Image v.4.2 (Perkin-Elmer, Waltham, MA) (6).

### **RNA Extraction, Sanger sequencing, and gene expression analyses**

Cellular RNA was isolated using Trizol (Thermo Fisher Scientific, Waltham, MA) followed by RNAeasy purification and genomic DNA removal (Qiagen, Hilden, Germany). For lung tissue RNA, tissues were passed through 70  $\mu$ m strainers (BD Biosciences, San Jose, CA) and  $10^7$  cells were subjected to RNA extraction. One  $\mu$ g RNA was reverse-transcribed using Oligo(dT)<sub>18</sub> and Superscript III (Thermo Fisher). *Kras*, and *Nras* cDNAs were amplified using specific primers (Supplementary table ST2) and Phusion Hot Start Flex polymerase (New England Biolabs, Ipswich, MA). DNA fragments were run on 2% agarose gels and were purified with NucleoSpin gel

and PCR clean-up columns (Macherey-Nagel, Düren, Germany) and were sequenced using the appropriate primers by VBC Biotech (Vienna, Austria). qPCR was performed using specific primers (Supplementary table ST2) 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 (2). mRNA abundance was determined relative to  $\beta$ -glucuronidase (*Gusb*) and is given as  $2^{-\Delta CT} = 2^{-(Ct \text{ of transcript})-(Ct \text{ of } Gusb)}$ . Microarrays were done as described elsewhere (10).

### **Flow cytometry and Immunoblotting**

Cell cytometry and data analysis were performed on a CyFlow ML instrument using FloMax Software (Partec GmbH, Münster, Germany). A CyFlow ML instrument with FloMax Software (Partec, Munster, Germany) was used for cell cytometry, sorting, and data analysis. Total protein extracts from cultured cells were prepared using  $Mg^{2+}$  lysis/wash buffer [25 mM HEPES (pH=7.5), 150 mM NaCl, 1% NP-40, 10 mM  $MgCl_2$ , 1 mM EDTA, 2% glycerol]. Proteins were separated by 8-15% Sodium Dodecyl Sulfate - Polyacrylamide Gel Electrophoresis (SDS-PAGE) and were electroblotted to PVDF membranes (Merck Millipore). Membranes were probed with primary antibodies followed by incubation with appropriate horseradish peroxidase-conjugated secondary antibodies (Supplementary table ST1) and were visualized with enhanced chemiluminescence (Merck Millipore).

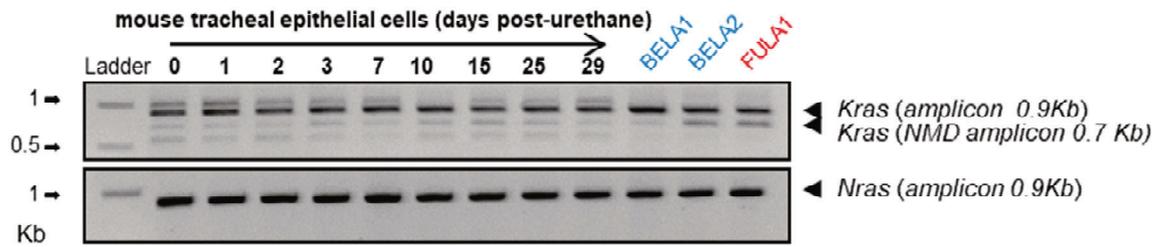
### **Statistics**

Sample size for *in vivo* experiments was calculated using G\*power [(<http://www.gpower.hhu.de/>) assuming  $\alpha = 0.05$ ,  $\beta = 0.8$ , and effect size  $d = 1.5$ . No

data were excluded. Data acquisition was blinded on samples previously coded by a non-blinded investigator. All data were examined for normality by Kolmogorov-Smirnov test. Values are given as mean  $\pm$  SD or median(interquartile range, IQR), as appropriate and indicated. Sample size ( $n$ ) refers to biological replicates. Differences in means were examined by t-test or one- or two-way ANOVA with Bonferroni post-hoc tests.  $P$  values are two-tailed and  $P < .05$  was considered significant. Analyses and plots were done on Prism v5.0 (GraphPad, La Jolla, CA).

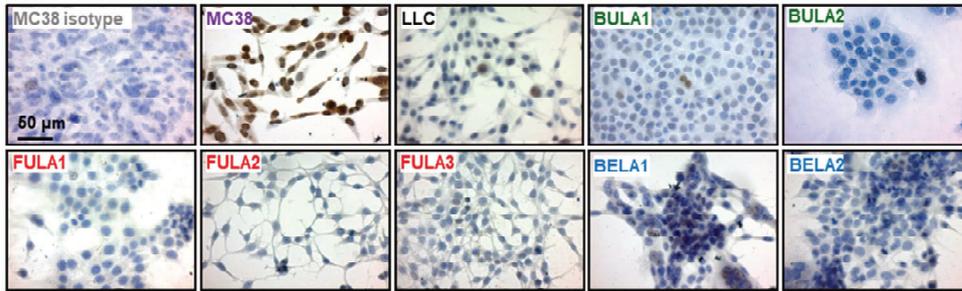
## Supplementary Figures

### Supplementary Figure SF1



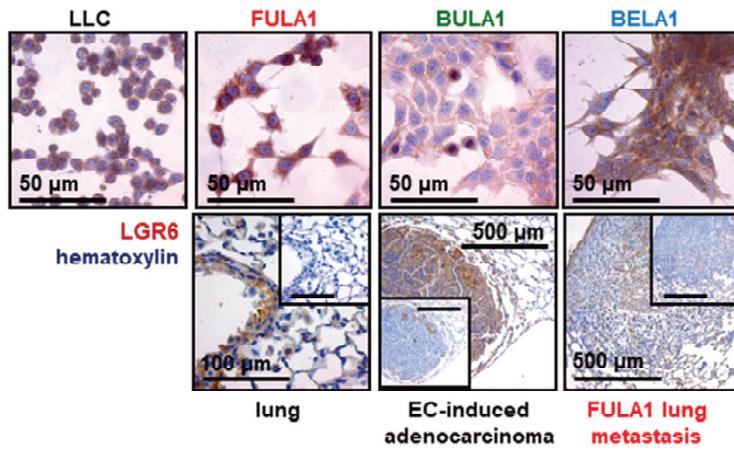
*Kras* and *Nras* mRNA expression by RT-PCR of mouse tracheal epithelial cells (mTECs) cultured from the lungs of urethane-exposed mice at various time-points post-injection and of select chemical-induced lung adenocarcinoma cell lines.

## Supplementary Figure SF2



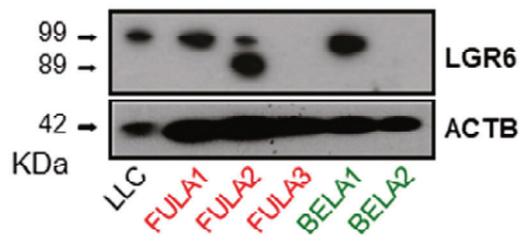
Immunoreactivity (brown color) of LLC, MC38, and chemical-induced LADC cell lines for TRP53 protein by immunocytochemistry. Note blue nuclear hematoxylin staining and the strong nuclear TRP53 immunoreactivity of MC38 cells.

Supplementary Figure SF3



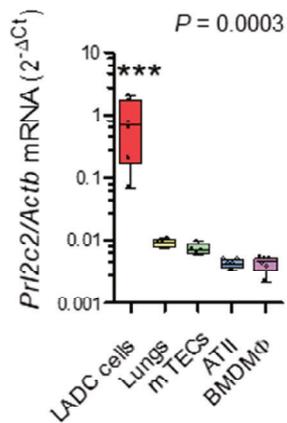
Immunoreactivity of LLC and urethane (EC)-induced LADC cell lines, naïve lungs, chemical-induced LADCs, and lung metastases induced by LADC cell lines for LGR6 (brown color). Blue color indicates hematoxylin counterstaining.

**Supplementary Figure SF4**



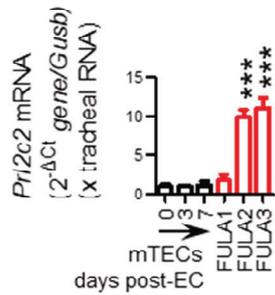
LGR6 and ACTB protein expression of LLC and select chemical-induced LADC cells by Western immunoblot.

## Supplementary Figure SF5



Proliferin (PRL) transcript Prl2c2 normalized to  $\beta$ -actin transcript (Actb) expression of chemical-induced lung adenocarcinoma (LADC) cell lines, total mouse lung RNA, mouse tracheal epithelial cells (mTECs) alveolar type II cells and bone marrow-derived macrophages by microarray (GEO Datasets accession IDs are given in Figure 4). Data are presented as median with Tukey's whiskers (boxes: interquartile range; bars: 50% extreme quartiles) and raw data points (dots) ( $n = 4-6$ /group).  $P$  denotes overall one-way ANOVA probability and \*\*\* denote  $P < 0.001$  for comparison of LADC cells with any other group by Bonferroni post-tests.

## Supplementary Figure SF6



Pri2c2 mRNA expression by qPCR of mouse tracheal epithelial cells (mTECs) cultured from the lungs of urethane-exposed mice at various time-points post-injection and of urethane-induced LADC cell lines relative to Gusb. Data are presented as mean±SD (n = 3/group). \*\*\* denotes P < 0.001 for comparison with mTECs by one-way ANOVA with Bonferroni post-tests.

## Supplementary Tables

### Supplementary Table ST1

#### Antibodies

<b>Antibody</b>	<b>Company</b>	<b>Product number</b>
Proliferin Antibody (E-10)	Santa Cruz Biotechnology	sc-271891
p53 Antibody (FL-393)	Santa Cruz Biotechnology	sc-6243
K-Ras Antibody (F234)	Santa Cruz Biotechnology	sc-30
N-Ras Antibody (F155)	Santa Cruz Biotechnology	sc-31
$\beta$ -Actin Antibody (C4)	Santa Cruz Biotechnology	sc-47778
Anti-Integrin beta 3 antibody [EPR2417Y]	Abcam	ab75872
Anti-EGFR antibody [EP38Y]	Abcam	ab52894
Anti-GPCR LGR6 antibody [EPR6874]	Abcam	ab126747
CD44 Monoclonal Antibody (IM7), PE-Cyanine5, eBioscience (Flow Cytometry)	Thermo Fisher Scientific	15-0441-82
Anti/TTF1 antibody	Thermo Fisher Scientific	MA5-16406

## Supplementary Table ST2

### PCR primers

Method <sup>a</sup>	Primer <sup>b</sup>	Sequence	Amplicon length
qPCR	Trp53F	CGCCGACCTATCCTTACCAT	bp 120
qPCR	Trp53R	TTCTTCTTCTGTACGGCGGT	
qPCR	EgfrF(1)	ATCAAAGTTCTGGGTTCTGGG	bp 156
qPCR	EgfrR(1)	CATCACATAGGCTTCGTCAAGG	
qPCR	EgfrF(2)	AACTGTACCTATGGATGTGCTG	bp 154
qPCR	EgfrR(2)	GGATTTGGAAGAACTGGAAGG	
qPCR	ProliferinF	CATCTCCAAAGCCACAGACAT	bp 145
qPCR	ProliferinR	GCGAGCATCTTCATTGTCAG	
qPCR	Itgb2F	GAATGCCTACTATAAACTCTCCTC	bp 117
qPCR	Itgb2R	GATTTGCCTATACTCGATGCT	
qPCR	Lgr6F	ATGACCTTGGCTCTCAACCA	bp 100
qPCR	Lgr6R	GCTGGATGCGGTTGTTATGT	
qPCR	GusbF	TTACTTTAAGACGCTGATCACC	bp 165
qPCR	GusbR	ACCTCCAAATGCCCATAGTC	
Sanger Seq	KrasF	CCATTTCCGACCCGGAG	bp 905
Sanger Seq	KrasR	CTTTAGTCTCTTCCACAGGCA	
Sanger Seq	NrasF	GCGCCTAGTGATTACGTAGC	bp 905
Sanger Seq	NrasR	TGAAGAGGTCTCAGGTTAGATGG	
RT-PCR	Trp53F	GTAGCTTCAGTTCATTGGGA	bp 1450
RT-PCR	Trp53R	TGAAGTCATAAGACAGCAAGGA	bp 1450

<sup>a</sup> Application: qPCR: quantitative (real-time) PCR, Sanger Seq: Sanger Sequencing, RT-PCR: Reverse transcription PCR <sup>b</sup> F: Forward, R Reverse

### Supplementary Table ST3

#### Oligonucleotides

Oligonucleotide	Company	Product number
<p><b>Integrin <math>\beta</math>3 shRNA</b> (m) Lentiviral Particles is a pool of 3 different shRNA plasmids:</p> <p>sc-35677-VA Hairpin sequence: GATCCGCTACAGTATGTGATGAAATTCAAG AGATTTTCATCACATACTGTAGCTTTTT</p> <p>sc-35677-VB Hairpin sequence: GATCCCATCCCATTTGCTAGTGTTCATCAAG AGAAACACTAGCAAATGGGATGTTTT</p> <p>sc-35677-VC Hairpin sequence: GATCCGTCAGTATGTGGGAATGTATTCAAG AGATACATTCCCACATACTGACTTTTT</p>	Santa Cruz Biotechnology	sc-35677-V
<p><b>Proliferin-1 shRNA</b> (m) Lentiviral Particles is a pool of 3 different shRNA plasmids:</p> <p>sc-61412-VA Hairpin sequence: GATCCGCTTCAGAATGGAGATGAATTCAAG AGATTCATCTCCATTCTGAAGCTTTTT</p> <p>sc-61412-VB Hairpin sequence: GATCCCCTGAAGTGTTACATGTTATTCAAG AGATAACATGTAACACTTCAGGTTTT</p> <p>sc-61412-VC Hairpin sequence: GATCCCTCTGCTTCTGAAATATCATTCAAG AGATGATATTTTCAGAAGCAGAGTTTT</p>	Santa Cruz Biotechnology	sc-61412-V
<p><b>K-Ras shRNA</b> (m) Lentiviral Particles is a pool of 3 different shRNA plasmids:</p> <p>sc-43876-VA Hairpin sequence: GATCCCTACAGGAAACAAGTAGTATTCAAG AGATACTACTTGTTCCTGTAGTTTT</p> <p>sc-43876-VB Hairpin sequence: GATCCGAACAGTAGACACGAAACATTCAAG AGATGTTTCGTGTCTACTGTTCTTTTT</p> <p>sc-43876-VC Hairpin sequence: GATCCCATTTCAGTTTCCATGTTATTCAAG AGATAACATGGAACTGAATGGTTTT</p>	Santa Cruz Biotechnology	sc-43876-V
<p><b>Mutant Kras plasmid</b> eGFP.KRASG12C-2B.retro.puro</p>	Addgene	64372

## References

1. Cao, Y.A., *et al.* (2004) Shifting foci of hematopoiesis during reconstitution from single stem cells. *Proc Natl Acad Sci U S A*, **101**, 221-6.
2. Jackson, E.L., *et al.* (2001) Analysis of lung tumor initiation and progression using conditional expression of oncogenic K-ras. *Genes Dev*, **15**, 3243-8.
3. Hecht, S.S. (2003) Tobacco carcinogens, their biomarkers and tobacco-induced cancer. *Nat Rev Cancer*, **3**, 733-44.
4. Stathopoulos, G.T., *et al.* (2007) Epithelial NF-kappaB activation promotes urethane-induced lung carcinogenesis. *Proc Natl Acad Sci U S A*, **104**, 18514-9.
5. Stathopoulos, G.T., *et al.* (2008) Host nuclear factor-kappaB activation potentiates lung cancer metastasis. *Mol Cancer Res*, **6**, 364-71.
6. Marazioti, A., *et al.* (2013) Beneficial impact of CCL2 and CCL12 neutralization on experimental malignant pleural effusion. *PLoS One*, **8**, e71207.
7. Giannou, A.D., *et al.* (2015) Mast cells mediate malignant pleural effusion formation. *J Clin Invest*, **125**, 2317-34.
8. You, Y., *et al.* (2002) Growth and differentiation of mouse tracheal epithelial cells: selection of a proliferative population. *Am J Physiol Lung Cell Mol Physiol*, **283**, L1315-21.
9. Manzanero, S. (2012) Generation of mouse bone marrow-derived macrophages. *Methods Mol Biol*, **844**, 177-81.
10. Agalioti, T., *et al.* (2017) Mutant KRAS promotes malignant pleural effusion formation. *Nat Commun*, **8**, 15205.
11. Fu, D.A., Campbell-Thompson, M. (2017) Periodic Acid-Schiff Staining with Diastase. *Methods Mol Biol*, **1639**, 145-49.
12. Schindelin, J., *et al.* (2012) Fiji: an open-source platform for biological-image analysis. *Nat Methods*, **9**, 676-82.
13. Johnson, S., *et al.* (2013) In vitro Tumorsphere Formation Assays. *Bio Protoc*, **3**.

### 3. Publication II: Club cells form lung adenocarcinomas and maintain the alveoli of adult mice

#### 3.1. Summary

Lung cancer and chronic lung diseases represent a major health problem and are caused by the inhalation of noxious substances, such as tobacco smoke. LUAD is mainly caused by chemical tobacco smoke carcinogens that induce alterations of *KRAS* in still unknown pulmonary cells<sup>7, 9, 10, 32</sup>. Previous studies of pulmonary lineage tracing have identified as LUAD progenitors both airway and alveolar cells<sup>23, 29</sup>. However, all those studies suffered of the peculiar promiscuity or incomplete lung cell lineage labeling of the existing lineage tracing mouse models. This resulted in failing in the complete identification of all cells belonging to a given lineage (false negative marking) or detecting other cells outside of the target lineage (false positive marking). Moreover, existing studies that tried to address the cellular origins of LUAD used genetically engineered mouse models based on the overexpression of oncogenes such as *KRAS*<sup>G12D</sup> in the lungs<sup>23, 29</sup>. Recently, it was shown that these genetic models do not mimic the mutational profiles of human LUAD as well as of chemical-induced mouse models<sup>18</sup>. Using the more human-relevant carcinogen-triggered LUAD models, we aimed to define the cell lineage/s that initiate tumorigenesis. For the purpose of cell lineage labeling, we crossed a CRE-reporter mouse strain that switches somatic cells from membranous mt/td Tomato to membranous GFP upon CRE-mediated recombination<sup>36</sup> with six different CRE-driver strains. This allowed permanent and accurate labeling of different lung cell lineages. Moreover, the co-localization of GFP-labeling with cell lineage protein markers showed that GFP in CCSP-CRE mice marks all airway epithelial cells including club and ciliated cells. In the same way, GFP in LYZ2-CRE mice marks some alveolar type 2 cell and alveolar macrophages. We next induced LUAD with urethane (also known as ethyl carbamate, EC) or 3-methylcholanthrene injections. Both carcinogens were able to give rise to preneoplastic lesions and neoplasia located in airways and in the alveolar regions. Amazingly, all tumors of GFP;CCSP-CRE mice showed GFP-labeled airway cells that did not express the club cell marker CCSP but acquired the expression of alveolar epithelial markers SFTPC with or without LYZ2. These results were then recapitulated using a single urethane hit on GFP;CCSP-

CRE, GFP;SFTPC-CRE, and GFP;LYZ2-CRE mice backcrossed to the susceptible FVB strain, which develops human-like alterations like  $KRAS^{Q61R}$ . We then established and applied digital droplet PCR (ddPCR) to detect in which lung lineage the  $KRAS^{Q61R}$  arose at early time points after single urethane exposures. To enable the simultaneous detection of  $KRAS$  and CRE-cassette status, an *ad hoc* technical assay was designed and tested. One week past the EC hit, both GFP; CCSP-CRE and GFP;LYZ2-CRE mouse strains showed  $KRAS^{Q61R}$  mutations, but  $KRAS^{Q61R}$  mutations selectively persisted in GFP-labeled airway cells in the lungs of GFP;CCSP-CRE mice at two weeks. In summary, we addressed in this publication the accomplice role of airway epithelial cells, and more specifically of club cells, as cells of origin in LUAD.

### 3.2. Contribution

My personal contribution to the publication consisted in developing an assay to quantitatively, spatially, and longitudinally define the presence of *Kras* mutations in carcinogen-induced mouse models of LUAD. Among different possibilities, we decided to utilize Digital droplet polymerase chain reaction (ddPCR)<sup>37</sup>. With a standard ddPCR assay it is possible to analyze a single genetic condition per experiment. My contribution was to repurpose this assay with a new approach to allow the detection of two different genetic conditions with a single assay and analyze the relative data. A detailed overview of technical method, statistical and analytical approaches implemented are provided in Appendix A. The ddPCR custom primers and probe are described in Spella, M. *et al.* 2019 materials and methods. The results are shown in Spella, M. *et al.* 2019 figure 2A and figure 2 supplement.

### 3.3. Publication

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

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

<sup>1</sup>Laboratory for Molecular Respiratory Carcinogenesis, Department of Physiology, Faculty of Medicine, University of Patras, Rio, Greece; <sup>2</sup>Comprehensive Pneumology Center (CPC), Institute for Lung Biology and Disease (iLBD), University Hospital, Ludwig-Maximilians University, Helmholtz Center Munich, The German Center for Lung Research (DZL), Munich, Germany; <sup>3</sup>Division of Molecular Thoracic Oncology, Translational Lung Research Center (TLRC), German Cancer Research Center (DKFZ), The German Center for Lung Research (DZL), Heidelberg, Germany; <sup>4</sup>Institute of Immunology, Biomedical Sciences Research Center "Alexander Fleming", Vari, Greece; <sup>5</sup>First Department of Critical Care Medicine and Pulmonary Services, School of Medicine, Evangelismos Hospital, National and Kapodistrian University of Athens, Athens, Greece

\*For correspondence:  
magasp@upatras.gr (MS);  
gstathop@upatras.gr (GTS)

†These authors contributed equally to this work

**Competing interests:** The authors declare that no competing interests exist.

**Funding:** See page 26

**Received:** 28 January 2019

**Accepted:** 24 May 2019

**Published:** 29 May 2019

**Reviewing editor:** Jody Rosenblatt, King's College London, United Kingdom

© Copyright Spella et al. This article is distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use and redistribution provided that the original author and source are credited.

**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.

DOI: <https://doi.org/10.7554/eLife.45571.001>

## Introduction

Chronic lung diseases present tremendous health burdens attributed to dysfunctional alveolar repair (*Barnes et al., 2015; Lozano et al., 2012; Spella et al., 2017*). 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 (*Torre et al., 2015; Forbes et al., 2011; Hecht, 1999; Campbell et al., 2016; Cancer Genome Atlas Research Network, 2014*). 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 (*Barbie et al., 2009; Seguin et al., 2014*). To this end,

**eLife digest** The deadliest form of lung cancer is called lung adenocarcinoma, or LUAD. Tobacco chemicals often cause the disease by damaging the genetic information of lung cells. The damage leads to harmful changes in the DNA sequence which prompt the cells to form tumors. For instance, the most common of these changes takes place in a gene called *KRAS*. However, it is still unclear exactly which type of lung cells are more likely to develop into a tumor.

In the lungs, airway epithelial cells cover the inside of the passages that bring the air inside little sacks called alveoli, which are lined by alveolar cells. Previous studies have used genetic methods to switch on the *KRAS* mutation in different compartments of the mouse lung. This showed that groups of airway cells, of alveolar cells, and of a class of cells located at the junction between airways and alveoli could all give rise to cancer. However, these experiments did not examine how tobacco chemicals could give rise to tumors in different groups of lung cells.

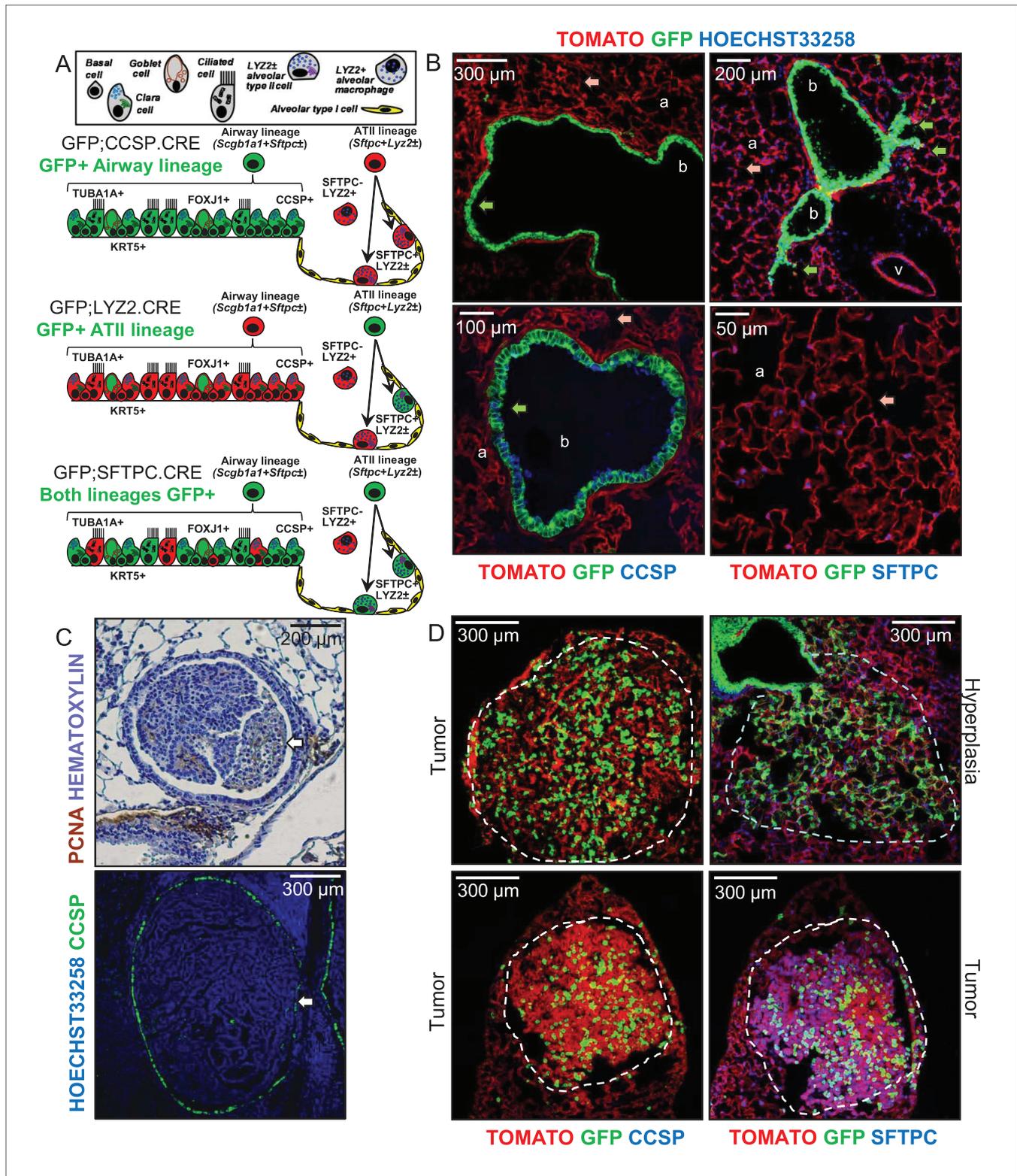
Here, Spella et al. triggered LUAD in adult mice by exposing them to the toxic chemicals found in tobacco smoke, but without making any change to the *KRAS* gene. These mice also had genetically engineered reporters that could be used to deduce where the resulting tumors came from. DNA sequencing showed that the airway epithelial cells gained *KRAS* mutations after the chemical treatment. When the airway epithelial cells were experimentally removed before the treatments with tobacco chemicals, these mice did not get LUAD tumors. Spella et al. also observed that the tobacco-induced tumors came from the epithelial cells in the airways, and not from the cells in the alveoli. Moreover, when the lung was damaged, airway cells could move to the alveoli and start adopting the identity of alveolar cells, thereby replenishing this population. Together, these experiments imply that tobacco-induced LUAD starts in the airway epithelial cells.

These findings suggest that airway epithelial cells could be targeted to stop lung cancer early on. Further studies should also examine how airway epithelial cells can transition to look more like alveolar cells when the lungs get harmed.

DOI: <https://doi.org/10.7554/eLife.45571.002>

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 (Imielinski et al., 2017).

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 (Zuo et al., 2015; Kim et al., 2005; Cho et al., 2011; Xu et al., 2012; Sutherland et al., 2014; Mainardi et al., 2014; Desai et al., 2014). 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 (AM $\Phi$ ) 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 AM $\Phi$ , are summarized in **Figure 1A** and **Figure 1—figure supplement 1**, and are extensively studied in Desai et al. (2014) and Treutlein et al. (2014). However, existing mouse models for lineage tracing feature incomplete and/or promiscuous lung cell labeling, that is cellular markings fail to identify all cells of a target lineage (false negative marking) or wrongly identify other cells outside of the target lineage (false positive marking) (Zuo et al., 2015; Kim et al., 2005; Cho et al., 2011; Xu et al., 2012; Sutherland et al., 2014; Mainardi et al., 2014; Desai et al., 2014). In addition, all studies that attempted to address the cellular origins of LUAD to date employed overexpression of oncogenes such as *KRAS*<sup>G12D</sup> in the lungs, to conclude that ATII cells or BASC are the most probable culprits of the disease (Kim et al., 2005; Cho et al., 2011; Xu et al., 2012; Sutherland et al., 2014; Mainardi et al., 2014; Desai et al., 2014). However, it was recently shown that oncogenic *KRAS*<sup>G12D</sup>-driven mouse lung tumors do not imitate the



**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 *Figure 1 continued on next page*

Figure 1 continued

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) and hematoxylin (blue)-stained (top) and CCSP (green) and Hoechst33258 (blue)-stained (bottom) lung tumor sections of urethane-treated C57BL/6 mice six months post-treatment ( $n = 5/\text{group}$ ), 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, top), and immunostained for the club cell marker CCSP (bottom left) and the alveolar type II cell marker SFTPC (bottom 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  $\alpha$ -tubulin; SFTPC, surfactant protein C; LY2Z, lysozyme 2; FOXJ1, forkhead box J1; KRT5, keratin 5.

DOI: <https://doi.org/10.7554/eLife.45571.003>

The following source data and figure supplements are available for figure 1:

**Figure supplement 1.** Table of pulmonary lineage markers and key abbreviations used in this study.

DOI: <https://doi.org/10.7554/eLife.45571.004>

**Figure supplement 2.** Genetic labeling of pulmonary lineages in eleven mouse strains and intercrosses: summary of results.

DOI: <https://doi.org/10.7554/eLife.45571.005>

**Figure supplement 3.** Genetic labeling of pulmonary lineages in seven lineage reporter strains on the C57BL/6 background: representative images.

DOI: <https://doi.org/10.7554/eLife.45571.006>

**Figure supplement 4.** Genetic labeling of pulmonary lineages in seven lineage reporter strains on the C57BL/6 background: data summary.

DOI: <https://doi.org/10.7554/eLife.45571.007>

**Figure supplement 4—source data 1.** Quantification of GFP+ alveolar and bronchial cells in our reporter mice.

DOI: <https://doi.org/10.7554/eLife.45571.008>

**Figure supplement 5.** Flow cytometric quantification of lineage-labeled cells in three lineage reporter strains on the C57BL/6 background.

DOI: <https://doi.org/10.7554/eLife.45571.009>

**Figure supplement 5—source data 1.** Flow cytometric quantification of GFP+ and TOMATO+ cells in three lineage reporter mice.

DOI: <https://doi.org/10.7554/eLife.45571.010>

**Figure supplement 6.** Genetic lineage labels of protein-marked cells in three lineage reporter strains on the C57BL/6 background: representative images.

DOI: <https://doi.org/10.7554/eLife.45571.011>

**Figure supplement 7.** Genetic lineage labels of protein-marked cells in seven lineage reporter strains on the C57BL/6 background: data summary.

DOI: <https://doi.org/10.7554/eLife.45571.012>

**Figure supplement 7—source data 1.** Quantification of GFP+/SFTPC+ and GFP+/CCSP+ cells in our reporter mice.

DOI: <https://doi.org/10.7554/eLife.45571.013>

**Figure supplement 8.** Protein markings of lineage-labeled cells in three lineage reporter strains on the C57BL/6 background: data summary.

DOI: <https://doi.org/10.7554/eLife.45571.014>

**Figure supplement 8—source data 1.** Quantification of protein marker expression of GFP+ cells in three lineage reporter mice.

DOI: <https://doi.org/10.7554/eLife.45571.015>

**Figure supplement 9.** Two carcinogen regimens for reproducible lung tumor induction in naturally resistant C57BL/6 mice.

DOI: <https://doi.org/10.7554/eLife.45571.016>

**Figure supplement 10.** Lung tumors induced in C57BL/6 mice by two carcinogen regimens.

DOI: <https://doi.org/10.7554/eLife.45571.017>

**Figure supplement 10—source data 1.** Quantification of data shown in **Figure 1—figure supplement 10**.

DOI: <https://doi.org/10.7554/eLife.45571.018>

**Figure supplement 11.** Airway links of urethane-induced lung adenocarcinomas.

DOI: <https://doi.org/10.7554/eLife.45571.019>

**Figure supplement 12.** Genetic labeling of urethane-induced lung adenocarcinomas in four lineage reporter strains on the C57BL/6 background: representative images.

DOI: <https://doi.org/10.7554/eLife.45571.020>

**Figure supplement 13.** Genetic labeling of urethane-induced lung adenocarcinomas in four lineage reporter strains on the C57BL/6 background: data summary.

DOI: <https://doi.org/10.7554/eLife.45571.021>

**Figure supplement 13—source data 1.** Quantification of GFP+ tumors/lung and GFP+ cells/tumor in four lineage reporter mice after urethane exposure.

DOI: <https://doi.org/10.7554/eLife.45571.022>

**Figure supplement 14.** Genetic labeling of MCA/BHT-induced lung adenocarcinomas in two lineage reporter strains on the C57BL/6 background: representative images.

DOI: <https://doi.org/10.7554/eLife.45571.023>

Figure 1 continued on next page

Figure 1 continued

**Figure supplement 15.** Protein marker expression of urethane-induced lung adenocarcinomas in three lineage-labeled mouse strains on the C57BL/6 background: representative images.

DOI: <https://doi.org/10.7554/eLife.45571.024>

**Figure supplement 16.** Genetic lineage labels of protein-marked cells in three lineage reporter strains on the FVB background: representative images.

DOI: <https://doi.org/10.7554/eLife.45571.025>

**Figure supplement 17.** A single-hit mouse model for urethane-induced lung adenocarcinoma induction in naturally susceptible FVB mice.

DOI: <https://doi.org/10.7554/eLife.45571.026>

**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.

DOI: <https://doi.org/10.7554/eLife.45571.027>

**Figure supplement 19.** Genetic labeling of urethane-induced lung adenocarcinomas in three lineage reporter strains on the FVB background: representative images.

DOI: <https://doi.org/10.7554/eLife.45571.028>

mutational landscape of human LUAD as closely as tobacco carcinogen-induced LUAD do (Campbell et al., 2016; Cancer Genome Atlas Research Network, 2014; Westcott et al., 2015).

