Article

Mapping spatially resolved transcriptomes in human and mouse pulmonary fibrosis

Check for updates

Lovisa Franzén [1,2,10], Martina Olsson Lindvall[1,10], Michael Hühn[3], Victoria Ptasinski [1], Laura Setyo[4], Benjamin P. Keith[5], Astrid Collin[6], Steven Oag[6], Thomas Volckaert[7], Annika Borde[7], Joakim Lundeberg [2], Julia Lindgren [8], Graham Belfield[8], Sonya Jackson[9], Anna Ollerstam[1], Marianna Stamou [1]✉, Patrik L. Ståhl [2]✉ & Jorrit J. Hornberg[1]

Idiopathic pulmonary fibrosis (IPF) is a progressive lung disease with poor prognosis and limited treatment options. Efforts to identify effective treatments are thwarted by limited understanding of IPF pathogenesis and poor translatability of available preclinical models. Here we generated spatially resolved transcriptome maps of human IPF ($n = 4$) and bleomycin-induced mouse pulmonary fibrosis ($n = 6$) to address these limitations. We uncovered distinct fibrotic niches in the IPF lung, characterized by aberrant alveolar epithelial cells in a microenvironment dominated by transforming growth factor beta signaling alongside predicted regulators, such as TP53 and APOE. We also identified a clear divergence between the arrested alveolar regeneration in the IPF fibrotic niches and the active tissue repair in the acutely fibrotic mouse lung. Our study offers in-depth insights into the IPF transcriptional landscape and proposes alveolar regeneration as a promising therapeutic strategy for IPF.

Idiopathic pulmonary fibrosis (IPF) is a chronic lung disease characterized by progressive and irreversible scarring of the lung. Treatment options are limited, and the development of new therapies is impeded by incomplete understanding of disease pathogenesis and translatability limitations of available preclinical models. Recent advances into mechanistic understanding of IPF pathogenesis reveal complex gene-environment interactions as key pathophysiological drivers[1–3].

Single-cell studies have revealed IPF-associated cell states, including atypical epithelial cells, fibroblasts[4,5] and profibrotic alveolar macrophages[6,7]. Interestingly, a novel KRT5⁻/KRT17⁺ aberrant basaloid (AbBa) epithelial cell population has been independently identified

in multiple studies[4,5,8,9], expressing epithelial, basal and mesenchymal markers and genes related to senescence and extracellular matrix (ECM) production. These cells probably originate from alveolar type 2 (AT2) or club cells[4,5,9,10], but their role in the fibrotic microenvironment remains elusive. A closely related Krt8⁺ alveolar differentiation intermediate (ADI) cell population is present in the widely used mouse model of bleomycin (BLM)-induced pulmonary fibrosis[11], which, in contrast to the IPF lung, features relatively rapid inflammatory onset, epithelial regeneration and fibrosis resolution[12].

Although recent single-cell RNA sequencing (scRNA-seq) studies have substantially advanced our understanding of the IPF lung

[1]Safety Sciences, Clinical Pharmacology and Safety Sciences, R&D, AstraZeneca, Gothenburg, Sweden. [2]Department of Gene Technology, KTH Royal Institute of Technology, Science for Life Laboratory, Stockholm, Sweden. [3]Translational Science and Experimental Medicine, Research and Early Development, Respiratory and Immunology, BioPharmaceuticals R&D, AstraZeneca, Gothenburg, Sweden. [4]Pathology, Clinical Pharmacology and Safety Sciences, R&D, AstraZeneca, Cambridge, UK. [5]Quantitative Biology, Discovery Sciences, R&D, AstraZeneca, Gothenburg, Sweden. [6]Animal Science and Technology, Clinical Pharmacology and Safety Sciences, R&D, AstraZeneca, Gothenburg, Sweden. [7]Bioscience In Vivo, Research and Early Development, Respiratory and Immunology, BioPharmaceuticals R&D, AstraZeneca, Gothenburg, Sweden. [8]Translational Genomics, Centre for Genomics Research, Discovery Sciences, R&D, AstraZeneca, Gothenburg, Sweden. [9]Late-Stage Development, Respiratory and Immunology, BioPharmaceuticals R&D, AstraZeneca, Gothenburg, Sweden. [10]These authors contributed equally: Lovisa Franzén, Martina Olsson Lindvall. ✉e-mail: marianna.stamou@astrazeneca.com; patrik.stahl@scilifelab.se

cellular composition[4–7,13–15], they lack insights into tissue architecture and cellular interplay in a spatial context. Spatially resolved transcriptomics (SRT) enables RNA profiling of intact tissue[16–19] and can illuminate dynamic cellular interactions in the lung[20–22]. However, a transcriptome-wide map of extensive areas of the fibrotic lung is currently missing.

We applied SRT to map the fibrotic lung in human IPF and the BLM mouse model. SRT was integrated with scRNA-seq data to characterize the AbBa cell microenvironment and delineate the dynamic crosstalk between alveolar epithelial cells, myofibroblasts, fibroblasts and profibrotic macrophages. These spatial atlases broaden our understanding of the IPF cellular interplay and unveil key convergent and divergent pathways in human IPF and the BLM mouse model.

Results

Spatial transcriptomics of healthy and IPF human lungs

We generated transcriptome-wide spatial profiles of freshly frozen human lung resection samples from four patients with IPF (IPF 1–4, collected during lung transplantation) and four subjects with no known lung disease (healthy controls (HCs) 1–4 and 'B0', collected postmortem) using the Visium Spatial Gene Expression platform (Fig. 1a,b and Supplementary Table 1). For each patient with IPF, three tissue blocks ('B1', 'B2' and 'B3') reflecting increasing extent of fibrotic injury within the same donor were selected (Fig. 1a and Extended Data Fig. 1a). The sample selection and study design are further described in Methods and Supplementary Note 1.

A total of 255–444 million sequencing reads (average 349 million) per sample were generated (Supplementary Table 1). We analyzed an average of ~4,000 spots per tissue section (each spot representing a transcriptome of the mixture of cells in the tissue that covers the spot), capturing an average of >1,500 unique genes per spot (Fig. 1c). We observed a higher average number of genes per spot and transcript count levels in IPF samples compared with HC, probably due to disease-associated differences in cellular density between conditions. A pseudobulk differential expression analysis (DEA) between IPF and HC samples identified 1,469 differentially expressed genes (DEGs) (Fig. 1d), including genes associated with fibroblasts, previously reported to be upregulated in IPF (*FNDC1*, *COL10A1* and *THY1*)[23], as well as matrix metalloproteinases[24] and genes involved in IPF-associated signaling pathways (*SFRP2*, *WNT10A* and *TGFBI*)[25,26] (Supplementary Table 2). While several upregulated genes in the IPF samples mapped to areas of remodeled fibrotic tissue, alterations were also seen in minimally remodeled alveolar tissue (Extended Data Fig. 1b). These areas were further examined by comparing annotated 'alveolar' or 'fibrotic' regions within the IPF samples to alveolar regions in the HC samples (Extended Data Fig. 2a,b and Supplementary Note 2). A high number of DEGs were found in fibrotic regions compared with IPF alveolar regions, with 95% of upregulated DEGs in IPF alveolar areas also found in fibrotic regions (Extended Data Fig. 2c and Supplementary Table 3). Among these, many DEGs displayed a relationship between expression and proximity to the fibrotic-alveolar interface (Extended Data Fig. 2d and Supplementary Table 3). At the fibrotic border, multiple chemokine ligands (*CCL18*, *CX3CL1*, *CXCL10* and *CXCL11*) displayed peaked expression that decreased into alveolar regions, suggesting recruitment of immune cells (for example, profibrotic macrophages) to the leading edge of fibrosis. Conversely, ECM-associated genes (*COL1A1*, *COL1A2* and *LUM*) had a decreased expression at the border, while expression was higher in the fibrotic and adjacent alveolar tissue (Supplementary Note 2).

Data deconvolution identifies structures and cell types.

The data were deconvoluted into 30 'factors' using non-negative matrix factorization (NMF)[27] (Supplementary Table 4, describing associated genes, pathways and cell types). Each factor was driven by expression of covarying genes with little overlap observed between the top contributing genes for most factors (Supplementary Fig. 1b,c). These factors revealed signatures of distinct cell types and structures including mixed bronchiolar epithelial cell types (factor 1 (F1)), smooth muscle cells (F10) and plasma and B cells (F6) (Fig. 1e). The spatial distribution of cell-type densities was further inferred by integration[28] with an IPF-derived scRNA-seq dataset[5] (referred to as 'Habermann[5]'). This revealed a distinct group (F1 and F21) that correlated with ciliated airway cell types, including basal cells, club cells, ciliated cells and *MUC5B*+ cells (Fig. 1f), in line with spatial mapping of F1 activity to bronchial epithelium. Other factors correlated specifically with the alveolar compartment (Fig. 1g), including alveolar macrophages (spatial overlap with F5), alveolar type 1 (AT1) cells (spatial overlap with F12 and annotated alveolar tissue) and AT2 cells. An additional group of factors corresponded to immune cells and stromal components, including lymphocytes, endothelial cells and fibroblasts (spatial overlap with F4 and areas labeled as fibrous or remodeled tissue). Several factors could not be clearly attributed to specific cell types or groups, probably representing a more complex mixture of cells, cell types not annotated in the reference dataset, and/or uncharacterized cell states. This included F16 in the alveolar compartment of HC and IPF lungs (Fig. 1f), dominated by prostaglandin signaling genes and AT1, AT2 and fibroblast markers (Supplementary Table 4).

Factor activity reveals pathways and cellular interactions.

Further examination of factor distribution across samples revealed 11 factors that were more prevalent in IPF compared with HC (Fig. 2a and Extended Data Fig. 2e). These factors are associated with important IPF cell morphologies and processes, including ECM-related pathways, and overlapped regions of fibrosis (F4 and F14) or IPF 'honeycomb' formations (F5 and F21) (Fig. 2b). F5 displayed markers of dendritic cells and macrophages, while F21 presented a *MUC5B*-expressing airway epithelial signature (Supplementary Table 4). The F21 profile might reflect a previously identified *MUC5B*+, *BPIFB1*+ and *SCGB3A1*+ IPF-associated cell population[29]. In line with our SRT data, MUC5B expression has previously been localized to honeycomb cysts, and *MUC5B* polymorphisms have been linked to IPF risk[30]. F9 was unique to IPF lungs and characterized by genes related to oxidative stress, inflammation, ECM remodeling and vascular changes. It was moreover particularly present in IPF samples with less dense fibrosis ('B1'), suggesting involvement in early-stage fibrotic processes (Supplementary Note 3).

Among the ECM/fibrosis-related factors, both F4 and F14 were notably active in IPF tissues, probably reflecting pathological remodeling (Fig. 2c and Supplementary Note 3). In addition to collagens and fibrosis-related genes, the F14 signature also included keratins such as *KRT7* and *KRT8* (Fig. 2d). F14 activity correlated with inferred cell-type densities of *KRT5*−/*KRT17*+ AbBa cells, myofibroblasts and the recently described *HAS1*-hi fibroblast subtype[5], specifically in the IPF samples (Fig. 2e). One lung (IPF donor 2) demonstrated a less pronounced correlation between F14 activity and the *KRT5*−/*KRT17*+ AbBa cell type. This deviation may stem from donor heterogeneity in cellular composition of the studied samples or more general variation in disease manifestations between patients.

Visual inspection confirmed that F14-positive spots coincided with the correlated cell types and revealed that F14 activity spatially aligned with fibroblastic foci (FF) (Fig. 2f), a histological feature of active tissue remodeling[21,31,32]. In many instances, spots with elevated *KRT5*−/*KRT17*+ AbBa cell densities appeared to be situated along the FF borders, confirming the previously proposed positioning of these cells within the fibrotic human lung[4]. Importantly, our NMF approach thus identified a signature encompassing the *KRT5*−/*KRT17*+ AbBa cell type independently of scRNA-seq data, placed in its spatial histological context across IPF samples.

Characterization of the AbBa niche in IPF

To better understand the cell-type heterogeneity in F14, we isolated its most active spots (denoted F14hi) and identified five distinct
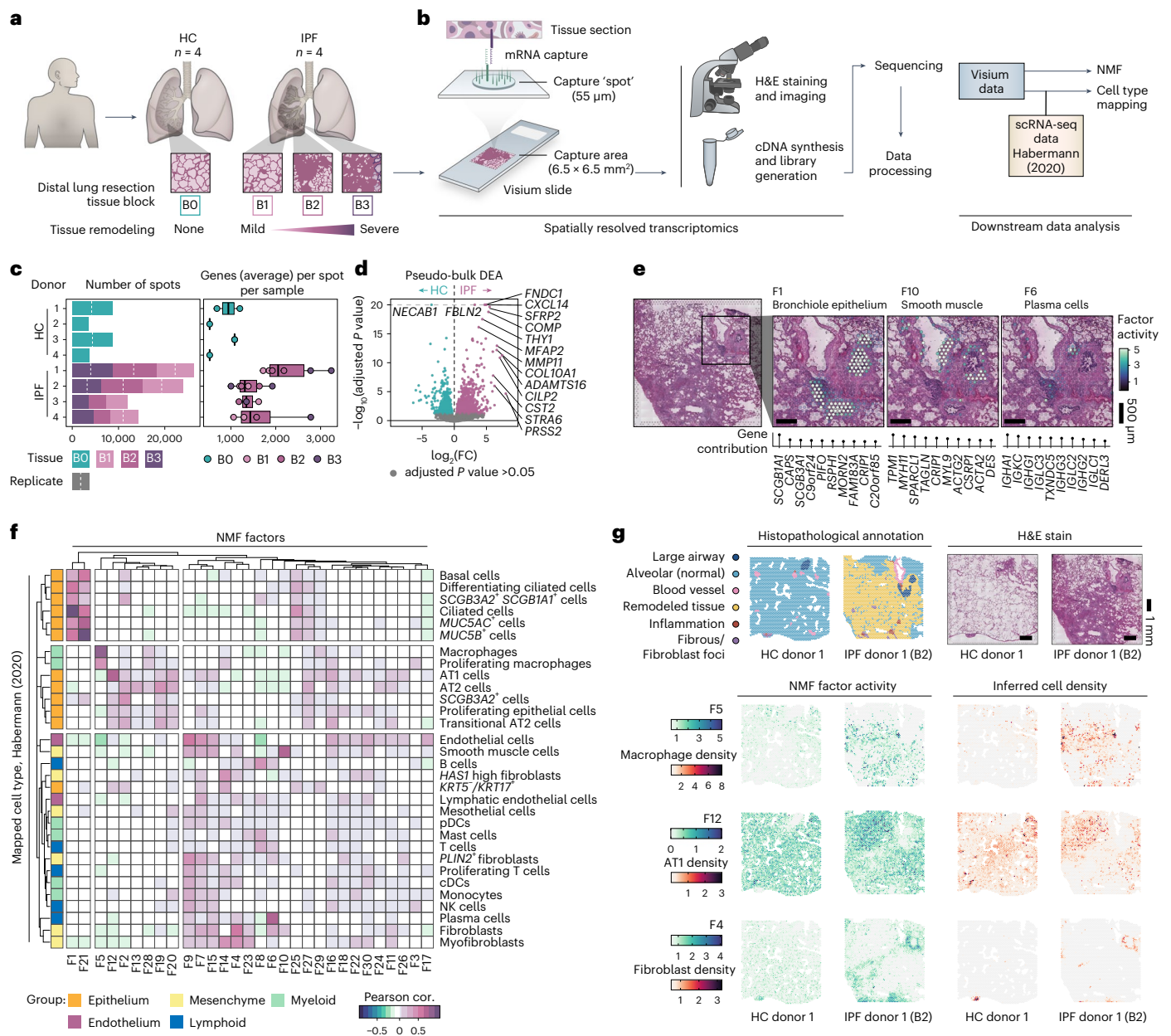
Fig. 1 | Spatial transcriptomic profiling of human pulmonary fibrosis.
a, Tissue sections from distal lung resection samples from HCs (B0; $n = 4$) and patients with IPF ($n = 4$), were sectioned and analyzed using the Visium Spatial Gene Expression technology. Three tissue blocks exhibiting progressive tissue remodeling (B1–3) were selected from each IPF donor. **b**, Schematic illustration of the Visium workflow and subsequent data processing steps. NMF was used for dimensionality reduction, generating 30 distinct factors. Cell-type distributions were inferred through integration with a scRNA-seq dataset published by Habermann et al.[5] (2020; GSE135893). **c**, Summarizing descriptions of the data, including the number of Visium capture spots per sample and the average number of unique genes detected per spot and sample. The box plot center line is the median, the box limits are the upper and lower quartiles, the whiskers are 1.5× interquartile range and the points are the value per Visium sample. **d**, Pseudobulk

DEA comparing pooled HC and IPF Visium data per donor to identify significant DEGs between conditions based on data from entire tissue sections. **e**, Spatial distribution maps for selected NMF factors that correspond histological and/or transcriptional profiles of bronchiole epithelium (F1), smooth muscle (F10) and plasma cells (F6). **f**, Pearson correlation (cor.) heat map of NMF factor activity and inferred cell-type densities, using the Habermann et al. scRNA-seq dataset, across all spots in all samples. **g**, Example of histopathological annotations performed on sections from each HC and IPF tissue block based on the H&E-stained Visium sections. Visualizing spatial NMF activity and inferred cell-type densities showcases the colocalization of highly correlated factor-cell pairs. NMF, non-negative matrix factorization; pDC/cDCs, plasmacytoid/classical dendritic cells; NK cells, natural killer cells.

