Alveolar epithelial cell plasticity and injury memory in human pulmonary fibrosis

Authors:

Taylor S Adams[1,*], Jonas C Schupp[1,2,3,4,*], Agshin Balayev[1], Johad Khoury[1], Aurélien Justet[1,5], Fadi Nikola[1], Laurens J De Sadeleer[6,7], Juan Cala-Garcia[1,8], Marta Zapata-Ortega[6], Panayiotis V Benos[9], John E McDonough[1,6,10,11], Farida Ahangari[1], Melanie Königshoff[12], Jun Ding[13,14,15,16], Robert J Homer[17], Ivan Rosas[8], Xiting Yan[1,18], Bart M Vanaudenaerde[6], Wim A Wuyts[6,7], Naftali Kaminski[1]

*Both authors contributed equally to this work.

Affiliations:

[1] Section of Pulmonary, Critical Care and Sleep Medicine, Yale School of Medicine, New Haven, CT, USA
[2] Department of Respiratory Medicine and Infectious Diseases, Hannover Medical School, Hannover, Germany
[3] Biomedical Research in Endstage and Obstructive Lung Disease Hannover (BREATH), German Center for Lung Research (DZL), Hannover, Germany
[4] Department of Clinical Airway Research, Fraunhofer Institute for Toxicology and Experimental Medicine (ITEM), Hannover, Germany
[5] Department of Respiratory Medicine, University Hospital of Caen, France; Unicaen, CEA, CNRS, UMR 6030, ISTCT Group, GIP Cyceron, Université de Normandie, Caen, France
[6] Laboratory of Respiratory Diseases and Thoracic Surgery (BREATHE), Department of Chronic Diseases and Metabolism, KU Leuven, Leuven, Belgium
[7] Unit for Interstitial Lung Diseases, Department of Pneumology, University Hospitals Leuven, Leuven, Belgium
[8] Section of Pulmonary, Critical Care and Sleep Medicine, Baylor College of Medicine
[9] Department of Epidemiology, University of Florida, Florida, USA
[10] Department of Medicine McMaster University Hamilton, Ontario, Canada.
[11] Firestone Institute for Respiratory Health St. Joseph's Healthcare Hamilton Hamilton, Ontario, Canada
[12] Department of Medicine, University of Pittsburgh, Pittsburgh, Pennsylvania, USA

[13] Department of Medicine, Faculty of Medicine and Health Sciences, McGill University, Montreal, Quebec, Canada
[14] Meakins-Christie Laboratories, Translational Research in Respiratory Diseases Program, Research Institute of the McGill University Health Centre, Montreal, Quebec, Canada
[15] School of Computer Science, McGill University, Montreal, Quebec, Canada
[16] Mila - Quebec AI Institute, Montreal, Quebec, Canada
[17] Department of Pathology, Yale School of Medicine, New Haven, CT, USA
[18] Department of Biostatistics, Yale School of Public Health, New Haven, CT, USA

Corresponding author:

Naftali Kaminski. Boehringer-Ingelheim Endowed Professor of Internal Medicine. Chief of Pulmonary, Critical Care and Sleep Medicine. Yale School of Medicine. 300 Cedar Street, TAC–441 South. P.O. Box 208057. New Haven, CT 06520-8057. Tel: 203-7854162, Fax: 203-7853826, E-Mail: naftali.kaminski@yale.edu

Abstract

Acute and repetitive lung epithelial injury can lead to irreversible and even progressive pulmonary fibrosis; Idiopathic pulmonary fibrosis (IPF) is a fatal disease and quintessential example of this phenomenon. The composition of epithelial cells in human pulmonary fibrosis – irrespective of disease etiology – is marked by the presence of Aberrant Basaloid cells: an abnormal cell phenotype with pro-fibrotic and senescent features, localized to the surface of fibrotic lesions. Despite their relevance to human pulmonary fibrosis, the exotic molecular profile of Aberrant Basaloid cells has obscured their etiology, preventing insights into how or why these cells emerge with fibrosis. Here we identify cellular intermediaries between Aberrant Basaloid and normal alveolar epithelial cells in human IPF tissue. We track the emergence of Aberrant Basaloid cells from alveolar epithelial cells *ex vivo* and uncover a role for similar cells in epithelial regeneration under normal conditions. Lastly, we characterize the epigenetic changes that distinguish Aberrant Basaloid cells from their progenitors and identify hallmarks of AP-1 injury memory retention. This study elucidates the phenomenon of maladaptive epithelial plasticity and regeneration in pulmonary fibrosis and re-contextualizes therapeutic strategies for epithelial dysfunction.

Main

The alveoli of the lung are the primary interface between the atmosphere and our circulation. Each alveolus is structurally optimized to maximize two critical functions: gas exchange and pulmonary compliance. On the alveolar surface, these tasks are delegated to two distinct types of epithelial cells: alveolar epithelial type I (ATI) cells enable gas-diffusion, while alveolar type II (ATII) cells produce surfactants that reduce surface tension[1,2]. Environmental exposure carries the risk of injury and ATII serve as the primary stem cells for alveolar epithelial regeneration. While alveolar epithelial injury and regeneration are a normal part of life, repeated injury can potentially lead to progressive and irreversible fibrosis. Understanding both functional and maladaptive responses to alveolar epithelial regeneration remain critical topics in pulmonary medicine[3,4].

Single cell RNA sequencing (scRNAseq) studies enabled the discovery of Aberrant Basaloid cells in tissue from patients with end-stage human lung fibrosis[5,6]. Aberrant Basaloid cells lack hallmarks of ATII and ATI cells and are characterized by a pronounced epithelial mesenchymal transition (EMT) signal (CDH2, VIM, COL1A1) and some overlapping features with airway basal cells (KRT17, TP63, LAMB3). Notably, this Aberrant Basaloid cell phenotype is found across many diseases including interstitial lung diseases like idiopathic pulmonary fibrosis (IPF)[5,6] and systemic sclerosis[7], COPD[6] and COVID-19 induced fibrosis[8,9]. A concurrent discovery was made in mice of Transitional Krt8+ cells: a cell phenotype which emerges from ATII following pulmonary injury and persist until resolution before becoming ATI cells[10–15]. Murine Transitional Krt8+ cells are a conserved phenotype following genotoxic, endotoxic and mechanical forms of lung injury: they lack hallmarks of ATII and ATI cells and are characterized by expression of Cldn4, Krt8, Krt19, Sfn and Fn1[10–15]. Despite their situational likeness, murine Transitional Krt8+ cells lack most defining features of human disease Aberrant Basaloid cells[16]; this lack of apparent molecular similarity raises questions regarding the degree of cellular orthology between the two epithelial phenotypes. Unlike Transitional Krt8+ cells, the cellular origin of Aberrant Basaloid cells in human disease remains unknown. More importantly, it remains unclear whether Aberrant Basaloid cells play a normal role in human epithelial injury repair or whether their extreme

expression profile and persistence through end-stage fibrosis are indicative of an abnormal phenomenon.

100 In this study, we perform single nuclei RNA sequencing (snRNAseq) on differentially fibrotic regions in human IPF lungs and identify intermediate states between Aberrant Basaloid cells and other epithelial cell types. We longitudinally profile a model of Aberrant Basaloid development and uncover a role for these cells in normal regeneration. Lastly, we dissect the epigenetic profile of Aberrant Basaloid cells and describe what makes them fundamentally different from normal
105 epithelial cells.

Results

Tissue specimens of differentially fibrotic IPF and control lungs were prepared and evaluated as described by McDonough *et al.*[17]. Explanted IPF lungs or donor lungs declined for transplantation
110 were inflated, frozen and CT scanned. Small cores of tissue were isolated and subject to microCT assessment of surface density: a proxy of localized fibrosis (Fig. 1A). snRNAseq was then performed on 54 control and IPF cores with variable levels of fibrosis, and cell types were classified into major lineages (Figures 1A, S1A). To provide further context, epithelial nuclei were integrated with epithelial cell data from two previously published scRNAseq datasets of lung
115 disease (Fig. 1A; S1A)[5,6] before cell types were collectively re-classified (Fig. 1B, Supp. Table S1).

A "Missing Link" Between Aberrant Basaloid and Alveolar Epithelial Cells

The unbiased cell representation afforded by snRNAseq (Figure S1A) enabled us to identify a
120 previously overlooked population of epithelial cells in IPF, herein referred to as Alveolar Intermediate (AT*i*) cells (Figure 1B). Analyzing the intersection of disease-associated cell marker genes with otherwise distinct markers from control epithelial cell types, we observe AT*i* cells to share some features with ATI (NTM, GPC5), ATII (SERPINA1, SFTA2) and Aberrant Basaloid

cells (Fig. 1C,D). The distinct combination of positive and negative features characterizing AT*i* cells (Fig. 1D) indicate that these are a unique population of cells straddling the phenotypic boundary of otherwise distinct cell states in health and disease. AT*i* cells share only a subset of features with Aberrant Basaloid cells (Fig. 1D): Overlapping features include DNA damage-related genes MDM2 and GDF15; elevated levels of TGFβ-activating integrin subunit ITGB6 and distinct expression of TGFβ signaling genes TGFBR2, SMAD4 and MIR100HG - a host gene for micro-RNA let-7a; hedgehog signaling repressor PTCHD4; the Angiotensin-1 converting protease CPA6, and CTSE – a metastatic hallmark of pancreatic ductal adenocarcinoma and Barretts' Esophagus[18,19]. In contrast, AT*i* cells lack basaloid features TP63 and KRT17; senescence hallmarks CDKN2A and CDKN2B; TGFβ-inducible NOX4, EPHB2, CASC15 and TGFBI; extracellular matrix (ECM) remodeling genes COL1A1, COL5A1, PRSS2, SPINK1 ADAMTS9, MMP7; the epithelial-mesenchymal transition (EMT) hallmark CDH2 and mutation-inducing transposase PGBD5 (Figure 1D).

