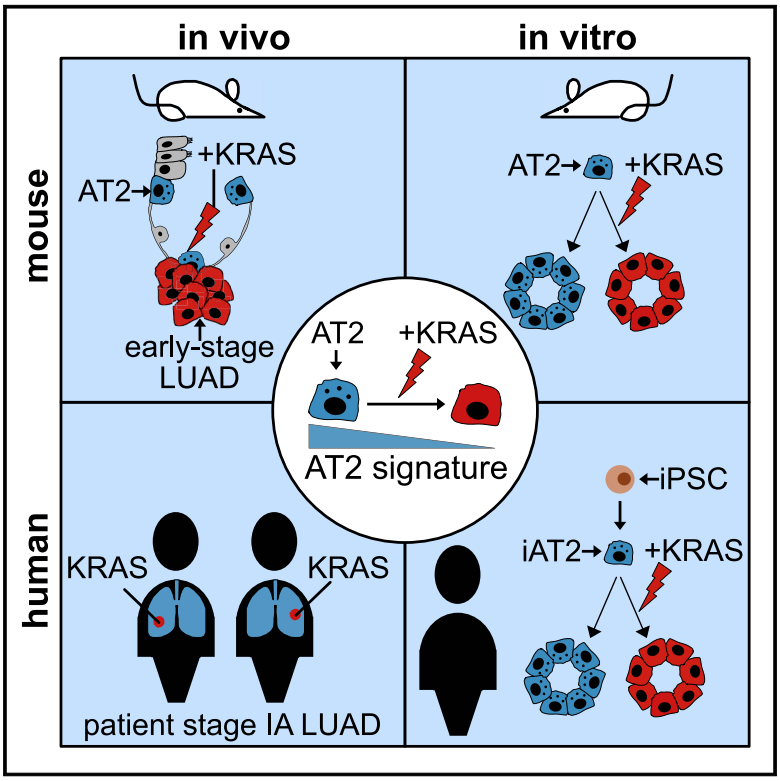
Cell Stem Cell

Resource

Organoids Model Transcriptional Hallmarks of Oncogenic KRAS Activation in Lung Epithelial Progenitor Cells

Graphical Abstract



Highlights

- Alveolar epithelial progenitor cells are transcriptionally distinct upon KRAS expression
- Alveolar epithelial organoids recapitulate early-stage lung adenocarcinoma
- Oncogenic KRAS leads to loss of lineage identity in AT2 cells
- Bulk, scRNA-seq, and proteomics data from murine and human KRAS mutant AT2 cells

Authors

Antonella F.M. Dost, Aaron L. Moye, Marall Vedaie, ..., Jane Yanagawa, Darrell N. Kotton, Carla F. Kim

Correspondence

jyanagawa@mednet.ucla.edu (J.Y.), dkotton@bu.edu (D.N.K.), carla.kim@childrens.harvard.edu (C.F.K.)

In Brief

Early-stage lung cancer is poorly understood. Dost, Moye et al. introduce new organoid systems to model lung cancer. KRAS-expressing alveolar progenitor cells had reduced expression of lineage genes in mouse and organoid models and stage IA cancers. This is the first report of loss of differentiation in early-stage lung cancer.

Dost et al., 2020, Cell Stem Cell *27*, 1–16
October 1, 2020
https://doi.org/10.1016/j.stem.2020.07.022

CellPress

Please cite this article in press as: Dost et al., Organoids Model Transcriptional Hallmarks of Oncogenic KRAS Activation in Lung Epithelial Progenitor Cells, Cell Stem Cell (2020), https://doi.org/10.1016/j.stem.2020.07.022

Cell Stem Cell

CellPress

Resource

Organoids Model Transcriptional Hallmarks of Oncogenic KRAS Activation in Lung Epithelial Progenitor Cells

Antonella F.M. Dost,[1,2,3,17] Aaron L. Moye,[1,2,3,17] Marall Vedaie,[4,5] Linh M. Tran,[6] Eileen Fung,[7] Dar Heinze,[4,8] Carlos Villacorta-Martin,[4] Jessie Huang,[4,5] Ryan Hekman,[9,10] Julian H. Kwan,[9,10] Benjamin C. Blum,[9,10] Sharon M. Louie,[1,2,3] Samuel P. Rowbotham,[1,2,3] Julio Sainz de Aja,[1,2,3] Mary E. Piper,[11] Preetida J. Bhetariya,[1,2,3,11] Roderick T. Bronson,[12] Andrew Emili,[9,10,13] Gustavo Mostoslavsky,[4,8] Gregory A. Fishbein,[14] William D. Wallace,[14,15] Kostyantyn Krysan,[6] Steven M. Dubinett,[6,16] Jane Yanagawa,[7,16,*] Darrell N. Kotton,[4,5,*] and Carla F. Kim[1,2,3,18,*]

[1]Stem Cell Program and Divisions of Hematology/Oncology and Pulmonary Medicine, Boston Children's Hospital, Boston, MA 02115, USA
[2]Harvard Stem Cell Institute, Cambridge, MA 02138, USA
[3]Department of Genetics, Harvard Medical School, Boston, MA 02115, USA
[4]Center for Regenerative Medicine of Boston University and Boston Medical Center, Boston, MA 02118, USA
[5]The Pulmonary Center and Department of Medicine, Boston University School of Medicine, Boston, MA 02118, USA
[6]Department of Medicine, David Geffen School of Medicine at UCLA, University of California, Los Angeles, Los Angeles, CA, USA
[7]Department of Surgery, David Geffen School of Medicine at UCLA, University of California, Los Angeles, Los Angeles, CA, USA
[8]Section of Gastroenterology and Department of Medicine, Boston University School of Medicine, Boston, MA 02118, USA
[9]Center for Network Systems Biology, Boston University, Boston, MA 02118, USA
[10]Department of Biochemistry, Boston University School of Medicine, Boston, MA 02118, USA
[11]Harvard T.H. Chan School of Public Health, Department of Biostatistics, Boston, MA 02115, USA
[12]Rodent Histopathology Core, Harvard Medical School, Boston, MA 02115, USA
[13]Department of Biology, Boston University, Boston, MA 02215, USA
[14]Department of Pathology and Laboratory Medicine, David Geffen School of Medicine at UCLA, University of California, Los Angeles, Los Angeles, CA 90095, USA
[15]Department of Pathology, Keck School of Medicine of USC, University of Southern California, Los Angeles, CA 90033, USA
[16]Jonsson Comprehensive Cancer Center, University of California, Los Angeles, Los Angeles, CA 90095, USA
[17]These authors contributed equally
[18]Lead Contact
*Correspondence: jyanagawa@mednet.ucla.edu (J.Y.), dkotton@bu.edu (D.N.K.), carla.kim@childrens.harvard.edu (C.F.K.)
https://doi.org/10.1016/j.stem.2020.07.022

SUMMARY

Mutant KRAS is a common driver in epithelial cancers. Nevertheless, molecular changes occurring early after activation of oncogenic KRAS in epithelial cells remain poorly understood. We compared transcriptional changes at single-cell resolution after KRAS activation in four sample sets. In addition to patient samples and genetically engineered mouse models, we developed organoid systems from primary mouse and human induced pluripotent stem cell-derived lung epithelial cells to model early-stage lung adenocarcinoma. In all four settings, alveolar epithelial progenitor (AT2) cells expressing oncogenic KRAS had reduced expression of mature lineage identity genes. These findings demonstrate the utility of our *in vitro* organoid approaches for uncovering the early consequences of oncogenic KRAS expression. This resource provides an extensive collection of datasets and describes organoid tools to study the transcriptional and proteomic changes that distinguish normal epithelial progenitor cells from early-stage lung cancer, facilitating the search for targets for KRAS-driven tumors.

INTRODUCTION

KRAS is one of the most frequently mutated oncogenes in epithelial cancers. Limited understanding of the biology of KRAS and its downstream effectors in epithelial cells likely contributes to the limited therapeutic targets for *KRAS* mutant cancers. Oncogenic KRAS is associated with poor prognosis and therapy resistance (Haigis, 2017). Tumor cell line experiments revealed that the rapidly accelerated fibrosarcoma (RAF)/mitogen-activated protein kinase (MAPK) and phosphatidylinositol 3-kinase (PI3K)/protein kinase B (AKT) pathways are activated upon overexpression of oncogenic KRAS, but pathway activation is distinct when oncogenic KRAS is expressed at physiological levels from its endogenous promoter (Tuveson et al., 2004; Zhu et al., 2014).

Oncogenic *KRAS* mutations are driving events in lung cancer and are present in 30% of lung adenocarcinomas (LUADs)

Please cite this article in press as: Dost et al., Organoids Model Transcriptional Hallmarks of Oncogenic KRAS Activation in Lung Epithelial Progenitor Cells, Cell Stem Cell (2020), https://doi.org/10.1016/j.stem.2020.07.022

(Collisson et al., 2014). Furthermore, expression of oncogenic $KRAS^{G12D}$ is sufficient to initiate LUAD in genetically engineered mouse models (GEMMs) (Jackson et al., 2001). Despite the significant effect of *KRAS* mutations in lung cancer, the effect of oncogenic KRAS on epithelial cells shortly after its activation, besides initiation of proliferation, has not been explored.

Recent advances in technologies such as single-cell RNA sequencing (scRNA-seq) and organoids make it possible to study transcriptional changes that follow oncogenic KRAS activation with single-cell resolution in a controlled environment. Previously published lung tumor organoids were derived from tumor cell lines or from tumors (Kaisani et al., 2014; Kim et al., 2019; Sachs et al., 2019) and, therefore, do not model the events in early-stage tumorigenesis. Efforts have been made to model all stages of cancer progression with organoids in non-lung tissues (Drost et al., 2015; Li et al., 2014; Matano et al., 2015; Seino et al., 2018). We demonstrated previously that primary murine lung progenitor cells survive *in vitro* activation of oncogenic KRAS in organoid cultures (Zhang et al., 2017a). However, the specific effect of oncogenic KRAS on transcriptional states was not studied in any of these reports.

To facilitate study of oncogenic KRAS-induced changes, we analyzed data from an early-stage $Kras^{G12D}$ GEMM, *in-vitro*-induced $Kras^{G12D}$ alveolar epithelial progenitor (AT2) cell-derived murine lung organoids, *in-vitro*-induced $KRAS^{G12D}$ human lung organoids derived from induced pluripotent stem cells (iPSCs), and lesions from stage IA LUAD patients, all at single-cell resolution. Characterization of the data revealed that a reduction in AT2 cell lineage marker gene expression is an early consequence of oncogenic KRAS. Our organoid systems are tools to rapidly and accurately model LUAD progression *in vitro*, and our datasets useful resources for the cancer research community.

RESULTS

scRNA-Seq of Distal Lung Epithelium Reveals Distinct Transcriptional Clusters of $KRAS^{G12D}$-Activated Cells during Early Tumorigenesis

We used scRNA-seq to define transcriptional changes in distal epithelial cell populations during early-stage LUAD in the yellow fluorescent protein (YFP) reporter containing $Kras^{LSL\text{-}G12D}$; $Rosa26^{LSL\text{-}YFP}$ (KY) LUAD GEMM (Jackson et al., 2001). KY mice were infected with an adenovirus 5 vector containing Cre recombinase driven by the ubiquitous cytomegalovirus (CMV) promoter (Ad5-CMV-Cre) (Figure 1A). After 7 weeks, we observed small clusters of YFP^+ cells consistent with atypical adenomatous hyperplasia (Figure S1A). Viable, epithelial cell adhesion molecule (EPCAM) positive, recombined ($CD31^-/CD45^-/EPCAM^+/YFP^+$ [YFP^+]) and non-recombined ($CD31^-/CD45^-/EPCAM^+/YFP^-$ [YFP^-]) cells were collected using fluorescence-activated cell sorting (FACS) (Figure S1B). We used 10X Genomics scRNA-seq to examine gene expression during early-stage LUAD and analyzed the data using ScanPy (Wolf et al., 2018). After pre-processing, we focused on clusters containing more than 100 cells, leaving four clusters for further analysis (Figures 1B, S1C, and S1D; STAR Methods). Cluster 1 (C1) was comprised primarily of YFP^+ cells and C0 of YFP^- cells, whereas C2 and C3 had equivalent contributions from YFP^+ and YFP^- cells (Figures 1C and 1D). Expression of the AT2 cell markers *Sftpc* and *Lyz2* was highest in C0 and C1, of the ciliated cell markers *Foxj1* and *Cd24a* in C2, and of the club cell markers *Scgb1a1* and *Scgb3a2* in C3 (Figure 1E). Although YFP^+ and YFP^- cells were present in C2 and C3, only C0 and C1 with elevated AT2 cell marker expression formed transcriptionally distinct YFP^- and YFP^+ clusters (Figures 1B–1D). Correlation analysis of all clusters revealed that C0 and C1 had some degree of similarity, whereas C2 and C3 were more distinct (Figure S1E).

AT2 cells have been proposed previously as the LUAD cells of origin (Lin et al., 2012; Xu et al., 2012) and are the only lung epithelial cell type that forms a transcriptionally distinct cluster upon $KRAS^{G12D}$ expression. Hence, we focused our studies on the consequences of KRAS activation in AT2 cells. To test whether the transcriptional changes in YFP^+ C1 agree with previously published data, we calculated *Z* scores using gene signatures we expected to be elevated in YFP^+ C1. Consistent with published observations, KRAS and nuclear factor κB (NF-κB) target gene signatures were elevated in C1 cells, as was a proliferation signature, indicating that the cluster is transcriptionally primed to proliferate (Figures 1F–1H; Table S1; Barbie et al., 2009; Bild et al., 2006; Meylan et al., 2009; Travaglini et al., 2019).

Next, we performed differential expression (DE) analysis to identify genes, transcription factors (TFs), and co-factors (TFCs) that define C0 and C1 (Figures S1F and S1G; Tables S1 and S2). We found that the lung fate TF *Nkx2-1* and the AT2 cell identity TF *Etv5* were enriched in C0 (Morrisey and Hogan, 2010; Zhang et al., 2017b). In contrast, the proto-oncogene *Myc* (Chen et al., 2018; Dang, 2012; Poole and van Riggelen, 2017) and *Id1*, a TF shown to promote non-small cell lung cancer (NSCLC) cell proliferation and metastasis, were upregulated in C1 (Antonângelo et al., 2016; Cheng et al., 2011; Pillai et al., 2011). Moreover, *Foxq1*, a TF found to be increased in NSCLC tumor tissue compared with paired adjacent tissue, was elevated (Li et al., 2020), and *Etv4*, a TF expressed during lung development (Herriges et al., 2015), and *Klf4*, important for inducing pluripotency in cells (Takahashi and Yamanaka, 2006), had elevated expression in C1. Hence, upon $KRAS^{G12D}$ expression, AT2 cells downregulate TF/TFCs that maintain AT2 cell identity, whereas factors known to promote cancer growth, important for developmental processes, and induce pluripotency have increased expression. We tested whether the expression of these TF/TFCs correlated with a transition to a less differentiated state, as often observed in late-stage cancers. Indeed, a signature consisting of 46 murine AT2 cell marker genes (Franzén et al., 2019) was significantly lower in C1 compared with C0 (Figure 1I; Table S1).

It has been shown recently that primary human LUAD contains cells that express multiple lineage-specific signatures (Laughney et al., 2020). Therefore, we looked for "lineage infidelity" in our early-stage GEMM data. We found that C1 had lower expression of the AT2 cell markers *Sftpc*, *Lyz2*, and *Etv5*, consistent with loss of AT2 cell identity. Strikingly, the alveolar type 1 (AT1) markers *Aqp5* and *Pdpn* and the club cell markers *Scgb1a1* and *Scgb3a2* were upregulated, indicating transcriptional priming for other lung epithelial cell types. Furthermore, *Ly6a* (SCA1), a marker of lung stem cells in mice (Kim et al., 2005) and tumor-propagating cells in the $Kras^{LSL\text{-}G12D/+}$; $p53^{fl/fl}$ (KP)

Please cite this article in press as: Dost et al., Organoids Model Transcriptional Hallmarks of Oncogenic KRAS Activation in Lung Epithelial Progenitor Cells, Cell Stem Cell (2020), https://doi.org/10.1016/j.stem.2020.07.022

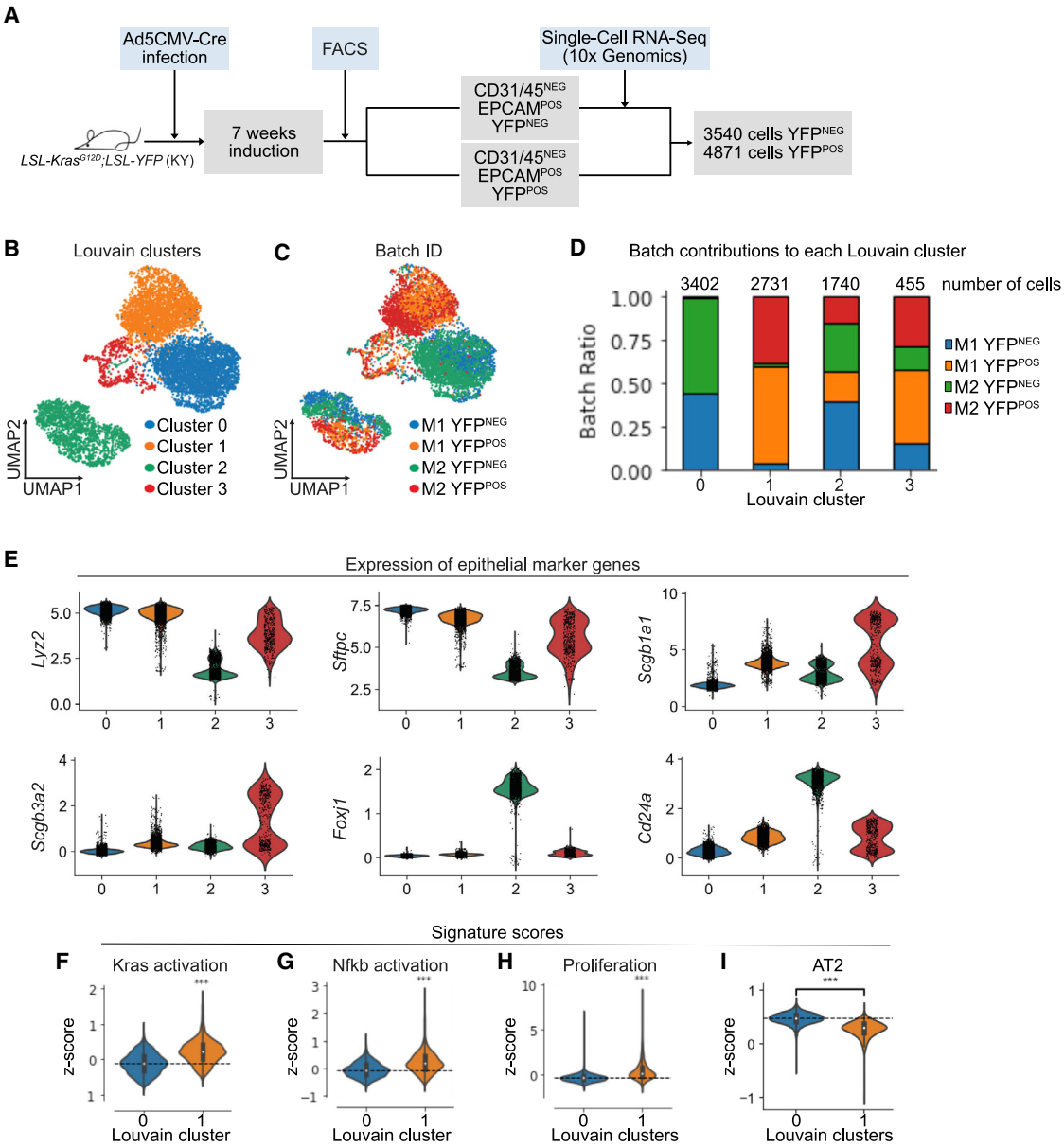


Figure 1. scRNA-Seq of Distal Lung Epithelium Reveals Distinct Transcriptional Clusters of KRAS^{G12D}-Activated Cells during Early Tumorigenesis

(A) Experimental strategy to analyze epithelial populations during early-stage LUAD *in vivo* using scRNA-seq.

(B and C) Clustering of transcriptomes using Uniform Manifold Approximation and Projection (UMAP). Cells are colored based on (B) Louvain clusters or (C) batch ID.

(D) Batch contributions to each Louvain cluster with the number of cells indicated.

(E) Log expression of lung epithelial cell marker genes in each Louvain cluster.

(F–I) *Z* scores of the indicated signatures in Louvain clusters 0 and 1. A dashed line marks the median of the reference sample. The p values were determined using a Mann-Whitney rank test. n.s. ***$p < 0.0005$.

See also Figure S1.

lung cancer model (Curtis et al., 2010) was also upregulated in some C1 cells (Figure S1H).

Finally, we performed Gene Ontology (GO) analysis on differentially expressed genes in C0 and C1 to identify pathways altered in AT2 cells after KRAS activation. In total, we found 8 common, 73 C0-specific and 160 C1-specific enriched GO pathways (Figure S1I; Table S2). Unique terms in C1 included "NIK/NF-κB signaling" (NIK: NF-κB-inducing kinase), consistent with our finding (Figure 1G), and terms that indicate upregulated ribosome biogenesis and translation. Unique terms in C0 included cholesterol, alcohol, and lipid metabolism pathways, suggesting that these processes have an essential role in AT2 cell biology.