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 (Miller et al., 2003; Malkinson et al., 1997; Stathopoulos et al., 2007), and corroborated the findings with the FVB strain that is susceptible to single-hit carcinogenesis (Westcott et al., 2015; Stathopoulos et al., 2007; Vreka et al., 2018). 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) (Muzumdar et al., 2007) to six different CRE-driver strains on the C57BL/6 background (Desai et al., 2014; Oikonomou et al., 2012; Okubo et al., 2005; Hayashi et al., 2002; Ogilvy et al., 1998; Tronche et al., 1999). 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, and in Materials and methods and in Appendix 1). Double heterozygote offspring at six postnatal weeks (i.e., after mouse lung development is complete [Zuo et al., 2015; Desai et al., 2014]) were examined for GFP-labeling (results are shown in Figure 1A, Figure 1—figure supplements 3 and 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 AM $\Phi$  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, 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

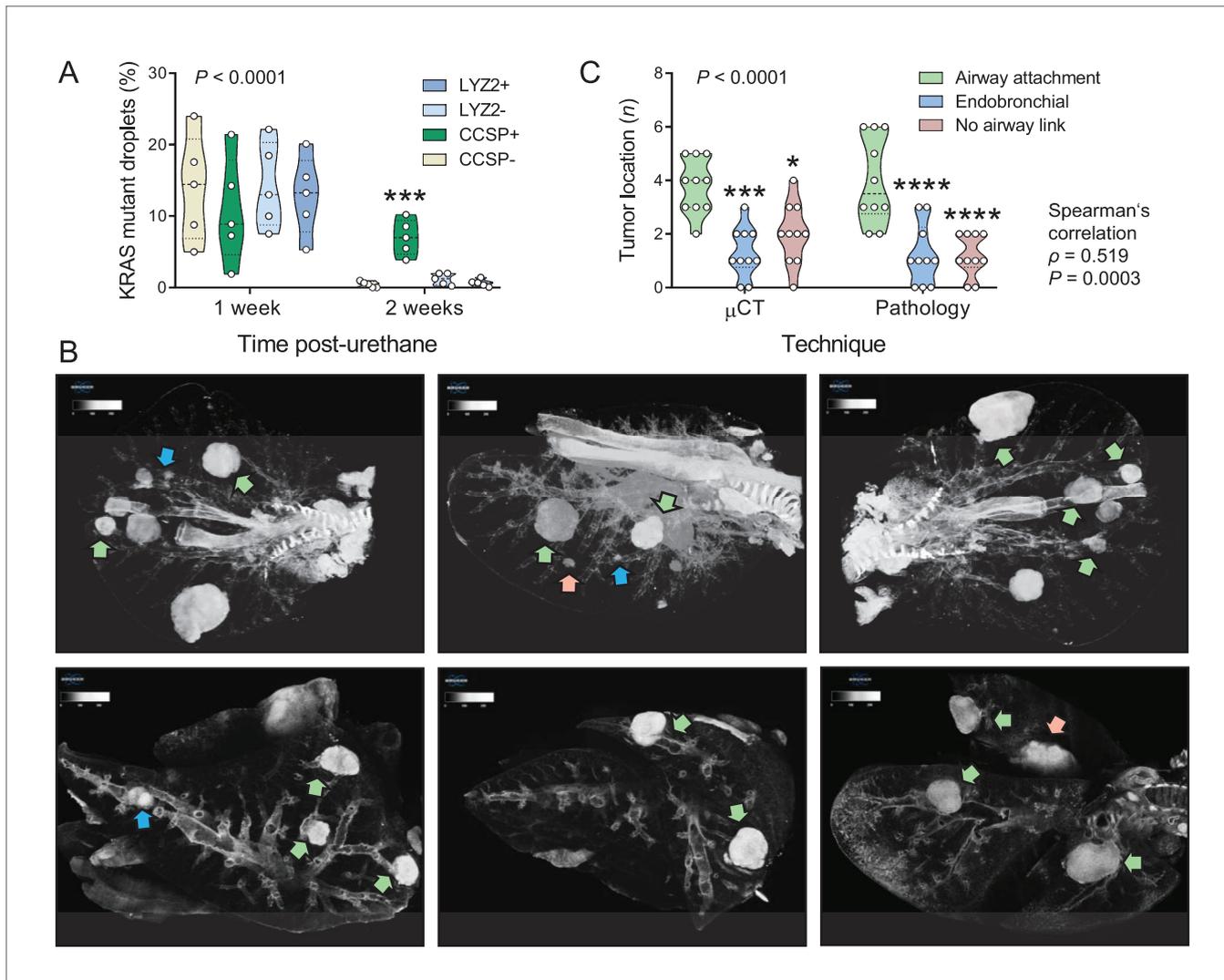
We next triggered LUAD in GFP;CCSP.CRE, GFP;SFTPC.CRE, and GFP;LYZ2.CRE mice on the C57BL/6 background using repetitive exposures to the tobacco carcinogens urethane (ethyl carbamate, EC; stand-alone mutagen and tumor promoter) (Westcott et al., 2015; Miller et al., 2003; Stathopoulos et al., 2007; Vreka et al., 2018) or 3-methylcholanthrene followed by butylated hydroxytoluene (MCA/BHT; a two-hit mutagen/tumor promoter regimen) (Malkinson et al., 1997) (Figure 1C, Figure 1—figure supplements 9 and 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 (Nikitin et al., 2004) 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*<sup>Q61R</sup> (Westcott et al., 2015; Vreka et al., 2018; Kanellakis et al., 2019) (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*<sup>Q61R</sup> mutations and give rise to juxtabronchial tumors

We next used digital droplet PCR (ddPCR) to determine the lung lineages that suffer *Kras*<sup>Q61R</sup> driver mutations at early time-points after single urethane hits (Westcott et al., 2015; Vreka et al., 2018; Kanellakis et al., 2019). 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*<sup>mT</sup> was performed one and two weeks later. Interestingly, GFP-labeled cells of both mouse strains had *Kras*<sup>Q61R</sup> mutations at one week post-urethane, but *Kras*<sup>Q61R</sup> 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 ( $\mu$ CT) revealed that most lung tumors were spatially linked with the airways, in accord with pathology results (Figure 2B and C, 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 (Franklin et al., 1997), 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 and 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 and 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 (Safran et al., 2003), a strain emitting light specifically from airway epithelia (Figure 3—figure supplement 5, and Figure 3—figure supplement 5—source data 2). In addition, co-staining of human LUAD



**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;LY2Z.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 ( $\mu$ CT) 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  $\mu$ CT (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 ( $\rho$ ) and probability ( $P$ ) for correlation of  $\mu$ CT and pathology results.

DOI: <https://doi.org/10.7554/eLife.45571.029>

The following source data and figure supplement are available for figure 2:

**Source data 1.** Quantification of  $Kras$  mutant droplets in duplexed digital droplet PCR (ddPCR).

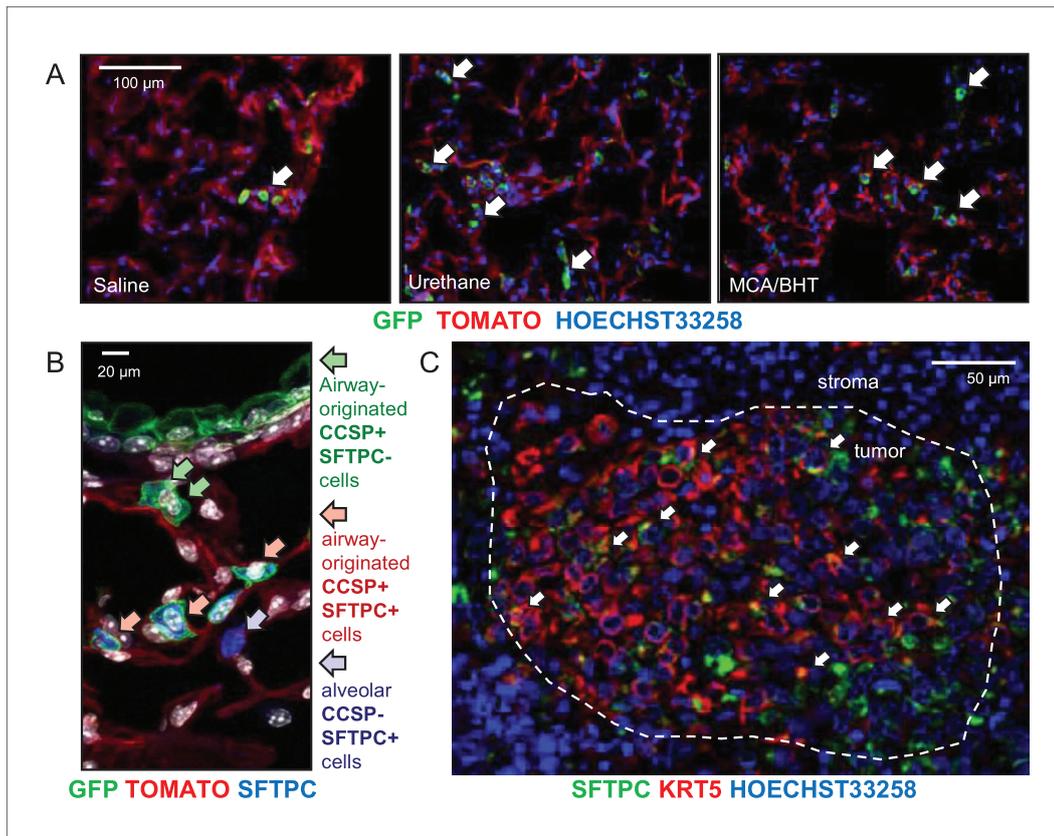
DOI: <https://doi.org/10.7554/eLife.45571.031>

**Source data 2.** Quantification of tumor airway link.

DOI: <https://doi.org/10.7554/eLife.45571.032>

**Figure supplement 1.** Airway cells sustain  $Kras^{Q61R}$  mutations inflicted by urethane.

DOI: <https://doi.org/10.7554/eLife.45571.030>



**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-dihydrobenzo[*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 and 2**. (B) Juxtbronchial 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.

DOI: <https://doi.org/10.7554/eLife.45571.033>

The following source data and figure supplements are available for figure 3:

**Figure supplement 1.** Airway-labeled cells in the alveoli of carcinogen-exposed C57BL/6 mice: representative images.

DOI: <https://doi.org/10.7554/eLife.45571.034>

**Figure supplement 2.** Airway-labeled cells in the alveoli of carcinogen-exposed C57BL/6 mice: data summary.

DOI: <https://doi.org/10.7554/eLife.45571.035>

**Figure supplement 2—source data 1.** Quantification of alveolar GFP+ cells in GFP;CCSP.CRE mice after carcinogen hit.

DOI: <https://doi.org/10.7554/eLife.45571.040>

**Figure supplement 3.** Airway-labeled cells in the alveoli of carcinogen-exposed mice express SFTPC.

DOI: <https://doi.org/10.7554/eLife.45571.036>

**Figure supplement 4.** Airway-labeled cells in environmental-induced lung tumors express SFTPC.

DOI: <https://doi.org/10.7554/eLife.45571.037>

**Figure supplement 5.** In vivo bioluminescent detection of the airway lineage in the lungs of saline- and carcinogen-treated mice.

DOI: <https://doi.org/10.7554/eLife.45571.038>

**Figure supplement 5—source data 2.** Quantification of chest bioluminescence signal in LUC;CCSP.CRE mice after urethane exposure.

DOI: <https://doi.org/10.7554/eLife.45571.041>

**Figure supplement 6.** Human lung adenocarcinomas co-express airway and alveolar markers.

DOI: <https://doi.org/10.7554/eLife.45571.039>

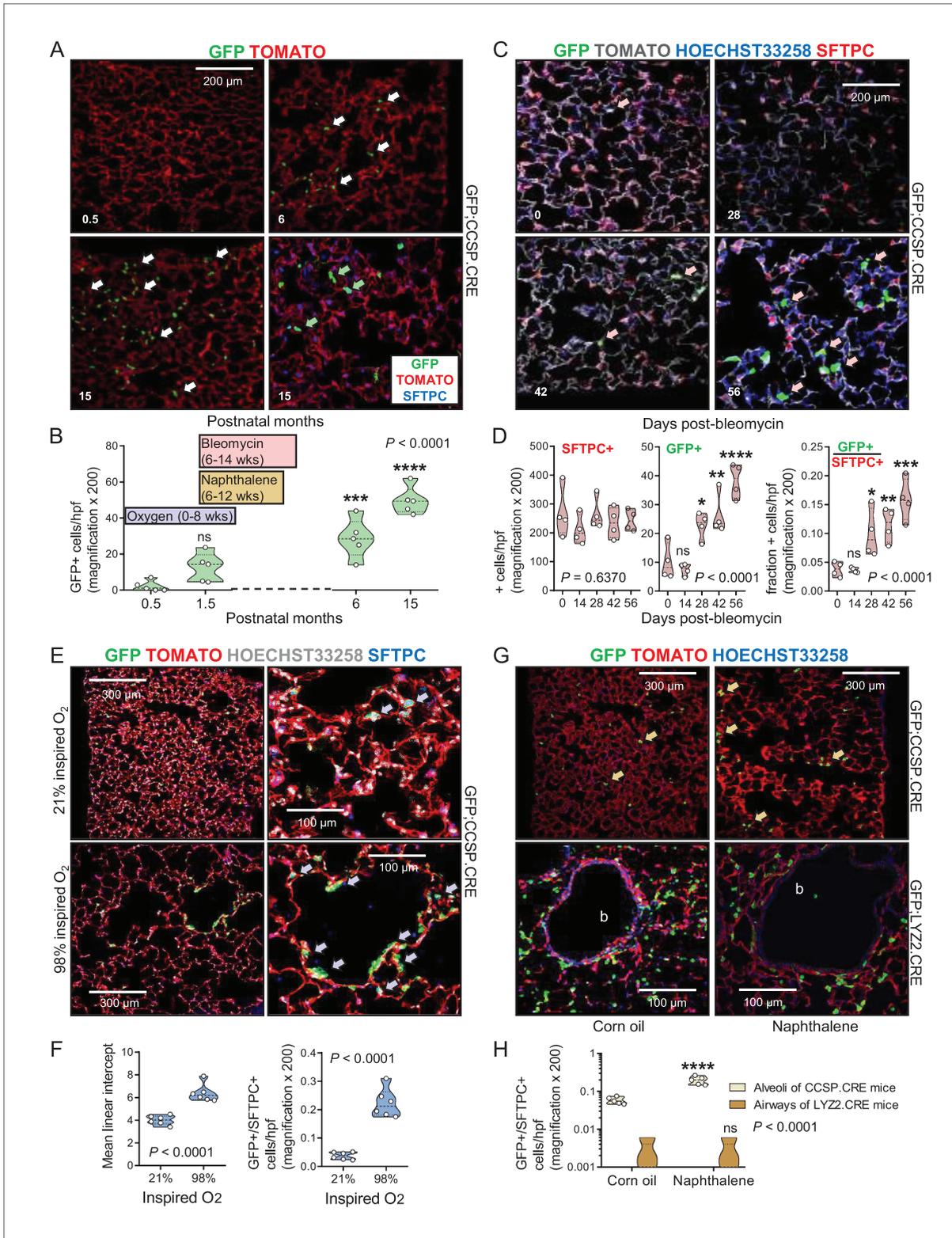
(*Giopanou et al., 2015*) for the alveolar marker SFTPC and the airway markers CCSP and KRT5 showed co-localization of SFTPC with KRT5 but not with CCSP (**Figure 3C** and **Figure 3—figure supplement 6**). These results suggest that airway epithelial cells expand to alveolar regions during field cancerization 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 (*Desai et al., 2014; Aberle et al., 2011; Mason et al., 2000; Lindskog et al., 2014; Sutherland and Berns, 2010*).

### 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 (**Figure 4A and B** and **Figure 4—source data 1**). Bleomycin treatment, which depletes alveolar type II cells (*Lawson et al., 2005*), accelerated the accumulation of GFP-labeled airway cells in the alveoli and in urethane-triggered LUAD (**Figure 4C and D**, **Figure 4—figure supplements 1 and 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 (*Rawlins et al., 2009*), and in the alveoli of GFP;CCSP.CRE mice treated with naphthalene that kills airway epithelial cells (*Sutherland and Berns, 2010; Rawlins et al., 2009*), 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 (**Figure 4E–4H**, **Figure 4—figure supplements 3 and 4**, **Figure 4—source data 3 and 4**, and **Figure 4—figure supplement 4—source data 2**). 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 (**Figure 4G and H**). 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 (*Desai et al., 2014; Rawlins et al., 2009*). 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) (*Voehringer et al., 2008*). 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  $>60$  offsprings 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 and 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 (**Figure 5B and C** and **Figure 5—source data 1**), mimicking human chronic



**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. (B) Data summary ( $n = 5$  mice/time-point) from (A) shown as violin plot. Color-coded boxes indicate time windows of experiments in (C-H).  $P$ , probability, one-way ANOVA. ns, \*\*\*, and \*\*\*\*:  $p > 0.05$ ,  $p < 0.001$ , and  $p < 0.0001$ . *Figure 4 continued on next page*

Figure 4 continued

$p < 0.0001$ , 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). See also **Figure 4—figure supplement 1** and **Figure 4—figure supplement 2**. (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%  $O_2$ . (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** and **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.

DOI: <https://doi.org/10.7554/eLife.45571.042>

The following source data and figure supplements are available for figure 4:

**Source data 1.** Quantification of alveolar GFP+ cells in GFP;CCSP.CRE mice during aging.

DOI: <https://doi.org/10.7554/eLife.45571.047>

**Source data 2.** Quantification of SFTPC+ and GFP+ cells in GFP;CCSP.CRE mice after bleomycin treatment.

DOI: <https://doi.org/10.7554/eLife.45571.048>

**Source data 3.** Data of mean linear intercept and GFP+/SFTPC+ cells in GFP;CCSP.CRE mice after hyperoxia treatment.

DOI: <https://doi.org/10.7554/eLife.45571.049>

**Source data 4.** Data of GFP+/SFTPC+ cells in GFP;CCSP.CRE and GFP;LYZ2.CRE mice after naphthalene treatment.

DOI: <https://doi.org/10.7554/eLife.45571.050>

**Figure supplement 1.** Alveolar type II cell ablation using bleomycin pre-treatment increases airway-labeled cells in urethane-induced lung tumors: representative images.

DOI: <https://doi.org/10.7554/eLife.45571.043>

**Figure supplement 2.** Alveolar type II cell ablation using bleomycin pre-treatment increases airway-labeled cells in urethane-induced lung tumors: data summary.

DOI: <https://doi.org/10.7554/eLife.45571.044>

**Figure supplement 2—source data 1.** Quantification of GFP+ tumors/lung and GFP+ cells/tumor in GFP;CCSP.CRE mice after bleomycin and urethane treatment.

DOI: <https://doi.org/10.7554/eLife.45571.051>

**Figure supplement 3.** Airway epithelial cell ablation using naphthalene is restored by airway-labeled cells: representative images.

DOI: <https://doi.org/10.7554/eLife.45571.045>

**Figure supplement 4.** Airway epithelial cell ablation by naphthalene: data summary.

DOI: <https://doi.org/10.7554/eLife.45571.046>

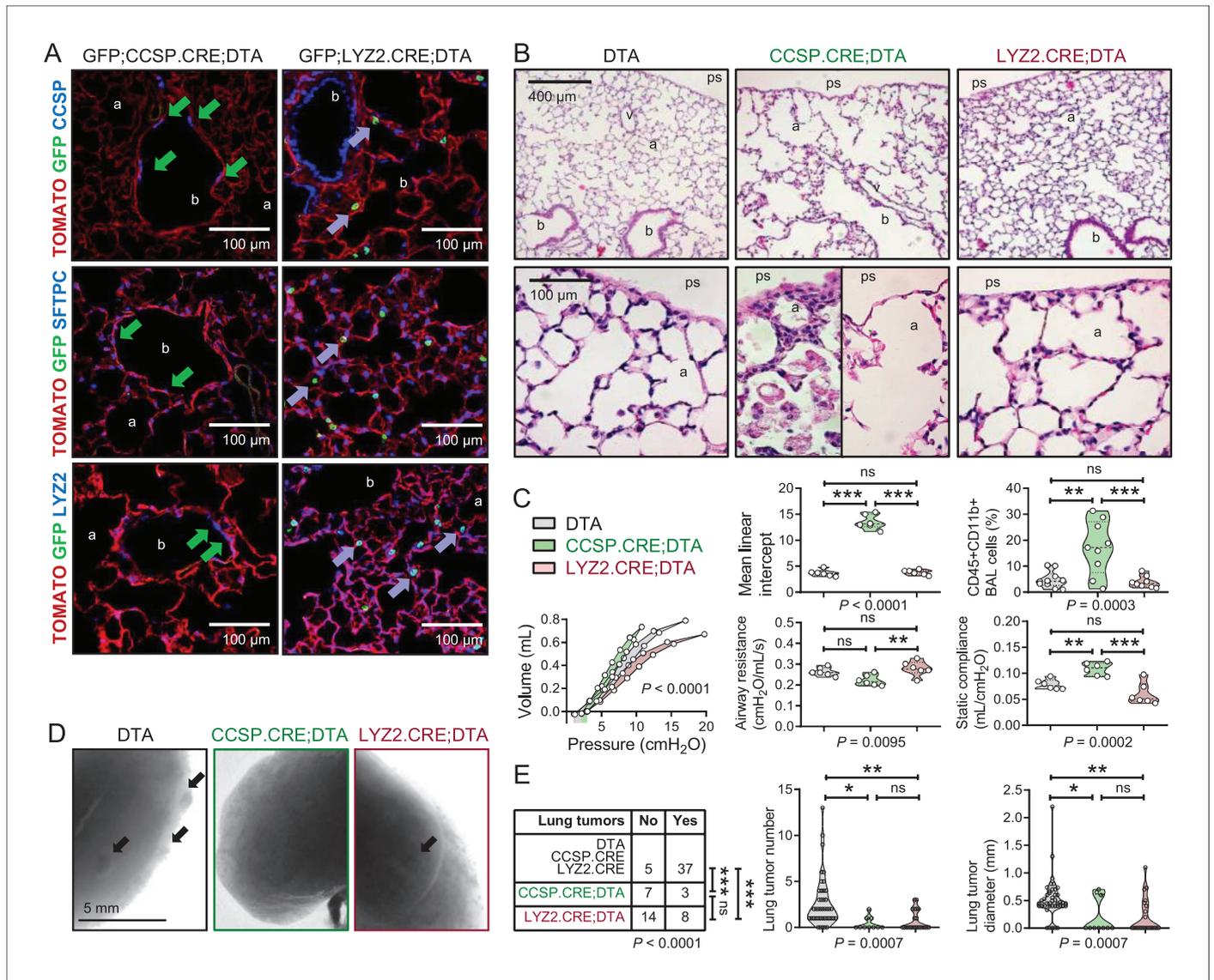
**Figure supplement 4—source data 2.** Quantification of GFP+ airway cells in GFP;CCSP.CRE mice after naphthalene treatment.

DOI: <https://doi.org/10.7554/eLife.45571.052>

obstructive pulmonary disease (Barnes et al., 2015). 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 (Figure 5D and E, 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 (Kanellakis et al., 2019) and of murine lungs with those of murine AEC isolated from tracheal explants, of murine ATII cells (Frank et al., 2016), and of murine bone-marrow-derived macrophages (BMDM). The AEC transcriptome was specifically enriched in LUAD cells compared with whole lungs (Figure 6A and B, 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 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 and 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,  $\chi^2$ -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.

DOI: <https://doi.org/10.7554/eLife.45571.053>

The following source data and figure supplements are available for figure 5:

**Source data 1.** Quantifications of data shown in **Figure 5C**.

DOI: <https://doi.org/10.7554/eLife.45571.056>

**Source data 2.** Quantifications of data shown in **Figure 5D and E**.

DOI: <https://doi.org/10.7554/eLife.45571.057>

Figure 5 continued on next page

Figure 5 continued

**Figure supplement 1.** Triple transgenic mouse models for validation of genetic pulmonary lineage ablation: representative images.

DOI: <https://doi.org/10.7554/eLife.45571.054>

**Figure supplement 2.** Triple transgenic mouse models for validation of genetic pulmonary lineage ablation: data summary.

DOI: <https://doi.org/10.7554/eLife.45571.055>

**Figure supplement 2—source data 1.** Data of GFP+ cells in airways and alveoli of GFP;CCSP.CRE, GFP;CCSP.CRE;DTA, GFP;LYZ2.CRE and GFP;LYZ2.CRE;DTA mice.

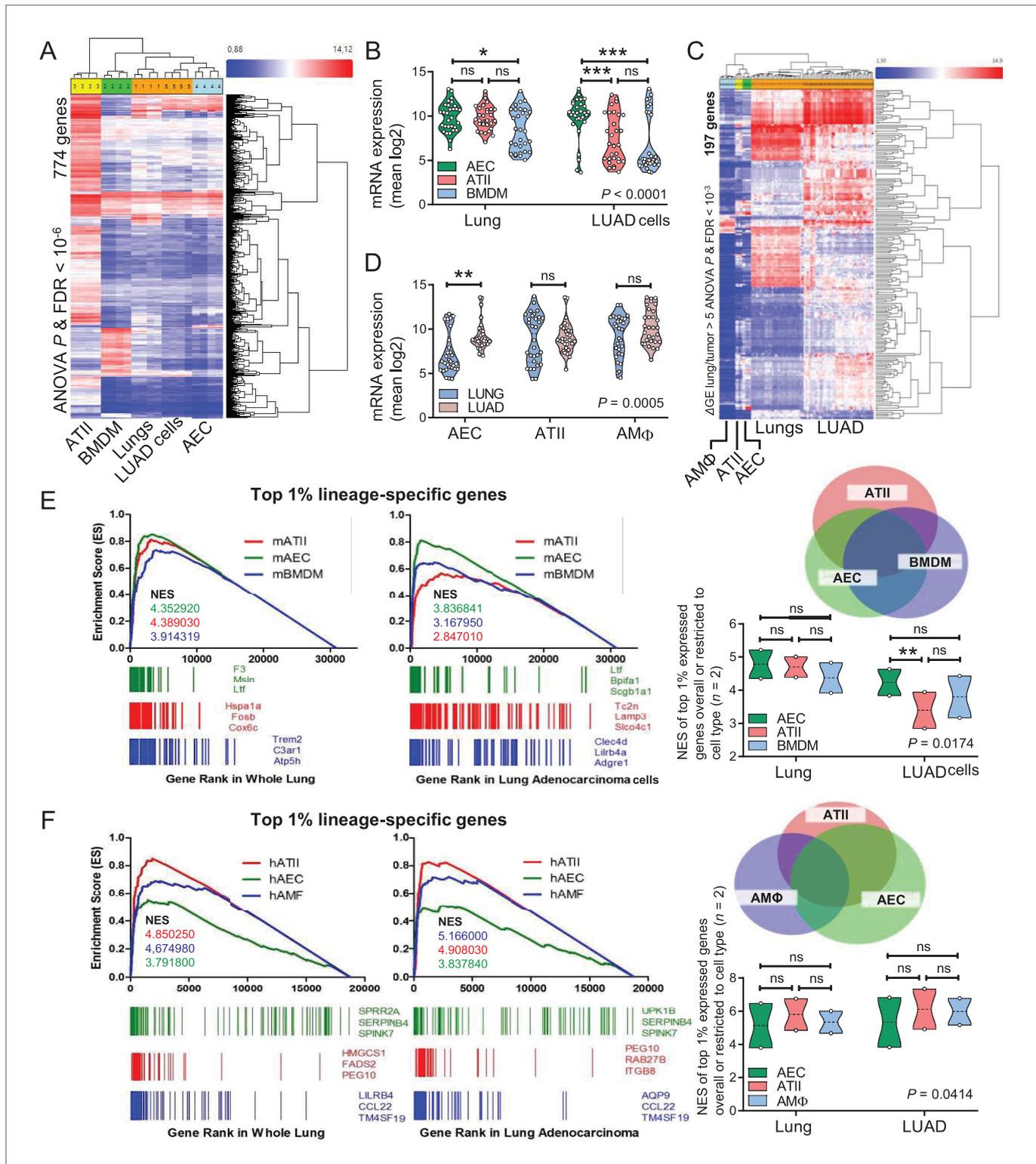
DOI: <https://doi.org/10.7554/eLife.45571.058>

**figure supplement 2—source data 1).** Similar analyses of the transcriptomes of human LUAD and corresponding healthy lungs (Kabbout et al., 2013), and of primary human AEC, ATII, and AM $\Phi$  (Clark et al., 2015; Dancer et al., 2015; Lee et al., 2009) also disclosed that the AEC transcriptome was significantly enriched in LUAD compared with healthy lungs (Figure 6C and D 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/AM $\Phi$  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 (Figure 7A). 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 (Figure 7B) 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 (Hecht, 1999; Campbell et al., 2016; Cancer Genome Atlas Research Network, 2014; Westcott et al., 2015; Miller et al., 2003; Malkinson et al., 1997; Alexandrov et al., 2016; Castelletti et al., 2019). 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*<sup>G12C</sup> or *Kras*<sup>G12D</sup> in the respiratory epithelium (Westcott et al., 2015). Although the latter transgenic tumors have been extensively studied (Kim et al., 2005; Cho et al., 2011; Xu et al., 2012; Sutherland et al., 2014; Mainardi et al., 2014; Desai et al., 2014), 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 (Franklin et al., 1997), 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



**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 **Figure 6 continued on next page**

Figure 6 continued

( $n = 30$ ), primary AEC ( $n = 5$ ), primary ATII ( $n = 4$ ), and alveolar macrophages (AM $\Phi$ ;  $n = 9$ ) was analyzed by Affymetrix Human Gene ST1.0 microarrays. (C) Heat map of genes significantly differentially expressed ( $\Delta GE > 5$  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 AM $\Phi$  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/AM $\Phi$  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/AM $\Phi$ ) signature compared with lung (nominal  $p < 0.0001$  for all, family-wise error rates FWER  $< 0.01$ ). Gene symbols indicate the top three lagging genes from each signature and shows loss of *Scgb1a1* (encoding CCSP) by LUAD. See also **Figure 6—figure supplements 5 and 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.

DOI: <https://doi.org/10.7554/eLife.45571.059>

The following source data and figure supplements are available for figure 6:

**Source data 1.** Cross-examination of signature genes of murine AEC, ATII cells, BMDM, LUAD cells and lungs.

DOI: <https://doi.org/10.7554/eLife.45571.066>

**Source data 2.** Cross-examination of signature genes of human AEC, ATII cells, BMDM, LUAD cells and lungs.

DOI: <https://doi.org/10.7554/eLife.45571.067>

**Source data 3.** Quantification of gene set enrichment analyses data shown in **Figure 6E**.

DOI: <https://doi.org/10.7554/eLife.45571.068>

**Source data 4.** Quantification of gene set enrichment analyses data shown in **Figure 6F**.

DOI: <https://doi.org/10.7554/eLife.45571.069>

**Figure supplement 1.** Lineage-specific gene expression in mouse lung adenocarcinoma cell lines induced by urethane compared with mouse lungs.

DOI: <https://doi.org/10.7554/eLife.45571.060>

**Figure supplement 2.** Loss of lineage marker expression in mouse lung adenocarcinoma cell lines induced by urethane.

DOI: <https://doi.org/10.7554/eLife.45571.061>

**Figure supplement 2—source data 1.** Quantification of gene expression levels of data shown in **Figure 6—figure supplement 2**.

DOI: <https://doi.org/10.7554/eLife.45571.070>

**Figure supplement 3.** Loss of lineage marker expression in mouse lung adenocarcinoma cell lines induced by urethane compared with mouse lungs: heat maps.

DOI: <https://doi.org/10.7554/eLife.45571.062>

**Figure supplement 4.** Loss of lineage marker expression in mouse lung adenocarcinoma cell lines induced by urethane compared with mouse lungs: volcano plot.

DOI: <https://doi.org/10.7554/eLife.45571.063>

**Figure supplement 5.** Mouse gene set enrichment analyses.

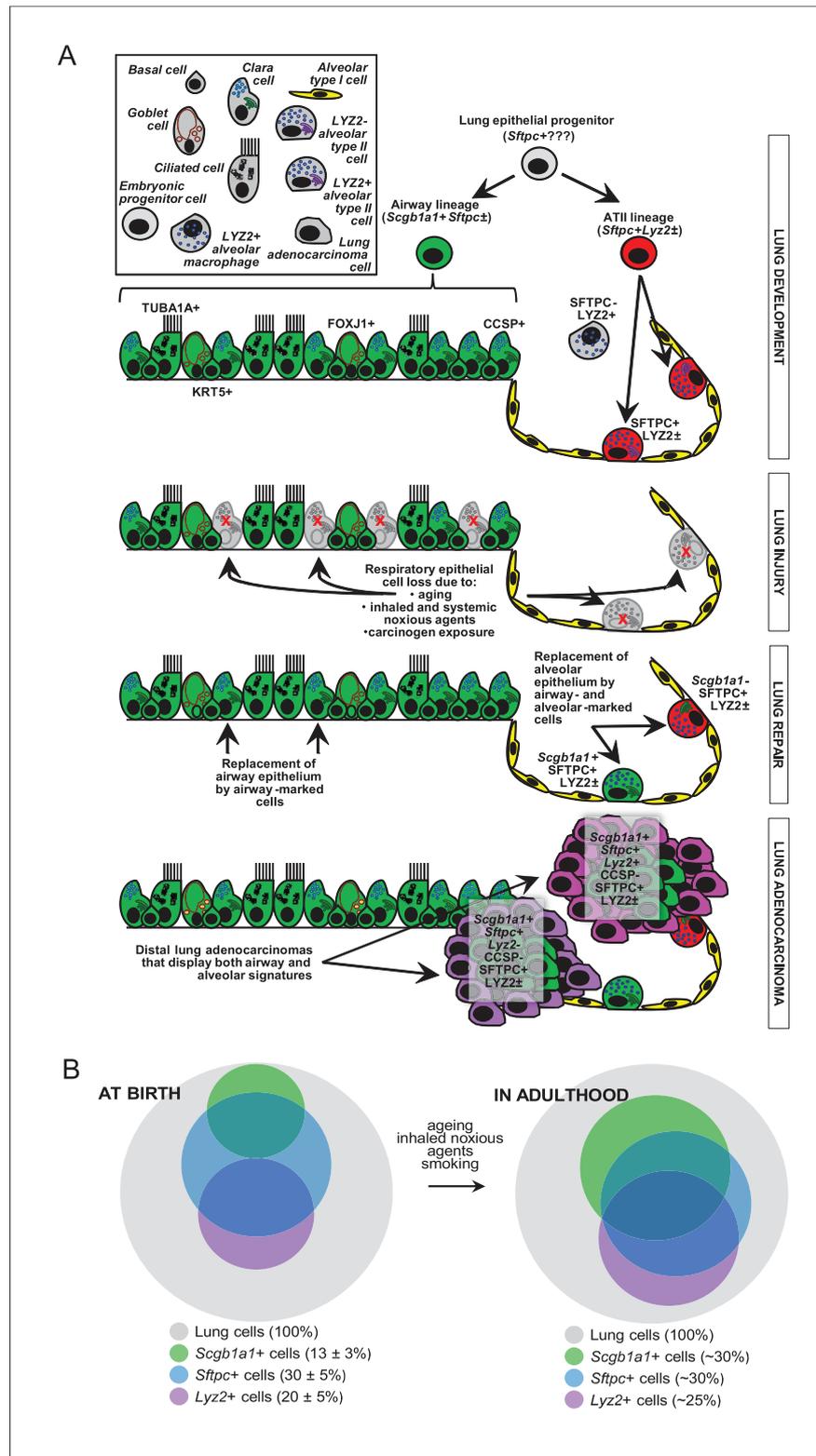
DOI: <https://doi.org/10.7554/eLife.45571.064>

**Figure supplement 6.** Human gene set enrichment analyses.

DOI: <https://doi.org/10.7554/eLife.45571.065>

tumors (*Kras*<sup>Q61R</sup>) compared to *KRAS*<sup>G12C/D</sup> 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 (**Westcott et al., 2015**). Our findings of airway epithelial cells being more sensitive than alveolar type II cells to *Kras*<sup>Q61R</sup> 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 LY2Z 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 (LY2Z+) alveolar transcriptomes; ii) alveolar cells transit through an obligate CCSP+ with or without a dispensable LY2Z+ 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 (**Desai et al., 2014; Frank et al., 2016; Xu et al., 2016**). 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  $\mu$ CT 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



**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 (ATI) 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, *Figure 7 continued on next page*

Figure 7 continued

ciliated, goblet, basal, and other cells, while the developmental ATII lineage (*Sftpc*<sup>+</sup> *Lyz2*<sup>±</sup>; 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*<sup>+</sup> *Sftpc*<sup>±</sup> marked cells ensues. Upon lung injury, this process is accelerated. Similarly, during carcinogenesis caused by chemical tobacco smoke carcinogens, *Scgb1a1*<sup>+</sup> *Sftpc*<sup>±</sup> 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*<sup>+</sup> 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  $\alpha$ -tubulin.