AT*i* and Aberrant Basaloid cells with the same features are also observed in publicly available datasets of end-stage COVID-19 lungs[9,20,21] (Supp. Fig. S2A, Supp. Table S8) extending their relevance from chronic to acute etiologies of pulmonary fibrosis.

Composition of Abnormal Epithelial Cells in IPF

The median percent of epithelial cells classified as Aberrant Basaloid are similar among IPF samples across all 3 datasets, ranging from 2.72% to 2.91% (Figure 1E). AT*i* cells are significantly enriched in IPF samples vs controls in snRNAseq (Wilcox pVal=2.165E-5) and in Adams *et al.* (Wilcox pVal=0.043), though not in Habermann *et al.* (Wilcox pVal=0.22; Fig. 1E). The proportion of AT*i* cells is significantly higher in IPF snRNAseq samples (median 3.87%) when compared to IPF samples from either scRNAseq datasets (1.32%, p=0.0015 in Adams *et al.*; 2.17%, p=0.038 in Habermann *et al.*; Figure 1E), suggesting that cell dissociation survivor bias negatively impacted the recovery of these cells in previous studies.

150 We explored the relationship between tissue surface density, a proxy for fibrosis, and the percent of abnormal epithelial cells. The decline in surface density across IPF tissues is significantly associated with a decline in the proportion of ATII cells (p=5.34e-4) alongside an increase in large airway-associated secretory (p=0.013) and ciliated cells (p=0.023) (Figure 1F; Supp. Table S2). Neither Aberrant Basaloid nor AT*i* cell proportions had significant relationships with tissue
155 surface density in IPF (Fig. 1F, Supp. Table S2). This suggests that the emergence of these cells pre-dates the development of pathophysiologic characteristics of fibrosis and it is thus a potential target of anti-fibrotic drugs.

Assessing co-occurrence patterns across all lung cores, the most positive compositional correlates of surface density are unsurprisingly ATI and ATII cells (Figure 1G). Notably, the strongest
160 abundance correlates of AT*i* cells are Aberrant Basaloid cells (Spearman's rho=0.813) followed by epithelial cells in cell cycle (Spearman's rho=0.442; Figure 1G), indicating spatial associations with both senescent and proliferating epithelial cells.

The Continuum of States Between ATI, ATII and Aberrant Basaloid Cells

165 After evaluating disease-relevant cell types in discrete terms, we assessed their phenotypic connectivity to healthy cell types to model continuous trajectories. Common epithelial cell types were re-clustered into granular sub-populations; inter-connectivity between sub-populations was evaluated with the lineage reconstruction technique "partition-based graph abstraction" (PAGA)[22]. Using an edge confidence threshold > 0.2, all healthy cell types remain connected with a plausible
170 lineage topology, and Aberrant Basaloid cells are exclusively connected to the graph via IPF AT*i* cells towards either ATI or ATII cells (Fig. 2A). To avoid spurious connections in trajectory analysis, low confidence edge pruning from the PAGA graph was extended to connections in the original nuclei-level graph, and a force-directed graph layout of pruned nuclei similarities was built to complement pseudotime analysis (Fig. 2A). Lacking certainty in the orientation of
175 trajectories, we chose to model independent trajectories from either ATII or ATI nuclei towards disease-restricted Aberrant Basaloid nuclei. Generalized additive models (GAMs) of genes were

fit to changes in pseudotime across nuclei from each trajectory and organized pseudotemporally by expression peak.

Of the two trajectories, nuclei along the trajectory from ATII towards Aberrant Basaloid features
180 a more gradual pseudotime mapping and more detailed model of pseudotemporal gene dynamics (Fig. 2B). While markers of ATII cell identity (SFTPC, LAMP3, PDE4D, ERBB4) decline, "Early Peak" features emerge including Let-7a host gene MIR100HG, unconventional myosin MYO16 and the peptidase CTSE (Fig. 2B,C). "Early Peak" features then decline with the loss of cell identity and rise of "Late Peak" genes – including: GDF15, FGF13, ITGAV and CPA6 – before
185 specific Aberrant Basaloid hallmarks emerge (TP63, CDH2, CASC15, CDKN2A; Fig. 2B,C). Trajectory analysis of ATI towards Aberrant Basaloid revealed a similar – though less detailed – sequence of gene expression patterns. ATI features (RTKN2, NCKAP5, CAV1) decline while "Early Peak" features (CTSE, MYO16, STK32A, NTM, GPC5) rise prior to "Late Peak" features (GDF15, CPA6), followed by peaks of Aberrant Basaloid hallmarks (CDH2, TP63, CDKN2A;
190 Fig. 2C,D). Though it remains unclear whether these trajectories represent real lineage fates, this analysis demonstrates that AT*i* cells are a diverse population occupying a branching continuum of phenotypes between ATI, ATII and IPF Aberrant Basaloid cells.

Immunofluorescent microscopy of control and IPF tissues confirmed the presence of both SFTPC+ ATII cells and AGER+ ATI cells with "Early Peak" AT*i* marker CTSE in affected regions of IPF
195 lungs (Fig. 3B) but not in control lungs (Fig. 3A). Despite "Late Peak" features like CPA6 transcriptionally emerging with the loss of cell-type markers (Fig. 2B,D), we observed many SFTPC+/CPA6+ and AGER+/CPA6+ cells in IPF lungs, as well as in hyper-keratinized SFTPC-/AGER-/CPA6+ cells (Fig. 3D), confirming the presence of AT*i* populations in IPF lungs.

200 Longitudinal Assessment of *ex vivo* Fibrosis

To study the emergence of Aberrant Basaloid cells in real time and compare *ex vivo* outcomes to *in vivo* disease, we took advantage of a recent observation that Aberrant Basaloid cells arise *ex vivo* from disease-free human precision cut lung slices (PCLS) following 5 days of exposure to a

"fibrotic cocktail" (FC: PDGF-αβ, TGF-β, LPA, and TNFα)[23–25]. We longitudinally assessed this model with snRNAseq (Fig. 4A,B) and observed that both cocktail and control PCLS undergo a decline in epithelial cells over the first 48 hours, before gradually rising (Fig. 4C, Supp. Table S3, Supp. Table S4). Concurrent with this gradual rise is a dramatic increase in epithelial cell proliferation: 3-8% of epithelial cells from 72 hours onwards are undergoing mitosis (Fig. 4C). Hereby, FC exposure significantly reduces the rate of epithelial mitosis between 72 and 120 hours (t-test pVal= 0.010607), from an average of 6.3% in controls to 3.8% with FC (Fig. 4C).

The phenotypic landscape of PCLS alveolar epithelial cells is characterized by many distinct phenotypic substates, associated with different points in time (Fig. 4D,E). All substates could be classified under *in vivo* conventions based on the conservation of marker genes: CAV1, COL4A1, COL4A2 in ATI and PDE4D, ACOXL and ABCC4 in ATII cells (Fig. 4E, Supp. Fig. S3A). The majority of different substates are associated with samples from the first 48 hours, reflecting a period of dynamic behavior with characteristic expression patterns. Early timepoint ATII subtypes exhibit features of stress (HSPA1A, STIP1) and metabolic change (ALDH1L2, GK, KYNU) (Fig. 4E, Supp. Fig. S3E). ATI substates enriched in the first 48 hours exhibit hallmarks of stress (FOS, JUN, HSPA1A), cell cycle arrest (CDKN1C, CCN2, ID2), endoplasmic reticular stress (ERO1B, CLPTM1L, PLA2G4) and cytoskeleton regulation (FSCN1, SAMSN1) (Fig. 4E, Supp. Fig. S3B). PCLS epithelial cells from 72 hours onwards are primarily classified into just three phenotypes: "ATIend", "ATIIend" and Aberrant Basaloid – suggesting that some degree of homeostasis has been achieved. Of interest, Aberrant Basaloid cells were detected in both control and FC PCLS as early as hour 24 onwards. Aberrant Basaloid cells are significantly more common in FC samples from 72 hours onwards vs control (tTest pVal=0.004385), increasing from an average 1.25% of epithelial cells in control PCLS to 30.4% following FC treatment (Fig. 4E).

A detailed description of temporal ATI and ATII gene expression dynamics is provided in the supplement (Supp. Discussion; Figs. S3, S4). These results depict a biphasic response of alveolar epithelial cells to PCLS experiment conditions where 48 hours of injury response and acclimation are followed by several days of phenotypic stability and tenacious regeneration.

Comparing control, fibrotic cocktail and IPF Aberrant Basaloid cells

To model the etiology of Aberrant Basaloid cells in PCLS, graph connectivity and pseudotime analysis was performed as described earlier. Similar to *in vivo* results, the strongest connectivities to PCLS Aberrant Basaloid cells are substates of both ATI and ATII cells (Supp. Fig. S3B). Unlike the *in vivo* results, the AT*i* gene signature was not particularly representative of any cell population detected in PCLS, with the prominent exception of FRMD5 and MIR100HG in the PCLS pseudotime trajectory from ATII towards Aberrant Basaloid cells (Supp. Fig. S3E).

Comparing markers of PCLS "ATIend", "ATIIend" and Aberrant Basaloid to their *in vivo* IPF equivalents, we find most *in vivo* ATI and ATII hallmarks are conserved *ex vivo* (Fig. 4F, Supp. Table S5). Although only 17.5% of the *in vivo* Aberrant Basaloid signal overlaps with ex vivo variants, this overlap encapsulates 54.6% of their defining ex vivo features (Fig. 4F). Control PCLS Aberrant Basaloid cells share many features with FC and *in vivo* variants, including KRT17, ADAMTS9, SLC35F3, PTGS2, HMGA2, FOSL2 and elevated levels of KRT7 and ITGAV (Fig. 4F). However, unlike FC PCLS Aberrant Basaloid cells, control PCLS variants notably lack core disease hallmarks of aberrancy (CDH2, COL1A1, SOX4, SERPINE1, EPHB2) and basal-likeness (TP63) (Fig. 4F). Though Aberrant Basaloid-like cells are observed in control PCLS, FC stimulation induces a more complete molecular representation of Aberrant basaloid cells described in IPF.