Please cite this article in press as: Dost et al., Organoids Model Transcriptional Hallmarks of Oncogenic KRAS Activation in Lung Epithelial Progenitor Cells, Cell Stem Cell (2020), https://doi.org/10.1016/j.stem.2020.07.022

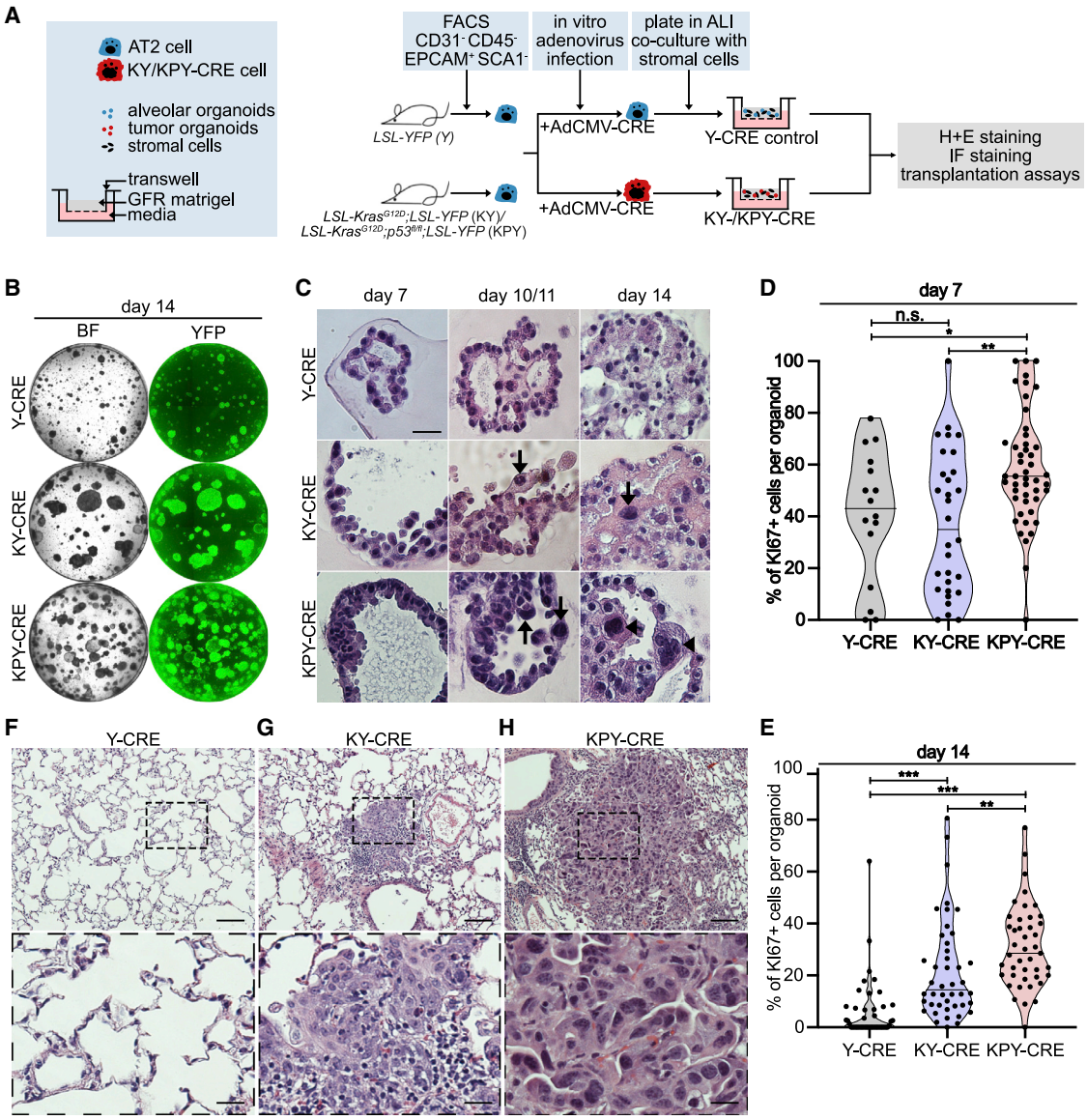


Figure 2. Inducible Organoids Rapidly Recapitulate *In Vivo* Tumor Progression and Form Tumors upon Transplantation

(A) Experimental strategy to grow air-liquid interphase (ALI) organoid cultures in growth factor-reduced (GFR) Matrigel.

(B) Representative whole-well bright-field (BF) and YFP channel images of organoid cultures. Images were stitched together to show whole wells.

(C) Representative H&E-stained organoid slides. Arrows: pleomorphic cells. Arrowheads: giant, multinucleated cells. Scale bar, 25 μm.

(D and E) Quantification of KI67$^+$ cells per organoid on (D) day 7 and (E) day 14 of organoid culture based on IF staining. Each dot represents one organoid.

(F–H) H&E staining of mouse lungs that were transplanted with organoid-derived cells. Scale bars, 100 μm (lower magnification) and 25 μm (higher magnification).

The p values were determined using a Mann-Whitney rank test. n.s., $p \geq 0.05$; *$p < 0.05$; **$p < 0.005$; ***$p < 0.0005$. See also Figure S2.

Inducible Organoids Rapidly Recapitulate *In Vivo* Tumor Progression and Form Tumors upon Transplantation

To better understand transcriptional programs that follow KRAS^{G12D} activation, we developed an *in vitro* organoid system that allowed us to rapidly model changes in primary lung AT2 cells shortly after induction of oncogenic KRAS. We hypothesized that *Kras^{G12D}* activation alone mimics an early tumor stage phenotype, whereas additional loss of the tumor suppressor *Tp53* models a more advanced stage, as is the case in GEMMs (Jackson et al., 2001, 2005). We generated organoids by dissecting lungs of adult KY, *Kras$^{LSL\text{-}G12D/+}$*; *p53$^{fl/fl}$*; *Rosa26$^{LSL\text{-}YFP}$* (KPY), and *Rosa26$^{LSL\text{-}YFP}$*(Y) control mice and used FACS to isolate AT2 cells (CD45$^-$/CD31$^-$/EPCAM$^+$/SCA1$^-$) (Kim et al., 2005; Lee et al., 2014; Figure 2A). Cells were infected with the Ad5-CMV-Cre (CRE) virus *in vitro* and cultured with stromal cells in our 3D organoid air-liquid interface (ALI) co-culturing system described previously (Lee et al., 2014, 2017). Upon Cre expression, almost all organoids were YFP$^+$, suggesting high Cre induction efficiency (Figure 2B).

Histological analysis revealed that our tumor organoid model recapitulated *in vivo* tumor progression. Hematoxylin and eosin

Please cite this article in press as: Dost et al., Organoids Model Transcriptional Hallmarks of Oncogenic KRAS Activation in Lung Epithelial Progenitor Cells, Cell Stem Cell (2020), https://doi.org/10.1016/j.stem.2020.07.022

(H&E)-stained sections of organoids demonstrated that Y-CRE control organoids maintained normal nuclei, whereas the nuclei of KY-CRE and KPY-CRE cells became enlarged and abnormal, with giant multinucleated cancer cells in KPY organoids (Figure 2C). This observation is reminiscent of documented *in vivo* tumor cell phenotypes in the *Kras*$^{LSL\text{-}G12D/+}$ and KP mouse models (Jackson et al., 2001, 2005).

Next, we interrogated the effect of KRAS^{G12D} on proliferation. On day 7 of organoid culture, there was no significant difference in the percentage of KI67^{+} cells per organoid between the Y-CRE control and KY-CRE, whereas there were 1.3-fold and 1.6-fold increases in KPY-CRE organoids compared to Y-CRE and KY-CRE, respectively (Figures 2D and S2A). On day 14, most of the Y-CRE control organoids stained negative for KI67, whereas KY-CRE and KPY-CRE organoids still contained cells that stained positive for KI67 (Figures 2E and S2B). Thus, organoids from all three genotypes contained a high number of proliferating cells on day 7, but although most of the cells in the control organoids had stopped proliferating by day 14, cells in KY and KPY organoids continued to proliferate.

To test whether KRAS^{G12D}-expressing organoids form tumors *in vivo*, we performed orthotopic transplantation assays. We transplanted single-cell suspensions from Y-CRE control, KY-CRE, and KPY-CRE organoids into the lungs of bleomycin-injured mice (n = 4, n = 6, and n = 4, respectively). After 4 weeks, we evaluated tumor formation by histology. The lungs of Y-CRE control-transplanted mice did not show any signs of aberrant epithelial cell growth or tumor formation (Figure 2F). In contrast, in KY-CRE- and KPY-CRE-transplanted lungs, we found tumors that contained cells with pleomorphic features and giant cancer cells in KPY-CRE-transplanted lungs, comparable with observations in the organoid cultures (Figures 2G and 2H). Immunofluorescence (IF) staining for YFP confirmed that these tumor lesions contained transplanted cells (Figures S2C and S2D). Hence, cells derived from our *in-vitro*-induced tumor organoids formed tumors within 4 weeks, dramatically reducing the time required to model lung cancer *in vivo* compared with traditional GEMMs.

KRAS^{G12D}-Activated Cells in Organoids Lose AT2 Cell Differentiation Markers and Express Developmental Lung Markers

To further investigate transcriptional changes following KRAS^{G12D} activation, we performed RNA-seq on cells from our organoid cultures. KY- and KPY-derived AT2 cells received the Ad5-CMV-Empty virus (Emp, control), no virus (control), or CRE (Figure 3A). Because we sought to reveal transcriptional changes that follow KRAS^{G12D} activation but not proliferation, we analyzed the organoids on day 7 of organoid culture, when proliferation was observed in all organoid types. After 7 days in culture, single-cell suspensions were enriched for epithelial cells by FACS for EPCAM^{+} cells (Figure S3A). 87% ± 7% and 95% ± 2% of EPCAM^{+} cells of the KY-CRE and KPY-CRE samples, respectively, were YFP^{+}, further confirming the high efficiency of *in vitro* Cre induction. Next we performed RNA-seq on EPCAM^{+} cells. Sample-sample correlation analysis revealed that all control samples were highly correlated, whereas KY-CRE and KPY-CRE samples had high correlation and were transcriptionally distinct from the controls (Figure S3B). To perform DE analysis, we compared the CRE samples with their respective Emp controls (henceforth, KY-Differential [KY-Dif] and KPY-Differential [KPY-Dif]; Table S3). To determine genes that were altered by KRAS^{G12D} expression, we compared KY-Dif with KPY-Dif and found 1,206 genes that were shared upregulated and 1,464 genes that were shared downregulated (Figure S3C; Table S3).

Because we saw downregulation of AT2 cell differentiation genes in our GEMM data, we investigated the expression of known AT2 cell markers and lung development genes. When we compared the top 100 up- and downregulated genes in our RNA-seq data, we found that *Cd74* and *Lyz2*, two AT2 cell marker genes, were among the top shared downregulated genes (Figures 3B and 3C). Conversely, the developmental genes *Hmga2* and *Sox9* were upregulated. Furthermore, we found increased expression of *Ly6a* (SCA1), consistent with our findings *in vivo* (Figures S1H, 3B, and 3C). Moreover, we found that other known AT2 cell markers, *Sftpc*, *Sftpd*, and *Nkx2-1*, were significantly downregulated in organoids from both genotypes (Figure 3C).

Next, we investigated whether these changes also occurred at the protein level. IF staining for surfactant protein C (SPC; *Sftpc*) showed that the percentage of SPC^{+} cells per organoid decreased 6.7-fold in KY-CRE and 20-fold in KPY-CRE compared with Y-CRE control organoids on day 7 (Figures 3D and 3E). On day 14, there was a 1.1-fold decrease in KY-CRE and a 1.6-fold decrease in KPY-CRE compared with Y control organoids (Figures S3E and S3F). Furthermore, staining for the lung epithelial marker NKX2-1 and the developmental marker HMGA2 was negatively correlated; individual cells that gained HMGA2 expression had reduced levels of NKX2-1 (Figure 3F). Thus, we demonstrated that transcriptional downregulation of AT2 cell markers and upregulation of developmental markers correlated with altered expression of the respective proteins.

KRAS^{G12D}-Expressing Organoid Cells Are Transcriptionally Distinct and Transition to a Developmental-like State

To further characterized our KY-CRE organoids, we performed scRNA-seq. As before, we characterized day 7 EPCAM^{+} cells from KY-CRE and KY-Emp organoids (Figures 4A and S4A). After filtering and preprocessing the data, we identified three clusters: C0^{org}, C1^{org}, and C2^{org} (Figures 4B, S4B, and S4C). C1^{org} was composed mostly of KY-Emp cells, representing the control cluster, whereas C0^{org} and C2^{org} mostly contained KY-CRE cells (Figures 4B–4D). Correlation analysis revealed that all three clusters were distinct and that C0^{org} and C1^{org} were negatively correlated (Figure S4D). As with our GEMM data, we checked the expression of previously published gene signatures upregulated in NSCLC. As expected, the KRAS activation signature was upregulated in C0^{org} and C2^{org} compared with control cluster C1^{org} (Figure 4E; Table S1). The NF-κB activation signature was lower in C2^{org} and higher in C0^{org} compared with C1^{org}, indicating that only one of the Cre clusters had upregulated NF-κB signaling (Figure 4F; Table S1). Interestingly, the proliferation signature was only elevated in C2^{org} but not in C0^{org}, indicating that only one of the Cre clusters had a higher proliferation signature than the control, despite high Kras activation signatures in both clusters (Figure 4G; Table S1).

Please cite this article in press as: Dost et al., Organoids Model Transcriptional Hallmarks of Oncogenic KRAS Activation in Lung Epithelial Progenitor Cells, Cell Stem Cell (2020), https://doi.org/10.1016/j.stem.2020.07.022

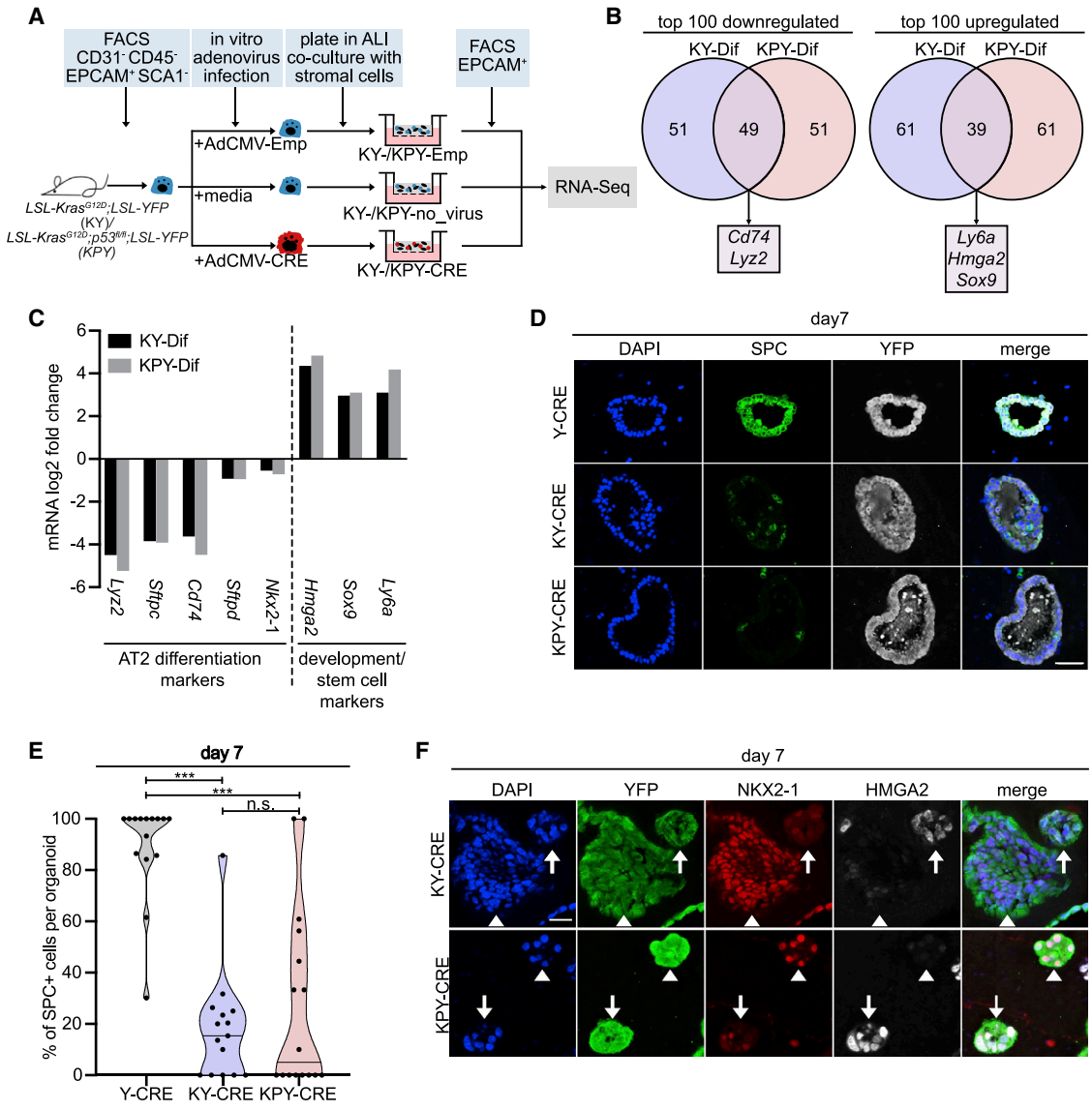


Figure 3. KRAS^{G12D}-Activated Cells in Organoids Lose AT2 Cell Differentiation Markers and Express Developmental Lung Markers

(A) Experimental strategy to grow ALI organoid cultures to perform RNA-seq.

(B) Venn diagram showing the overlap of the top 100 differentially expressed genes in KY-CRE and KPY-CRE compared with their respective Emp controls.

(C) Log2 fold change expression of selected genes compared with their control from RNA-seq results.

(D) Representative pictures of IF staining on day 7 of organoid culture. Scale bar, 100 μm.

(E) Quantification of SPC$^+$ cells per organoid on day 7 of organoid culture. Each dot represents one organoid.

(F) Representative pictures of IF staining on day 7 of organoid culture. Scale bar, 25 μm.

The p values were determined using a Mann-Whitney rank test. n.s., $p \geq 0.05$; ***$p < 0.0005$. See also Figure S3.

Next, we performed DE analysis followed by identification of TF/TFCs (Figures S4E and S4F; Tables S1 and S4). Similar to our GEMM data, control C1^{org} had elevated expression of *Etv5*, providing additional evidence of loss of AT2 cell transcriptional identity. One TF highly expressed in both Cre clusters compared with the control was *Foxq1*, and C2^{org} had high expression of *Id1*, two TFs we had also detected in our GEMM. Interestingly, C0^{org} had high expression of the lung development TF *Sox9*, confirming the observations in our RNA-seq analysis (Figure 3C). In the same cluster, *Smad7* and *Trp53*, indicative of Tgfb and p53 signaling, respectively, were also upregulated.

In agreement with our RNA-seq and IF results, we observed a reduced AT2 cell signature in the two KY-CRE clusters C0^{org} and C2^{org}, similar to our GEMM data (Figures 4H and 1I). Consistent with that, the AT2 cell markers *Lyz2* and *Sftpc* and the lung identity TF *Nkx2-1* had reduced expression (Figure 4I). In contrast, the lung development genes *Hmga2* and *Sox9* were upregulated in both KY-CRE clusters (Figures 4I and 4J; Kim et al., 2005; Liu et al., 2019; Salwig et al., 2019; Singh et al., 2014). Furthermore, we found that a *Sox9* target gene signature was upregulated in C0^{org}, suggesting that *Sox9* is highly expressed and active in this cluster (Figure 4K;

Please cite this article in press as: Dost et al., Organoids Model Transcriptional Hallmarks of Oncogenic KRAS Activation in Lung Epithelial Progenitor Cells, Cell Stem Cell (2020), https://doi.org/10.1016/j.stem.2020.07.022

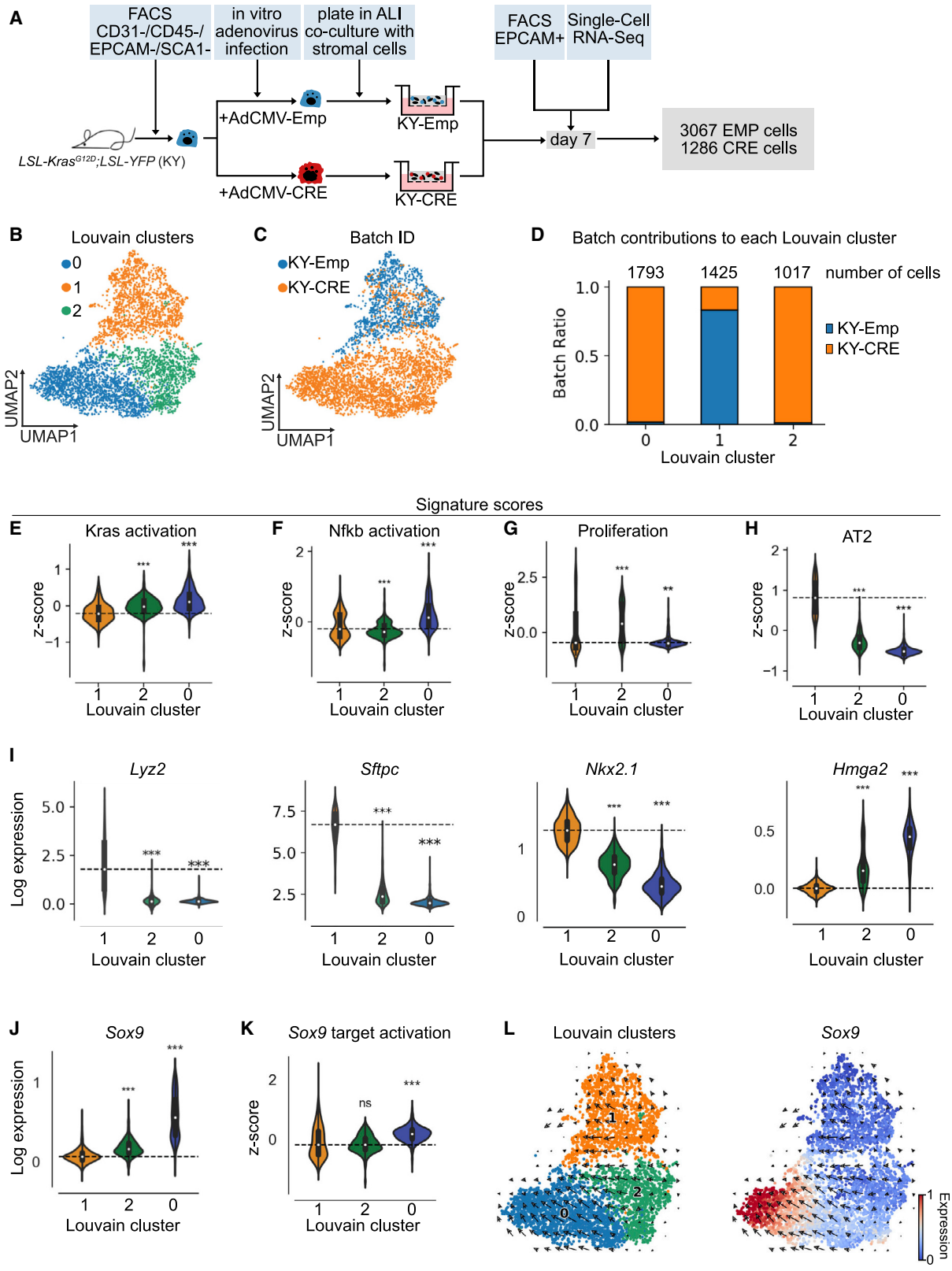


Figure 4. KRAS^{G12D}-Expressing Organoid Cells Are Transcriptionally Distinct and Transition to a Developmental-like State

(A) Experimental strategy to grow ALI organoid cultures followed by scRNA-seq.

