Temporo-spatial cellular atlas of the regenerating alveolar niche in idiopathic pulmonary fibrosis

Praveen Weeratunga[1], Bethany Hunter[2], Martin Sergeant[3], Joshua Bull,[4] Colin Clelland[5], Laura Denney[1], Chaitanya Vuppusetty[1], Rachel Burgoyne[2], Jeongmin Woo[1,3], Tian Hu[1], Lee Borthwick[2], James Shaw[7,8], Agne Antanaciuvete[1,3], Andrew Filby[2], Helen Byrne[4,6], Andrew Fisher[7,8], Ling-Pei Ho[1,8*]

1. MRC Translational Immune Discovery Unit, Weatherall Institute of Molecular Medicine, University of Oxford, Oxford UK
2. Newcastle University Biosciences Institute, Newcastle upon Tyne, UK
3. MRC WIMM Centre for Computational Biology, Weatherall Institute of Molecular Medicine, University of Oxford, UK
4. Wolfson Centre for Mathematical Biology, Mathematical Institute, University of Oxford, UK
5. Anatomic Pathology, Weill Cornell Medical College, Doha, Qatar
6. Ludwig Institute for Cancer Research, University of Oxford, UK
7. Newcastle University Translational and Clinical Research Institute, University of Newcastle, UK
8. Institute of Transplantation, Newcastle upon Tyne Hospitals NHS Foundation Trust, Newcastle upon Tyne, UK
9. Respiratory Medicine Unit, Nuffield Department of Medicine, University of Oxford, Oxford UK.

*Corresponding author – Ling-pei.ho@imm.ox.ac.uk

Abstract

Healthy repair of the alveoli requires alveolar stem cells to differentiate into cells designed for gas exchange. In chronic lung fibrotic disease like idiopathic pulmonary fibrosis (IPF), alveolar epithelial cells regenerate abnormally. The cause of this is unknown but its highly cellular, inflamed and structurally altered regenerating niche is likely to be relevant. Here, in unique sets of human lung tissues capturing advancing fibrosis, and with a 33-plex single cell imaging mass cytometry (IMC), we provide a high resolution and comprehensive temporo-spatial cell atlas of the regenerating alveolar niches. Using a suite of mathematical tools, we expose an organized immune network and identify CD206^{hi} alveolar macrophages as a central immune cell in the immune-alveolar epithelial interactome. A spatially-directed receptor-ligand analysis offers an in-silico mechanism by which these macrophages influenced alveolar regeneration. Our study unravels a complex cellular environment and identifies key interactions that influence alveolar regeneration in a fibrotic lung.

Introduction

Tissue regeneration and the function of progenitor cells rely on a local microenvironment that provides optimal signals for adult stem cells to proliferate and renew(*1, 2*). After injury, healthy lungs possess remarkable ability to regrow, served by adult stem cells including type II alveolar epithelial cells (ATII), club cells, KRT5^{-} bronchoalveolar stem cells (BASCs) and the rarer airway-based KRT5^{hi} basal stem cells(*3*). In contrast, regeneration potential is low and often abnormal in chronic lung diseases like idiopathic pulmonary fibrosis (IPF).

Idiopathic pulmonary fibrosis is a chronic progressive fibrotic disease with a devastating outcome. Half of patients die within 5 years of diagnosis and there are currently only two drugs, which slow down but do not halt deterioration(*4*). Activated fibroblasts lay down collagen as part of an exaggerated response to minor alveolar insults and expand the interstitium around the alveoli. There is a significant immune cell presence in the interstitium (*5*)and it is now established that some parts of the alveolar epithelium regenerates into bronchiolar-like rather than the type I alveolar epithelial (ATI)cells designed for optimal gas exchange (*6, 7*).

With single cell transcriptomic technologies, several investigators have recently delineated a group of abnormal alveolar epithelial cells with 'basaloid' or bronchial epithelial features (*8-10*) These cells could hold the key to understanding why alveoli regenerate abnormally in IPF lungs(*10*). In addition to being ectopically found in the alveoli (instead of bronchi), these alveolar basaloid intermediate (ABI) cells are transcriptomically distinct from normal alveolar and airway stem and epithelial cells (*8, 9*). Human alveolar organoid studies simulating IPF conditions suggest that these intermediate basaloid cells differentiate to KRT5^{hi}KRT17^{+} basal cells, and could be the differentiation path that converts alveoli to non-functioning honeycomb-like cysts typically seen in IPF lungs(*11*). However the factors influencing this differentiation trajectory is unknown.

We hypothesize that their diseased microenvironment, with new inflammatory signals, cross talk opportunities and growth factors, sustains the presence of ABIs in IPF lungs and alter the differentiation trajectory to basal cells. Understanding the immune cellular composition, their temporal organization and interaction within the regenerating alveolar tissue niche, is a crucial prerequisite to delineating how immune cells influence ABIs and regeneration. It has the potential to alter how we tackle some of the key barriers to stem cell regeneration therapy(*12*).

In this paper, we provide a high resolution temporo-spatial cell atlas of the immune and epithelial cells, and their interaction in the regenerating alveoli in IPF lungs. We use a 33-plex tissue mass cytometry method to define the identity and location of each cell (and abundance of each cell type) in a unique series of lung samples obtained over an area of advancing lung disease, providing an in-tissue method of understanding the temporal evolution of disease. Then, with our bespoke suite of mathematical tools (SpOOx 2.0, available as an open access resource) we investigate the interactions between the immune cells and regenerating alveolar epithelial cells at single cell level and in networks, their cellular neighborhood and identify the immune cells that interact spatially with each other and with different subtypes of alveolar epithelial cells above chance interaction.

Results

Defining the temporo-spatial niche of regenerating alveoli in IPF lungs

The overall study steps are shown in Fig 1A-K and described in detail in Methods.

Lung biopsies were obtained from patients with IPF (n=7 patients; 3 lung biopsies per patient) undergoing lung transplantation. Whole explanted IPF lungs were coronally sectioned and examined by an experienced thoracic pathologist to identify a lobe with macroscopic evidence of advancing disease stages. A 1x1cm^2 biopsy was taken from each of three visually distinct lung areas representing sequential progression from early to advanced fibrotic disease [(Adv, or 'A'); relatively normal lung (Early, or 'E') and the lung region in between (Intermediate or 'Intm/I') which represents the leading edge of disease] (Fig 1B). Disease-free tissue obtained from left upper lobe of un-used lungs from healthy organ donors(n=2) was obtained for comparison (table S1).

Formal histopathological analysis showed an overall progressive increase in the amount of collagen, smooth muscle and a reduction in the numbers of fibroblastic foci as disease progressed from Early to Intm and Adv stages (Fig 1L, table S2). To address our aim of examining the immune interactome around regenerating alveolar epithelium, all ROIs (1mm^2) were selected by an expert thoracic lung pathologist and two pulmonologists to fit the following criteria – histopathologicallyidentified alveoli with (i) presence of cells <u>and</u> (ii) presence of type II alveolar metaplasia (a feature of alveolar regeneration) <u>and</u> (iii) lack of bronchi and bronchioles (Fig 1M). The 'regenerating alveolar niche' in this paper is defined as the area around type II alveolar metaplasia-containing alveoli. In total, 53 ROIs were selected (n=19 'E', n=20 'I' and n=14 'A').

Sections were stained with a 33-plex metal-tagged antibody panel (Fig 1N, table S3), designed to capture the breadth of the immune cell profile, and to identify regenerating alveolar epithelium, including the 'ABIs' and basal cells described by Adams, Habermann and Kathiriya et al (*8, 9, 11*). Single cell segmentation, dimensionality reduction, cell clustering and annotation were performed according to the SpOOx 2.0 (Spatial Omics Oxford) pipeline, an extension of the SpOOx pipeline which we developed previously (*13*) (Fig 1F-K and Methods). A total lung area of 61.89mm^2 was ablated and 313,818 single cells were obtained.

With our step-wise and iterative annotation method (fig. S1), we identified 29 cell clusters – 12 structural and 17 immune cell types (Fig 1O and table S4). Detailed annotation decisions, including criteria, and likely cells for 'undefined'/UD and 'adjacent' ADJ cell clusters are provided in fig. S1, and table S4.

This part of the work establishes our alveolar regeneration niche and the basis for single cell resolution characterization of this microenvironment. It also shows that alveolar regeneration niches are abundant and found in all stages of disease.

Cellular composition of regenerating alveolar niche in IPF

In total, five alveolar epithelial cell types were annotated in the regenerating alveolar niche (Fig 2A-E). By morphology (type II alveolar metaplasia by histology) and presence of basaloid markers, and in conjunction with relative expression of ProSP-C, KRT5, EPCAM and KRT17/7 (in order of hierarchy of importance), three groups of cells were referred to as ABIs. These were classified as follows – 'ABI_a' (KRT5^{neg-lo} ProSP-C^{neg-lo}), 'ABI_b' (KRT5^{lo-mid} proSpC $^{neg-lo}$) and 'ABI_b-DC ADJ' (Fig 2F-G). ABI_b-DC ADJ subcluster showed markers of ABI_b and were the only alveolar epithelial cells to show CD11c and low levels of CD45 expression in keeping with presence of adjacent dendritic cells (DC) (Fig 1F, and fig. S2A). An alveolar-based 'basal cells' (KRT5^{hi} KRT17/7^{hi} ProSP-C^{neg}), and a cell cluster containing AT II cells ('ATII')(KRT5$^{neg/lo}$ ProSpC$^{mid/hi}$) were also identified (Fig 2F, table S4, fig. S2B).

We propose a link between our alveolar epithelial cells and points along the trajectory of ABI differentiation suggested by amalgamation of protein markers and transcriptomic data from Kathiriya, Adams and Haberman(*8, 9, 11*), shown in Fig 2H. ABI_b has lower ProSP-C expression compared to ABI_a; so is likely to be later than ABI_a in the differentiation trajectory. This is also supported by slightly higher levels of KRT5 in ABI_b compared to ABI_a (Fig 2F and H). Our manual matching analysis between our and published data (fig. S3) suggests that the alveolar basaloid intermediate states (i.e. our ABI_a, ABI_b; Adam's AB; Haberman's KRT5^{-}KRT17^{+} and transitional AT II; and, Kathiriya's ABI1 and ABI2) are a mixture of cells along an approximate trajectory of differentiation suggested by Kathiriya's organoid studies(*11*).

16 clusters of immune cells were identified, representing the breadth of adaptive and innate immune cells, including rarer, but relevant, specialised cells like Tregs and $\gamma\delta$ T cells, and subclusters of monocytes and macrophages (table S4).

In terms of abundance, ABIs and basal cells formed a large proportion of alveolar epithelial cells in the regenerating alveolar niche (63.7%) (Fig 2I left panel) (table S4]). There was a near 1:1 ratio for ATII:ABI_a.

Of the immune cells, CD14^{lo} monocytes [17% of all immune cells] were the most abundant immune cells (Fig 2I, right panel). These monocytes were distinct from CD14^{hi} monocytes, both in location and association with other myeloid cells (fig. S4). CD14^{lo} monocytes were scattered throughout the lung sections, co-located with endothelial cells, most likely from alveolar capillaries, making these most likely to be patrolling monocytes(*14*). In contrast, CD14^{hi} (classical) monocytes were found closely associated with DCs, forming the CD14^{hi} mono-DC ADJ cell cluster, and were less abundant (5.6% of immune cells) (Fig 2I and fig. S4B).

Macrophages can be divided into three subgroups according to CD206 and CD14 expression (table S4), reflecting their differentiation trajectory. CD206^{hi} and CD206^{mid} macrophages were found in the alveolar space whereas CD206^{neg} macrophages were found in the interstitium (fig. S5). There was also morphological differences between CD206^{mid-hi} macrophages and the tissue-based CD206^{neg} macrophages (fig. S5B-C). CD206^{mid} and CD206^{hi} macrophages have a distinctive basophilic cytoplasm on H&E staining and were invariably found as tightly packed cell

clusters in the alveolar space. These macrophage subsets also differed in their CD11c, CD1c and CD206 expression levels in keeping with different maturation stages (table S4).

Notably, neutrophils numbers (5.2%) were lower than monocyte and macrophage subsets (Fig 2I).

In addition to myeloid cells, there was an unexpectedly high number of CD4 T cells (17.9 % of all immune cells), CD8 T cells (11.8%), and B cells (7.9%) (Fig 2I). We detected CD103^{+} resident CD4 T cells but not CD103^{+} CD8 T cells [in keeping with prior findings that resident CD8 T cells are found predominantly in airways (rather than alveoli)](*15*)

There was no statistical difference in the overall proportion, distribution, or abundance of these immune and epithelial cells in ROIs from Early, Intm and Adv lung sections (Fig 2J). The exception was ABI_b which was significantly more abundant in the advanced stages (Fig 2J).