Among proliferating epithelial cells, ATII were the most commonly detected, though SFTPB+ Secretory, Aberrant Basaloid and even ATI cells with mitotic hallmarks are found. Strikingly, 30% of all control Aberrant Basaloid cells are observed undergoing mitosis compared to just 0.5% of FC stimulated variants (Fig. 4G); for reference, 7% of all ATII cells and 1.2% of ATI cells are observed mitotic from 24 hours onwards. From hour 72 onwards, FC induced a dramatic decline in the percent of mitotic Aberrant Basaloid (tTest pVal=0.04077) from a per-sample average of 32.6% in controls to just 0.49% (Fig. 4G). FC did not significantly affect the percent of ATII cells undergoing mitosis over this time (tTest pVal=0.6304), indicating that the overall decline in

epithelial proliferation induced by FC arose not from direct effects on ATII proliferation, but rather their displacement with senescent Aberrant Basaloid cells.

FC stimulation thus confers a more pathologically relevant presentation of Aberrant Basaloid cells in PCLS in terms of abundance, molecular signature and cell cycle arrest. The observations that control PCLS aberrant basaloid cells are rare yet frequently found dividing, suggest they are normally a transient phenotype associated with homoeostatic repair response. This contrasts with abundant yet rarely dividing Aberrant basaloid Cells found in FC-stimulated PCLS, patterns indicative of a change towards persistence and senescence. Beyond highlighting FC's ability to emulate the disease phenotype of Aberrant Basaloid cells, these results place a spotlight on control PCLS Aberrant Basaloid cells, understanding their role and plasticity may be key to distinguishing normal repair from maladaptive outcomes.

Chromatin Accessibility Patterns of Aberrant Basaloid Cells and ATi cells

To explore the epigenetic landscape of epithelial cells in lung fibrosis, we subjected three control and seven IPF lung samples to snMultiOme analysis, where each nucleus is simultaneously profiled with snRNAseq and snATAC-seq chemistry (Fig. 5A). Cell types were classified by transcriptional identity, and ATAC fragments from each cell type per sample were pooled for pseudo bulk peak calling using the R package ArchR[26]; fragments from merged peaks were used in downstream analyses (Fig. 5A). UMAP embeddings of gene expression and chromatin-accessibility show remarkable topological similarities among epithelial cells (Fig. 5B,C). ATI, ATII and Aberrant Basaloid cells occupy unique regions of both RNA and ATAC feature-space. AT*i* straddle the feature-space between these cell types in both UMAP embeddings, however their ATAC profile was far less distinct (Supp. Table S6). The tight co-localization of Aberrant Basaloid cells in unsupervised ATAC-space indicates a divergence from normal epithelial cell chromatin accessibility.

We compared differentially expressed genes to differentially accessible genes using ArchR's Gene Score[26] (Supp. Table S6). Across epithelial cell types, distinct marker genes often have

correspondingly distinct accessibility patterns – MIR205HG in basal cells, FOXJ1 in ciliated, LAMP3 and SFTPC in ATII, RTKN2 and NCKAP5 in ATI (Fig. 5D). Likewise, many uniquely expressed genes in Aberrant Basaloid cells are associated with uniquely accessible chromatin, including TGF-B signaling mediators SKIL, FSTL1, LTBP1 and HMGA2, as well as EMT hallmarks CDH2, COL6A1 and COL18A1 (Fig. 5D). AT*i* cells have no significantly distinct ATAC Gene Scores, but – alongside Aberrant Basaloid cells – they share ATI features like KRT7, NTM and collagen IV and show elevated accessibility of pathological features like CPA6 and the cytoskeleton regulators FGD6 and TMEM132D (Fig. 5D). Collectively, Aberrant Basaloid cells are both epigenetically and transcriptionally distinct from normal lung epithelial cells. The chromatin accessibility profile of AT*i* cells is less categorically distinct, consistent with their diverse intermediate transcriptional states.

AP-1 Transcription Factor Accessibility is a Defining Feature of Aberrant Basaloid Cells

We explored patterns of transcription factor (TF) motif enrichment across differentially accessible peaks using the TF database CIS-BP[27]. TF motifs with the largest contributions to variance in the data are primarily associated with specific epithelial cell types (Fig. 5E). The top TF families associated with ATII cells are C/EBP, NKX-2 and FOXA1/2; in ATI they are TEAD, KLF, GATA (Supp. Table S7) – observations consistent with previous studies[28–30]. Differentially accessible motif positions from these families have cis-regulatory links to over 60% of their respective transcriptional features (Supp. Fig. S4A,B).

TF motifs for JUND, FOSL2, SMARCC1, JUNB, FOSL1, FOS, JUN, BACH2, BACH1, FOSB and JDP2 are outliers in terms of their contribution to overall variance in the data (Fig. 5E, left) and in terms of the significance of their enrichment among Aberrant Basaloid relative to other cells (pVal<1E-20; Figure 5SE, right). Further examination of the top 15 significantly enriched TF motifs among Aberrant Basaloid cells revealed that they all shared a common sequence: TGA G/C

310 TCA (Fig. 5F, Supp. Table S7) – the consensus sequence for AP-1 – we subsequently focused on this particular sequence motif.

Though 12.5% of all merged peaks contain one or more AP-1 consensus sequences, they are found in over 50% of all differentially accessible peaks in Aberrant Basaloid cells (hypergeometric test p-value < 1e-50; Fig. 5G). Cell-level AP-1 enrichment z-scores are profoundly elevated in
315 Aberrant Basaloid compared to other cell types (Wilcoxon test pVal=2.2e-16; Fig. 5I), where the median cell's AP-1 enrichment zScore is 5.29; Alveolar Intermediate cells have the second highest median zScore of 1.09. Footprint analysis of accessible chromatin surrounding AP-1 sequences reveals a prominent dip at the binding site among Aberrant Basaloid cells (Fig. 5I), indicating AP-1 TF binding. AP-1 TF motif sequences are found in the majority of differentially accessible
320 chromatin in Aberrant Basaloid cells, these cells have the highest breadth of accessible AP-1 binding sites of any cell in the lung and these sites are actively occupied.

Discussion

This study provides an unprecedented view of human alveolar epithelial plasticity and unearths
325 important context behind the identity of Aberrant Basaloid cells and their role in regeneration and disease. Analyzing differentially affected IPF tissues, we observe similar frequencies of aberrant basaloid cells across disease tissues, irrespective of their level of fibrosis. We uncover AT*i* cells: a diverse spectrum of cell phenotypes between Aberrant Basaloid cells and ATII and ATI cells – we verify their presence across multiple cohorts and diseases, and localize them in fibrotic tissues.
330 Tracking the emergence of these cells in human PCLS, we observe that fibrotic cocktail stimulation reduces the overall rate of epithelial cell mitosis while increasing the abundance of Aberrant Basaloid cells and imparting them with a molecular profile more aligned with human pulmonary fibrosis. Finally, we establish that IPF Aberrant Basaloid cells have a common epigenetic profile which is distinct from all other lung cell types, characterized by extreme levels
335 of AP-1 family TF chromatin accessibility and binding site occupation.

Upon their discovery, the exotic molecular profile of Aberrant Basaloid cells precluded the ability to infer where these cells originated, a limitation that undercut their translational impact[5,6]. Not only was it unclear which cells should be prophylactically targeted to prevent Aberrant Basaloid emergence, but interventions aimed at targeting the behavior of Aberrant Basaloid cells lacked a phenotypic compass to orient the desired outcome. The discovery of AT*i* cells in IPF and fibrotic COVID-19 lungs represents the first example of a "missing link" between these pathological cells and normal epithelial lung cell phenotypes in human lung disease. In both *in vivo* and *ex vivo* models of fibrosis, we observe phenotypic intermediate states between Aberrant Basaloid cells and both ATI and ATII cells – however the PCLS intermediate phenotypes are largely different from the AT*i* cells found in human disease. In addition, studies of cells exposed to fibrosis-associated signaling molecules have shown that human – but not mouse - ATII[31] and Basal cells[32] are each capable of achieving an Aberrant Basaloid feature set. Overall, this suggests that the Aberrant Basaloid cell phenotype is an example of cellular convergence: a singular state that different cell types acquire upon injury.

A key limitation of this study is its inability to provide evidence regarding the orientation of changes between these cell types, beyond the assumption that Aberrant Basaloid cells emerge from normal epithelial cells. It is further possible that AT*i* are not transitioning at all. In multiple murine models of fibrosis, it is well established that Transitional Krt8+ cells temporarily emerge from ATII before becoming ATI cells [10–13]; human disease Aberrant Basaloid cells could reflect malfunction in this process. Perhaps the better analogue to transient murine Transitional Krt8+ cells are the human control PCLS Aberrant Basaloid cells, serving conceptually similar roles in epithelial alveolar regeneration – a form of cellular homoplasy.

Genome-wide chromatin accessibility profiling reveals that IPF Aberrant Basaloid cells share a common epigenetic profile distinct from any other cell type in the lung, forming a unique valley in Waddington's epigenetic landscape model[33]. Their defining molecular hallmarks are often in regions of the genome uniquely accessible to them, denoting a direct relationship between their functional profile and epigenetic programming.

A defining property of the IPF Aberrant Basaloid epigenome is the degree of AP-1 TF accessibility on display. Gene expression of AP-1 is widely associated with cellular responses to injury and stress, including in lung epithelial cells following cigarette smoke exposure[34], viral[35] or bacterial infection[36], or in the stress response of adjusting to the PCLS environment (Supp. Fig. S4B). However, FOSL2 is the only AP-1 family gene we observe constitutively expressed in either IPF or *ex vivo* Aberrant Basaloid variants, although Footprint analysis indicates that TFs are occupying these binding sites. It is conceivable that these changes arose from transcriptional AP-1 responses to injuries which occurred long before the fibrotic lungs described herein were profiled. Although the AP-1 motif has been among many TF motifs enriched in murine Transitional Krt8+ cells[10,11] and aged human AT2 cells[37], this is the first time that genome-wide AP-1 epigenetics are implicated in epithelial cell dysfunction of human non-cancerous lung disease.