(B and C) Clustering of transcriptomes using UMAP. Cells are colored based on (B) Louvain clusters or (C) Batch ID.

(legend continued on next page)

Please cite this article in press as: Dost et al., Organoids Model Transcriptional Hallmarks of Oncogenic KRAS Activation in Lung Epithelial Progenitor Cells, Cell Stem Cell (2020), https://doi.org/10.1016/j.stem.2020.07.022

Table S1). Next we tested whether *Sox9* was also upregulated in our YFP$^+$ cluster in our GEMM. Strikingly, the *Sox9* and *Sox9* target activation signatures were significantly upregulated in the YFP$^+$ C1 cluster compared with the YFP$^-$ C0 cluster (Figures S4G and S4H). Notably, the changes in the GEMM model were much more subtle and the expression levels lower compared with the organoid data.

We wondered whether the two KY-CRE clusters represent two different stages in cancer cell progression and whether there is a transition from one cluster to the other. To address this, we analyzed the organoid and GEMM scRNA-seq datasets using RNA velocity, a computational pipeline that infers expression dynamics and directionality based on RNA splicing (La Manno et al., 2018). In the organoid data, RNA velocity indicated that KRAS^{G12D}-expressing AT2 cells transition from *Sox9*^{LOW} to *Sox9*^{HIGH} cells (Figure 4L). In contrast, although the C1 GEMM cluster shows a clear direction of transition, it is not solely directed toward *Sox9*$^+$ cells (Figure S4I). This observed difference might be due to the significantly lower expression levels of *Sox9* in the GEMM.

Next, we tested whether the cells expressed differentiation markers of other cell types, as observed in our GEMM data (Figure S1H). As expected, the two Cre clusters C2^{org} and C0^{org} had lower expression of the AT2 cell markers *Sftpc*, *Lyz2*, and *Etv5*, consistent with loss of AT2 cell identity (Figure S4J). In contrast to our GEMM data, the AT1 marker *Aqp5* had higher expression in the Cre clusters, whereas *Pdpn* expression was elevated in the control cluster. Furthermore, some cells in the Cre clusters had high expression of the ciliated cell markers *Cd24a* and *Foxj1* and the progenitor marker *Ly6a* (SCA1). SCA1 and CD24 mark tumor-propagating cells in the KP mouse model (Lau et al., 2014). The club cell markers *Scgb1a1* and *Scgb3a2* were upregulated in some Cre-expressing cells, similar to our observations in the GEMM.

Last, GO enrichment analysis was performed to identify unique pathways for each of the KY organoid clusters (Figure S4K; Table S4). Pathways enriched in C0^{org} included "Regulation of I-κB kinase/NF-κB signaling," consistent with the increased NF-κB signature (Figure 4F), and "ERBB signaling," demonstrated to facilitate KRAS^{G12D} lung tumorigenesis (Kruspig et al., 2018). C2^{org} was enriched for pathways related to translation, mRNA processing, and G1/S transition, potentially connected to the increased proliferation signature identified in this cluster (Figure 4G). Control C1^{org}, much like YFP$^-$ AT2 cells in our GEMM scRNA-seq dataset (Table S2), was enriched for cholesterol, alcohol, and lipid metabolism pathways (Table S4).

Overall, we found many similarities between our GEMM and *in-vitro*-induced tumor organoid system. Most notably, we found that AT2 cell lineage genes are downregulated and developmental and progenitor genes are upregulated in both models, providing evidence that loss of differentiation occurs during early-stage LUAD.

Human iAT2s Downregulate Differentiation and Maturation Markers and Upregulate Progenitor Markers upon KRAS^{G12D} Expression

To test whether loss of AT2 differentiation markers early after KRAS^{G12D} induction can also be observed in human cells, we engineered an iPSC line to allow doxycycline (dox) regulated activation of KRAS^{G12D} in iPSC-derived AT2 (iAT2) cells. Using the iPSC line BU3 NKX2-1-GFP; SFTPC-tdTomato (NGST) (Jacob et al., 2017), which includes GFP and tdTomato reporters targeted to the endogenous *NKX2-1* and *SFTPC* loci, respectively, we integrated the KRAS^{G12D} cassette together with a dox-inducible promoter into the "safe harbor" AAVS1 locus (Figure 5A; Tiyaboonchai et al., 2014). Next we differentiated the iPSCs into NKX2-1$^+$ lung epithelial progenitors, sorted for NKX2-1^{GFP+} cells by FACS, and generated distal lung alveolospheres using our lung-directed differentiation protocol (Figure 5B; Jacob et al., 2019). To test the dox-inducible KRAS^{G12D} construct, we treated the alveolospheres with control vehicle (DMSO) or dox and performed deep proteomics and phosphoproteomics analyses (n = 4 replicates per condition; Figures 5SA and 5SB; Table S5). As expected, we observed upregulation of KRAS protein in dox-treated cells (Figure 5C), and increased phosphorylation of KRAS targets such as MAPK1, RPS6KA1, and MAPK3 (Figure 5C). Gene set enrichment analysis (GSEA) revealed RAS signaling as the top enriched pathway in dox-treated iAT2 cells (Figure 5D). Therefore, our proteomics and phosphoproteomics analyses confirmed that iAT2 KRAS^{G12D} cells upregulated KRAS and components of the RAS/MAPK signaling pathway upon dox treatment, indicating successful dox-regulated functional activation of KRAS in the human iAT2 cell *in vitro* model system.

To assess the downstream consequences of this signaling in iAT2 cells, we sorted pure NKX2-1^{GFP+} SFTPC^{tdTomato} double-positive cells and treated them with dox or DMSO (Figure 5B). After 2 weeks of treatment, flow analysis revealed that, although the majority of cells maintained NKX2-1^{GFP} expression under both conditions, there was a reduction of SFTPC^{tdTomato} expression frequency and intensity under the dox condition, which was sustained through multiple passages (Figure 5E).

To better understand the loss of *SFTPC*, we performed scRNA-seq (Figure 5B). Using the 10X Chromium platform, we profiled the transcriptomes of 775 DMSO- and 1,322 dox-treated cells and performed DE analysis (Table S5). Unbiased analysis of all cells revealed 3 cell clusters, with control iAT2 cells grouped as a single cluster (Figures S5C and S5D). DE analysis showed significant upregulation of KRAS in both dox-treated clusters, one of which also exhibited significant upregulation of proliferation markers (e.g., *MKI67*, *TOP2A*, and *CDK1*) (Figures 5F and S5E). In contrast, multiple AT2 cell genes were significantly upregulated in the control cluster (e.g., *LPCAT1*, *SFTPB*, *SFTPC*, *CRLF1*, *CTSH*, *SLC34A2*, *NAPSA*, and *PGC*) (Figures 5F and S5E). Consistent with this observation, previously published iAT2 cell differentiation (*SFTPB*, *SFTPC*, *SFTPD*,

(D) Batch contributions to each Louvain cluster with the number of cells indicated.
(E–H) *Z* scores of the indicated signatures in each Louvain cluster. A dashed line marks the median of the reference sample.
(I and J) Log2 expression of the indicated genes. A dashed line marks the median expression of the reference sample.
(K) *Z* score of the indicated signature in each Louvain cluster. A dashed line marks the median of the reference sample.
(L) RNA velocity analysis of the KRAS^{G12D} organoid scRNA-seq dataset. Louvain clusters are shown on the left. *Sox9* expression is visualized on the right.
The p values were determined using a Mann-Whitney rank test. ***p > 0.001, **p > 0.01. See also Figure S4.

Please cite this article in press as: Dost et al., Organoids Model Transcriptional Hallmarks of Oncogenic KRAS Activation in Lung Epithelial Progenitor Cells, Cell Stem Cell (2020), https://doi.org/10.1016/j.stem.2020.07.022

CellPress

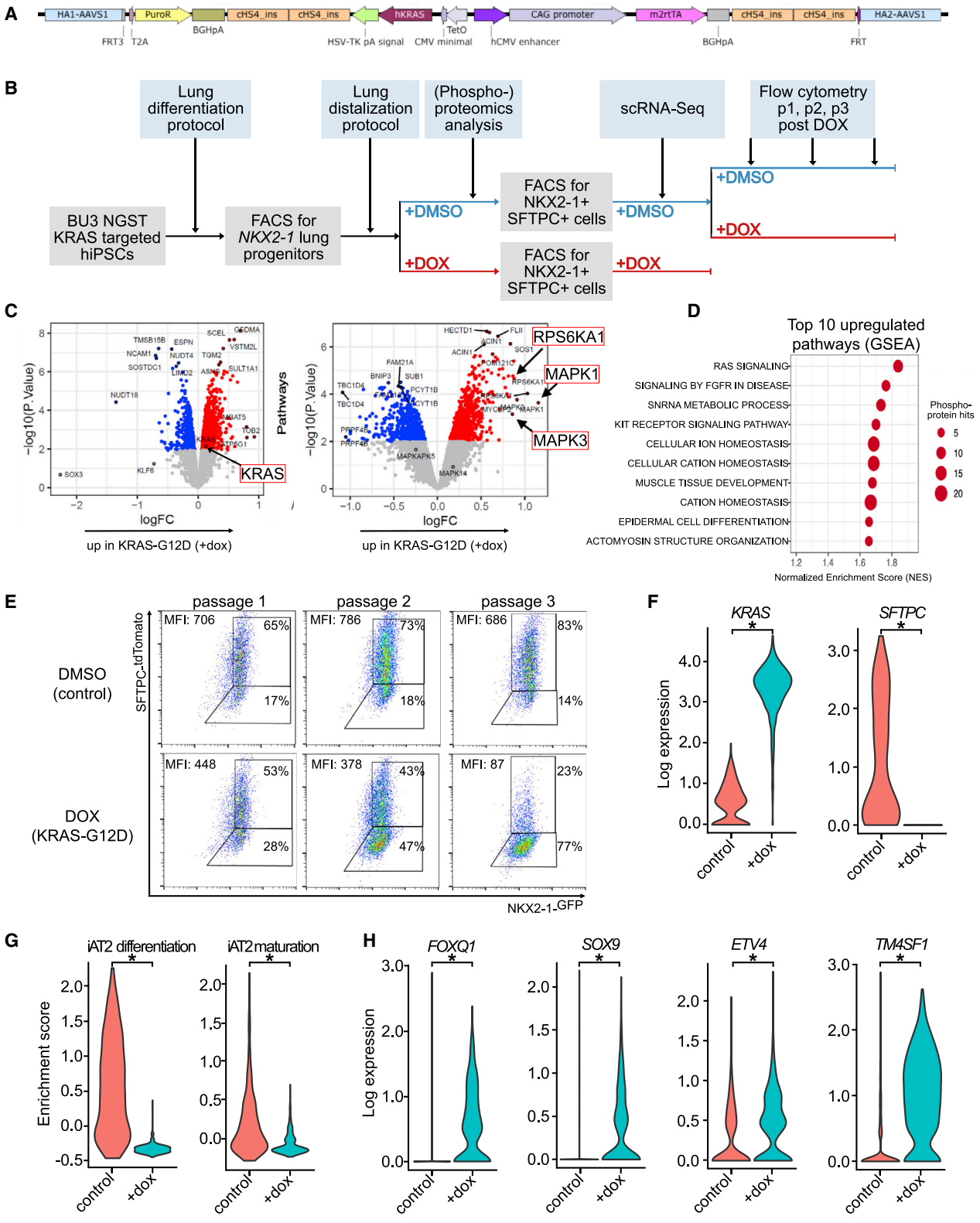


Figure 5. Human iAT2 Cells Downregulate Differentiation and Maturation Markers and Upregulate Progenitor Markers upon KRAS^{G12D} Expression

(A) Schematic of the AAVS1 locus with integrated dox-inducible KRAS^{G12D}.

(B) Experimental strategy and timeline to grow and analyze KRAS^{G12D}-inducible iAT2 cells. Dox, doxycycline (1 μg/mL); p1, p2, and p3, passages 1, 2, and 3.

(legend continued on next page)

Please cite this article in press as: Dost et al., Organoids Model Transcriptional Hallmarks of Oncogenic KRAS Activation in Lung Epithelial Progenitor Cells, Cell Stem Cell (2020), https://doi.org/10.1016/j.stem.2020.07.022

CLDN18, *LAMP3*, *SLC34A2*, *IL8*, and *NAPSA*) and maturation (*SFTPA1*, *SFTPA2*, *PGC*, *CXCL5*, and *SLPI*) gene signatures (Hurley et al., 2020) and 20 AT2 cell markers shared between mouse and human from the Panglao database (Table S1) were significantly downregulated in dox-treated iAT2 cells (Figures 5G and S5F), as was the TF *ETV5*, which we had also identified in our GEMM and murine organoid data (Figure S5G). Moreover, the TFs *FOXQ1* and *ID1* were upregulated, together with the developmental and progenitor genes *SOX9* and *ETV4*, which is also consistent with our murine data (Figures 5H and S5G). An additional notable upregulated transcript in dox-treated iAT2 cells was *TM4SF1*, reported recently as an AT2 marker, enriched in Wnt-responsive cells during regeneration *in vivo* (Zacharias et al., 2018; Figure 5H). In keeping with an increased Wnt response, the Wnt target gene *LEF1* (McCauley et al., 2017; Zacharias et al., 2018) was upregulated in dox-exposed cells (Figure S5G). As indicated by our FACS results, *NKX2-1* was still expressed by our $KRAS^{G12D}$-expressing cells but slightly downregulated, consistent with our RNA-seq data and IF staining in murine organoids (Figures 3C, 3F, and S5G).

Taken together, our human iAT2 cell results indicated that $KRAS^{G12D}$ results in downregulation of iAT2 differentiation and maturation markers and upregulation of progenitor and developmental markers, corroborating the results from our GEMM and murine organoid model.

Differentiation and Maturation Markers Are Downregulated in AT2 Cells from Human Early-Stage LUAD

To assess whether the loss of AT2 cell identity observed in our GEMM, murine organoid, and human iAT2 cell models also occurs in lung cancer patients, we performed scRNA-seq of LUAD specimens with activating *KRAS* mutations and associated distal normal lung tissues (>2 cm from the tumor) from two stage IA LUAD patients (Figure 6A). Unsupervised clustering of non-immune cells identified epithelial ($EPCAM^+$), fibroblast ($COL1A1^+$), and endothelial ($PECAM1^+$) cell clusters (Figures 6B, 6E, and S6A). The epithelial cells were further divided into AT1 ($PDPN^+$), club ($SCGB1A1^+$), ciliated ($FOXJ1^+$), and two distinct AT2 cell clusters, one comprised of AT2 cells from normal lung tissue and the second one from LUAD (Figures 6C–6E). AT2 cells from normal lung were characterized by high *SFTPB* and *SFTPD* expression, whereas AT2 cells from stage 1A LUAD had decreased *SFTPD* expression (Figure 6E). Interestingly, AT2 cells were the only epithelial cell type that formed distinct clusters in LUAD and associated normal lung tissues, whereas other cell types aggregated regardless of their origin (Figures 6B and 6C), consistent with our observations in the $KRAS^{G12D}$ GEMM (Figures 1B and 1C). Next, we checked the expression of 20 AT2 cell markers shared between mouse and human from the Panglao database (Table S1) in AT2 cells from LUAD patients. All 20 markers were highly expressed in normal lung and LUAD AT2 cells but not in the other cell types, confirming that AT2 cell cluster annotation was appropriate in normal lung and LUAD (Figures 6F and S6B; Table S6). However, AT2 cells from stage IA LUAD expressed reduced levels of these markers compared with AT2 cells from normal lung, which was consistent with our findings in the GEMM, murine organoid, and human iAT2 cell model systems (Figures 6F and S6B). To our knowledge, this is the first documentation of loss of AT2 cell identity in human early-stage LUAD patient samples.

Comparison of GEMM and Murine and Human Organoid Models and Early-Stage LUAD Patient Datasets

Comparison of the transcriptional profiles in the model systems showed that our murine and human organoid systems recapitulated transcriptional changes in the GEMM and in early-stage lung cancer patients, revealing shared downregulation of alveolar differentiation markers. To further compare all four scRNA-seq datasets with each other, we calculated *Z* scores for each cell in our murine $Kras^{G12D}$ organoid dataset using gene signatures derived from our DE analysis and previously published signatures. As expected, control AT2 cells from the GEMM correlated with control murine organoid AT2 cells; organoid control AT2 cluster C1 was most similar to the AT2 cell YFP^- cluster signature from the GEMM model (Figure 6G). Furthermore, murine organoid cells with oncogenic KRAS and GEMM cells with oncogenic KRAS were transcriptionally similar; murine organoid C0 and C2 correlated with the AT2 cell YFP^+ GEMM signatures. The human *KRAS* iAT2 and *KRAS* mutant patient datasets were similar to murine $KRAS^{G12D}$ organoid cells, with iAT2 and patient cells most closely resembling C0. Moreover, we found that our murine organoids correlated with the gene expression signature of lung cancer progression (“LUAD progression”) from a report that used the $Kras^{G12D}$ GEMM (Neidler et al., 2019). The organoid datasets also correlated with the HALLMARK_WNT signature (The Molecular Signatures Database), demonstrating how our organoids recapitulate the GEMM, because the Wnt pathway has been shown to play an important role in lung cancer progression (Tammela et al., 2017).

Taken together, our resource provides omics analyses of three models and patient-derived *KRAS*-driven LUAD at its earliest stages. Our data suggest that reduction of the mature AT2 transcription program is an important early step in *KRAS*-driven LUAD initiation. Furthermore, we demonstrated in GEMM and human stage IA LUAD patients that only AT2 cells transition to a transcriptionally distinct state during the early stage of KRAS

(C) Volcano plots indicating differential protein (left) and phosphoprotein (right) expression between dox-induced and control iAT2 cells.
(D) Top 10 upregulated pathways in dox-induced compared with control iAT2 cells based on phosphoproteomics analysis.
(E) FACS analysis of iAT2 cells over three passages following initiation of dox versus control vehicle (DMSO) treatment. The mean fluorescence intensity (MFI) of tdTomato is indicated.
(F) Log expression of the indicated genes. The p values were determined using a Model-based Analysis of Single-cell Transcriptomics (MAST) single-cell test. *$p < 0.05$.
(G) Log expression of the indicated gene signatures. The p values were determined using a Welch two-sample t test. *$p < 0.05$.
(H) Log expression of the indicated genes. The p values were determined using a MAST single-cell test. *$p < 0.05$.
See also Figure S5.

Please cite this article in press as: Dost et al., Organoids Model Transcriptional Hallmarks of Oncogenic KRAS Activation in Lung Epithelial Progenitor Cells, Cell Stem Cell (2020), https://doi.org/10.1016/j.stem.2020.07.022

CellPress

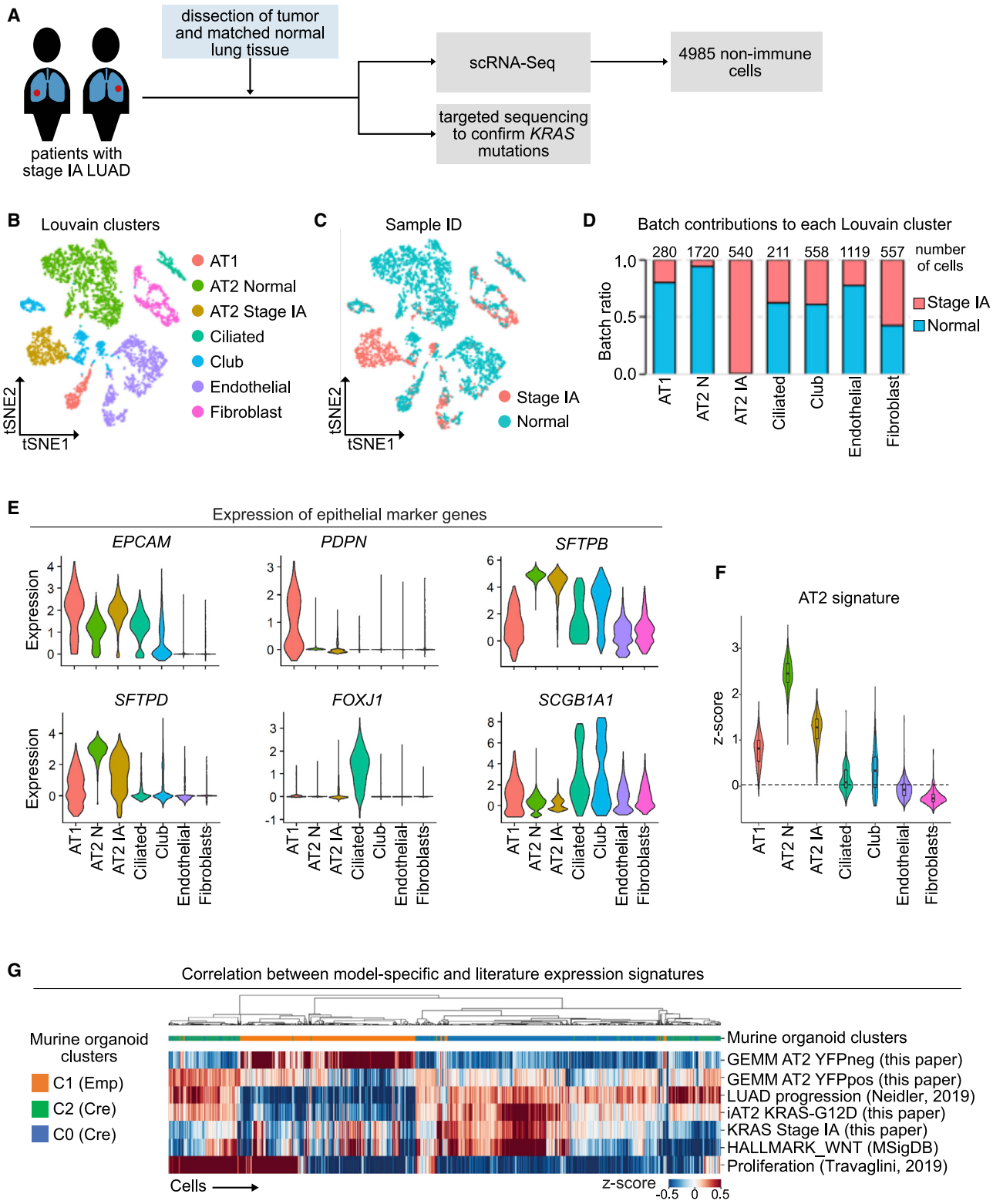


Figure 6. Differentiation and Maturation Markers Are Downregulated in AT2 Cells from Human Early-Stage LUAD

(A) Experimental strategy to obtain cells from human early-stage IA LUAD for scRNA-seq.