DOI: <https://doi.org/10.7554/eLife.45571.071>

sensitize human airway epithelial cells to a single *KRAS* mutation (Vaz et al., 2017). 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 (Desai et al., 2014).

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 (Nagel et al., 2016). To this end, insertions and deletions in lineage-restricted genes were recently shown to occur in human lung adenocarcinoma (Imielinski et al., 2017). Moreover, integrin  $\beta_3$  and TANK-binding kinase one partner with oncogenic *KRAS* signaling to mediate cancer stemness and drug resistance (Barbie et al., 2009; Seguin et al., 2014). 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.

## Materials and methods

### Key resources table

Reagent type (species) or resource	Designation	Source or reference	Identifiers	Additional information
Strain, strain background ( <i>Mus musculus</i> )	C57BL/6	Jackson Laboratory	Stock #: 000664; RRID:IMSR_JAX:000664	

Continued on next page

Continued

Reagent type (species) or resource	Designation	Source or reference	Identifiers	Additional information
Strain, strain background ( <i>M. musculus</i> )	FVB	Jackson Laboratory	Stock #: 001800; RRID:IMSR_JAX:001800	
Genetic reagent ( <i>M. musculus</i> )	TOMATO	Jackson Laboratory	Stock #: 007676; RRID:IMSR_JAX:007676	<b>Muzumdar et al., 2007</b>
Genetic reagent ( <i>M. musculus</i> )	LUC	Jackson Laboratory	Stock #: 005125; RRID:IMSR_JAX:005125	<b>Safran et al., 2003</b>
Genetic reagent ( <i>M. musculus</i> )	DTA	Jackson Laboratory	Stock #: 009669; RRID:IMSR_JAX:009669	<b>Voehringer et al., 2008</b>
Genetic reagent ( <i>M. musculus</i> )	LYZ2.Cre	Jackson Laboratory	Stock #: 004781; RRID:IMSR_JAX:004781	PMID: 10621974
Genetic reagent ( <i>M. musculus</i> )	SOX2.Cre	Jackson Laboratory	Stock #: 008454; RRID:IMSR_JAX:008454	<b>Hayashi et al., 2002</b>
Genetic reagent ( <i>M. musculus</i> )	VAV.Cre	Jackson Laboratory	Stock #: 008610; RRID:IMSR_JAX:008610	<b>Ogilvy et al., 1998</b>
Genetic reagent ( <i>M. musculus</i> )	NES.Cre	Jackson Laboratory	Stock #: 003771; RRID:IMSR_JAX:003771	<b>Tronche et al., 1999</b>
Genetic reagent ( <i>M. musculus</i> )	CCSP.Cre	European Mouse Mutant Archive	Stock #: EM:04965; RRID:IMSR_M231009	<b>Oikonomou et al., 2012</b>
Genetic reagent ( <i>M. musculus</i> )	SFTPC.Cre	Mouse Genome Informatics	RRID:MGI:3574949	<b>Okubo et al., 2005</b>
Cell line ( <i>M. musculus</i> )	LUAD cells	PMID: 30828726		Derived from urethane models
Biological sample ( <i>Homo sapiens</i> )	Lung adeno carcinomas	<b>Giopanou et al., 2015</b>		Archival samples of patients with LUAD
Antibody	rabbit polyclonal 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 $\alpha$ -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 $\mu$ g)
Antibody	mouse monoclonal anti-CD11b PE conjugated	eBioscience	Cat. #: 12-0112-82; RRID:AB_2734869	FC (0,05 $\mu$ g)

Continued on next page

Continued

Reagent type (species) or resource	Designation	Source or reference	Identifiers	Additional information
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 <sup>Q61R</sup> mutation detection	Forward: ATCTGACGTGCTTTGCCTGT, Reverse: CCCTCCCCAGTTCTCATGTA
Sequence-based reagent	Digital droplet PCR probe	This paper	Kras <sup>Q61R</sup> mutation detection	sequence: GACACAGCAGGT CAAGAGGAGTACA
Sequence-based reagent	Digital droplet PCR primers and probe	Bio-Rad Laboratories	Registration #: dCNS5685684912	Tomato allele detection
Sequence-based reagent	Quantitative PCR	This paper	<i>Scgb1a1</i> gene	Forward: ATCACTGTGGTCATGCTGTCC, Reverse: GCTTCAGGGATGCCACATAAC
Sequence-based reagent	Quantitative PCR	This paper	<i>Sftpc</i> gene	Forward: TCGTTGTCGTGGTGATTGTAG, Reverse: TCGTTGTCGTGGTGATTGTAG
Sequence-based reagent	Quantitative PCR	This paper	<i>Gusb</i> gene	Forward: TTACTTTAAGACGCTGATCACC, Reverse: ACCTCCAAATGCCCATAGTC
Commercial assay or kit	GenElute Mammalian Genomic DNA Minipreps Kit	Sigma-Aldrich	Cat. #: G1N70	
Commercial assay or kit	RNeasy Mini Kit	Qiagen	Cat. #: 74106	
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-methyl cholanthrene (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

Continued on next page

Continued

Reagent type (species) or resource	Designation	Source or reference	Identifiers	Additional information
Chemical compound, drug	Bleomycin A2	Calbiochem	Cat. #: 203401	0.08 units
Software, algorithm	Transcriptome Analysis Console Software	<a href="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">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</a>	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	<a href="http://software.broadinstitute.org/gsea/index.jsp">http://software.broadinstitute.org/gsea/index.jsp</a>		<b>Subramanian et al., 2005</b>
Software, algorithm	NRECON software	Bruker		
Software, algorithm	CT analysis (Ctan) software	Bruker		
Software, algorithm	CTVox software	Bruker		
Software, algorithm	QuantaSoft	Bio-Rad Laboratories ( <a href="http://www.bio-rad.com/en-gr/sku/1864011-quantasoft-software-regulatory-edition?ID=1864011">http://www.bio-rad.com/en-gr/sku/1864011-quantasoft-software-regulatory-edition?ID=1864011</a> )		
Software, algorithm	G*power	<a href="http://www.gpower.hhu.de/">http://www.gpower.hhu.de/</a>	RRID:SCR_013726	<b>Faul et al., 2007</b>
Software, algorithm	GraphPad Prism	<a href="http://www.graphpad.com/">http://www.graphpad.com/</a>	RRID:SCR_002798	Version 8
Software, algorithm	Fiji	<a href="http://fiji.sc">http://fiji.sc</a>	RRID:SCR_002285	PMID: 22743772
Software, algorithm	Living Image software	Perkin-Elmer ( <a href="http://www.perkinelmer.com/catalog/category/id/living%20image%20software">http://www.perkinelmer.com/catalog/category/id/living%20image%20software</a> )	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	<i>M. musculus</i> ATII cells; <i>H. sapiens</i> AEC cells; <i>H. sapiens</i> ATII cells; <i>H. sapiens</i> AMΦ; <i>H. sapiens</i> non-smokers lung and LUAD

Continued on next page

Continued

Reagent type (species) or resource	Designation	Source or reference	Identifiers	Additional information
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 (*Giopanou et al., 2015*) 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-dihydrobenzo[*a*]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<sup>tm4(ACTB-tdTomato,-EGFP)<sup>Luo</sup>/J</sup> [mT/mG; TOMATO; #007676; (*Muzumdar et al., 2007*)], FVB.129S6(B6)-Gt(ROSA)26Sor<sup>tm1(Luc)<sup>Kael</sup>/J</sup> [LUC; #005125; (*Safran et al., 2003*)], B6.129P2-Gt(ROSA)26Sor<sup>tm1(DTA)<sup>Lky</sup>/J</sup> [DTA; #009669; (*Voehringer et al., 2008*)], B6.129P2-Lyz2<sup>tm1(cre)<sup>lfe</sup>/J</sup> [LYZ2.CRE; #004781; (*Desai et al., 2014*)], B6.Cg-Tg(Sox2-cre)1Amc/J [SOX2.CRE; #008454; (*Hayashi et al., 2002*)], B6.Cg-Tg(Vav1-cre)A2Kio/J [VAV.CRE; #008610; (*Ogilvy et al., 1998*)], and B6.Cg-Tg(Nes-cre)1Kln/J [NES.CRE;

#003771; (Tronche et al., 1999)] 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 (Oikonomou et al., 2012) and *Tg(Sftpc-cre)1Blh* (SFTPC.CRE; Mouse Genome Informatics #MGI:3574949) mice were donated by their founder (Okubo et al., 2005). 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  $\mu$ L saline) and were sacrificed 6–7 months after the first injection, or four consecutive weekly intraperitoneal MCA (15 mg/Kg in 100  $\mu$ L saline) followed by eight consecutive weekly intraperitoneal BHT injections (200 mg/Kg in 100  $\mu$ L 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  $\mu$ L saline) and were sacrificed 6–7 months later (Westcott et al., 2015; Miller et al., 2003; Malkinson et al., 1997; Stathopoulos et al., 2007; Vreka et al., 2018).

### Mouse models of lung injury

Six-week-old mice (C57BL/6 background) received intratracheal bleomycin A2 (0.08 units in 50  $\mu$ L saline) or intraperitoneal naphthalene (250 mg/Kg in 100  $\mu$ L corn oil) (Lawson et al., 2005; Rawlins et al., 2009). 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 (Rawlins et al., 2009; Yee et al., 2009). 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  $\mu$ m filter before passage into a sealed Lexan polycarbonate chamber measuring 40  $\times$  25  $\times$  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 hr 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°C in 5% CO<sub>2</sub>–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 (Kanellakis et al., 2019). 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<sub>2</sub>. 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 (Giopanou et al., 2015). 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  $\mu$ CT 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 hr incubation in Hexamethyldisilazane (HMDS; Sigma, St. Louis, MO) and 2 hr air-drying. The dehydrated lungs were then scanned in a Bruker SkyScan 1172 scanner at 41kV without filtration and with 5.94  $\mu$ m 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 5 s camera (Zeiss, Jena, Germany) in trans-illumination mode, allowing for visualization of both superficial and deeply-located lung tumors (Stathopoulos *et al.*, 2007; Vreka *et al.*, 2018). Tumor location was charted and diameter ( $\delta$ ) 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  $\mu\text{m}$ -thick lung sections ( $n = 10/\text{lung}$ ), 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 (Hsia *et al.*, 2010). 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 (Hsia *et al.*, 2010). 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:  $\Sigma\{2 \times (\text{length of line}/\text{number of intercepts})\}/\text{total number of lines}$ . 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  $\text{cmH}_2\text{O}$  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 (Hsia *et al.*, 2010). Subsequently, lungs were fixed with 10% formalin overnight and were embedded in paraffin. Five- $\mu\text{m}$ -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  $-80^\circ\text{C}$ . Ten- $\mu\text{m}$  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 hr. 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  $\alpha$ -tubulin (1:2000 dilution, T7451, Sigma-Aldrich, St. Louis, 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 5 s 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 (*Manali et al., 2011*), 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<sub>2</sub>O, 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<sub>2</sub>O pressure. Measurements were repeated thrice at 60 s intervals and were averaged. Data were fit into the constant phase model in order to fractionate total respiratory input impedance into airways resistance ( $R_{aw}$ ) 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 ( $C_{st}$ ), 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/ $\mu$ L)/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) (*Mazaika and Homsy, 2014*). 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: ATC TGACGTGCTTTGCCTGT,  $Kras^{Q61R}$  reverse: CCCTCCCAGTTCTCATGTA, and  $Kras^{Q61R}$  probe: GACACAGCAGGTCAAGAGGAGTACA. The  $Rosa^{mT}$  assay is registered as dCNS685684912 (BioRad) with MIQE context: seq1:195–315:+CCAGTTCATGTACGGCTCCAAGGCGTACGTGAAG-CACCCGCGGACATCCCCGATTACAAGAAGCTGTCTTCCCCGAGGGCTTCAAGTGGGAGCGCG

TGATGAACTTCGAGGACGGCGGTCT. Primers and fluorescently labeled probes were combined in a mixture containing 18  $\mu$ M forward and reverse primers and 5  $\mu$ M labeled probes (20x primer/Taqman probe mix). Reactions were assembled to contain 12.5  $\mu$ L 2x ddPCR mix no-UTP, 1.25  $\mu$ L 20x *Kras*<sup>O61R</sup> primer/Taqman probe Mix, 1.25  $\mu$ L 20x *Rosa*<sup>mT</sup> custom primer/Taqman probe Mix and 10  $\mu$ L 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 s and 40 cycles of annealing at 62.5°C for 60 s, 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*<sup>mT</sup> and at 10000 for the *Kras*<sup>O61R</sup> probe, according to the manufacturer's instructions.

### Bronchoalveolar lavage (BAL)

BAL was performed using three sequential aliquots of 1000  $\mu$ L 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 (Stathopoulos et al., 2007; Vreka et al., 2018).

### Bioluminescence imaging

LUC;CCSP.CRE mice, bioluminescent reporters of CCSP-labeled cell mass, received one intraperitoneal injection of saline (100  $\mu$ L saline) or urethane (1 g/Kg in 100  $\mu$ L 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  $\mu$ L of sterile water to the retro-orbital vein, and data were analyzed using Living Image v.4.2 (Perkin-Elmer, Waltham, MA) (Stathopoulos et al., 2007; Vreka et al., 2018).

### qPCR and microarrays

Triplicate cultures of 10<sup>6</sup> 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  $\mu$ g) 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 (Frank et al., 2016; Kabbout et al., 2013; Clark et al., 2015; Dancer et al., 2015; Lee et al., 2009) murine A11 and human AEC, A11, AM $\Phi$ , 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<sup>- $\Delta$ CT</sup> method relative to *Gusb*.

### Flow cytometry

BAL cells were suspended in 50  $\mu$ L PBS with 2% FBS and 0.1% NaN<sub>3</sub>, 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  $\mu$ L 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 (Kanellakis et al., 2019). 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  $\mu$ g/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>) (Subramanian *et al.*, 2005). In detail, genes significantly expressed ( $\log_2$  normalized expression  $>8$ ) in murine tracheal airway cells, ATII cells (Frank *et al.*, 2016), and BMDM were cross-examined against the murine lung and chemical-induced LUAD cell line transcriptomes. In addition, previously reported human AEC, ATII, and AM $\Phi$  cellular signatures (Clark *et al.*, 2015; Dancer *et al.*, 2015; Lee *et al.*, 2009) were cross-examined against the previously described transcriptomes of human normal lung tissue from never-smokers and of LUAD (Kabbout *et al.*, 2013).

### 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  $d = 1.5$  (Faul *et al.*, 2007). 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  $\pm$  SD, and sample size ( $n$ ) always refers to biological and not technical replicates. Differences in frequency were examined by Fischer's exact and  $\chi^2$  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).

### 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, AM $\Phi$ , non-smokers lung, and LUAD microarray data are available at GEO using Accession IDs GSE82154, GSE55459, GSE46749, GSE18816, and GSE43458).

### Acknowledgements

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  $\mu$ CT facility.

### Additional information

#### Funding

Funder	Grant reference number	Author
H2020 European Research Council	260524	Georgios T Stathopoulos
Hellenic State Scholarships Foundation	Post-doctoral Research Fellowship	Magda Spella
Howard Hughes Medical Institute	International Research Scholars Award	Rocio Sotillo
German Center for Lung Research		Kristina AM Arendt Laura V Klotz Georgios T Stathopoulos

Hellenic Thoracic Society	PhD Fellowship	Malamati Vreka Anastasios D Giannou
H2020 European Research Council	281614	Rocio Sotillo
H2020 European Research Council	679345	Georgios T Stathopoulos

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

### Author contributions

Magda Spella, Conceptualization, Data curation, Formal analysis, Methodology, Writing—original draft; Ioannis Lilis, Mario AA Pepe, Yuanyuan Chen, Maria Armaka, Anne-Sophie Lamort, Dimitra E Zazara, Fani Roumelioti, Malamati Vreka, Nikolaos I Kanellakis, Darcy E Wagner, Anastasios D Giannou, Kristina AM Arendt, Dimitrios Toumpanakis, Vassiliki Karavana, Spyros G Zakynthinos, Ioanna Giopanou, Antonia Marazioti, Data curation, Formal analysis; Vasileios Armenis, Laura V Klotz, Data curation; Vassilis Aidinis, Resources; Rocio Sotillo, Resources, Formal analysis, Supervision, Funding acquisition, Project administration; Georgios T Stathopoulos, Conceptualization, Resources, Data curation, Formal analysis, Supervision, Funding acquisition, Validation, Investigation, Visualization, Methodology, Writing—original draft, Project administration, Writing—review and editing

### Author ORCIDs

Magda Spella  <https://orcid.org/0000-0003-2505-7778>

Rocio Sotillo  <https://orcid.org/0000-0002-0855-7917>

Georgios T Stathopoulos  <https://orcid.org/0000-0002-9215-6461>

### Ethics

Human subjects: Archival tissue samples of patients with lung adenocarcinoma were used in this study. The observational protocol for the original 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.

Animal experimentation: 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>).

### Decision letter and Author response

Decision letter <https://doi.org/10.7554/eLife.45571.087>

Author response <https://doi.org/10.7554/eLife.45571.088>

## Additional files

### Supplementary files

- Transparent reporting form

DOI: <https://doi.org/10.7554/eLife.45571.072>

### 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 (Frank et al., 2016; Clark et al., 2015; Dancer et al., 2015; Lee et al., 2009; Kabbout et al., 2013) 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).

The following dataset was generated:

Author(s)	Year	Dataset title	Dataset URL	Database and Identifier
Stathopoulos GT	2017	Epithelial signatures of chemical-induced lung adenocarcinoma	<a href="https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE94981">https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE94981</a>	Gene Expression Omnibus , GSE94981

The following previously published datasets were used:

Author(s)	Year	Dataset title	Dataset URL	Database and Identifier
Frank DB, Peng T, Zepp JA, Snitow M, Vincent TL, Penkala IJ, Cui Z, Herriges MJ, Morley MP, Zhou S, Lu MM, Morrissey EE	2016	Emergence of a Wave of Wnt Signaling that Regulates Lung Alveologenesis by Controlling Epithelial Self-Renewal and Differentiation.	<a href="https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE82154">https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE82154</a>	NCBI Gene Expression Omnibus, GSE82154
Clark JG, Kim KH, Basom RS, Gharib SA	2015	Plasticity of airway epithelial cell transcriptome in response to flagellin.	<a href="https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE55459">https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE55459</a>	NCBI Gene Expression Omnibus, GSE55459
Dancer RC, Parekh D, Lax S, D'Souza V, Zheng S, Bassford CR, Park D, Bartis DG, Mahida R, Turner AM, Sappey E, Wei W, Naidu B, Stewart PM, Fraser WD, Christopher KB, Cooper MS, Gao F, Sansom DM, Martineau AR, Perkins GD, Thickett DR	2015	Vitamin D deficiency contributes directly to the acute respiratory distress syndrome (ARDS).	<a href="https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE46749">https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE46749</a>	NCBI Gene Expression Omnibus, GSE46749
Lee SM, Gardy JL, Cheung CY, Cheung TK, Hui KP, Ip NY, Guan Y, Hancock RE, Peiris JS	2009	Systems-level comparison of host-responses elicited by avian H5N1 and seasonal H1N1 influenza viruses in primary human macrophages	<a href="https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE18816">https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE18816</a>	NCBI Gene Expression Omnibus, GSE18816
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	2013	ETS2 mediated tumor suppressive function and MET oncogene inhibition in human non-small cell lung cancer.	<a href="https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE43458">https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE43458</a>	NCBI Gene Expression Omnibus, GSE43458

## References

- Aberle DR**, Adams AM, Berg CD, Black WC, Clapp JD, Fagerstrom RM, Gareen IF, Gatsonis C, Marcus PM, Sicks JD, National Lung Screening Trial Research Team. 2011. Reduced lung-cancer mortality with low-dose computed tomographic screening. *The New England Journal of Medicine* **365**:395–409. DOI: <https://doi.org/10.1056/NEJMoa1102873>, PMID: 21714641
- 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. 2016. Mutational signatures associated with tobacco smoking in human cancer. *Science* **354**:618–622. DOI: <https://doi.org/10.1126/science.aag0299>, PMID: 27811275
- 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, et al. 2009. Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. *Nature* **462**:108–112. DOI: <https://doi.org/10.1038/nature08460>, PMID: 19847166
- Barnes PJ**, Burney PG, Silverman EK, Celli BR, Vestbo J, Wedzicha JA, Wouters EF. 2015. Chronic obstructive pulmonary disease. *Nature Reviews Disease Primers* **1**:15076. DOI: <https://doi.org/10.1038/nrdp.2015.76>, PMID: 27189863

- 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, et al. 2016. Distinct patterns of somatic genome alterations in lung adenocarcinomas and squamous cell carcinomas. *Nature Genetics* **48**:607–616. DOI: <https://doi.org/10.1038/ng.3564>, PMID: 27158780
- Cancer Genome Atlas Research Network**. 2014. Comprehensive molecular profiling of lung adenocarcinoma. *Nature* **511**:543–550. DOI: <https://doi.org/10.1038/nature13385>, PMID: 25079552
- Castelletti N**, Kaiser JC, Simonetto C, Furukawa K, Küchenhoff H, Stathopoulos GT. 2019. Risk of lung adenocarcinoma from smoking and radiation arises in distinct molecular pathways. *Carcinogenesis* **3**:pii: bgz036. DOI: <https://doi.org/10.1093/carcin/bgz036>
- Cho HC**, Lai CY, Shao LE, Yu J. 2011. Identification of tumorigenic cells in kras(G12D)-induced lung adenocarcinoma. *Cancer Research* **71**:7250–7258. DOI: <https://doi.org/10.1158/0008-5472.CAN-11-0903>, PMID: 22088965
- Clark JG**, Kim KH, Basom RS, Gharib SA. 2015. Plasticity of airway epithelial cell transcriptome in response to flagellin. *PLoS ONE* **10**:e0115486. DOI: <https://doi.org/10.1371/journal.pone.0115486>, PMID: 25668187
- Dancer RC**, Parekh D, Lax S, D'Souza V, Zheng S, Bassford CR, Park D, Bartis DG, Mahida R, Turner AM, Sapay E, Wei W, Naidu B, Stewart PM, Fraser WD, Christopher KB, Cooper MS, Gao F, Sansom DM, Martineau AR, et al. 2015. Vitamin D deficiency contributes directly to the acute respiratory distress syndrome (ARDS). *Thorax* **70**:617–624. DOI: <https://doi.org/10.1136/thoraxjnl-2014-206680>, PMID: 25903964
- Desai TJ**, Brownfield DG, Krasnow MA. 2014. Alveolar progenitor and stem cells in lung development, renewal and cancer. *Nature* **507**:190–194. DOI: <https://doi.org/10.1038/nature12930>, PMID: 24499815
- Faul F**, Erdfelder E, Lang AG, Buchner A. 2007. G\*power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods* **39**:175–191. DOI: <https://doi.org/10.3758/BF03193146>, PMID: 17695343
- 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. 2011. COSMIC: mining complete cancer genomes in the catalogue of somatic mutations in cancer. *Nucleic Acids Research* **39**:D945–D950. DOI: <https://doi.org/10.1093/nar/gkq929>, PMID: 20952405
- Frank DB**, Peng T, Zepp JA, Snitow M, Vincent TL, Penkala IJ, Cui Z, Herriges MJ, Morley MP, Zhou S, Lu MM, Morrissey EE. 2016. Emergence of a wave of wnt signaling that regulates lung alveologenesis by controlling epithelial Self-Renewal and differentiation. *Cell Reports* **17**:2312–2325. DOI: <https://doi.org/10.1016/j.celrep.2016.11.001>, PMID: 27880906
- Franklin WA**, Gazdar AF, Haney J, Wistuba II, La Rosa FG, Kennedy T, Ritchey DM, Miller YE. 1997. Widely dispersed p53 mutation in respiratory epithelium. A novel mechanism for field carcinogenesis. *Journal of Clinical Investigation* **100**:2133–2137. DOI: <https://doi.org/10.1172/JCI119748>, PMID: 9329980
- Giopanou I**, Lilis I, Papaleonidopoulos V, Marazioti A, Spella M, Vreka M, Papadaki H, Stathopoulos GT. 2015. Comprehensive evaluation of nuclear Factor- $\kappa$ B expression patterns in Non-Small cell lung cancer. *PLoS ONE* **10**:e0132527. DOI: <https://doi.org/10.1371/journal.pone.0132527>, PMID: 26147201
- Hayashi S**, Lewis P, Pevny L, McMahon AP. 2002. Efficient gene modulation in mouse epiblast using a Sox2Cre transgenic mouse strain. *Mechanisms of Development* **119**:S97–S101. DOI: [https://doi.org/10.1016/S0925-4773\(03\)00099-6](https://doi.org/10.1016/S0925-4773(03)00099-6), PMID: 14516668
- Hecht SS**. 1999. Tobacco smoke carcinogens and lung cancer. *JNCI Journal of the National Cancer Institute* **91**: 1194–1210. DOI: <https://doi.org/10.1093/jnci/91.14.1194>, PMID: 10413421
- Hsia CC**, Hyde DM, Ochs M, Weibel ER. 2010. 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. *American Journal of Respiratory and Critical Care Medicine* **181**:394–418. DOI: <https://doi.org/10.1164/rccm.200809-1522st>
- Imielinski M**, Guo G, Meyerson M. 2017. Insertions and deletions target Lineage-Defining genes in human cancers. *Cell* **168**:460–472. DOI: <https://doi.org/10.1016/j.cell.2016.12.025>, PMID: 28089356
- 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. 2013. ETS2 mediated tumor suppressive function and MET oncogene inhibition in human non-small cell lung cancer. *Clinical Cancer Research* **19**:3383–3395. DOI: <https://doi.org/10.1158/1078-0432.CCR-13-0341>, PMID: 23659968
- Kanellakis NI**, Giannou AD, Pepe MAA, Agaloti T, Zazara DE, Giopanou I, Psallidas I, Spella M, Marazioti A, Arendt KAM, Lamort AS, Champeris Tsaniras S, Taraviras S, Papadaki H, Lilis I, Stathopoulos GT. 2019. Tobacco chemical-induced mouse lung adenocarcinoma cell lines pin the prolactin orthologue proliferin as a lung tumour promoter. *Carcinogenesis* **25**:pii: bgz047. DOI: <https://doi.org/10.1093/carcin/bgz047>
- Kim CF**, Jackson EL, Woolfenden AE, Lawrence S, Babar I, Vogel S, Crowley D, Bronson RT, Jacks T. 2005. Identification of bronchioalveolar stem cells in normal lung and lung cancer. *Cell* **121**:823–835. DOI: <https://doi.org/10.1016/j.cell.2005.03.032>, PMID: 15960971
- Lawson WE**, Polosukhin VV, Zoia O, Stathopoulos GT, Han W, Plieth D, Loyd JE, Neilson EG, Blackwell TS. 2005. Characterization of fibroblast-specific protein 1 in pulmonary fibrosis. *American Journal of Respiratory and Critical Care Medicine* **171**:899–907. DOI: <https://doi.org/10.1164/rccm.200311-1535OC>, PMID: 15618458
- Lee SM**, Gardy JL, Cheung CY, Cheung TK, Hui KP, Ip NY, Guan Y, Hancock RE, Peiris JS. 2009. Systems-level comparison of host-responses elicited by avian H5N1 and seasonal H1N1 influenza viruses in primary human macrophages. *PLoS ONE* **4**:e8072. DOI: <https://doi.org/10.1371/journal.pone.0008072>, PMID: 20011590

- Lindskog C**, Fagerberg L, Hallström B, Edlund K, Hellwig B, Rahnenführer J, Kampf C, Uhlén M, Pontén F, Micke P. 2014. The lung-specific proteome defined by integration of transcriptomics and antibody-based profiling. *The FASEB Journal* **28**:5184–5196. DOI: <https://doi.org/10.1096/fj.14-254862>, PMID: 25169055
- Lozano R**, Naghavi M, Foreman K, Lim S, Shibuya K, Aboyans V, Abraham J, Adair T, Aggarwal R, Ahn SY, AlMazroa MA, Alvarado M, Anderson HR, Anderson LM, Andrews KG, Atkinson C, Baddour LM, Barker-Collo S, Bartels DH, Bell ML, et al. 2012. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the global burden of disease study 2010. *The Lancet* **380**:2095–2128. DOI: [https://doi.org/10.1016/S0140-6736\(12\)61728-0](https://doi.org/10.1016/S0140-6736(12)61728-0)
- Mainardi S**, Mijimolle N, Francoz S, Vicente-Dueñas C, Sánchez-García I, Barbacid M. 2014. Identification of cancer initiating cells in K-Ras driven lung adenocarcinoma. *PNAS* **111**:255–260. DOI: <https://doi.org/10.1073/pnas.1320383110>, PMID: 24367082
- Malkinson AM**, Koski KM, Evans WA, Festing MF. 1997. Butylated hydroxytoluene exposure is necessary to induce lung tumors in BALB mice treated with 3-methylcholanthrene. *Cancer Research* **57**:2832–2834. PMID: 9230183
- Manali ED**, Moschos C, Triantafyllidou C, Kotanidou A, Psallidas I, Karabela SP, Roussos C, Papiris S, Armaganidis A, Stathopoulos GT, Maniatis NA. 2011. Static and dynamic mechanics of the murine lung after intratracheal bleomycin. *BMC Pulmonary Medicine* **11**:33. DOI: <https://doi.org/10.1186/1471-2466-11-33>, PMID: 21627835
- Mason RJ**, Kalina M, Nielsen LD, Malkinson AM, Shannon JM. 2000. Surfactant protein C expression in urethane-induced murine pulmonary tumors. *The American Journal of Pathology* **156**:175–182. DOI: [https://doi.org/10.1016/S0002-9440\(10\)64717-7](https://doi.org/10.1016/S0002-9440(10)64717-7), PMID: 10623665
- Mazaika E**, Homsy J. 2014. Digital droplet PCR: cnv analysis and other applications. *Current Protocols in Human Genetics* **82**. DOI: <https://doi.org/10.1002/0471142905.hg0724s82>, PMID: 25042719
- Miller YE**, Dwyer-Nield LD, Keith RL, Le M, Franklin WA, Malkinson AM. 2003. Induction of a high incidence of lung tumors in C57BL/6 mice with multiple ethyl carbamate injections. *Cancer Letters* **198**:139–144. DOI: [https://doi.org/10.1016/S0304-3835\(03\)00309-4](https://doi.org/10.1016/S0304-3835(03)00309-4), PMID: 12957351
- Muzumdar MD**, Tasic B, Miyamichi K, Li L, Luo L. 2007. A global double-fluorescent cre reporter mouse. *Genesis* **45**:593–605. DOI: <https://doi.org/10.1002/dvg.20335>, PMID: 17868096
- Nagel R**, Semenova EA, Berns A. 2016. Drugging the addict: non-oncogene addiction as a target for cancer therapy. *EMBO Reports* **17**:1516–1531. DOI: <https://doi.org/10.15252/embr.201643030>, PMID: 27702988
- 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, et al. 2004. Classification of proliferative pulmonary lesions of the mouse: recommendations of the mouse models of human cancers consortium. *Cancer Research* **64**:2307–2316. DOI: <https://doi.org/10.1158/0008-5472.can-03-3376>, PMID: 15059877
- Ogilvy S**, Elefanty AG, Visvader J, Bath ML, Harris AW, Adams JM. 1998. Transcriptional regulation of vav, a gene expressed throughout the hematopoietic compartment. *Blood* **91**:419–430. PMID: 9427694
- Oikonomou N**, Mouratis MA, Tzouveleki A, Kaffe E, Valavanis C, Vilaras G, Karameris A, Prestwich GD, Bouras D, Aidinis V. 2012. Pulmonary autotaxin expression contributes to the pathogenesis of pulmonary fibrosis. *American Journal of Respiratory Cell and Molecular Biology* **47**:566–574. DOI: <https://doi.org/10.1165/rcmb.2012-0004OC>, PMID: 22744859
- Okubo T**, Knoepfler PS, Eisenman RN, Hogan BL. 2005. Nmyc plays an essential role during lung development as a dosage-sensitive regulator of progenitor cell proliferation and differentiation. *Development* **132**:1363–1374. DOI: <https://doi.org/10.1242/dev.01678>, PMID: 15716345
- Rawlins EL**, Okubo T, Xue Y, Brass DM, Auten RL, Hasegawa H, Wang F, Hogan BL. 2009. The role of Scgb1a1+ clara cells in the long-term maintenance and repair of lung airway, but not alveolar, epithelium. *Cell Stem Cell* **4**:525–534. DOI: <https://doi.org/10.1016/j.stem.2009.04.002>, PMID: 19497281
- Safran M**, Kim WY, Kung AL, Horner JW, DePinho RA, Kaelin WG. 2003. Mouse reporter strain for noninvasive bioluminescent imaging of cells that have undergone Cre-mediated recombination. *Molecular Imaging* **2**:297–302. DOI: <https://doi.org/10.1162/153535003322750637>, PMID: 14717328
- 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, Cheresch DA. 2014. An integrin  $\beta_3$ -KRAS-RalB complex drives tumorigenesis and resistance to EGFR inhibition. *Nature Cell Biology* **16**:457–468. DOI: <https://doi.org/10.1038/ncb2953>, PMID: 24747441
- Spella M**, Lilis I, Stathopoulos GT. 2017. Shared epithelial pathways to lung repair and disease. *European Respiratory Review* **26**:170048. DOI: <https://doi.org/10.1183/16000617.0048-2017>, PMID: 28659498
- Stathopoulos GT**, Sherrill TP, Cheng DS, Scoggins RM, Han W, Polosukhin VV, Connolly L, Yull FE, Fingleton B, Blackwell TS. 2007. Epithelial NF-kappaB activation promotes urethane-induced lung carcinogenesis. *PNAS* **104**:18514–18519. DOI: <https://doi.org/10.1073/pnas.0705316104>, PMID: 18000061
- Subramanian A**, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. 2005. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *PNAS* **102**:15545–15550. DOI: <https://doi.org/10.1073/pnas.0506580102>, PMID: 16199517
- Sutherland KD**, Song JY, Kwon MC, Proost N, Zevenhoven J, Berns A. 2014. Multiple cells-of-origin of mutant K-Ras-induced mouse lung adenocarcinoma. *PNAS* **111**:4952–4957. DOI: <https://doi.org/10.1073/pnas.1319963111>, PMID: 24586047
- Sutherland KD**, Berns A. 2010. Cell of origin of lung cancer. *Molecular Oncology* **4**:397–403. DOI: <https://doi.org/10.1016/j.molonc.2010.05.002>, PMID: 20594926

- Torre LA**, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. 2015. Global cancer statistics, 2012. *CA: A Cancer Journal for Clinicians* **65**:87–108. DOI: <https://doi.org/10.3322/caac.21262>, PMID: 25651787
- Treutlein B**, Brownfield DG, Wu AR, Neff NF, Mantalas GL, Espinoza FH, Desai TJ, Krasnow MA, Quake SR. 2014. Reconstructing lineage hierarchies of the distal lung epithelium using single-cell RNA-seq. *Nature* **509**:371–375. DOI: <https://doi.org/10.1038/nature13173>, PMID: 24739965
- Tronche F**, Kellendonk C, Kretz O, Gass P, Anlag K, Orban PC, Bock R, Klein R, Schütz G. 1999. Disruption of the glucocorticoid receptor gene in the nervous system results in reduced anxiety. *Nature Genetics* **23**:99–103. DOI: <https://doi.org/10.1038/12703>, PMID: 10471508
- 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. 2017. Chronic cigarette Smoke-Induced epigenomic changes precede sensitization of bronchial epithelial cells to Single-Step transformation by KRAS mutations. *Cancer Cell* **32**:360–376. DOI: <https://doi.org/10.1016/j.ccell.2017.08.006>, PMID: 28898697
- Voehringer D**, Liang HE, Locksley RM. 2008. Homeostasis and effector function of lymphopenia-induced "memory-like" T cells in constitutively T cell-depleted mice. *The Journal of Immunology* **180**:4742–4753. DOI: <https://doi.org/10.4049/jimmunol.180.7.4742>, PMID: 18354198
- 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. 2018. Ikb kinase  $\alpha$  is required for development and progression of KRAS-Mutant Lung Adenocarcinoma. *Cancer Research* **78**:2939–2951. DOI: <https://doi.org/10.1158/0008-5472.CAN-17-1944>, PMID: 29588349
- 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. 2015. The mutational landscapes of genetic and chemical models of Kras-driven lung cancer. *Nature* **517**:489–492. DOI: <https://doi.org/10.1038/nature13898>, PMID: 25363767
- Xu X**, Rock JR, Lu Y, Futtner C, Schwab B, Guinney J, Hogan BL, Onaitis MW. 2012. Evidence for type II cells as cells of origin of K-Ras-induced distal lung adenocarcinoma. *PNAS* **109**:4910–4915. DOI: <https://doi.org/10.1073/pnas.1112499109>, PMID: 22411819
- 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. 2016. Single-cell RNA sequencing identifies diverse roles of epithelial cells in idiopathic pulmonary fibrosis. *JCI Insight* **1**:e90558. DOI: <https://doi.org/10.1172/jci.insight.90558>, PMID: 27942595
- Yee M**, Chess PR, McGrath-Morrow SA, Wang Z, Gelein R, Zhou R, Dean DA, Notter RH, O'Reilly MA. 2009. Neonatal oxygen adversely affects lung function in adult mice without altering surfactant composition or activity. *American Journal of Physiology-Lung Cellular and Molecular Physiology* **297**:L641–L649. DOI: <https://doi.org/10.1152/ajplung.00023.2009>, PMID: 19617311
- 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. 2015. p63(+)/Krt5(+) distal airway stem cells are essential for lung regeneration. *Nature* **517**:616–620. DOI: <https://doi.org/10.1038/nature13903>, PMID: 25383540

## Appendix 1

DOI: <https://doi.org/10.7554/eLife.45571.073>

### Abbreviations list and master legend

AEC, airway epithelial cells; AM $\Phi$ , 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;  $\mu$ CT, 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.