In contrast, lung sections from non-diseased donor lungs (HC) showed paucity of immune cells (as expected) (fig. S6A). Immunofluorescence staining of these healthy lung sections confirms scarcity of immune cells, and lack of ABIs in normal alveoli (fig. S6B-D).

A key finding from this part is that aberrant alveolar epithelium (ABI and basal cells) make up a large proportion (at least 64%) of the regenerating alveoli in IPF lungs. This means that very few cells are functioning ATI cells in regenerating alveolar niches in IPF (less than 5%; compared to 1:2 ATI:ATII in healthy lungs). Another biologically significant result is that immune cells in these regenerating niches were abundant even in the most advanced disease stage, dominated by CD14^{lo} monocytes, CD4 T cells, CD8 T cells and CD206^{hi} macrophage, and the composition of immune cells did not change significantly in the different stages of disease.

Strong positive correlation in numbers of specific immune cells with ABIs in regenerating alveolar niche

Although there was no difference in the overall composition of immune cells between Early, Intm and Adv disease stages, we found striking and significant correlations in numbers between specific immune cells and alveolar epithelial cells at different disease stages (Fig 2K, fig. S7). The highest number of significant ABI-immune cell correlates were found in the Intm stage (Fig 2K). Here, ABIs correlated positively with CD103^{+} CD4 T cells, CD4 T cells, CD14^{lo} monocytes, CD206^{hi} macrophages, NK cells, proliferating monocyte-macrophages and neutrophils(r between 0.50 to 0.85; $p < 0.001$ to 0.03)(Fig 2K-L).

Notably, correlations in numbers between ABI and neutrophils and NK cells were found uniquely in Intm (disease leading edge) stage (Fig 2K). CD206^{hi} macrophages were the only cell type whose numbers increased with ABI numbers in all stages of disease (with ABI_b-DC ADJ) (Fig 2K, fig. S7). There was only one statistically significant negative correlation between ABIs and alveolar epithelial cells – CD206^{neg} macrophages with ABI_b-DC ADJ ($r = 0.71$; $p = 0.02$).

These findings suggest possible roles for specific immune cells (CD103$^+$ CD4 T cells, CD4 T cells, CD14^{lo} monocytes, proliferating monocyte-macrophage cells, CD206^{hi} macrophages and neutrophils) in the pathobiology of ABIs or vice versa. It is noteworthy that apart from the correlation between CD206^{hi} macrophages and ATII (in Early disease stage) all correlation in immune cell numbers were with ABIs.

CD103$^+$ CD4 T cells, CD206^{mid} macrophages and CD206^{hi} macrophages spatially co-locate with aberrant regenerating alveolar epithelial cells

We next determined, without *a priori* specification of neighborhood or types of cells, which cell types were spatially associated in each ROI and compared this across the three disease stages. We used an updated suite of mathematical tools (SpOOx 2.0) to: (a) identify pairs of cells types that are more frequently co-located with each other than would be expected under complete spatial randomness (CSR) [via a 2-step sequential method - quadrat correlation matrix (QCM) then, cross pair correlation function (cross-PCF)] (b) determine whether three cell types are co-located with each other more frequently than under CSR, [using a mathematical function called neighborhood correlation function (NCF)] and (c) construct a network of contacting cells [via calculation of 'adjacency contact network(ACN) and then contacting cell network (CCN)] (Fig 1 F-J). This sequential pipeline is described in Methods.

The cross-PCF, $g(r)$, quantifies co-location between cells of type A and B by measuring how frequently cells of type B are found in an annulus of inner radius r and width $dr = 10\mu m$, centred around cells of type A, compared to expected in CSR (Fig 1G) (*16*). We compute the cross-PCF at a fixed inner radius of r=20μm (and width *dr* = 10μm) as the diameter of any cell of interest in our study ranges from a mean of 8mm (T lymphocytes) to 21μm for macrophages (Fig 1G). We also calculate the 95% confidence intervals around $g(r = 20)$, following the bootstrapping process described by Loh (*17*). To provide greater detail of the spatial distribution of co-located cells, a '$g(r)$ plot' (fig. S8A) provides information for the entire range of radii from 5 to 300μm and a 'topographical correlation map' ('TCM') allows visualization of how spatial co-location between cells of types A and B changes across a single ROI (fig. S8B).

In total, 450 (from a possible 2187) pairs of cells passed the first step of analysis (QCM) and were submitted for cross-PCF analysis (Fig 3A) (table S5). Of these, 184 pairs showed co-location above expected at CSR [$g(r = 20) > 1$, (table S5)]. When these pairs were organized into disease stages, with epithelial cell types assigned as 'anchor cells' (around which the number of co-locating cells were counted), only three immune cell types were found to be co-located with any ABIs (ABI_b-DC ADJ and ABI_a) - CD206^{mid} macrophages, CD206^{hi} macrophages and CD103$^+$ CD4 T cells (Fig 3B-C). All other interactions were epithelial-to-epithelial cell type or immune-to-immune cell type co-locations. Notably, no immune cells were significantly co-located with the ATII cell cluster, nor with the more differentiated ABI (ABI_b) or basal cells (Fig 3B).

We next examined how co-localisation of immune cells varied with disease stage. CD206^{hi} macrophages were found significantly co-located with ABI_b-DC ADJ at all disease stages. The only other significant co-locations were between early CD103$^+$ CD4 T cells and ABI_b-DC ADJ cells in the Early stage of disease, and CD206^{mid} macrophages with ABI_a in the Adv stage (Fig

3B). Thus, despite highly significant and strong correlation in abundance between ABIs and the immune cells, CD103$^+$ CD4 T cells and neutrophils at the leading edge of disease (Intm stage) (Fig 2L-M), neither cell types were spatially associated with ABIs in this disease stage. $g(r)$ plot for CD206^{hi} macrophages - ABI_b-DC ADJ pairs showed the same qualitative distribution across the three disease stages (Fig 3C) – the peak in co-location ('g_max') for CD206^{hi} macrophages-ABI_b-DC ADJ cell pair was found near the centroid of ABI_b-DCs, indicating stronger co-location above CSR nearer to ABI. The highest values for g_max was observed in ROIs from the most advanced disease stage (Adv) (Fig 3C), in keeping with the strongest co-location between the cell types. Examination of the individual cell centroid maps and topographical correlation map corroborated the $g(r)$ plot findings (Fig 3D-F).

Remarkably, despite their high numbers, CD4 and CD8 T cells, and B cells were not found significantly co-located around any alveolar epithelial cells (Fig 3B). Similarly, although there were striking correlations in numbers between ABI cell types and CD14^{lo} monocytes (Fig 3B), there were no significant spatial associations between these cell types. The highest g(r=20), as expected, were found for alveolar epithelial cells amongst themselves (e.g. ABI_b-DC ADJ and basal cells) (Fig 3B, table S15).

We also detected co-locations between the vasculature cell cluster and some immune cell types (Fig 3B) - NK cells and neutrophils were co-located around the vasculature cluster in Early disease stage, and CD14^{lo} monocytes around vasculature in all disease stages. This is consistent with NK and neutrophil influx into areas of regenerating alveoli in the early stages of disease and CD14^{lo} monocyte (purportedly patrolling monocytes) presence in the blood vessels at all stages of disease.

Altogether, these mathematical tools drew out statistically supported spatial changes in immune cells-ABI connections over the pseudotime provided by the in-tissue linear progression of disease, which would not have been possible with visual inspection or correlation analysis alone. Specifically, we found that CD206^{hi} macrophages were co-located with ABI_b- DC ADJ cell clusters across all disease stages, while CD103$^+$ CD4 T cells were found co-located with this ABI earlier in disease. NK cells and neutrophils were found spatially associated with 'vasculature' in Early disease, and by the Intm stage, were found correlated in numbers with ABI_a and ABI_b respectively, though not co-located with them.

CD206^{mid} macrophages, CD206^{hi} macrophages and ABI-DC cells form strongest spatially-connected foci at the leading edge of disease.

Having identified CD206^{hi} macrophages, CD206^{mid} macrophages and CD103$^+$ CD4 T cells as the main immune cell types paired spatially with ABI_a and ABI_b-DC ADJs, we reasoned that these immune and ABI cell types could also be spatially connected with each other (beyond pair-wise connections). To test this, we applied a mathematical method which examines co-location of triplets rather than pairs of cell types(*18*). The neighbourhood correlation function (NCF), $NCF_{C_1C_2C_3}(r)$, uses the concept of a minimum enclosing circle (MEC) – the smallest circle drawn which encompasses all three specified cells, of types C_1, C_2 and C_3 described in Methods and fig.

S9). The smaller the radius of MEC, the closer the three cell types spatially (fig. S9). The NCF is a numerical representation of the strength of co-location of these triplets, obtained by comparing the distribution of MECs with radius r against the distribution expected if all three cell types were distributed under CSR. Our relevant MECs are those with low radii but these radii have to be greater than the combined maximal diameter of the three cell types [which ranged from 8μm (T lymphocytes) to 21μm for macrophages] if they were found in closest proximity to each other (fig. S9). Accordingly, NCF for MECs with 20μm and 30 μm were considered most relevant biologically. All combination of triplets of CD206^{hi} and CD206^{mid} macrophages, CD103^{+} CD4 T cells, ABI_b-DC ADJ and ABI_a were examined.

Using this method, we observed spatial co-location above CSR (FDR q < 0.05) for all triplet combinations of the specified ABI -immune cells, across all disease stages, apart from 9 triplets (Fig 3G). The co-localisation was less pronounced (lower NCF) for triplets of CD103^{+} CD4 T cells, macrophage and ABIs (Fig 3G and J). All triplets containing macrophages and ABI subsets, and without CD103^{+} CD4 T cells were co-located in all disease stages but most strongly in the Adv stage (NCF >2, MEC with radii of 20 and 30μm). The CD206^{mid} macrophage, CD206^{hi} macrophage and ABI_b-DC ADJ triplet at the leading edge of disease showed the highest NCF values (Fig 3G-I).

These analyses augment and support the spatial picture gained from our cross-PCF analysis. It shows that amongst the co-locating pairs detected by cross-PCF method, most cell types are also co-located with each other. The most strongly co-located were ABI_b-DC ADJ, CD206^{mid} macrophages and CD206^{hi} macrophages at the leading edge of disease (Intm stage) and ABI_b-DC ADJ, ABI_a and CD206^{hi} macrophages at Adv stage. Thus at the leading edge of disease, cellular activity between ABI and immune cells could centre around CD206 $^{mid-hi}$ macrophages, DC and ABI_b.

Neighborhood analysis for ABI cells supports involvement of CD206^{hi} macrophages in differentiation of ABI cells to basal cells

To provide a complete picture of the neighborhood of ABIs, we collated all cell types (i.e. not just immune cells) that are significantly co-located with ABI_a and ABI_b-DC ADJ, revealing two further information about these ABIs. Firstly, ABI_a cells (purportedly, the earlier cells in the differentiation trajectory) are co-localized with the 'vasculature' cell cluster in Early and Intm stages [albeit a relatively weak co-location ; g(r=20)=1.1]. Basal cells are a co-located cell type for ABI_a in the Adv stage only , together with CD206^{mid} macrophages (Fig 4A). In contrast, ABI_b-DC ADJ cell cluster form foci of cells with consistent presence of CD206^{hi} macrophages and basal cells from Early to Adv stage (Fig 4B).

The differences in neighborhoods for these two ABIs supports the concept that ABI_b-DC ADJ clusters are a more established aberrant alveolar regeneration foci, particularly with the presence of basal cells across all stages of disease. Contemporaneous presence of CD206^{hi} macrophage with basal cells in the ABI_b-DC ADJ's but its absence around the younger ABI_a cells supports a role for CD206^{hi} macrophages in ABI differentiation

Network analysis of contacting cell pairs shows significantly more CD206^{hi} macrophages in direct contact with ABIs and reveals networks of immune and aberrant epithelial cells which evolves as disease progresses.

In the final part of the spatial analysis, we questioned which pairs of cell types are in direct physical contact [rather than merely in higher abundance within a radius of 20μm around these cells, [as shown by the cross-PCF, g(r=20)], since this will enable us to infer a potential receptor-ligand interaction.

To do this, we used the cell segmentation masks to identify co-located cells that are also in direct physical contact with each other, as previously described (we termed this the Adjacency cell network or ACN) (*13*) (fig. S10). In an extension to the ACN, for each cell type pair (e.g., cell types A and B), we then calculate the proportion of type A cells that is in contact with at least one of cell type B, and then the *z*-score [i.e. the deviation of the number of A- B contacts observed in our tissue compared to CSR (CSR here is derived by shuffling all cell type labels computationally by 1000 permutations, without moving the cell positions in the cell mask)]. A p value is computed for the *z*-score associated with each pair of cell types. The data are then filtered to retain pairs of cell types with g(r=20)>1, positive *z*-scores and $p < 0.05$ to create a network in which each connection ('edge') between cell pairs encompasses these three calculations. To differentiate this final network from the ACN (which only determines which cells are in contact with each other), we termed this network the contacting cell network (CCN).