(B and C) Louvain clustering of transcriptomes of non-immune LUAD cell types and matching normal lung tissue. Cells are colored based on (B) Louvain clusters or (C) sample ID.

(D) Batch contributions to each Louvain cluster shown in (B) with the number of cells indicated.

(legend continued on next page)

Please cite this article in press as: Dost et al., Organoids Model Transcriptional Hallmarks of Oncogenic KRAS Activation in Lung Epithelial Progenitor Cells, Cell Stem Cell (2020), https://doi.org/10.1016/j.stem.2020.07.022

tumorigenesis. Additionally, the *in-vitro*-induced human and murine organoid systems, which recapitulate core components of early-stage LUAD progression, provide rapid and easily perturbed models for investigation of lung cancer biology.

DISCUSSION

In our studies, we showed that developmental gene signatures are present in early-stage, non-metastasizing LUAD, indicating that alveolar cells lose differentiation markers early after activation of oncogenic KRAS. To our knowledge, this is the first time it has been shown that loss of differentiation occurs in early-stage LUAD. It is a long-held notion that tumor cells hijack developmental programs. However, this process has been thought to occur in late-stage, metastasizing tumors (Kulesa et al., 2013; Nieto, 2013; Thiery, 2002; Yang and Weinberg, 2008). In humans, SOX9 protein levels are correlated with a higher NSCLC tumor stage and worse survival (Jiang et al., 2010; Zhou et al., 2012). In mouse models, primary tumors that have metastasized contain cells that have lost NKX2-1 and express HMGA2 (Winslow et al., 2011). *Sox9*, *Nkx2-1*, and *Hmga2* are genes with important functions during embryonic lung development (Alanis et al., 2014; Maeda et al., 2007; Singh et al., 2014). SOX9 has been shown to work together with KRAS in lung development to maintain a balance between branching morphogenesis and alveolar differentiation (Chang et al., 2013). Although NKX2-1 is still present in adult lung epithelial cells, SOX9 and HMGA2 are not found in healthy adult lung epithelium (Nikolić et al., 2017; Pfannkuche et al., 2009). Our organoid systems can be used to identify transcriptional states in cells bearing oncogenic KRAS, which distinguishes them from their normal adult epithelial counterparts, shedding light on new ways to intervene in lung cancer progression.

Our findings present a murine organoid system that can be used as a tool to study tumor initiation and progression in a controlled environment. We directly compared our KY AT2 cell-derived organoids with AT2 cells with activated $KRAS^{G12D}$ *in vivo* at an early-stage time point. We observed corresponding transcriptional changes in day 7 organoids and *in vivo* in cells 7 weeks after induction. Therefore, we hypothesize that the tumor organoids recapitulate LUAD progression in an accelerated manner. Furthermore, our transplantation studies showed that KRAS tumor organoids can be orthotopically transplanted. Therefore, our organoid system can be used for *in vitro* manipulation and subsequent transplantation to facilitate study of potential therapeutic targets on lung cancer development and progression. This creates an exciting opportunity to model lung cancer tumorigenesis on an accelerated timescale while maintaining the core transcriptional signatures that appear during tumor progression in a manner that is compatible with genetic or chemical perturbations prior to transplantation.

Our murine organoid data show remarkable similarities to our GEMM and human datasets. However, there we also found differences. Some cells from the organoids have high expression levels of *Sox9* and *Hmga2* and stain positive for HMGA2, whereas transcriptional upregulation of *Sox9* and its targets in our GEMM data is rather modest. Furthermore, we see strong downregulation of the AT2 cell signature in murine and human organoids with almost complete loss of SPC expression in murine organoids, whereas downregulation of the signature is more subtle in our GEMM data. One explanation is that the organoids are in a state of unrestrained proliferation. Therefore, it is conceivable that our organoids progress fast, whereas cells in the GEMM model receive inhibition cues from the microenvironment or are being cleared by the immune system. Indeed, it is difficult to compare the timelines of the organoids with the timeline of tumor progression *in vivo*. Nevertheless, because of the defined and easy-to-manipulate culture conditions, we think that organoids are an advantageous system to study the direct effect of $KRAS^{G12D}$ expression on AT2 cells.

Our work provides murine and human organoid systems to study LUAD progression rapidly *in vitro*. We analyzed our murine organoids, human iAT2 cell organoids, $KRAS^{G12D}$ GEMM, and stage IA patient data and provided these datasets to the research community. Our comparison of the single-cell datasets revealed a common loss of AT2 cell identity as an early-occurring event following KRAS pathway activation in all four contexts. These comparisons also revealed the utility of our murine tumor organoid system in modeling human lung cancer driven by *KRAS* mutagenesis in its earliest stage. Bulk RNA-seq, proteomics, and phosphoproteomics validate our findings in the single-cell datasets and are an additional resource for data mining. Our data may be a useful component of cancer atlas projects and screening candidate drug targets to prevent progression of early stage LUAD in *KRAS* mutant patients followed by proof-of-principle testing. Additionally, the organoid tools could have utility in the cancer modeling field and drug screening.

Limitations of Study

Although the observed loss of alveolar identity markers was validated at the protein level in our organoid cultures, we only present evidence of downregulation of these markers in our GEMM and stage IA patient data on a transcriptional level. In future work, we will examine GEMM and patient samples by immunohistochemistry to confirm the changes in AT2 cell marker expression and altered expression of TFs and their targets. Further studies are also required to determine whether decreased *NKX2-1* expression is an important early consequence of expression of oncogenic KRAS in human cells. In a related manner, there are multiple upstream signaling pathways connected to genes in our analysis that could cause loss of the AT2 cell differentiation phenotype, including *Sox9*, Wnt, and *Nkx2-1*. Determining the role of AT2 cell lineage identity and other observed transcriptional changes in LUAD progression will be important. In-depth analysis of lineage plasticity and assessment of the transcriptomics, proteomics, and functional heterogeneity in cells expressing oncogenic KRAS in early-stage LUAD will be another interesting topic of

(E) Violin plots showing gene expression values of selected genes in the annotated clusters shown in (B).
(F) *Z* score of gene signature comprised of AT2 cell signature genes shared between mouse and human from the Panglao database. A dashed line marks y = 0.
(G) Transcriptional comparison of KRAS LUAD models and correlation heatmap of individual cells of the organoid scRNA-seq data (x axis) and z-normalized gene signatures (y axis). Cells are ordered based on correlation distance calculation. Louvain clusters are annotated.
See also Figure S6.

future work. Finally, although our organoid models of oncogenic KRAS activation provide rapid ways to identify possible therapeutic avenues for early-stage LUAD, we have not yet validated a new therapeutic lead generated from our data.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead Contact
  - Materials Availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Mouse cohorts
  - Stage IA LUAD patient information
- METHOD DETAILS
  - Mouse studies
  - Human iPSC studies
  - Patient stage IA lung cancer studies
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Statistics

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.stem.2020.07.022.

ACKNOWLEDGMENTS

We thank members of the Kim, Kotton, Dubinett, and Yanagawa labs for helpful discussions; the Boston Children's Hospital Flow Cytometry Core Facility; the Harvard Medical School Single-Cell Core Facility for use of their 10X Genomics Chromium Controller; the Harvard Medical School Biopolymers Facility for Illumina NextSeq500 Sequencing and cDNA/library quality control experiments; Yuriy Alekseyev and M.J. Mistretta of the Boston University School of Medicine (BUSM) Single Cell Sequencing Core; Brian R. Tilton of the BUSM Flow Cytometry Core; Greg Miller (The Center for Regenerative Medicine (CReM) Laboratory Manager), Marianne James (CReM iPSC Core Manager), Lauren Winter, and Tamara Silva for administrative support; the UCLA Translational Pathology Core Laboratory and the UCLA Jonsson Comprehensive Cancer Center for shared resources; and the Rodent Histopathology Facility at Harvard Medical School and the Molecular Biology Core Facility at Dana-Faber Cancer Institute for library preparation, sequencing, and data analysis of bulk RNA-seq. A.F.M.D is supported by a Boehringer Ingelheim Fonds PhD fellowship. A.L.M. was supported by an NIH R01 postdoctoral supplement and, currently, by a Damon Runyon Cancer Research Foundation postdoctoral fellowship (DRG:2368-19) and a Postdoctoral Enrichment Program Award from the Burroughs Wellcome Fund. S.M.L. is supported by a Hope Funds for Cancer Research postdoctoral fellowship and S.P.R. by an IASLC Young Investigator Fellowship. This work was supported in part by R01 HL090136, R01 HL132266, R01 HL125821, U01 HL100402, RFA-HL-09-004, R35HL150876-01, American Cancer Society Research Scholar Grant RSG-08-082-01-MGO, the V Foundation for Cancer Research, the Thoracic Foundation, the Ellison Foundation, American Lung Association LCD-619492, and the Harvard Stem Cell Institute (to C.F.K.). CReM was supported by grants R24HL123828 and U01TR001810. D.N.K. is supported by R01HL128172, R01HL095993, R01HL122442, U01HL134745, and U01HL134766. iPSC model development and characterization (D.N.K., G.M., A.E., C.V., D.H., M.V., R.H., R.H.K., and B.C.B.) was sponsored by the Lung Cancer Initiative at Johnson & Johnson. A.E. acknowledges generous start-up funding and ongoing support from Boston University to support the operations of the CNSB. This work was also supported by a Stand Up To Cancer-LUNGevity-American Lung Association Lung Cancer Interception Dream Team translational cancer research grant (SU2C-AACR-DT23-17). Stand Up To Cancer is a division of the Entertainment Industry Foundation. Research grants are administered by the American Association for Cancer Research, the scientific partner of SU2C (S.M.D.). This work was further supported by NIH/NCI Molecular Characterization Laboratory 5U01CA196408-04 (to S.M.D.), UC Tobacco-Related Disease Research Program (TRDRP) 27IR-0036 (to K.K.), and a Thoracic Surgery Foundation research award (to E.F.).

AUTHOR CONTRIBUTIONS

Conceptualization, A.F.M.D., A.L.M., M.V., J.H., D.N.K., A.E., G.M., J.Y., S.M.D., K.K., L.M.T., and C.F.K.; Methodology, A.F.M.D., A.L.M., M.V., D.H., J.H., D.N.K., A.E., E.F., L.M.T., K.K., and G.S.; Software, A.L.M. and M.E.P.; Validation, P.J.B.; Investigation, A.F.M.D., A.L.M., S.M.L., S.P.R., J.S.d.A., M.V., J.H., D.H., R.H., B.C.B., E.F., and J.H.K.; Formal Analysis, C.V.-M., R.H., L.M.T., and B.C.B.; Writing – Original Draft, A.F.M.D., A.L.M., M.V., D.N.K., C.F.K.; Writing – Review & Editing, A.F.M.D., A.L.M., S.M.L., S.P.R., P.J.B., M.V., D.H., C.V.-M., D.N.K., R.H., A.E., J.Y., L.M.T., S.M.D., K.K., and G.M.; Visualization, A.F.M.D. and A.L.M.; Resources, J.Y., W.D.W., and G.A.F.; Funding Acquisition, A.F.M.D., A.L.M., D.N.K., G.M., A.E., S.M.D., K.K., and C.F.K.; Supervision, R.T.B., D.N.K., G.M., A.E., and C.F.K.

DECLARATION OF INTERESTS

W.D.W. is a member of the Leica Biosystems Medical Imaging Advisory Board. S.M.D. is on the Scientific Advisory Boards of EarlyDiagnostics; Johnson & Johnson Lung Cancer Initiative; LungLife AI; and T-Cure Bioscience. He has received research funding from Johnson & Johnson Lung Cancer Initiative and Novartis. C.F.K. has a sponsored research agreement from Celgene/BMS, but this funding did not support the research described in this manuscript.

Received: May 26, 2020
Revised: July 9, 2020
Accepted: July 29, 2020
Published: September 4, 2020

REFERENCES

Alanis, D.M., Chang, D.R., Akiyama, H., Krasnow, M.A., and Chen, J. (2014). Two nested developmental waves demarcate a compartment boundary in the mouse lung. Nat. Commun. *5*, 3923.

Antonângelo, L., Tuma, T., Fabro, A., Acencio, M., Terra, R., Parra, E., Vargas, F., Takagaki, T., and Capelozzi, V. (2016). Id-1, Id-2, and Id-3 co-expression correlates with prognosis in stage I and II lung adenocarcinoma patients treated with surgery and adjuvant chemotherapy. Exp. Biol. Med. (Maywood) *241*, 1159–1168.

Barbie, D.A., Tamayo, P., Boehm, J.S., Kim, S.Y., Moody, S.E., Dunn, I.F., Schinzel, A.C., Sandy, P., Meylan, E., Scholl, C., et al. (2009). Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. Nature *462*, 108–112.

Bild, A.H., Yao, G., Chang, J.T., Wang, Q., Potti, A., Chasse, D., Joshi, M.B., Harpole, D., Lancaster, J.M., Berchuck, A., et al. (2006). Oncogenic pathway signatures in human cancers as a guide to targeted therapies. Nature *439*, 353–357.

Blondel, V.D., Guillaume, J.-L., Lambiotte, R., and Lefebvre, E. (2008). Fast unfolding of communities in large networks. J. Stat. Mech. Theory Exp. *2008*, P10008.

Bourgon, R., Gentleman, R., and Huber, W. (2010). Independent filtering increases detection power for high-throughput experiments. Proc. Natl. Acad. Sci. USA *107*, 9546–9551.

Chang, D.R., Martinez Alanis, D., Miller, R.K., Ji, H., Akiyama, H., McCrea, P.D., and Chen, J. (2013). Lung epithelial branching program antagonizes alveolar differentiation. Proc. Natl. Acad. Sci. USA *110*, 18042–18051.

Chen, H., Liu, H., and Qing, G. (2018). Targeting oncogenic Myc as a strategy for cancer treatment. Signal Transduct. Target. Ther. *3*, 5.

Cheng, Y.J., Tsai, J.W., Hsieh, K.C., Yang, Y.C., Chen, Y.J., Huang, M.S., and Yuan, S.S. (2011). Id1 promotes lung cancer cell proliferation and tumor growth through Akt-related pathway. Cancer Lett. *307*, 191–199.

Collisson, E.A., Campbell, J.D., Brooks, A.N., Berger, A.H., Lee, W., Chmielecki, J., Beer, D.G., Cope, L., Creighton, C.J., Danilova, L., et al. (2014). Comprehensive molecular profiling of lung adenocarcinoma: The cancer genome atlas research network. Nature *511*, 543–550.

Cornwell, M., Vangala, M., Taing, L., Herbert, Z., Köster, J., Li, B., Sun, H., Li, T., Zhang, J., Qiu, X., et al. (2018). VIPER: Visualization Pipeline for RNA-seq, a Snakemake workflow for efficient and complete RNA-seq analysis. BMC Bioinformatics *19*, 135.

Cox, J., Neuhauser, N., Michalski, A., Scheltema, R.A., Olsen, J.V., and Mann, M. (2011). Andromeda: a peptide search engine integrated into the MaxQuant environment. J. Proteome Res. *10*, 1794–1805.

Curtis, S.J., Sinkevicius, K.W., Li, D., Lau, A.N., Roach, R.R., Zamponi, R., Woolfenden, A.E., Kirsch, D.G., Wong, K.K., and Kim, C.F. (2010). Primary tumor genotype is an important determinant in identification of lung cancer propagating cells. Cell Stem Cell *7*, 127–133.

Dang, C.V. (2012). MYC on the path to cancer. Cell *149*, 22–35.

Deutsch, E.W., Bandeira, N., Sharma, V., Perez-Riverol, Y., Carver, J.J., Kundu, D.J., García-Seisdedos, D., Jarnuczak, A.F., Hewapathirana, S., Pullman, B.S., et al. (2020). The ProteomeXchange consortium in 2020: enabling 'big data' approaches in proteomics. Nucleic Acids Res. *48* (D1), D1145–D1152.

Diaz-Papkovich, A., Anderson-Trocmé, L., Ben-Eghan, C., and Gravel, S. (2019). UMAP reveals cryptic population structure and phenotype heterogeneity in large genomic cohorts. PLoS Genet. *15*, e1008432.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21.

Drost, J., van Jaarsveld, R.H., Ponsioen, B., Zimberlin, C., van Boxtel, R., Buijs, A., Sachs, N., Overmeer, R.M., Offerhaus, G.J., Begthel, H., et al. (2015). Sequential cancer mutations in cultured human intestinal stem cells. Nature *521*, 43–47.

DuPage, M., Dooley, A.L., and Jacks, T. (2009). Conditional mouse lung cancer models using adenoviral or lentiviral delivery of Cre recombinase. Nat. Protoc. *4*, 1064–1072.

Finak, G., McDavid, A., Yajima, M., Deng, J., Gersuk, V., Shalek, A.K., Slichter, C.K., Miller, H.W., McElrath, M.J., Prlic, M., et al. (2015). MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. Genome Biol. *16*, 278.

Franzén, O., Gan, L.-M., and Björkegren, J.L.M. (2019). PanglaoDB: a web server for exploration of mouse and human single-cell RNA sequencing data. Database (Oxford) *2019*, baz046.

Hafemeister, C., and Satija, R. (2019). Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. Genome Biol. *20*, 296.

Haigis, K.M. (2017). KRAS Alleles: The Devil Is in the Detail. Trends Cancer *3*, 686–697.

Han, H., Cho, J.W., Lee, S., Yun, A., Kim, H., Bae, D., Yang, S., Kim, C.Y., Lee, M., Kim, E., et al. (2018). TRRUST v2: an expanded reference database of human and mouse transcriptional regulatory interactions. Nucleic Acids Res. *46* (D1), D380–D386.

Herriges, J.C., Verheyden, J.M., Zhang, Z., Sui, P., Zhang, Y., Anderson, M.J., Swing, D.A., Zhang, Y., Lewandoski, M., and Sun, X. (2015). FGF-Regulated ETV Transcription Factors Control FGF-SHH Feedback Loop in Lung Branching. Dev. Cell *35*, 322–332.

Hu, H., Miao, Y.-R., Jia, L.-H., Yu, Q.-Y., Zhang, Q., and Guo, A.-Y. (2019). AnimalTFDB 3.0: a comprehensive resource for annotation and prediction of animal transcription factors. Nucleic Acids Res. *47* (D1), D33–D38.

Hunter, J.D. (2007). Matplotlib: A 2D graphics environment. Comput. Sci. Eng. *9*, 99–104.

Hurley, K., Ding, J., Villacorta-Martin, C., Herriges, M.J., Jacob, A., Vedaie, M., Alysandratos, K.D., Sun, Y.L., Lin, C., Werder, R.B., et al. (2020). Reconstructed Single-Cell Fate Trajectories Define Lineage Plasticity Windows during Differentiation of Human PSC-Derived Distal Lung Progenitors. Cell Stem Cell *26*, 593–608.e8.

Jackson, E.L., Willis, N., Mercer, K., Bronson, R.T., Crowley, D., Montoya, R., Jacks, T., and Tuveson, D.A. (2001). Analysis of lung tumor initiation and progression using conditional expression of oncogenic K-ras. Genes Dev. *15*, 3243–3248.

Jackson, E.L., Olive, K.P., Tuveson, D.A., Bronson, R., Crowley, D., Brown, M., and Jacks, T. (2005). The differential effects of mutant p53 alleles on advanced murine lung cancer. Cancer Res. *65*, 10280–10288.

Jacob, A., Morley, M., Hawkins, F., McCauley, K.B., Jean, J.C., Heins, H., Na, C.L., Weaver, T.E., Vedaie, M., Hurley, K., et al. (2017). Differentiation of Human Pluripotent Stem Cells into Functional Lung Alveolar Epithelial Cells. Cell Stem Cell *21*, 472–488.e10.

Jacob, A., Vedaie, M., Roberts, D.A., Thomas, D.C., Villacorta-Martin, C., Alysandratos, K.D., Hawkins, F., and Kotton, D.N. (2019). Derivation of self-renewing lung alveolar epithelial type II cells from human pluripotent stem cells. Nat. Protoc. *14*, 3303–3332.

Jiang, S.S., Fang, W.T., Hou, Y.H., Huang, S.F., Yen, B.L., Chang, J.L., Li, S.M., Liu, H.P., Liu, Y.L., Huang, C.T., et al. (2010). Upregulation of SOX9 in lung adenocarcinoma and its involvement in the regulation of cell growth and tumorigenicity. Clin. Cancer Res. *16*, 4363–4373.

Kaisani, A., Delgado, O., Fasciani, G., Kim, S.B., Wright, W.E., Minna, J.D., and Shay, J.W. (2014). Branching morphogenesis of immortalized human bronchial epithelial cells in three-dimensional culture. Differentiation *87*, 119–126.

Kim, C.F., Jackson, E.L., Woolfenden, A.E., Lawrence, S., Babar, I., Vogel, S., Crowley, D., Bronson, R.T., and Jacks, T. (2005). Identification of bronchioalveolar stem cells in normal lung and lung cancer. Cell *121*, 823–835.

Kim, M., Mun, H., Sung, C.O., Cho, E.J., Jeon, H.J., Chun, S.M., Jung, D.J., Shin, T.H., Jeong, G.S., Kim, D.K., et al. (2019). Patient-derived lung cancer organoids as in vitro cancer models for therapeutic screening. Nat. Commun. *10*, 3991.

Kruspig, B., Monteverde, T., Neidler, S., Hock, A., Kerr, E., Nixon, C., Clark, W., Hedley, A., Laing, S., Coffelt, S.B., et al. (2018). The ERBB network facilitates KRAS-driven lung tumorigenesis. Sci. Transl. Med. *10*, eaao2565.

Kulesa, P.M., Morrison, J.A., and Bailey, C.M. (2013). The neural crest and cancer: a developmental spin on melanoma. Cells Tissues Organs (Print) *198*, 12–21.

Kuleshov, M.V., Jones, M.R., Rouillard, A.D., Fernandez, N.F., Duan, Q., Wang, Z., Koplev, S., Jenkins, S.L., Jagodnik, K.M., Lachmann, A., et al. (2016). Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res. *44* (W1), W90-7.

La Manno, G., Soldatov, R., Zeisel, A., Braun, E., Hochgerner, H., Petukhov, V., Lidschreiber, K., Kastriti, M.E., Lönnerberg, P., Furlan, A., et al. (2018). RNA velocity of single cells. Nature *560*, 494–498.