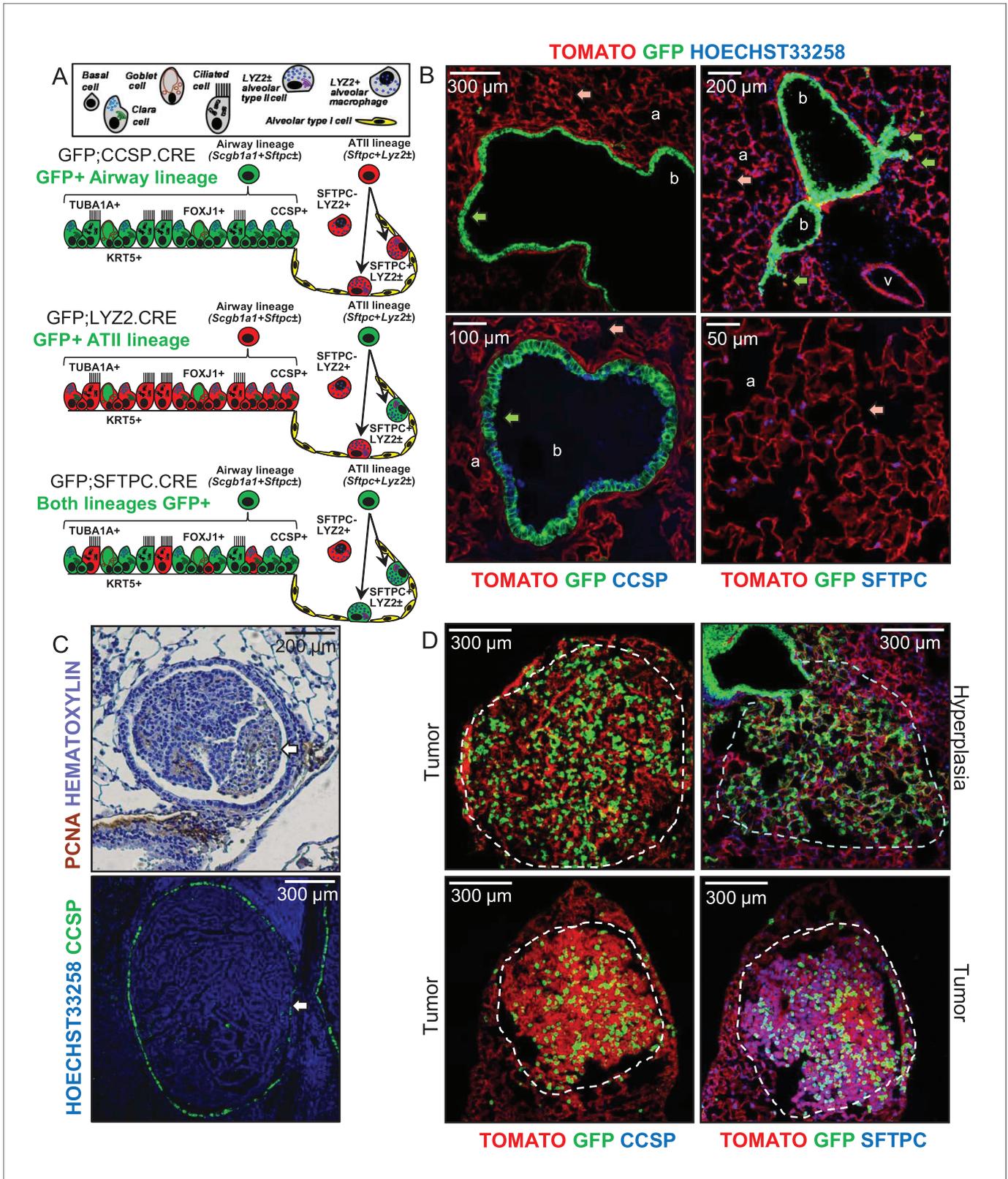


---

## Figures and figure supplements

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

**Magda Spella *et al***



**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 Figure 1 continued on next page

*Figure 1 continued*

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) and hematoxylin (blue)-stained (top) and CCSP (green) and 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, top), and immunostained for the club cell marker CCSP (bottom left) and the alveolar type II cell marker SFTPC (bottom 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  $\alpha$ -tubulin; SFTPC, surfactant protein C; LY2Z, lysozyme 2; FOXJ1, forkhead box J1; KRT5, keratin 5.

DOI: <https://doi.org/10.7554/eLife.45571.003>

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

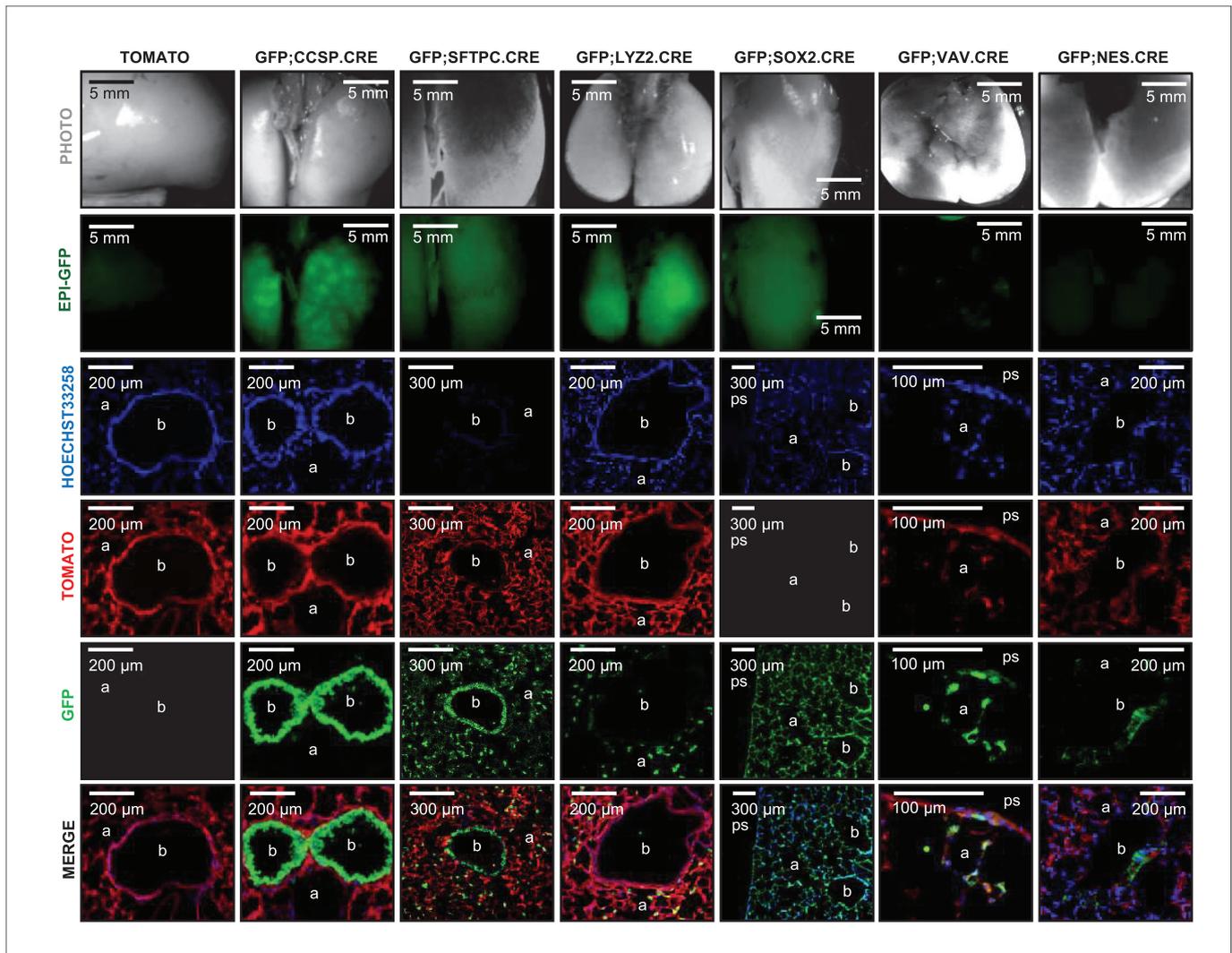
**Figure 1—figure supplement 1.** Table of pulmonary lineage markers and key abbreviations used in this study. TUBA1A, Tubulin alpha 1a or acetylated tubulin; KRT5, Keratin 5; FOXJ1, Forkhead box J1; CCSP, Secretoglobulin, family 1A, member 1 (uteroglobin) or Clara cell secretory protein or Clara cell 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.

DOI: <https://doi.org/10.7554/eLife.45571.004>

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/F1mg	Tg(Sftpc-cre)1Blh	B6.129P2-Lyz2tm1(cre)lfo/J	B6.Cg-Tg(Sox2-cre)1Amc/J	B6.Cg-Tg(Vav1-icre)A2Kio/J	B6.Cg-Tg(Nes-cre)1Kln/J
Reference		Muzumdar et al., 2007	Oikonomou et al., 2012	Okubo et al., 2005	Desai et al., 2014	Hayashi et al., 2002	Ogilvy et al., 1998	Tronche et al., 1999
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	-	-	-	±	+	+	-

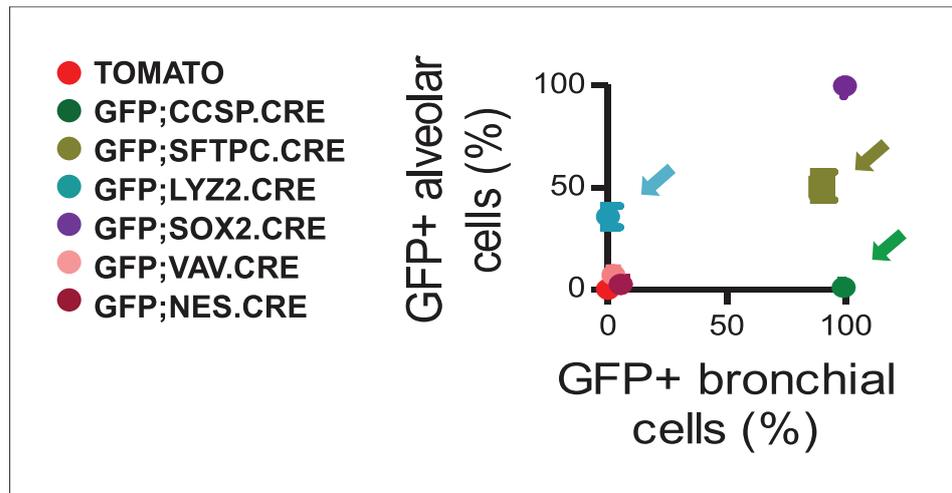
**Figure 1—figure supplement 2.** Genetic labeling of pulmonary lineages in eleven mouse strains and intercrosses: summary of results. CRE, causes recombination; TOMATO, tdTomato; GFP, green fluorescent protein; CCSP, Clara cell secretory protein; SFTPC, surfactant protein C; LYZ2, lysozyme 2; SOX2, 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.

DOI: <https://doi.org/10.7554/eLife.45571.005>



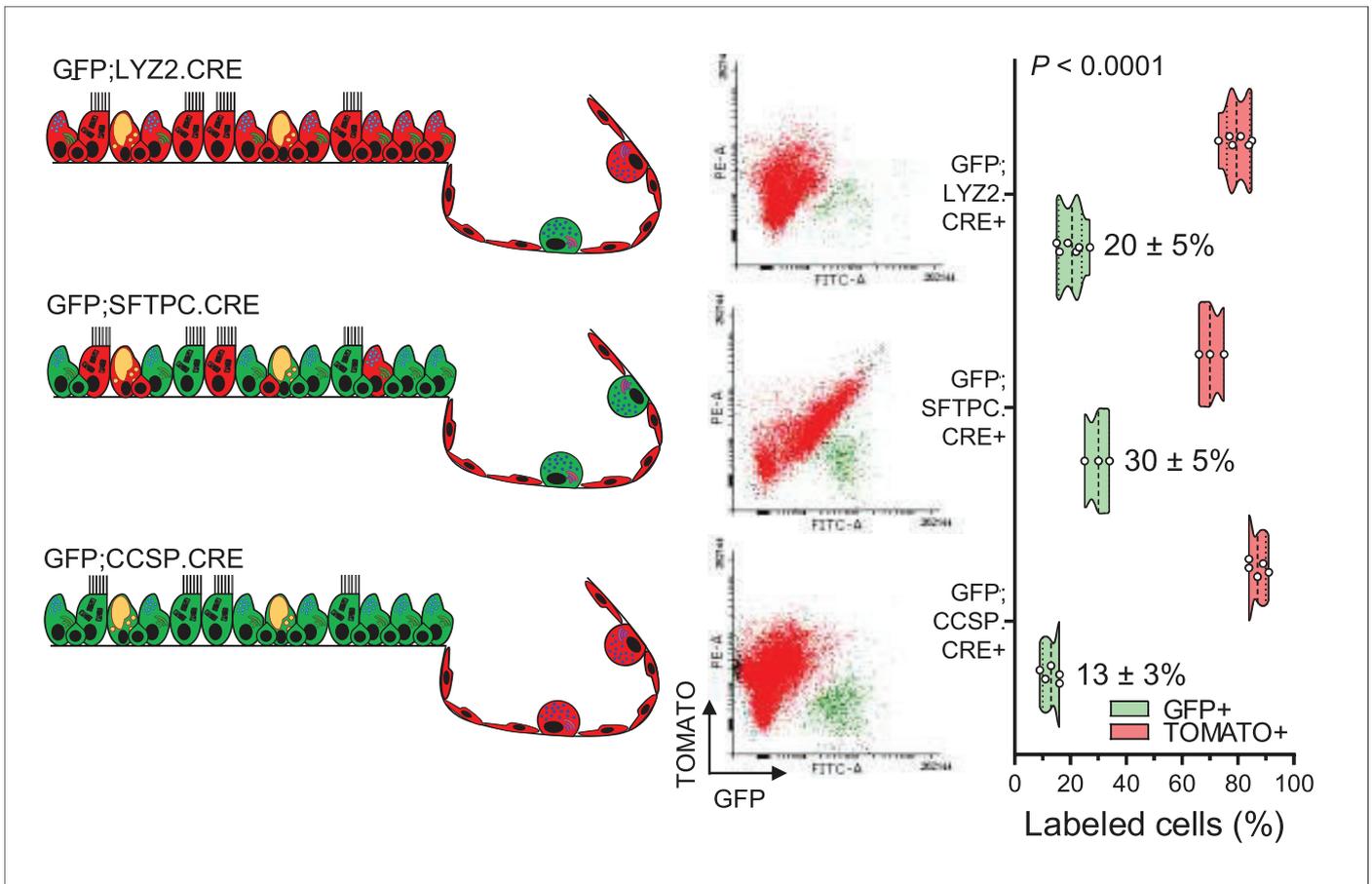
**Figure 1—figure supplement 3.** Genetic labeling of pulmonary lineages in seven lineage reporter strains on the C57BL/6 background: representative images. Representative photographs (top row) and green epifluorescence images (second row) of whole lungs, as well as fluorescent microscopic images of 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.

DOI: <https://doi.org/10.7554/eLife.45571.006>



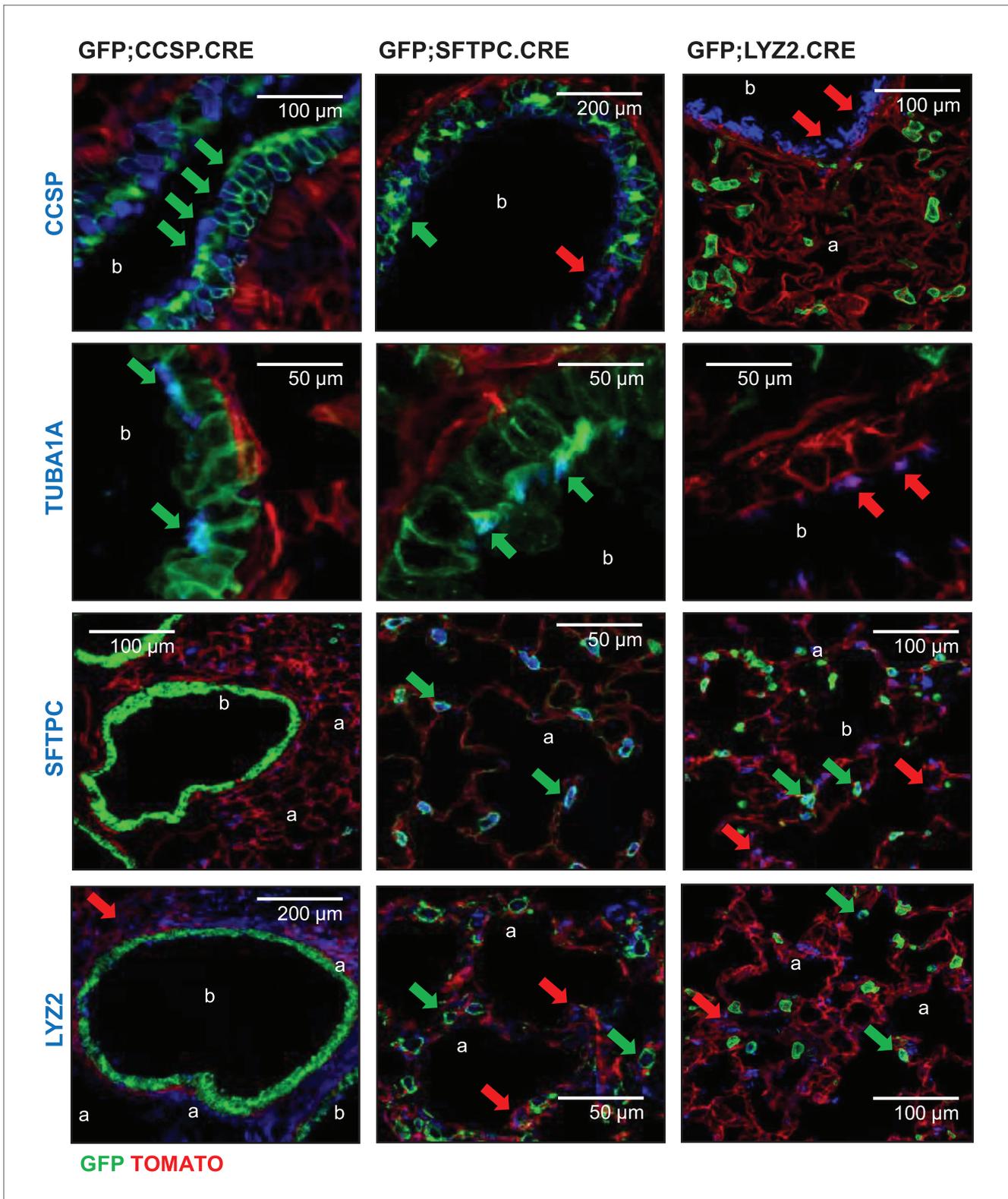
**Figure 1—figure supplement 4.** Genetic labeling of pulmonary lineages in seven lineage reporter strains on the C57BL/6 background: data summary. XY plot of GFP-labeled airway versus alveolar cells from  $n = 5$  mice/mouse strain. Arrows denote the three lineage-reporter strains selected for further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

DOI: <https://doi.org/10.7554/eLife.45571.007>



**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 six mice/strain (right). Numbers are mean  $\pm$  SD.  $P$ , overall probability, two-way ANOVA.

DOI: <https://doi.org/10.7554/eLife.45571.009>

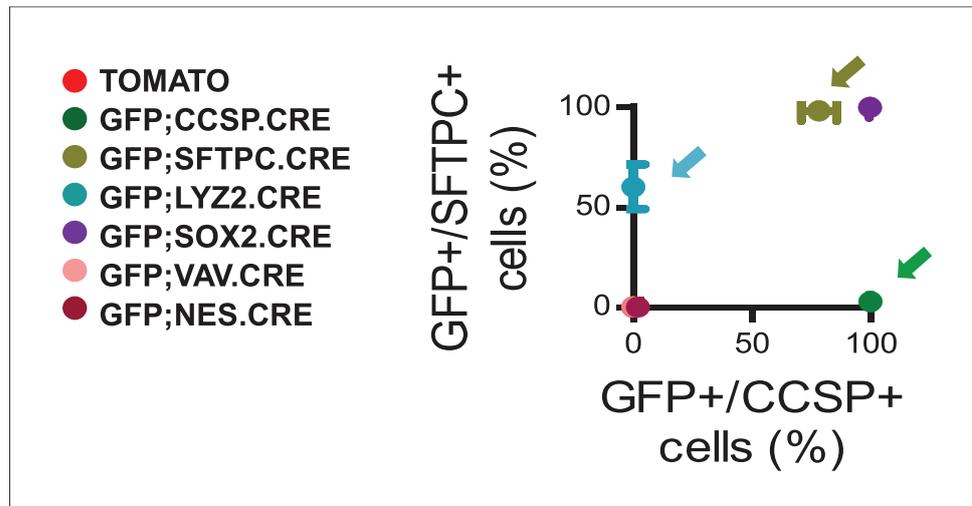


**Figure 1—figure supplement 6.** Genetic lineage labels of protein-marked cells in three lineage reporter strains on the C57BL/6 background: representative images. Representative merged fluorescent microscopic images from lineage marker-stained lung sections of 6-week-old lineage-  
 Figure 1—figure supplement 6 continued on next page

*Figure 1—figure supplement 6 continued*

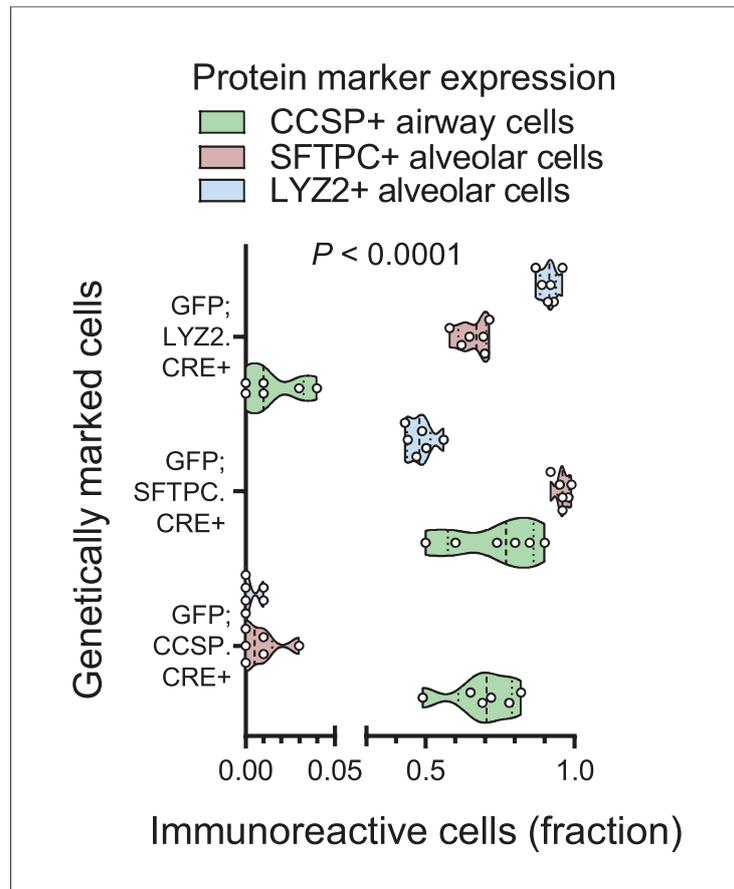
labeled mice ( $n = 5/\text{group}$ ). Arrows indicate cells expressing the respective marker protein with (green) or without (red) genetic lineage-labeling. CCSP, Clara cell secretory protein; TUBA1A, acetylated tubulin; SFTPC, surfactant protein C; LYZ2, lysozyme 2; b, bronchi; a, alveoli.

DOI: <https://doi.org/10.7554/eLife.45571.011>



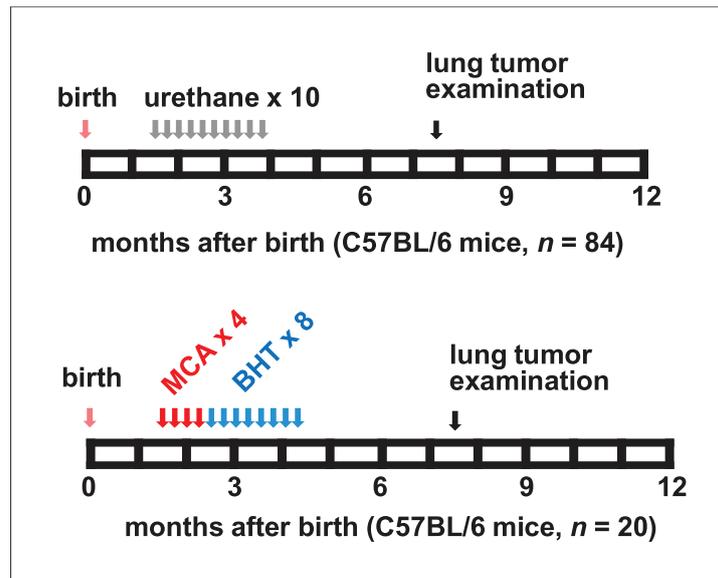
**Figure 1—figure supplement 7.** Genetic lineage labels of protein-marked cells in seven lineage reporter strains on the C57BL/6 background: data summary. XY plot of ratios of genetic GFP-labeled to protein marker CCSP and SFTPC-immunoreactive cells ( $n = 5/\text{group}$ ). Arrows denote the three 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  $\pm$  SD.

DOI: <https://doi.org/10.7554/eLife.45571.012>



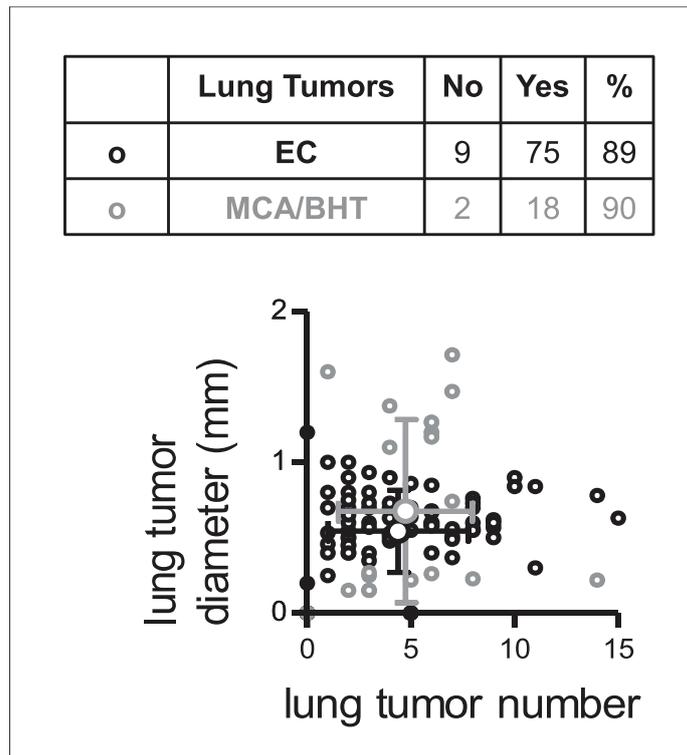
**Figure 1—figure supplement 8.** Protein markings of lineage-labeled cells in three lineage reporter strains on the C57BL/6 background: data summary. Quantification of protein marker expression of genetic-labeled cells of GFP; CCSP.CRE, GFP;LYZ2.CRE, and GFP;SFTPC.CRE mice ( $n = 6$ /strain) for Clara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

DOI: <https://doi.org/10.7554/eLife.45571.014>



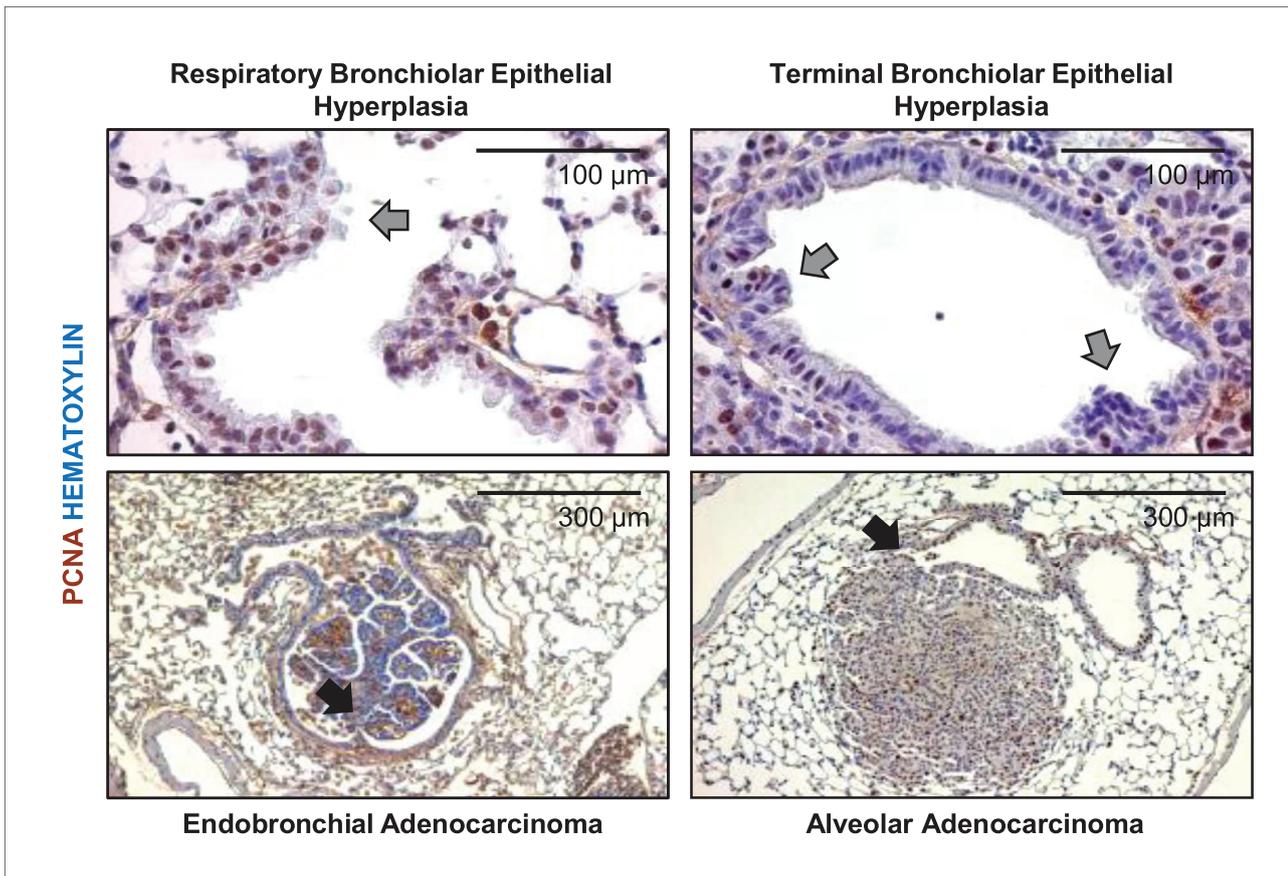
**Figure 1—figure supplement 9.** Two carcinogen regimens for reproducible lung tumor induction in naturally resistant C57BL/6 mice. Top: schematic of multi-hit urethane administration tailored to yield 90% tumor incidence in C57BL/6 mice: ten weekly intraperitoneal injections of 1 g/Kg urethane (ethyl carbamate, EC; gray 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-dihydrobenzo[*a*]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).

DOI: <https://doi.org/10.7554/eLife.45571.016>



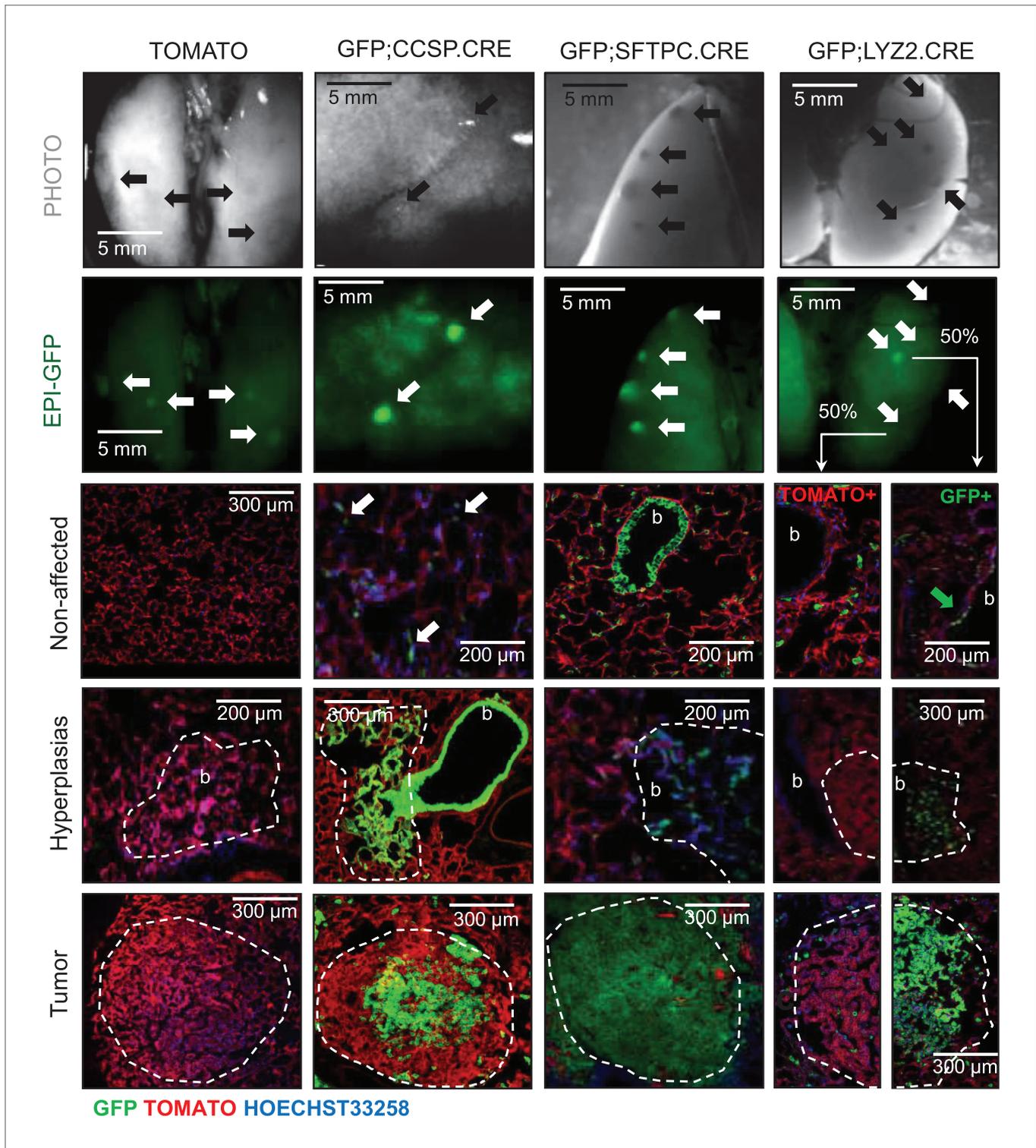
**Figure 1—figure supplement 10.** Lung tumors induced in C57BL/6 mice by two carcinogen regimens. Eighty-four C57BL/6 mice received ten weekly intraperitoneal injections of 1 g/Kg urethane (ethyl carbamate, EC) initiated at six weeks of age and lungs were examined six months after the first urethane injection (black font and symbols). Twenty C57BL/6 mice received four weekly intraperitoneal injections of 15 mg/Kg 3-methyl-1,2-dihydrobenzo[*a*]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 (gray 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.