CCN revealed that two immune cell types are in direct contact with ABIs above CSR - CD206^{hi} macrophages and CD206^{mid} macrophages, with ABI_b-DC ADJ and ABI_a cells respectively(Fig 5A-C). CD206^{hi} macrophages-ABI_b DC ADJ pairs were found in direct contact across all disease stages while CD206^{mid} macrophage- ABI_a pairs were only observed in the Adv disease stage, adding and corroborating findings from the neighbourhood analysis (Fig 4A-B).

Additional cell networks, associated with distinct disease stages, were also uncovered by the CCN. In the Early disease stage, ABIs, basal cells and ATII were found in direct contact with each other (Fig 5D). CD206^{mid} and CD206^{hi} macrophages were the only immune cells found within this alveolar epithelial cell community. All other immune cells in the Early disease stage were spatially co-located forming a separate 'community' of cells (Fig 5D). The pattern of spatial co-location changed at the leading edge of disease (Intm), where CD206^{hi} macrophages appear to be the central connector for the ABI and immune cell communities (Fig 5E), linking the entire ABI network to the immune network. The highest numbers of directly contacting pairs of cell types were also found at the leading edge (Intm stage). This mega ABI-immune networks was not seen in Adv (Fig 5F). Here, the immune cell community split into two main sub-CCNs: a myeloid-dominant network (comprising CD206^{hi} macrophages, CD14^{hi} monocyte-DC ADJ cluster, proliferating monocyte-macrophage cluster and CD206^{mid} macrophages), a T cell-dominant network (comprising CD8 T cells, CD4 T cells, CD103^{+} CD4 T cells, Tregs, CD11b^{mid} UD and CD14^{lo} monocytes) (Fig 5F). Persistence of a myeloid-dominant network with the ABI network in the Adv disease stage, signifies the co-existence of an active network of immune cells with aberrant alveolar epithelial cells, even at the most advance stage of fibrosis. Of note, while co-location of

CD103^{+} CD4 T cells with ABIs was identified by cross PCF and NCF analyses, they were not found to be in direct contact with the ABIs, suggesting that any cross talk is likely to be via humoral communications.

In summary, analysis of contacting cell types at pair-wise level shows significantly more CD206^{hi} macrophages in direct contact with ABIs than CSR, and they (and to a lesser degree, the CD206^{mid}-ABI_a pairings in Adv stage only) were the only pairs of cell types that achieved this. In addition, the networks of contacting cell pairs exposes networks of immune and aberrant epithelial cells which evolved temporally as disease progresses. Importantly, at the leading edge of disease, CD206^{hi} macrophages connect communities of immune cells with communities of aberrant alveolar epithelial cells. These findings elevate the potential importance of CD206^{hi} macrophages and ABIs interaction and prioritises ligand-receptor cross talk between these two cell types.

Fibronectin and MIF signalling is are key signalling pathways between CD206hi macrophages and ABIs.

In the last part of our study, we explored the potential functional significance of direct contact between CD206^{hi} macrophages and ABIs. To do this, we obtained publicly single cell transcriptomic data (which provides depth in cell identity but had no information on location or spatial interaction) from Habermann et al [single cell RNA sequencing data on explanted lungs, from six groups of patients, with lung diseases including IPF (n = 12), and non-fibrotic controls from lung donors (HC; n = 10)](*9*). We chose Habermann's data as some of our trajectory analysis of ABIs was based on Kathiriya's organoid studies (*11*), and they matched their findings with Habermann's transcriptomic dataset (*9*).

We extracted the single cell metadata from the IPF and HC datasets, and first, re-clustered monocyte and macrophage clusters (Fig 6A). This was then resolved transcriptomically to identify a subset with a clear, discriminating level of CD206 expression (which we termed CD206^{hi} macrophages) (Fig 6B-C). Habermann's KRT5^{-}KRT17^{lo} ABI subset and transitional AT II subset were extracted for analysis, the former having been matched to our ABIs (fig. S3), and the latter as a comparator. Habermann's dataset did not allow determination of whether their 'basal cells' were the airway-based basal stem cells or the alveolar-based basaloid cells which were ectopically sited in alveoli. Therefore, we did not include this cell type in the analysis. As we had a population of ABI closely associated with DCs (ABI_b-DC ADJ), we also extracted the plasmacytoid DC (pDC) and classical DC (cDC) populations from the IPF dataset. The CD206^{hi} macrophages, pDCs and cDCs were then submitted for receptor-ligand analysis using CellChat(*19*). We directed CellChat to examine interactions between these three immune cell types and KRT5^{-}KRT17^{lo} ABI and transitional AT II in both directions (e.g., with CD206^{hi} macrophages as senders (ligand) and ABI as receivers (receptors) and vice versa).

Communication probabilities (calculated in CellChat based primarily on receptor-ligand gene expression profile)(*19*) identified several receptor-ligand pathways between ABIs and their directly contacting immune cells. Amongst these, those with 'high' communication probabilities

(defined by us, as those higher than 2 S.D. from the mean communication probability value) were (a) macrophage migration inhibitory factor (MIF) (as the ligand produced by ABIs) that engages the CD74-CD44 and CD74-CXCR4 complexes on CD206^{hi} macrophages, and CXCR4 on DCs (Fig 6D) and (b) fibronectin (as a ligand produced by CD206^{hi} macrophages) that engages with integrin (ITGA3, ITGAV, ITGB 1 and 6), Syndecan (SDC1 and 4) or CD44 receptors on ABI (Fig 6E).). Identification of APP (amyloid beta precursor protein), a cell surface receptor is intriguing, with very little known apart from its role in Alzheimer disease and amyloidosis(*20, 21*).

To place this CD206^{hi} macrophages cluster in the context of known macrophage subclusters in IPF, we re-interrogated the transcriptome of the macrophages in Habermann's dataset and identified the two established macrophage subsets found in single cell analysis of IPF lungs - the FABP4^{+} macrophages [purportedly resident alveolar macrophages (*22*)] and the pro-fibrogenic SPP1^{+}TREM2^{+} macrophages (*8, 22, 23*) (clusters 0 and 1 respectively in Fig 6C). CD206^{hi} macrophages (cluster 9) were distinct from these two macrophage subsets and, interestingly, expressed conspicuously higher levels of a fibronectin (Fig 6F). We validated this FN expression by immunofluorescent staining in our IPF samples and found differential presence of FN expression on CD206^{+} macrophages, supporting the transcriptomic findings (Fig 6G). The potential functional consequence of receptor-ligand engagement is summarised in Fig 6H.

Our data suggest that the CD206^{hi} macrophage subset identified in our IPF lung tissue is the CD206^{hi}FN^{hi} alveolar macrophage subset in Habermann's lung digest dataset. This CD206^{hi}FN^{hi} alveolar macrophage is distinct from previously established resident and pro-fibrotic macrophage subsets in IPF lungs. Our in silico analysis highlights fibronectin and MIF as potential ligands in this interaction. **Discussion**

In this paper, we present spatial data in regenerating fibrotic alveoli from the largest number of single cells to date. We use a unique set of lung sections capturing a timeline in progression as disease advances, and specifically characterized ROIs to denote regenerating fibrotic alveoli. Critically, our unbiased spatial analytical methods scan the entire cellular interactome in these ROIs to identify mathematically determined cellular interactions, networks and spatial organization between cells.

We show that just over half (51%) of the alveolar epithelium is made up of ABIs, with an immune landscape that is numerically dominated by CD14^{lo} monocytes, CD206^{hi} alveolar macrophages, B cells, CD4 and CD 8 T cells. The overall proportions of these alveolar epithelial and immune cell types do not change from one disease stage to another but the way they are organized does. In the early stage or where there is less disease, resident CD103^{+} CD4 T cells feature prominently in spatial association with ABIs, whether in pair-wise interaction, as triplets with CD206^{hi} macrophages and ABI_b-DC ADJ cell clusters or within co-located cells in the immediate neighborhood of ABI_b-DC ADJ cells. As the disease advances, the only interacting cells are the CD206^{mid} and CD206^{hi} macrophages with ABI_a and ABI_b-DC ADJ cells. In both cases, the neighborhood contained co-located basal cells in keeping with an epithelial differentiation trajectory from ABI to basal cells as disease progressed from early to advance stage.

A key outcome from our analysis is the co-location of specific immune cells with aberrant basaloid intermediates but not any other epithelial cells. This means that ABIs are likely to be key points in the differentiation trajectory where immune cell cross talk occurs, and a priority target point for therapeutics. ABI cells are more prevalent than hitherto known (*8, 9*), and found in all regenerating alveolar epithelium, into advanced disease where there is scarce fibroblastic foci and maximal collagen deposition in the interstitium. This means that even at the point of severe end stage disease, alveolar epithelium is attempting to regenerate, and ABIs are still interacting with CD206^{hi} macrophages; possibly providing opportunities for therapeutically targeting even at this late point to prevent progression to basal cells.

The receptor-ligand analysis takes current similar analyses forward by providing a spatial dimension where cell types that are actually directly in contact with each other are selected for receptor-ligand interrogation. This significantly enhances the precision of receptor-ligand analysis compared to single cell studies where all possible receptor-ligand interactions are put forward in silico. Of the several receptor-ligand signalling pathways identified by CellChat, a highly relevant one is the FN- $\alpha_v\beta_6$ integrin pathway. Integrins are a family of cell adhesion molecules which mediate many key cell–cell and cell-matrix interactions during fibrosis(*24*). They are a major node of communication between the stroma, immune cells and fibroblasts and closely involved in the initiation, maintenance and resolution of tissue fibrosis. $\alpha_v\beta_6$ integrin is required for fibrosis, and has been a therapeutic target in several clinical trials in IPF(*25, 26*) .It is specifically found on epithelial cells and upregulated after injury, activating TGF-β, a key profibrogenic cytokine(*27*) which activates fibroblasts and facilitates epithelial-mesenchymal transition (EMT). Very little is known of FN or $\alpha_v\beta_6$ integrin's impact on the differentiation trajectory of ABIs but here, a possible consequence is the alteration of the epithelial regeneration and downstream effect on fibroblastic activity. Another receptor-ligand signaling pathway that may be relevant to ABI regeneration is the syndecan-1(SDC1)-FN-mediated ABI-macrophage cross talk. SDC1 is involved in regulating epithelial cell migration and adhesion(*28, 29*) and could augment Wnt signaling(*30*). In acute lung injury soluble Syndecan -1 mediates chemokine mobilization, and influx and activation of neutrophils, whose numbers are found to correlate with ABI_b-DC numbers at the leading edge of disease (Fig 2K-L) (*29, 31, 32*). The converse cross-talk, with production of macrophage migration inhibitory factor (MIF) by ABIs and engagement of CD74, CXCR4 and CD44 receptors on CD206^{hi} macrophages, could be important in maintenance of ABIs. MIF sustains pro-inflammatory immune responses by inhibiting p53-dependent apoptosis of macrophages (*33*) and controlling the expression of NLRP3, an essential step for activation of the NLRP3 inflammasome and subsequent activation of the IL-1β family of cytokines(*34, 35*). Its level have been shown to be increased in IPF lungs(*36*). This may be significant as IL1-β -mediated inflammation appears critical in accumulation of basaloid intermediates (or damage-associated transient progenitors, DATP) in murine models of alveolar injury (*10*).

Numerically, the immune cell landscape is populated equally by adaptive and innate immune cells but there are clear differences in the way immune cells interact spatially with ABIs, and how they are organized spatially at different stages of disease. Our mathematical analysis allow us to generate an interactome which can be organized by a hierarchy of interactions - direct contact [positive correlation in numbers, g(r=20)>1 and CCN FDR < 0.05 - CD206^{mid} and CD206^{hi}

macrophages], short range interactions [positive correlation in numbers and g(r=20)>1 but CCN > 0.05 e.g. resident ($CD103^+$) CD4 T cells] and potentially long range interactions (positive correlation in numbers only e.g. CD4 T cells, $CD14^{lo}$ monocytes, proliferating monocyte-macrophages, neutrophil and NK cells) (Fig 3A, 3D,4C). Thus, foci of ABI-$CD206^{hi}$ macrophages together with less spatially intimate immune cells (e.g. NK and neutrophil) could drive the growth of ABI at the leading edge of disease. Studies showing resident $CD103^+$ CD4 T cells interacting with MHCII-expressing epithelial cells (*37*) and augmenting homing of neutrophils to epithelium are (*38*) particularly interesting in this context since they suggest possibilities around initiation of injury in early disease stage by the $CD103^+$CD4 T cells and ABI_b_DC ADJs as trigger for formation of ABIs.