Lau, A.N., Curtis, S.J., Fillmore, C.M., Rowbotham, S.P., Mohseni, M., Wagner, D.E., Beede, A.M., Montoro, D.T., Sinkevicius, K.W., Walton, Z.E., et al. (2014). Tumor-propagating cells and Yap/Taz activity contribute to lung tumor progression and metastasis. EMBO J. *33*, 468–481.

Laughney, A.M., Hu, J., Campbell, N.R., Bakhoum, S.F., Setty, M., Lavallée, V.P., Xie, Y., Masilionis, I., Carr, A.J., Kottapalli, S., et al. (2020). Regenerative lineages and immune-mediated pruning in lung cancer metastasis. Nat. Med. *26*, 259–269.

Lee, J.-H., Bhang, D.H., Beede, A., Huang, T.L., Stripp, B.R., Bloch, K.D., Wagers, A.J., Tseng, Y.-H., Ryeom, S., and Kim, C.F. (2014). Lung stem cell differentiation in mice directed by endothelial cells via a BMP4-NFATc1-thrombospondin-1 axis. Cell *156*, 440–455.

Lee, J.-H., Tammela, T., Hofree, M., Choi, J., Marjanovic, N.D., Han, S., Canner, D., Wu, K., Paschini, M., Bhang, D.H., et al. (2017). Anatomically and Functionally Distinct Lung Mesenchymal Populations Marked by Lgr5 and Lgr6. Cell *170*, 1149–1163.e12.

Leutert, M., Rodríguez-Mias, R.A., Fukuda, N.K., and Villén, J. (2019). R2-P2 rapid-robotic phosphoproteomics enables multidimensional cell signaling studies. Mol. Syst. Biol. *15*, e9021.

Li, X., Nadauld, L., Ootani, A., Corney, D.C., Pai, R.K., Gevaert, O., Cantrell, M.A., Rack, P.G., Neal, J.T., Chan, C.W.M., et al. (2014). Oncogenic transformation of diverse gastrointestinal tissues in primary organoid culture. Nat. Med. *20*, 769–777.

Li, L., Xu, B., Zhang, H., Wu, J., Song, Q., and Yu, J. (2020). Potentiality of forkhead box Q1 as a biomarker for monitoring tumor features and predicting prognosis in non-small cell lung cancer. J. Clin. Lab. Anal. *34*, e23031.

Lin, C., Song, H., Huang, C., Yao, E., Gacayan, R., Xu, S.-M., and Chuang, P.-T. (2012). Alveolar type II cells possess the capability of initiating lung tumor development. PLoS ONE *7*, e53817.

Liu, Q., Liu, K., Cui, G., Huang, X., Yao, S., Guo, W., Qin, Z., Li, Y., Yang, R., Pu, W., et al. (2019). Lung regeneration by multipotent stem cells residing at the bronchioalveolar-duct junction. Nat. Genet. *51*, 728–738.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550.

Maeda, Y., Davé, V., and Whitsett, J.A. (2007). Transcriptional control of lung morphogenesis. Physiol. Rev. *87*, 219–244.

Matano, M., Date, S., Shimokawa, M., Takano, A., Fujii, M., Ohta, Y., Watanabe, T., Kanai, T., and Sato, T. (2015). Modeling colorectal cancer using CRISPR-Cas9-mediated engineering of human intestinal organoids. Nat. Med. *21*, 256–262.

McCauley, K.B., Hawkins, F., Serra, M., Thomas, D.C., Jacob, A., and Kotton, D.N. (2017). Efficient Derivation of Functional Human Airway Epithelium from Pluripotent Stem Cells via Temporal Regulation of Wnt Signaling. Cell Stem Cell *20*, 844–857.e6.

Meylan, E., Dooley, A.L., Feldser, D.M., Shen, L., Turk, E., Ouyang, C., and Jacks, T. (2009). Requirement for NF-kappaB signalling in a mouse model of lung adenocarcinoma. Nature *462*, 104–107.

Morrisey, E.E., and Hogan, B.L.M. (2010). Preparing for the first breath: genetic and cellular mechanisms in lung development. Dev. Cell *18*, 8–23.

Neidler, S., Kruspig, B., Hewit, K., Monteverde, T., Gyuraszova, K., Braun, A., Clark, W., James, D., Hedley, A., Nieswandt, B., et al. (2019). Identification of a clinically relevant signature for early progression in KRAS-driven lung adenocarcinoma. Cancers (Basel) *11*, 600.

Nieto, M.A. (2013). Epithelial plasticity: A common theme in embryonic and cancer cells. Science *342*, 1234850.

Nikolić, M.Z., Caritg, O., Jeng, Q., Johnson, J.A., Sun, D., Howell, K.J., Brady, J.L., Laresgoiti, U., Allen, G., Butler, R., et al. (2017). Human embryonic lung epithelial tips are multipotent progenitors that can be expanded in vitro as long-term self-renewing organoids. eLife *6*, 6.

Ordovás, L., Boon, R., Pistoni, M., Chen, Y., Wolfs, E., Guo, W., Sambathkumar, R., Bobis-Wozowicz, S., Helsen, N., Vanhove, J., et al. (2015). Efficient recombinase-mediated cassette exchange in hPSCs to study the hepatocyte lineage reveals AAVS1 locus-mediated transgene inhibition. Stem Cell Reports *5*, 918–931.

Perez-Riverol, Y., Csordas, A., Bai, J., Bernal-Llinares, M., Hewapathirana, S., Kundu, D.J., Inuganti, A., Griss, J., Mayer, G., Eisenacher, M., et al. (2019). The PRIDE database and related tools and resources in 2019: improving support for quantification data. Nucleic Acids Res. *47* (D1), D442–D450.

Pfannkuche, K., Summer, H., Li, O., Hescheler, J., and Dröge, P. (2009). The high mobility group protein HMGA2: a co-regulator of chromatin structure and pluripotency in stem cells? Stem Cell Rev. Rep. *5*, 224–230.

Phipson, B., Lee, S., Majewski, I.J., Alexander, W.S., and Smyth, G.K. (2016). Robust hyperparameter estimation protects against hypervariable genes and improves power to detect differential expression. Ann. Appl. Stat. *10*, 946–963.

Pillai, S., Rizwani, W., Li, X., Rawal, B., Nair, S., Schell, M.J., Bepler, G., Haura, E., Coppola, D., and Chellappan, S. (2011). ID1 facilitates the growth and metastasis of non-small cell lung cancer in response to nicotinic acetylcholine receptor and epidermal growth factor receptor signaling. Mol. Cell. Biol. *31*, 3052–3067.

Poole, C.J., and van Riggelen, J. (2017). MYC—master regulator of the cancer epigenome and transcriptome. Genes (Basel) *8*, 142.

Sachs, N., Papaspyropoulos, A., Zomer-van Ommen, D.D., Heo, I., Böttinger, L., Klay, D., Weeber, F., Huelsz-Prince, G., Iakobachvili, N., Amatngalim, G.D., et al. (2019). Long-term expanding human airway organoids for disease modeling. EMBO J. *38*, e100300.

Salwig, I., Spitznagel, B., Vazquez-Armendariz, A.I., Khalooghi, K., Guenther, S., Herold, S., Szibor, M., and Braun, T. (2019). Bronchioalveolar stem cells are a main source for regeneration of distal lung epithelia *in vivo*. EMBO J. *38*, e102099.

Schneider, C.A., Rasband, W.S., and Eliceiri, K.W. (2012). NIH Image to ImageJ: 25 years of image analysis. Nat. Methods *9*, 671–675.

Seino, T., Kawasaki, S., Shimokawa, M., Tamagawa, H., Toshimitsu, K., Fujii, M., Ohta, Y., Matano, M., Nanki, K., Kawasaki, K., et al. (2018). Human Pancreatic Tumor Organoids Reveal Loss of Stem Cell Niche Factor Dependence during Disease Progression. Cell Stem Cell *22*, 454–467.e6.

Sergushichev, A.A. (2016). An algorithm for fast preranked gene set enrichment analysis using cumulative statistic calculation. bioRxiv. https://doi.org/10.1101/060012.

Singh, I., Mehta, A., Contreras, A., Boettger, T., Carraro, G., Wheeler, M., Cabrera-Fuentes, H.A., Bellusci, S., Seeger, W., Braun, T., and Barreto, G. (2014). Hmga2 is required for canonical WNT signaling during lung development. BMC Biol. *12*, 21.

Soneson, C., and Robinson, M.D. (2018). Bias, robustness and scalability in single-cell differential expression analysis. Nat. Methods *15*, 255–261.

Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., 3rd, Hao, Y., Stoeckius, M., Smibert, P., and Satija, R. (2019). Comprehensive Integration of Single-Cell Data. Cell *177*, 1888–1902.e21.

Takahashi, K., and Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell *126*, 663–676.

Tammela, T., Sanchez-Rivera, F.J., Cetinbas, N.M., Wu, K., Joshi, N.S., Helenius, K., Park, Y., Azimi, R., Kerper, N.R., Wesselhoeft, R.A., et al. (2017). A Wnt-producing niche drives proliferative potential and progression in lung adenocarcinoma. Nature *545*, 355–359.

Thiery, J.P. (2002). Epithelial-mesenchymal transitions in tumour progression. Nat. Rev. Cancer *2*, 442–454.

Tirosh, I., Izar, B., Prakadan, S.M., Wadsworth, M.H., Treacy, D., Trombetta, J.J., Rotem, A., Rodman, C., Lian, C., Murphy, G., et al. (2016). Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. Science *352*, 189–196.

Tiyaboonchai, A., Mac, H., Shamsedeen, R., Mills, J.A., Kishore, S., French, D.L., and Gadue, P. (2014). Utilization of the AAVS1 safe harbor locus for hematopoietic specific transgene expression and gene knockdown in human ES cells. Stem Cell Res. (Amst.) *12*, 630–637.

Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J., and Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat. Biotechnol. *28*, 511–515.

Travaglini, K.J., Nabhan, A.N., Penland, L., Sinha, R., Gillich, A., Sit, R.V., Chang, S., Conley, S.D., Mori, Y., Seita, J., et al. (2019). A molecular cell atlas of the human lung from single cell RNA sequencing. bioRxiv. https://doi.org/10.1101/742320.

Tuveson, D.A., Shaw, A.T., Willis, N.A., Silver, D.P., Jackson, E.L., Chang, S., Mercer, K.L., Grochow, R., Hock, H., Crowley, D., et al. (2004). Endogenous oncogenic K-ras(G12D) stimulates proliferation and widespread neoplastic and developmental defects. Cancer Cell *5*, 375–387.

van Dijk, D., Sharma, R., Nainys, J., Yim, K., Kathail, P., Carr, A.J., Burdziak, C., Moon, K.R., Chaffer, C.L., Pattabiraman, D., et al. (2018). Recovering Gene Interactions from Single-Cell Data Using Data Diffusion. Cell *174*, 716–729.e27.

Winslow, M.M., Dayton, T.L., Verhaak, R.G.W., Kim-Kiselak, C., Snyder, E.L., Feldser, D.M., Hubbard, D.D., DuPage, M.J., Whittaker, C.A., Hoersch, S.,

et al. (2011). Suppression of lung adenocarcinoma progression by Nkx2-1. Nature *473*, 101–104.

Wolf, F.A., Angerer, P., and Theis, F.J. (2018). SCANPY: large-scale single-cell gene expression data analysis. Genome Biol. *19*, 15.

Xu, X., Rock, J.R., Lu, Y., Futtner, C., Schwab, B., Guinney, J., Hogan, B.L.M., and Onaitis, M.W. (2012). Evidence for type II cells as cells of origin of K-Ras-induced distal lung adenocarcinoma. Proc. Natl. Acad. Sci. USA *109*, 4910–4915.

Yang, J., and Weinberg, R.A. (2008). Epithelial-mesenchymal transition: at the crossroads of development and tumor metastasis. Dev. Cell *14*, 818–829.

Zacharias, W.J., Frank, D.B., Zepp, J.A., Morley, M.P., Alkhaleel, F.A., Kong, J., Zhou, S., Cantu, E., and Morrisey, E.E. (2018). Regeneration of the lung alveolus by an evolutionarily conserved epithelial progenitor. Nature *555*, 251–255.

Zhang, H., Brainson, C.F., Koyama, S., Redig, A.J., Chen, T., Li, S., Gupta, M., Garcia-de-alba, C., Paschini, M., Herter-sprie, G.S., et al. (2017a). Lkb1 inactivation drives lung cancer lineage switching governed by Polycomb Repressive Complex 2. Nat. Commun. *8*, 1–14.

Zhang, Z., Newton, K., Kummerfeld, S.K., Webster, J., Kirkpatrick, D.S., Phu, L., Eastham-Anderson, J., Liu, J., Lee, W.P., Wu, J., et al. (2017b). Transcription factor Etv5 is essential for the maintenance of alveolar type II cells. Proc. Natl. Acad. Sci. USA *114*, 3903–3908.

Zhou, C.H., Ye, L.P., Ye, S.X., Li, Y., Zhang, X.Y., Xu, X.Y., and Gong, L.Y. (2012). Clinical significance of SOX9 in human non-small cell lung cancer progression and overall patient survival. J. Exp. Clin. Cancer Res. *31*, 18.

Zhu, Z., Golay, H.G., and Barbie, D.A. (2014). Targeting pathways downstream of KRAS in lung adenocarcinoma. Pharmacogenomics *15*, 1507–1518.

Please cite this article in press as: Dost et al., Organoids Model Transcriptional Hallmarks of Oncogenic KRAS Activation in Lung Epithelial Progenitor Cells, Cell Stem Cell (2020), https://doi.org/10.1016/j.stem.2020.07.022

CellPress

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Rat monoclonal anti-CD45 APC [30-F11, BD]	Thermo Fisher Scientific	Cat#BDB559864
Rat monoclonal anti-CD31 APC [MEC 13.3, BD]	Thermo Fisher Scientific	Cat# BDB551262
Rat monoclonal anti-CD326 (EP-CAM) PE/Cy7 [G8.8]	BioLegend	RRID:AB_1236471; Cat#118216
Rat monoclonal anti-Ly-6A/E (Sca1) APC/Cy7 [D7]	Thermo Fisher Scientific	RRID:AB_1727552; Cat#560654
Rabbit monoclonal anti-SP-C [EPR19839]	Abcam	Cat#ab211326
Rat monoclonal anti-Ki67 [SolA15]	Thermo Fisher Scientific	RRID:AB_10854564;Cat#14-5698-82
Rabbit monoclonal anti-TTF1 (Nkx2-1) [8G7G3/1]	Abcam	RRID:AB_1310784; Cat#ab76013
Mouse monoclonal anti-Hmga2 [GT763]	GeneTex	Cat#GTX629478
Goat polyclonal anti-GFP (YFP)	Abcam	RRID:AB_305643; Cat#ab6673
Donkey anti-rat Alexa 594	Invitrogen	RRID:AB_2535795;Cat#A-21209
Donkey anti-goat Alexa Fluor 488	Invitrogen	RRID:AB_2534102; Cat#A-11055
Donkey anti-goat Alexa Fluor 647	Invitrogen	RRID:AB_141844; Cat#A-21447
Donkey anti-rabbit Alexa Fluor 488	Invitrogen	RRID:AB_141708; Cat#A-21206
Donkey anti-rabbit Alexa Fluor 594	Invitrogen	RRID:AB_141637; Cat#A-21207
Donkey anti-mouse Alexa Fluor 647	Invitrogen	RRID:AB_162542; Cat#A-31571
Mouse monoclonal antibody to human CKIT, allophycocyanin (APC) conjugated	Life Technologies	Cat#CD11705; RRID: AB_1463361
Mouse monoclonal IgG2a antibody against human, rhesus, cynomolgus CD184(CXCR4) Clone 12G5	Stem Cell Technologies	Cat #60089PE
Mouse IgG1 isotype, APC conjugated	Life Technologies	Cat#MA5-18093; RRID: AB_2539476
Mouse IgG2a isotype, PE-conjugated	Stem Cell Technologies	Cat#60108PE
Bacterial and Virus Strains		
Ad5CMVempty	Viral Vector Core University of Iowa	Lot:Ad4154; Cat#VVC-U of Iowa-272
Ad5CMVCre	Viral Vector Core University of Iowa	Lot: Ad4117; Cat#VVC-U of Iowa-5
Chemicals, Peptides, and Recombinant Proteins		
GFR Matrigel	Corning	Cat#356231
Bleomycin Sulfate	Sigma-Aldrich	Cat#B2434
Dispase	Corning	Cat#CB-40235
Collagenase/Dispase	Roche	Cat#10269638001
DNase	Sigma-Aldrich	Cat#D4527
PmeI	New England Biolabs	Cat# R0560S
MluI	New England Biolabs	Cat# R0198S
Puromycin	Old stock, unknown	N/A
EcoRV	New England Biolabs	Cat# R0195S
Growth Factor Reduced Matrigel (3D Matrigel)	Corning	Cat# 356230
Human embryonic stem cell (hESC)-qualified Matrigel (2D Matrigel)	Corning	Cat# 354277
CHIR99021 (CHIR)	Tocris	Cat# 4423

(*Continued on next page*)

Please cite this article in press as: Dost et al., Organoids Model Transcriptional Hallmarks of Oncogenic KRAS Activation in Lung Epithelial Progenitor Cells, Cell Stem Cell (2020), https://doi.org/10.1016/j.stem.2020.07.022

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Recombinant Human Keratinocyte Growth Factor (KGF)	R&D Systems	Cat# 251-KG-010
Recombinant Human BMP4 (rhBMP4)	R&D Systems	Cat# 314-BP
Hyclone Fetal Bovine Serum (characterized; FBS)	GE Healthcare Life Sciences	Cat# SH30071.03
Rho-associated kinase inhibitor (Y-27632 dihydrochloride; Y)	Tocris	Cat# 1254
0.05% Trypsin-EDTA	GIBCO	Cat# 25-300-062
Dexamethasone (Dex)	Sigma Aldrich	Cat# D4902
3-Isobutyl-1-methylxanthine (IBMX)	Sigma Aldrich	Cat# I5879
8-Bromoadenosine 3′, 5′-cyclic monophosphate sodium salt (cAMP)	Sigma Aldrich	Cat# B7880
Retinoic Acid (Ra)	Sigma Aldrich	Cat# R2625
Doxycycline Hydrochloride (Dox)	Sigma Aldrich	Cat# D3072
Dimethyl Sulfoxide (DMSO)	Sigma Aldrich	Cat# D2650
Dorsomorphin (DS)	Stemgent	Cat# 04-0024
SB431542 (SB)	Tocris	Cat# 1614
Dispase II	Thermo Fisher Scientific	Cat# 17105-041
Ascorbic Acid	Sigma Aldrich	Cat# A4544
1-Thioglycerol (MTG)	Sigma Aldrich	Cat# M6145
BSA 7.5% Stock	Thermo Fisher Scientific	Cat# 15260037
4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (EDTA)	Sigma Aldrich	Cat# E7889
N-(2-Hydroxyethyl)piperazine-N′-(2-ethanesulfonic acid) Solution (HEPES)	Sigma Aldrich	Cat# H0887
Critical Commercial Assays		
Chromium Single Cell 3′ Library & Gel Bead Kit v2, 16 rxns	10X Genomics	Cat#120237
Chromium Single Cell A Chip Kit, 48 rxns	10X Genomics	Cat#120236
Chromium i7 Multiplex Kit, 96 rxns	10X Genomics	Cat#120262
Amaxa P3 Primary Cell Kit	Lonza	Cat#V4XP-3024
Stem Diff Definitive Endoderm Kit	StemCell Technologies	Cat#05210
RNeasy Mini Kit	QIAGEN	Cat#741404
Qiazol Lysis Reagent	QIAGEN	Cat#79306
TaqMan Fast Universal PCR Master Mix (2X), no AmpErase UNG	Thermo Fisher Scientific	Cat#4364103
High-Capacity cDNA Reverse Transcription Kit	Applied Biosystems	Cat#4368814
Deposited Data		
Jupyter notebooks for GEMM and organoid single-cell RNA-Seq analysis	This paper	https://github.com/alm8517/Kras_invivo_organoid
Single cell RNA-seq raw data (GEMM / organoid)	This paper	GEO: GSE149813 / GEO: GSE149909
Single cell RNA-seq features/matrix/barcode files (GEMM / organoid)	This paper	GEO: GSE149813 / GEO: GSE149909
Bulk RNA-Seq raw data	This paper	GEO: GSE150425
iAT2 single cell RNA-Seq data	This paper	GEO: GSE150263 www.kottonlab.com
Code for iAT2 scRNA-Seq analysis	This paper	https://github.com/cvillamar/Vedaie_CReM
Human patient stage IA singe cell RNA-Seq data	This paper	GEO: GSE149655
Mass spectrometry proteomics data iAT2	This paper	PRIDE: PXD019240

(*Continued on next page*)

Please cite this article in press as: Dost et al., Organoids Model Transcriptional Hallmarks of Oncogenic KRAS Activation in Lung Epithelial Progenitor Cells, Cell Stem Cell (2020), https://doi.org/10.1016/j.stem.2020.07.022

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Experimental Models: Cell Lines		
Human: Normal donor iPSC line targeted with NKX2-1^{GFP} SFTPC^{tdTomato} (BU3 NGST)	Kotton Lab (Jacob et al., 2017)	RRID: CVCL_WN82
Experimental Models: Organisms/Strains		
Gt(ROSA)26Sor$^{tm1(EYFP)Cos}$	The Jackson Laboratory	Cat#006148
Kras$^{LSL-G12D/+}$	Jackson et al., 2001	N/A
Kras$^{LSL\ G12D/+}$; p53$^{fl/fl}$	Jackson et al., 2005	N/A
Hsd:Athymic Nude-Foxn1^{nu}	ENVIGO	Cat#6903F
Oligonucleotides		
hKRAS mutG12D PmeI: gtggcaagtttaaacATGACTGAA TATAAACTTGTGGTAG	Mostoslavsky Lab	N/A
hKRAS mut G12D MluI: ccaatcaggccacgcgtTTA CATAATTACACACTTTGTC	Mostoslavsky Lab	N/A
Z-AV-4: gccggaactctgccctctaacgct	Kotton Lab	N/A
T2A R: GATTCTCCTCCACGTCACCGC	Mostoslavsky Lab	N/A
Taqman Gene Expression Assay Primer/ Probe Set: KRAS	Thermo Fisher Scientific	Hs00364284_g1
Taqman Gene Expression Assay Primer/ Probe Set: NKX2-1	Thermo Fisher Scientific	Hs00968940_m1
Taqman Gene Expression Assay Primer/ Probe Set: SFTPC	Thermo Fisher Scientific	Hs00161628_m1
Recombinant DNA		
pBabe-Kras G12D	Channing Der	Addgene plasmid # 58902 RRID:Addgene_58902
pZ P 4X(cHS4) TetON-3XFLAG-tdT CAGG-m2rtTA v2	Ordovás et al., 2015	N/A
AAVS1 Zinc Finger R		N/A
AAVS1 Zinc Finger L		N/A
Software and Algorithms		
ImageJ	Schneider et al., 2012	https://imagej.nih.gov/ij/
GraphPad Prism for MacOS version 8.2.1	GraphPad Software	https://www.graphpad.com/scientific-software/prism/
FlowJo version 10.5.3	Becton, Dickinson & Company	https://www.flowjo.com/
Scanpy 1.4.4	Wolf et al., 2018	https://github.com/theislab/scanpy
Velocyto 0.17.16	La Manno et al., 2018	https://github.com/velocyto-team/velocyto.py
scVelo 0.1.25	Theis lab	https://github.com/theislab/scvelo
CellRanger 3.0.0	10X Genomics	https://support.10xgenomics.com/single-cell-gene-expression/software/pipelines/latest/installation
Matplotlib 3.0.2	Hunter, 2007	https://matplotlib.org/index.html
Seaborn 0.9.0		https://seaborn.pydata.org/#
Enrichr in gseapy 0.9.13	Kuleshov et al., 2016	https://github.com/zqfang/GSEApy/blob/master/docs/index.rst
Markov Affinity-based Graph Imputation of Cells (MAGIC) 1.5.5	van Dijk et al., 2018	https://github.com/KrishnaswamyLab/MAGIC
Other		
ProLong Gold Antifade Mountant with DAPI	Invitrogen	Cat#P36935
DAPI	Sigma-Aldrich	Cat#D9542

(*Continued on next page*)

Please cite this article in press as: Dost et al., Organoids Model Transcriptional Hallmarks of Oncogenic KRAS Activation in Lung Epithelial Progenitor Cells, Cell Stem Cell (2020), https://doi.org/10.1016/j.stem.2020.07.022

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Transwells	Corning	Cat#3470
SPRI Select Reagent	Beckman Coulter	Cat#NC0406407
Qubit dsDNA HS Assay Kit	Invitrogen	Cat#Q32851
Calcein Blue AM	Life Technologies	Cat#C1429
Hank's Buffered Saline Solution (HBSS; no calcium, no magnesium, no phenol red)	GIBCO	Cat#14175095
Gentle Cell Dissociation Reagent	StemCell Technologies	Cat#07174
GlutaMAX (100x)	Thermo Fisher Scientific	Cat#35050-061
Ham's F12 Medium	Cellgro	Cat#10-080-CV
Iscove's Modified Dulbecco's Medium (IMDM)	Thermo Fisher Scientific	Cat#12440053
N2 Supplement	Invitrogen	Cat#17502-048
B27 Supplement	Invitrogen	Cat#15260-037
Primocin	Invitrogen	Cat#NC9141851
mTeSR1	StemCell Technologies	Cat#05850

RESOURCE AVAILABILITY

Lead Contact

Further information and requests for resources and reagents should be directed to the Lead Contact, Carla F. Kim (carla.kim@childrens.harvard.edu).