Similar AP-1 epigenetics are central to the first recognized instance of innate memory in non-immune cells: skin epithelial stem cells with a history of injury exposure maintained enhanced AP-1 chromatin accessibility and sustained AP-1 binding[38,39]. Retaining this inflammation memory allows cells to respond to future injuries faster and more robustly[40]. Maladaptive outcomes of AP-1 injury memory have been implicated in several diseases where chronic injury leads to epithelial stem cell metaplasia: from benign cases like nasal polyps in chronic rhinosinusitis to precancerous conditions like Barrett's Esophagus and pancreatic intra-epithelial neoplasia[18,19]. Our study describes the pulmonary disease rendition of AP-1 epigenetic associated dysfunction: where abnormal Aberrant Basaloid cells retain the memory of injuries responsible for triggering lung fibrosis.

Taken together, our study provides a holistic characterization of alveolar epithelial plasticity in the human fibrotic lung. We describe and explore cellular plasticity during injury and fibrosis, in which epithelial cell types converge on Aberrant Basaloid cells via AT*i* upon injury. The discovery of spontaneously occurring control Aberrant Basaloid cells opens new lines of research in lung regeneration in general. Finally, by uncovering the presence of human AT*i* cells in the IPF lung, and the unique nature of the epigenetic AP-1 injury memory that characterizes Aberrant Basaloid

390 Cells, we shift the view of the role of alveolar epithelial cells in IPF from 'innocent bystanders' to drivers of disease – and position Aberrant Basaloid Cells as potential targets for cell-based therapies in IPF.

References

395 1. Travaglini, K. J. *et al.* A molecular cell atlas of the human lung from single-cell RNA sequencing. *Nature* **587**, 619–625 (2020).

2. Wang, Y., Wang, L., Ma, S., Cheng, L. & Yu, G. Repair and regeneration of the alveolar epithelium in lung injury. *FASEB J* **38**, e23612 (2024).

3. Jones, D. L. *et al.* An injury-induced mesenchymal-epithelial cell niche coordinates
400 regenerative responses in the lung. *Science* **386**, eado5561 (2024).

4. Parimon, T. *et al.* Senescence of alveolar epithelial progenitor cells: a critical driver of lung fibrosis. *Am J Physiol Cell Physiol* **325**, C483–C495 (2023).

5. Habermann, A. C. *et al.* Single-cell RNA sequencing reveals profibrotic roles of distinct epithelial and mesenchymal lineages in pulmonary fibrosis. *Science Advances* **6**, eaba1972
405 (2020).

6. Adams, T. S. *et al.* Single-cell RNA-seq reveals ectopic and aberrant lung-resident cell populations in idiopathic pulmonary fibrosis. *Science Advances* **6**, eaba1983 (2020).

7. Valenzi, E. *et al.* Disparate Interferon Signaling and Shared Aberrant Basaloid Cells in Single-Cell Profiling of Idiopathic Pulmonary Fibrosis and Systemic Sclerosis-Associated
410 Interstitial Lung Disease. *Front Immunol* **12**, 595811 (2021).

8. Narasimhan, H. *et al.* An aberrant immune-epithelial progenitor niche drives viral lung sequelae. *Nature* **634**, 961–969 (2024).
9. Bharat, A. *et al.* Lung transplantation for patients with severe COVID-19. *Sci Transl Med* **12**, eabe4282 (2020).
10. Choi, J. *et al.* Inflammatory Signals Induce AT2 Cell-Derived Damage-Associated Transient Progenitors that Mediate Alveolar Regeneration. *Cell Stem Cell* **27**, 366-382.e7 (2020).
11. Kobayashi, Y. *et al.* Persistence of a regeneration-associated, transitional alveolar epithelial cell state in pulmonary fibrosis. *Nat Cell Biol* **22**, 934–946 (2020).
12. Jiang, P. *et al.* Ineffectual Type 2-to-Type 1 Alveolar Epithelial Cell Differentiation in Idiopathic Pulmonary Fibrosis: Persistence of the KRT8hi Transitional State. *Am J Respir Crit Care Med* **201**, 1443–1447 (2020).
13. Strunz, M. *et al.* Alveolar regeneration through a Krt8+ transitional stem cell state that persists in human lung fibrosis. *Nat Commun* **11**, 3559 (2020).
14. Niethamer, T. K. *et al.* Longitudinal single-cell profiles of lung regeneration after viral infection reveal persistent injury-associated cell states. *Cell Stem Cell* **32**, 302-321.e6 (2025).
15. Konkimalla, A. *et al.* Transitional cell states sculpt tissue topology during lung regeneration. *Cell Stem Cell* **30**, 1486-1502.e9 (2023).
16. Xie, T. *et al.* Abnormal respiratory progenitors in fibrotic lung injury. *Stem Cell Res Ther* **13**, 64 (2022).
17. McDonough, J. E. *et al.* Transcriptional regulatory model of fibrosis progression in the human lung. *JCI Insight* **4**, 131597 (2019).

18. Fisher, O. M. *et al.* High Expression of Cathepsin E in Tissues but Not Blood of Patients with Barrett's Esophagus and Adenocarcinoma. *Ann Surg Oncol* **22**, 2431–2438 (2015).
19. Azuma, T. *et al.* Expression of cathepsin E in pancreas: a possible tumor marker for pancreas, a preliminary report. *Int J Cancer* **67**, 492–497 (1996).
20. Delorey, T. M. *et al.* COVID-19 tissue atlases reveal SARS-CoV-2 pathology and cellular targets. *Nature* **595**, 107–113 (2021).
21. Melms, J. C. *et al.* A molecular single-cell lung atlas of lethal COVID-19. *Nature* **595**, 114–119 (2021).
22. Wolf, F. A. *et al.* PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. *Genome Biol* **20**, 59 (2019).
23. Lang, N. J. *et al.* Ex vivo tissue perturbations coupled to single-cell RNA-seq reveal multilineage cell circuit dynamics in human lung fibrogenesis. *Sci Transl Med* **15**, eadh0908 (2023).
24. Alsafadi, H. N. *et al.* An ex vivo model to induce early fibrosis-like changes in human precision-cut lung slices. *American Journal of Physiology-Lung Cellular and Molecular Physiology* **312**, L896–L902 (2017).
25. Lehmann, M. *et al.* Precision Cut Lung Slices: Emerging Tools for Preclinical and Translational Lung Research. An Official American Thoracic Society Workshop Report. *Am J Respir Cell Mol Biol* **72**, 16–31 (2024).
26. Granja, J. M. *et al.* ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. *Nat Genet* **53**, 403–411 (2021).

27. Weirauch, M. T. *et al.* Determination and inference of eukaryotic transcription factor sequence specificity. *Cell* **158**, 1431–1443 (2014).
28. Hassan, D. & Chen, J. CEBPA restricts alveolar type 2 cell plasticity during development and injury-repair. *bioRxiv* 2023.10.10.561625 (2023) doi:10.1101/2023.10.10.561625.
29. Penkala, I. J. *et al.* Age-dependent alveolar epithelial plasticity orchestrates lung homeostasis and regeneration. *Cell Stem Cell* **28**, 1775-1789.e5 (2021).
30. Zepp, J. A. *et al.* Genomic, epigenomic, and biophysical cues controlling the emergence of the lung alveolus. *Science* **371**, eabc3172 (2021).
31. Kathiriya, J. J. *et al.* Human alveolar type 2 epithelium transdifferentiates into metaplastic KRT5+ basal cells. *Nat Cell Biol* **24**, 10–23 (2022).
32. Uthaya Kumar, D. B. *et al.* Bronchial epithelium epithelial-mesenchymal plasticity forms aberrant basaloid-like cells in vitro. *Am J Physiol Lung Cell Mol Physiol* **322**, L822–L841 (2022).
33. *The Strategy of the Genes: A Discussion of Some Aspects of Theoretical Biology*. (Routledge, Abingdon, Oxon New York, New York, 2014).
34. Zhong, C.-Y., Zhou, Y.-M., Douglas, G. C., Witschi, H. & Pinkerton, K. E. MAPK/AP-1 signal pathway in tobacco smoke-induced cell proliferation and squamous metaplasia in the lungs of rats. *Carcinogenesis* **26**, 2187–2195 (2005).
35. Ludwig, S. *et al.* Influenza virus-induced AP-1-dependent gene expression requires activation of the JNK signaling pathway. *J Biol Chem* **276**, 10990–10998 (2001).
36. Rajasekaran, S. *et al.* Visualization of Fra-1/AP-1 activation during LPS-induced

inflammatory lung injury using fluorescence optical imaging. *Am J Physiol Lung Cell Mol Physiol* **309**, L414-424 (2015).

37. Liu, X. *et al.* Transcriptomics Analysis Identifies the Decline in the Alveolar Type II Stem Cell Niche in Aged Human Lungs. *Am J Respir Cell Mol Biol* **71**, 229–241 (2024).
38. Naik, S. *et al.* Inflammatory memory sensitizes skin epithelial stem cells to tissue damage. *Nature* **550**, 475–480 (2017).
39. Larsen, S. B. *et al.* Establishment, maintenance, and recall of inflammatory memory. *Cell Stem Cell* **28**, 1758-1774.e8 (2021).
40. Gonzales, K. A. U. *et al.* Stem cells expand potency and alter tissue fitness by accumulating diverse epigenetic memories. *Science* **374**, eabh2444 (2021).
41. Setty, M. *et al.* Characterization of cell fate probabilities in single-cell data with Palantir. *Nat Biotechnol* **37**, 451–460 (2019).
42. Van den Berge, K. *et al.* Trajectory-based differential expression analysis for single-cell sequencing data. *Nat Commun* **11**, 1201 (2020).