subclusters, F14^hi C0–C4 (Fig. 3a and Extended Data Fig. 3a). Defining genes of C0 corresponded to markers of the *KRT5⁻/KRT17⁺* AbBa cell type (for example, *PRSS2* and *KRT7*)[5], characteristically devoid of the basal cell marker *KRT5*. The remaining four F14^hi clusters expressed genes corresponding to fibroblasts and myofibroblasts (C1 and C2), macrophages (C3) and basal and secretory airway epithelial cells (C4). Based on the marker gene profiles, C1 and C2 appeared to represent

fibrotic populations with distinct roles. C1 displayed a matrix deposition and scar formation profile whereas C2 had markers indicative of stress responses (metallothioneins), immune modulation (*CCL2* and *FCN3*) and vascular interactions (*ENG* and *THBD*), probably reflecting diverse fibroblast phenotypes within the fibrotic niche of the IPF lungs. F14^hi C0 was markedly present in three out of four IPF donors (lowly abundant in donor 2), consistent with AbBa cell detection rates from
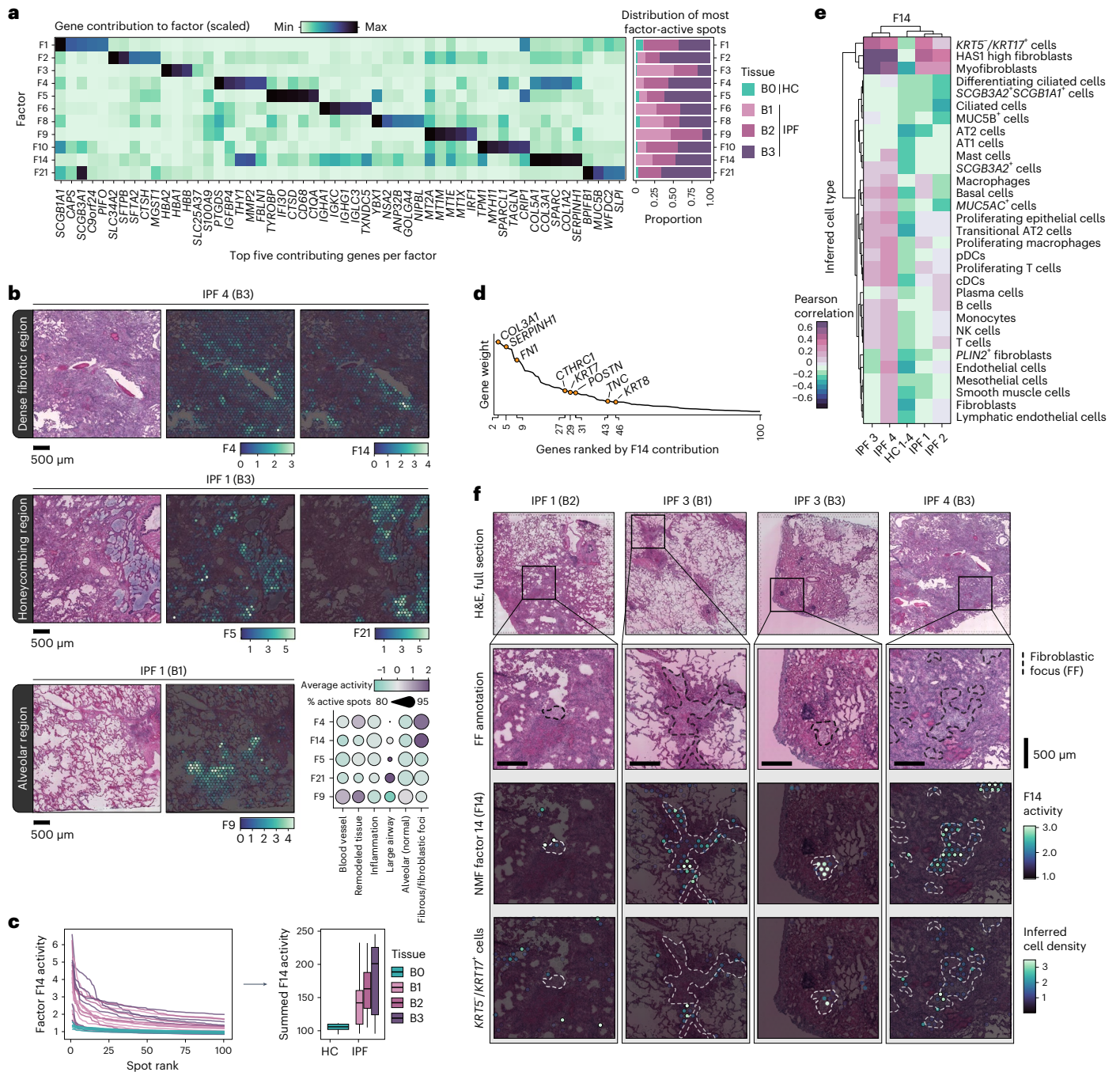
Fig. 2 | Disease-associated signatures revealed by NMF. a, NMF identified signatures over-represented in IPF tissue. Their relative contribution to each factor (scaled) is displayed for the top five contributing genes per factor, and the proportion of spots with the highest activity (99th percentile) by condition and tissue severity grade. **b**, Spatial representation of selected NMF factors across IPF lung sections, demonstrating distinct localization patterns that was observed across all IPF samples. F4 and F14 marked heavily fibrotic regions, F5 and F21 associated with honeycombing structures and F9 were seen in alveolar regions. The dot plot displays the average activity (scaled and centered) and detection rate (percentage of spots with increased activity) within the annotated histological regions across all tissue blocks. **c**, Activity profile of the top 100 ranked spots per sample based on F14 activity (line chart), highlighting a

consistent distinction between HC and IPF tissue samples, further summarized by summing the F14 activity levels and grouped based on tissue remodeling extent (B0, $n = 6$; B1, $n = 7$; B2, $n = 6$; B3, $n = 6$) (box plot). The center line is the median, the box limits are the upper and lower quartiles and the whiskers are 1.5× interquartile range. **d**, Ranking of the top 100 genes for F14 based on gene weight (contribution) to the factor, with keratins, collagens and other fibrosis-related genes emphasized. **e**, Correlation heat map between F14 activity and densities of inferred cell types within spatial spots, capturing potential colocalization of F14 and cell types (strong correlation suggests spatial co-occurrence). **f**, Visualization of FF annotations, F14 activity and the distribution of inferred $KRT5^-/KRT17^+$ cells in selected samples and regions in all IPF donor lungs.

prior scRNA-seq studies[4,5] (Supplementary Note 4 and Supplementary Fig. 2). Upon visual inspection, C0 spots were often seen to colocalize with edges of FF (Fig. 3b and Extended Data Fig. 3b). Histopathological annotations of FF-overlapping spots were used to compute the

overrepresentation of each F14^{hi} cluster at a distance to the FF borders (Extended Data Fig. 3c,d). This revealed significant enrichment of cluster C0 around the edges and within the FF (Extended Data Fig. 3e). Given the multicellular resolution of Visium, it is challenging to capture the

precise lining of FF; however, our results support previous reports of AbBa cell localization at the FF periphery[4,5]. In comparison, the cluster marked by fibroblast genes (C1) was highly concentrated within the FF.

Crosstalk and signaling in the AbBa microenvironment

We found a higher abundance of AT2 cells and transitional AT2 cells around the F14[hi]C0 AbBa niche, compared with more distant regions (Fig. 3c). Notably, the density of transitional AT2 cells peaked slightly closer to the AbBa niche than AT2 cells, suggesting a possible differentiation lineage from AT2 to transitional AT2 cells and toward AbBa cells, consistent with a previously proposed cell trajectory[5], captured in space. Additionally, the proximity of SCGB3A2[+] secretory cells to F14[hi]C0 spots aligns with previous findings suggesting them as another potential source for AbBa cells[5,29].

We observed a decline in matrix remodeling and fibrosis-associated genes (for example, MMP11, POSTN and COL1A2) with increasing distance from F14[hi]C0 (Fig. 3d), indicating elevated fibrotic activity around AbBa cells. Conversely, genes linked to alveolar function and immune response (for example, SFTPA2, SFTPC and SLPI) showed lower expression within C0 compared with its immediate surroundings. A group of immunoglobulin-related genes (for example, IGLC1, IGKC and PIGR) were expressed near AbBa cell-dense areas but not within, implying a differential immune response or possible exclusion of certain immune elements from the AbBa microenvironment.

Analysis of areas neighboring (n.b.) the F14[hi]C0 spots identified subclusters containing alveolar epithelial cells (n.b. cluster 0), fibroblasts/myofibroblasts (n.b. cluster 1), alveolar macrophages (n.b. cluster 2) and plasma cells (n.b. cluster 3) (Fig. 3e), allowing us to study regulatory molecules and signaling within and between clusters. Upstream regulator and pathway enrichment analyses predicted upstream activation in F14[hi]C0 and n.b. cluster 1 of molecules (including TGF-β1, TGF-β2, MRTFB, TEAD1-4 and ISG15) known to be involved in fibrosis (Fig. 3f and Supplementary Table 5). The canonical profibrotic cytokine transforming growth factor beta (TGF-β) (encoded by TGF-β1, TGF-β2 and TGF-β3) plays an important role in IPF[25,33] and has been implicated in ADI cell formation and inhibition of differentiation toward AT1 cells[34]. MRTFB regulates myofibroblast differentiation[35], while TEAD family members (part of YAP/TAZ coactivator complex) are key effectors of profibrotic pathways including Hippo, TGF-β and WNT signaling[36–38], implicated in tissue regeneration and fibrosis[26,39]. ISG15, a modulator of p53, has been implicated in age-related signaling pathways[40]. Enrichment of IPF-, glycoprotein VI (GP6)- and wound-healing signaling pathways, along with pathways associated with cell movement and migration, further supports an active fibrogenic node.

Prediction analysis of ligand–target interaction[41], with directional information preserved (Methods), identified further cell–cell communications within the F14[hi]C0 microenvironment, including TGFB1, IL1B and SFRP2 (Fig. 3g and Supplementary Table 5). SFRP2 (a WNT signaling modulator) expression distinctly originated from the neighboring fibroblast cluster, implicating potential autocrine and paracrine WNT signaling between (myo)fibroblasts, alveolar epithelial and AbBa cells. The predicted ligand apolipoprotein E (encoded by APOE) was highly expressed in the macrophage cluster, alluding to a monocyte-derived and M2-like profile of the neighboring macrophage population[42,43]. By analyzing the expression of all annotated APOE receptors across the F14[hi]C0 region distance, we identified an inverse expression pattern between APOE and several of its receptors (Fig. 3h). Glypican 1 (encoded by GPC1), LDL receptor-related protein 1 (LRP1) and syndecan 1 (SDC1) were more highly expressed within, and in close proximity to, the F14[hi]C0 region. These observations suggest a potentially under-recognized role for apolipoprotein signaling within the AbBa cell fibrotic niche in IPF.

Spatial transcriptomics in a mouse model of pulmonary fibrosis

The BLM mouse model is the most established preclinical model for pulmonary fibrosis[44,45]. To increase understanding of the translational predictivity of this model for human IPF, we generated SRT data from mouse lung samples collected at day 7 (d7) and day 21 (d21) following single-dose BLM or saline (vehicle) administration (Fig. 4a, Extended Data Fig. 4a and Supplementary Table 6). These time points were selected to cover the acute inflammatory and early fibrogenic phase (d7), as well as the established fibrotic stage (d21) (Supplementary Notes 5 and 7).

A total of 151–571 million sequencing reads (average 325 million) per sample were generated (Supplementary Table 6). Healthy alveolar regions accounted for 80–90% and 30–50% of the total number of spots in saline and BLM-challenged lungs, respectively. Remaining spots in the BLM-challenged samples were labeled as areas of tissue damage or remodeling (Extended Data Fig. 4b). A pseudobulk DEA between BLM and vehicle controls identified 3214 and 3787 DEGs at d7 and d21, respectively (Extended Data Fig. 4c).

Comparative analysis of gene expression and cellular composition of pulmonary fibrosis in mice. We identified DEGs in annotated fibrotic areas compared with control samples in the mouse model and analyzed their overlap with DEGs in human IPF (Fig. 4b and Supplementary Table 7). Numerous DEGs overlapped between mouse and human (178 between IPF and d7 BLM fibrotic regions and 93 between IPF and d21 BLM), with eight DEGs displaying contrasting fold-change directionality. Among the latter, most are involved in ECM organization (COL17A1 (ref. 46)), inflammatory signaling (CX3CL1 (ref. 47)) and apoptosis regulation and cellular adhesion (S100A14 (ref. 48) and FAIM2 (ref. 49)). While these genes may play a role in fibrosis in both conditions, the inverse expression patterns suggest divergent roles in human IPF and the mouse BLM model.

Cell-type deconvolution was performed using a lung scRNA-seq dataset generated in the BLM mouse model[11] (referred to as 'Strunz'[11]). Spatial visualization of cell-type densities demonstrated accurate mapping to relevant tissue regions, where alveolar epithelial cells were inferred in healthy alveolar tissue (Fig. 4c). Pronounced differences in cell-type densities were observed between BLM and vehicle groups, including for resolution (M2-polarized) macrophages and Krt8[+] ADI cells (Fig. 4d). AT2 cell abundance decreased at d7 but showed recovery by d21. The apparent influx of recruited (proinflammatory) macrophages at d7 normalized by d21, confirming resolution of acute inflammation.

Fig. 3 | Cellular and molecular deconvolution of the AbBa niche. a, Clustering of the 99th percentile of F14 active spots (F14[hi]; yellow shaded area) into five clusters (F14[hi]C0–4), visualized in the UMAP space with the top gene markers listed and expression of AbBa cell markers PRSS2 and KRT17 and basal cell marker KRT5. The box plot center line is the median, the box limits are the upper and lower quartiles and the whiskers are 1.5× interquartile range. **b,** Spatial location of F14[hi] clusters within FF in representative IPF tissue sections. **c,** Inferred cell-type densities around F14[hi]C0 boundaries (dashed line), with radial distance. The distance below zero indicates F14[hi]C0 positive spots (shaded area). **d,** Pearson correlation (cor.) and gene expression (expr.) of top 20 genes (positive and negative) across a 0–500 μm distance (dist.) radius from the F14[hi]C0 border. **e,** Clustering of neighboring F14[hi]C0 spots (~200 μm) into six clusters (n.b. cluster 0–5) with average expression (avg. expr.) of top marker genes and selected average inferred cell-type densities highlighted in the dot plot. The bar chart displays distribution per donor. **f,** Enrichment analysis (IPA) of n.b. clusters 0–2 using marker genes (adjusted P < 0.05). The heat maps display activation Z-scores of top predicted upstream regulators enriched pathways and diseases and functions. **g,** NicheNet analysis of cell–cell communication showing prioritized ligands acting upon F14[hi]C0 and n.b. clusters 0–2 regions (left) with their average expression (avg. expr.) levels (right). **h,** Average expression of APOE and its receptors across F14[hi]C0 and over radial distance (three spots, ~300 μm) and background expression in remaining spots ('rest'). The directional arrows show correlation (Pearson) trends across spot distance. Prolif. T cells, proliferating T cells; Sign. pathway, signaling pathway.

Dynamic lung tissue remodeling in response to BLM in mice. Colocalization analysis revealed dynamic spatial compartmentalization of cell types within spots, capturing the spatiotemporal dynamics of fibrogenesis and indicating lung tissue remodeling in response to BLM injury (Fig. 4e). In vehicle control lungs, we identified two compartments consisting of bronchial epithelial (A) and alveolar (B) tissue,