Another noteworthy mention is the abundance of $CD14^{lo}$ monocytes, its correlation with that of ABI numbers, co-location with vascular endothelium in Early and Intm disease stages. By their markers, $CD14^{lo}$ monocytes are likely patrolling (or non-classical) monocytes ($CX3CR1^{high}CD14^{dim}CD16^+$ in humans)(*14, 39, 40*). Their co-location with vessels in our tissue further supports this. Resting, patrolling monocytes actively scan the vasculature. Inflamed or injured endothelium releases further chemoattractant to attract these monocytes(*14*). Presence of $CD14^{lo}$ monocytes with blood vessels is indicative of inflamed vasculature in IPF, and supports an involvement, or association, between aberrant vascular health and aberrant alveolar regeneration.

Our main limitations are the following. We were unable to detect ATI cells, in main because at the time of study, we did not have specific ATI protein markers. Some ATI cells could be found within clusters of $KRT5^{neg\text{-}lo}$ ABIs (e.g. ABI_a) but the morphology of these cell types were examined on H&E sections (Fig 2G) and were not compatible with ATI cell types. We believe they are probably found within the 27 UD cluster which formed 2.6% of epithelial cells. We focused on breadth of immune cells rather than depth and functional protein markers within the 33-plex panel so our immune cell profile lacked the deeper layer of markers that could indicate the subtypes of adaptive immune cells in particular. However as these cell types did not co-locate with alveolar epithelial cells. Finally our studies are descriptive and further work will be required to definitively determine the interactions observed and in silico mechanisms.

In conclusion, we present a robust mathematical method of spatial analysis which, without a priori specification of type of immune cells or alveolar epithelial cells, identifies $CD206^{hi}$ macrophages as a key immune influence in the differentiation trajectory of alveolar regeneration in IPF lungs, and expands our knowledge of the regenerating alveolar niche in IPF. Our results emphasize the importance of ABI cells as an immune interaction point in the differentiation of alveolar regeneration, directing focus to these cells and its accompanying immune cells and pathways, underlining these cellular interactions as new avenues for therapeutic explorations to alter abnormal stem cell differentiation, even in the most advanced stages of IPF.

References

1. C. A. Chacón-Martínez, J. Koester, S. A. Wickström, Signaling in the stem cell niche: regulating cell fate, function and plasticity. *Development* **145**, (2018).
2. David T. Scadden, Nice Neighborhood: Emerging Concepts of the Stem Cell Niche. *Cell* **157**, 41-50 (2014).
3. K. D. Alysandratos, M. J. Herriges, D. N. Kotton, Epithelial Stem and Progenitor Cells in Lung Repair and Regeneration. *Annu Rev Physiol* **83**, 529-550 (2021).
4. D. J. Lederer, F. J. Martinez, Idiopathic Pulmonary Fibrosis. *N Engl J Med* **378**, 1811-1823 (2018).
5. K. Shenderov, S. L. Collins, J. D. Powell, M. R. Horton, Immune dysregulation as a driver of idiopathic pulmonary fibrosis. *The Journal of Clinical Investigation* **131**, (2021).
6. A. E. Vaughan *et al.*, Lineage-negative progenitors mobilize to regenerate lung epithelium after major injury. *Nature* **517**, 621-625 (2015).
7. W. Zuo *et al.*, p63(+)Krt5(+) distal airway stem cells are essential for lung regeneration. *Nature* **517**, 616-620 (2015).
8. T. S. Adams *et al.*, Single-cell RNA-seq reveals ectopic and aberrant lung-resident cell populations in idiopathic pulmonary fibrosis. *Sci Adv* **6**, eaba1983 (2020).
9. A. C. Habermann *et al.*, Single-cell RNA sequencing reveals profibrotic roles of distinct epithelial and mesenchymal lineages in pulmonary fibrosis. *Sci Adv* **6**, eaba1972 (2020).
10. J. Choi *et al.*, Inflammatory Signals Induce AT2 Cell-Derived Damage-Associated Transient Progenitors that Mediate Alveolar Regeneration. *Cell Stem Cell* **27**, 366-382.e367 (2020).
11. J. J. Kathiriya *et al.*, Human alveolar type 2 epithelium transdifferentiates into metaplastic KRT5+ basal cells. *Nature Cell Biology* **24**, 10-23 (2022).
12. K. L. McKinley, M. T. Longaker, S. Naik, Emerging frontiers in regenerative medicine. *Science* **380**, 796-798 (2023).
13. P. Weeratunga *et al.*, Single cell spatial analysis reveals inflammatory foci of immature neutrophil and CD8 T cells in COVID-19 lungs. *Nature Communications* **14**, 7216 (2023).
14. C. Auffray *et al.*, Monitoring of Blood Vessels and Tissues by a Population of Monocytes with Patrolling Behavior. *Science* **317**, 666-670 (2007).
15. M. Z. M. Zheng, L. M. Wakim, Tissue resident memory T cells in the respiratory tract. *Mucosal Immunology* **15**, 379-388 (2022).
16. E. Gabriel, A. Baddeley, E. Rubak, R. Turner: Spatial Point Patterns: Methodology and Applications with R. *Mathematical Geosciences* **49**, 815-817 (2017).
17. J. M. Loh, A Valid and Fast Spatial Bootstrap for Correlation Functions. *The Astrophysical Journal* **681**, 726 (2008).
18. J. A. Bull, E. J. Mulholland, S. J. Leedham, H. M. Byrne, Extended correlation functions for spatial analysis of multiplex imaging data. *Biol Imaging* **4**, e2 (2024).
19. S. Jin *et al.*, Inference and analysis of cell-cell communication using CellChat. *Nature Communications* **12**, 1088 (2021).
20. D. Obregon *et al.*, Soluble amyloid precursor protein-α modulates β-secretase activity and amyloid-β generation. *Nat Commun* **3**, 777 (2012).
21. P. Cras *et al.*, Senile plaque neurites in Alzheimer disease accumulate amyloid precursor protein. *Proc Natl Acad Sci U S A* **88**, 7552-7556 (1991).
22. C. Morse *et al.*, Proliferating SPP1/MERTK-expressing macrophages in idiopathic pulmonary fibrosis. *Eur Respir J* **54**, (2019).

23. P. A. Reyfman *et al.*, Single-Cell Transcriptomic Analysis of Human Lung Provides Insights into the Pathobiology of Pulmonary Fibrosis. *American journal of respiratory and critical care medicine* **199**, 1517-1536 (2019).
24. D. Sheppard, The role of integrins in pulmonary fibrosis. *European Respiratory Review* **17**, 157-162 (2008).
25. G. Raghu *et al.*, Randomized Phase IIa Clinical Study of an Anti-α(v)β(6) Monoclonal Antibody in Idiopathic Pulmonary Fibrosis. *American journal of respiratory and critical care medicine* **206**, 1166-1168 (2022).
26. G. Raghu *et al.*, A Phase IIb Randomized Clinical Study of an Anti-α(v)β(6) Monoclonal Antibody in Idiopathic Pulmonary Fibrosis. *American journal of respiratory and critical care medicine* **206**, 1128-1139 (2022).
27. J. S. Munger *et al.*, The integrin alpha v beta 6 binds and activates latent TGF beta 1: a mechanism for regulating pulmonary inflammation and fibrosis. *Cell* **96**, 319-328 (1999).
28. W. A. Altemeier, S. Y. Schlesinger, C. A. Buell, W. C. Parks, P. Chen, Syndecan-1 controls cell migration by activating Rap1 to regulate focal adhesion disassembly. *J Cell Sci* **125**, 5188-5195 (2012).
29. W. A. Altemeier *et al.*, Transmembrane and extracellular domains of syndecan-1 have distinct functions in regulating lung epithelial migration and adhesion. *J Biol Chem* **287**, 34927-34935 (2012).
30. C. A. Pataki, J. R. Couchman, J. Brábek, Wnt Signaling Cascades and the Roles of Syndecan Proteoglycans. *J Histochem Cytochem* **63**, 465-480 (2015).
31. S. E. Gill *et al.*, Shedding of Syndecan-1/CXCL1 Complexes by Matrix Metalloproteinase 7 Functions as an Epithelial Checkpoint of Neutrophil Activation. *Am J Respir Cell Mol Biol* **55**, 243-251 (2016).
32. T. Parimon *et al.*, Syndecan-1 promotes lung fibrosis by regulating epithelial reprogramming through extracellular vesicles. *JCI Insight* **5**, (2019).
33. I. Kang, R. Bucala, The immunobiology of MIF: function, genetics and prospects for precision medicine. *Nature Reviews Rheumatology* **15**, 427-437 (2019).
34. T. Lang *et al.*, Macrophage migration inhibitory factor is required for NLRP3 inflammasome activation. *Nat Commun* **9**, 2223 (2018).
35. M. S. Shin *et al.*, Macrophage migration inhibitory factor regulates U1 small nuclear RNP immune complex–mediated activation of the NLRP3 Inflammasome. *Arthritis & Rheumatology* **71**, 109-120 (2019).
36. C. Olivieri *et al.*, Macrophage migration inhibitory factor in lung tissue of idiopathic pulmonary fibrosis patients. *Experimental Lung Research* **42**, 263-266 (2016).
37. A. T. Shenoy *et al.*, Antigen presentation by lung epithelial cells directs CD4+ TRM cell function and regulates barrier immunity. *Nature Communications* **12**, 5834 (2021).
38. C. Zheng, T. Cao, C. Ye, Y. Zou, Neutrophil recruitment by CD4 tissue-resident memory T cells induces chronic recurrent inflammation in atopic dermatitis. *Clinical Immunology* **256**, 109805 (2023).
39. J. Cros *et al.*, Human CD14dim monocytes patrol and sense nucleic acids and viruses via TLR7 and TLR8 receptors. *Immunity* **33**, 375-386 (2010).
40. L. M. Carlin *et al.*, Nr4a1-dependent Ly6Clow monocytes monitor endothelial cells and orchestrate their disposal. *Cell* **153**, 362-375 (2013).

41. G. Raghu *et al.*, Diagnosis of Idiopathic Pulmonary Fibrosis. An Official ATS/ERS/JRS/ALAT Clinical Practice Guideline. *American journal of respiratory and critical care medicine* **198**, e44-e68 (2018).
42. P. Bankhead *et al.*, QuPath: Open source software for digital pathology image analysis. *Sci Rep* **7**, 16878 (2017).
43. B. Hunter *et al.*, OPTIMAL: An OPTimized Imaging Mass cytometry AnaLysis framework for benchmarking segmentation and data exploration. *Cytometry A* **105**, 36-53 (2024).
44. C. A. Schneider, W. S. Rasband, K. W. Eliceiri, NIH Image to ImageJ: 25 years of image analysis. *Nature Methods* **9**, 671-675 (2012).
45. J. Schindelin *et al.*, Fiji: an open-source platform for biological-image analysis. *Nature Methods* **9**, 676-682 (2012).
46. D. A. Van Valen *et al.*, Deep learning automates the quantitative analysis of individual cells in live-cell imaging experiments. *PLoS computational biology* **12**, e1005177 (2016).
47. Jacob H. Levine *et al.*, Data-Driven Phenotypic Dissection of AML Reveals Progenitor-like Cells that Correlate with Prognosis. *Cell* **162**, 184-197 (2015).
48. L. M. Weber, M. Nowicka, C. Soneson, M. D. Robinson, diffcyt: Differential discovery in high-dimensional cytometry via high-resolution clustering. *Communications Biology* **2**, 183 (2019).
49. N. Morueta-Holme *et al.*, A network approach for inferring species associations from co-occurrence data. *Ecography* **39**, 1139-1150 (2016).
50. B. D. Ripley, Modelling Spatial Patterns. *Journal of the Royal Statistical Society: Series B (Methodological)* **39**, 172-192 (1977).
51. D. Moldoveanu *et al.*, Spatially mapping the immune landscape of melanoma using imaging mass cytometry. *Science Immunology* **7**, eabi5072 (2022).
52. P. Virtanen *et al.*, SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods* **17**, 261-272 (2020).
53. S. Seabold, J. Perktold, in *SciPy*. (2010).
54. Y. Hao *et al.*, Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587.e3529 (2021).
55. I. Korsunsky *et al.*, Fast, sensitive and accurate integration of single-cell data with Harmony. *Nature Methods* **16**, 1289-1296 (2019).

Acknowledgements:

Funding: The study is funded by combination of grants from MRC UKRMP Grant MR/S020918/1; NIHR Oxford Biomedical Research Centre grant; University of Oxford Medical Science Division and Newcastle University Flow Core Facility. LPH is supported by MRC (grant CFR01480) and the NIHR Oxford Biomedical Research Centre. AJF is supported by the National Institute for Health and Care Research (NIHR) Blood and Transplant Research Unit in Organ Donation and Transplantation (NIHR203332), a partnership between NHS Blood and Transplant, University of Cambridge and Newcastle University. The views expressed are those of the author(s) and not necessarily those of the NIHR, NHS Blood and Transplant or the Department of Health and Social Care.

Author contribution: PW led the conceptual and intellectual development of the SpOOx V2.O pipeline with LPH, and ran all analyses, writing, interpretation of data, data organization, figures, tables, and writing of the paper with LPH.