Data availability

Raw sequencing data generated for this study was deposited at Gene Expression Omnibus (GEO) under the accession number GSE286182.

Acknowledgments

We are grateful to all patients and control subjects who participated in this study. Special thanks to Mei Zhong from the Yale Stem Cell Center genomics core facility who performed sequencing,

and Joerg Niklaus from the Yale West Campus Imaging Core for provided microscope training and guidance.

Authors Contribution

NK supervised the study. NK and JCS conceptualized the study and acquired funding. JM performed the micro-CT. JCS, AB and TSA performed snRNAseq and analysis. TSA, AB and AJ performed snMultiOme and analysis. TSA performed immunofluorescent microscopy experiments. JK and AJ performed the PCLS time course experiments. AB, TA and JCS analyzed the COVID19 scRNAseq data. LDS, BMV, MZO and WAW assembled the IPF cohort and supervised the biobanking. IR, JCG provided healthy lung tissue. LDS, JD, FN, JCG, BVP, FA, RJH, IR, XY, BMV, WAW provided critical interpretation, review, and commentary on data and the manuscript. The manuscript was drafted by TSA, JCS and NK and reviewed and edited by all authors.

Funding

This project was supported by NIH grants R01HL127349, R01HL141852, U01HL145567 to NK, and a Three Lakes Foundations grant to NK and IOR; NIH grants R01HL159805, R01DK130294 to PVB; NIH grant R01LM014087 to XY; by the by DOD grant HT9425-23-1-0126 and Boehringer Ingelheim discovery award grant to JK; DOD grant W81XWH-19-1-0131, the Else Kröner-Fresenius Foundation (2023_EKCS.18) and the German Center for Lung research (FKZ 82DZL002B1, FKZ 82DZL002C1 & FKZ 82DZLT82C1) to JCS. JCS was supported by CORE100Pilot Advanced Clinician Scientist Program of Hannover Medical School funded by Else-Kröner-Fresenius Foundation (EKFS, 2020_EKSP.78) and the Ministry of Science and Culture of Lower Saxony. BMV and WAW are funded by the KULeuven research fund (C16/19/005) supporting the biobanking and lung core imaging.

Competing interests

NK reports consulting to Boehringer Ingelheim, Pliant, GSK, Three Lake Partners, Merck, Astra-Zeneca, RohBar, BMS, Galapagos, Chiesi, Sofinnova, Fibrogen, reports Equity in Pliant and grants from Astra Zeneca and BMS. NK has IP on novel biomarkers and therapeutics in IPF licensed to Biotech. JCS served as a consultant to Boehringer Ingelheim, Merck/MSD, GSK, AOP Health, and received lecture honoraria from Boehringer Ingelheim, GSK, Kinevant. JCS has IP on basal cell-targeted therapies in IPF.

Figure Legends and Tables

Figure 1: Study design and epithelial cell atlas of interstitial lung disease.

A) Overview of tissue procurement and experiment design. B) UMAPs of epithelial cells from integrated analysis of multiple datasets. C) Euler diagram of intersecting conserved gene expression markers between normal and disease-associated cell types. D) Heatmap of corresponding features. Each column represents the average normalized gene expression of one patient. Within each dataset, values for each gene are min-max normalized to a common range. E) Boxplots of cell type abundances normalized to the total number of epithelial cells per subject. F) Left: Stacked bar plots of epithelial cell compositions per core in descending rank order of tissue surface density. Right: Linear regression models of changes in epithelial cell proportions as a function of surface density. G) Heatmap of spearman distances among cell-type proportions and surface density, across lung cores.

Figure 2: Continuous changes between healthy and disease alveolar epithelial cells

A) From left to right: UMAPs of epithelial nuclei labelled by cell type, Louvain cluster and disease; PAGA graph of Louvain clusters with edges colored by confidence levels above (black) or below (red) the threshold of 0.2; pruned PAGA graph in Force Atlas 2 (FA2) layout; FA2 layout

of nuclei with corresponding edge pruning labelled by Louvain cluster, disease and cell type. B) Upper left: FA2 layout of pseudotime trajectory from ATII to Aberrant Basaloid. Upper right: Generalized additive model fits of genes changing as a function of pseudotime. Lower: heatmap of corresponding genes; each column is one nucleus randomly sampled from fixed intervals in pseudotime, gaps correspond to pseudotime intervals with missing representation. C) FA2 layouts colored by gene expression. D) Pseudotime trajectory from ATI to Aberrant Basaloid, represented in the same manner as Fig 2B.

Figure 3: Immunofluorescent Imaging of AT*i* Cells in IPF

A) CTSE expression in control alveolar epithelial cells. B) CTSE expression in IPF alveolar epithelial cells. C) CPA6 expression in control alveolar epithelial cells. D) CPA6 expression in IPF alveolar epithelial cells.

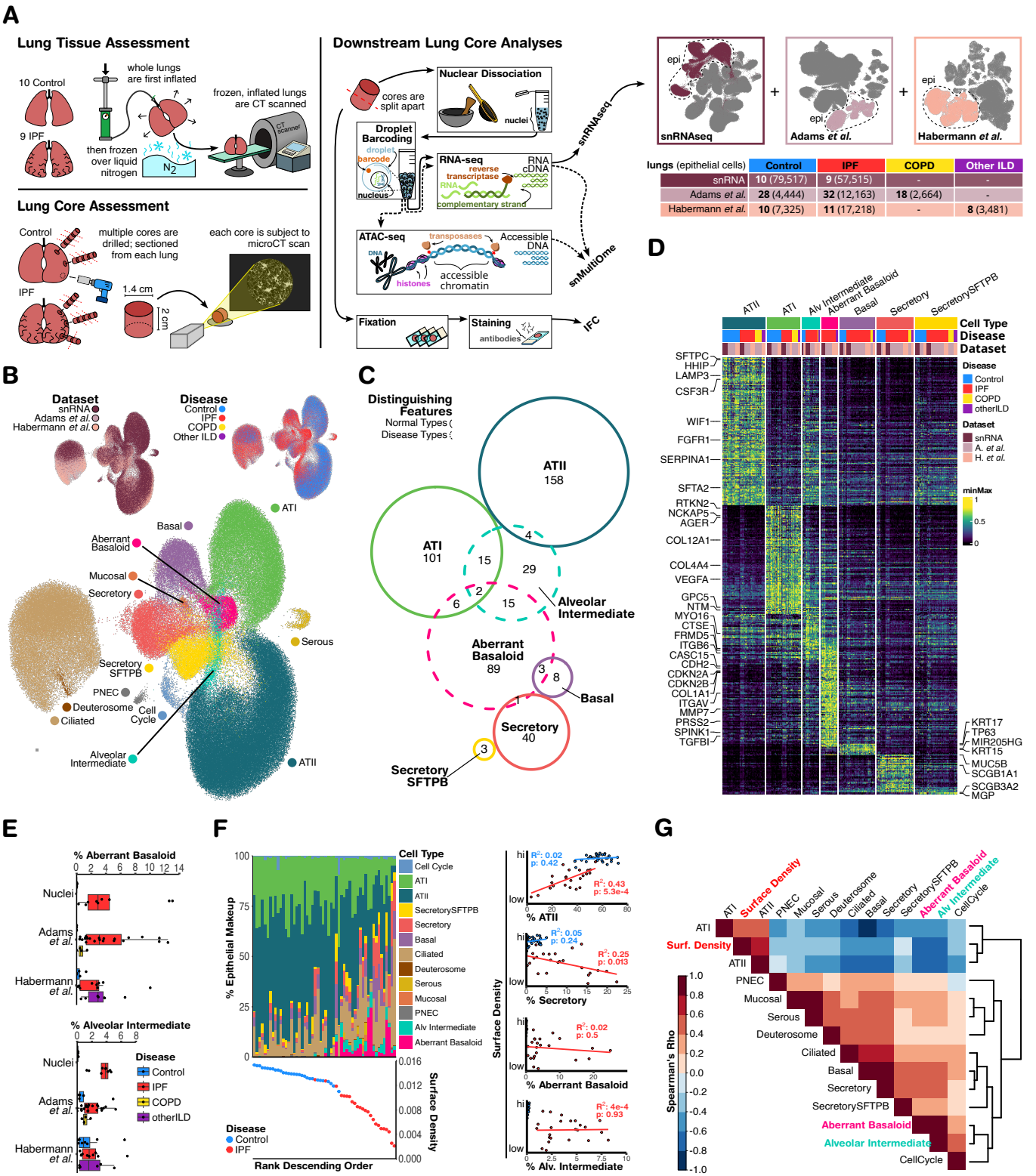
Figure 4: Longitudinal snRNAseq Assessment of PCLS