DOI: <https://doi.org/10.7554/eLife.45571.017>



**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 (gray) and lung adenocarcinomas (black) arising within a bronchus (left) and apparently in an alveolar region but adjacent to a bronchus (right).

DOI: <https://doi.org/10.7554/eLife.45571.019>

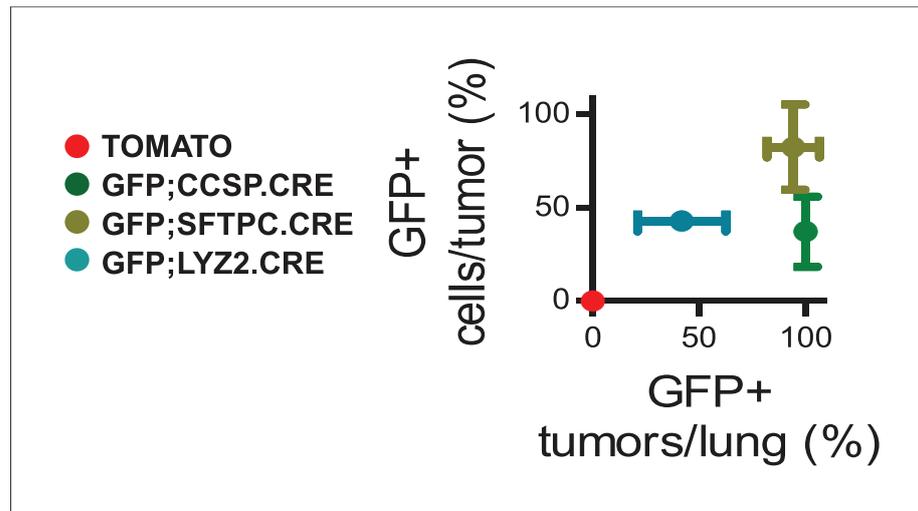


**Figure 1—figure supplement 12.** Genetic labeling of urethane-induced lung adenocarcinomas in four lineage reporter strains on the C57BL/6 background: representative images. Representative photographs (top row) and green epifluorescence images (second row), as well as merged fluorescent microscopic 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 **Figure 1—figure supplement 12 continued on next page**

Figure 1—figure supplement 12 continued

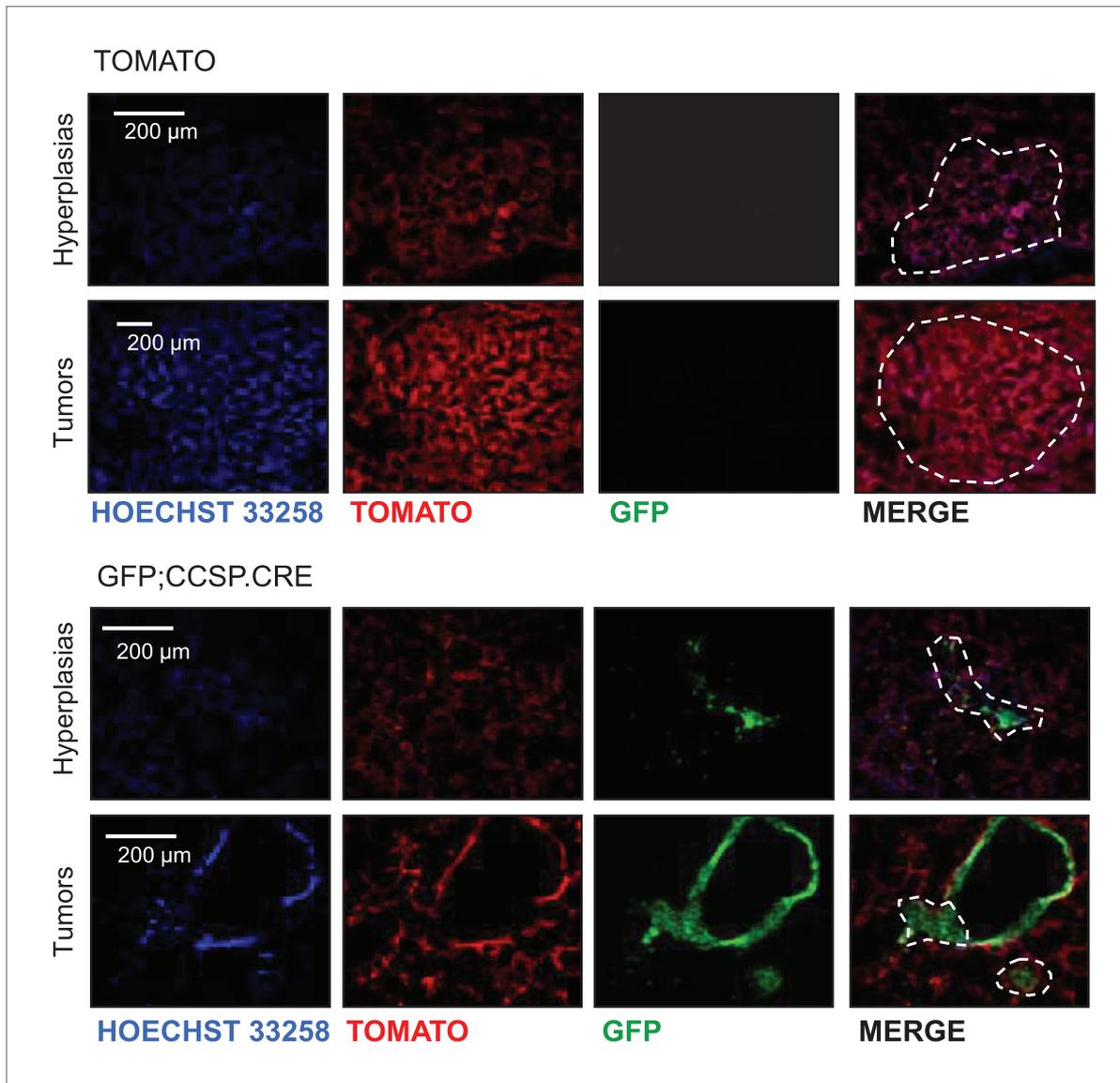
lung tumors. Bottom three rows: white arrows indicate GFP-labeled cells in apparently non-affected alveolar areas of GFP;CCSP.CRE mice; green arrow indicates rare GFP+ cell in non-affected central airway of GFP;LYZ2.CRE mouse. Note the absence of GFP-labeling of lung tumors in TOMATO mice, the complete GFP-labeling in GFP;CCSP.CRE and GFP;SFTPC.CRE mice, and the partial GFP-labeling in GFP;LYZ2.CRE mice.

DOI: <https://doi.org/10.7554/eLife.45571.020>



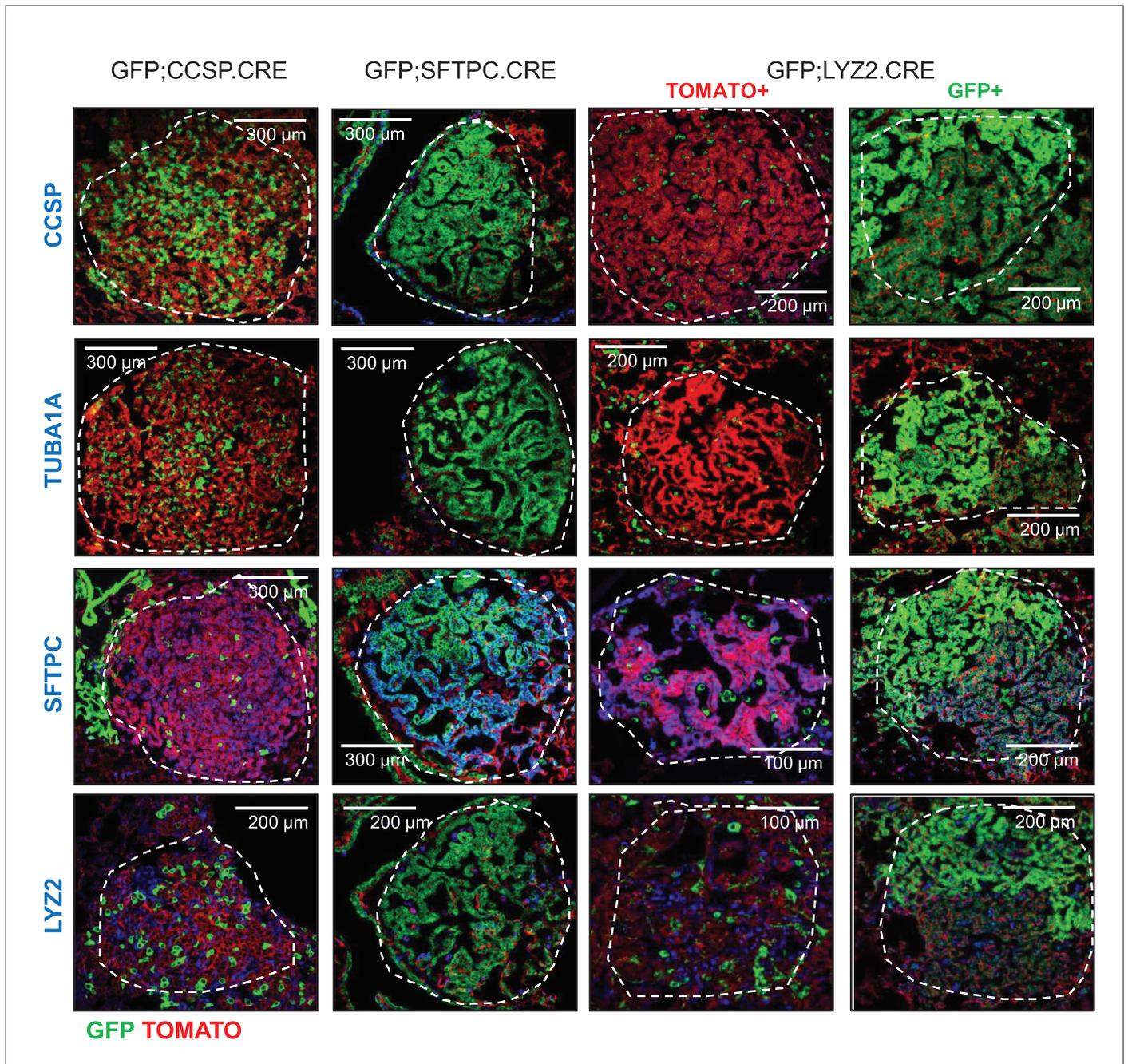
**Figure 1—figure supplement 13.** Genetic labeling of urethane-induced lung adenocarcinomas in four lineage reporter strains on the C57BL/6 background: data summary. XY plot of percentage of GFP-labeled tumors/lung versus GFP-labeled tumor cells/tumor averaged per lung in strains from **Figure 1—figure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

DOI: <https://doi.org/10.7554/eLife.45571.021>

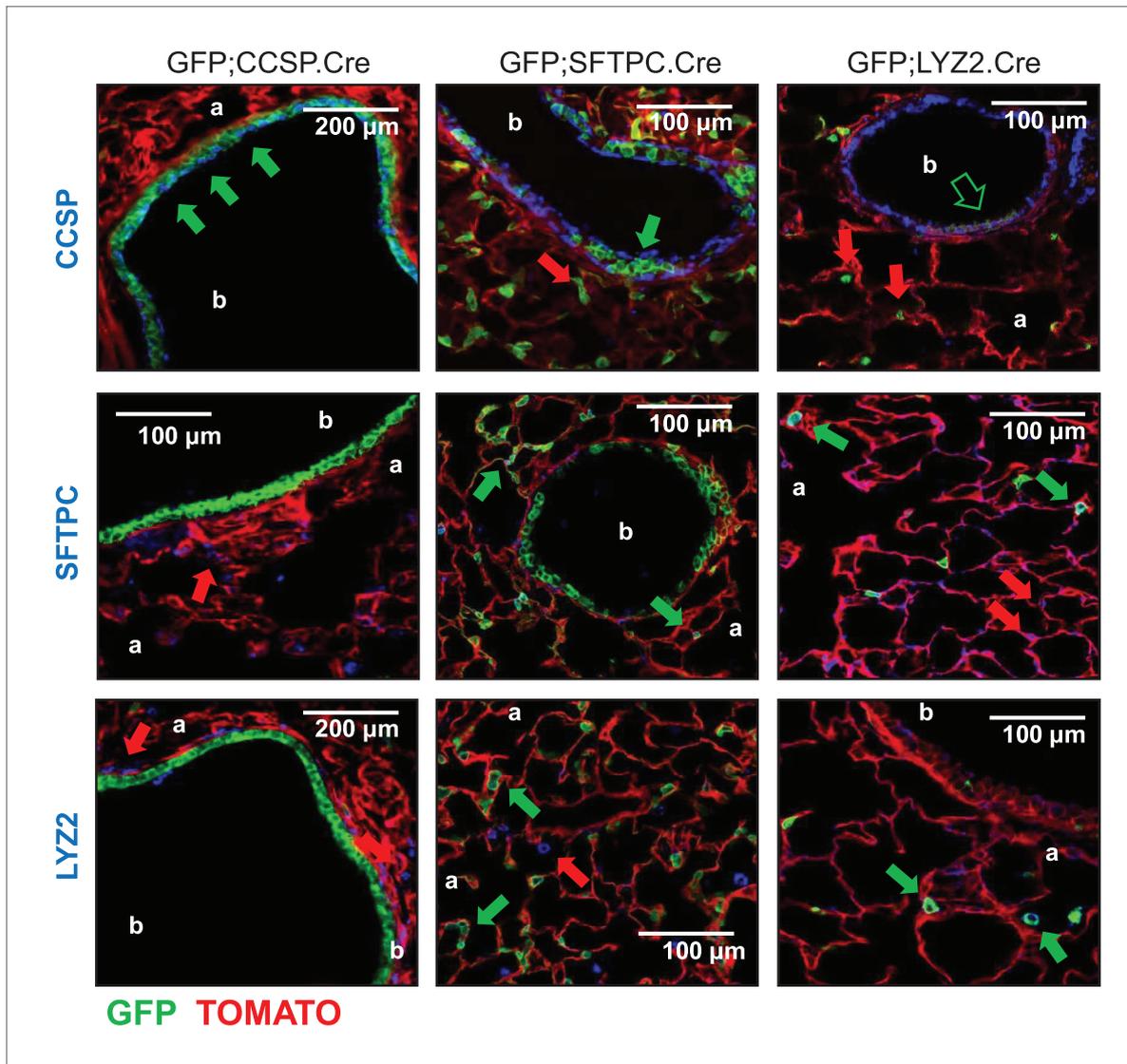


**Figure 1—figure supplement 14.** Genetic labeling of MCA/BHT-induced lung adenocarcinomas in two lineage reporter strains on the C57BL/6 background: representative images. Single-channel (endogenous TOMATO and GFP labeling and Hoechst 33258 nuclear stain) and merged images of lung hyperplasias and 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) ( $n = 8/\text{group}$ ). Note the absence of GFP-labeling in lesions of TOMATO mice and the GFP-labeled lesions of GFP;CCSP.CRE mice.

DOI: <https://doi.org/10.7554/eLife.45571.023>

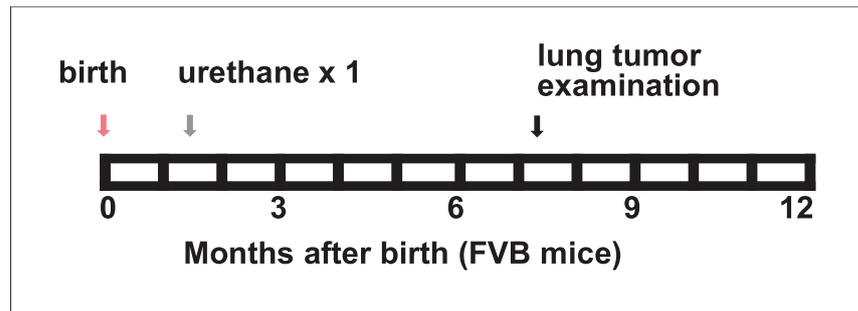


**Figure 1—figure supplement 15.** Protein marker expression of urethane-induced lung adenocarcinomas in three lineage-labeled mouse strains on the C57BL/6 background: representative images. Lineage marker protein-stained lung adenocarcinomas (dashed outlines) from genetically marked mice ( $n = 10/\text{group}$ ). 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  $\alpha$ -tubulin; SFTPC, surfactant protein C; LYZ2, lysozyme 2.  
DOI: <https://doi.org/10.7554/eLife.45571.024>

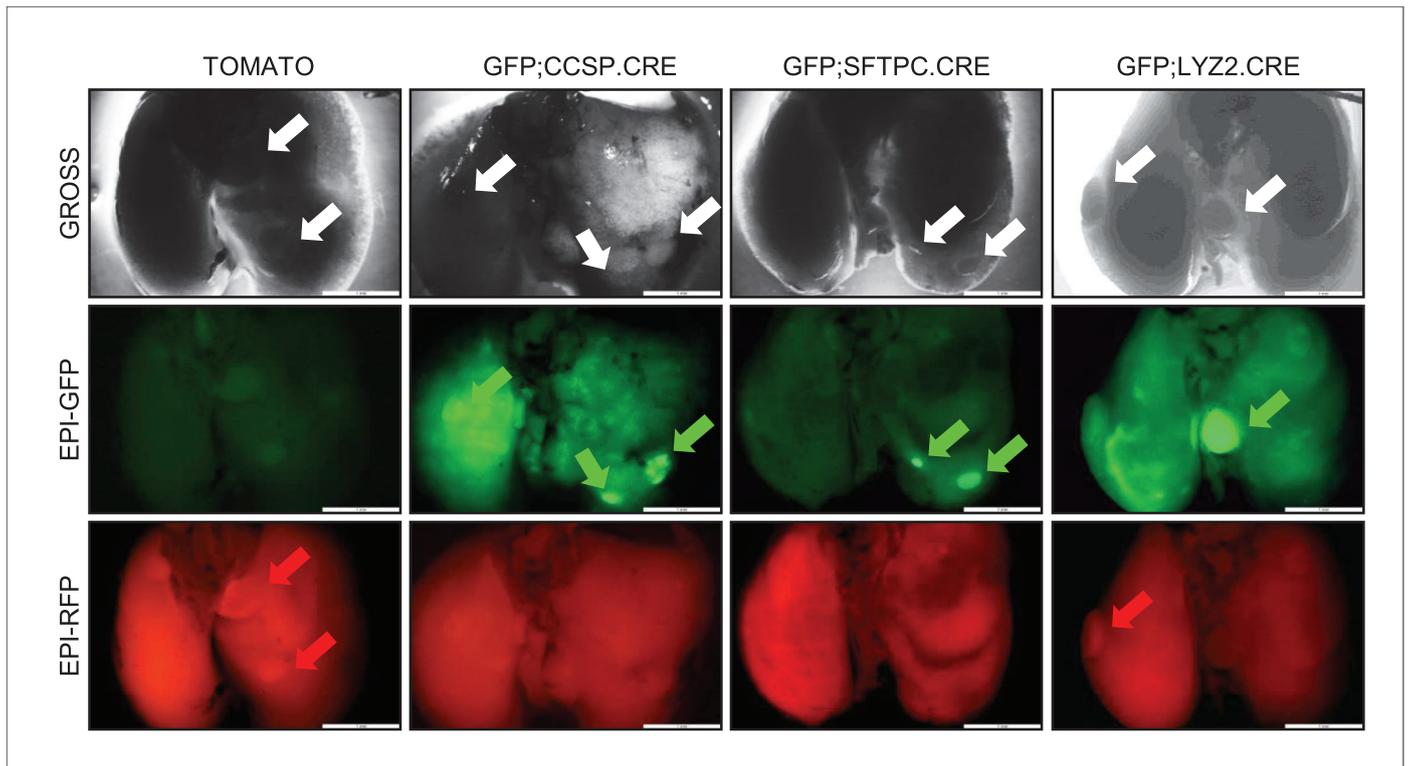


**Figure 1—figure supplement 16.** Genetic lineage labels of protein-marked cells in three lineage reporter strains on the FVB background: representative images. Representative merged fluorescent microscopic images from lineage marker-stained lung sections of 6-week-old lineage reporter mice ( $n = 5/\text{group}$ ). 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.

DOI: <https://doi.org/10.7554/eLife.45571.025>

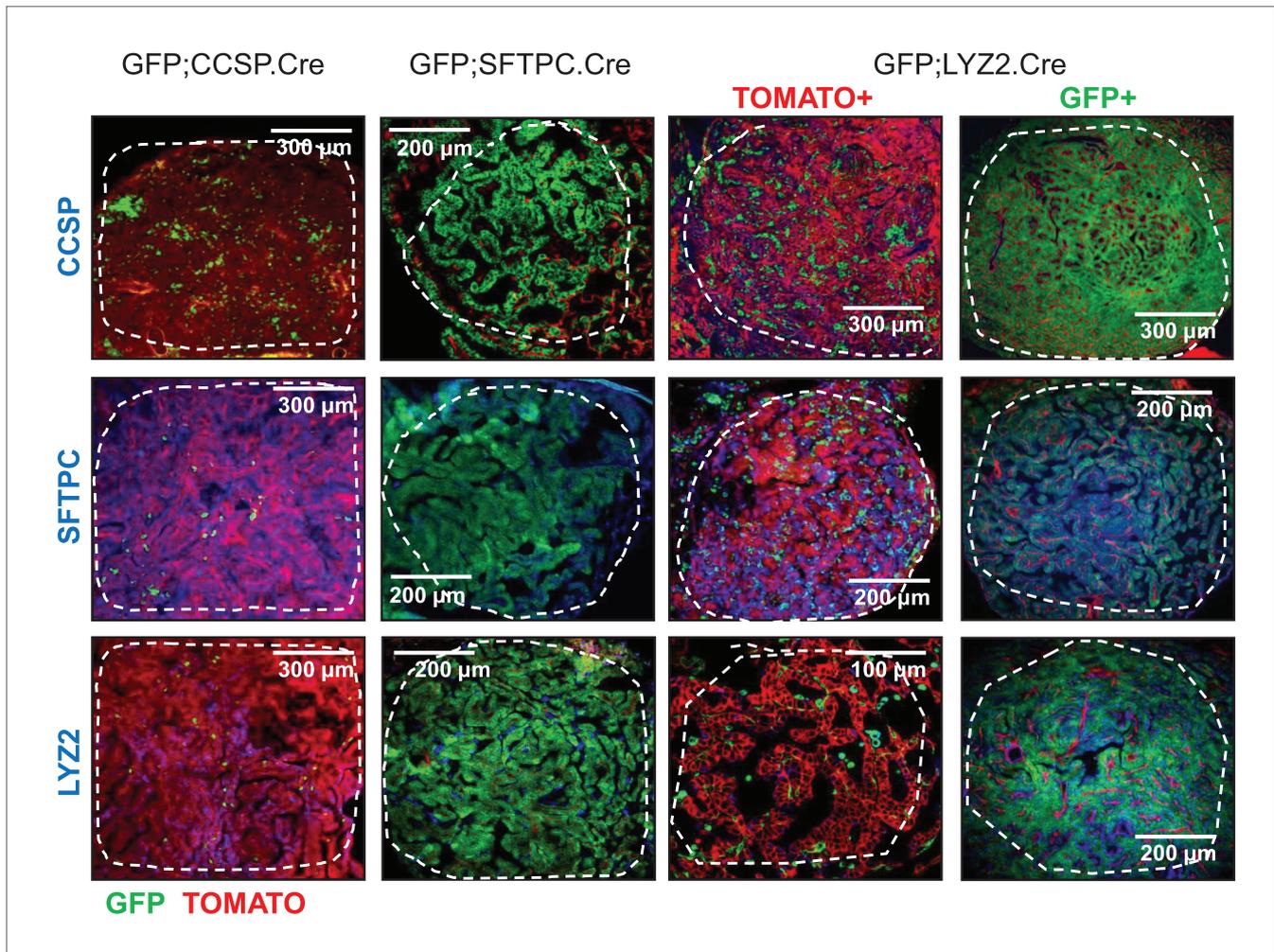


**Figure 1—figure supplement 17.** A single-hit mouse model for urethane-induced lung adenocarcinoma induction in naturally susceptible FVB mice. Schematic of single-hit urethane administration tailored to yield 100% tumor incidence in FVB mice: one intraperitoneal injection of 1 g/Kg urethane (ethyl carbamate, EC; gray arrow) is delivered at six weeks after birth (pink arrow) and lungs are examined six months later (black arrow).  
DOI: <https://doi.org/10.7554/eLife.45571.026>



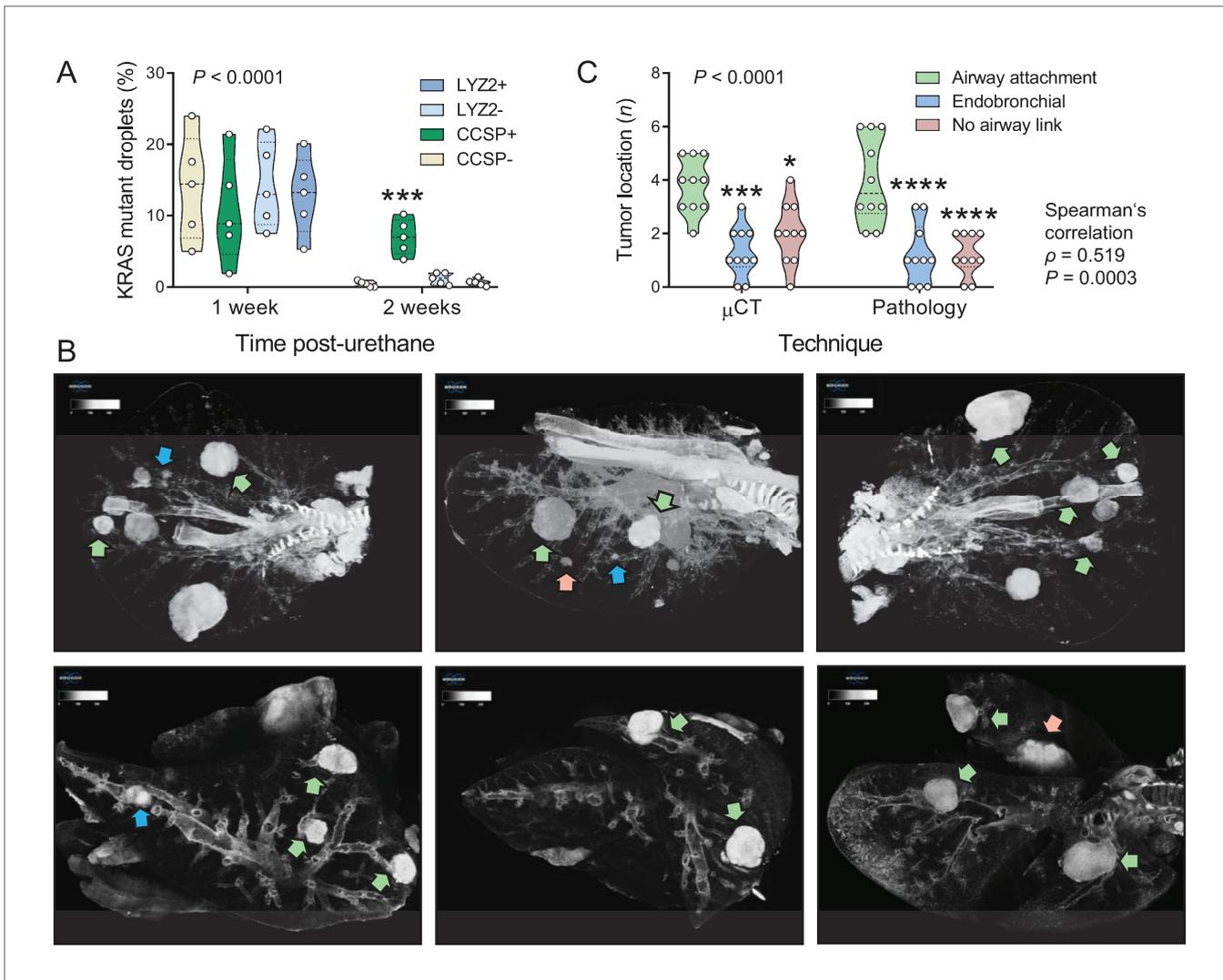
**Figure 1—figure supplement 18.** High-throughput epifluorescent detection of genetic labeling of urethane-induced lung adenocarcinomas in four lineage reporter strains on the FVB background: representative images. Representative photographs (top) and green (middle) and red (bottom) epifluorescence 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.

DOI: <https://doi.org/10.7554/eLife.45571.027>



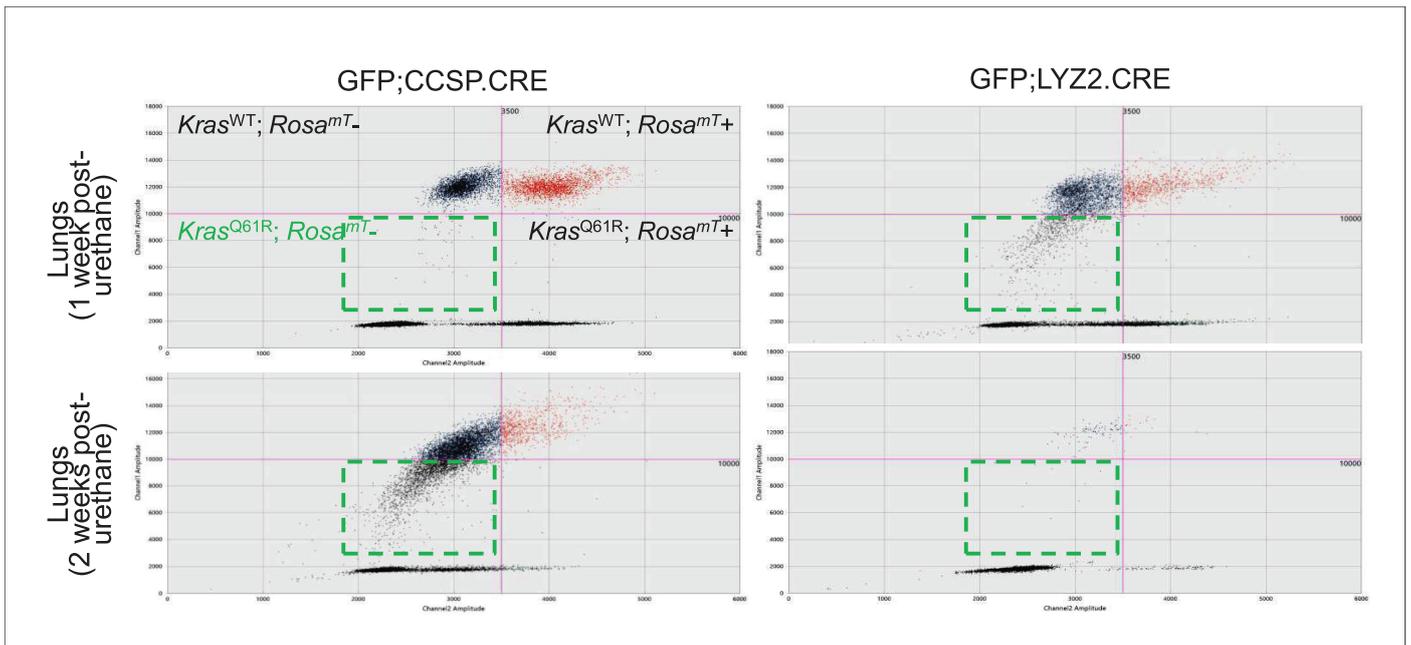
**Figure 1—figure supplement 19.** Genetic labeling of urethane-induced lung adenocarcinomas in three lineage reporter strains on the FVB background: representative images. Representative merged fluorescent microscopic images of lineage marker protein-stained lung tumors (dashed outlines) from genetically 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.

DOI: <https://doi.org/10.7554/eLife.45571.028>



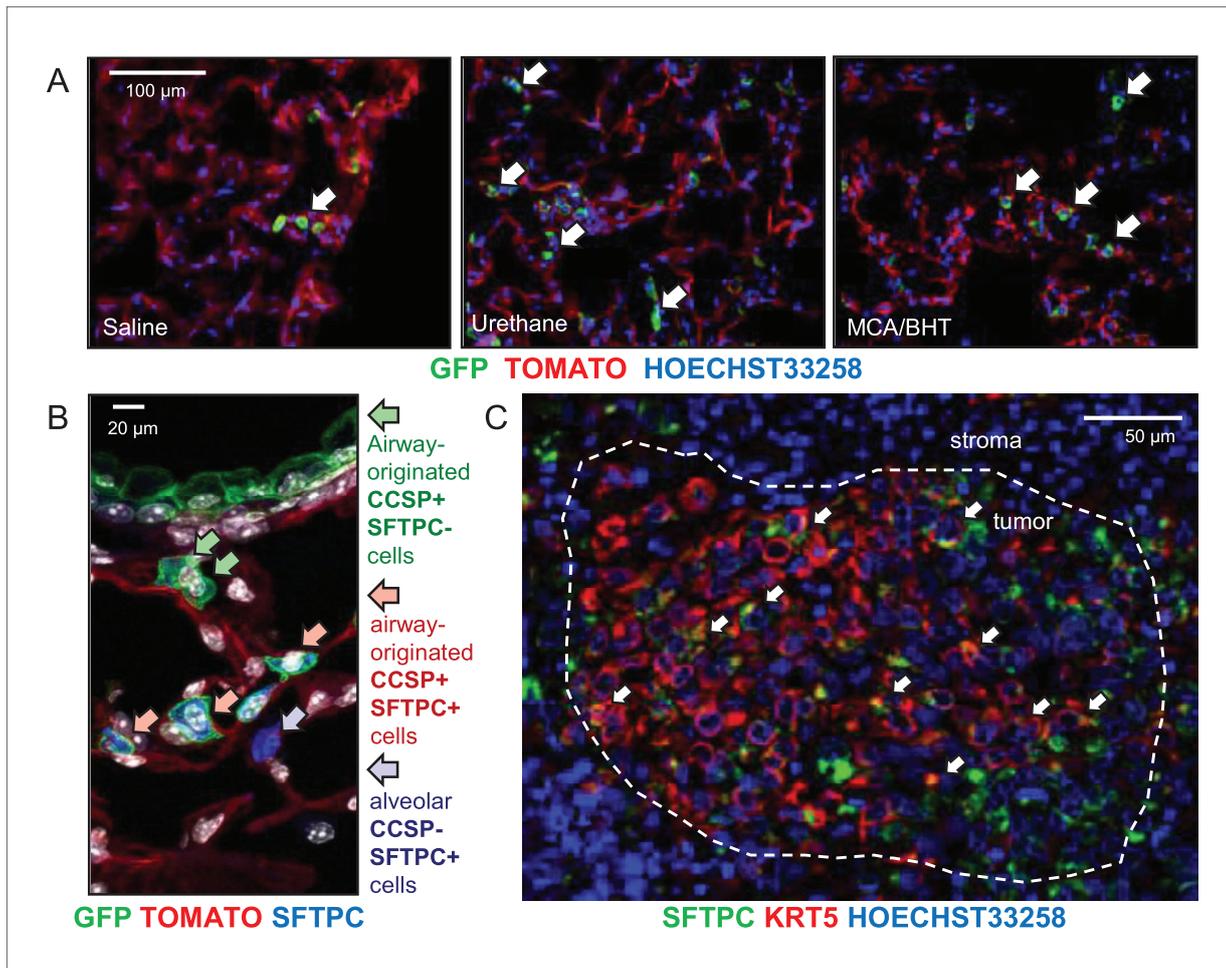
**Figure 2.** Airway cells sustain *Kras*<sup>Q61R</sup> mutations inflicted by urethane and give rise to juxtabronchial lung adenocarcinomas. (A) DNA was extracted from the lungs of GFP;CCSP.CRE and GFP;LYZ2.CRE mice (FVB strain) one and two weeks post-urethane treatment ( $n = 5/\text{group}$ ). Summary of duplexed digital droplet PCR (ddPCR) results using primers and probes specific for the *Rosa*<sup>mT</sup> and the *Kras*<sup>WT</sup> sequences. Note that all cell types equally suffer initial *Kras*<sup>Q61R</sup> mutations, but only GFP-labeled cells of GFP;CCSP.CRE mice (i.e. airway cells) maintain the *Kras*<sup>Q61R</sup> 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 ( $\mu\text{CT}$ ) 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  $\mu\text{CT}$  (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 ( $\rho$ ) and probability ( $P$ ) for correlation of  $\mu\text{CT}$  and pathology results.