PW and MS performed the MDV analyses and computational software and pipeline extension .

MS generated all computational output utilizing the SpOOx 2.0 pipeline and was involved in data organization and visualization on MDV in conjunction with PW, JB and HB

JB designed the mathematical algorithms, tested these in silico, applied these to study and critically appraised interpretation of mathematical tools with PW, HB and LPH

HB led mathematical development, interpretation and provided deep critical analyses of overall final synthesis of data

CC performed all the expert histopathology analyses in discussions with LPH and AF.

LD designed the IMC panel with LPH, and contributed to generating and analyses of the MCD images during analysis of overall data and critical review of data interpretation and methods text

CV performed all optimisation of immunofluorescence (IF) validation staining, stained all samples and acquired further control samples for IF staining

TH analysed and contributed to annotation the aberrant alveolar epithelial regeneration, and ran transcriptomic analysis to inform annotation and matching of ABIs to prevailing literature

JMW, AA and PW performed the single cell transcriptomic analysis including CellChat analyses and generated graphical output

JS was involved in conception of study, acquisition of parts of funding of study and review of results

AFilby and BH optimised and performed the IMC staining and generation of MCD files

AFisher was involved in conception of study, acquisition of parts of funding of study, sample acquisition, ethics application, governance of sample harvesting, review of results and writing

LB leads the Newcastle lung sampling programme, and ensured governance of sample harvesting with AFisher

RB collected explanted lungs and processed lung tissue samples

LPH conceived and led the study and the conceptual and intellectual development of SpOOx v2.0 pipeline with PW, JB and HB, analysed all aspects of the data and interpreted all the output with PW, JB and HB, wrote the paper with PW, performed extended analyses of histopathology with CC, corroborated all IMC and IF staining with PW and CV, interpreted all mathematical applications and output with PW, JB and HB, synthesised and amalgamated overall output, and obtained funding for all parts of the project.

All authors read the paper and were involved in corrections and refining the final paper.

Material and Methods

Table of antibodies and reagents used in imaging mass cytometry and immunofluorescence

All antibodies, their catalogue numbers, final dilutions, and source are documented in Table S6.

Patients, samples, and ethical approvals

Diseased lung tissue was obtained from patients with a diagnosis of Idiopathic pulmonary Fibrosis (IPF) based on standard diagnostic criteria (*41*) and multi-disciplinary team (MDT) consensus, at the point of lung transplantation at the Institute of Transplantation, Newcastle Upon Tyne Hospitals NHS Foundation Trust. All patients provided informed consent for use of their tissue via the Cellular and Molecular Mechanisms in Chronic Lung Diseases (EXPLANT) study which was approved by the NHS Research Ethics Service (11/NE/0291) and was sponsored by Newcastle Upon Tyne Hospitals NHS Foundation Trust (R&D ref 5885).

Formal histopathological analysis showed an overall progressive increase in the amount of collagen, smooth muscle and a reduction in the numbers of fibroblastic foci as disease progressed from Early to Intm and Adv stages (Fig 1L, Suppl Table 2). In keeping with histopathology features of 'usual interstitial pneumonia' (UIP) in the IPF disease(*41*), lung sections displayed temporal and geographic heterogeneity in the fibrotic and inflammatory features across the Early, Intm and Adv regions. Given this, we subsequently treated each ROI as a single independent entity within each disease stage in our analysis.

Sections were stained with a 33-plex metal-tagged antibody panel (Fig 1N, Suppl Table 3), which we designed to capture the breadth of the immune cell profile, and to identify regenerating alveolar epithelium, including the 'ABIs' and basal cells described by Adams, Habermann and Kathiriya et al (*8, 9, 11*). Single cell segmentation, dimensionality reduction, cell clustering and annotation were performed according to the SpOOx 2.0 (Spatial Omics Oxford) pipeline, an extension of the SpOOx pipeline which we developed previously (*13*) (Fig 1F-K and Methods). A total lung area of 61.89mm^2 was ablated and 313,818 single cells were obtained.

Disease-free control tissue was obtained from healthy donor lungs which were not used in transplantation and where the donor's next of kin had consented for NHS REC approved research on the tissue (REC 16/NE/0230). Sections were taken again from the upper lobe as a direct comparison to the IPF tissue samples. Further samples from non-cancerous areas of lung cancer patients from the Oxford Biobank were also obtained under the NHS REC Number 14/SC/1060 for H&E exemplar staining.

Slide preparation

Tissue sections for each donor were cut serially, at an IMC optimum thickness of 8 µm and mounted onto SuperFrost Plus™ Adhesion slides (Epredia). The primary slide from each serial deck was subject to H+E staining, followed by dual IHC staining of the following serial slide, with KI67 and PROSPC using the Discovery Ultra Research Staining System (Roche). Staining was performed by Novopath Research Service (NovoPath, Department of Pathology, Newcastle Hospitals NHS Foundation Trust, Newcastle upon Tyne, UK). Stained slides were scanned using the Aperio A2 slide scanner (Leica Biosystems). In combination, both H&E and IHC stains were

used to navigate to areas of epithelial damage and repair in order to aid ROI selection. Slides were viewed and annotated in QuPath(*42*)

Antibody Validation and Metal Labelling

A 33-plex antibody panel was created to target the immune, structural and functional components of the IPF lung samples, and to identify the regenerating alveolar epithelial niche (Table 1). Antibody clones were pre-validated as described in Hunter *et al*. (*43*) In brief, validation and optimisation was carried out using two-colour immune-fluorescence staining of appropriate positive control and lung tissue. Here, slides were imaged using a Axio Scan Z1 fluroscence microscope (Zeiss). Using Zeiss software, the microscope was configured for AF405, AF488, AF568, AF647, and AF750, and the entire area of the tissue section is selected using the software for higher resolution scanning, utilizing a plan apochromat 40x 0.95 korr M27 objective. Images were saved on a computer for further processing using custom Fiji/Image J macros (*44, 45*).

Following verification of staining pattern and performance quality, approved antibodies were subject to lanthanide metal conjugation using a Maxpar X8 metal conjugation kit following manufacturer's protocol (Standard Biotools). Successful metal conjugation was verified by binding the antibody to iridium labelled antibody capture beads AbC™ Total Antibody Compensation Beads (Thermo Fisher) and acquiring on a Helios system (Standard Bio-tools).

IMC Staining

Sections for IMC analysis were baked at 60°C for 2 hours, deparaffinised in two changes of Histological grade xylene (Merck), followed by rehydration in decreasing grades of ethanol ; 100%, 90%, 75% and 50%. Samples were washed twice in MiliQ ultrapure water and then submerged in Tris-EDTA 0.05% Tween 20, pH 9.0. Heat induced epitope retrieval was performed in the microwave for 20 mins. Slides were cooled to 70°C, washed twice in MiliQ water followed by two washes in PBS (Gibco). Blocking of non-specific binding sites was performed using 3% BSA solution in PBS for 45 mins at room temperature. Sections were then subject to staining with a cocktail of 2 DNA probe and 31 conjugated antibodies diluted in PBS containing 0.5% BSA and left to incubate overnight. Slides were washed in 0.2% Triton X-100 (ThermoFisher), followed by two washes in PBS. Nuclei were stained with Cell-ID™ Intercalator-Ir (Standard Biotools) for 30mins at room temperature. Slides were washed in MiliQ water and left to air dry before IMC analysis.

Hyperion (IMC) set up, quality control (QC) and sample acquisition

ROIs were ablated using the Hyperion Imaging System using CyTOF7 Software v7.0 (Standard BioTools) as described in Hunter *et al*(*43*). Sample data was acquired over 4 balanced batches. Prior to each sample acquisition, the Hyperion Tissue Imager was calibrated and rigorously quality controlled to achieve reproducible sensitivity based on the detection of ^{175}Lutetium. Ablations were performed at 200Hz laser frequency to create a resultant MCD file per sample containing all data from ROIs. Raw MCD files were opened in MCD Viewer (Standard BioTools) where visual QC of the staining intensity and pattern was performed. All images were exported as 16-bit single multi-level TIFFs using the "export" function.

Immunofluorescence

Paraffin-embedded human lung tissue sections were deparaffinized and each section was pre-treated using heat-mediated antigen epitope retrieval with sodium citrate buffer (pH 6) for 20 minutes. Then sections were blocked in 10% normal goat serum (Thermo Fischer Scientific, 50062Z) for 20 minutes and then incubated with Alexa Fluor® 488 Anti-Cytokeratin 17 antibody 1:200 dilution (Abcam, AB185032), CD15 antibody 1:200 dilution (Cell signalling Technology, 4744S), Pro-Surfactant Protein C Antibody 1:400 dilution (Abcam AB90716), Alexa Fluor® 647 Anti-Cytokeratin 5 antibody 1:200 dilution (AB193895), Anti-MRC1 Antibody 1:500 dilution (Atlas antibodies, AMAB90746), Anti-BDCA-2 antibody 1:100 dilution (Merck Millipore, MABF94), Alexa Fluor® 647 Anti-Fibronectin antibody 1:200 dilution (Abcam, AB198934) overnight at 4˚C. Each section is washed three times in TBS-T (0.1% Tween) and stained with Alexa Fluor 750 conjugated Goat anti Rabbit IgG or Alexa Fluor 488 or 568 conjugated goat anti-mouse IgM secondary antibody or Alexa Fluor 568 conjugated goat anti-mouse IgG1 for 30 minutes and washed three times in TBS-T (0.1% Tween) and mounted with Prolong platinum antifade Mountant with DAPI (Fischer Scientific) and the section slides were imaged using a Axio Scan Z1 fluroscence microscope (Zeiss).

Data analysis

Software and algorithms

We utilized the SpOOx pipeline for data processing and analysis (*13*) The SpOOx pipeline incorporates Python and R based command line tools implemented as a semi-automated pipeline within a Ruffus framework. This pipeline provides end to end processing of imaging mass cytometry data encompassing processing of .TIFF files, segmentation and cell mask generation, extraction of signal intensity, dimensionality reduction, clustering and spatial analysis. For this work we improved the prior version of the SpOOx pipeline to include additional functionality for neighbourhood correlation function analysis and generation of statistically robust adjacency cell networks.

Processing of MCD to .TIFF files

MCD files were checked for staining and data acquisition quality using MCD viewer (provided by Standard BioTools). Following this initial QC assessment, images were converted to OME-TIFF format for segmentation.

Segmentation and mask generation

Cell segmentation was performed with the Mesmer library in DeepCell(*46*). Nuclear markers DNA1 and DNA3 and cytoplasmic markers (CD45, CD68, CD206, HLA DR, CD14, CD15, CD11b, BDCA2, CD1c, CD11c, CD103, CD20, CD3, CD8a, CD4, CD69, CD56, EPCAM, KRT17, KRT5, CD31, aSMA, ProSP-C, CD127, Amphireg, CD161) were extracted and Z projected to generate single channel nuclear images and cytoplasmic TIFF images. These were contrast adjusted and processed with the Mesmer library reference as nuclear and cytoplasmic channels to generate a final single cell segmentation mask.

Extraction of signal intensities

Mean arcsinh transformed marker signal intensities were extracted from the segmentation masks with no filter for cell size area.

Dimensionality reduction and clustering

Clustering was performed with R phenograph (*47*) with parameter k = 15 on data following integration for condition (IPF) with Harmony for correction of sample and batch effects. Markers used for clustering are shown in Suppl Table 3.

Annotation workflow

Cell phenotype annotation was performed using a sequential approach as described in Weeratunga et al. 2023.(*13*) Briefly, clusters were initially defined using i) heatmaps showing median marker expression, ii) scaled expression density histograms and iii) cluster distribution plots. Based on this analysis we excluded clusters with uniformly low/negative marker expression and clusters found in only one sample. We termed a cluster undefined (UD) if they expressed markers which were found in more than two cell types e.g. variably, epithelial, several immune cell type and endothelial markers. We then examined the spatial location of these annotated clusters using cell centroid maps mapped on to the corresponding segmentation mask and an adjacent H and E region of interest. Some clusters demonstrated markers expressed uniquely on 2 different cell types and these markers showed adjacent expression in IMC images and IF. These clusters were denoted with 'ADJ' (*13*).

Analysis of differential cell abundance

Differential cell abundance was examined between the defined states of IPF (early, intermediate and advanced) using the glmFit and glmLRT functions from the diffcyt package (*48*). To account for differences in area of the analysed ROIs, area was used as a normalizing factor.

Spatial analysis

The spatial analysis pipeline in this manuscript follows (*13*), with the addition of Neighbour Correlation Functions (NCF)(*18*). Here we give a brief technical overview of the components of this pipeline described in (*13*) (the Quadrat Correlation Matrix, cross-Pair Correlation Function, Adjacency Cell Network and Topographical Correlation Matrix), before a more detailed description of the NCF.