outlining the uninjured lung architecture. In the d7 BLM-challenged lungs, prominent cell densities consisted of bronchial epithelium (C), alveolar epithelium and alveolar capillary endothelium (D) and remodeled alveolar tissue marked by fibroblasts and myofibroblasts (E). At d21, the cellular composition of the compartments was altered, so that in addition to bronchial epithelium (F) and fibrotic, remodeled alveolar
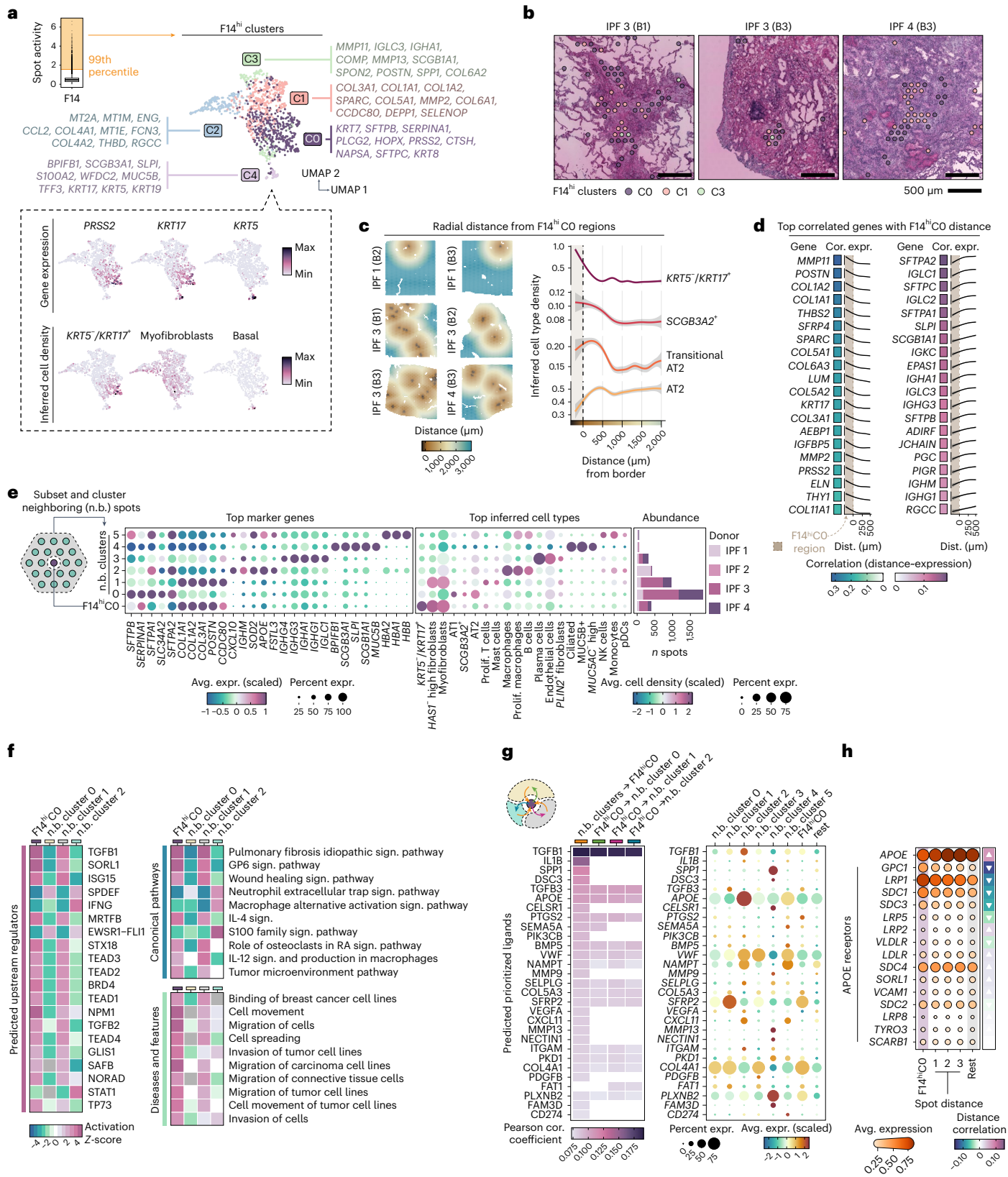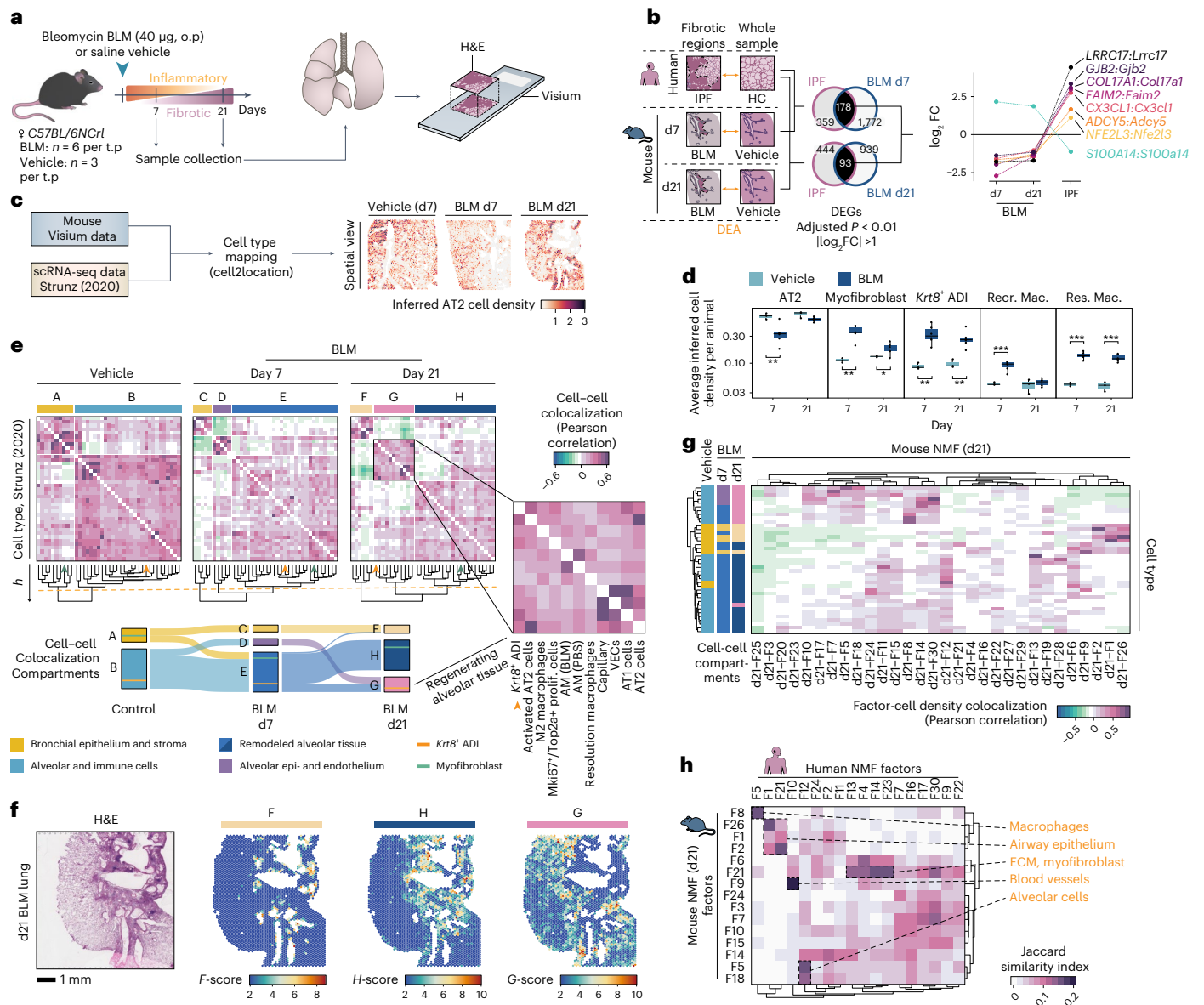
Fig. 4 | Comparative spatial analysis of pulmonary fibrosis in mouse and human. a, Study design for the mouse BLM lung injury model, analyzing lungs collected at days 7 (d7) and 21 (d21) post oropharyngeal (o.p.) administration (*n* = 6 for BLM, *n* = 3 for vehicle per time point (t.p.)), used for Visium experiments. **b**, DEA of fibrotic regions in human IPF and BLM-treated mice versus controls, with Venn diagrams of DEGs unique and common to human IPF and mouse BLM at d7 or d21 and highlighting genes with inverse expression patterns. **c**, Integration of Visium and scRNA-seq data (Strunz et al.[11]; GSE141259) to infer spot cell-type densities, exemplified by inferred AT2 density. **d**, Averaged cell-type abundance per animal, comparing time points and treatments for selected cell types (Welch two sample *t*-test, two-sided; $n_{Vehicle}$ = 3 and n_{BLM} = 6, per time point). *P < 0.05, **P < 0.01 and ***P < 0.001 (AT2, d7: P = 0.0036; myofibroblast, d7: P = 0.0011, d21: P = 0.0306; *Krt8*[+] ADI, d7: P = 0.0042, d21: P = 0.0045; Recr. Mac., d7: P = 0.0006; Res. Mac: d7: P < 0.0001, d21: P < 0.0001). The center line is the median, the box limits are the upper and lower quartiles,

the whiskers are 1.5× interquartile range and the points are the value per animal. **e**, Cell–cell correlation heat maps display cellular colocalization compartments (defined by tree height cutoff, h = 1.5, orange dashed line) across condition and time point. The Sankey diagram illustrates cell type shifts within compartments from vehicle to BLM d7 to BLM d21, with *Krt8*[+] ADI (orange line) and myofibroblast (green line) populations highlighted. **f**, Computed compartment scores (F–H) based on cell-type densities for a BLM d21 lung section. **g**, Correlation (Pearson) of BLM d21 NMF factor activity and cell-type densities in all spots. The cell-type group colors match their respective compartments (A–H) from prior analysis. **h**, Comparison between human and mouse d21 NMF analyses using the top 100 factor-driving genes, filtered for factor pairs with a Jaccard index >0.1 to highlight major overlaps. AM, alveolar macrophages; Recr. Mac., recruited macrophages; Res. Mac., resolution macrophages; VECs, vascular endothelial cells; Prolif., proliferating.

tissue (H), we observed a compartment characterized by alveolar epithelium macrophages and *Krt8*[+] ADI cells (G), exhibiting a profile of regenerating alveolar tissue (Supplementary Table 8). Spatial mapping confirmed that compartment F aligned with bronchial structures and H coincided with fibrotic/remodeled tissue, while G was present along the borders of fibrotic areas and extending into intact tissue (Fig. 4f).

Regenerative signatures reveal distinct epithelial responses in mouse BLM-induced fibrosis. Given that d21 in the mouse model reflected an established stage of fibrosis with minimal acute inflammation, we focused on this time point for comparison with IPF. NMF application to the mouse d21 data (mmNMF$_{d21}$) generated 30 factors (Supplementary Table 9). Factor activity and cell-type abundance colocalization analysis largely reflected the d21 BLM compartmentalization,

affirming that NMF effectively captures patterns comparable with the cell-type deconvolution approach (Fig. 4g). The regenerating alveolar epithelial compartment (G) was represented by a set of factors primarily reflecting AT2 cells (F30), alveolar and resolution macrophages (F8) or activated AT2 and $Krt8^+$ ADI cells (F14). Factors F18, F5 and F7 predominantly represented AT1 and endothelial cells (Supplementary Table 9).

We compared the mouse $mmNMF_{d21}$ factors with the factors identified in the human IPF NMF analysis (hsNMF) (Fig. 4h). The top contributing genes showed overall weak overlap between human and mouse factors. However, factors associated with distinct morphological features, such as smooth muscle cells, blood vessels and ciliated airway epithelium, demonstrated more pronounced overlap, highlighting conserved signatures in normal lung structures, compared with disease or injury responses. Notably, factors containing transcriptional signatures for the human $KRT5^-/KRT17^+$ AbBa cells (hsNMF-F14) and mouse $Krt8^+$ ADI cells ($mmNMF_{d21}$-F14) had a limited overlap.

Translation of the fibrotic microenvironment
Contrasting responses in human IPF and mouse BLM model. We analyzed the spatial correlation of the factors containing the $KRT5^-/KRT17^+$ AbBa cells (hsNMF-F14) in human IPF samples and the $Krt8^+$ ADI cells ($mmNMF_{d21}$-F14) in mouse BLM samples. hsNMF-F14 activity predominantly correlated with fibroblasts ($HAS1$-hi), myofibroblasts and $KRT5^-/KRT17^+$ AbBa cells. Conversely, $mmNMF_{d21}$-F14 activity primarily correlated with $Krt8^+$ ADI cells and AT2 cells, while showing a weaker correlation with myofibroblasts. Additionally, $mmNMF_{d21}$-F14 showed correlation (albeit weaker) with AT1 cells, unlike hsNMF-F14 (Fig. 5a), in line with the distinct fibrogenic environment in the human AbBa niche.

To further compare the AbBa (IPF) and ADI (BLM) niches, we refined $mmNMF_{d21}$-F14 and identified four subclusters ($mmNMF_{d21}$-F14^{hi} C0–3), where cluster 0 ($mmNMF_{d21}$-F14^{hi} C0) exhibited strongest association with $Krt8^+$ ADI cells (Fig. 5b). We detected shared markers between hsNMF-F14^{hi} C0 and $mmNMF_{d21}$-F14^{hi} C0, including several collagens and ECM-related genes (for example, $COL1A1$, $FN1$, TNC and $CTHRC1$), epithelial cell markers ($CDH1$) and markers for human AbBa cells ($KRT17$) and mouse ADI cells ($KRT8$) (Fig. 5c and Supplementary Table 10). This suggests shared traits between AbBa and ADI regions involving ECM remodeling and a basaloid phenotype, which was further supported by pathway analysis (Extended Data Fig. 4d). However, the ADI-related gene signature predominantly engaged pathways related to inflammation and repair, whereas the AbBA signature reflected the chronic and progressive nature of IPF, dominated by immune responses and pathways governing long-term tissue remodeling.

For better understanding of the aberrant fibrotic niche drivers, we performed an upstream regulator analysis for hsNMF-F14^{hi} C0 and $mmNMF_{d21}$-F14^{hi} C0 (Fig. 5d and Supplementary Table 11). Both groups had predicted activation of TGFB1, TP53 and SMAD3, suggesting a conserved TGF-β-related mechanism[25,50,51] and cellular senescence[52]. hsNMF-F14^{hi} C0-specific regulators included the antifibrotic growth factor BMP7, APOE receptor SORL1 and GLIS1, a component of the Notch signaling pathway. $mmNMF_{d21}$-F14^{hi} C0 showed activation of oxidative stress and inflammation regulators including HIF1A, IL4, YAP1 and NFE2L2 (NRF2). Contrary to our findings in the human samples suggesting an activating role for apolipoprotein signaling upon hsNMF-F14^{hi} C0 (Fig. 3f–h), APOE and its receptor LDLR were predicted to be inhibited for $mmNMF_{d21}$-F14^{hi} C0 in the mouse samples.

Next, we examined the histological context of the $mmNMF_{d21}$-F14^{hi} clusters and quantified their presence in relation to fibrosis (Fig. 5e and Extended Data Fig. 4e–g). The $mmNMF_{d21}$-F14^{hi} C0 cluster was found within fibrotic regions, often near the junction between healthy and fibrotic tissue, comparable with the spatial relation of FF and hsNMF-F14^{hi} C0 in the IPF samples and suggesting a transitional state.

The BLM $Krt8^+$ ADI transitional cell population is predicted to originate from either AT2 cells or club cells and differentiate into AT1 cells[11]. By assessing the cell-type densities in relation to radial distance from the borders of the $mmNMF_{d21}$-F14^{hi} C0 niche, we identified high densities of AT2, activated AT2 and AT1 cells close to the niche (Fig. 5f). These observations shared similarities with the corresponding IPF hsNMF-F14^{hi} C0 analysis (Fig. 3c), highlighting the absence of AT1 cells around the human AbBa niche.

In BLM mouse data, spatial trajectory analysis of the alveolar epithelial cell types/states (Fig. 5g) showed a single path from AT2 cells through activated AT2 and ADI cells to AT1 cells. Conversely, in IPF lungs, the trajectory branched from AT2 cells through transitional AT2 cells and diverged into either $KRT5^-/KRT17^+$ AbBa or AT1 cells. Spatial visualization of the trajectories further illustrated the continuous colocalization of these cell types in BLM-injured mouse lungs and spatial separation of transitional AT2–AbBa and AT2–AT1 niches in IPF (Fig. 5h).

Immune cell dynamics in pulmonary fibrosis. Our NMF analysis revealed factors in the IPF and the d7 and d21 BLM datasets that shared key marker genes indicative of macrophages ($SPP1$, $CD68$ and $APOE$) (Extended Data Fig. 5a). $SPP1^+$ profibrotic macrophages, displaying an M2-polarization phenotype, have previously been implicated in ECM remodeling and fibrosis[6,53]. In the two IPF donor lungs containing bronchial honeycombing, the macrophage factor (hsNMF-F5) activity was particularly pronounced within these regions (Extended Data Fig. 5b). The hsNMF-F5 activity varied across samples and donors, potentially owing to tissue heterogeneity or constraints from the chosen number of NMF factors.

A shared histological feature between the IPF and BLM-injured lungs was presence of dense immune cell infiltrates embedded within the fibrotic tissue (Fig. 6a). In the time-point-separated mouse NMF analyses, $mmNMF_{d7}$ and $mmNMF_{d21}$ (Supplementary Table 9), we identified factors prevalent in these regions. In the human NMF, a similar factor was not consistently detected. Therefore, more targeted donor-specific NMF analyses were performed, which identified factors in three of the donors (IPF 1–3) overlapping spatially with the observed immune infiltrates. Gene expression (Fig. 6b) and cell-type composition (Fig. 6c) within the immune-dense regions revealed notable differences between IPF and BLM lungs. In the BLM model, enrichment of genes such as $Cd74$ and $Coro1a$ indicated presence of antigen-presenting cells and lymphocytes[54]. Additionally, a distinct plasma cell factor, driven by IgA heavy chain ($Igha$) expression (Extended Data Fig. 5c), was identified in proximity to the d21 BLM immune-dense structures (Fig. 6a). Overall, the BLM regions demonstrated a relatively balanced mixture of B, T and dendritic cells, in contrast to the human IPF samples, where a pronounced expression of $CXCL13$ suggested a B cell-dominated immune response[55], which was further supported by the cell-type composition data. Given recent descriptions of their presence in both healthy and diseased lungs[56], lymphocytes and plasma cells may modulate fibrosis progression.

Taken together, our analyses delineate distinct cellular trajectories and molecular mechanisms in the fibrotic niche of human IPF and the BLM mouse model (Fig. 6d). We highlight the arrested alveolar cell regeneration in IPF versus the active repair in the BLM model, alongside distinct signaling molecules, such as TGF-β, APOE, YAP1 and TEAD, and differences in immune cell presence. While we recognize the complexity and heterogeneity of IPF and aspects that remain unexplored (Supplementary Note 6), these comparative insights underscore the unique characteristics of fibrosis in human IPF.