DOI: <https://doi.org/10.7554/eLife.45571.029>



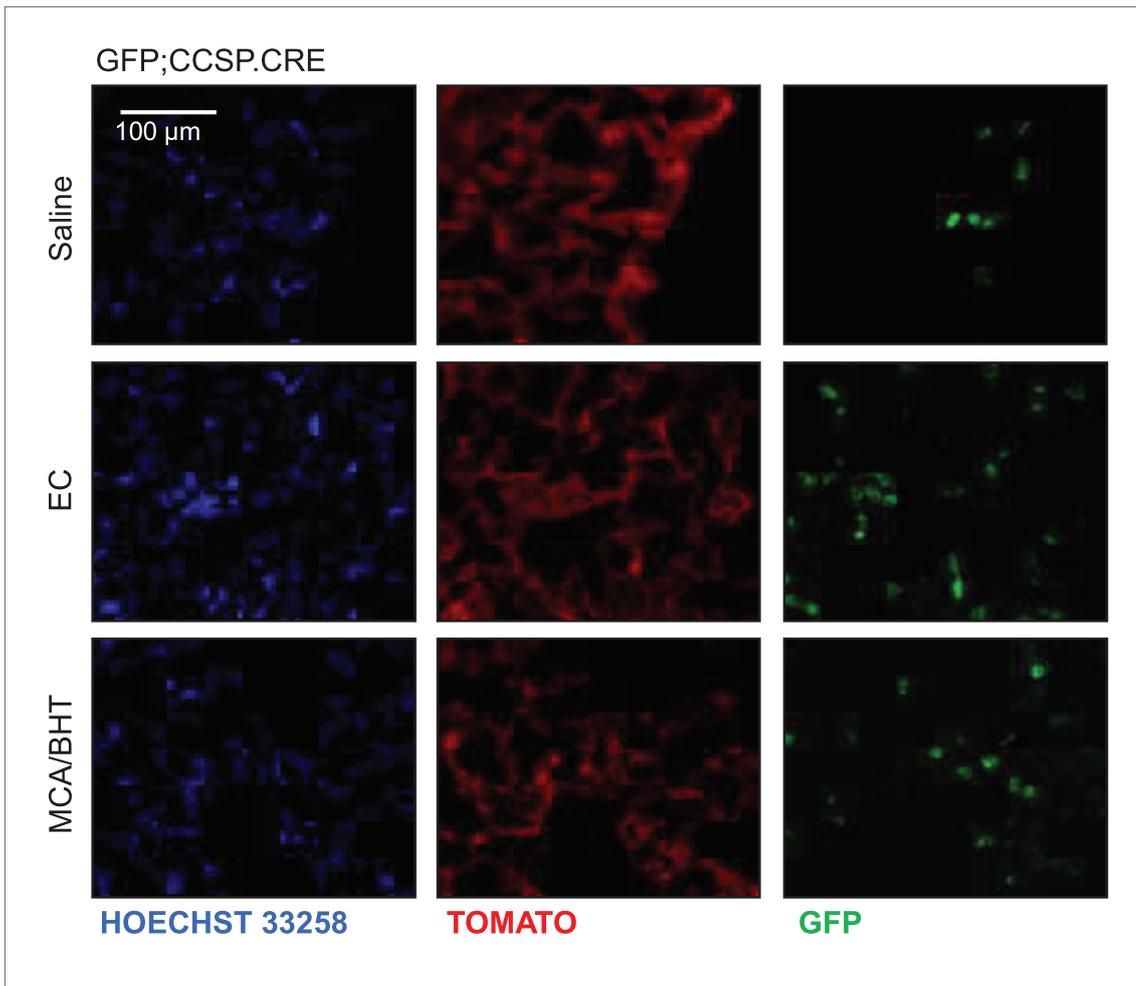
**Figure 2—figure supplement 1.** Airway cells sustain  $Kras^{Q61R}$  mutations inflicted by urethane. DNA was extracted from the lungs of GFP;CCSP.CRE and GFP;LYZ2.CRE mice (FVB strain) one and two weeks post-urethane treatment ( $n = 5/\text{group}$ ). Representative gating strategy of digital droplet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

DOI: <https://doi.org/10.7554/eLife.45571.030>



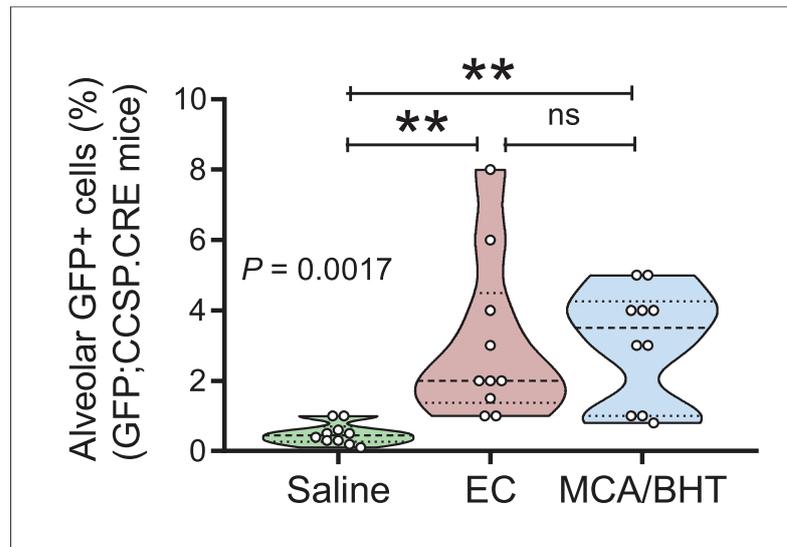
**Figure 3.** Expansion of airway cells in the tumor-initiated lung. (A) Non-neoplastic alveolar regions from lung sections of saline-, urethane (ethyl carbamate, EC)-, and 3-methyl-1,2-dihydrobenzo[*j*]aceanthrylene/butylated hydroxytoluene (MCA/BHT)-treated GFP;CCSP.CRE mice at six months into treatment ( $n = 8$  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 and 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.

DOI: <https://doi.org/10.7554/eLife.45571.033>



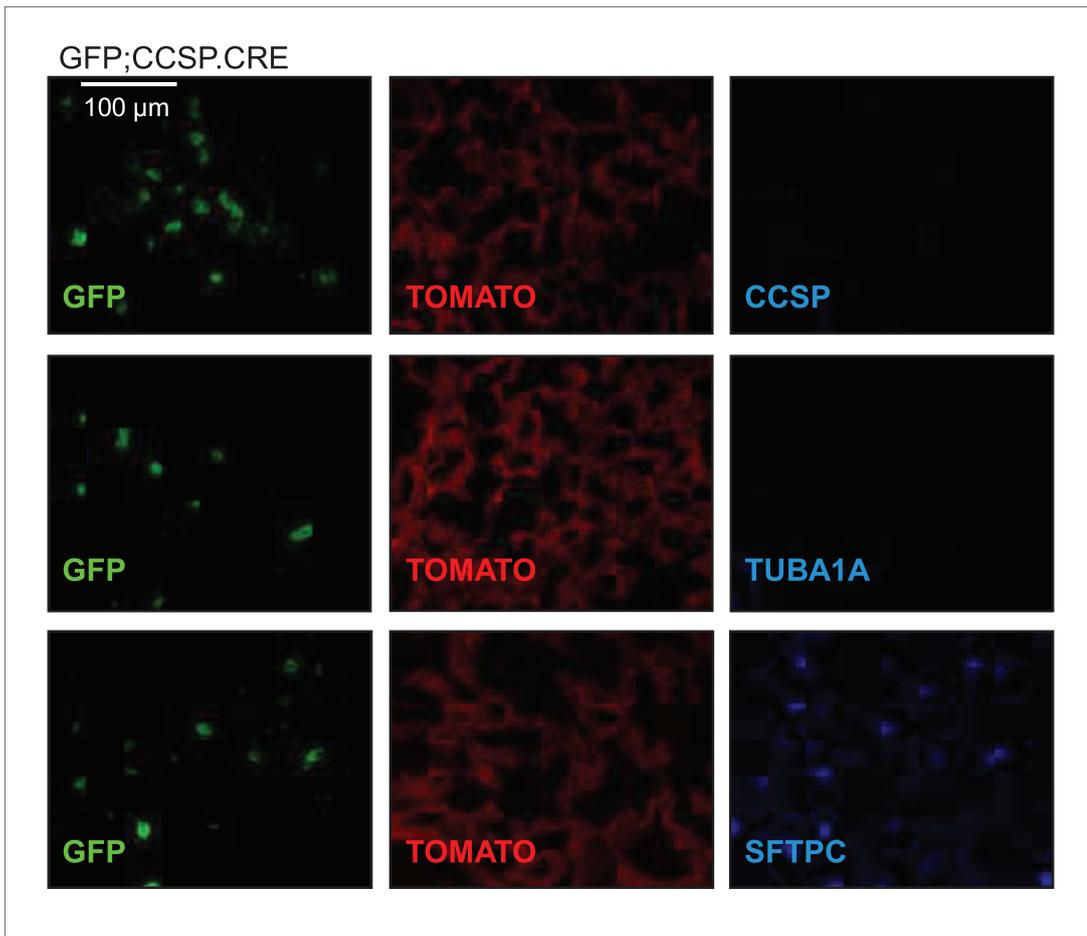
**Figure 3—figure supplement 1.** Airway-labeled cells in the alveoli of carcinogen-exposed C57BL/6 mice: representative images. Single-channel microscopy images (endogenous TOMATO and GFP labeling with Hoechst 33258 nuclear stain) of non-neoplastic alveolar regions of GFP;CCSP.CRE mice treated as in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

DOI: <https://doi.org/10.7554/eLife.45571.034>



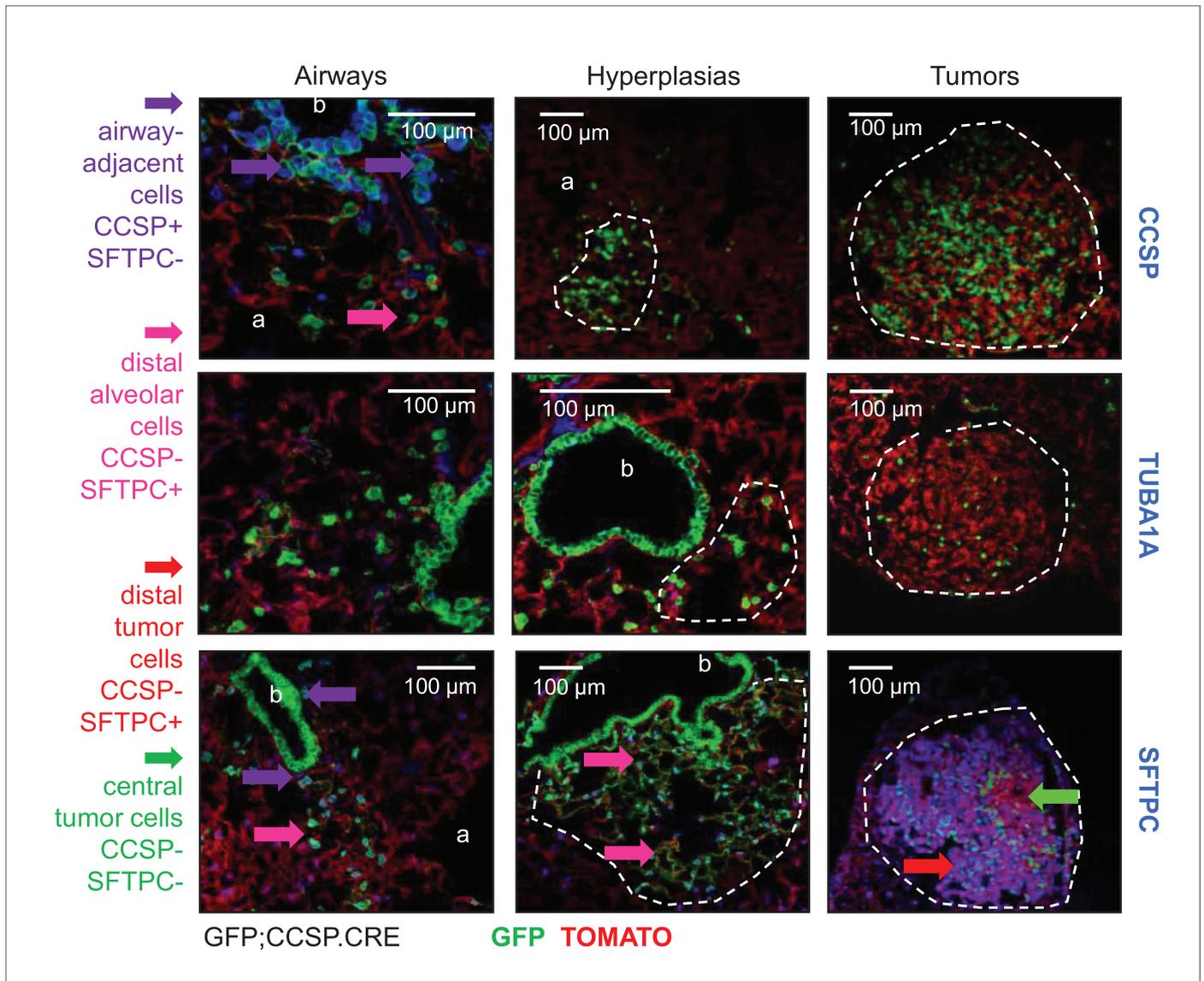
**Figure 3—figure supplement 2.** Airway-labeled cells in the alveoli of carcinogen-exposed C57BL/6 mice: data summary. Data summary (shown as violin plot) from GFP;CCSP.CRE mice treated as in **Figure 3A** ( $n = 10/\text{group}$ ).  $P$ , overall probability, one-way ANOVA. ns and \*\*:  $p > 0.05$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

DOI: <https://doi.org/10.7554/eLife.45571.035>



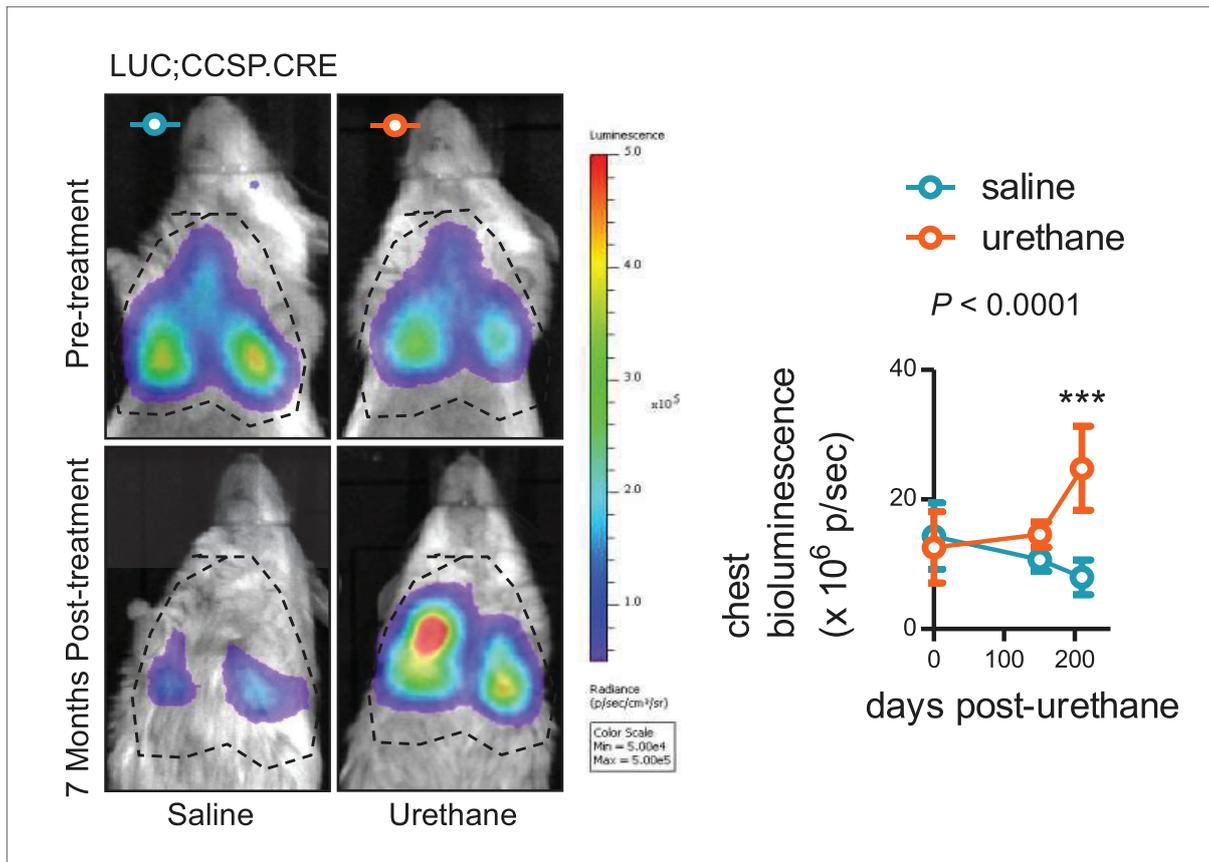
**Figure 3—figure supplement 3.** Airway-labeled cells in the alveoli of carcinogen-exposed mice express SFTPC. Single-channel images of non-neoplastic distal lung regions of urethane-treated GFP;CCSP.CRE mice at six months into treatment ( $n = 22$ ), stained for the lung cell markers Clara cell secretory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

DOI: <https://doi.org/10.7554/eLife.45571.036>



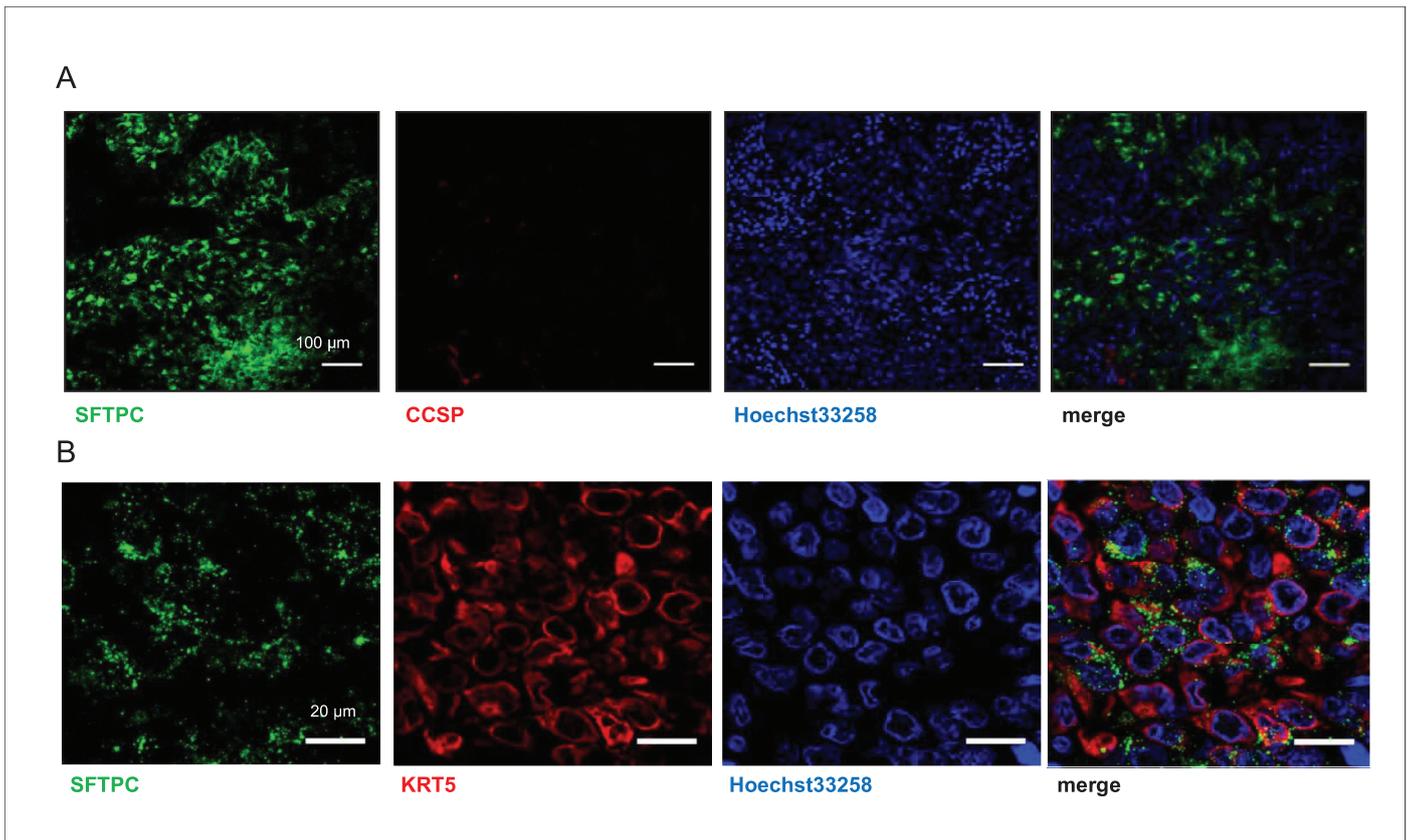
**Figure 3—figure supplement 4.** Airway-labeled cells in environmental-induced lung tumors express SFTPC. Juxtabronchial regions, alveolar hyperplasias, and tumors (dashed lines) of lungs from urethane-treated GFP;CCSP.CRE mice at six months into treatment ( $n = 22$ ) stained for lineage marker proteins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

DOI: <https://doi.org/10.7554/eLife.45571.037>



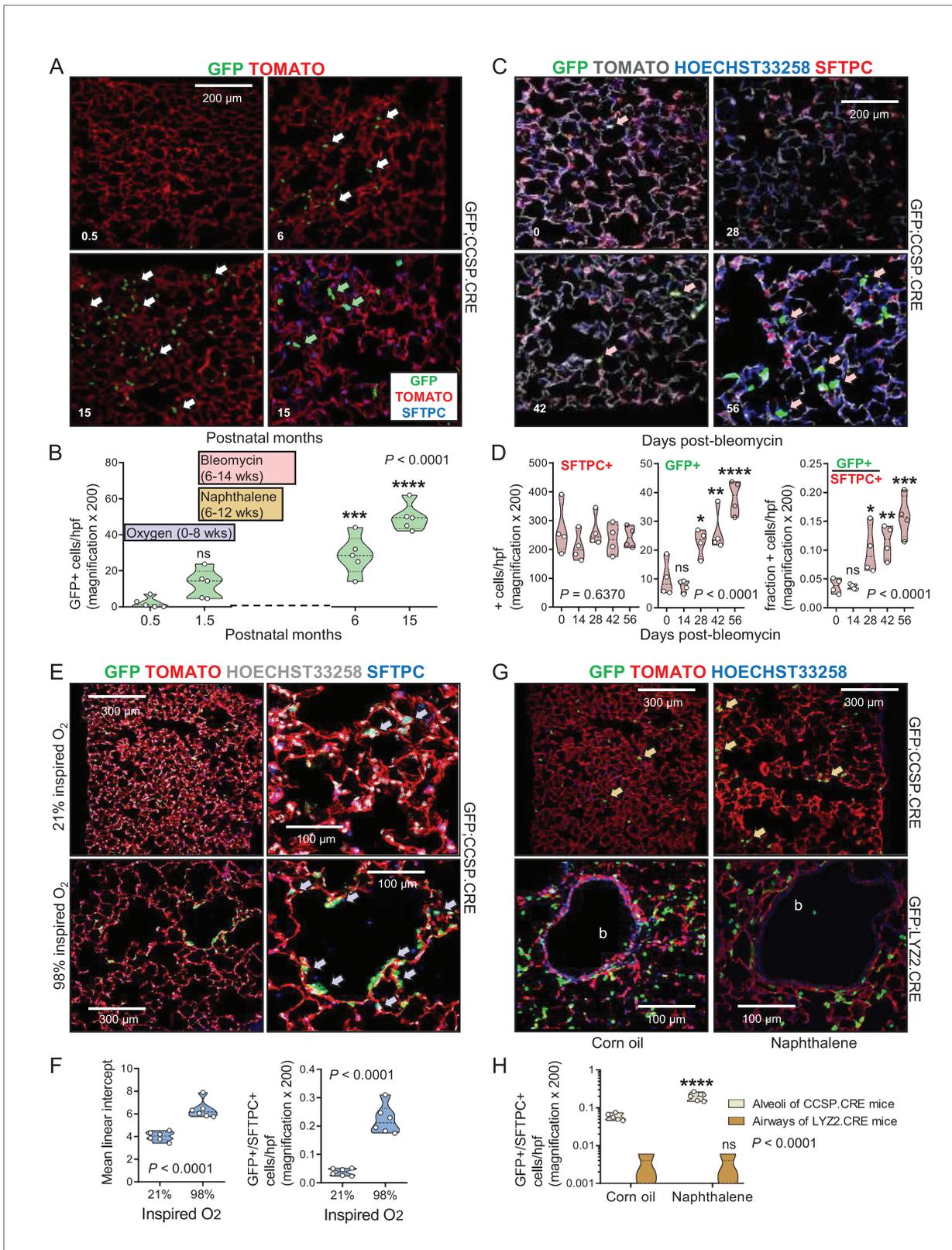
**Figure 3—figure supplement 5.** In vivo bioluminescent detection of the airway lineage in the lungs of saline- and carcinogen-treated mice. Representative merged bioluminescence/photographic images (left) and data summary (right) of LUC;CCSP.CRE mice (FVB background) before and seven months after saline (one intraperitoneal injection of 100  $\mu$ L;  $n = 6$ ) or urethane (one intraperitoneal injection of 1 g/Kg in 100  $\mu$ L 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  $\pm$  SD.  $P$ , overall probability, two-way ANOVA. \*\*\*:  $p < 0.001$  for comparison with saline, Bonferroni post-test.

DOI: <https://doi.org/10.7554/eLife.45571.038>



**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.

DOI: <https://doi.org/10.7554/eLife.45571.039>

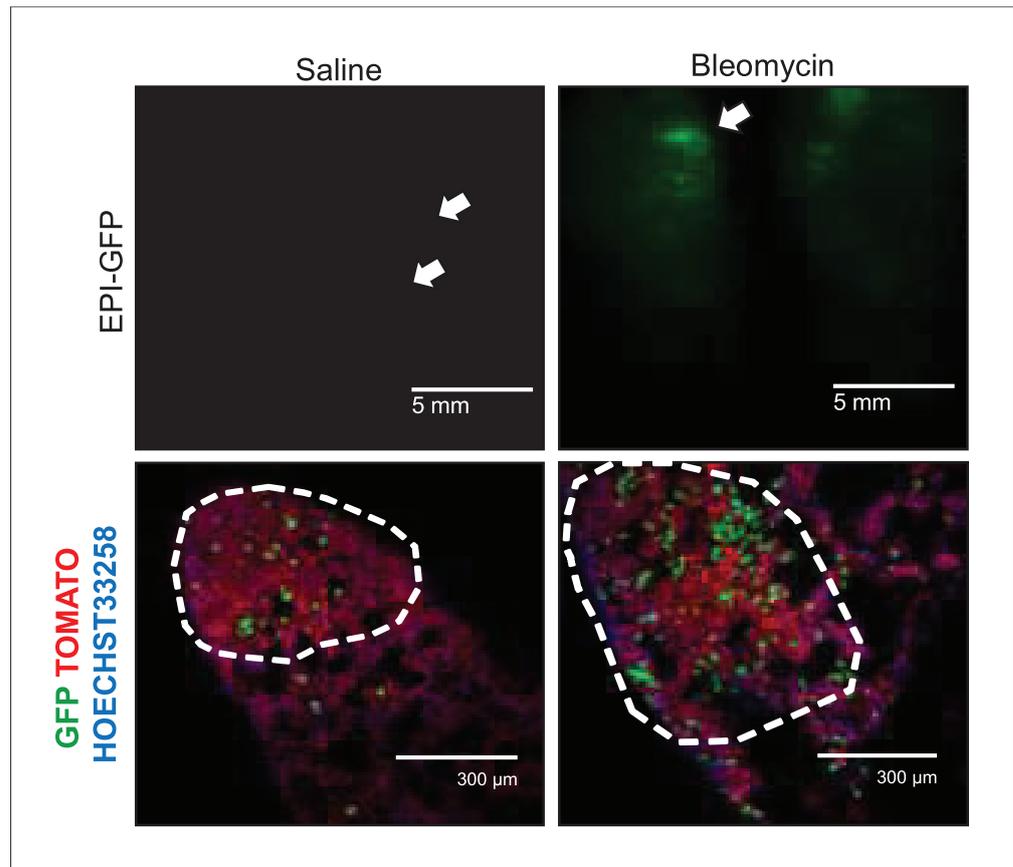


**Figure 4.** Airway cells in alveolar repair. (A) Non-neoplastic alveolar regions from lung sections of aging GFP;CCSP:CRE mice (bottom right section is also SFTPC-immunostained) show increasing numbers of alveolar GFP-labeled cells with age (arrows). Green arrows: genetically GFP-labeled, SFTPC-  
 Figure 4 continued on next page

## Figure 4 continued

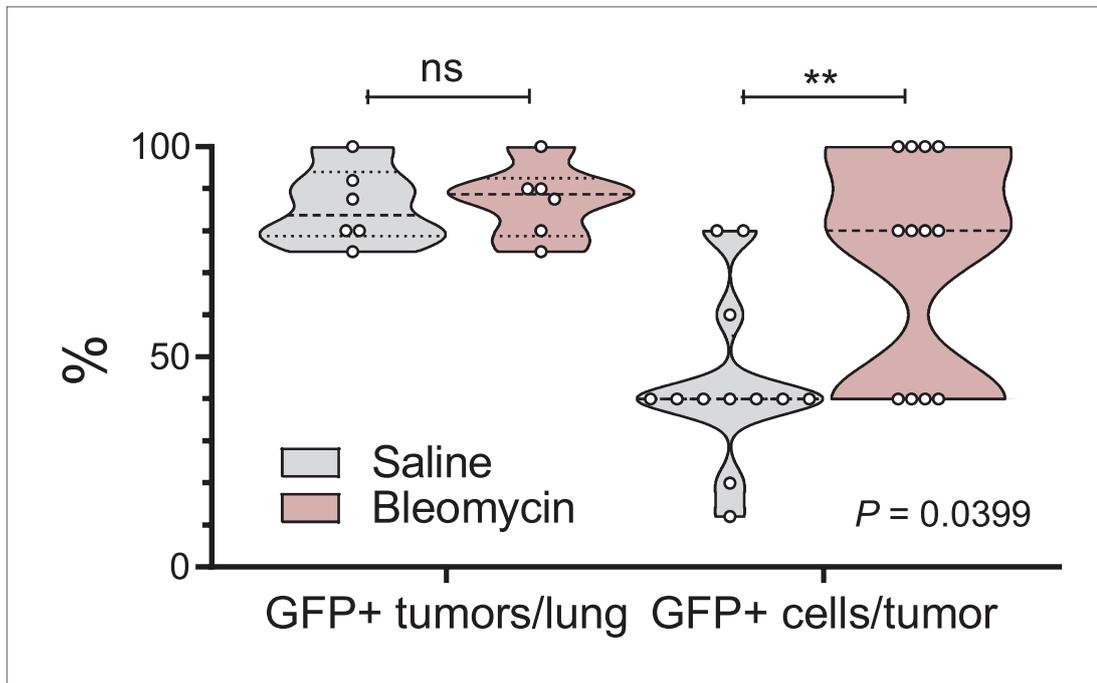
immunoreactive airway cells in alveolus of 15-month-old GFP;CCSP.CRE mouse. (B) Data summary ( $n = 5$  mice/time-point) from (A) shown as violin plot. Color-coded boxes indicate time windows of experiments in (C–H).  $P$ , probability, one-way ANOVA. ns, \*\*\*, and \*\*\*\*:  $p > 0.05$ ,  $p < 0.001$ , 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). See also **Figure 4—figure supplement 1** and **Figure 4—figure supplement 2**. (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%  $O_2$ . (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** and **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.

DOI: <https://doi.org/10.7554/eLife.45571.042>



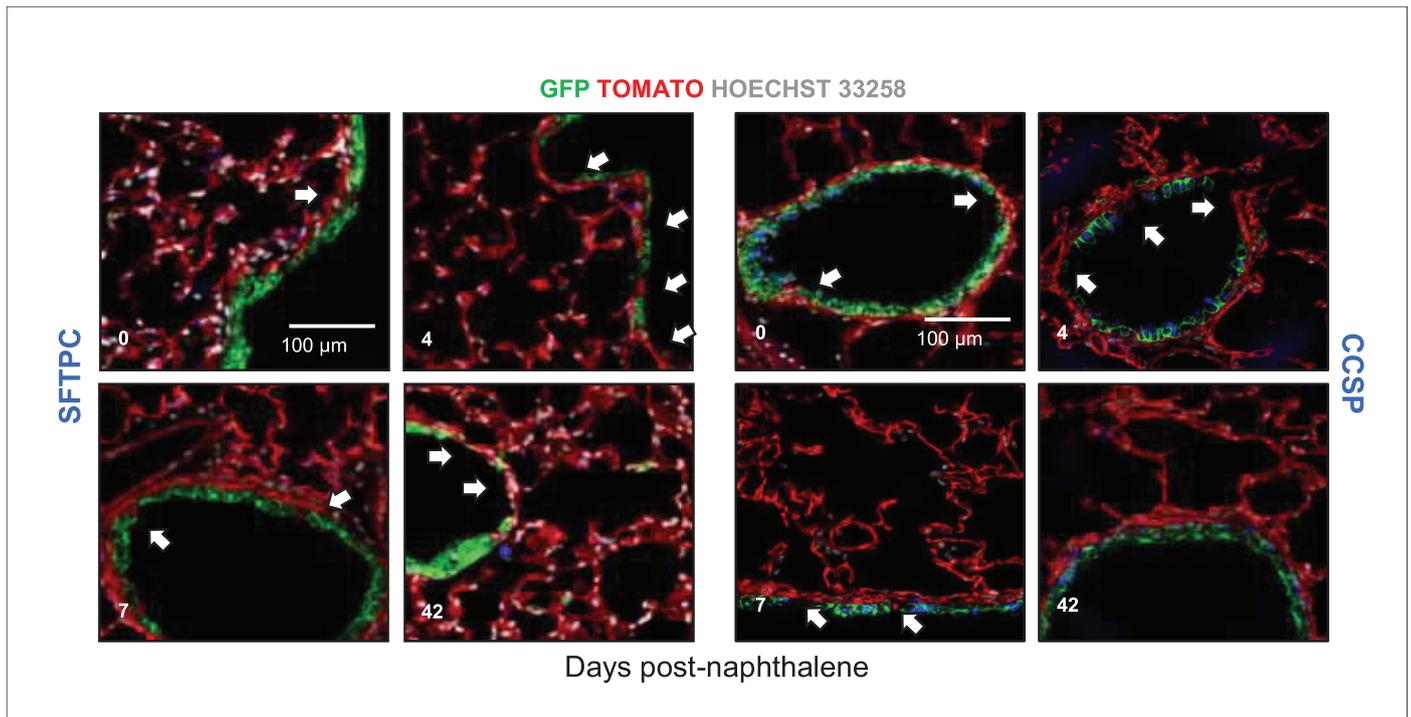
**Figure 4—figure supplement 1.** Alveolar type II cell ablation using bleomycin pre-treatment increases airway-labeled cells in urethane-induced lung tumors: representative images. Representative epifluorescence (top) and merged fluorescent microscopy (bottom) images of tumor-bearing lungs and lung 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.