Quadrat Correlation Matrix

The QCM follows methods described by Morueta – Holme et al for inferring species associations from co – occurrence data (*49*), and provides a high-level overview of whether counts of each pair of cell types are observed to be correlated at short-range (a square of edge length 100 microns) more or less frequently than would be anticipated. Briefly, the QCM uses random shuffling of cell labels to generate a null distribution reflecting the correlation coefficients between each pair of cell types under random labelling. Observed correlation coefficients are compared against this distribution, to identify whether correlations occur by chance due to the underlying spatial structure of the points. The QCM produces p-values for the significance of the correlation, which account for multiple comparisons between a large number of different cell types. See Morueta Home 2014 (*49*) and Weeratunga 2023 (*13*) for further details.

Cross Pair Correlation Function

Cross pair correlation function was performed as described previously(*13*). Briefly, the cross pair correlation function (PCF), $g(r)$, is a spatial statistic which describes the correlation between pairs of points with specified labels A and B(*18*). It describes this correlation at multiple length scales according to the variable r, with $g(r) > 1$ indicating that more pairs of points with label A and label B are observed separated by distance r than would be expected under an assumption that the points are distributed by complete spatial randomness. Similarly, $g(r) < 1$ indicates that cell pairs are observed less frequently than expected at this distance. For full details of the PCF and its mathematical definition, see (*13*) or (*18*).

Topographical Correlation Map (TCM)

TCM was performed as described in (*13*). The TCM is an example of a local indicator of spatial association (LISA)(*18*) , and overlays a topographic "map" onto the ROI to show the local strength of positive or negative spatial association between cells of types A and B across the ROI(*13, 18*). It can be considered to represent the contribution of each cell of type A to the PCF for values of r up to a given threshold, here 100 μm, or equivalently as a spatially resolved representation of contributions to Ripley's K function at $K(r)$ (*50*).

Neighbourhood Correlation Function

The Neighbourhood Correlation Function (NCF) extends the standard pair correlation function to identify correlation between three (or more) classes of cell simultaneously(*18*). We consider a point pattern in a finite 2D domain Ω such that each point *i* has coordinates $\boldsymbol{x_i} = (x_i, y_i)$. Suppose that each point is associated with a categorical label A, B or C and there are N_A, N_B and N_C points with labels A, B and C respectively. Then there are $N_A \times N_B \times N_C$ possible ways to select a set of three points such that we have one point with each label. For each such set of three points ζ, we calculate the smallest circle which can enclose all three points, which we call a "neighbourhood" and define to have radius $R(\zeta)$.

The NCF at radius *r* compares the total number of neighbourhoods of radius *r* against the number of neighbourhoods of size *r* that would be expected if points were distributed according to complete spatial randomness (CSR). Thus, the NCF for points with labels A, B and C is defined as:

$$NCF_{ABC}(r) = \sum_{a=1}^{N_A}\sum_{b=1}^{N_B}\sum_{c=1}^{N_C}\frac{I_r(R(\{\boldsymbol{x_a}, \boldsymbol{x_b}, \boldsymbol{x_c}\}))}{N_A \times N_B \times N_C \times p_3(r)}$$

where $I_r(R) = 1$ if $R \in [r, r + dr)$, $dr > 0$, and $p_3(r)$ is the probability of three points chosen under CSR being found within a neighbourhood of radius r. $p_3(r)$ is defined within Ω as $p_3(r) = \lim_{M\to\infty}\frac{\sum_{i=1}^{M} I_r(R(\{\boldsymbol{x_i^1}, \boldsymbol{x_i^2}, \boldsymbol{x_i^3}\}))}{M}$, where $\boldsymbol{x_i^1}$, $\boldsymbol{x_i^2}$ and $\boldsymbol{x_i^3}$ are locations sampled under CSR within Ω. Practically, $p_3(r)$ is approximated by taking large M (here $M = 10^7$). Values of the NCF greater than 1 are consistent with clustering of the three cell types together at distance *r*, while values less than 1 are consistent with exclusion of one or more cell types from the others at the length scale *r*. For more details, please see(*18*).

The resulting radii were the binned into 5 µm intervals to produce a histogram. Bootstrapping was performed by randomly shuffling the location of the three cell types and calculating the appropriate histogram. The average and 0.05 and .95 percentiles for each interval from the 1000 bootstrapped histograms were used for comparison with the observed value.

To calculate NCFs for a condition, comprising several ROIs, the radii for all ROIs in a condition were combined and the histogram recalculated. For bootstrapping, the bootstrap histograms for each ROI in a condition were combined resulting in 1000 histograms. Again, the average and 0.05 and .95 percentiles for each interval from these histograms were used for comparison with the observed value.

Contact Cell Networks

Adjacency cell networks were calculated as previously described(*13*) – for each co-locating cell pair A/B [g(r=20)>1], the total number of B cells in contact with A cells was calculated. For the contact cell network (CCN), we ask if contacts were significantly different. Z -scores were calculated by randomly shuffling cell labels and recalculating contacts(*51*). The average and standard deviation (sd) of a thousand bootstrap values were the used to calculate z-scores.

$$z_{AB} = (O_{AB} - \mu(N_{AB}))/\sigma(N_{AB})$$

where:

- z_{AB} is the z-score between cell types A and B
- O_{AB} is the number of observed contacts
- $\mu(N_{AB})$ and $\sigma(N_{AB})$ represent the mean and standard deviation of the number of contacts in the bootstrapped data

P-values were calculated using the *norm.sf* function from *scipy* package(*52*) and corrected for multiple comparisons with *multitest* from the *statsmodel* package(*53*):-

```
from scipy.stats import norm
from statsmodel.stats import multitest
pvalues = norm.sf(np.asbsolute(z_scores)) *2
fdr_corrected_pvalues = multitest.fdrcorrection(pvals)
```

To obtain the z-score for an A/B interaction in a condition, comprising several ROIs, a weighted average was used, based on the number of A Cells in in the interaction. This was calculated by the sum of

*A/B zscore * A count / sum of A counts*

p-values for each condition were calculated combining the individual p-value for each ROI using the *stouffer* method and weighted by the A counts. This was implemented by the *combine_pvalue* method in the in *scipy* package(*52*)

```
from scipy.stats import combine_pvalue
combine_pvalues(pvalues,"stouffer",Acounts)
```

Again, the p-values were corrected using the *multitest* function from the *statsmodel* package (see above)

Single cell transcriptomic interrogation

Within the single cell data from Haberman et al. (*9*) macrophage, monocytes, cDCs and pDCs from control and IPF patients were further sub clustered using Seurat R package (ver 4.3.0)(*54*). Highly variable genes were identified by fitting the mean variance relationship for each sample to avoid sample-to-sample variation and to retain variance within the samples. We then performed principal-component analysis using top 2,000 highly variable genes. Scree plots and Jackstraw permutation tests were used to determine significant principal components (with P cut-off <0.01) in the data. Subsequently, we performed batch correction using Harmony (ver 0.1.1) (*55*). A k-nearest neighbor (kNN) graph was constructed from cells in the reduced dimension space in Seurat based on the top 20 batch-corrected components. Cells were then clustered in this on this neighborhood graph using the Leiden algorithm for modularity optimization where the resolution parameter was set to 0.7.

Based on cluster marker gene expression Plasmacytoid dendritic cells (pDC), classical DCs (cDC), monocytes and macrophages were identified, and the monocyte and macrophage clusters re-clustered as a single entity to identify the CD206hi macrophage subset. We then merged these with annotated KRT5-/KRT17+ cells and transitional AT-II for downstream receptor ligand analysis.

Potential ligand receptor (LR) interaction between different celltypes was analysed using CellChatDB statistical analysis (ver 1.6.1) with built in human ligand-receptor database(*19*). Potential ligand–receptor pairs, cell–cell interactions were identified based on mass action models, along with differential expression analysis and statistical tests on cell groups. We used the standard workflow; Preprocessing functions, identifyOverExpressedGenes and identifyOverExpressedInteractions with the default parameters were applied followed by communication probability computation between interacting cell groups with tri-mean set at 0.1. Subsequently we ran the core functions computeCommunProbPathway and aggregateNet using the standard parameters. Finally, the functions, netVisual_bubble were applied to the network to determine the senders and receivers and to generate a LR interaction plot.

Data availability

The analysis results are presented as a dynamic online resource in Multi-Dimensional Viewer (MDV) (https://mdv.molbiol.ox.ac.uk/projects/mdv_project/7430). All source data for the figures are found within the project 7430 in the MDV link.

Code availability

The complete code for the Spatial Omics Oxford (SpOOx) pipeline and the SpOOx v2.0 (as an update) is available as a GitHub repository under the GPL license – https://github.com/LingPeiHo/Ho_Taylor_Byrne-SpOOx-2.0. The Multi-Dimensional Viewer code is available under the GPL license: https://github.com/Taylor-CCB-Group/MDV.

Software source and identifier:

Name of software	Source	Identifier
imctools	https://github.com/BodenmillerGroup/imctools	RRID:SCR_017132
Deepcell	https://vanvalen.github.io/about/	RRID:SCR_022197
Phenograph	https://github.com/JinmiaoChenLab/Rphenograph	RRID:SCR_016919
Harmony	https://github.com/slowkow/harmonypy	RRID:SCR_022206
Ruffus	http://www.ruffus.org.uk/	RRID:SCR_022196
QuPath	https://qupath.github.io/	https://doi.org/10.1038/s41598-017-17204-5
MCD	https://www.standardbio.com/products-services/software	RRID:SCR_023007
Catalyst R	http://bioconductor.org/packages/CATALYST/	RRID:SCR_017127
Harmony	https://github.com/immunogenomics/harmony	RRID:SCR_022206
diffcyt R package (version 1.8.8)	https://www.bioconductor.org/packages/release/bioc/html/diffcyt.html	RRID:SCR_023006
Cellchat	https://github.com/jinworks/CellChat	RRID:SCR_021946
Seurat (version 4.0)	https://satijalab.org/seurat/	RRID:SCR_007322

Legends

Figure 1. Overall outline of study, the regenerating alveolar niche in IPF lungs and its cellular constituents . A-K overall workflow of study from macroscopic selection of lung biopsy (B) to final analysis(K). **L.** Examples of ROIs from lung sections obtained in Early, Intermediate (Intm) and Advance (Adv) lung regions (shown in B) showing histopathology features of collagen deposition, alveolar metaplasia, inflammatory infiltrate and smooth muscle hyperplasia (top panel). Matching MCD images stained with metal tagged antibodies – five of 33 markers shown. **M.** Representative ROI from an H&E-stained IPF lung section showing criteria for ROI selection (presence of AT II metaplasia, immune cell infiltrate and lack of airways. **N**. Metal-tagged antibody panel used for Imaging Mass Cytometry (IMC) staining **O.** UMAP (front and back of 3 dimensional projection) of annotated cell clusters derived from 33 plex IMC staining. H&E – hematoxylin and eosin, FFPE – formalin fixed paraffin embedded, AT II – type 2 alveolar epithelial cells, Neut – neutrophils, mac – macrophages, mono – monocytes, SM – smooth muscle, ABI – aberrant basal intermediates.