Discussion
Our study presents comprehensive comparative genome-wide spatial transcriptome maps of the diverse cellular ecosystems and distinct molecular signatures in the human IPF lung and the BLM mouse model. Investigations into prefibrotic alveolar tissue and the leading edge of fibrotic remodeling offer potential for identification of targetable early disease-associated molecules for therapeutic intervention.
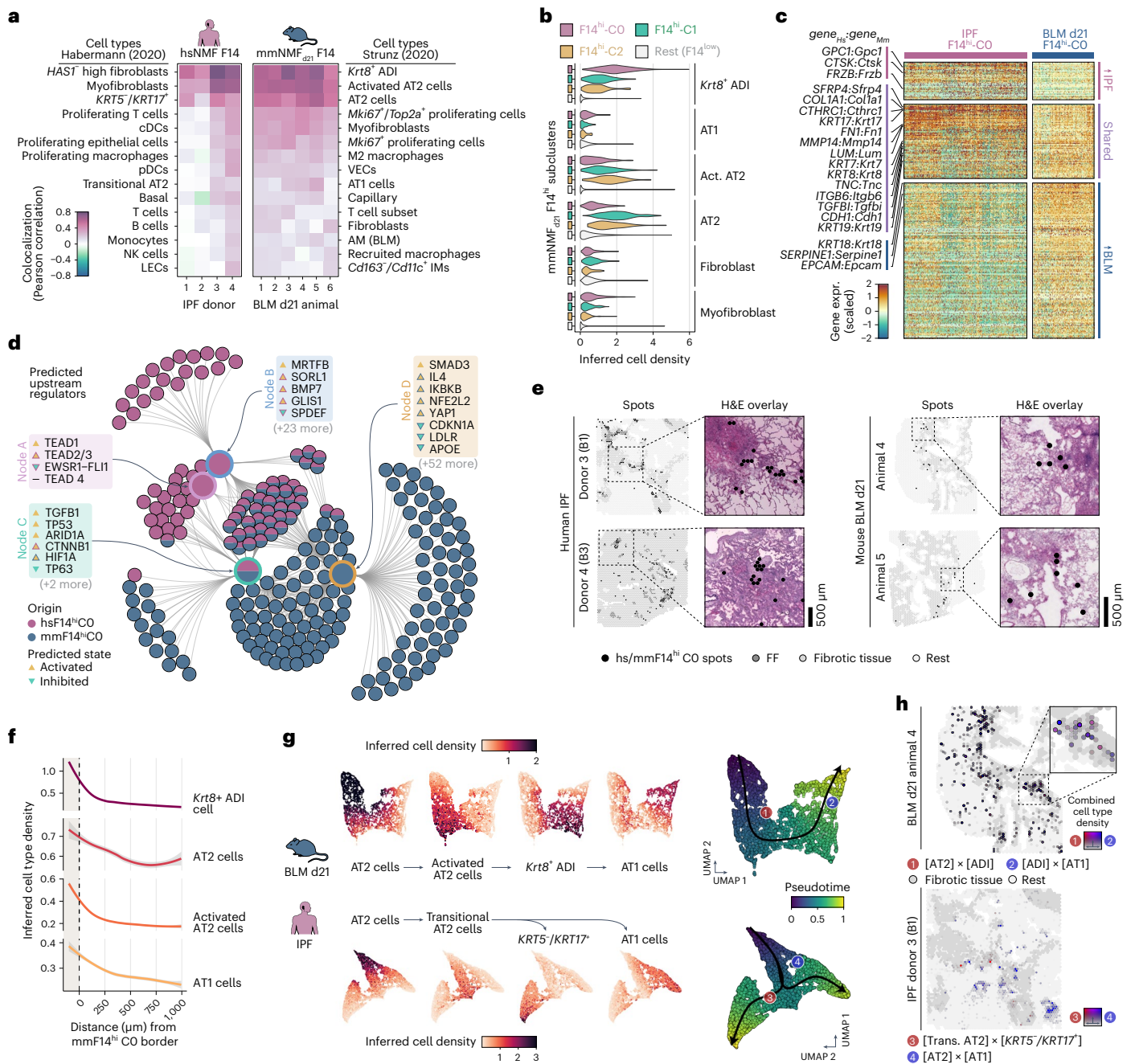
Fig. 5 | Translational dissection of the fibrotic niche and cellular dynamics. a, Correlation of hsNMF-F14^{hi} and mmNMF-F14^{hi} factor activity with the top 15 correlating cell types from Habermann[5] (human) and Strunz[11] (mouse) scRNA-seq datasets. **b**, Cell density distribution in mmNMF$_{d21}$-F14^{hi} subclusters showing high abundance of *Krt8*$^+$ ADI cells in mmNMF$_{d21}$-F14hi C0 and AT2 cells in C1 and C2. **c**, Heat map with scaled and centered hsNMF-F14^{hi}C0 and mmNMF$_{d21}$-F14^{hi}C0 marker gene expression (expr.), grouped by higher in IPF, shared or higher in BLM (adjusted *P* < 0.01, avgerage log$_2$FC >0). **d**, Comparative network plot of top significant regulators (*P* < 10^{-7}, right-tailed Fisher's exact test) from IPA upstream analyses of marker genes (adjusted *P* < 0.05) in hsNMF-F14^{hi}C0 and mmNMF$_{d21}$-F14^{hi}C0. The inner nodes illustrate groups of regulators sharing genetic influences and the outer nodes represent contributing marker genes. **e**, Spatial mapping of hsNMF-F14^{hi}C0 and mmNMF$_{d21}$-F14^{hi}C0 spots within tissue sections illustrating

the relationship with fibrotic regions. **f**, Radial distribution line graphs of cell densities around mmNMF$_{d21}$-F14^{hi}C0. The gray shading corresponds to the 95% confidence interval. **g**, Spatial trajectories of spots with high inferred densities of AT2, activated AT2, *Krt8*+ ADI and AT1 cells (mouse) or AT2, transitional AT2, *KT5-/KRT17*$^+$ and AT1 cells (human). **h**, Spatial colocalization of AT2-to-*Krt8*$^+$ ADI (1, red) and ADI-to-AT1 (2, blue) inferred cell densities in a mouse lung section and transitional AT2-to-*KRT5$^-$/KRT17*$^+$ (3, red) and AT2-to-AT1 (4, blue) densities in one human IPF lung section. Colocalization intensities are visualized on a red–blue spectrum, with mixtures appearing purple and spot opacity reflecting intensity level. Tissue background and areas of fibrosis (gray) are illustrated for context. VECs, vascular endothelial cells; pDC/cDCs, plasmacytoid/classical dendritic cells; NK cells, natural killer cells; AM, alveolar macrophages; IM, interstitial macrophages; Act. AT2, activated AT2 cells; Trans. AT2, transitional AT2 cells.

We propose a central involvement of TGF-β signaling in IPF, alongside other mediators such as TP53, SMAD3, BMP7, MRTFB, TEAD, GLIS1 and APOE, which are linked to senescence, myofibroblast activation and differentiation, Notch and Wnt signaling, apoptosis and cell migration.

Using data factorization, we identify the *KRT5$^-$/KRT17*$^+$ AbBa population in IPF and the *Krt8*$^+$ ADI population in the BLM model. Analyzing their proximate neighborhoods delineated critical regions within the fibrotic landscape, and the complex signaling networks
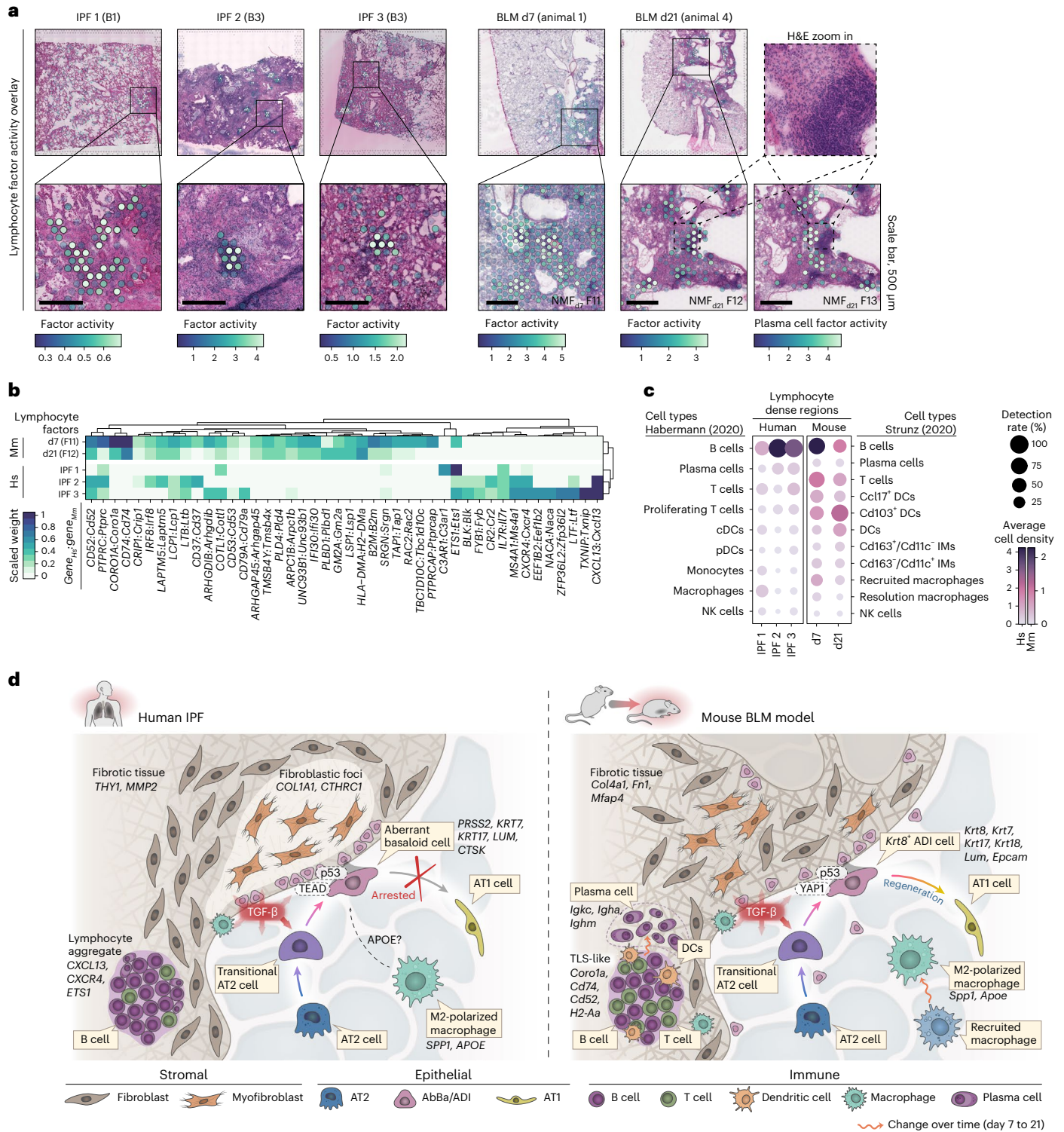
Fig. 6 | Immune cell signatures and comparative overview of the fibrotic niche in IPF and the BLM mouse model. a, Spatial visualization of NMF factors overlapping dense lymphocyte or immune cell aggregates in selected human and mouse samples. Scale bars, 500 μm. Imaged at 20× magnification. **b**, Heat map displays the top contributing factor genes across condition, filtered to show genes with a summed scaled weight above 0.5 across the groups. **c**, Dot plot with inferred cell-type densities, for selected immune cell types from the 'Habermann'[5] and 'Strunz'[11] datasets, in the most active spots of the selected human and mouse factors. **d**, Schematic summary of the fibrotic niche in human IPF lungs and in mouse BLM-injured lungs, illustrating the proposed cellular interplay within the fibrotic lungs. A key distinction between IPF and the BLM mouse model was centered around the diverging regenerative properties of the IPF-associated AbBa cells versus the mouse Krt8+ ADI cells. While both populations exhibit signs of

senescence (p53), the mouse ADI state appears to maintain a functional balance that still prompts it to differentiate into AT1 cells. TGF-β and Wnt-related (TEAD, YAP1) signaling pathways were central within the fibrotic niche, and the presence of immune cells in proximity to, or within, the severely remodeled tissue implies active fibrogenic modulatory roles. Pro-fibrotic M2-polarized ('resolution') macrophages with similar gene signatures, expressing SPP1 (Spp1) and APOE (Apoe) were detected in both human IPF and mouse BLM-injured lungs. In contrast to human IPF AbBa regions, a predicted negative APOE upstream signaling was identified in mouse ADI regions. In mouse, the recruited proinflammatory macrophages seen at the early time point post BLM-installation were absent by d21. Establishment of plasma cells adjacent to TLS-like areas in the BLM-injured mice occurred at the later time point. pDC/cDCs, plasmacytoid/classical dendritic cells; IMs, interstitial macrophages; NK cells, natural killer cells; TLS, tertiary lymphoid structure.

identified within these niches suggest these cells participate in the fibrotic response through paracrine signaling and local microenvironment modulation. TGF-β1, a profibrotic cytokine with a major role in IPF pathogenesis[25,33] and SMAD3, integral to the TGF-β signaling pathway[57], were predicted as upstream regulators in human AbBa and mouse ADI fibrotic niches, pointing to a shared TGF-β-driven fibrotic signaling pathway. Furthermore, our data suggest that the role of APOE signaling within the IPF fibrotic niche is more substantial than previously appreciated. This warrants further exploration into the potential regulatory function of APOE in IPF, given its well-documented function in lipid metabolism and its emerging role in immunomodulation and fibrosis[58,59].

Through tracing the alveolar epithelial cell spatial trajectory, we observed an AT2–ADI–AT1 lineage in the mouse model that is preserved in situ, supporting previous single-cell and in vitro studies[11,60,61] and indicating an ongoing postinjury repair mechanism. In contrast, human IPF lungs depicted a divergent path, with AT2 cells branching into either AbBa cells or AT1 cells. The apparent disruption in the IPF lung regenerative process is in line with descriptions of AbBa cell persistence as an intermediate, nonregenerative state[5], potentially driving the progressive, irreversible fibrosis in IPF, as opposed to the resolution of fibrosis following acute injury in the mouse model. These findings highlight key challenges in translating animal models to human disease and suggest that the acute BLM mouse model might offer valuable insight into alveolar regeneration. Application of SRT to other pulmonary fibrosis animal models, such as the repeat BLM instillation model[62] in which a more persistent, senescent *Krt8*+ transitional alveolar cell state has been identified[52], could potentially provide more insights into disease progression in the IPF lung.

Taken together, our study illustrates the potential for spatial transcriptomics to deepen our understanding of IPF pathology and offers rich datasets to further probe the complex cellular interplay in pulmonary fibrosis. This work provides resolution of key mechanisms underpinning IPF and proposes a divergent cellular trajectory toward arrested regeneration in the human IPF lung, as a potential target for the discovery of innovative disease modifying therapies.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41588-024-01819-2.

References

1. Richeldi, L., Collard, H. R. & Jones, M. G. Idiopathic pulmonary fibrosis. *Lancet* **389**, 1941–1952 (2017).
2. Spagnolo, P. et al. Idiopathic pulmonary fibrosis: disease mechanisms and drug development. *Pharm. Ther.* **222**, 107798 (2021).
3. Mei, Q., Liu, Z., Zuo, H., Yang, Z. & Qu, J. Idiopathic pulmonary fibrosis: an update on pathogenesis. *Front. Pharm.* **12**, 797292 (2021).
4. Adams, T. S. et al. Single-cell RNA-seq reveals ectopic and aberrant lung-resident cell populations in idiopathic pulmonary fibrosis. *Sci. Adv.* **6**, eaba1983 (2020).
5. Habermann, A. C. et al. Single-cell RNA sequencing reveals profibrotic roles of distinct epithelial and mesenchymal lineages in pulmonary fibrosis. *Sci. Adv.* **6**, eaba1972 (2020).
6. Morse, C. et al. Proliferating SPP1/MERTK-expressing macrophages in idiopathic pulmonary fibrosis. *Eur. Respir. J.* **54**, 1802441 (2019).
7. Reyfman, P. A. et al. Single-cell transcriptomic analysis of human lung provides insights into the pathobiology of pulmonary fibrosis. *Am. J. Respir. Crit. Care Med* **199**, 1517–1536 (2019).
8. Melms, J. C. et al. A molecular single-cell lung atlas of lethal COVID-19. *Nature* **595**, 114–119 (2021).
9. Kathiriya, J. J. et al. Human alveolar type 2 epithelium transdifferentiates into metaplastic KRT5+ basal cells. *Nat. Cell Biol.* **24**, 10–23 (2022).
10. Ptasinski, V. et al. Modeling fibrotic alveolar transitional cells with pluripotent stem cell-derived alveolar organoids. *Life Sci. Alliance* **6**, e202201853 (2023).
11. Strunz, M. et al. Alveolar regeneration through a *Krt8*+ transitional stem cell state that persists in human lung fibrosis. *Nat. Commun.* **11**, 3559 (2020).
12. Degryse, A. L. et al. Repetitive intratracheal bleomycin models several features of idiopathic pulmonary fibrosis. *Am. J. Physiol. Lung Cell Mol. Physiol.* **299**, L442–L452 (2010).
13. Kadur Lakshminarasimha Murthy, P. et al. Human distal lung maps and lineage hierarchies reveal a bipotent progenitor. *Nature* **604**, 111–119 (2022).
14. Sikkema, L. et al. An integrated cell atlas of the lung in health and disease. *Nat. Med* **29**, 1563–1577 (2023).
15. Heinzelmann, K. et al. Single-cell RNA sequencing identifies G-protein coupled receptor 87 as a basal cell marker expressed in distal honeycomb cysts in idiopathic pulmonary fibrosis. *Eur. Respir. J.* **59**, 2102373 (2022).
16. Fu, X. et al. Polony gels enable amplifiable DNA stamping and spatial transcriptomics of chronic pain. *Cell* **185**, 4621–4633.e17 (2022).
17. Liu, Y. et al. High-spatial-resolution multi-omics sequencing via deterministic barcoding in tissue. *Cell* **183**, 1665–1681.e18 (2020).
18. Rodriques, S. G. et al. Slide-seq: a scalable technology for measuring genome-wide expression at high spatial resolution. *Science* **363**, 1463–1467 (2019).
19. Ståhl, P. L. et al. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* **353**, 78–82 (2016).
20. Eyres, M. et al. Spatially resolved deconvolution of the fibrotic niche in lung fibrosis. *Cell Rep.* **40**, 111230 (2022).
21. Guillotin, D. et al. Transcriptome analysis of IPF fibroblastic foci identifies key pathways involved in fibrogenesis. *Thorax* **76**, 73–82 (2021).
22. Blumhagen, R. Z. et al. Spatially distinct molecular patterns of gene expression in idiopathic pulmonary fibrosis. *Respir. Res* **24**, 287 (2023).
23. Hanmandlu, A. et al. Transcriptomic and epigenetic profiling of fibroblasts in idiopathic pulmonary fibrosis. *Am. J. Respir. Cell Mol. Biol.* **66**, 53–63 (2022).
24. Pardo, A., Cabrera, S., Maldonado, M. & Selman, M. Role of matrix metalloproteinases in the pathogenesis of idiopathic pulmonary fibrosis. *Respir. Res.* **17**, 23 (2016).
25. Frangogiannis, N. Transforming growth factor-β in tissue fibrosis. *J. Exp. Med* **217**, e20190103 (2020).
26. Piersma, B., Bank, R. A. & Boersema, M. Signaling in fibrosis: TGF-β, WNT, and YAP/TAZ converge. *Front. Med. (Lausanne)* **2**, 59 (2015).
27. Lin, X. & Boutros, P. C. Optimization and expansion of non-negative matrix factorization. *BMC Bioinformatics* **21**, 7 (2020).
28. Kleshchevnikov, V. et al. Cell2location maps fine-grained cell types in spatial transcriptomics. *Nat. Biotechnol.* **40**, 661–671 (2022).
29. Zuo, W. L. et al. Dysregulation of club cell biology in idiopathic pulmonary fibrosis. *PLoS ONE* **15**, e0237529 (2020).
30. Seibold, M. A. et al. A common MUC5B promoter polymorphism and pulmonary fibrosis. *N. Engl. J. Med* **364**, 1503–1512 (2011).
31. Herrera, J. A. et al. The UIP/IPF fibroblastic focus is a collagen biosynthesis factory embedded in a distinct extracellular matrix. *JCI Insight* **7**, e156115 (2022).