DOI: <https://doi.org/10.7554/eLife.45571.043>



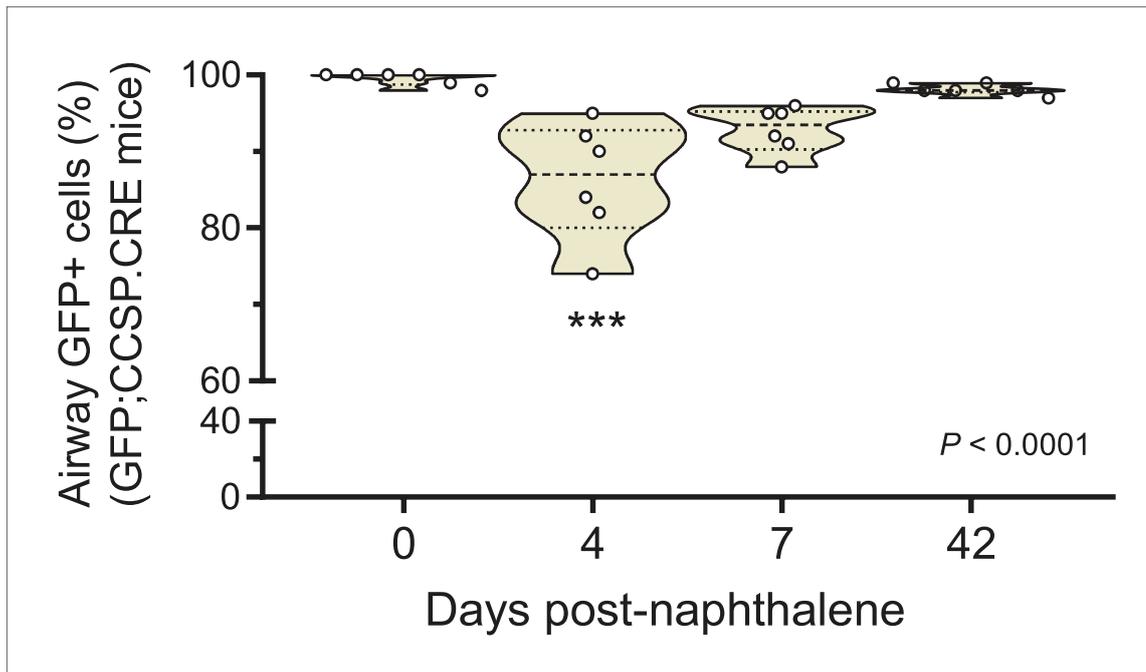
**Figure 4—figure supplement 2.** Alveolar type II cell ablation using bleomycin pre-treatment increases airway-labeled cells in urethane-induced lung tumors: data summary. Violin plot of GFP-labeled tumors/mouse ( $n = 6$  mice/group) and GFP-labeled cells/tumor ( $n = 12$  tumors/group;  $n = 2$  tumors/mouse 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.

DOI: <https://doi.org/10.7554/eLife.45571.044>



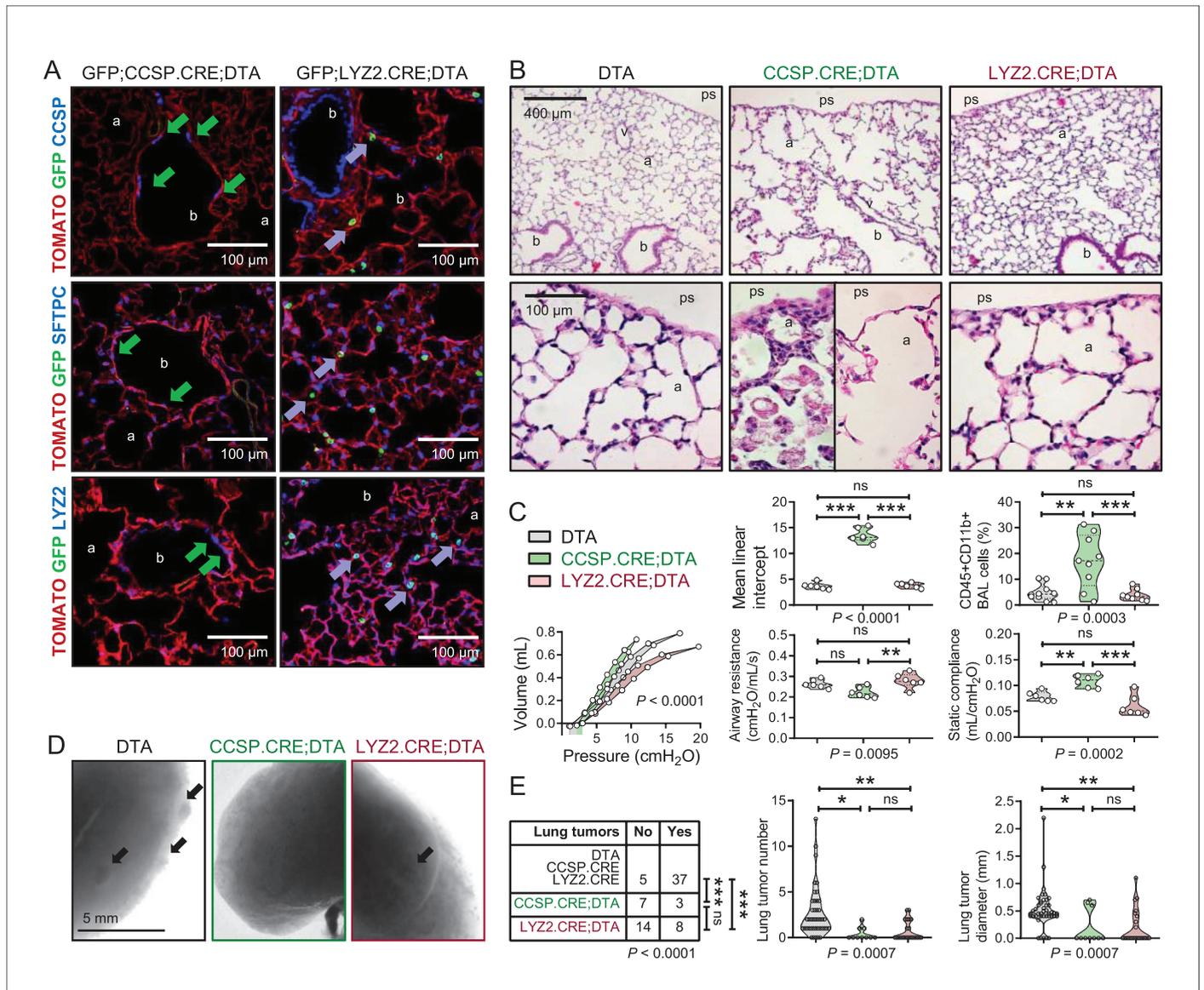
**Figure 4—figure supplement 3.** Airway epithelial cell ablation using naphthalene is restored by airway-labeled cells: representative images. Representative fluorescent microscopic images of lungs of GFP;CCSP.CRE mice at different time-points after intraperitoneal injection of 250 mg/Kg naphthalene 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.

DOI: <https://doi.org/10.7554/eLife.45571.045>



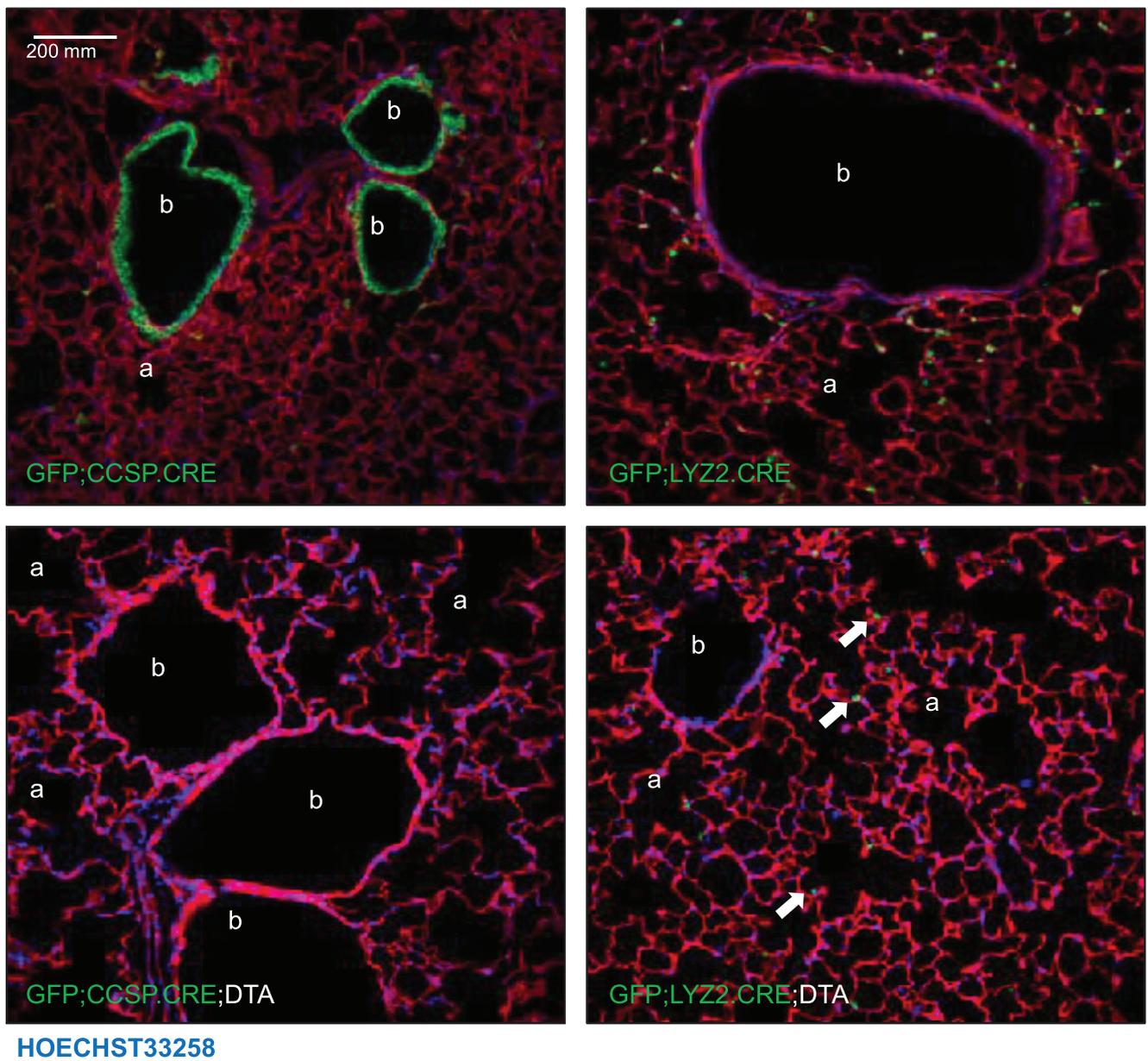
**Figure 4—figure supplement 4.** Airway epithelial cell ablation by naphthalene: data summary. Violin plot of percentage of GFP-labeled airway cells from experiment described in **Figure 4—figure supplement 3** ( $n = 6$  mice/time-point).  $P$ , overall probability, one-way ANOVA. \*\*\*:  $p < 0.001$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

DOI: <https://doi.org/10.7554/eLife.45571.046>



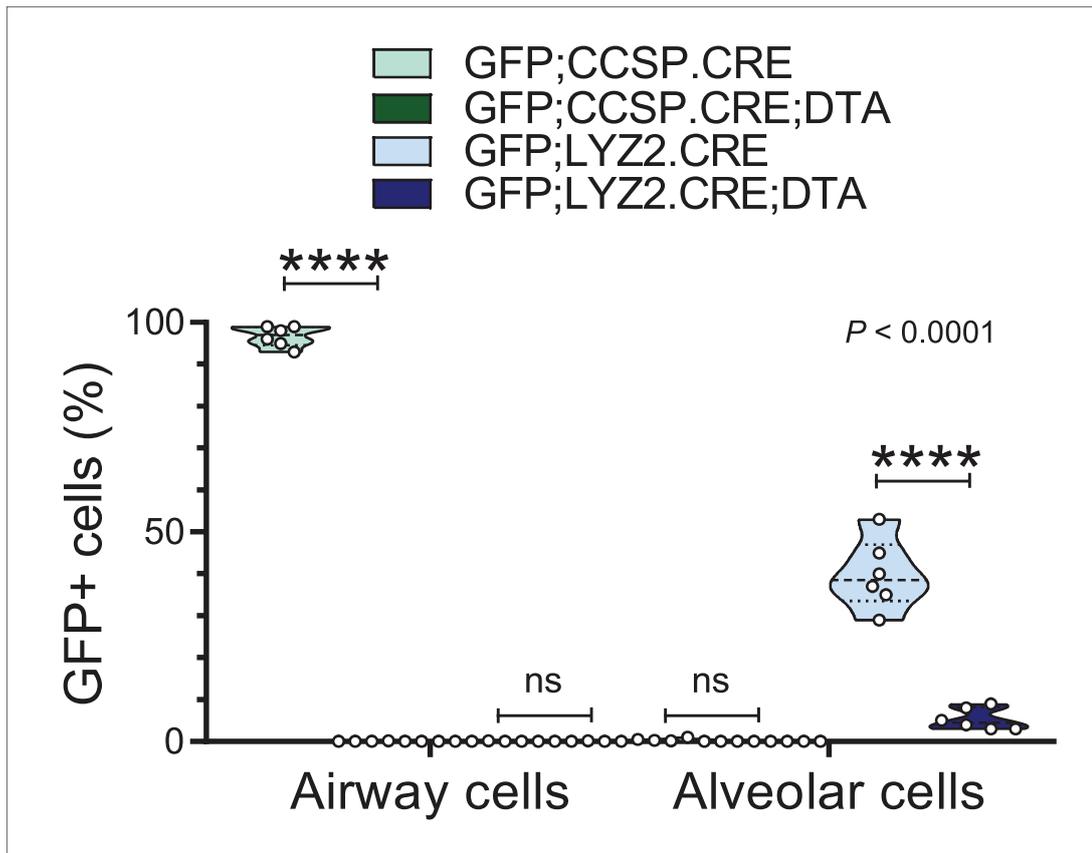
**Figure 5.** Airway cell-ablated mice display alveolar destruction and are protected from carcinogenesis. (A) Lineage marker-immunostained lung sections of 12-week-old GFP;CCSP.CRE;DTA and GFP;LYZ2.CRE;DTA mice ( $n = 6$ /group) show increased bronchial and alveolar size and flat CCSP + SFTPC + LYZ2+ 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** and **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,  $\chi^2$ -test (table) and one-way ANOVA (graphs). ns, \*, \*\*, and \*\*\*:  $p > 0.05$ ,  $p < 0.01$ , 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.

DOI: <https://doi.org/10.7554/eLife.45571.053>



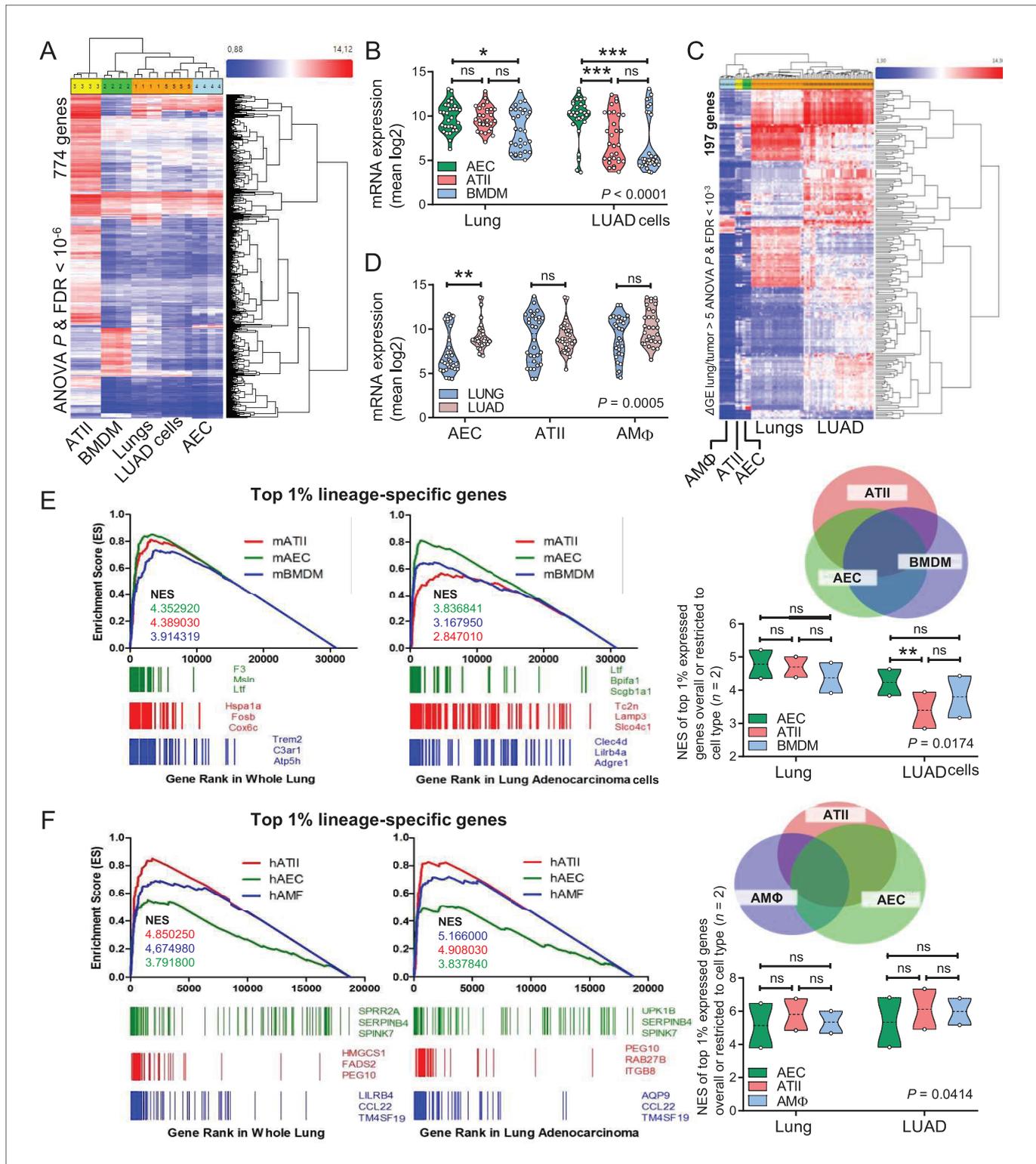
**Figure 5—figure supplement 1.** Triple transgenic mouse models for validation of genetic pulmonary lineage ablation: representative images. Representative lung sections of 12-week-old GFP;CCSP.CRE, GFP;LYZ2.CRE, GFP;CCSP.CRE;DTA, and GFP;LYZ2.CRE;DTA mice ( $n = 6/\text{group}$ ). Shown are merges 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.

DOI: <https://doi.org/10.7554/eLife.45571.054>



**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** ( $n = 6/\text{group}$ ). 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.

DOI: <https://doi.org/10.7554/eLife.45571.055>



**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 lung and LUAD cells. (C) Heat map of genes significantly differentially expressed (overall ANOVA and FDR  $p < 10^{-3}$ ) shows accurate hierarchical clustering. (D) Expression of the 30 top-represented transcripts of AEC, ATII, and AM $\phi$  in lung and LUAD cells. (E) Gene Rank plots for top 1% lineage-specific genes in murine lung and LUAD cells. (F) Gene Rank plots for top 1% lineage-specific genes in human lung and LUAD cells. Venn diagrams and box plots show gene overlap and NES between cell types. *Figure 6 continued on next page*

Figure 6 continued

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 (AM $\Phi$ ;  $n = 9$ ) was analyzed by Affymetrix Human Gene ST1.0 microarrays. (C) Heat map of genes significantly differentially expressed ( $\Delta GE > 5$  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 AM $\Phi$  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/AM $\Phi$  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/AM $\Phi$ ) signature compared with lung (nominal  $p < 0.0001$  for all, family-wise error rates FWER  $< 0.01$ ). Gene symbols indicate the top three lagging genes from each signature and shows loss of *Scgb1a1* (encoding CCSP) by LUAD. See also **Figure 6—figure supplements 5 and 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.

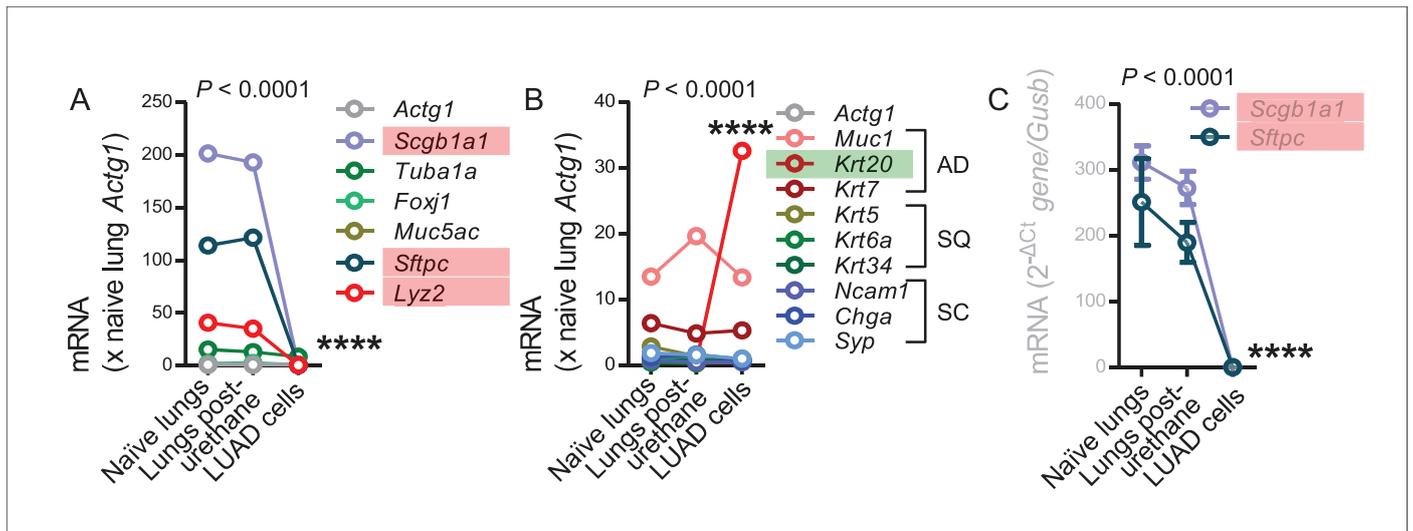
DOI: <https://doi.org/10.7554/eLife.45571.059>

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

$P = 0.1168$

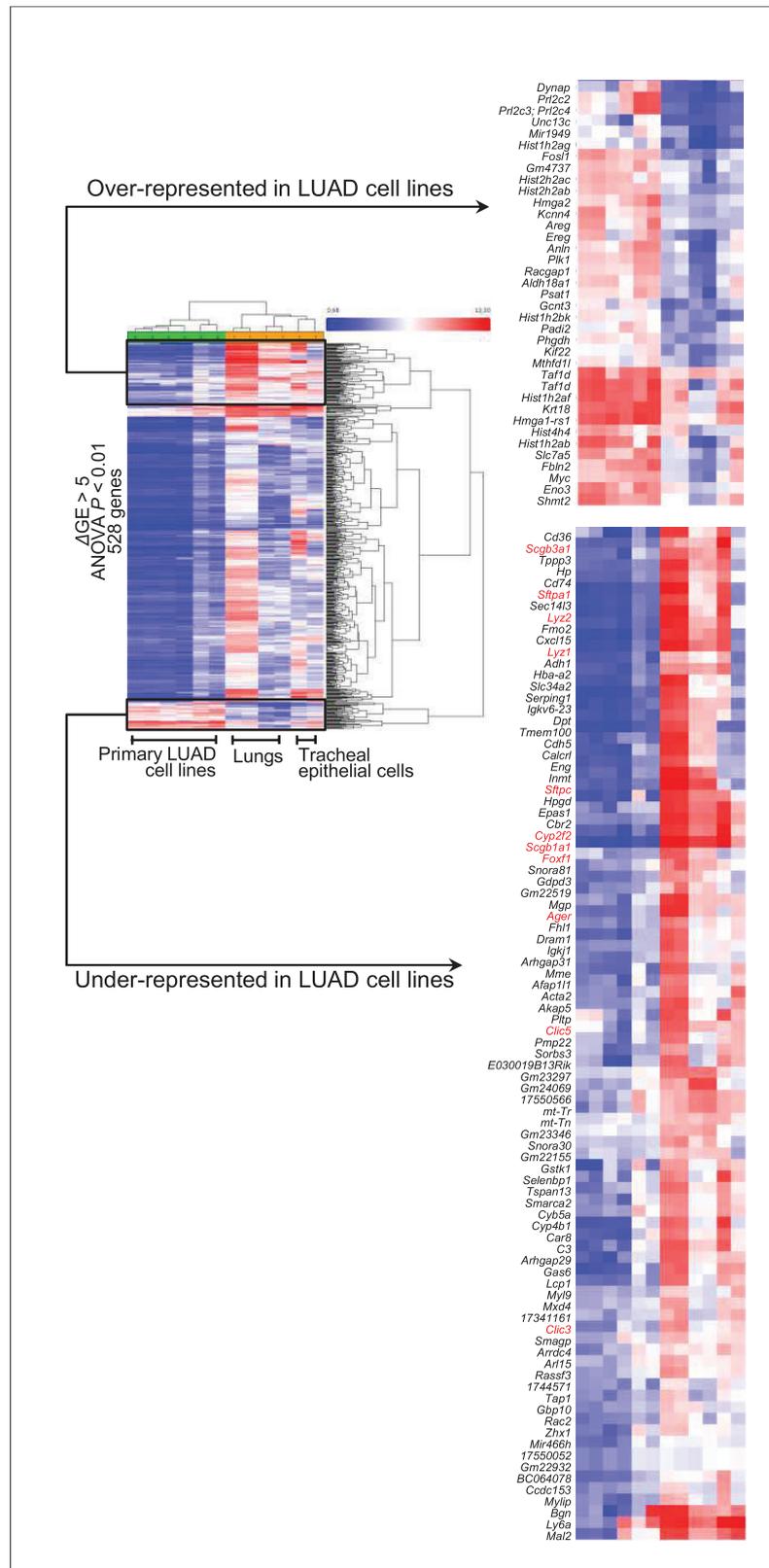
**Figure 6—figure supplement 1.** Lineage-specific gene expression in mouse lung adenocarcinoma cell lines induced by urethane compared with mouse lungs. RNA of mouse urethane-induced lung adenocarcinoma (LUAD) cell lines, lungs obtained pre- and one week post-urethane treatment, and airway epithelial cells (AEC), alveolar type II cells (ATII), and bone marrow-derived macrophages (BMDM) was examined by Affymetrix Mouse Gene ST2.0 microarrays ( $n = 4/\text{group}$ ). 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.

DOI: <https://doi.org/10.7554/eLife.45571.060>



**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**,  $n = 2/\text{group}$ ; C, qPCR,  $n = 3/\text{group}$ ). 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).

DOI: <https://doi.org/10.7554/eLife.45571.061>

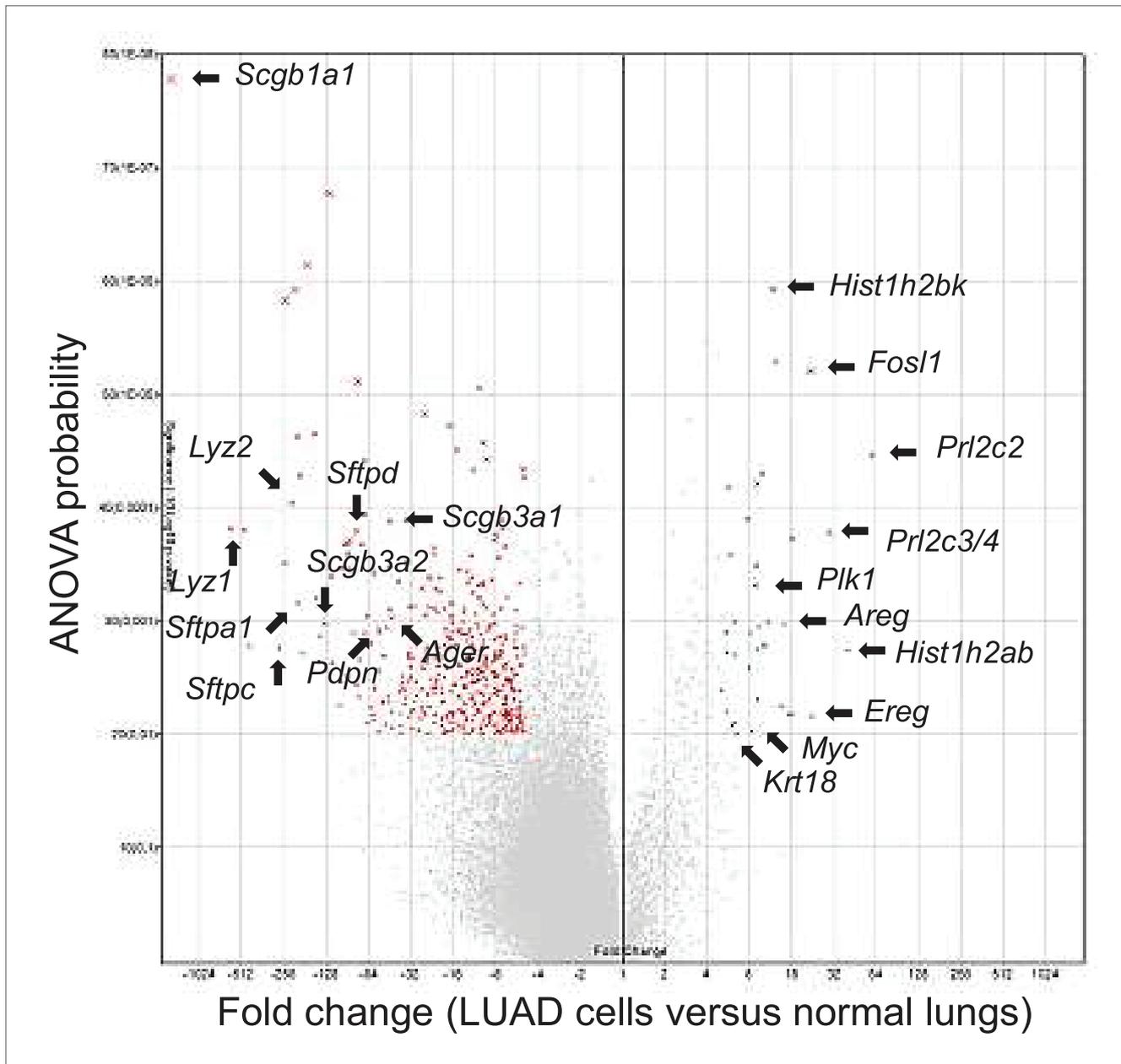


**Figure 6—figure supplement 3.** Loss of lineage marker expression in mouse lung adenocarcinoma cell lines induced by urethane compared with mouse lungs: heat maps. 528 genes differentially expressed between six Figure 6—figure supplement 3 continued on next page

*Figure 6—figure supplement 3 continued*

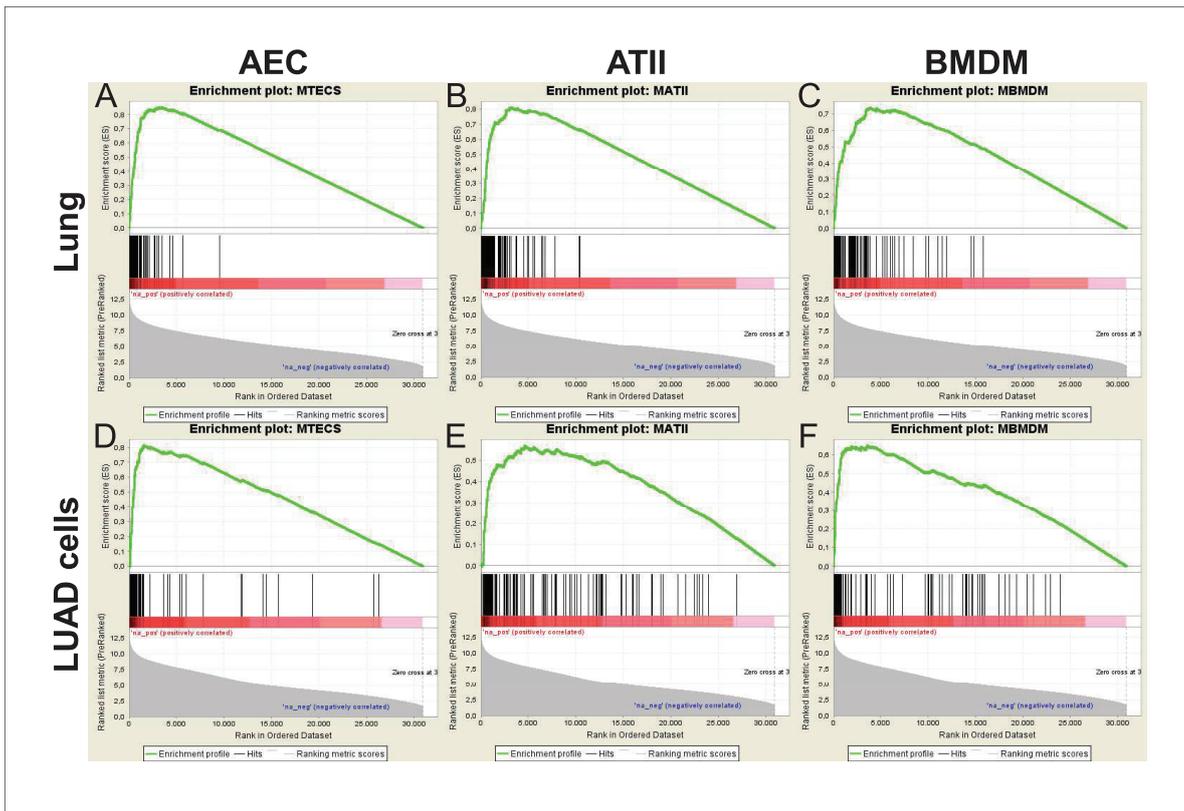
different lung adenocarcinoma cell lines cultured from urethane-induced lung tumors and six benign respiratory mouse samples, including lungs of saline- and urethane-treated mice obtained at one week post-treatment, as well as primary mouse tracheal epithelial cells using the cut-offs indicated. Whole 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.