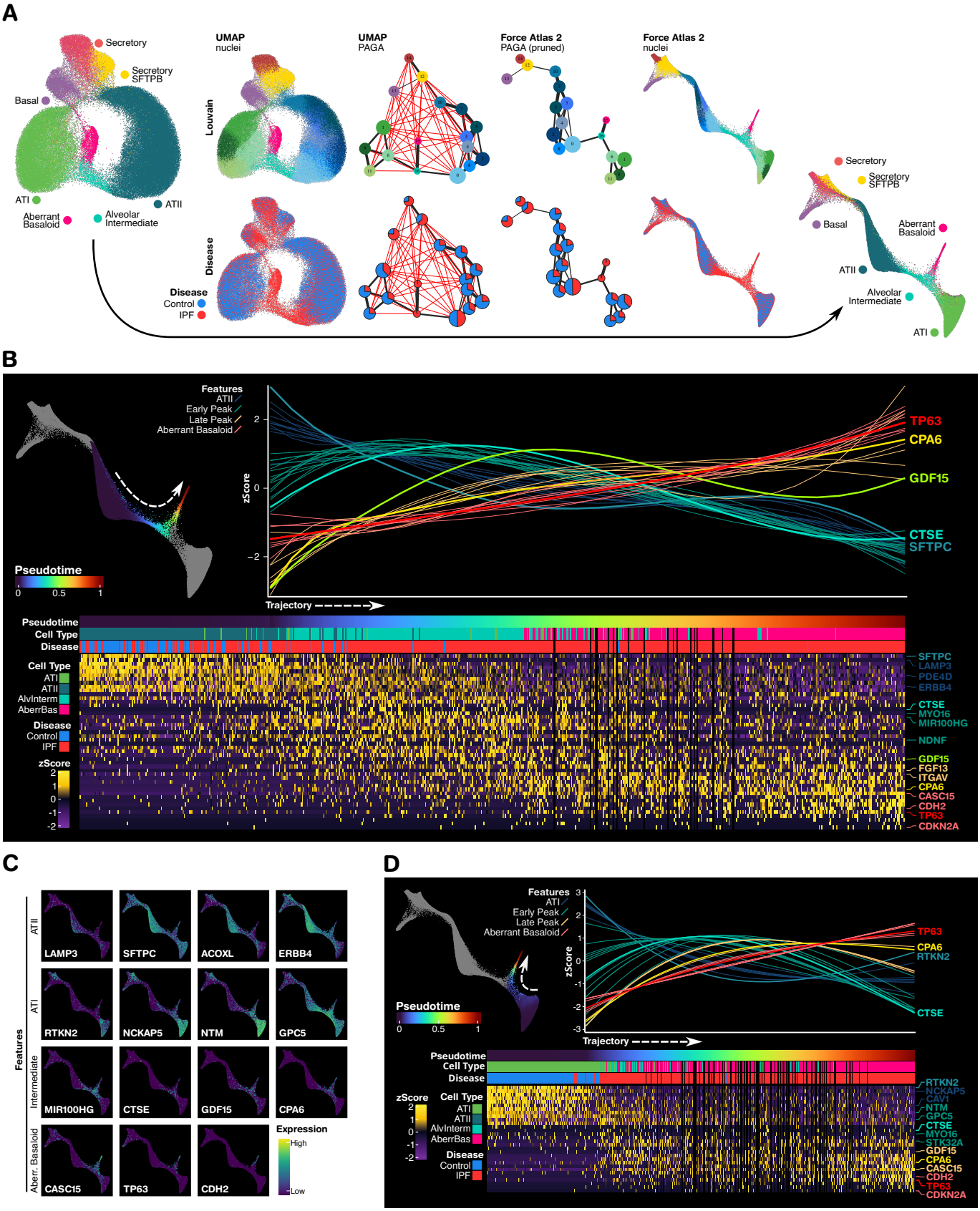
A) Overview of experiment design. B) UMAP of snRNAseq data labeled by condition, timepoint and cell type. C) Upper: Stacked bar plots of cell type composition. Middle: Line plot of cell type makeup changes as a function of time. Lower: Line plot of the percent of mitotic cells per cell type as a function of time. D) UMAP of epithelial cells labeled by condition, timepoint and cell subtype. E) Heatmap of cell subtype markers and corresponding percent makeup values. The average normalized gene expression value for each sample and subtype are normalized between 0 and 1 across samples. The corresponding percent makeup value is shown in a bar graph above. F) Heatmap of intersecting hallmarks between *ex vivo* and *in vivo* cells. G) Above: UMAPs of epithelial cells in cell cycle labeled by cell type, condition or timepoint. Middle) Heatmap of mitosis marker genes (above) and cell type marker genes for cells undergoing mitosis in cell cycle (CC) or not undergoing mitosis (nonCC). Below: Line plot of the proportion of cells found in cell cycle for each epithelial cell type.

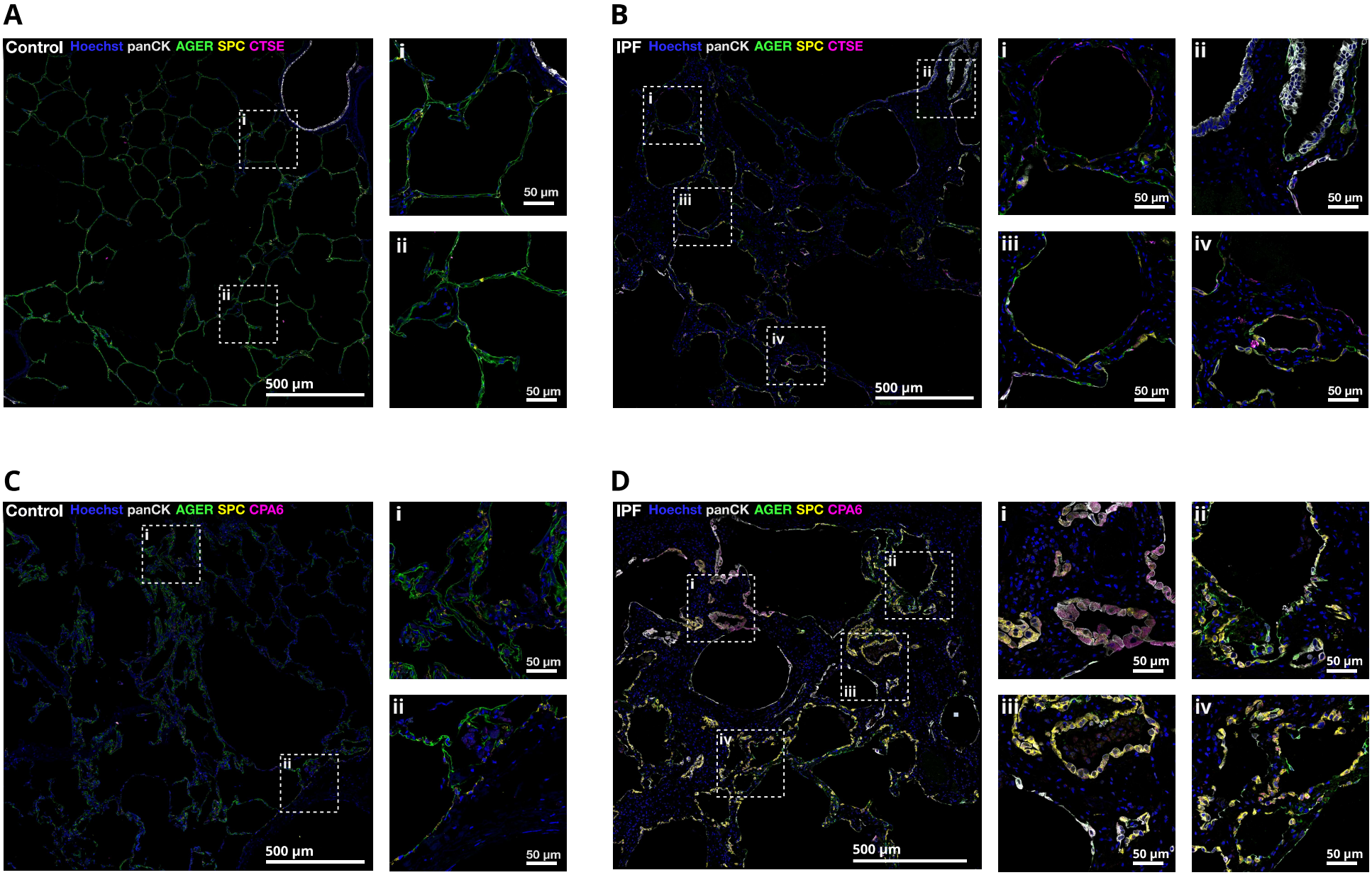
570

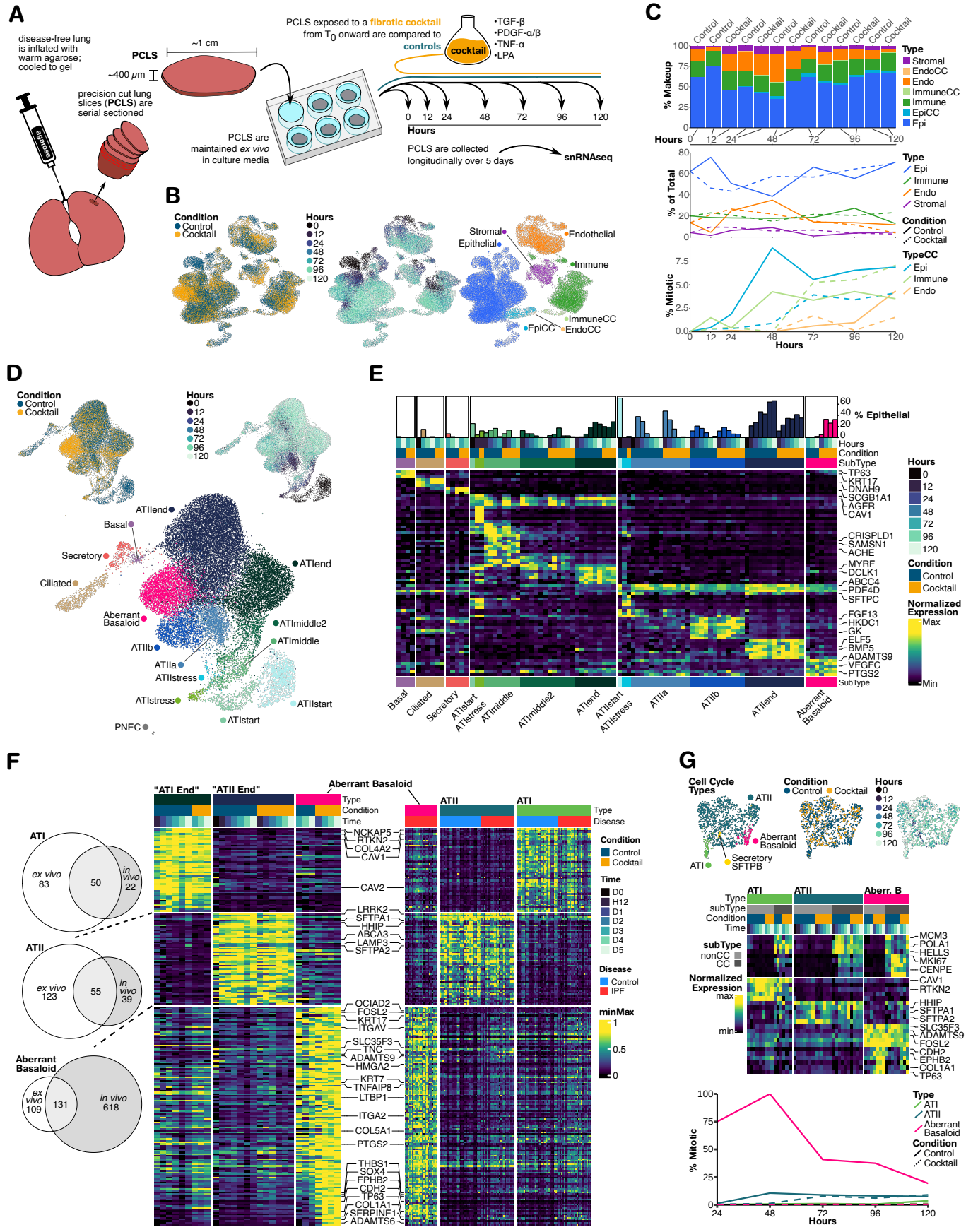
Figure 5: snMultiOme Analysis of IPF and Control Lung Tissue