32. Yamaguchi, M. et al. Fibroblastic foci, covered with alveolar epithelia exhibiting epithelial-mesenchymal transition, destroy alveolar septa by disrupting blood flow in idiopathic pulmonary fibrosis. *Lab. Invest.* **97**, 232–242 (2017).

33. Ye, Z. & Hu, Y. TGF-β1: gentlemanly orchestrator in idiopathic pulmonary fibrosis (review). *Int J. Mol. Med.* **48**, 132 (2021).

34. Riemondy, K. A. et al. Single cell RNA sequencing identifies TGF-β as a key regenerative cue following LPS-induced lung injury. *JCI Insight* **5**, e123637 (2019).

35. Gau, D. & Roy, P. SRF'ing and SAP'ing—the role of MRTF proteins in cell migration. *J. Cell Sci.* **131**, jcs218222 (2018).

36. Lin, K. C., Park, H. W. & Guan, K. L. Regulation of the hippo pathway transcription factor TEAD. *Trends Biochem. Sci.* **42**, 862–872 (2017).

37. Park, H. W. et al. Alternative Wnt signaling activates YAP/TAZ. *Cell* **162**, 780–794 (2015).

38. Speight, P., Kofler, M., Szászi, K. & Kapus, A. Context-dependent switch in chemo/mechanotransduction via multilevel crosstalk among cytoskeleton-regulated MRTF and TAZ and TGF-β-regulated Smad3. *Nat. Commun.* **7**, 11642 (2016).

39. Zheng, Y. & Pan, D. The hippo signaling pathway in development and disease. *Dev. Cell* **50**, 264–282 (2019).

40. González-Amor, M., Dorado, B. & Andrés, V. Emerging roles of interferon-stimulated gene-15 in age-related telomere attrition, the DNA damage response, and cardiovascular disease. *Front. Cell Dev. Biol.* **11**, 1128594 (2023).

41. Browaeys, R., Saelens, W. & Saeys, Y. NicheNet: modeling intercellular communication by linking ligands to target genes. *Nat. Methods* **17**, 159–162 (2020).

42. Cui, H. et al. Monocyte-derived alveolar macrophage apolipoprotein E participates in pulmonary fibrosis resolution. *JCI Insight* **5**, e134539 (2020).

43. Zheng, P. et al. Tumor-associated macrophages-derived exosomes promote the migration of gastric cancer cells by transfer of functional apolipoprotein E. *Cell Death Dis.* **9**, 434 (2018).

44. Jenkins, R. G. et al. An official american thoracic society workshop report: use of animal models for the preclinical assessment of potential therapies for pulmonary fibrosis. *Am. J. Respir. Cell Mol. Biol.* **56**, 667–679 (2017).

45. Moore, B. B. & Hogaboam, C. M. Murine models of pulmonary fibrosis. *Am. J. Physiol. Lung Cell Mol. Physiol.* **294**, L152–L160 (2008).

46. Nishimura, M. et al. Extracellular cleavage of collagen XVII is essential for correct cutaneous basement membrane formation. *Hum. Mol. Genet* **25**, 328–339 (2016).

47. Bazan, J. F. et al. A new class of membrane-bound chemokine with a CX3C motif. *Nature* **385**, 640–644 (1997).

48. Liu, F. et al. LRRc17 controls BMSC senescence via mitophagy and inhibits the therapeutic effect of BMSCs on ovariectomy-induced bone loss. *Redox Biol.* **43**, 101963 (2021).

49. Somia, N. V. et al. LFG: an anti-apoptotic gene that provides protection from Fas-mediated cell death. *Proc. Natl Acad. Sci. USA* **96**, 12667–12672 (1999).

50. Elston, R. & Inman, G. J. Crosstalk between p53 and TGF-β signalling. *J. Signal Transduct.* **2012**, 294097 (2012).

51. Hata, A. & Chen, Y. G. TGF-β signaling from receptors to smads. *Cold Spring Harb. Perspect. Biol.* **8**, a022061 (2016).

52. Wang, F. et al. Regulation of epithelial transitional states in murine and human pulmonary fibrosis. *J. Clin. Invest* **133**, e165612 (2023).

53. Liu, J. et al. CTHRC1+ fibroblasts are stimulated by macrophage-secreted SPP1 to induce excessive collagen deposition in keloids. *Clin. Transl. Med.* **12**, e1115 (2022).

54. Shin, J. S. et al. Surface expression of MHC class II in dendritic cells is controlled by regulated ubiquitination. *Nature* **444**, 115–118 (2006).

55. Ansel, K. M. et al. A chemokine-driven positive feedback loop organizes lymphoid follicles. *Nature* **406**, 309–314 (2000).

56. Madissoon, E. et al. A spatially resolved atlas of the human lung characterizes a gland-associated immune niche. *Nat. Genet.* **55**, 66–77 (2023).

57. Schmierer, B. & Hill, C. S. TGF-β-SMAD signal transduction: molecular specificity and functional flexibility. *Nat. Rev. Mol. Cell Biol.* **8**, 970–982 (2007).

58. Nascimento, J. C. R. et al. Impact of apolipoprotein E genetic polymorphisms on liver disease: an essential review. *Ann. Hepatol.* **19**, 24–30 (2020).

59. Nishimura, K. et al. Dual disruption of eNOS and ApoE gene accelerates kidney fibrosis and senescence after injury. *Biochem Biophys. Res Commun.* **556**, 142–148 (2021).

60. 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).

61. Kobayashi, Y. et al. Persistence of a regeneration-associated, transitional alveolar epithelial cell state in pulmonary fibrosis. *Nat. Cell Biol.* **22**, 934–946 (2020).

62. Redente, E. F. et al. Persistent, progressive pulmonary fibrosis and epithelial remodeling in mice. *Am. J. Respir. Cell Mol. Biol.* **64**, 669–676 (2021).

Methods

Ethical considerations

Human IPF lung samples were acquired with approval by the Human Research Ethics Committee in Gothenburg, Sweden (permit number 1026-15) and HC lung samples were acquired with approval by the Human Research Ethics Committee in Lund, Sweden (permit number Dnr 2016/317). All participants gave written informed consent before inclusion and did not receive any compensation for their participation.

All mouse experiments were approved by the Gothenburg Ethics Committee for Experimental Animals in Sweden and conformed to Directive 2010/63/EU. The present study was approved by the Research Ethics Committee in Gothenburg, Sweden (permit number 31-5373/11). Animal handling conformed to standards established by the Council of Europe ETS123 AppA, the Helsinki Convention for the Use and Care of Animals, Swedish legislation and AstraZeneca global internal standards.

All investigations were performed in accordance with the declaration of Helsinki.

Experimental methods

Human lung tissues. IPF lung tissue was obtained from patients that had undergone lung transplant. Healthy lung tissue was obtained from deceased donors without known lung disease. Human tissues were collected by resection from peripheral lung. Fresh-frozen tissues were obtained from four HC subjects and four patients with IPF. For each patient with IPF, three different tissues were selected to represent areas of mild ('B1'), moderate ('B2') or severe ('B3') fibrosis within the same donor, as determined by histological inspection (see Supplementary Table 1 for additional details regarding donors and samples).

Mice and BLM challenge. Female C57BL/6NCrl mice (Charles River, Germany) were 8 weeks old on the day of arrival. After a 5-day acclimatization period, the mice were challenged with 30 µl saline or BLM (Apollo Scientific, BI3543, Chemtronica Sweden; 40 µg per mouse) dissolved in saline via oropharyngeal route administration. Lungs were collected at d7 ($n = 6$ for BLM, $n = 3$ for vehicle) or d21 ($n = 6$ for BLM, $n = 3$ for vehicle) following administration.

The mice were housed in Macrolon III cages with poplar chips (Rettenmeier and Sohne), shredded paper, gnaw sticks and a paper house in 12 h–12 h light–dark cycle at $21 \pm 1\,°C$, $55 \pm 15\%$ relative humidity with free access to food (R70, Lantmannen AB) and tap water.

Mouse tissue collection. Mice were anesthetized with isoflurane (5%, air flow ~2 l min⁻¹, maintained at 3% with ~0.7 l min⁻¹ air flow). An incision was made from the stomach up to the chin. Heparin (0.1 ml) was injected intracardially through the diaphragm, and the abdominal aorta and apex of the heart was cut. Pulmonary circulation was perfused with 0.8 ml 37 °C saline followed by 0.6 ml 37 °C low-temperature melt agarose (SeaPlaque). The left lung lobe was inflated with 0.4–0.5 ml 37 °C low-melt agarose via the trachea and tied off, collected, snap frozen in prechilled NaCl over dry ice and stored at −80 °C.

The methods describing animal monitoring and sample collection and analyses for pre-study evaluation of BLM time points are outlined in Supplementary Note 7.

Generation of SRT. Optimal cutting temperature (OCT)-embedded human lung tissue blocks and agarose-inflated mouse lungs (left lobe) were cryosectioned at 10 µm (mouse) or 12 µm (human) with cryostat temperatures of −20 °C. RNA quality for human samples were assessed by extracting RNA from ten sections using the RNeasy Plus Mini kit (Qiagen) and estimating RNA integrity numbers (RIN) with a 2100 Bioanalyzer (Agilent), which ranged from 5.4 to >8, except for one sample with a RIN of 3 (IPF donor 2, B3). Despite lower RINs for some samples, tissues with satisfactory histological integrity were included in further analyses. For mouse sections, RNA was extracted similarly

with the RNeasy micro kit (Qiagen). The RIN values were estimated to >9 for all samples using a 5300 Fragment Analyzer (Agilent).

The tissue sections were placed on Visium Gene Expression slides and stored at −80 °C until further processing. The methanol fixation, hematoxylin and eosin (H&E) staining and imaging Visium protocol (10x Genomics) were followed. Human lung sections were imaged using Axio Imager.Z2 (ZEISS) light microscope at 20× magnification and stitched using Vslide (MetaSystems). Mouse lung sections were imaged at 20× magnification using an Aperio Digital Pathology Slide Scanner (Leica Biosystems).

The sequencing libraries were prepared using the Visium Spatial Gene Expression Slide and Reagent Kit (cat. no. 1000184, 10x Genomics) with the Dual Index Kit TT Set A (cat. no. 1000215, 10x Genomics) according to the Visium Spatial Gene Expression User Guide (Rev C). The human and mouse libraries were pooled separately and sequenced on a NovaSeq 6000 (Illumina) with an S4 flowcell.

Histopathology annotations. Histopathological assessments were performed on H&E-stained tissue sections using the Loupe Browser (10x Genomics) software and were manually annotated based on tissue morphology. The human lung data were classified into 'blood vessel', 'large airway', 'diseased (remodeled) tissue', 'fibroblastic foci/fibrous tissue', 'inflammation' and 'within normal limits' (alveolar), where 'inflammation' was distinguished as areas with aggregations of immune cells and 'diseased tissue' corresponding to clearly recognizable changes in normal lung architecture. The 'fibroblastic foci/fibrous tissue' was annotated based on microscopic features, including nuclear density and shape, and increased collagenous matrix. Since this selection also encompassed fibrous tissue that did not classify as FF, a second layer of annotations were performed to specifically identify the FF. Mouse data were annotated as 'blood vessel', 'large airway', 'within normal limits' (alveolar), 'inflammation (d7)', 'inflammation (d21)' and 'suspect fibrosis/fibroplasia (d21)'. The 'inflammation (d7)' areas comprised an indistinguishable mix of inflammatory and fibrotic tissue, while 'inflammation (d21)' labeled dense immune cell aggregates. Consequently, both spots labeled as 'inflammation (d7)' and 'suspect fibrosis/fibroplasia (d21)' contains fibrotic tissue.

Computational processing and analysis

Processing of Visium sequencing data. FastQ files were processed using the Space Ranger 1.2.2 (10x Genomics) pipeline. Sequence reads were mapped to the respective reference genomes GRCh38 (human) and mm10 (mouse). Manual alignment to the fiducial frame and identification of tissue-covered spots was performed using Loupe Browser (v.6, 10x Genomics).

Spatial cell-type mapping. For human samples, spatial deconvolution was performed using cell2location[28] (v. 0.1) against a previously published scRNA-seq dataset (Gene Expression Omnibus (GEO) accession GSE135893)[5]. The method uses signatures from the provided scRNA-seq data to infer absolute numbers (density) of cell types within each spatial spot. The single-cell regression model was trained with max_epochs = 250 after selecting genes with parameters nonz_mean_cutoff = 1.25, cell_count_cutoff = 5 and cell_percent_cutoff = 0.05. The cell2location model was thereafter obtained with parameters max_epochs = 10,000, detection_alpha = 20 and $n = 7$. For the mouse data, a scRNA-seq dataset from the BLM mouse model (collected at multiple time points including d7 and d21) was used (GEO accession GSE141259)[11]. max_epochs = 400 was set for single-cell model generation using the parameters nonz_mean_cutoff = 1.10, cell_count_cutoff = 4 and cell_percent_cutoff = 0.02 for gene selection. For model training, max_epochs = 15000, detection_alpha = 20 and $n = 7$ was applied.

Downstream data quality control and processing. Filtering, processing and analyses of the Visium data were performed in R (v.4.0.5) using

the STUtility (v.1.1.1)[63], Seurat (v.4.1.1)[64] and 'semla'[65] (v. 1.1.6, under R v. 4.2.3 and Seurat v. 4.3.0.1) packages.

For the human IPF and HC samples, tissue-covered spots were selected, and data were imported into R via the STUtility function 'InputFromTable'. Initial filtering settings were: minimum 350 UMIs per spot, 100 UMIs per gene, 10 genes per spot and five spots per gene. Additional filtering excluded spots with more than 30% mitochondria and/or hemoglobin gene expression. Gene information was retrieved via biomaRt (ref. [66]) and used to select for 'protein coding', 'IG' (immunoglobulin) and 'TR' (T cell receptor) gene biotypes, as well as to flag X and Y chromosome genes for removal to avoid sex biases. Normalization and scaling was performed with 'SCTransform'[67] (Seurat package), specifying sample ID and donor as variables to regress out, to remove major effects of technical and interindividual differences.

The mouse data were filtered in a similar manner, apart from omitting the number of genes per spot ('minGenesPerSpot') cutoff, and filtering for number of UMIs per spot was set to 300. 'SCTransform' was used for normalization, specifying animal ID as a variable to regress out.

All thresholds for filtering were set based on initial examination of the raw data to exclude low quality spots (or spots outside of tissue areas) and genes with low expression.

Visium pseudobulk DEA. For DEA between conditions, pseudobulk datasets were generated by aggregating raw counts per gene across all spots belonging to a donor or animal. DESeq2 (ref. [68]) (v. 1.30.1) was used for differential gene testing by specifying 'condition', with 'control' as reference in the design.

For comparison of the normal-appearing alveolar and the fibrotic tissue in IPF samples, gene count data were aggregated into pseudobulk based on histological annotations of 'within normal limits (alveolar)', 'diseased (remodeled) tissue' and 'fibroblastic foci/fibrous tissue'. These regions were compared against HC alveolar regions using DESeq2 in two analyses (HC−alveolar versus IPF−alveolar and HC−alveolar versus IPF−fibrotic). Overlapping upregulated DEGs ($n = 223$; adjusted P value <0.01, log$_2$ fold change (log$_2$FC) >0) between the two analyses were used to compute Pearson correlation between expression values in alveolar tissue and distance (0−500 μm) from the fibrotic region border. This analysis was performed on all IPF samples with available histological annotation ($n = 12$). Radial distances from fibrotic areas were calculated using the 'semla::RadialDistance' function with 'remove_singletons=FALSE', after which only data from fibrotic and alveolar regions were kept before computing gene expression versus distance correlations. The P values were adjusted using the Benjamini−Hochberg method, and expression of significantly (adjusted $P < 0.05$) correlated genes were visualized along the fibrosis−distance axis (distances −1,000 to 1,000 μm) as line plots using a generalized additive model (GAM) for smoothing.

For translational comparison of fibrotic regions between IPF and BLM d7 or d21, pseudobulk data were generated by pooling counts on donor/animal level from spots labeled as fibrotic, FF, remodeled or inflamed (BML d7) in disease condition samples. These data were compared with control samples using DESeq2, with 'condition' set as design, for each species and time point. Orthogene[69] (v. 1.4.2) was used to identify mouse orthologues of human genes, and the results were filtered to include only genes with available orthologues and present in all datasets (total of 12,611 genes) (Supplementary Table 7).

NMF. Deconvolution through NMF was applied to the Visium gene expression data using the 'RunNMF' function in STUtility, which relies on the NNLM package (v. 0.4.4). The method decomposes data into a set number of factors that are expressed as non-negative values (activity) within each data point (spot) along with a feature (gene) loading matrix, describing the contribution (weight) of each gene to the factors. The full human (HC and IPF) dataset was deconvolved into 30 factors ('hsNMF'), while the mouse data (vehicle control and BLM) were split by time point (d7, d21) before each subset was deconvolved into 30 factors (mmNMF$_{d7}$, mmNMF$_{d21}$). To describe each factor, functional enrichment analysis of the top 25 most contributing genes for each factor was performed using the 'gost' function in the gprofiler2 (v. 0.2.1) R package[70], with the 'hsapiens' (human) or 'mmusculus' (mouse) organism specified. All factors were further annotated by examining the top contributing genes, the spatial localization of factor activity and their abundance in different samples (diseased or control).