Materials Availability

Pluripotent stem cell lines generated in this study are available from the CReM Biobank at Boston University and Boston Medical Center and can be found at http://www.bu.edu/dbin/stemcells/.

Data and code availability

Raw and processed single-cell and bulk RNA-seq data were deposited to the NCBI Gene Expression Omnibus (GEO) and Sequencing Read Archive (SRA) under the following accession codes:

- GEMM single cell RNA-Seq data: GEO: GSE149813
- Murine organoid single cell RNA-Seq data: GEO: GSE149909
- Murine organoid bulk RNA-Seq data: GEO: GSE150425
- Jupyter notebooks for GEMM and organoid single cell RNA Seq data are available on Github https://github.com/alm8517/Kras_invivo_organoid
- iAT2 single cell RNA-Seq data: GEO: GSE150263
- Code for iAT2 scRNA-Seq analysis available at: https://github.com/cvillamar/Vedaie_CReM
- The single cell RNA-seq data from human iPSC derived lung organoids discussed in this publication are available for free interactive analysis through the bioinformatics portal at http://www.bumc.bu.edu/kottonlab/.
- Human patient stage IA singe cell RNA-Seq data: GEO: GSE149655

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the Proteomics Identifications (PRIDE) partner repository (Deutsch et al., 2020; Perez-Riverol et al., 2019) with the dataset identifier PRIDE: PXD019240.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Mouse cohorts

$Kras^{LSL\text{-}G12D/WT}$ (Jackson et al., 2001) and $Kras^{LSL\text{-}G12D/WT}$;$p53^{flox/flox}$ (Jackson et al., 2005) mice were crossed to $Rosa26^{LSL\text{-}eYFP}$ mice to obtain $Kras^{LSL\text{-}G12D/WT}$; $Rosa26^{LSL\text{-}eYFP}$ (KY) and $Kras^{LSL\text{-}G12D/WT}$;$p53^{flox/flox}$; $Rosa26^{LSL\text{-}eYFP}$ (KPY) mice. $Rosa26^{LSL\text{-}eYFP}$ (Y) control mice were littermates of the KY mice. Mice were maintained in virus-free conditions. All mouse experiments were approved by the BCH Animal Care and Use Committee, accredited by AAALAC, and were performed in accordance with relevant institutional and national guidelines and regulations.

Please cite this article in press as: Dost et al., Organoids Model Transcriptional Hallmarks of Oncogenic KRAS Activation in Lung Epithelial Progenitor Cells, Cell Stem Cell (2020), https://doi.org/10.1016/j.stem.2020.07.022

Stage IA LUAD patient information

Samples of two patients with the diagnosis stage IA LUAD were analyzed in these studies. One patient was female, 74 years old, with a KRAS-G12F mutation identified as driver mutation. The other patient was female, 77 years old, with a KRAS-G12V mutation identified as driver mutation. All patients provided written informed consent. The studies were approved by the UCLA institutional review board.

METHOD DETAILS

Mouse studies

In vivo adenovirus infection

8-week-old mice were infected with $2.5x10^7$ PFU adenovirus by intratracheal instillation as described previously (DuPage et al., 2009). A 1:1 ratio of male and female mice was used.

Lung preparation and FACS

Mice were anesthetized with avertin, perfused with 10 mL PBS, followed by intratracheal instillation of 2 mL dispase (Corning). Lungs were iced, minced and incubated in 0.0025% DNase (Sigma Aldrich) and 100 mg/ml collagenase/dispase (Roche) in PBS for 45 min at 37°C, filtered through 100 μm and 40 μm cell strainers (Fisher Scientific), and centrifuged at 1000 rpm, 5 min at 4°C. Cells were resuspended in red blood cell lysis buffer (0.15 M NH4Cl, 10mM KHCO3, 0.1 mM EDTA) for 1.5 min, washed with advanced DMEM (GIBCO), and resuspended in PBS/10% FBS (PF10) at 1 million/100 μl. Depending on the experiment, cells were incubated for 10 min on ice with DAPI as a viability dye and the following antibodies: anti-CD31 APC, anti-CD45 APC, anti-Ly-6A/E (SCA1) APC/Cy7 (all Thermo Fisher Scientific), anti-CD326 (EP-CAM) PE/Cy7 (Biolegend) (all 1:100). Single stain controls and fluorophore minus one (FMO) controls were included for each experiment. FACS was performed on a FACSAria II and analysis was done with FlowJo.

In vitro virus infection and organoid culture

Murine lung CD31- CD45- EPCAM+ SCA1- cells isolated by FACS as described in section "Lung preparation and FACS" were split into 2 or 3 equal aliquots, or not split, depending on the experiment, pelleted by pulse spin and resuspended in 100 μL MTEC/Plus media (Zhang et al., 2017a) containing 6×10^7 PFU/ml of Ad5CMV-Cre, Ad5CMV-Empty, or no virus in 100 μL per 100,000 cells. The cells were incubated for 1 h at 37°C, 5% CO2 in 1.5 mL tubes. Cells were then pelleted by pulse spin and resuspended in 1x phosphate-buffered saline (PBS). This step was repeated twice for a total of three washing steps. Cells were resuspended in Dulbecco's Modified Eagle's Medium/F12 (Invitrogen) supplemented with 10% FBS, penicillin/streptomycin, 1 mM HEPES, and insulin/transferrin/selenium (Corning) (3D media) at a concentration of 5,000 live cells (trypan blue negative) per 50 μl. As supporting cells, a mix of neonatal stromal cells was isolated as described elsewhere (Lee et al., 2014). The stromal cells were pelleted and resuspended in growth factor reduced (GFR) Matrigel at a concentration of 50,000 cells per 50 μl. Equal volumes of cells in 3D media and supporting cells in GFR Matrigel were mixed and 100 μL were pipetted into a Transwell (Corning). Plates were incubated for 20 min at 37°C, 5% CO2 until Matrigel solidified. Finally, 500 μL of 3D media was added to the bottom of the well. 3D media was changed every other day.

Staining and IF of organoid cultures

To image whole wells, multiple overlapping images of live organoid cultures were taken and stitched together using EvosTM FL Auto2 software. To prepare organoid slides, organoid cultures were fixed with 10% neutral-buffered formalin overnight at room temperature. After rinsing with 70% ethanol, the organoid cultures containing Matrigel plug was immobilized with Histogel (Thermo Scientific) for paraffin embedding. Paraffin blocks were cut into 5 μm sections and adhered to glass slides. For deparaffinization, slides were incubated in xylene and then rehydrated in 100%, 95%, 70% ethanol successively. Slides were then stained with hematoxylin and eosin, or further processed for IF staining. For IF staining, antigen was retrieved by incubating the slides in citric acid buffer (pH 6) at 95°C for 20 min. After washing slides with PBS containing 0.2% Triton-X (PBS-T) and blocking with 10% normal donkey serum for 1 h at room temperature, slides were incubated with antibodies for Ki67 (EBioscience 1:100), YFP (Abcam, 1:400), SPC (Abcam, 1:1,000), Nkx2-1 (Abcam, 1:250) Hmga2 (GeneTex, 1:200), in a humidified chamber at 4°C overnight. Secondary antibodies were added following three washing steps with PBS-T and included donkey anti-rat Alexa 594, donkey anti-goat Alexa 488/647, donkey anti-rabbit Alexa 488/594, donkey anti-mouse Alexa 647 (all Invitrogen, 1:200). Slides were mounted using Prolong Gold with DAPI (Invitrogen).

Preparing single cell suspensions of organoid cultures

At day 7 of organoid culture, 100 μL dispase (Fisher Scientific) was added to the transwells on top of the Matrigel and incubated for 1 h at 37°C, 5% CO2. After digestion of Matrigel, the wells were washed with PBS and the organoids were pipetted into 15 mL conical tubes. The tubes were filled with PBS to dilute the remaining Matrigel and dispase. After pelleting the organoids at 300 g for 5 min, the organoids were resuspended in 37°C warm Trypsin EDTA (0.25%, Invitrogen) and incubated for 7-10 min at room temperature to obtain a single cell suspension. Trypsin was quenched by adding PBS + 10% FBS (PF10).

Transplantation assays of organoids

To ensure engraftment, 8-10 weeks old Athymic Nude mice were injured by injecting 1.5U/kg bleomycin intratracheally one day before transplantation. For transplantation assays, single cell suspensions were obtained from day 14-21 of passage 0 organoid cultures as described in section "Preparing single cell suspensions of organoid cultures." To ensure transplantation of equal numbers of Cre-activated cells across samples, YFP+ cells were counted under the fluorescence microscope and 33,000-130,000 YFP+ cells resuspended in 45 μL PBS were administered into the lungs of the injured Athymic Nude intratracheally. For

Please cite this article in press as: Dost et al., Organoids Model Transcriptional Hallmarks of Oncogenic KRAS Activation in Lung Epithelial Progenitor Cells, Cell Stem Cell (2020), https://doi.org/10.1016/j.stem.2020.07.022

histology evaluation, mice were sacrificed after 4 weeks and lungs were fixed by injecting 10% neutral-buffered formalin into the lungs through the trachea.

FACS to prepare organoid cultures for RNA-Seq

Single cell suspensions were obtained from day 7 organoid cultures as described in section "Preparing single cell suspensions of organoid cultures." For FACS staining, cells were incubated with EPCAM-PeCy7 (BioLegend) and DAPI (Sigma-Aldrich) for 10 min on ice. A DAPI only control served as the fluorophore minus one (FMO) control for EPCAM. FACS was performed on a FACSAria II and analysis was done with FlowJo.

RNA extraction and bulk RNA-Seq of organoids

EPCAM+ cells were obtained from organoid cultures as described in section "FACS to prepare organoid cultures for RNA-Seq." RNA was extracted using the Absolutely RNA Microprep Kit (Agilent). After RNA extraction, all downstream quality control steps, library preparation, sequencing, and differential gene expression analysis was performed by the Molecular Biology Core Facilities at Dana-Farber Cancer Institute. Complementary DNA (cDNA) was synthesized with Clontech SmartSeq v4 reagents from 2ng of RNA. Full length cDNA was fragmented to a mean size of 150bp with a Covaris M220 ultrasonicator and Illumina libraries were prepared from 2ng of sheared cDNA using Takara Thruplex DNaseq reagents according to manufacturer's protocol. The finished double strand DNA libraries were quantified by Qubit fluorometer, Agilent TapeStation 2200, and RT-qPCR using the Kapa Biosystems library quantification kit. Uniquely indexed libraries were pooled in equimolar ratios and sequenced on an Illumina NextSeq500 run with single-end 75bp reads at the Dana-Farber Cancer Institute Molecular Biology Core Facilities.

Bioinformatic analysis of bulk RNA-Seq

Sequenced reads were aligned to the UCSC hg19 reference genome assembly and gene counts were quantified using STAR (v2.5.1b) (Dobin et al., 2013) Differential gene expression testing was performed by DESeq2 (v1.10.1) (Love et al., 2014) and normalized read counts (FPKM) were calculated using cufflinks (v2.2.1) (Trapnell et al., 2010). RNaseq analysis was performed using the VIPER snakemake pipeline (Cornwell et al., 2018).

ScRNA-Sequencing of GEMM and organoids

ScRNA-Seq was performed using the 10X Genomics platform (10X Genomics, Pleasanton, CA). FACS sorted cells from either mice or organoid cultures were encapsulated with a 10X Genomics Chromium Controller Instrument using the Chromium Single Cell A Chip Kit. Encapsulation, reverse transcription, cDNA amplification, and library preparation reagents are from the Chromium Single Cell 3′ Library & Gel Bead Kit v2. Briefly, single cells were resuspended in PF10 at a concentration of 1000 cells μ^{-1}. The protocol was performed as per 10X Genomics protocols without modification (chromium single cell 3 reagent kits user guide v2 chemistry). Total cDNA and cDNA quality following amplification and clean-up was determined using a QubitTM dsDNA HS assay kit and the Agilent TapeStation High Sensitivity D5000 ScreenTape System. Library quality pre-sequencing was determined using Agilent TapeStation and QPCR prior to sequencing. TapeStation analysis and library QPCR was performed by the Biopolymers Facility at Harvard Medical School. Libraries were sequenced using an Illumina NextSeq500 using paired-end sequencing with single indexing (Read 1 = 26 cycles, Index (i7) = 8 cycles, and Read 2 = 98 cycles). Reads were aligned to the mm10 reference genome and count matrices were generated using CellRanger3.0.0 (10X Genomics).

Bioinformatics for GEMM and organoid scRNA-Seq

Count matrices generated by CellRanger3.0.0 were read into the Python single cell analysis environment Scanpy (v 1.4.4) (Wolf et al., 2018). In brief, cells with > 10% mitochondrial content, which correlated with low read count, were removed. The data was normalized, logarithmized, and the significant number of principle components determined using in-built Scanpy functions. Data was denoised using Markov Affinity-based Graph Imputation (v 1.5.5) using the following settings (Gene to return = all, k = 3, t = 3, n_pca = 30) (van Dijk et al., 2018). Gene Ontology enrichment analysis was performed with Enrichr (Kuleshov et al., 2016) using the GSEAPY (v 0.9.13) python wrapper. A reference list of murine transcription factors and transcription co-factors is from the Animal Transcription Factor Database (Hu et al., 2019). Lists of genes activated by specific transcription factors were from the TRRUST database (Han et al., 2018). KRAS activation signature was previously described (Barbie et al., 2009; Bild et al., 2006). Murine AT2 marker genes are from PanglaoDB (Franzén et al., 2019). All gene lists can be found in table S1. Data was visualized using in-built Scanpy plotting functions, Seaborn (v0.9.0) (https://seaborn.pydata.org/), and Matplotlib (v 3.0.2) (Hunter, 2007).

RNA Velocity

Velocyto (0.17.16) was run on the KY GEMM and KY organoid CellRanger output files using the run10X shortcut and the mm10 genome annotation file provided with the CellRanger pipeline. Loom files generated by Velocyto for each sample were concatenated into an anndata object. To visualize velocity on the original UMAP embedding a new anndata was created by merging the velocity and original anndata objects using the utils.merge() function in scVelo (0.1.25). Velocity was calculated using the merged anndata object and in-built velocity functions.

Human iPSC studies

Generation of BU3 NGST-TetOn:KRAS^{G12D} line

To generate a dox-inducible KRAS^{G12D} cassette targeted to the AAVS1 locus by gene editing, the previously published BU3 NGST human iPSC line was used (Jacob et al., 2017). PZ P 4X(cHS4) TetON-3XFLAG-tdT CAGG-m2rtTA v2, an optimized targeting vector for the AAVS1 locus was obtained as the kind gift of Laura Ordovas (Ordovás et al., 2015). This vector has the addition of two cHS4 insulators on either side of the transgene to reduce the potential for silencing. In addition, the construct contains an m2rtTA under the control of a CAG promoter and a T2A:puromycin resistance gene that should only be active when inserted near a coding sequence,

Please cite this article in press as: Dost et al., Organoids Model Transcriptional Hallmarks of Oncogenic KRAS Activation in Lung Epithelial Progenitor Cells, Cell Stem Cell (2020), https://doi.org/10.1016/j.stem.2020.07.022

improving the selection specificity. Human $KRAS^{G12D}$ was PCR amplified from pBabe-Kras G12D, a gift from Channing Der (Addgene plasmid # 58902 ; http://addgene.org/58902 ; RRID:Addgene_58902) using primers hKRAS mutG12D PmeI and hKRAS mutG12D MluI. The resulting PCR product was cloned into PZ P 4X(cHS4) TetON-3XFLAG-tdT CAGG-m2rtTA v2 using EcoRV and MluI restriction sites to generate a new vector named AAVS1-TetOn:$KRAS^{G12D}$. For targeting the BU3NGST iPSC line, 4 $x10^6$ live cells were resuspended in Amaxa P3 primary cell nucleofection solution containing 1υg/10^6 cells of the AAVS1-TetOn:$KRAS^{G12D}$ plasmid and the left and right zinc finger plasmids targeting the AAVS1 locus. The cells were then nucleofected using the human embryonic stem cell (hESC), H9 standard program on the Lonza 4D-nucleofector. The cells were then resuspended in mTeSR with 10υM Y27632 and plated on a 10cm hESC Matrigel coated plate. Cells were selected using puromycin at 500-700 ng/ml starting a minimum of 96hrs after nucleofection. Selection was maintained for 7-10 days as the resistant colonies emerged and grew. Successful colonies were manually picked into 24-well hESC Matrigel coated plates in mTeSR with 10υM Y27632. Genomic DNA from each clone was screened for insertion using primers Z-AV-4 (binds in the AAVS1 locus outside the donor arm)/ T2A R and correct insertion validated by sequencing. Positive clones were expanded, re-selected with puromycin and frozen, and a single clone was carried forward after G-banding analysis to confirm normal 46XY karyotype.

Lung differentiation and flow cytometry

Lung differentiation of the iPSC line (BU3 NGST-TetOn:$KRAS^{G12D}$) into alveolar type 2 cells was performed according to the detailed protocol previously published by Jacob et al. (2017, 2019). Briefly, iPSC-derived NKX2-1^{GFP+} lung epithelial progenitors generated after 15 days of directed differentiation were purified by GFP+ flow cytometry sorting and replated for further distal lung/alveolar differentiation in 3D Matrigel cultures and the resulting monolayered epithelial spheres were maintained as self-renewing distal alveolar epithelial cells by serial passaging approximately every 2 weeks in serum-free, feeder-free 3D culture ("CK+DCI" media as detailed in Jacob et al., 2019). Quality and phenotype of the cultures was monitored at each passage by flow cytometry quantitation of NKX2-1^{GFP} and $SFTPC^{tdTomato}$ expression as shown in the text. Detailed protocols for cell preparation for flow cytometry and analysis of these reporters has been previously published (Jacob et al., 2019).Briefly, for flow cytometry analysis, cells were resuspended in FACS buffer (PBS with 2% FBS and 10 nM calcein blue AM (ThermoFisher)) and analyzed on an S1000EXi flow cytometer (Stratedigm San Jose, CA). For cell sorting, cells were resuspended in FACS buffer plus 10 uM Y-27632 to support viability in replated cells. Live cells were sorted on a high speed cell sorter (MoFlo Legacy, Beckman Coulter) at the Boston University Medical Center Flow Cytometry Core Facility based on NKX2-1^{GFP} expression. All differentiation and passaging protocols for iAT2s are also available for free download from the protocols webpage of http://www.bumc.bu.edu/kottonlab/.