Figure 2. Composition of alveolar epithelial and immune cells and their correlations in regenerating alveolar niche in IPF lungs. A. Representative ROI from Early disease stage showing KRT17/7 and CD206 protein expression (IMC staining). Two areas of interest (AOIs) ('a' and 'b') are highlighted for later morphological analysis. **B-E.** Corresponding cell centroid maps for ROI in (A) generated using cell outline mask for the ROI, highlighting the location of ABIs, basal cells and AT II cells. No ABI_b is visible in keeping with very low number of ABI_b. **F**. Expression density plot (scaled between 0 and 1) for markers EPCAM, KRT17/7, KRT5, Ki67 and ProSP-C for all 5 defined epithelial cell clusters in alveolar niche. Blue broken vertical lines represent point above background staining. Note ATII cells are the only cells with high ProSP-C expression and have no or low KRT5 staining by further gene expression analysis (see fig. S3B-C) and immunofluorescence staining (fig. S6), hence the higher level of background staining for KRT5. **G**. High magnification of H&E images from AOI 'a' and 'b' from figure 2A showing the morphology of ABI_a and ABI_b-DC ADJ cells which are in keeping with formal histopathology feature of type II alveolar metaplasia, representing abnormal alveolar regeneration. AM- alveolar macrophages **H.** Proposed points for ABIs, basal cells and ATII cells on differentiation trajectory of abnormal alveolar epithelium as shown by Kathiriya's (*11*) human alveolar organoid studies. Broken blue boxes identifies markers used in our study. Diagram adapted from Kathiriya JJ et al 2021(*11*); AEC 2 – type 2 alveolar epithelial cells. **I**. Proportion (mean, for all 53 ROIs) of different epithelial and immune cells (as % of total epithelial and immune cells respectively) **J.** Number (cells/mm^2 of tissue) of immune (left graph) and epithelial (right graph) cells in Early, Intm and Adv disease stages. *** $p<0.001$ **$p<0.01$; comparison derived using generalised linear model (GLM) model, with area as a normalizing factor. **K.** Correlation in abundance of cells (per mm^2 of tissue) between alveolar epithelial cell types and immune cells. Only significant correlation shown; values are 'r' derived using Pearson's correlation. Only positive correlation shown. **$p<0.01$; ***$p<0.001$; Pearson's correlation with Benjamini-Hochberg correction for multiple comparisons**.** Note no significant correlations were observed between ATII cells and any immune cells. **I.** Correlation plots for the pairs of ABI-immune cell pairs with top three highest r values for each disease stage

Figure 3. Spatial co-location between aberrant regenerating alveolar epithelial cells and immune cells. A. Work flow and findings for each step, from derivation of 313,818 single cells and their location on lung tissue to identification of three co-locating pairs of alveolar epithelial-immune cell types. **B.**

Heat map showing g(r=20) values for all structural cell types vs all immune cells. Red rectangular outline indicates g(r=20) for alveolar epithelial vs immune cell types. Broken white circles indicate alveolar epithelial - immune cell pairs with g(r=20) >1. Red filled-in arrows indicate these pairs. Black squares indicate no co-location above complete spatial randomness [g(r=20)<1]. Only targeted cell pairs shown for ease of reading. Full spatial data in table S6. **C.** g(r) plots for the 5 co-locating alveolar epithelial-immune cell pairs which shows g(r) for r=0-300μm from cell centroid of specified ABIs. g_max is the peak g(r) value in the area around cell centroid of specific cell type described by r=0-300um; g(r=20) is value of g(r) at 20μm. **D-F.** Cell centroid maps for CD206^{hi} macrophage-ABI_b-DC ADJ pairs across the three disease stages. This is matched to the topographical correlation map (TCM) which visualises how the spatial correlation between cells of types A and B changes across an ROI. Γ_{ab} indicate spatial proximity of cells of type B to cells of type A within a radius of r, in this case 100μm; high values indicate close spatial proximity and negative values indicate that there are few cells of type B in proximity to cells of type A. In this TCM, our cell of type A is ABI_b-DC ADJ and cell type B is CD206^{hi} macrophages. **G.** Output for Neighbourhood Correlation Function (NCF) which interrogates co-location of 3 (rather than 2) cell types; here, any three cell types of the cell types ABI_b-DC ADJ ,ABI_a, CD206^{hi} macrophages, CD206^{mid} macrophages and CD103^{+} CD4 T cells, selected due to their co-location as pairs (Fig 3B). Graphs shows $NCF_{C_1C_2C_3}(r)$ for all triplet ($C_1C_2C_3$) combinations of ABI (ABI_b-DC ADJ and ABI_a) to immune cells (CD206^{hi} macrophages, CD103^{+} CD4 T cells) at MEC radius of 20 μm (top panel) and 30 μm (bottom panel). Green pane – triplets containing CD103^{+} CD4 T cells; blue pane – triplets without CD103^{+} CD4 T cells **H-I.** Individual $NCF_{C_1C_2C_3}(r)$plots for r = 0 to 150μm for triplets that are most strongly spatially associated with each other compared to one which did not show association (**J**)

Figure 4. Cellular neighborhood of aberrant regenerating alveolar epithelial cells. A-B. Donut charts show g(r=20) values for all structural, epithelial and immune cell types that are co-located above CSR with ABI_a cell type (A) and ABI_b-DC ADJ cell type (B) for the different stages of disease. Size of segment is proportional to the relative value of g(r=20). **C.** Cell centroid maps showing the location of all cell types depicted in A and B in one representative ROI from Early, Intm and Adv for ABI_b(C) and ABI_b-DC ADJ (D). Colour of the different cell types are matched to the colour of the cell types in the donut charts in A and B.

Figure 5. Cell contact networks. A-C Output for Cell Contact Network (CCN) for alveolar epithelial - immune cell pairs, depicting pairs of cell type (A; here alveolar epithelial cell) which are in contact with at least one of cell type B (immune cell types) at a frequency greater than expected under complete spatial randomness (CSR), and where z score between observed and random contact has p<0.05. Presence of a bubble for any pairs of contact above CSR indicates that there is a significant increase in frequency. Size of bubble is proportional to z score and intensity of colour, to the FDR for the z score. Broken circle identifies pairs of alveolar epithelial cell-immune cell types. All other bubbles picks out immune-immune or epithelial- epithelial pairings. Pairs with 'UD' cell clusters are not circled. Left upper quadrant - 'anchor cells' are alveolar epithelial cell; right lower quadrant output mirrors the left upper quadrant but 'anchor cells' are immune cell types. **D-F**. CCN network for all cell type pairs showing 'communities' of directly contacting cell pairs (grey and lime coloured spaces). Connecting lines join cell type pairs that are in direct contact with each other with greater frequency than complete spatial randomness, and with z scores that have FDR p values <0.05. Size of nodes represents number of cells per mm^2 , thickness of lines represents number of contacts between cell types A and

B. Grey spaces – communities of immune cells (purple nodes); lime spaces – communities of alveolar epithelial cells (yellow nodes). Note we have included the UD's in the network to show all cell types . Broken circles highlight ABI_b-DC ADJ and CD206^{hi} macrophage pairs, which at all stages, link up aberrant alveolar epithelial communities with immune cell communities.

Figure 6. Receptor-ligand analysis between aberrant basaloid intermediates and CD206^{hi} macrophages. **A.** Data extracted from of single cell RNA sequencing of lung digest cells from IPF (n=12) and healthy control (n=12) lungs deposited by Habermann et al (Sci Adv 2020) (*9*) reproduced here to demonstrate starting point for further clustering of macrophages. This UMAP retains Habermann's annotations. Text in red identifies Habermann's aberrant basaloid intermediate cells (KRT5^{-}KRT17^{+} cells) and transitional AT II cells, and in blue the two clusters of DCs. Monocytes (Mono) and Macrophage (Mac) cell clusters in broken blue circle are re-clustered (using 0.7 resolution in Seurat) to discriminate levels of CD206 (MRC1) gene expression and where there is a clear CD206^{hi} subcluster (subcluster 9) shown in (**B**) and (**C**). Cluster 9 (in broken box) **(C)** showing the highest CD206 expression is annotated as CD206^{hi} macrophages. **D-E.** Receptor-ligand analysis using CellChat, with ABI as sender and CD206^{hi} macrophages, cDC and pDC as receivers (D) and vice versa (E). Red circles - highest probability of communication (>0.5); large bubbles indicate probability of communication >0.3, and small bubble <0.3; no bubble – no communication detected. Colour of the bubble represents levels of communication probability. Y axis – ligand-receptor; x axis ABI - immune cell pairs. tAT2 – Transitional AT II; prob – communication probability **F**. Marker gene list for Cluster 9 (CD206^{hi} macrophages) compared to the established macrophage subclusters (FABP4+ macrophage and SPP1+ macrophage)(*8, 22, 23*) found in IPF. Note that cluster 9 (CD206^{hi} macrophages) shows a distinct difference in fibronectin (FN) expression compared to FABP4+ macrophage and SPP1+ macrophage subsets. **G.** Confocal picture of immunofluorescent staining of FN in CD206^{+} macrophages in one of our IPF lungs (one of 16 ROIs shown, n=3 patients). **H.** Schema proposing location of ABIs and CD206^{hi} macrophages, and functional consequences of receptor-ligand engagement. MIF – macrophage migration inhibitory factor, mac – macrophages.

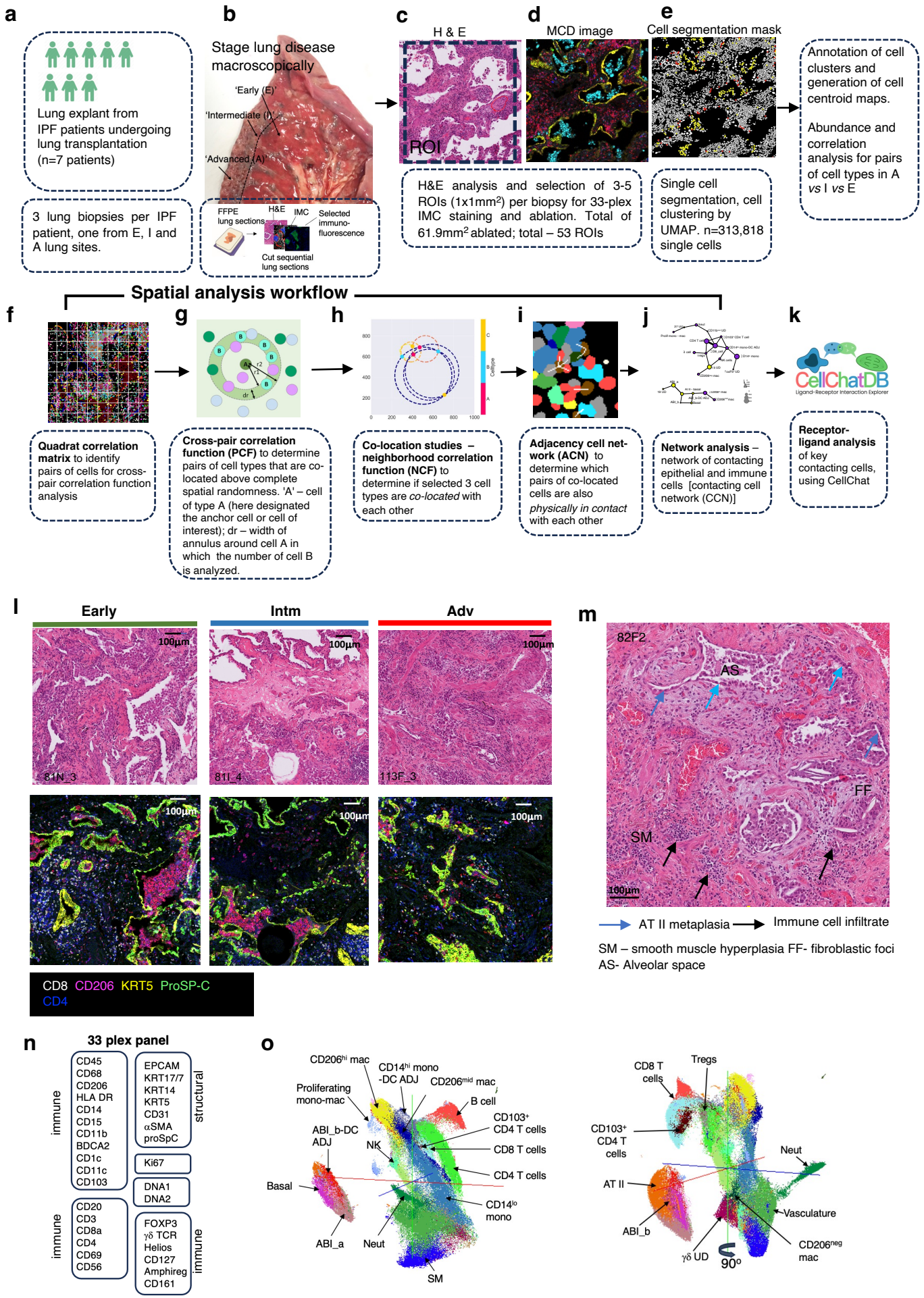
a
Lung explant from IPF patients undergoing lung transplantation (n=7 patients)
3 lung biopsies per IPF patient, one from E, I and A lung sites.
b
Stage lung disease macroscopically
'Early (E)'
'Intermediate (I)'
'Advanced (A)'
FFPE lung sections
H&E
IMC
Selected immuno-fluorescence
Cut sequential lung sections
c
H & E
ROI
H&E analysis and selection of 3-5 ROIs (1x1mm²) per biopsy for 33-plex IMC staining and ablation. Total of 61.9mm² ablated; total – 53 ROIs
d
MCD image
e
Cell segmentation mask
Single cell segmentation, cell clustering by UMAP. n=313,818 single cells
Annotation of cell clusters and generation of cell centroid maps.
Abundance and correlation analysis for pairs of cell types in A vs I vs E
Spatial analysis workflow
f
Quadrat correlation matrix to identify pairs of cells for cross-pair correlation function analysis
g
Cross-pair correlation function (PCF) to determine pairs of cell types that are co-located above complete spatial randomness. 'A' – cell of type A (here designated the anchor cell or cell of interest); dr – width of annulus around cell A in which the number of cell B is analyzed.
h
Co-location studies – neighborhood correlation function (NCF) to determine if selected 3 cell types are co-located with each other
i
Adjacency cell network (ACN) to determine which pairs of co-located cells are also physically in contact with each other
j
Network analysis – network of contacting epithelial and immune cells [contacting cell network (CCN)]
k
CellChatDB
Ligand-Receptor Interaction Explorer
Receptor-ligand analysis of key contacting cells, using CellChat
l
Early
Intm
Adv
100µm
81N_3
81I_4
113F_3
CD8 CD206 KRT5 ProSP-C
CD4
m
82F2
AS
FF
SM
AT II metaplasia
Immune cell infiltrate
SM – smooth muscle hyperplasia FF- fibroblastic foci
AS- Alveolar space
n
33 plex panel
immune
CD45
CD68
CD206
HLA DR
CD14
CD15
CD11b
BDCA2
CD1c
CD11c
CD103
structural
EPCAM
KRT17/7
KRT14
KRT5
CD31
αSMA
proSpC
Ki67
DNA1
DNA2
immune
CD20
CD3
CD8a
CD4
CD69
CD56
immune
FOXP3
γδ TCR
Helios
CD127
Amphireg
CD161
o
CD206hi mac
CD14hi mono -DC ADJ
CD206mid mac
Proliferating mono-mac
B cell
CD103+ CD4 T cells
ABI_b-DC ADJ
NK
CD8 T cells
CD4 T cells
Basal
CD14lo mono
ABI_a
Neut
SM
CD8 T cells
Tregs
CD103+ CD4 T cells
Neut
AT II
Vasculature
ABI_b
γδ UD
CD206neg mac
90°