To compare factors across species, the Jaccard similarity index was computed using the top 100 contributing genes for each factor, calculated as the intersection over the union of each gene set pair.

The distribution of each factor within the human samples (hsNMF) was assessed by calculating the frequency of spots in the 99th percentile of factor activity (F^{hi}) versus the total number of spots in each tissue block category (B0−3). Spatial colocalization of factors and cell types was analyzed using pairwise Pearson correlation between spot factor activity and inferred cell-type density. To explore donor variability in colocalization, the human data were grouped into HC (all HC donors), IPF donor 1, IPF donor 2, IPF donor 3 and IPF donor 4, before correlation analysis.

For subclustering of hsNMF-F14, a principal component (PC) analysis was performed on the most active (99th quantile) spots (F14^{hi}) and PCs 1−8 were used as input for 'FindNeighbors' and 'FindClusters' (resolution of 0.4), which generated five clusters. The mmNMF$_{d21}$ F14^{hi} spots were subclustered using the same approach but with PCs 1−14 as input and clustering resolution of 0.5, obtaining three clusters.

Spatial enrichment in fibrotic regions. For IPF samples, we used the histopathological annotations to extract the radial distances, d, from the areas of FF, using the 'RadialDistance' function in 'semla'[65] (v. 1.1.6; R v. 4.2.3; Seurat v. 4.3.0.1) with kept singletons. Distances of 0 μm represent the outer line of spots immediately exterior to the FF-labeled spots. Distances were binned into groups of 100 μm (corresponding to one row of spots), and the F14^{hi} spot count was registered in each bin. To produce a null distribution, the F14^{hi} cluster labels were shuffled across all spots within each sample, $k = 100$ times. Distance bins >3,000 μm were discarded to exclude distant tissue areas. For each cluster, the observed count at every bin (x_d), the mean (μ_d) and the standard deviation (σ_d) across all randomization rounds were used to compute a Z-score per distance bin as: $Z_d = \frac{x_d - \mu_d}{\sigma_d}$. The same approach was applied to the d21 BLM mouse data to analyze spatial enrichment of the mmNMF$_{d21}$ F14^{hi} clusters in relation to areas of fibrosis (annotated 'suspect fibrosis/fibroplasia'). The P values were computed from absolute Z-scores using a two-tailed t-test.

Expression radial distance analysis from F14^{hi} clusters. Fluctuations in gene expression and cell-type densities along a radial distance from the hsNMF-F14^{hi}C0 or mmNMF$_{d21}$ F14^{hi}C0 regions (region of interest, ROI) were computed. Distance information from all tissues containing a ROI was extracted using 'semla'[65] (v. 1.1.6; R v. 4.2.3; Seurat v. 4.3.0.1) with the 'RadialDistance' function, where singletons were excluded in the human analysis. For human IPF data, distance correlation coefficients (Pearson) were computed for the 1,000 most variable genes within 500 μm from the ROI. P values were corrected using the Benjamini−Hochberg method. The cell-type density correlations were similarly analyzed. The cell density and gene expression fluctuations were visualized using the 'geom_smooth' function (ggplot2, v. 3.4.0) with GAM and formula 'y - s(x, bs = 'cs')'. For mouse BLM d21 data, 'geom_smooth' with the locally estimated scatterplot smoothing method ('loess') was used for visualization of cell density data.

IPF fibrotic niche regulators and cell–cell communication. To investigate the microenvironment surrounding hsNMF-F14^{hi}C0 spots, the nearest n.b. (≤2 spot distance) were identified using the 'RegionNeighbours' STUtility function. The neighboring spots were clustered using

PCA (PCs 1–9) through 'FindNeighbors' and 'FindClusters' (resolution of 0.2), resulting in six clusters (n.b. clusters 0–5). The marker genes were determined using 'FindAllMarkers', comparing each cluster to all others. Due to their low abundancies, n.b. clusters 3–5 were omitted from some downstream analyses.

Upstream regulators and active pathways for n.b. clusters (0–2) were predicted with Ingenuity Pathway Analysis (IPA; version 90348151, Ingenuity Systems, Qiagen) based on marker genes (adjusted $P < 0.01$). The marker genes for hsNMF-F14^{hi}C0, generated by comparing hsNMF-F14^{hi}C0 spots against all others in the IPF Visium subset using 'FindMarkers' ('min.pct = 0.25' and 'min.diff.pct = 0.1'), were also analyzed. Results were assessed with multienrichjam (v. 0.0.72.900)[71], visualizing the top 20 upstream regulators and top ten pathways and diseases/functions.

Directional cell–cell communication between n.b. clusters was analyzed using NicheNet (v. 1.1.1)[41]. This method predicts ligand–target links from gene expression and incorporates intracellular signaling. Ligand–receptor interactions ('lr_network.rds'), ligand–target gene regulatory potential scores ('ligand_target_matrix.rds') and weighted ligand-signaling and gene regulatory networks ('weighted_networks. rds') were retrieved from the NicheNet data repository (https://doi.org/10.5281/zenodo.3260758). Analyses were conducted in four rounds, alternating receiver and sender populations: (1) F14^{hi}C0 with n.b. clusters 0–5 as senders, (2) n.b. cluster 0 with F14^{hi}C0 as sender, (3) n.b. cluster 1 with F14^{hi}C0 as sender and (4) n.b. cluster 2 with F14^{hi}C0 as sender. In each round, receiver genes were identified using data from spots not included in the analysis (of IPF and HC origin). The results across all rounds were compiled, and the top prioritized ligands (average correlation >0.075) from round 1 were used to visualize corresponding results in subsequent rounds and across all relevant clusters.

Spatial cell-type compartmentalization in mouse. Cell-type colocalization compartments were identified in mouse Visium data using the Strunz (2020) cell2location results. Cell types labeled as 'NA' and 'low.quality.cells' were excluded. Visium spot data were grouped into vehicle (d7 and d21), BLM d7 and BLM d21 subsets, and pairwise correlations (Pearson) were computed for each cell type across all spots within each group. Hierarchical clustering defined compartments with a tree height (h) cutoff of 1.5. A Sankey diagram visualized cell types in each compartment for each subset. For BLM d21 compartments F, H and G, spot-wise compartment scores were calculated by summing the inferred densities for all cell types within each compartment.

Translational analyses of AbBa clusters. Samples IPF 3 B1–B3, IPF 4 B1–B3 and BLM d21 animals 1–5 were selected for analysis, based on their pronounced fibrosis and AbBa cell-dense regions. The raw counts were filtered to include only orthologous genes, identified by orthogene[69]. A new assay was created from the filtered data for separate normalization of human and mouse datasets, followed by integration using Seurat's 'FindIntegrationAnchors' and 'IntegrateData'. Anchor features were selected by 'SelectIntegrationFeatures'. The marker genes for hsNMF-F14^{hi}C0 and mmNMF$_{d21}$ F14^{hi}C0 were separately identified using 'FindMarkers', comparing against all other same-species spots, using the integrated genes (Supplementary Table 10).

Marker genes (Bonferroni adjusted $P < 0.05$) were analyzed for upstream regulator and canonical pathway enrichment using IPA. Comparisons across species were conducted with multienrichjam (ref. 71) (v.0.0.72.900) to identify shared and unique regulators and pathways. The most significant findings ($P < 10^{-7}$ for regulators and $P < 10^{-4}$ for pathways, right-tailed Fisher's exact test) were visualized in clustered network (cnet) plots, grouping predicted molecules in clusters ('Nodes') based on shared marker genes.

Lymphocyte aggregate comparison. To identify gene signatures overlapping histological findings of lymphocyte aggregates, the human IPF

data was split by donor and individually processed using 'SCTransform' and NMF, to produce 30 new subject-specific factors for each donor (hsNMF$_{IPF}$). For the mouse data, the NMF results produced for each time point was used (mmNMF$_{d7}$ and mmNMF$_{d21}$).

Activities of the top 100 contributing genes for each factor were scaled between 0-1. Cell densities and detection rates were estimated for spots within the 99th percentile of factor activity. Some mouse cell types that lacked human comparatives or relevant signals were excluded from visualization ('AM (BLM)', 'AM (PBS)', 'non classical monocytes (*Ly6c2*-)', '*Fn1*$^+$ macrophages', 'M2 macrophages', 'Themis T cells' and 'T cell subset'). Average cell densities and detection rates in spots with density score >0.5 were calculated and visualized.

Spatial cell colocalization trajectory analysis. Inferred cell-type densities of 'AT2 cells', 'activated AT2 cells', '*Krt8*$^+$ ADI' and 'AT1 cells' (mouse) or 'AT2', 'Transitional AT2', '*KRT5*$^-$/*KRT17*$^+$' and 'AT1' (human) were used to select spots in the 95th percentile of abundance for uniform manifold approximation and projection (UMAP) dimensionality reduction (n.neighbors = 30, min.dist = 0.1) and to generate low-resolution clusters with 'FindNeighbors' and 'FindClusters' (mouse: resolution of 0.2, human: resolution of 0.1), identifying clusters corresponding to AT2-dense spots ('AT2-cluster'). Slingshot[72] (v. 1.8) trajectory analyses were applied to each UMAP embedding with 'getLineages' assigning AT2-clusters as starting points. The curves were extrapolated using 'getCurves' (approx_points = 300, thresh = 0.01, stretch = 0.8, allow. breaks = FALSE, shrink = 0.99). Pseudotime was estimated by integrating gene counts (detected (greater than five transcripts) in ≥1% of spots) and Slingshot curves into a negative binomial GAM using 'fitGAM' (tradeSeq, v. 1.4.0)[73]. Two curves were identified in the human data, and the visualized pseudotime represents the maximum values of the two curves.

Statistics and reproducibility

Statistical methods for each analysis are detailed in their respective method sections. No statistical method was used to predetermine sample size. For human samples, it was maximized based on availability, tissue quality, cost and throughput, and the number of donors aligns with similar spatial transcriptomics studies[20,56,74]. The number of BLM-challenged mice exceeded that of controls to account for possible study dropouts due to adverse effects; however, all mice met the inclusion criteria and were retained. Technical replicates were included for most human and mouse samples and produced consistent results, ensuring reproducibility. Investigators were not blinded to disease status or treatment during experiments and analyses.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

All Visium datasets, including Space Ranger output and high-resolution images, needed to replicate and further build upon the presented analyses have been deposited to BioStudies under accession numbers S-BSST1410 for the human data and S-BSST1409 for the mouse data. These repositories also include sample metadata, processed Seurat/STUtility objects containing deconvolution and clustering results, and cell2location results for all samples. The raw and processed RNA-sequencing data for the mouse samples have been deposited to the GEO under accession number GSE267904. The processed data (Space Ranger output) from the human samples have also been deposited to ArrayExpress under accession number E-MTAB-14121. Due to Swedish law, patient consent and the risk of finding personally identifiable information within the sequencing data, the raw RNA-sequencing read data for the human samples cannot be shared. Source data are provided with this paper.

Code availability

The code used for downstream data analyses performed in R that was used to produce all analyses and figures presented in the article is freely available at the GitHub repository via https://github.com/lfranzen/spatial-lung-fibrosis, with a mirrored deposition to Zenodo under https://doi.org/10.5281/zenodo.11193764 (ref. 75).

References

63. Bergenstråhle, J., Larsson, L. & Lundeberg, J. Seamless integration of image and molecular analysis for spatial transcriptomics workflows. *BMC Genomics* **21**, 482 (2020).
64. Stuart, T. et al. Comprehensive integration of single-cell data. *Cell* **177**, 1888–1902.e21 (2019).
65. Larsson, L., Franzén, L., Ståhl, P. L. & Lundeberg, J. Semla: a versatile toolkit for spatially resolved transcriptomics analysis and visualization. *Bioinformatics* **39**, btad626 (2023).
66. Smedley, D. et al. BioMart—biological queries made easy. *BMC Genomics* **10**, 22 (2009).
67. Hafemeister, C. & Satija, R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol.* **20**, 296 (2019).
68. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
69. Schilder, B.M. & Skene, N.G. Orthogene: an R package for easy mapping of orthologous genes across hundreds of species. *Bioconductor* https://doi.org/10.18129/B9.bioc.orthogene (2022).
70. Kolberg, L., Raudvere, U., Kuzmin, I., Vilo, J. & Peterson, H. gprofiler2–an R package for gene list functional enrichment analysis and namespace conversion toolset g:Profiler. *F1000Res* **9**, ELIXIR-709 (2020).
71. Ward, J. M. Multienrichjam: analysis and visualization of multiple gene set enrichments. *GitHub* https://github.com/jmw86069/multienrichjam (2023).
72. Street, K. et al. Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genomics* **19**, 477 (2018).
73. Van den Berge, K. et al. Trajectory-based differential expression analysis for single-cell sequencing data. *Nat. Commun.* **11**, 1201 (2020).
74. Sounart, H. et al. Dual spatially resolved transcriptomics for human host-pathogen colocalization studies in FFPE tissue sections. *Genome Biol.* **24**, 237 (2023).
75. Franzén, L. lfranzen/spatial-lung-fibrosis: Zenodo v1.0.0. *Zenodo* https://doi.org/10.5281/zenodo.11193764 (2024).

Author contributions

J.H., M.S., A.O., G.B., S.J., P.L.S. and J. Lundeberg, conceived the study. L.F., M.O.L., S.J. and M.S. planned and designed the experiments. T.V. and A.B. provided animals from the BLM mouse model. A.C., S.O. and M.O.L. collected the mouse lung tissue. M.O.L., V.P. and L.F. assessed tissue quality and performed initial expression analyses. S.J. and L.F. identified and selected the human tissues. L.F. and M.O.L. carried out tissue sectioning and spatial gene expression experiments. J. Lindgren and M.O.L. sequenced the samples. L.S. performed histopathological annotations. B.P.K. processed the raw Visium data and performed cell-type deconvolution. L.F. and M.O.L. carried out computational analyses of the Visium data. Data interpretation was done by M.O.L., L.F., M.H., V.P., M.S., A.O., P.L.S. and J.H. M.S., P.L.S., J.H. and A.O. supervised the project. L.F. created the final figures and illustrations. M.O.L. and L.F. drafted the manuscript with input from M.H., V.P., M.S., A.O., P.L.S., L.S., G.B., B.K., T.V., A.B., S.J. and J.H. All authors read and approved the manuscript.

Competing interests

The authors declare the following competing interests: P.L.S. and J. Lundeberg are scientific consultants to 10x Genomics. The remaining authors (L.F., M.O.L., M.H., V.P., L.S., B.P.K., A.C., S.O., T.V., A.B., J. Lindgren, G.B., S.J., A.O., M.S. and J.H.) are employees and/or stockholders of AstraZeneca.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41588-024-01819-2.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41588-024-01819-2.

Correspondence and requests for materials should be addressed to Marianna Stamou or Patrik L. Ståhl.