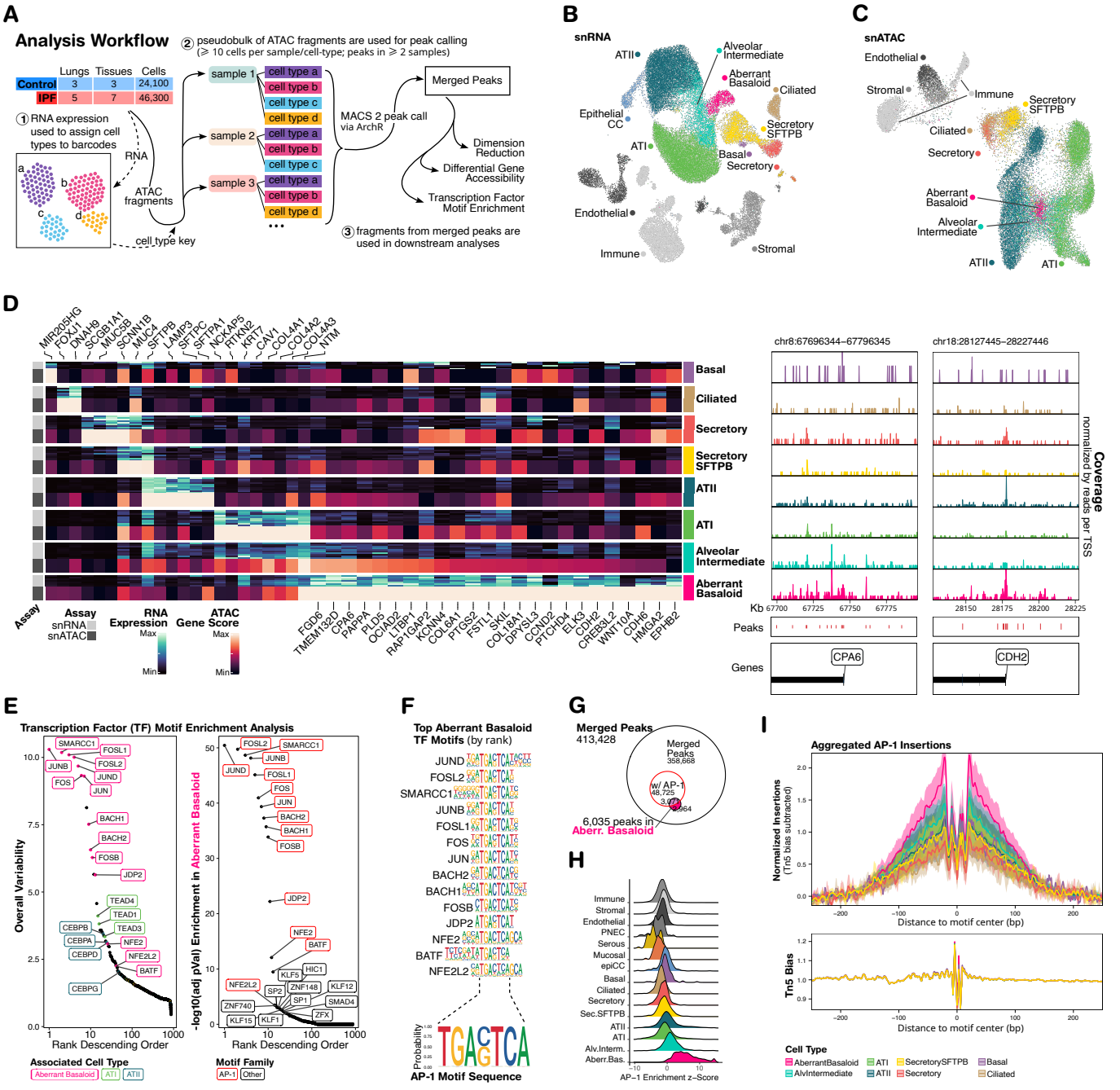
A) Overview of analysis strategy. B) UMAP of snRNAseq data labeled by cell type. C) UMAP of corresponding snATAC-seq data colored by cell type. D) Heatmap of epithelial gene expression markers (blue) with elevated chromatin accessibility geneScores (red). Average gene expression
575 per sample and average cell type geneScore values for each gene are independently normalized between 0 and 1. E) Left: The top transcription factor (TF) motifs ranked by global contributions to variance; TFs associated with a particular cell type are highlighted. Right: The top TF motifs enriched in Aberrant Basaloid cells compared to other cell types. F) Above: Sequence logos of the top TF motifs in Aberrant Basaloid cells. Below: Sequence logo for AP-1 motif. G) Euler plot
580 showing intersections of all merged peaks, merged peaks significantly enriched in Aberrant Basaloid cells and merged peaks containing at least one AP-1 motif sequence. H) Density plots of AP-1 enrichment z-scores per cell, organized by cell type. I) Footprint analysis of AP-1 transcription factor occupation across epithelial cell types.


A
Lung Tissue Assessment
10 Control
9 IPF
whole lungs are first inflated
then frozen over liquid nitrogen
N2
frozen, inflated lungs are CT scanned
CT scanner
Lung Core Assessment
Control
IPF
multiple cores are drilled; sectioned from each lung
1.4 cm
2 cm
each core is subject to microCT scan
Downstream Lung Core Analyses
cores are split apart
Nuclear Dissociation
nuclei
Droplet Barcoding
droplet
barcode
nucleus
RNA-seq
reverse transcriptase
RNA
cDNA
complementary strand
snRNAseq
ATAC-seq
transposases
DNA
histones
accessible chromatin
Accessible DNA
snMultiOme
Fixation
Staining
antibodies
IFC
epi
snRNAseq
Adams et al.
Habermann et al.
lungs (epithelial cells) | Control | IPF | COPD | Other ILD
snRNA | 10 (79,517) | 9 (57,515) | - | -
Adams et al. | 28 (4,444) | 32 (12,163) | 18 (2,664) | -
Habermann et al. | 10 (7,325) | 11 (17,218) | - | 8 (3,481)
B
Dataset
snRNA
Adams et al.
Habermann et al.
Disease
Control
IPF
COPD
Other ILD
ATI
Basal
Aberrant Basaloid
Mucosal
Secretory
Serous
Secretory SFTPB
PNEC
Deuterosome
Ciliated
Cell Cycle
Alveolar Intermediate
ATII
C
Distinguishing Features
Normal Types
Disease Types
ATII
158
ATI
101
4
15
29
2
6
15
Alveolar Intermediate
Aberrant Basaloid
89
3
8
Basal
1
Secretory
40
3
Secretory SFTPB
D
ATII
ATI
Alv Intermediate
Aberrant Basaloid
Basal
Secretory
SecretorySFTPB
Cell Type
Disease
Dataset
SFTPC
HHIP
LAMP3
CSF3R
WIF1
FGFR1
SERPINA1
SFTA2
RTKN2
NCKAP5
AGER
COL12A1
COL4A4
VEGFA
GPC5
NTM
MYO16
CTSE
FRMD5
ITGB6
CASC15
CDH2
CDKN2A
CDKN2B
COL1A1
ITGAV
MMP7
PRSS2
SPINK1
TGFBI
KRT17
TP63
MIR205HG
KRT15
MUC5B
SCGB1A1
SCGB3A2
MGP
Disease
Control
IPF
COPD
otherILD
Dataset
snRNA
A. et al.
H. et al.
minMax
1
0.5
0
E
% Aberrant Basaloid
0 2 4 6 8 10 12 14
Nuclei
Adams et al.
Habermann et al.
% Alveolar Intermediate
0 2 4 6 8
Disease
Control
IPF
COPD
otherILD
F
% Epithelial Makeup
100
75
50
25
0
Cell Type
Cell Cycle
ATI
ATII
SecretorySFTPB
Secretory
Basal
Ciliated
Deuterosome
Serous
Mucosal
PNEC
Alv Intermediate
Aberrant Basaloid
Surface Density
0.016
0.012
0.008
0.004
0.000
Disease
Control
IPF
Rank Descending Order
hi
low
R2: 0.02
p: 0.42
R2: 0.43
p: 5.3e-4
% ATII
R2: 0.05
p: 0.24
R2: 0.25
p: 0.013
% Secretory
R2: 0.02
p: 0.5
% Aberrant Basaloid
R2: 4e-4
p: 0.93
% Alv. Intermediate
G
ATI
Surface Density
ATII
PNEC
Mucosal
Serous
Deuterosome
Ciliated
Basal
Secretory
SecretorySFTPB
Aberrant Basaloid
Alv Intermediate
CellCycle
Surf. Density
Spearman's Rho
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
-1.0