Proteomic and phosphoproteomic analysis

After 3 passages as NKX2-1^{GFP+} sorted alveolospheres, iAT2s were treated with dox (1mcg/ml) or DMSO for 15 days. Four replicates of each condition were dissociated and sorted on live, NKX2-1^{GFP+} cells using the previously described protocol (Jacob et al., 2017, 2019), and collected as cell pellets. In order to interrogate the proteome and phosphoproteome of ± dox-exposed KRASG12D targeted iAT2s, the cell pellets collected were resuspended in lysis buffer composed of 6M GuHCl (guanidinium chloride), 100mM Tris pH 8, 40mM chloroacetamide, 10mM TCEP (tris(2-carboxyethyl)phosphine), and phosphatase inhibitors (PhosStop, Roche), and sonication via a Branson probe. Total protein content was quantified and equal amounts of denatured protein was allocated from each sample, diluted with 7 volumes of 100mM tris, and trypsin digested into peptides. The peptide mixtures from control iAT2s versus dox-exposed iAT2s were individually isotopically-labeled with a distinct isobaric TMT-10plex reagent. After pooling, the mixture was injected onto a reverse-phase Waters Xbridge C18 HPLC column to fractionate the multiplexed peptides, which markedly increased depth of coverage. Peptides were eluted in 12 fractions over 48 min. For the total proteome analysis, 5% of each fraction was analyzed directly by LC/MS. The remaining 95% was set aside for phospho-peptide enrichment using Fe-NTA magnetic beads (Cube Biotech) (Leutert et al., 2019), totaling 24 injections analyzed by precision mass spectrometry (LC/MS). We used the MaxQuant (1.6.7.0; https://maxquant.org/) software package for protein identification by searching with the UniProt Human database (accessed April 2019) and relative quantification of the TMT reporter labels (Cox et al., 2011). Standard search parameters included allowing for two missed trypsin cleavage sites, variable modifications of methionine oxidation, and N-terminal acetylation, and fixed modification of carbamidomethylation of cysteine residues. Protein phosphorylation at S, T, and Y residue data was included as a variable modification for the phosphoproteomic data. Ion tolerances of 20 and 4.5 ppm were set for first and second searches, respectively. After stringent filtering (peptide and protein level FDR of 1% as determined by reverse decoy search), cognate proteins were identified using strict matching parameters guided by principles of parsimony to account for all observed peptide hits. Matches were pruned by filtering out candidates supported by only a single unique peptide. For the identification of phosphopeptides, only modified peptides with unambiguous single site-localization probabilities of at least 0.7 was retained for downstream (differential and pathway enrichment) analyses. For quantitative comparisons of the samples, summed protein intensities were log transformed, LoessF normalized, and statistically significant changes determined using empirical Bayes analysis implemented in the limma package (Phipson et al., 2016) in R: A language and environment for Statistical Computing (R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org). Gene Set Enrichment Analysis (GSEA) was performed using the fgsea package in R (Sergushichev, 2016).

ScRNA-Seq

Parallel wells of iAT2s (derived from BU3 NGST-TetOn:$KRAS^{G12D}$ iPSCs, beginning at sphere passage P3) were treated with either control vehicle (DMSO) or doxycycline (dox; 1ug/mL) to induce expression of $KRAS^{G12D}$. After 4 more passages and 69 days of exposure to Dox or DMSO (total differentiation time = 127 days), cells were dissociated from 3D Matrigel (as described in Jacob et al., 2019), and sorted for Calcein Blue+ live cells. scRNA-seq of all calcein blue-stained live cells was performed using the 10X Chromium

Please cite this article in press as: Dost et al., Organoids Model Transcriptional Hallmarks of Oncogenic KRAS Activation in Lung Epithelial Progenitor Cells, Cell Stem Cell (2020), https://doi.org/10.1016/j.stem.2020.07.022

system with v3 chemistry as previously published (McCauley et al., 2017). Library preparation and sequencing was done at the Boston University Microarray and Sequencing Resource (BUMSR) Core using an Illumina NextSeq 500 instrument.

Bioinformatics analysis of scRNA-Seq

Reads were demultiplexed and aligned to the human genome assembly (GRCh38, Ensembl) with the CellRanger pipeline v.3.0.2 (10X Genomics). Further analyses were done using Seurat v. 3.1.4 (Stuart et al., 2019). Cells with more than 25% of mitochondrial content or less than 800 detected genes were excluded from downstream analyses (leaving 775 controls and 1322 dox-treated cells). We then filtered out the non-lung endoderm population from the control sample (149 cells), leaving a total of 626 cells in the control population and 1322 cells in the dox+ population. We normalized and scaled the UMI counts using the regularized negative binomial regression (SCTransform; Hafemeister and Satija, 2019). Following the standard procedure in Seurat's pipeline, we performed linear dimensionality reduction (principal components analysis; PCA), and used the top 20 principal components to compute both the UMAP (Diaz-Papkovich et al., 2019) and the clusters (Louvain method; Blondel et al., 2008), which were computed at a range of resolutions from 1.5 to 0.05 (more to fewer clusters). For downstream analyses, we refer to the 3 clusters identified at resolution 0.1 (Figure S5C). Cell cycle scores and classifications were done with Seurat using the method from Tirosh et al. (2016). The same method was used to calculate the enrichment in the iAT2 differentiation and maturation signatures from Hurley et al. (2020). The cut-offs for independent filtering (Bourgon et al., 2010) prior to DE testing required genes: a) being detected in at least 10% of the cells of either population and b) having a natural log fold change of at least 0.25 between populations. The tests were performed using Seurat's wrapper for the MAST framework (Finak et al., 2015), identifying 393 differentially expressed genes between control and dox-treated cells (Table S5). For a comparison of the performance of methods for single-cell DE, see Soneson and Robinson (2018). The top 20 genes upregulated and ranked by their fold-change in each clustered population with FDR < 0.05 are represented in a heatmap (Figure S5E).

Patient stage IA lung cancer studies

Sample collection and preparation for scRNA-Seq

Lung cancer resection specimens were obtained from patients with the radiographic diagnosis of stage IA lung cancer. All patients provided written informed consent. KRAS mutation status of tumors was determined by targeted sequencing. Resected tissues were placed on ice in RPMI medium immediately after resection and delivered to the lab for tissue dissociation. Dissociation was performed in RPMI medium supplemented with 10% FBS. Briefly, tissues were sliced to approximately 1 mm^3 pieces and dissociated in 1 mg/ml collagenase (Sigma Aldrich, #C9407) and 1000 U/ml DNase I (Sigma Aldrich, #D4263-1VL) at 37°C for approximately 1 hour until homogeneity followed by passing through a 40 μm strainer to remove cell aggregates and red blood cell lysis with 1 mL of ACK buffer (Sigma Aldrich, #11814389001). Cells were resuspended in 5 mL DPBS + 0.04% BSA, counted and immediately used to prepare the sequencing libraries.

ScRNA-Seq and read alignment

The 10X Genomics platform (10X Genomics, Pleasanton, CA) was utilized for assessing human single cell transcriptome. Single cell encapsulation, library construction and sequencing were performed at Technology Center for Genomics and Bioinformatics at UCLA according to the manufacturer's protocols. The Chromium Single Cell 3′ Library & Gel Bead Kit v2 and v3 were used for library preparation. Libraries were sequenced utilizing Illumina NovaSeq 6000 instrument. CellRanger 3.0.0 software (10X Genomics) was utilized to align reads to human GRCh38 reference and generate count matrices.

Bioinformatics analysis

Human single cell transcriptome data was analyzed by following Seurat pipeline (Stuart et al., 2019). Poor quality cells with > 15% mitochondrial content and less than 500 detected features were filtered out. The data was normalized and batch-adjusted based on Seurat Standard workflow. Cell clustering analyses were performed on the adjusted data to first separate immune cells from non-immune cells in-silico, and then to identify lung specific cell subtypes among non-immune cells. Pseudobulk approach was utilized to identify differentially expressed genes (DEGs) in AT2 cells from tumor and the associated normal lung tissue. Patient-associated variation was included in modeling DEG using edgeR package.

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistics

Statistical testing was performed using GraphPad Prism or Scipy 1.3.0 statistical functions (scipy.stats). The tests used to determine statistical significance are quoted in the appropriate figure legends. P values are indicated in the figures, and P values < 0.05 were considered significant.

Cell Stem Cell, Volume *27*

Supplemental Information

Organoids Model Transcriptional Hallmarks of Oncogenic KRAS Activation in Lung Epithelial Progenitor Cells

Antonella F.M. Dost, Aaron L. Moye, Marall Vedaie, Linh M. Tran, Eileen Fung, Dar Heinze, Carlos Villacorta-Martin, Jessie Huang, Ryan Hekman, Julian H. Kwan, Benjamin C. Blum, Sharon M. Louie, Samuel P. Rowbotham, Julio Sainz de Aja, Mary E. Piper, Preetida J. Bhetariya, Roderick T. Bronson, Andrew Emili, Gustavo Mostoslavsky, Gregory A. Fishbein, William D. Wallace, Kostyantyn Krysan, Steven M. Dubinett, Jane Yanagawa, Darrell N. Kotton, and Carla F. Kim

Figure S1

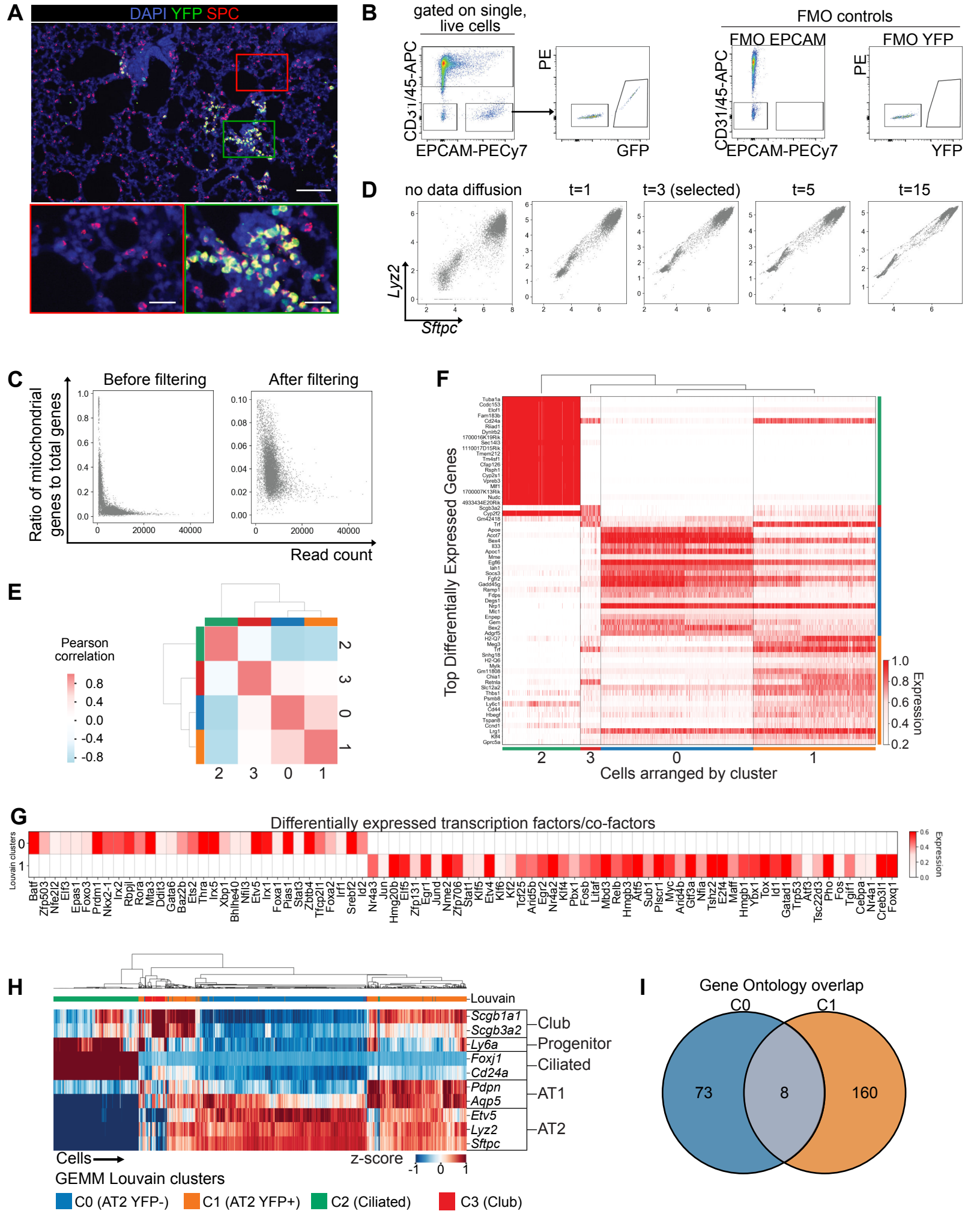

A
DAPI YFP SPC
B
gated on single, live cells
CD31/45-APC
EPCAM-PECy7
PE
GFP
FMO controls
FMO EPCAM
FMO YFP
YFP
D
no data diffusion
t=1
t=3 (selected)
t=5
t=15
Lyz2
Sftpc
C
Before filtering
After filtering
Ratio of mitochondrial genes to total genes
Read count
E
Pearson correlation
F
Top Differentially Expressed Genes
Cells arranged by cluster
Expression
G
Differentially expressed transcription factors/co-factors
Louvain clusters
H
Louvain
Scgb1a1
Scgb3a2
Ly6a
Foxj1
Cd24a
Pdpn
Aqp5
Etv5
Lyz2
Sftpc
Club
Progenitor
Ciliated
AT1
AT2
Cells
z-score
GEMM Louvain clusters
C0 (AT2 YFP-)
C1 (AT2 YFP+)
C2 (Ciliated)
C3 (Club)
I
Gene Ontology overlap
C0
C1
73
8
160

Figure S1, related to Figure 1: Single-cell RNA-Seq of distal lung epithelium reveals distinct transcriptional clusters of KRAS^{G12D} activated cells during early tumorigenesis

(A) Representative IF image of lungs 7 weeks after KRAS^{G12D} induction. Shown is a normal alveolar region (red box) and a hyperplastic region with clusters of YFP+ cells (green box). Scale bar low magnification = 100 μm. Scale bar insets = 25um.

(B) Representative FACS plot showing cell sorting strategy starting with whole lung tissue from a KY mouse. FMO = fluorophore minus one.

(C) Correlation between mitochondrial gene expression and read count before and after filtering cells with > 10% mitochondrial gene expression.

(D) Correlation between *Sftpc* and *Lyz2* expression in individual cells after different numbers of data diffusion (t) (Van Dijk et al., 2018).

(E) Similarity between each Louvain cluster calculated using the Pearson's correlation co-efficient.

(F) Top differentially expressed genes in each Louvain cluster, visualized with a heatmap.

(G) Heatmap depicting expression of transcription factors/co-factors identified amongst differentially expressed genes for Louvain clusters 0 and 1.

(H) Correlation heatmap of individual cells of the GEMM scRNA-Seq data (x-axis) and z-normalized gene expression of indicated lineage genes (y-axis). Cells are ordered based on correlation distance calculation. Louvain clusters are annotated.

(I) Identification of common, C0 specific, and C1 specific Gene Ontology Biological Process terms. Data was visualized as a Venn diagram.

Figure S2

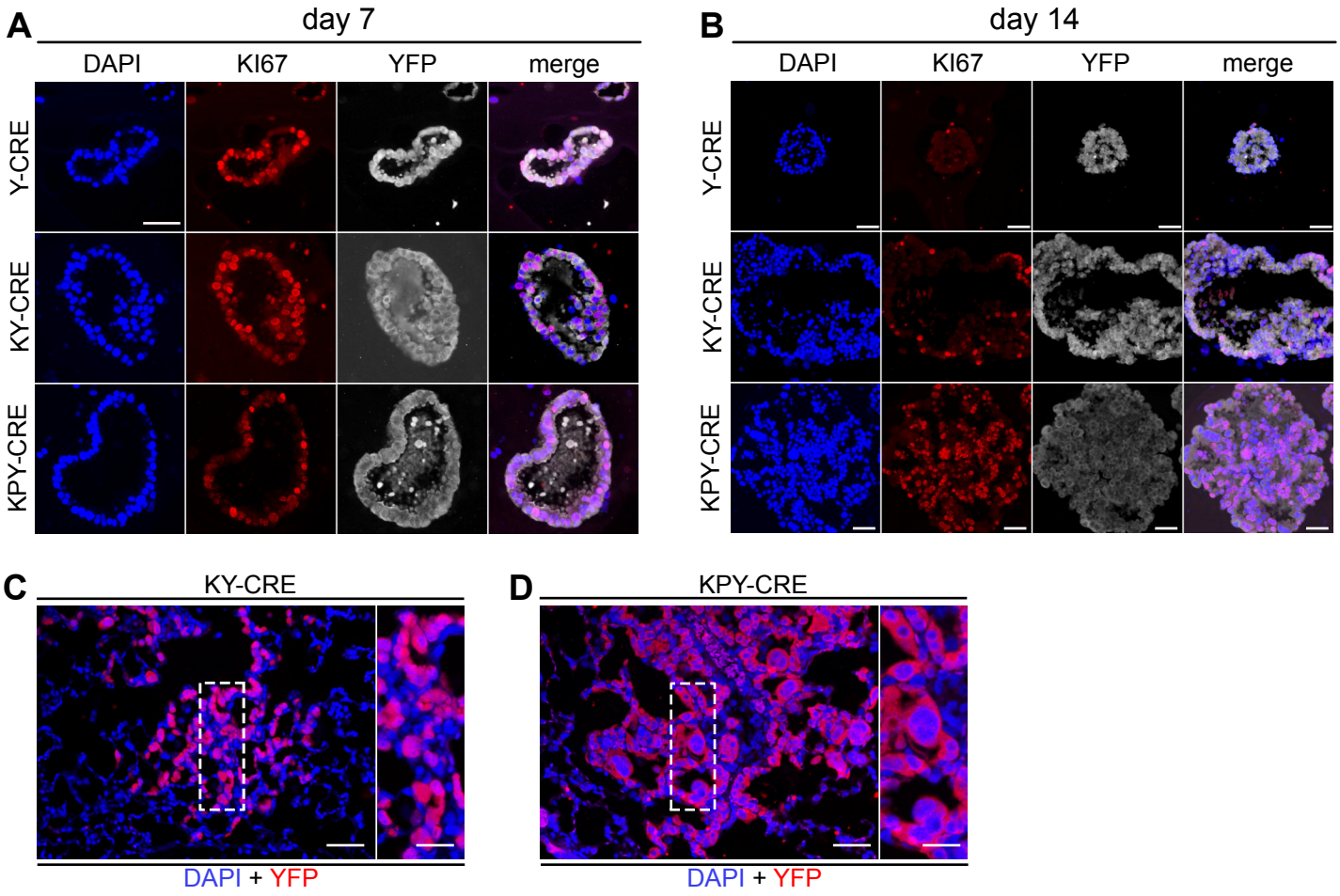

A
day 7
DAPI
KI67
YFP
merge
Y-CRE
KY-CRE
KPY-CRE
B
day 14
DAPI
KI67
YFP
merge
Y-CRE
KY-CRE
KPY-CRE
C
KY-CRE
DAPI + YFP
D
KPY-CRE
DAPI + YFP

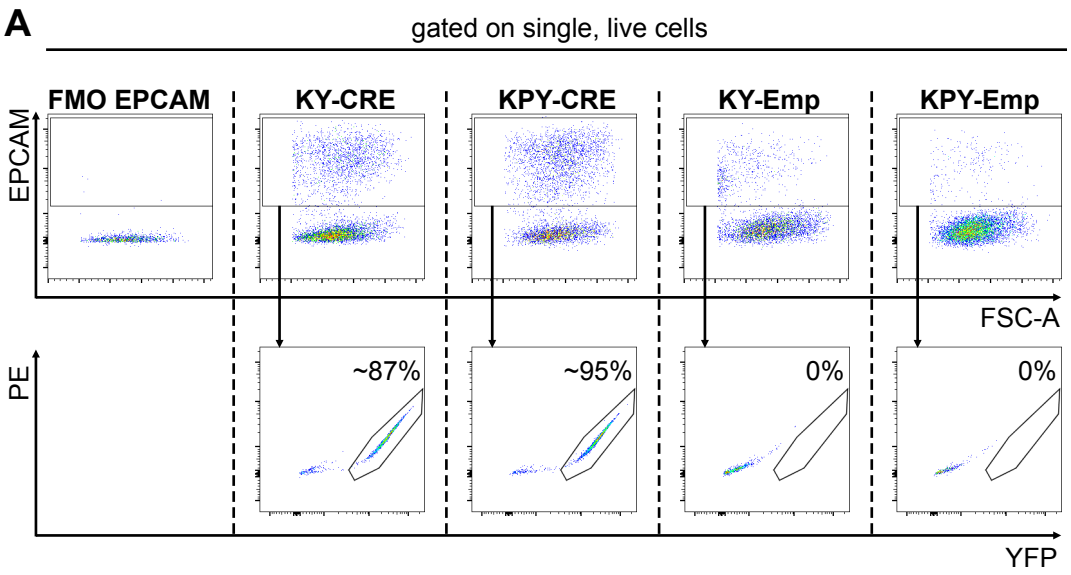
Figure S2, related to Figure 2: Inducible organoids rapidly recapitulate *in vivo* tumor progression and form tumors upon transplantation

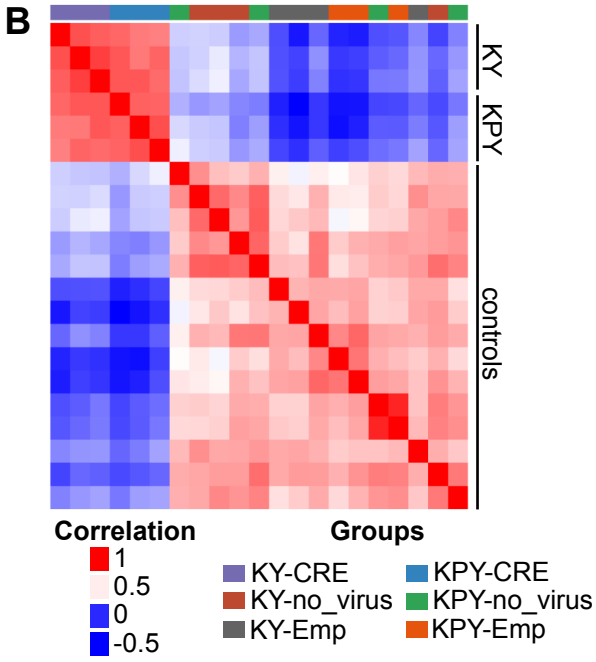
(A) Representative pictures of immunofluorescence staining on day 7 and

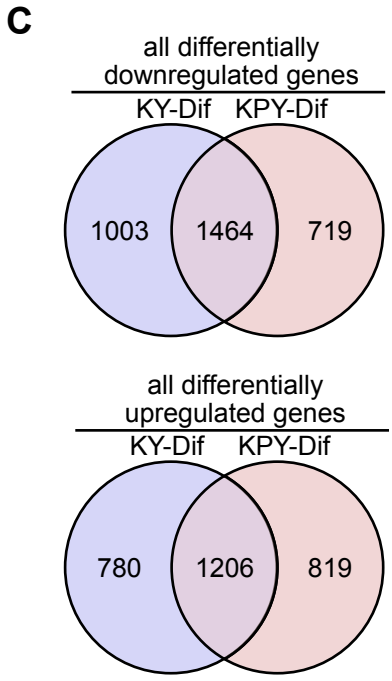
(B) day 14 of organoid culture. Scale bar = 50 μm.