Peer review information *Nature Genetics* thanks Kerstin Meyer and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.
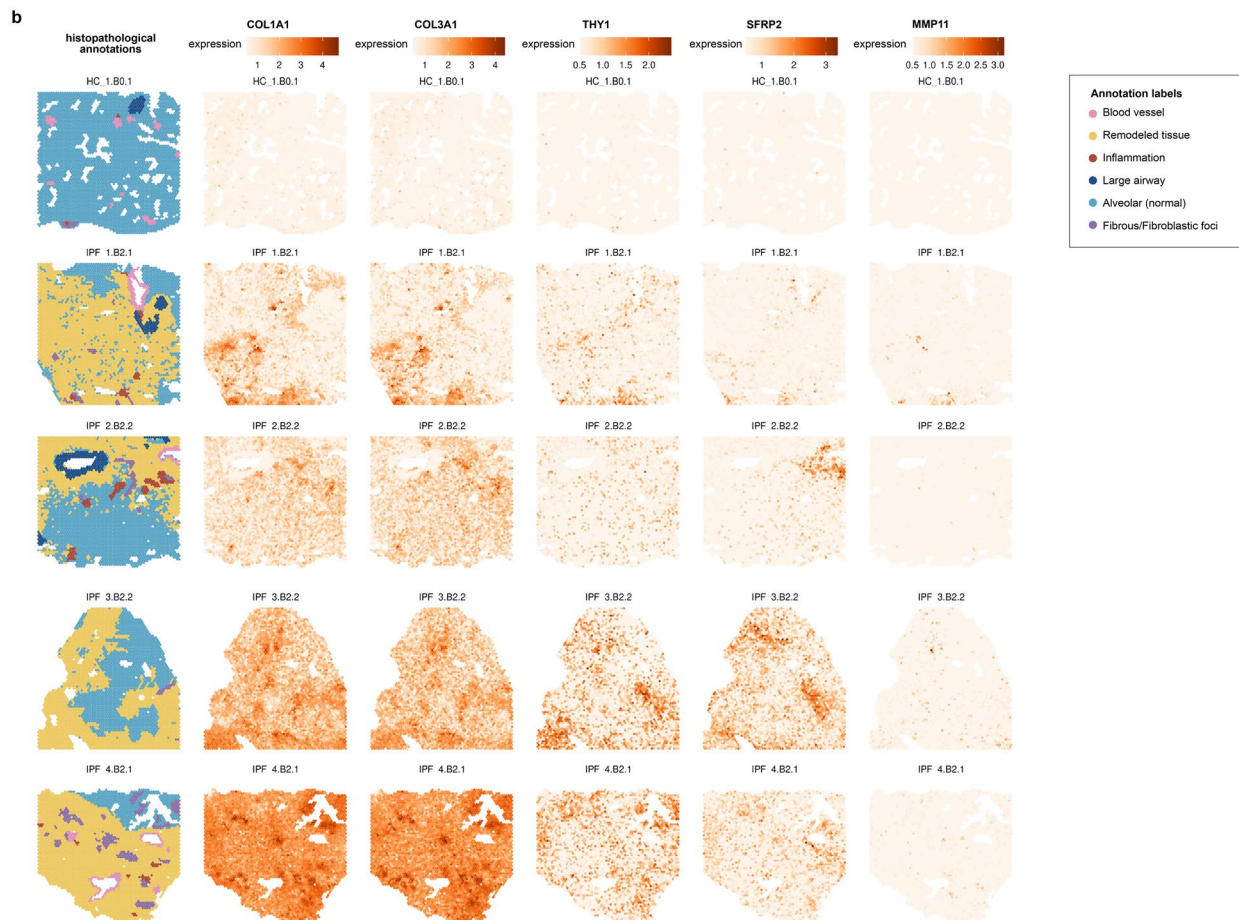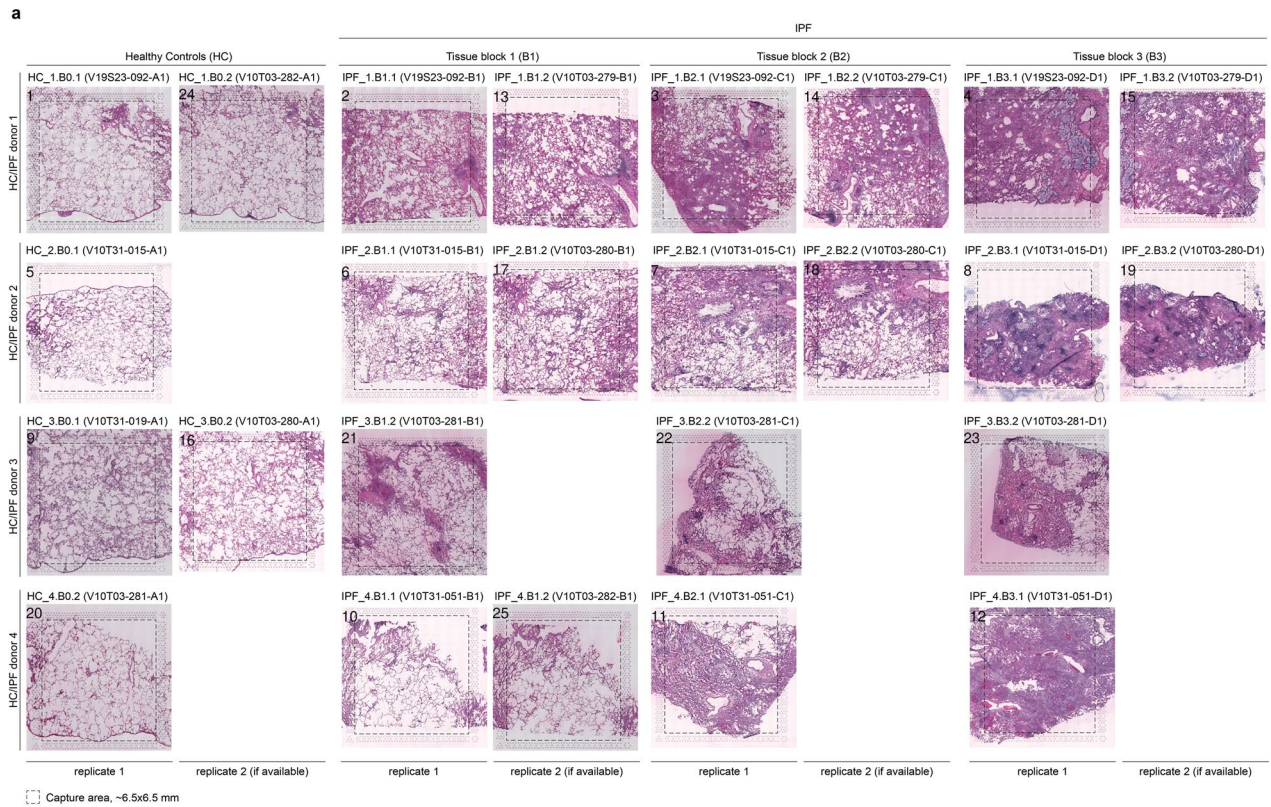
Extended Data Fig. 1 | See next page for caption.

Extended Data Fig. 1 | Histological and spatial gene expression analysis of human lung tissues. a) Hematoxylin and eosin (H&E) stained sections of lung tissue samples from four healthy controls (HC 1-4; 'B0') and four IPF patients (IPF 1-4), used for Visium experiments. Each IPF patient sample includes three samples ('B1', 'B2', 'B3') representing a gradient in the extent of fibrotic injury, from mild to severe tissue remodeling. Visium data from a total of 25 tissue sections were analyzed, including replicate sections for some tissues to control for technical variability. **b)** Histopathological annotations were performed on the H&E-stained sections, here showing the annotations in sections from one HC and each IPF donor, displaying the extent of tissue remodeling. Differential expression analysis performed on pseudo-bulk Visium data between IPF and HC identified several extracellular matrix and fibrosis-related genes, and when plotted spatially, their expression largely originates from regions of tissue fibrosis.
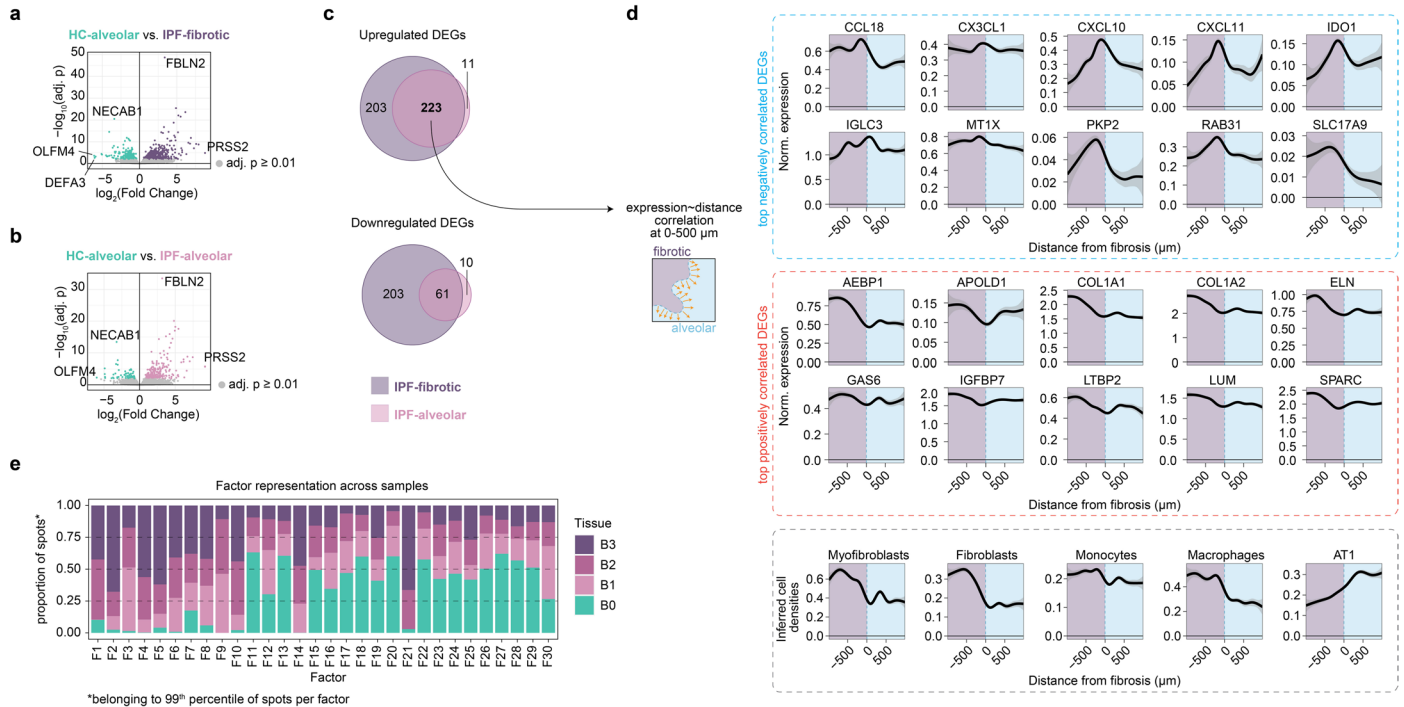
Extended Data Fig. 2 | Differential gene expression in IPF and NMF factor distribution. a) Differential gene expression between annotated fibrotic regions of IPF lungs and HC samples. **b)** Differential gene expression between annotated alveolar regions from IPF lungs and samples from healthy control (HC) lungs. **c)** Overlap of significantly (FDR-adjusted p-value < 0.01) upregulated (top) and downregulated (bottom) DEGs in IPF alveolar and fibrotic regions compared to HC. **d)** The spatial gene expression of the 223 overlapping upregulated DEGs were used to calculate correlation (Pearson) values from the fibrotic border distances using gene expression against the 0-500 µm distances. The gene expression for the top 10 significantly (Benjamini–Hochberg adj. p < 0.05) up and down

correlated genes, as well as the inferred cell densities of selected cell types, were visualized for the 2000 µm distances centered around the fibrotic border using a generalized additive model (GAM) for the regression line. Gray areas correspond to 95% confidence intervals. **e)** NMF factorization was performed on all human Visium samples. Spots with the highest activity (99th percentile; F^{hi}) for each factor were identified and used to calculate the proportion of F^{hi} spots among all spots, grouped by tissue block grading (B0-B1) and displayed as a stacked bar chart. NMF, non-negative matrix factorization; DEG, differentially expressed gene; HC, healthy controls.
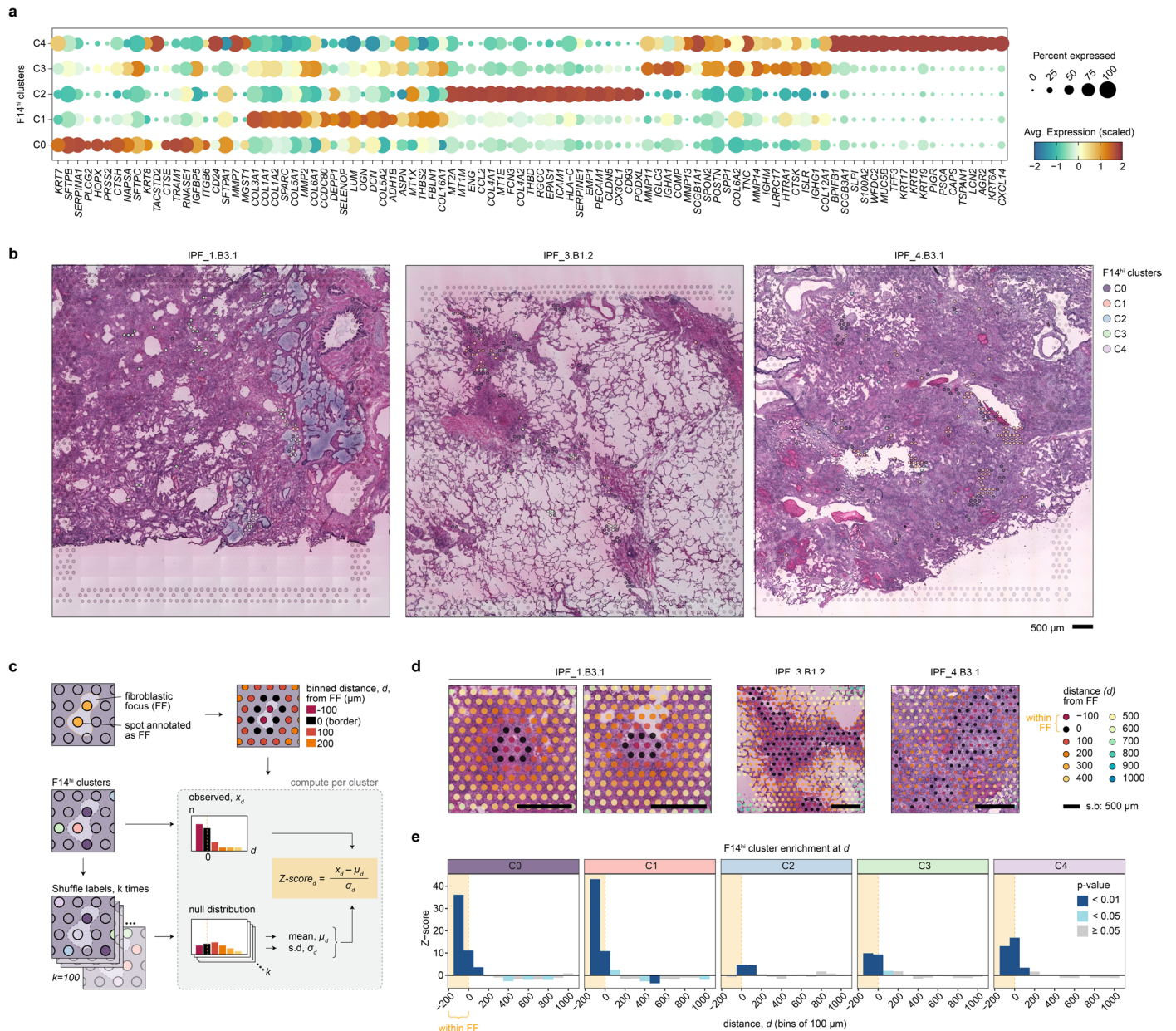
Extended Data Fig. 3 | F14^hi clusters in human lung tissues. a) Gene expression profiles for F14^hi cluster (C0-C4) top marker genes. Color scale represents average scaled expression, and dot size reflects the proportion of spots expressing each gene. b) Spatial distribution of F14^hi clusters C0-C4 overlaid on H&E images of three human IPF lung sections. Similar spatial patterns were observed in all IPF sections. c) Approach for calculating spatial enrichment of each F14^hi clusters at distances from FF. The radial distances, d, from the FF were extracted for all spots, where 0 μm corresponds to the outer FF border. The observed F14^hi cluster spot locations were registered at each distance (binned into 100 μm to cover each row of spots), and the F14^hi cluster labels were shuffled across all spots within in each sample k=100 times to produce a null distribution

(that is, how each cluster would be distributed if there were no spatial patterns). For each cluster, the observed count at every distance (x_d) and the mean (μ_d) and standard deviation (s.d., σ_d) across all randomization rounds were used to compute a Z-score per d. d) Spatial overlay of the FF distance-colored spots over the H&E-stained tissue image for a few selected areas. e) Enrichment of each F14^hi cluster at distances (μm) from FF using a Z-score computed with a randomized spatial cluster distribution as the baseline. Two-sided p-values were calculated using the Z-scores, and the bars are colored based on their significance levels (dark blue: p < 0.01; light blue: p < 0.05). FF, fibroblastic focus; H&E, Hematoxylin and eosin.

a

day 7 (d7)

day 21 (d21)

vehicle control

1 d7.Veh.1 (V10A20-072-A1) | 2 d7.Veh.2 (V10A20-074-A1) | 3 d7.Veh.3 (V10A20-073-A1) | 13 d21.Veh.1 (V10A20-071-A1) | 14 d21.Veh.2 (V10A20-052-A1) | 15 d21.Veh.3 (V10A20-053-A1)

4 d7.BLM.1 (V10A20-072-B1) | 5 d7.BLM.2 (V10A20-072-C1) | 6 d7.BLM.3 (V10A20-072-D1) | 16 d21.BLM.1 (V10A20-071-B1) | 17 d21.BLM.2 (V10A20-071-C1) | 18 d21.BLM.3 (V10A20-071-D1)

bleomycin

7 d7.BLM.4a (V10A20-074-B1) | 10 d7.BLM.4b (V10A20-073-B1) | 19 d21.BLM.4a (V10A20-052-B1) | 22 d21.BLM.4b (V10A20-053-B1)

8 d7.BLM.5a (V10A20-074-C1) | 11 d7.BLM.5b (V10A20-073-C1) | 20 d21.BLM.5a (V10A20-052-C1) | 23 d21.BLM.5b (V10A20-053-C1)

9 d7.BLM.6a (V10A20-074-D1) | 12 d7.BLM.6b (V10A20-073-D1) | 21 d21.BLM.6a (V10A20-052-D1) | 24 d21.BLM.6b (V10A20-053-D1)

1 mm

b

c Differential Expression Analysis (pseudo-bulk)

d

Node D

Cell surface interactions at the vascular wall
IL-17A Signaling in Fibroblasts
Apelin Liver Signaling Pathway
Atherosclerosis Signaling
Binding and Uptake of Ligands by Scavenger Receptors
Immunoreg. interactions between Lymphoid and non-Lymphoid cell
Signaling by PDGF
IL-4 Signaling
NCAM signaling for neurite out-growth
Signaling by MET
Syndecan interactions
Degradation of the extracellular matrix
Activin Inhibin Signaling Pathway
HOTAIR Regulatory Pathway
Tumor Microenvironment Pathway

Node B

Post-translational protein phosphorylation
Regulation of IGF transport and uptake by IGFBPs
Surfactant metabolism
p53 Signaling
Transcriptional regulation by RUNX2
ID1 Signaling Pathway
Interleukin-4 and Interleukin-13 signaling
Inhibition of Matrix Metalloproteases
Role of Osteoblasts in Rheumatoid Arthritis (RA) Sign. Pathway
Role of Osteoblasts, Osteoclasts and Chondrocytes in RA
Elastic fibre formation
Glycosaminoglycan metabolism
Granulocyte Adhesion and Diapedesis

Node A

Pulmonary Fibrosis Idiopathic Sign. Pathway
Role of Osteoclasts in Rheumatoid Arthritis Sign. Pathway
Assembly of collagen fibrils and other multimeric structures
GP6 Signaling Pathway
Wound Healing Signaling Pathway
Collagen degradation
Pathogen Induced Cytokine Storm Signaling Pathway
Collagen biosynthesis and modifying enzymes
Collagen chain trimerization
Microautophagy Signaling Pathway
Hepatic Fibrosis / Hepatic Stellate Cell Activation
Neutrophil Extracellular Trap Signaling Pathway

Canonical pathways

● hs F14hiC0
● mm F14hiC0

Node C

Extracellular matrix organization
Integrin cell surface interactions

e d21 BLM mouse lungs

animal 1 (d21.BLM.1) | animal 2 (d21.BLM.2)

animal 3 (d21.BLM.3) | animal 4 (d21.BLM.4a)

animal 5 (d21.BLM.5b) | animal 6 (d21.BLM.6a)

● mmF14hi C0   ○ fibrotic tissue
● mmF14hi C1   ○ rest
● mmF14hi C2

f

d21 BLM animal 4

mmF14hi cluster

distance (d) from fibrosis (100 μm bins)

-300 / -200 / -100 / 0 / 100 / 200 / 300 / 400 / 500 / 600 / 700 / 800 / 900 / 1000 / 1100

within fibrosis — border — outside fibrosis

Extract distances from fibrotic areas

compute per mmF14hi cluster

observed, x_d

○ fibrotic tissue ○ rest
● mmF14hi C0

● C0 ● C2
● C1 ○ other

500 μm

Shuffle spot label identities (k=100)

Count spots per mmF14hi cluster, per d

$$Z\text{-}score_d = \frac{x_d - \mu_d}{\sigma_d}$$

null distribution

→ mean, μ_d
→ s.d, σ_d

g Enrichment of clusters at a distance from fibrotic tissue

mmF14hi C0

mmF14hi C1

mmF14hi C2

p-value
■ < 0.01
■ < 0.05
■ ≥ 0.05

Z-score

distance, d (bins of 100 μm)

within fibrotic tissue

Extended Data Fig. 4 | See next page for caption.