A
Secretory
Secretory SFTPB
Basal
ATI
ATII
Aberrant Basaloid
Alveolar Intermediate
UMAP nuclei
UMAP PAGA
Force Atlas 2 PAGA (pruned)
Force Atlas 2 nuclei
Louvain
Disease
Disease
Control
IPF
B
Features
ATII
Early Peak
Late Peak
Aberrant Basaloid
Pseudotime
0
0.5
1
zScore
Trajectory
TP63
CPA6
GDF15
CTSE
SFTPC
Pseudotime
Cell Type
Disease
Cell Type
ATI
ATII
AlvInterm
AberrBas
Disease
Control
IPF
zScore
SFTPC
LAMP3
PDE4D
ERBB4
CTSE
MYO16
MIR100HG
NDNF
GDF15
FGF13
ITGAV
CPA6
CASC15
CDH2
TP63
CDKN2A
C
Features
ATII
ATI
Intermediate
Aberr. Basaloid
LAMP3
SFTPC
ACOXL
ERBB4
RTKN2
NCKAP5
NTM
GPC5
MIR100HG
CTSE
GDF15
CPA6
CASC15
TP63
CDH2
Expression
High
Low
D
Features
ATI
Early Peak
Late Peak
Aberrant Basaloid
Pseudotime
0
0.5
1
zScore
Trajectory
TP63
CPA6
RTKN2
CTSE
Pseudotime
Cell Type
Disease
zScore
Cell Type
ATI
ATII
AlvInterm
AberrBas
Disease
Control
IPF
RTKN2
NCKAP5
CAV1
NTM
GPC5
CTSE
MYO16
STK32A
GDF15
CPA6
CASC15
CDH2
TP63
CDKN2A


A
Control Hoechst panCK AGER SPC CTSE
i
ii
500 µm
50 µm
B
IPF Hoechst panCK AGER SPC CTSE
i
ii
iii
iv
500 µm
50 µm
C
Control Hoechst panCK AGER SPC CPA6
D
IPF Hoechst panCK AGER SPC CPA6


A
disease-free lung is inflated with warm agarose; cooled to gel
PCLS
~1 cm
~400 μm
precision cut lung slices (PCLS) are serial sectioned
agarose
PCLS exposed to a fibrotic cocktail from T0 onward are compared to controls
cocktail
•TGF-β
•PDGF-α/β
•TNF-α
•LPA
PCLS are maintained ex vivo in culture media
0 12 24 48 72 96 120
Hours
PCLS are collected longitudinally over 5 days
snRNAseq
B
Condition
Control
Cocktail
Hours
0
12
24
48
72
96
120
Stromal
Epithelial
Endothelial
Immune
ImmuneCC
EndoCC
EpiCC
C
Control
Cocktail
% Makeup
Type
Stromal
EndoCC
Endo
ImmuneCC
Immune
EpiCC
Epi
Hours
% of Total
Epi
Immune
Endo
Stromal
Condition
Control
Cocktail
% Mitotic
TypeCC
Epi
Immune
Endo
D
Condition
Control
Cocktail
Hours
0
12
24
48
72
96
120
ATIIend
Basal
Secretory
Ciliated
ATIend
Aberrant Basaloid
ATImiddle2
ATIIb
ATImiddle
ATIIa
ATIIstress
ATIstress
ATIIstart
ATIstart
PNEC
E
% Epithelial
Hours
Condition
SubType
TP63
KRT17
DNAH9
SCGB1A1
AGER
CAV1
CRISPLD1
SAMSN1
ACHE
MYRF
DCLK1
ABCC4
PDE4D
SFTPC
FGF13
HKDC1
GK
ELF5
BMP5
ADAMTS9
VEGFC
PTGS2
Normalized Expression
Max
Min
Basal
Ciliated
Secretory
ATIstart
ATIstress
ATImiddle
ATImiddle2
ATIend
ATIIstart
ATIIstress
ATIIa
ATIIb
ATIIend
Aberrant Basaloid
F
ATI
ex vivo 83
50
in vivo 22
ATII
ex vivo 123
55
in vivo 39
Aberrant Basaloid
ex vivo 109
131
in vivo 618
"ATI End"
"ATII End"
Aberrant Basaloid
ATII
ATI
Type
Condition
Time
Disease
NCKAP5
RTKN2
COL4A2
CAV1
CAV2
LRRK2
SFTPA1
HHIP
ABCA3
LAMP3
SFTPA2
OCIAD2
FOSL2
KRT17
ITGAV
SLC35F3
TNC
ADAMTS9
HMGA2
KRT7
TNFAIP8
LTBP1
ITGA2
COL5A1
PTGS2
THBS1
SOX4
EPHB2
CDH2
TP63
COL1A1
SERPINE1
ADAMTS6
Condition
Control
Cocktail
Time
D0
H12
D1
D2
D3
D4
D5
Disease
Control
IPF
minMax
1
0.5
0
G
Cell Cycle Types
ATII
Aberrant Basaloid
ATI
Secretory
SFTPB
Condition
Control
Cocktail
Hours
0
12
24
48
72
96
120
ATI
ATII
Aberr. B
Type
subType
Condition
Time
subType
nonCC
CC
Normalized Expression
max
min
MCM3
POLA1
HELLS
MKI67
CENPE
CAV1
RTKN2
HHIP
SFTPA1
SFTPA2
SLC35F3
ADAMTS9
FOSL2
CDH2
EPHB2
COL1A1
TP63
% Mitotic
Hours
Type
ATI
ATII
Aberrant Basaloid
Condition
Control
Cocktail


A
Analysis Workflow
② pseudobulk of ATAC fragments are used for peak calling
(≥ 10 cells per sample/cell-type; peaks in ≥ 2 samples)
Lungs Tissues Cells
Control 3 3 24,100
IPF 5 7 46,300
① RNA expression used to assign cell types to barcodes
RNA
ATAC fragments
cell type key
sample 1
sample 2
sample 3
cell type a
cell type b
cell type c
cell type d
MACS 2 peak call via ArchR
Merged Peaks
Dimension Reduction
Differential Gene Accessibility
Transcription Factor Motif Enrichment
③ fragments from merged peaks are used in downstream analyses
B
snRNA
ATII
Alveolar Intermediate
Aberrant Basaloid
Ciliated
Epithelial CC
Secretory SFTPB
ATI
Basal
Secretory
Endothelial
Immune
Stromal
C
snATAC
Endothelial
Stromal
Immune
Secretory SFTPB
Ciliated
Secretory
Aberrant Basaloid
Alveolar Intermediate
ATII
ATI
D
MIR205HG FOXJ1 DNAH9 SCGB1A1 MUC5B SCNN1B MUC4 SFTPB LAMP3 SFTPC SFTPA1 NOXAP5 RTKN2 KRT17 CAV1 COL4A1 COL4A2 COL4A3 NTM
Basal
Ciliated
Secretory
Secretory SFTPB
ATII
ATI
Alveolar Intermediate
Aberrant Basaloid
Assay
snRNA
snATAC
RNA Expression
Max
Min
ATAC Gene Score
Max
Min
FGD6 TMEM132D CPA6 PAPPA PLD5 OCIAD2 LTBP1 RAP1GAP2 KCNN4 COL6A1 PTGS2 FSTL1 SKIL COL18A1 DPYSL3 CCND2 PTCHD4 ELK3 CDH2 CREB3L2 WNT10A CDH6 HMGA2 EPHB2
chr8:67696344–67796345
chr18:28127445–28227446
Coverage
normalized by reads per TSS
Kb 67700 67725 67750 67775
28150 28175 28200 28225
Peaks
Genes
CPA6
CDH2
E
Transcription Factor (TF) Motif Enrichment Analysis
Overall Variability
Rank Descending Order
-log10(adj pVal) Enrichment in Aberrant Basaloid
Rank Descending Order
SMARCC1 FOSL1 FOSL2 JUNB JUND FOS JUN BACH1 BACH2 FOSB JDP2 TEAD4 TEAD1 TEAD3 CEBPB CEBPA NFE2 CEBPD NFE2L2 BATF CEBPG
FOSL2 SMARCC1 JUNB JUND FOSL1 FOS JUN BACH2 BACH1 FOSB JDP2 NFE2 BATF HIC1 KLF5 NFE2L2 SP2 ZNF148 KLF12 SP1 SMAD4 ZNF740 ZFX KLF15 KLF1
Associated Cell Type
Aberrant Basaloid ATI ATII
Motif Family
AP-1 Other
F
Top Aberrant Basaloid TF Motifs (by rank)
JUND
FOSL2
SMARCC1
JUNB
FOSL1
FOS
JUN
BACH2
BACH1
FOSB
JDP2
NFE2
BATF
NFE2L2
Probability
AP-1 Motif Sequence
G
Merged Peaks
413,428
Merged Peaks
358,668
w/ AP-1
48,725
6,035 peaks in
Aberr. Basaloid
H
Immune
Stromal
Endothelial
PNEC
Serous
Mucosal
epiCC
Basal
Ciliated
Secretory
Sec.SFTPB
ATII
ATI
Alv.Interm.
Aberr.Bas.
AP−1 Enrichment z−Score
I
Aggregated AP-1 Insertions
Normalized Insertions
(Tn5 bias subtracted)
Distance to motif center (bp)
Tn5 Bias
Distance to motif center (bp)
Cell Type
AberrantBasaloid ATI SecretorySFTPB Basal
AlvIntermediate ATII Secretory Ciliated