(C, D) immunofluorescence staining of nuclei (DAPI, blue), and YFP (red) of mouse lungs that were transplanted with organoid-derived cells. Scale bar lower magnification = 50 μm; scale bar higher magnification of inlets = 25 um.

Figure S3

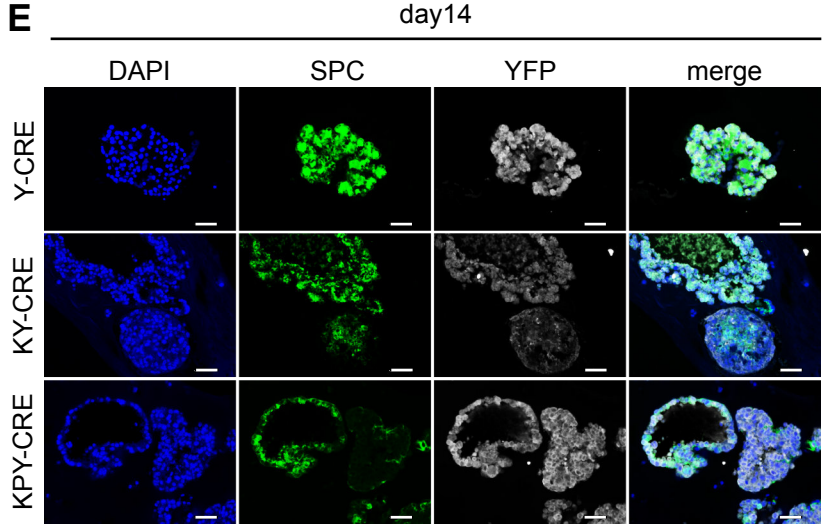






D

upregulated	downregulated
Ly6a	Sec14l3
Pthlh	Mlc1
Emp1	Gas6
Phgdh	Aox3
B3gnt3	Mettl7a1
Procr	Ces1d
Msln	Ces1g
Ecm1	Pmp22
F2r	Galnt18
Hnf4a	Kcnk2
Hmga2	Ak1
Slc2a1	H2-Eb1
Slc16a3	2610028H24Rik
Krt7	Cyp2b10
St8sia6	Akr1c14
Tigit	Scnn1b
Tubb6	Adcy7
Tpi1	Cd74
Pglyrp1	Mme
Itga6	H2-Ab1
Ptges	Nkd1
Fam107b	H2-Aa
Aqp3	Tppp3
F3	Nrep
Errfi1	Cyp4b1
Ldha	Ptn
Sox9	Gsta3
B4galt6	Col6a1
Anxa2	Hnmt
Tnfrsf23	Lyz2
Mcpt2	Cytip
Adss	Sepp1
Kcnq1	Lgi3
Ero1l	Cyp4v3
Lad1	Gsap
Fabp5	H2-DMb1
Pfkl	Creg1
Prss22	Ppp2r2b
Spp1	Itm2a
	Ldhb
	Fgf1
	Tmem116
	Gstt1
	Dynlrb2
	Ddo
	S100g
	Tspan11
	Serpinb9
	Scnn1g



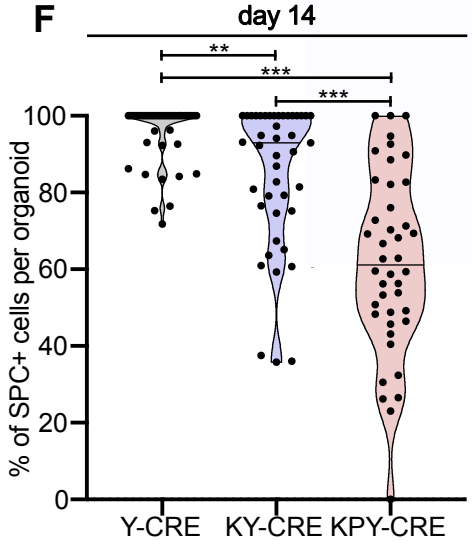


Figure S3, related to Figure 3: KRAS^{G12D} activated cells in organoids lose AT2 differentiation markers and express developmental lung markers

(A) Representative FACS plot showing cell sorting strategy of organoids. FMO = fluorophore minus one.

(B) Heat map showing sample-sample-correlations of RNA-Seq samples.

(C) Venn diagram showing the overlap of all differentially expressed genes in KY-CRE and KPY-CRE compared to their respective -Emp controls.

(D) List of differentially expressed genes that are amongst the top 100 hits of both KY-CRE and KPY-CRE compared to their respective -Emp controls (genes in overlap of Venn diagram from Figure 3B)

(E) Representative pictures of immunofluorescence staining on day 14 of organoid culture. Scale bars = 50 µm.

(F) Quantification of SPC+ cells per organoid on day 7 of organoid culture. Each point represents one organoid.

P-values were determined using the Mann-Whitney rank test. **=p<0.005, ***=p<0.0005.

Figure S4

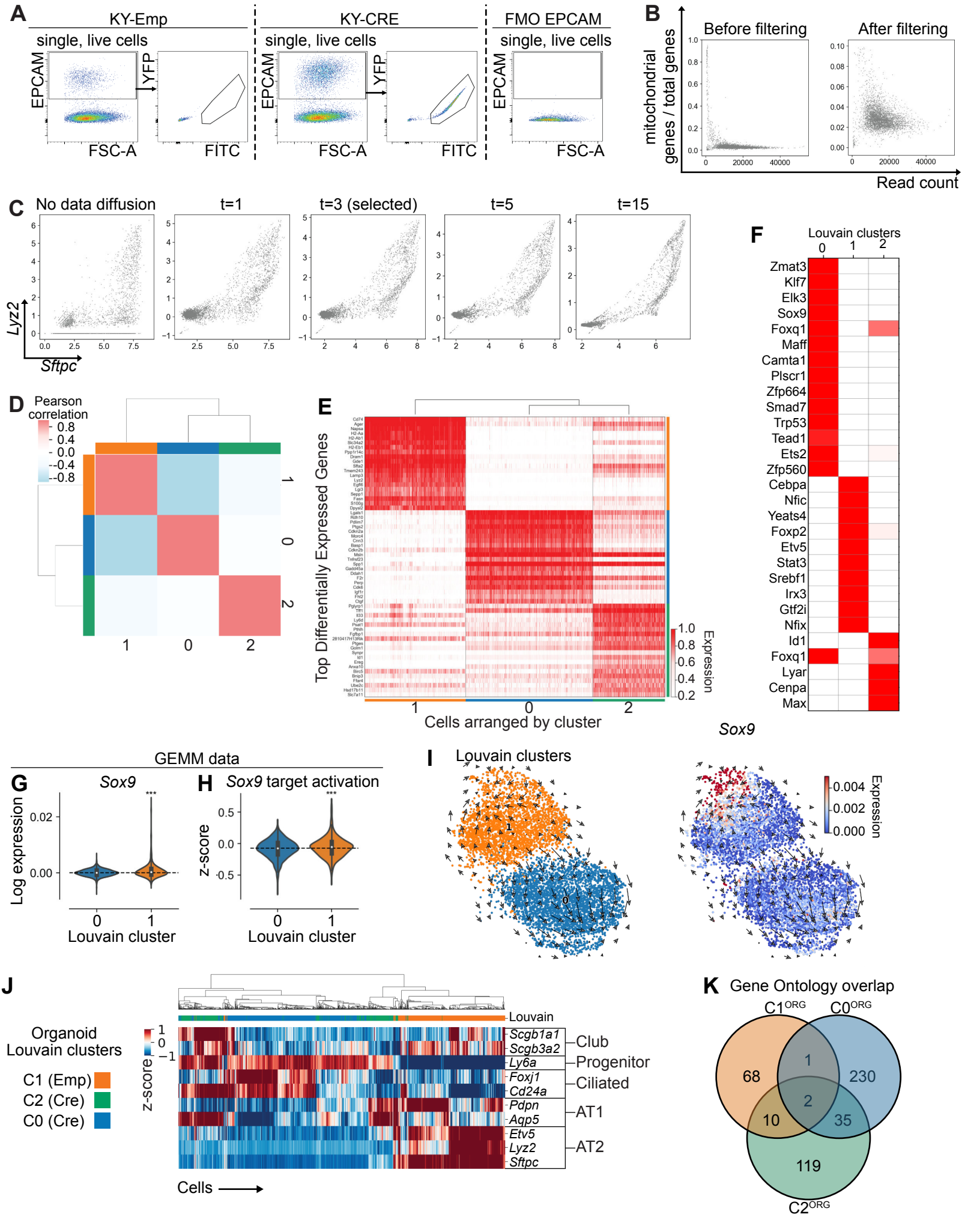

A
KY-Emp
KY-CRE
FMO EPCAM
single, live cells
EPCAM
YFP
FSC-A
FITC
B
Before filtering
After filtering
mitochondrial genes / total genes
Read count
C
No data diffusion
t=1
t=3 (selected)
t=5
t=15
Lyz2
Sftpc
D
Pearson correlation
0.8
0.4
0.0
-0.4
-0.8
E
Top Differentially Expressed Genes
Cells arranged by cluster
Expression
F
Louvain clusters
Zmat3
Klf7
Elk3
Sox9
Foxq1
Maff
Camta1
Plscr1
Zfp664
Smad7
Trp53
Tead1
Ets2
Zfp560
Cebpa
Nfic
Yeats4
Foxp2
Etv5
Stat3
Srebf1
Irx3
Gtf2i
Nfix
Id1
Foxq1
Lyar
Cenpa
Max
Sox9
GEMM data
G
Sox9
Log expression
Louvain cluster
H
Sox9 target activation
z-score
Louvain cluster
I
Louvain clusters
Expression
J
Organoid Louvain clusters
C1 (Emp)
C2 (Cre)
C0 (Cre)
z-score
Louvain
Scgb1a1
Scgb3a2
Ly6a
Foxj1
Cd24a
Pdpn
Aqp5
Etv5
Lyz2
Sftpc
Club
Progenitor
Ciliated
AT1
AT2
Cells
K
Gene Ontology overlap
C1ORG
C0ORG
C2ORG
68
1
230
2
10
35
119

Figure S4, related to Figure 4: $KRAS^{G12D}$ expressing organoid cells are transcriptionally distinct and transition to a developmental-like state

(A) Representative FACS plot showing cell sorting strategy of organoids. FMO = fluorophore minus one.

(B) Correlation between mitochondrial gene expression and read count before and after filtering cells with > 10% mitochondrial gene expression.

(C) Correlation between *Sftpc* and *Lyz2* expression in individual cells after different numbers of data diffusion (t) (Van Dijk et al., 2018).

(D) Similarity between each Louvain cluster calculated using the Pearson's correlation co-efficient.

(E) Top differentially expressed genes in each Louvain cluster, visualized with a heatmap.

(F) Heatmap depicting expression of transcription factors/co-factors identified amongst differentially expressed genes.

(G) Log expression of indicated gene in GEMM Louvain clusters 0 and 1. Dashed line marks median expression of the reference sample.

(H) Z-score of indicated signature in GEMM Louvain clusters 0 and 1. Dashed line marks median of reference sample

(F)(G)(H) Log expression of indicated genes. Dashed line marks median expression of the reference sample.

(I) RNA velocity analysis of the early-stage $KRAS^{G12D}$ GEMMs scRNA-Seq dataset. Louvain clusters for YFP^{-} (C0) and YFP^{+} (C1) AT2 clusters is shown on the left. *Sox9* expression is visualized on the right.

(J) Correlation heatmap of individual cells of the organoid scRNA-Seq data (x-axis) and z-normalized gene expression of indicated lineage genes (y-axis). Cells are ordered based on correlation distance calculation. Louvain clusters are annotated.

(K) Number of overlapping and unique Gene Ontology terms in each Louvain cluster in the KY organoid dataset.

P-values were determined using a Mann-Whitney rank test *** = p-value > 0.001.

Figure S5

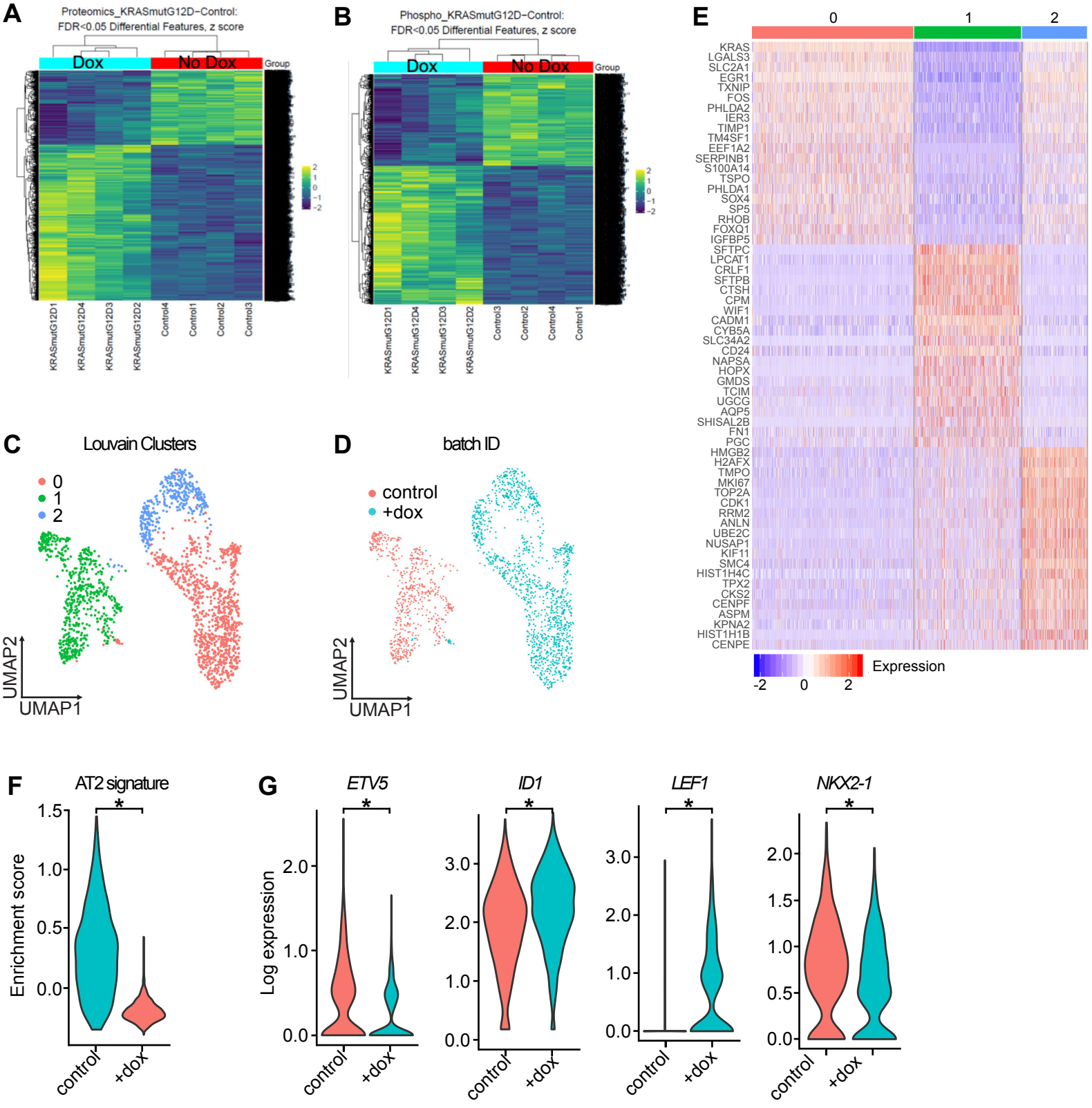

A
Proteomics_KRASmutG12D−Control:
FDR<0.05 Differential Features, z score
Dox
No Dox
Group
KRASmutG12D1
KRASmutG12D4
KRASmutG12D3
KRASmutG12D2
Control4
Control1
Control2
Control3
B
Phospho_KRASmutG12D−Control:
FDR<0.05 Differential Features, z score
Dox
No Dox
Group
KRASmutG12D1
KRASmutG12D4
KRASmutG12D3
KRASmutG12D2
Control3
Control2
Control4
Control1
E
0
1
2
KRAS
LGALS3
SLC2A1
EGR1
TXNIP
FOS
PHLDA2
IER3
TIMP1
TM4SF1
EEF1A2
SERPINB1
S100A14
TSPO
PHLDA1
SOX4
SP5
RHOB
FOXQ1
IGFBP5
SFTPC
LPCAT1
CRLF1
SFTPB
CTSH
CPM
WIF1
CADM1
CYB5A
SLC34A2
CD24
NAPSA
HOPX
GMDS
TCIM
UGCG
AQP5
SHISAL2B
FN1
PGC
HMGB2
H2AFX
TMPO
MKI67
TOP2A
CDK1
RRM2
ANLN
UBE2C
NUSAP1
KIF11
SMC4
HIST1H4C
TPX2
CKS2
CENPF
ASPM
KPNA2
HIST1H1B
CENPE
Expression
-2
0
2
C
Louvain Clusters
0
1
2
UMAP2
UMAP1
D
batch ID
control
+dox
UMAP2
UMAP1
F
AT2 signature
Enrichment score
1.5
1.0
0.5
0.0
control
+dox
G
ETV5
ID1
LEF1
NKX2-1
Log expression
2.0
1.0
0.0
3.0
2.0
1.0
0.0
3.0
2.0
1.0
0.0
2.0
1.5
1.0
0.5
0.0
control
+dox

Figure S5, related to Figure 5: Human iAT2s downregulate differentiation and maturation markers and upregulate progenitor markers upon KRAS^{G12D} expression

(A) Heatmap indicating differentially expressed proteins between dox induced and control iAT2s.

(B) Heatmap indicating differentially expressed phosphoproteins between dox induced and control iAT2s.

(C) Clustering of transcriptomes using UMAP. Cells are colored based on Louvain clusters

(D) and batch ID.

(E) Heatmap visualizing the top 20 differentially expressed genes in each Louvain cluster.

(F) Enrichment score of gene signature comprised of AT2 signature genes shared between mouse and human from the Panglao database (table S2). P-values were determined using a Welch Two Sample t-test. *p<0.05.

(G) Log expression of indicated genes. P-values were determined using the MAST single-cell test. *p<0.05.

Figure S6

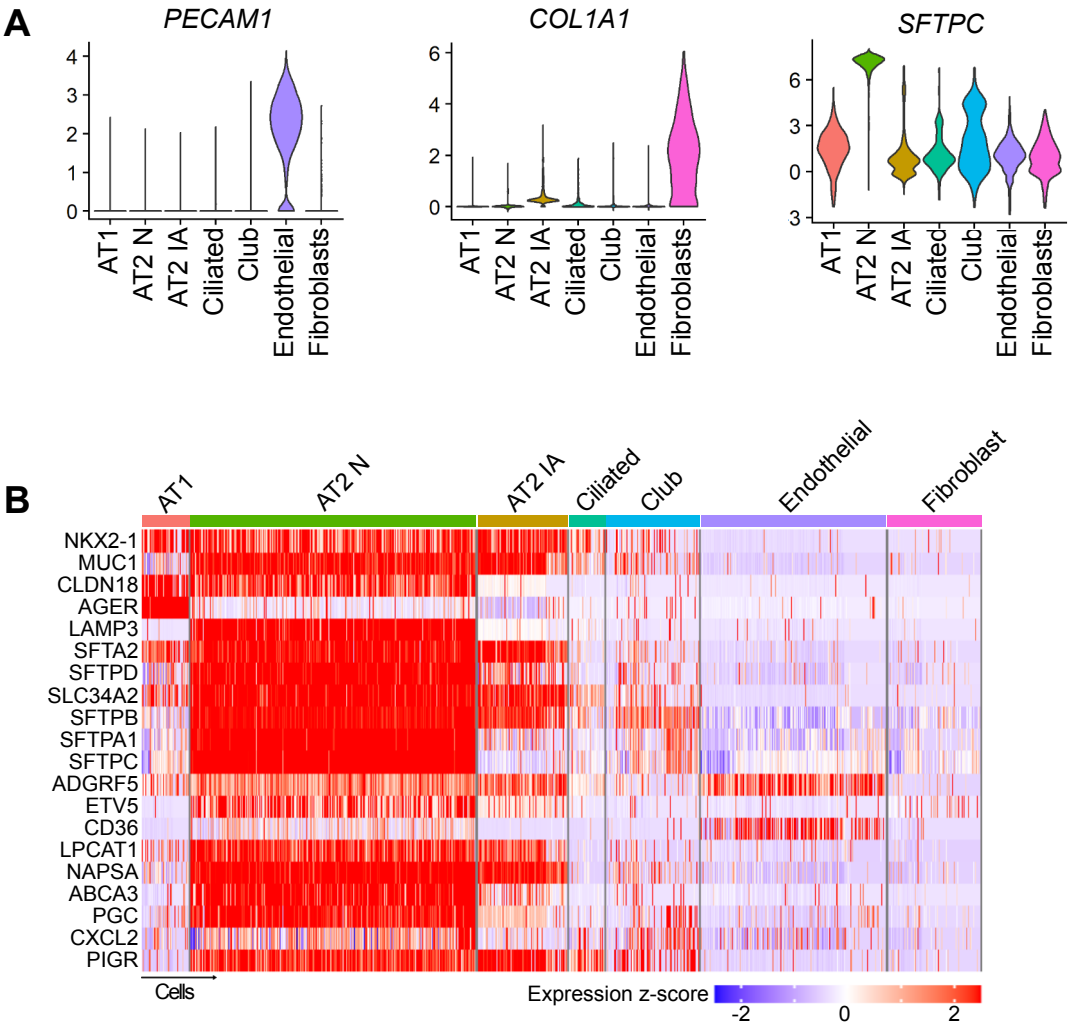

A
PECAM1
COL1A1
SFTPC
AT1
AT2 N
AT2 IA
Ciliated
Club
Endothelial
Fibroblasts
B
AT1
AT2 N
AT2 IA
Ciliated
Club
Endothelial
Fibroblast
NKX2-1
MUC1
CLDN18
AGER
LAMP3
SFTA2
SFTPD
SLC34A2
SFTPB
SFTPA1
SFTPC
ADGRF5
ETV5
CD36
LPCAT1
NAPSA
ABCA3
PGC
CXCL2
PIGR
Cells
Expression z-score
-2
0
2

Figure S6, related to Figure 6: Differentiation and maturation markers are downregulated in AT2 cells from human early stage LUAD

(A) Violin plots showing gene expression values of selected genes in annotated clusters.

(B) Heatmap showing expression of AT2 signature genes shared between mouse and human from the Panglao database.