Extended Data Fig. 4 | Spatial and molecular analysis of BLM-induced mouse lung fibrosis. a) H&E-stained sections of mouse lung tissues collected at day 7 (d7) and day 21 (d21) post-BLM or saline vehicle treatment that were used for Visium analysis. **b)** Summarizing statistics of Visium spot count used and histopathological annotations. Center line, median; box limits, upper and lower quartiles; whiskers, 1.5× interquartile range. **c)** Volcano plots depicting differential expression analysis results between BLM and vehicle-treated lungs at d7 and d21, highlighting genes with significant (adj. p < 0.01) changes in expression. **d)** Network of top significant (p value < 0.0001) canonical pathway enrichment results based on human hsNMF-F14^{hi} C0 and mouse mmNMF$_{d21}$-F14^{hi} C0 marker genes (adj. p < 0.05), with specific pathways within each node cluster annotated. Inner nodes illustrate groups of regulators sharing genetic influences, and outer nodes represent contributing marker genes. **e)** Spatial plots of mmNMF-F14^{hi} C0-C2 and regions histologically annotated as fibrotic tissue (gray) for all the BLM-challenged lungs collected at d21. **f)** To quantify the spatial localization of the mmNMF-F14^{hi} clusters in relation to fibrotic tissue, the same strategy as presented in Extended Data Fig. 2h was employed, except for using regions annotated as fibrotic to extract radial distances. **g)** Z-scores for each cluster plotted across each distance bin, relating to the distance from fibrotic regions (yellow shaded areas) across all tissue samples. P-values were computed from the Z-scores using a two-tailed test (dark blue: p < 0.01; light blue: p < 0.05; gray: ≥ 0.05). mmF14^{hi} C0 was significantly enriched within the border of fibrotic areas, especially at 150-350 μm distances into the fibrotic tissue. On the other hand, there was an underrepresentation of mmF14^{hi} C0 spots in areas outside the fibrosis. H&E, Hematoxylin and eosin.
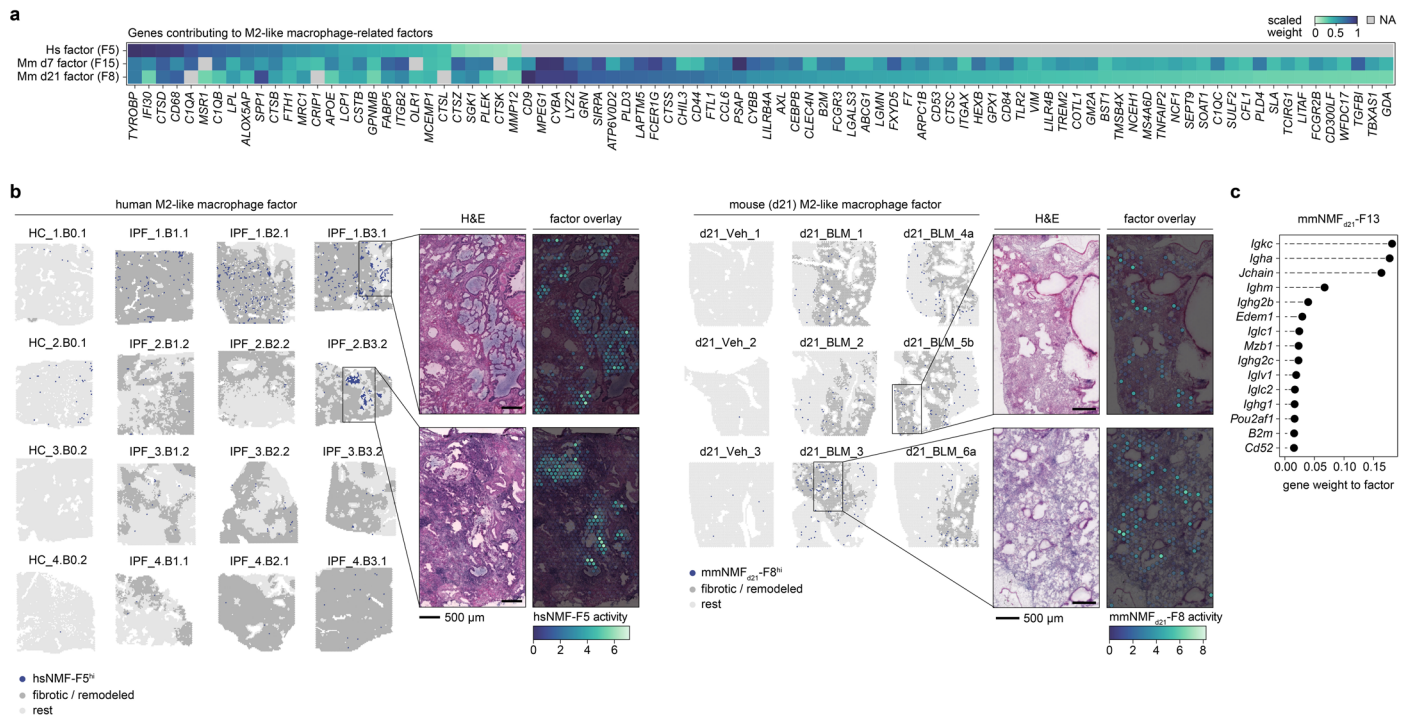
Extended Data Fig. 5 | Profibrotic macrophages and immune cell profiles.
a) Heatmap showing the gene contributions to the identified M2-like macrophage-related factors across human lungs (hsNMF-F5) and mouse lungs at day 7 (d7) (mmNMF$_{d7}$-F15) and day 21 (d21) (mmNMF$_{d21}$-F8). The visualized genes are those, among the top 100 contributing genes, that overlapped between at least two of the three groups. Color represents each gene's factor contribution as their scaled weight, where 1 is assigned to the strongest contributing gene to the factor out of all genes in the data set. **b)** Spatial distribution of macrophage-associated factors in human (hsNMF-F5) and mouse d21 (mmNMF$_{d21}$-F8) tissue

sections. The top 99th percentile of either factor was labelled as hsNMF-F5^{hi} or mmNMF$_{d21}$-F8^{hi} and visualized together with the histopathological spot annotations grouped into 'fibrotic / remodeled' tissue (gray) and remaining tissue (light gray). Zoomed in graphs display factor activity overlaid the H&E images for selected regions that demonstrated high factor activity in areas of bronchial epithelium within fibrotic tissue. **c)** Top 15 contributing genes to mmNMF$_{d21}$-F13, which was identified to contain a strong plasma cell profile driven by *Igha*, among other immunoglobulin-related genes. H&E, Hematoxylin and eosin.

nature portfolio

Corresponding author(s):	Marianna Stamou
	Patrik L Ståhl
Last updated by author(s):	May 23, 2024

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's d, Pearson's r), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	No specific software was used for data collection. The datasets were processed and analyzed using the open source software listed below.
Data analysis	The following software, analysis packages, and code was used: Space Ranger (v. 1.2.2) cell2location (v. 0.1) R (v. 4.0.5 and v. 4.2.3) STUtility (v. 1.1.1) Seurat (v. 4.1.1 and v. 4.3.0.1) NNLM (v. 0.4.4) DESeq2 (v. 1.30.1) gprofiler2 (v. 0.2.1) orthogene (v. 1.4.2) semla (v. 1.1.6) ggplot2 (v. 3.4.0) NicheNet (v. 1.1.1) multienrichjam (v.0.0.72.900) Slingshot (v. 1.8) tradeSeq (v. 1.4.0) Ingenuity Pathway Analysis (v. 90348151) Loupe Browser (v. 6)

All the code used for downstream data analyses performed in R (used to produce all figures presented in the article) are available at the GitHub repository https://github.com/lfranzen/spatial-lung-fibrosis and deposited to Zenodo with DOI 10.5281/zenodo.11193764.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:
- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Sequencing data was generated using the NovaSeq 6000 (Illumina) system with the following set-up: Read1: 28 bp, Index 1: 10 bp, Index 2: 10 bp, Read2: 90 bp. Human and mouse libraries were sequenced separately over S4 flowcells. FastQ files were processed using the Space Ranger 1.2.2 (10x Genomics) pipeline. Sequencing reads were mapped to their respective reference genomes GRCh38 (human) and mm10 (mouse).

The publicly available single cell data used for cell type deconvolution were retrieved from Gene Expression Omnibus (GEO) database, from studies GSE135893 (human, IPF) and GSE141259 (mouse, bleomycin).

All processed Visium datasets reflecting the "minimum datasets" from the human and mouse lung samples have been deposited to BioStudies under accession numbers S-BSST1410 (https://doi.org/10.6019/S-BSST1410) (human data) and S-BSST1409 (https://doi.org/10.6019/S-BSST1409) (mouse data). The deposited data includes Space Ranger output files, full resolution H&E images, spot alignment files, sample meta data, cell2location results, and Seurat/STUtility objects.

Raw and processed RNA-sequencing data (FASTQ files) for the mouse Visium samples have been deposited to Gene Expression Omnibus (GEO) under accession number GSE267904. The processed Space Ranger output data from all human samples have also been deposited to ArrayExpress under accession number E-MTAB-14121. Raw sequencing data for human samples cannot be made available due to Swedish law, patient consent, and the risk of finding personally-identifiable information within the sequencing data.

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

| Reporting on sex and gender | Six donors were male and two donors (one control, one IPF) were female. Due to the restricted availability of human biological samples, we were not in a position to select samples based on sex. Information about donor characteristics, including gender, is available in Supplementary Table 1.

In the analysis of the human Visium data, genes mapped to the X and Y chromosomes were excluded from the downstream analyses in order to avoid biases in the results based on donor sex. |
|---|---|
| Reporting on race, ethnicity, or other socially relevant groupings | No information regarding race or ethnicity was documented |
| Population characteristics | Information regarding age, sex, and smoking status of the organ donors have been listed in Supplementary Table 1. |
| Recruitment | Tissues from lung organ donors with healthy lungs were collected post-mortem and tissue from patient donors with pulmonary fibrosis were collected during lung transplant or resection at the Sahlgrenska University Hospital (Gothenburg Sweden). Lung tissues included in the current study were thereafter selected based on tissue availability and quality. |
| Ethics oversight | Samples from pulmonary fibrosis patients were acquired with approval by the local human research ethics committee (Gothenburg, permit number 1026-15) and participants gave written informed consent prior to inclusion. Healthy lung tissue was obtained from deceased donors acquired with approval by the local human research ethics committee (Lund, permit number Dnr 2016/317). All donors were anonymized and any derived samples or data cannot be traceable to the donor. |

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

| Sample size | No statistical method was used to predetermine sample size. Instead, the size was optimized based on practical constraints including availability of suitable tissue, cost considerations, and experimental throughput. This approach is in line with similar spatial transcriptomics studies. For the human samples, a motivation for sample size selection is outlined in Supplementary Note 1.

For the human study, samples from eight organ donors were used, with an equal split between controls and idiopathic pulmonary fibrosis (IPF) patients, ensuring balanced representation. Within each IPF patient, three distinct tissue samples representing mild, moderate, and severe fibrosis were collected to capture the heterogeneity of the disease across different stages within the same individual. All tissues were analyzed in duplicate sections, with the highest quality samples selected for further analysis. This led to the inclusion of 25 Visium sections representing 16 unique tissues, resulting in 101,075 spatial spots analyzed in the human dataset.

Mouse lung samples from a total of 18 animals were collected. Tissues were collected at day 7 (bleomycin n=6; control n=3) or day 21 (bleomycin n=6; control n=3) following bleomycin or saline (vehicle control) challenge. The inclusion of 6 mice in the bleomycin groups per timepoint was selected to take hight for the potentially drop-out rate of mice injured by bleomycin, however, all mice fulfilled the inclusion criteria and could be used for the study. Three bleomycin-injured lungs per timepoint were analyzed in technical replicates, resulting in a total of 24 Visium sections. This yielded a total of 91,311 spatial spots across all samples available for analysis in the mouse cohort.

We aimed for a minimum of three donors/animals per condition and group, consistent with typical experimental design in spatial transcriptomics, to ensure sufficient statistical power without predetermined calculations. |

| Data exclusions | Visium data was filtered on a spot and gene basis, based on selected thresholds and gene lists outline in detail in the Methods.

For the human lung tissue Visium data, six replicate sections were excluded from downstream analyses due to poor technical quality (low sequencing mapping results and high degree of lateral diffusion of transcripts). The remaining dataset included data from all eight individuals, and all selected tissues (B1-B3) in the IPF lung samples. |

| Replication | We implemented several measures to verify the reproducibility of the experimental findings. Both technical and biological replicates were included in the Visium spatial gene expression analysis for human and mouse samples. Specifically, technical replicates consisted of tissue sections processed on different Visium slides. Consistency in results across these technical replicates was confirmed, with comparable data clustering outputs observed, in the cases where both replicates passed the quality control check. Biological replication was achieved using samples from multiple donors or animals, and key findings such as gene expression patterns and spatial distribution were reproducible across these biological replicates where similar histopathological features were captured within the studied tissue section.

All replication attempts described were successfully reproduced, affirming the reliability of our findings. We have further supported reproducibility by providing open access to all processed data and custom code necessary to replicate our results and figures (see Data and Code Availability statements), ensuring that our findings can be verified by independent analyses. |

| Randomization | Randomization was not applicable to the human study due to the specific constraints of working with biobanked samples. The samples from IPF patients and healthy controls were selected based on the availability of fresh-frozen lung tissues that met our quality criteria and for which explicit consent had been obtained to conduct genetic studies. As these samples were derived from a biobank (AstraZeneca Biobank), the selection was inherently dependent on what was available and suitable at the time of the study, rather than being randomly assigned from a larger pool.

For the mouse bleomycin study, conditions allowed for implementation of randomization. Upon arrival, mice were randomly assigned to different treatment groups to ensure that the experimental conditions were evenly distributed and that the results could be generalized across the population studied |

| Blinding | Blinding was not implemented in the Visium spatial gene expression analyses due to the intrinsic characteristics of the experimental design and the analysis techniques used. The nature of the Visium technology involves spatial mapping of gene expression directly on tissue sections, where the outcomes (i.e., gene expression patterns) are inherently linked to visible, histological features of the tissue. In our experimental setup, the researchers needed to be aware of the specific tissue regions being analyzed to accurately interpret the spatial gene expression data in relation to the tissue morphology. Moreover, the impact of fibrosis in the tissue causes visible histopathological alterations which makes it impossible for blinding of disease condition in a sample. Blinding would further obscure the link between histopathology and gene expression patterns, potentially compromising the accuracy and relevance of the spatial annotations and the subsequent data interpretation. |

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒ ☐	Antibodies
☒ ☐	Eukaryotic cell lines
☒ ☐	Palaeontology and archaeology
☐ ☒	Animals and other organisms
☒ ☐	Clinical data
☒ ☐	Dual use research of concern
☒ ☐	Plants

Methods

n/a	Involved in the study
☒ ☐	ChIP-seq
☒ ☐	Flow cytometry
☒ ☐	MRI-based neuroimaging

Animals and other research organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research, and [Sex and Gender in Research](#)

Laboratory animals	C57BL/6NCrl mice were purchased from Charles River, Germany. At arrival to AstraZeneca R&D Gothenburg (Sweden), the mice were 8 weeks old. After an acclimatization period of 5 days, the study was initiated and the mice were challenged with bleomycin dissolved in saline or saline. Lung samples were thereafter collected at day 7 or day 21 following bleomycin or saline challenge.
Wild animals	The study did not involve wild animals.
Reporting on sex	All mice included in the study (n=18) were of female sex. Sex-based variation in the bleomycin mouse model of pulmonary fibrosis has been seen reported previously, and within AstraZeneca R&D Gothenburg this model has been optimized for female mice.
Field-collected samples	The study did not involve samples collected from the field.
Ethics oversight	Animal handling conformed to standards established by the Council of Europe ETS123 AppA, the Helsinki Convention for the Use and Care of Animals, Swedish legislation, and AstraZeneca global internal standards. All mouse experiments were approved by the Gothenburg Ethics Committee for Experimental Animals in Sweden and conformed to Directive 2010/63/EU. The present study was approved by the local Ethical committee in Gothenburg (EA000680-2017) and the approved site number is 31-5373/11.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Plants

Seed stocks	*Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.*
Novel plant genotypes	*Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.*
Authentication	*Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosiacism, off-target gene editing) were examined.*