DOI: <https://doi.org/10.7554/eLife.45571.062>



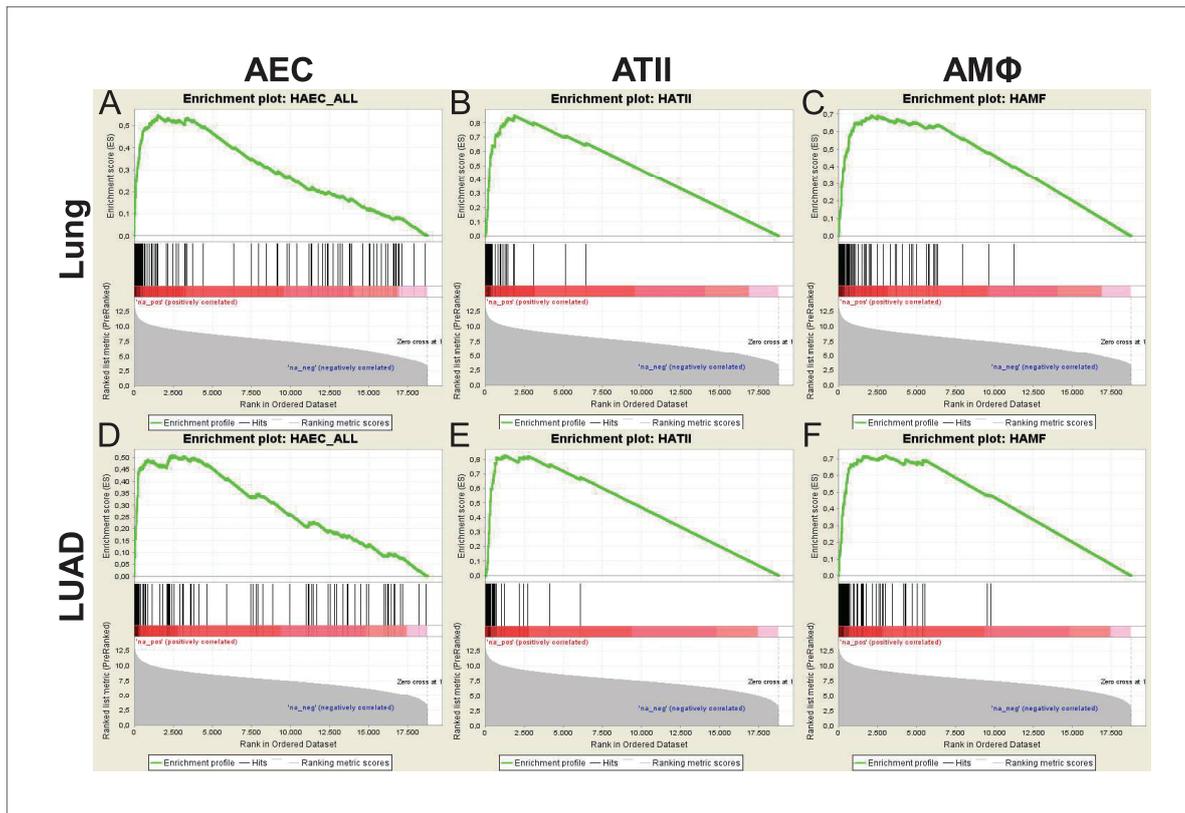
**Figure 6—figure supplement 4.** Loss of lineage marker expression in mouse lung adenocarcinoma cell lines induced by urethane compared with mouse lungs: volcano plot. Shown are selected top over- and under-represented genes (arrows) from microarrays from **Figure 6—figure supplement 2**.

DOI: <https://doi.org/10.7554/eLife.45571.063>



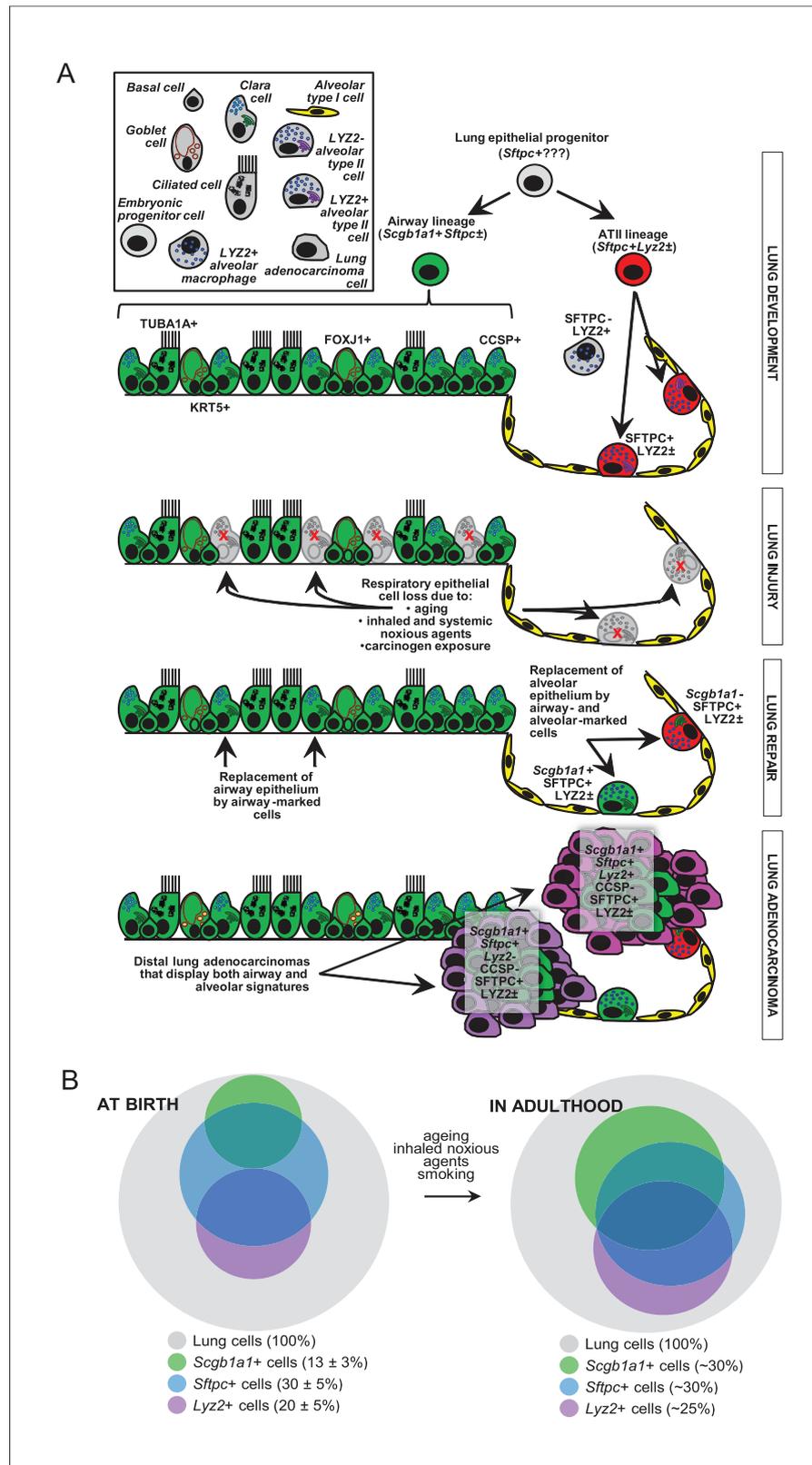
**Figure 6—figure supplement 5.** Mouse gene set enrichment analyses. Shown are gene set enrichment analyses of airway epithelial cell (AEC), alveolar type II cell (ATII), and bone marrow-derived macrophage (BMDM) transcriptome signatures in mouse lungs (top) and urethane-induced lung adenocarcinoma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

DOI: <https://doi.org/10.7554/eLife.45571.064>



**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 (AM $\Phi$ ;  $n = 9$ ) were cross-examined. Shown are gene set enrichment analyses of AEC, ATII, and AM $\Phi$  signatures in lung (top) and LUAD (bottom) transcriptomes. The data were used to design **Figure 6F**.

DOI: <https://doi.org/10.7554/eLife.45571.065>



**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 cells. (B) Venn diagrams showing the overlap of cell populations in the lung at birth and in adulthood. *Figure 7 continued on next page*

*Figure 7 continued*

II (ATII) cells, notwithstanding their common descent from an early (possibly *Sftpc*<sup>+</sup>) lung epithelial progenitor. The developmental airway lineage (*Scgb1a1*<sup>+</sup> *Sftpc*<sup>±</sup>; green) gives rise to all types of airway cells, including club, ciliated, goblet, basal, and other cells, while the developmental ATII lineage (*Sftpc*<sup>+</sup> *Lyz2*<sup>±</sup>; 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*<sup>+</sup> *Sftpc*<sup>±</sup> marked cells ensues. Upon lung injury, this process is accelerated. Similarly, during carcinogenesis caused by chemical tobacco smoke carcinogens, *Scgb1a1*<sup>+</sup> *Sftpc*<sup>±</sup> 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*<sup>+</sup> 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  $\alpha$ -tubulin.

DOI: <https://doi.org/10.7554/eLife.45571.071>

## 4. Conclusion

From the publications included in this dissertation it becomes evident that a broad range of mouse and *in vitro* models was implemented to study carcinogen-induced alterations in the lungs. It was demonstrated how chemical carcinogenesis can be used to deepen our current understanding of LUAD. Moreover, a new version of a cutting-edge technology was implemented (ddPCR<sup>Adv</sup>) and adapted to mouse models of tobacco-induced LUAD.

The generation of seven carcinogen-induced LUAD cell lines from carcinogen-induced tumors in different mouse strains was presented in the first publication. These cell lines represent a multitasking tool to study different aspects of smoking-induced LUAD since they are able to mimic the different aspects of tumors-initiating cells and to exhibit a fully malignant state. Furthermore, due to their gene expression profiles and mutational burden in key cancer genes overlapping with smoker's LUAD profiles, these cell lines epitomize a road map to study the different aberrations induced by tobacco carcinogens. As shown in the publication, these cell lines presented a unique gene expression profile that is not only displaying cancer related deregulation but also suggests the existence of specific carcinogen signatures. This possibility is examined in depth in a study still unpublished, in which our collection of cell lines was expanded to a larger number. We used different mouse strains and carcinogens, which were subsequently deployed in order to develop a new approach based on RNAseq. We were able to fingerprint, in this study, the carcinogen-induced mutation status and gene expression profiles of our cell lines.

In the second publication, the role of club cells in LUAD carcinogenesis was elucidated. Using different combinations of cell lineage labeling, it became possible to pinpoint the role of airway cells in LUAD development. The use of ddPCR<sup>Adv</sup> elucidated the persistence of *Kras* mutation in club cells during early steps of carcinogenesis. Furthermore, the combination of cell lineage labeling and ddPCR<sup>Adv</sup> permitted for the first time analyses of mutational processes in space and time on the single copy level. Consistent with the tumor initiation phase, where a mutation arises in a cell and is clonally expanded among its progeny, this approach allowed us to finely track the development of a mutation in a specific cell lineage and to quantitatively

assess its expansion over time. Expanding the spatial and longitudinal mapping from one single mutation to the full spectrum of carcinogen-induced mutations would require an experimental approach designed toward the massive parallel sequencing. Nevertheless, the approach described in Spella *et al* 2019 opens a targeted path for further explorations.

In conclusion, this dissertation combines different biological approaches designed to map and characterize tobacco carcinogen-induced alterations in the lungs. Importantly, the relative contributions of the candidate towards the two included publications were critical in showing the faithful mimicry of smokers' LUAD by our carcinogen-induced cell lines in the first manuscript, and in proving the early persistence of *Kras* mutations specifically in airway epithelial cells in the second manuscript.

## 5. References

1. Bray, F. *et al.* Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* **68**, 394–424 (2018).
2. Fossum, C. C. *et al.* Characterization of the oropharynx: anatomy, histology, immunology, squamous cell carcinoma and surgical resection. *Histopathology* **70**, 1021–1029 (2017).
3. Travis, W. D. *et al.* The 2015 World Health Organization Classification of Lung Tumors: Impact of Genetic, Clinical and Radiologic Advances Since the 2004 Classification. *J Thorac Oncol.* **10**, 1243–1260 (2015).
4. Weissferdt, A. Large cell carcinoma of lung: On the verge of extinction? *Semin Diagn Pathol.* **31**, 278–288 (2014).
5. Lu, T. *et al.* Trends in the incidence, treatment, and survival of patients with lung cancer in the last four decades. *Cancer Manag Res.* **11**, 943–953 (2019).
6. Collison, E. *et al.* Comprehensive genomic characterization of squamous cell lung cancers. *Nature* **489**, 519–525 (2012).
7. Imielinski, M. *et al.* Mapping the Hallmarks of Lung Adenocarcinoma with Massively Parallel Sequencing. *Cell* **150**, 1107–1120 (2012).
8. Network, T. C. G. A. R. Comprehensive molecular profiling of lung adenocarcinoma. *Nature* **511**, 543–550 (2014).
9. Alexandrov, L. B. *et al.* Mutational signatures associated with tobacco smoking in human cancer. *Science* **354**, 618–622 (2016).
10. Malhotra, J. *et al.* Risk factors for lung cancer worldwide. *Eur Respir J.* **48**, 889–902 (2016).
11. Hecht, S. S. Tobacco Smoke Carcinogens and Lung Cancer. *J Natl Cancer Inst.* **91**, 1194–1210 (1999).
12. Klotz, L.V. *et al.* Comprehensive clinical profiling of the Gauting locoregional lung adenocarcinoma donors. *Cancer Med.* **8**, 1486–1499 (2019).
13. Campbell, J.D. *et al.* Distinct patterns of somatic genome alterations in lung adenocarcinomas and squamous cell carcinomas. *Nat Genet.* **48**, 607–616 (2016).
14. Network, T. C. G. A. R. Pan-cancer analysis of whole genomes. *Nature* **578**, 82–93 (2020).

15. Bailey, H.M. *et al.* Comprehensive Characterization of Cancer Driver Genes and Mutations. *Cell* **173**, 371-385 (2018).
16. Alexandrov, L. B. *et al.* Signatures of mutational processes in human cancer. *Nature* **500**, 415–421 (2013).
17. Shlien, A. *et al.* Direct Transcriptional Consequences of Somatic Mutation in Breast Cancer. *Cell Rep* **16**, 2032-2046 (2016).
18. Westcott, P. M. K. *et al.* The mutational landscapes of genetic and chemical models of Kras-driven lung cancer. *Nature* **517**, 489–492 (2015).
19. Martincorena, I. & Campbell, P. J. Somatic mutation in cancer and normal cells. *Science* **349**, 1483 (2015).
20. Zhang, J. *et al.* Intratumor heterogeneity in localized lung adenocarcinomas delineated by multiregion sequencing. *Science* **346**, 256-259 (2014).
21. Barbara, Y. *et al.* *Wheater's functional histology: a text and colour atlas.* Churchill Livingstone (2006).
22. Stanke, F. The Contribution of the Airway Epithelial Cell to Host Defense. *Mediators Inflamm* **2015**, 463016 (2015).
23. Kim, C. F. B. *et al.* Identification of Bronchioalveolar Stem Cells in Normal Lung and Lung Cancer. *Cell* **121**, 823–835 (2005).
24. Kwon, M.C., Berns A. Mouse models for lung cancer. *Mol Oncol.* **7**, 165-177 (2013).
25. Guerra, C *et al.* Tumor induction by an endogenous K-ras oncogene is highly dependent on cellular context. *Cancer Cell.* **4**, 111-120 (2003).
26. Jackson, E.L. *et al.* Analysis of lung tumor initiation and progression using conditional expression of oncogenic K-ras. *Genes Dev.***15**, 3243-3248 (2001).
27. Prior, I.A. *et al.* The frequency of Ras mutations in cancer. *Cancer Res.* **80**, 2969-2974 (2020).
28. O'Hagan, R.C, Heyer J. KRAS Mouse Models: Modeling Cancer Harboring KRAS Mutations. *Genes Cancer.* **3**, 335-343 (2011).
29. Desai, T. J., Brownfield, D. G. &Krasnow, M. A. Alveolar progenitor and stem cells in lung development, renewal and cancer. *Nature* **507**, 190-194 (2014).

30. Vreka, M et al. I $\kappa$ B Kinase  $\alpha$  Is Required for Development and Progression of KRAS-Mutant Lung Adenocarcinoma. *Cancer Res* 78, 2939-2951 (2018).
31. Agalioti, T. et al. Mutant KRAS promotes malignant pleural effusion formation. *Nat Commun.* 8, 15205 (2017).
32. Sutherland, K. D. et al. Multiple cells-of-origin of mutant K-Ras–induced mouse lung adenocarcinoma. *PNAS* 111, 4952-4957 (2014).
33. Govindan, R. et al. Genomic Landscape of Non-Small Cell Lung Cancer in Smokers and Never-Smokers. *Cell* 150, 1121-1134 (2012).
34. de Seranno, S., Meuwissen, R. Progress and applications of mouse models for human lung cancer. *Eur. Respir. J.*35, 426-443 (2010).
35. Kabbout, M. et al. ETS2 mediated tumor suppressive function and MET oncogene inhibition in human non-small cell lung cancer. *Clin Cancer Res* 19, 3383-3395 (2013).
36. Muzumdar, M. D., Tasic, B., Miyamichi, K., Li, L. & Luo, L. A global double-fluorescent Cre reporter mouse. *Genesis* 45, 593-605 (2007).
37. Mazaika, E. & Homsy, J. Digital Droplet PCR: CNV Analysis and Other Applications. *CurrProtoc Hum Genet* 82, 1-13 (2014).

## Appendix A

The ddPCR is a reliable and powerful method that allows detection and quantification of nucleic acids with exceptional sensitivity and high accuracy<sup>32</sup>. The ddPCR method is based on a single PCR reaction trapped inside an oil droplet and the usage of fluorescent labeled oligos (TaqMan® probe) to identify a specific DNA fragment. In general, the ddPCR reaction is set up as a duplex PCR, where one TaqMan® probe is designed to target the region of interest (ROI) and a second one for any standard reference fragment (REF). The two TaqMan® probes are always labeled with two different fluorophores to allow the differential detection (Figure 4).

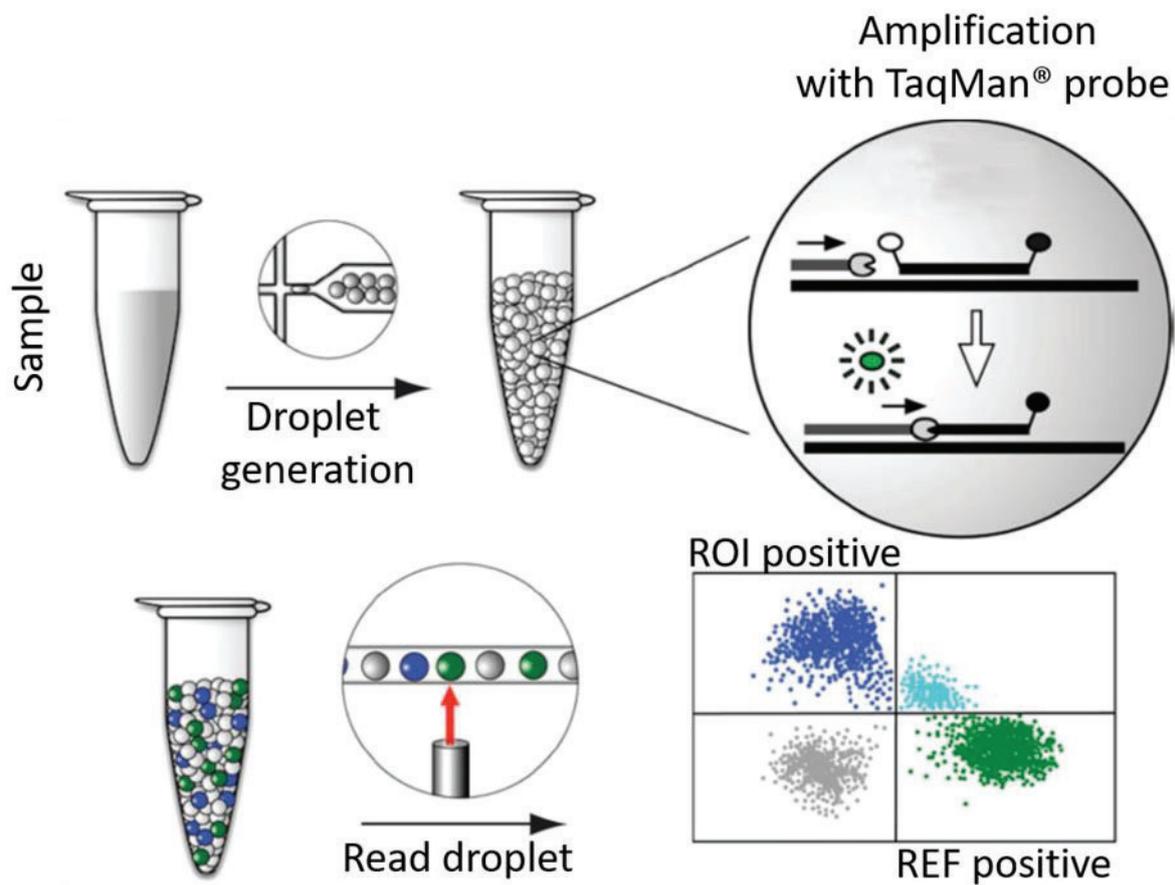


Figure 4: ddPCR workflow overview (adapted from Mazaika E. et al. 2011)

For our purposes, we wanted to overcome the technological limitation of two detections (one ROI and one REF), and be able to detect and quantify two different ROI with their relative REF, for a total of four different detections, in the same ddPCR reaction. For this aim, we had to tackle three major tasks:

- To perform ddPCR on a single genome copy
- To identify a ROI-REF couple with one single TaqMan® probe
- To quantify two different fragments emitting the same fluorescence wavelength

## Performing the ddPCR on a single genome copy

For preparing a ddPCR assay, the researcher needs to set up a PCR reaction (composed by DNA template material, buffer, dNTPs, primers, and DNA polymerase plus the two TaqMan® probes). The number of total droplets prepared per reaction is 20,000. To achieve filling every droplet with a single genome copy, the researcher needs to know the weight in grams of the specific species' genome s/he is analyzing. According to this principle, the sample weight should be equal to the weight of 20,000 genome copies. Despite preparing the sample according to the above, it is unlikely that a perfect distribution of one single genome copy per droplet can be achieved. To overcome this issue, I succeeded in defining the different fluorescence emitted from an empty droplet respective to a loaded one. This procedure was set up based on comparisons between a droplet loaded with the full ddPCR reaction and another missing the DNA template. In this way, I was able to quantify the specific fluorescence of a droplet missing the template and define the gating threshold. Combining this approach with the maximal number of droplets per reaction, I was able to precisely define the number of empty or loaded droplets (Figure 5).

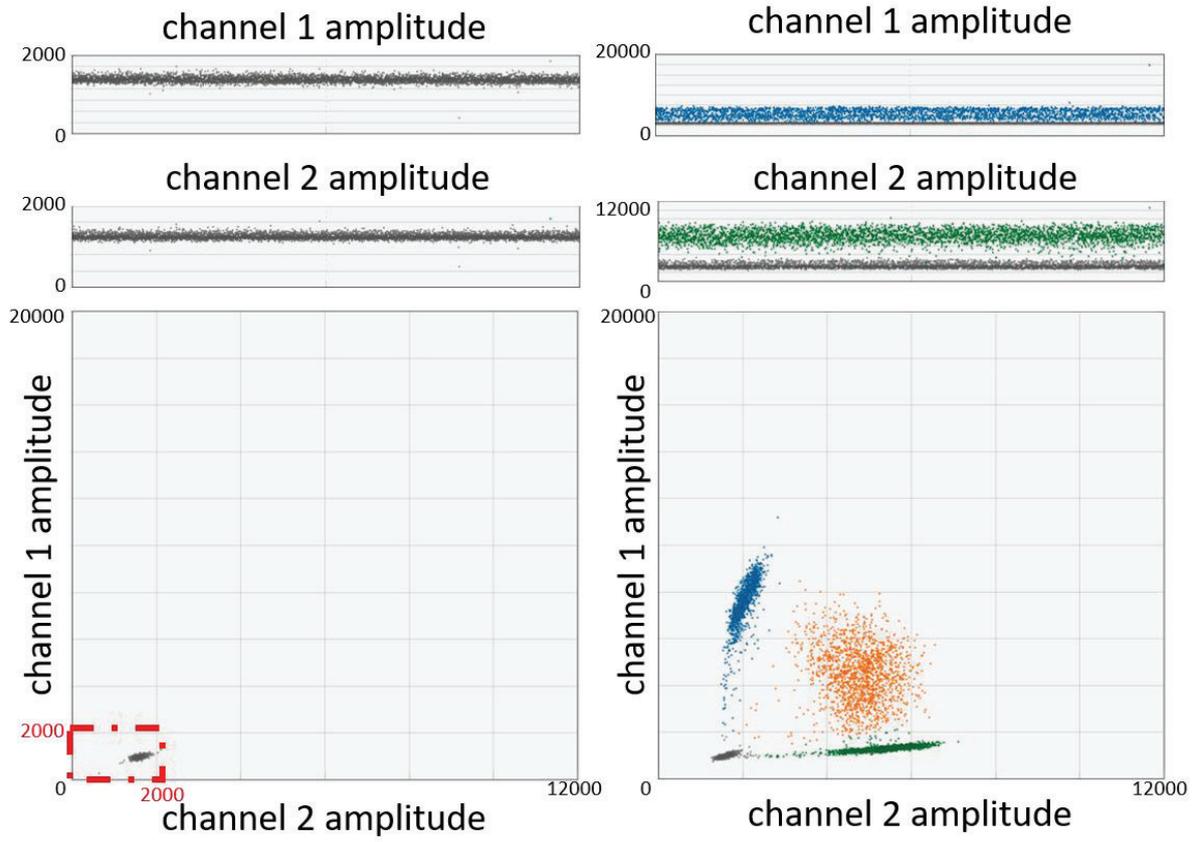


Figure 5: Empty droplet gating. On the left, a representative plot of no template sample is shown; on the right, a representative plot of a DNA-loaded sample is depicted. The red box shows the specific gating threshold used.

## Identifying a ROI-REF couple with one single TaqMan® probe

In general, a TaqMan® probe has the size and a melting temperature (depending on base composition and length of the probe) compatible with the primer set. The modulation of these two parameters is the key to allow a differential annealing of a TaqMan® probe to two different DNA fragments. Given the fact that the TaqMan® probe will anneal specifically to one fragment, the unspecific annealing to the second fragment relies on a sequence homology exceeding 85% with the unspecific target sequence and an annealing temperature of 2-4°C lower than the optimal one (depending on TaqMan® probe sequence composition). Combined with the droplet technology, this approach allows identification of a ROI-REF couple (Figure 6).

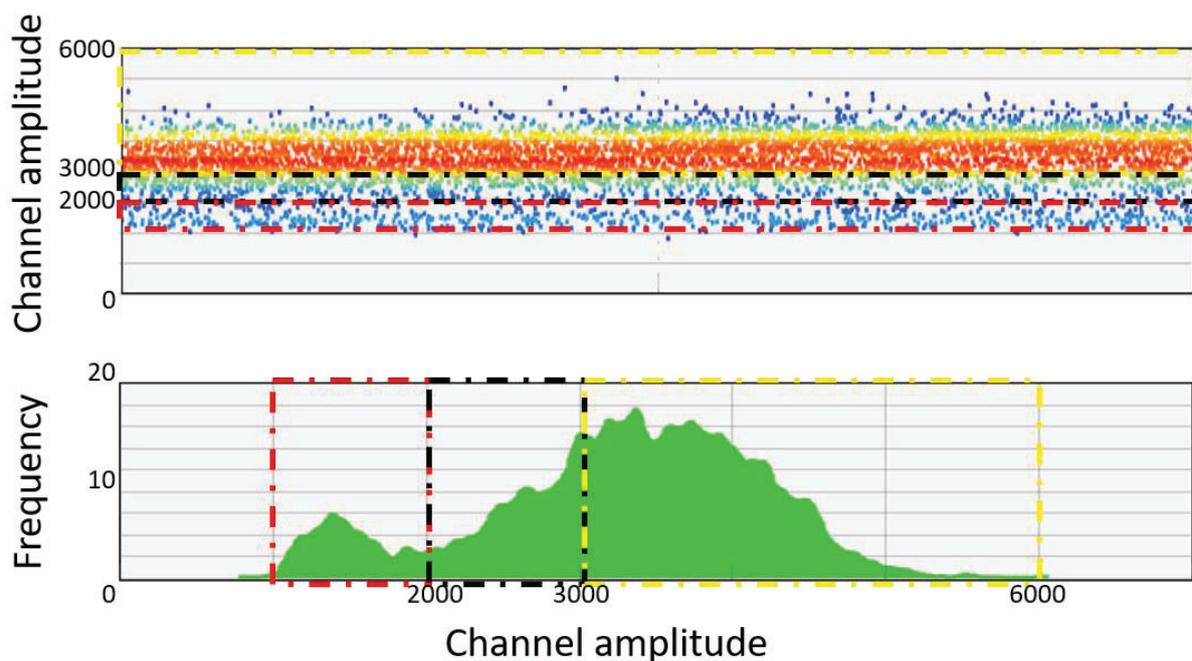


Figure 6: Identification and gating of different emission spectra of a TaqMan probe: shown on the top is the channel amplitude plot and in the bottom the droplet frequency plot. Boxes represent in red the empty droplet gating, in black the unspecific fluorescence gating, and in yellow the specific fluorescence gating.

When this approach is extended to two TaqMan® probes with different fluorophores, it becomes possible to identify four different fragments (two ROI-REF couples).

## Quantifying two different fragments emitting the same fluorescence

All ddPCR analysis software solutions are based on the quantification of fluorescent droplets. Although with such software I was able to identify two different fragments emitting the same fluorescence, quantification was not possible, due to the several combinations of fluorescence thresholds to examine (Figure 7).

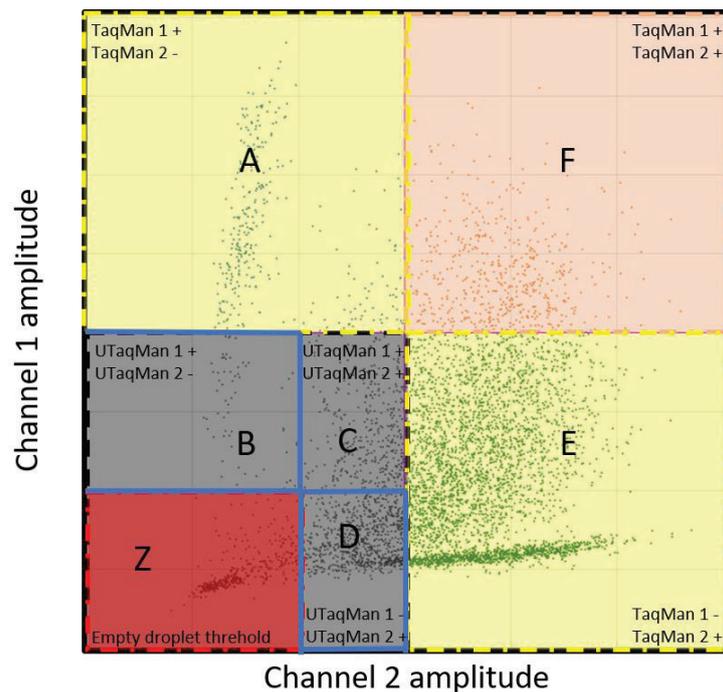


Figure 7: 2D amplitude plot depicting the different combinations of fluorescent droplets: every box with a different color and letter labelling represents a different threshold; unspecific (U), positive (+), and negative (-).

To analyze all different combinations of droplet fluorescence I developed an R language computational script. This script is based on the formula

$$n \text{ of positive droplets} = \frac{x}{(x + y) - Z}$$

Where:

x= n droplets positive for the first TaqMan® probe (specific and unspecific)

y= n droplets positive for the second TaqMan® probe (specific and unspecific)

Z= n of empty droplets

The possible combination of thresholds (x and y) were calculated using the formula

$$C_{n,k} = \frac{n!}{k!(n-k)!}$$

This approach allowed to specifically quantify the 7-specific combinations of fluorescence for each fragment (both REF and ROI).

With the implementations of all these approaches, I repurposed ddPCR for our specific aims. For further reference to this technically advanced ddPCR I will use ddPCRAdv.

## Spatial and longitudinal quantification of $Kras^{Q61R}$ mutation

After developing and testing the ddPCR<sup>Adv</sup>, I applied it to quantify the  $Kras^{Q61R}$  mutation over time and specifically detect in which cell lineage it aroused. For these purposes, we selected GFP;CCSP-CRE, and GFP;LYZ2-CRE mice on the FVB background, as sources for airway epithelial-labeled cells, and the second alveolar epithelial-labeled cells, respectively. Urethane was used to induce  $Kras^{Q61R}$  mutations. The injected mice were divided in two cohorts, one of which was harvested one week (GFP;CCSP-CRE,  $n=5$  and GFP;LYZ2-CRE,  $n=5$ ) after the injection and another that was sacrificed two weeks after the injection (GFP;CCSP-CRE,  $n=5$  and GFP;LYZ2-CRE,  $n=5$ ). The goal of this experiment was to detect simultaneously:

- $Kras$  status (ROI= $Kras^{Q61R}$  and REF= $Kras^{WildType}$ )
- mt/mg locus status (Rosa locus status) in CRE-reporter strain mt/TdTomato mice (ROI=mt and REF=mg) (Figure 5)

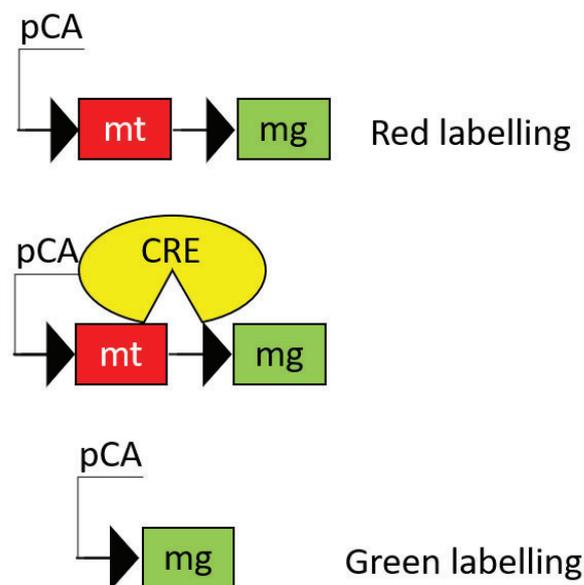


Figure 7: schematic diagram of the mt/mg locus: from top to bottom is shown the CRE recombinase action. Actin B core promoter (pCA), mt/Tomato (mt) in red, green fluorescent protein gene (mg) in green, Cre recombinase (CRE) in yellow. The figure is readapted from Muzumdar, M. D. et al 2007

For this ddPCR<sup>Adv</sup>, a specific TaqMan® probe was conceptualized and designed to detect mutations in the  $Kras$  gene (ROI-REF couple 1) and the cells in which it occurred (green or red labeled) (ROI-REF couple 2) (Spella M. et al 2019 materials and methods). In this way, a spatially and

quantitative definition of the *Kras* status could be obtained. Subsequently, I analyzed the trend over time finally allowing us to address the *Kras* status longitudinally.

## Acknowledgments

This Ph.D. thesis relies on travel of 5 years full of events and experiences. I still remember the happiness of the day I received the e-mail with the news that my interview was successful.

With this chapter in my thesis, I would like to thank you in the really beginning to my supervisor, Doctor Georgos Stathopoulos, and his group for hosting me in this period. Moreover, a special mention to Professor Silke Meiners for accepting to be my university supervisor.

On the personal side, I would like to thank you my family to be always of support to me with all possibilities they have and the help I needed. My friends that were there in the happy and sad moments of this period. My girlfriend, Dilyana, to push me to complete this path even when I was near to give up. Despite my weird and bad moods, she was always there cheering me up and doing the hard work that a life together brings.

A really unique mention goes to Doctor Anne-Sophie Lamort, with her working was fun and productive and same time. Moreover, we grow up a friendship that goes over the walls of a lab and lands in the personal life. We have spent a piece of life together that I will always keep in my hearth and even if now we are far, our friendship will never end.

Finally, I would like to thank all the people that crossed my path and in different ways and left a sign in this experience, and let me grow professionally and personally.