Figure 1

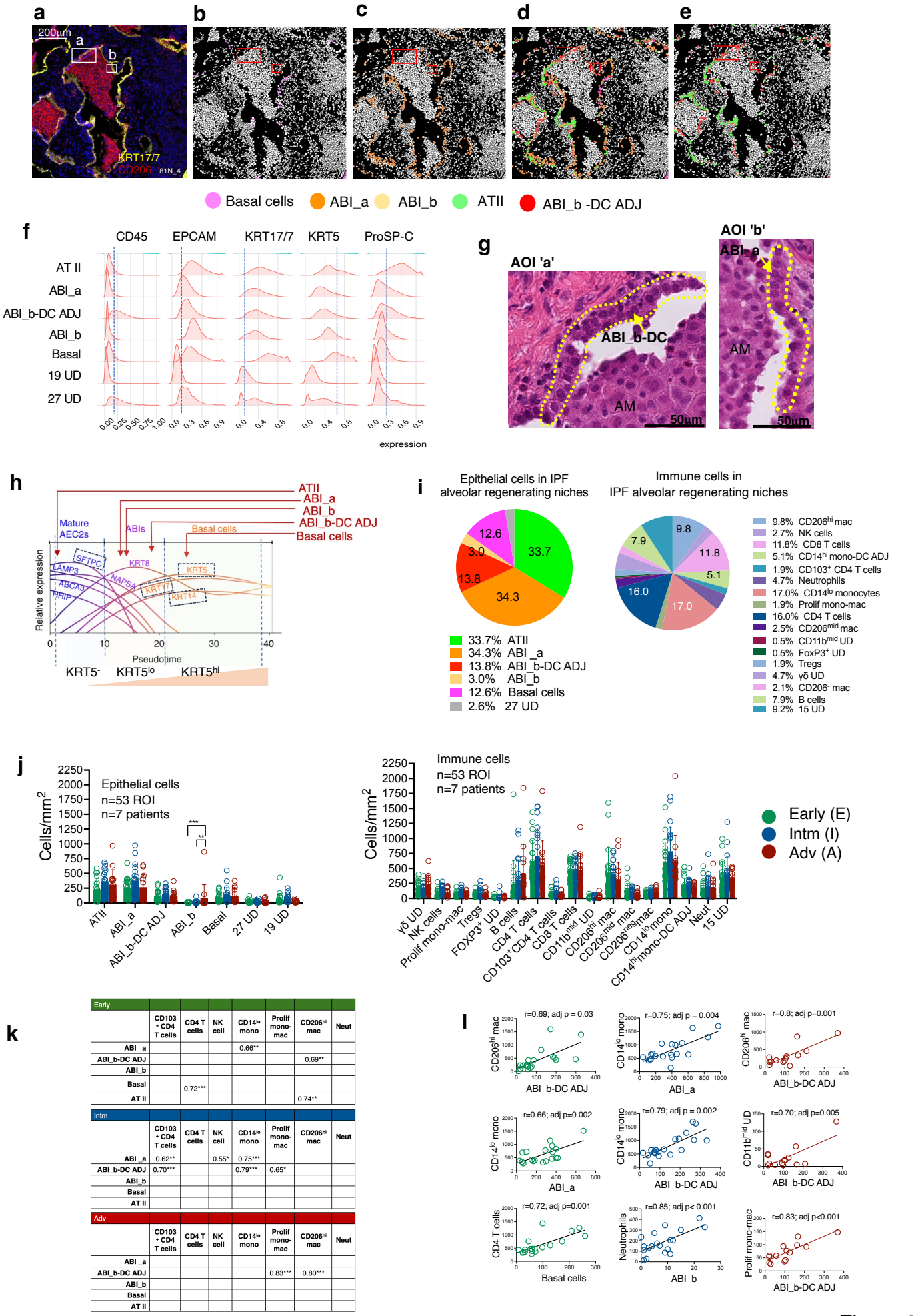
a
b
c
d
e
200µm
KRT17/7
CD206
81N_4
Basal cells
ABI_a
ABI_b
ATII
ABI_b -DC ADJ
f
CD45
EPCAM
KRT17/7
KRT5
ProSP-C
AT II
ABI_a
ABI_b-DC ADJ
ABI_b
Basal
19 UD
27 UD
expression
g
AOI 'a'
AOI 'b'
ABI_b-DC
AM
50µm
ABI_a
h
ATII
ABI_a
ABI_b
ABI_b-DC ADJ
Basal cells
Mature AEC2s
ABIs
Basal cells
SFTPC
LAMP3
ABCA3
HTIP
KRT8
NAPSA
KRT17
KRT14
KRT5
Relative expression
Pseudotime
KRT5-
KRT5lo
KRT5hi
i
Epithelial cells in IPF alveolar regenerating niches
Immune cells in IPF alveolar regenerating niches
33.7% ATII
34.3% ABI _a
13.8% ABI_b-DC ADJ
3.0% ABI_b
12.6% Basal cells
2.6% 27 UD
9.8% CD206hi mac
2.7% NK cells
11.8% CD8 T cells
5.1% CD14hi mono-DC ADJ
1.9% CD103+ CD4 T cells
4.7% Neutrophils
17.0% CD14lo monocytes
1.9% Prolif mono-mac
16.0% CD4 T cells
2.5% CD206mid mac
0.5% CD11bmid UD
0.5% FoxP3+ UD
1.9% Tregs
4.7% γδ UD
2.1% CD206- mac
7.9% B cells
9.2% 15 UD
j
Epithelial cells
n=53 ROI
n=7 patients
Cells/mm2
ATII
ABI_a
ABI_b-DC ADJ
ABI_b
Basal
27 UD
19 UD
Immune cells
n=53 ROI
n=7 patients
Early (E)
Intm (I)
Adv (A)
γδ UD
NK cells
Prolif mono-mac
Tregs
FOXP3+ UD
B cells
CD4 T cells
CD103+CD4 T cells
CD8 T cells
CD11bmid UD
CD206hi mac
CD206mid mac
CD206neg mac
CD14lo mono
CD14hi mono-DC ADJ
Neut
15 UD
k
Early
Intm
Adv
l
r=0.69; adj p = 0.03
r=0.75; adj p = 0.004
r=0.8; adj p=0.001
r=0.66; adj p=0.002
r=0.79; adj p = 0.002
r=0.70; adj p=0.005
r=0.72; adj p=0.001
r=0.85; adj p< 0.001
r=0.83; adj p<0.001
CD206hi mac
CD14lo mono
CD11bmid UD
CD4 T cells
Neutrophils
Prolif mono-mac
ABI_b-DC ADJ
ABI_a
ABI_b
Basal cells

Figure 2

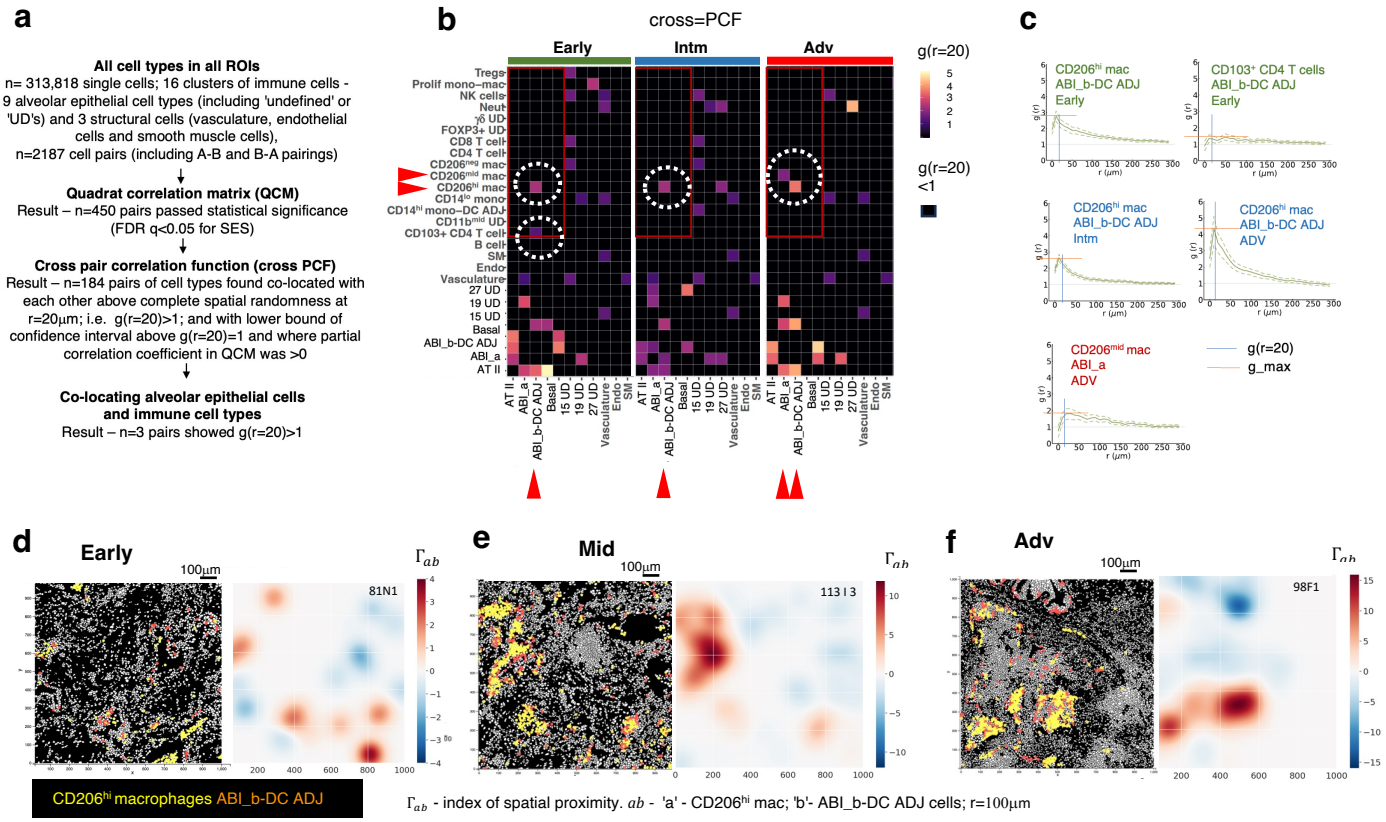
a
All cell types in all ROIs
n= 313,818 single cells; 16 clusters of immune cells - 9 alveolar epithelial cell types (including 'undefined' or 'UD's) and 3 structural cells (vasculature, endothelial cells and smooth muscle cells),
n=2187 cell pairs (including A-B and B-A pairings)
Quadrat correlation matrix (QCM)
Result – n=450 pairs passed statistical significance (FDR q<0.05 for SES)
Cross pair correlation function (cross PCF)
Result – n=184 pairs of cell types found co-located with each other above complete spatial randomness at r=20μm; i.e. g(r=20)>1; and with lower bound of confidence interval above g(r=20)=1 and where partial correlation coefficient in QCM was >0
Co-locating alveolar epithelial cells and immune cell types
Result – n=3 pairs showed g(r=20)>1
b
cross=PCF
Early
Intm
Adv
g(r=20)
g(r=20) <1
c
CD206hi mac ABI_b-DC ADJ Early
CD103+ CD4 T cells ABI_b-DC ADJ Early
CD206hi mac ABI_b-DC ADJ Intm
CD206hi mac ABI_b-DC ADJ ADV
CD206mid mac ABI_a ADV
g(r=20)
g_max
d
Early
100μm
81N1
e
Mid
113 I 3
f
Adv
98F1
CD206hi macrophages ABI_b-DC ADJ
Γab - index of spatial proximity. ab - 'a' - CD206hi mac; 'b'- ABI_b-DC ADJ cells; r=100μm
g
Early
Mid
Adv
colocation observed for 3 cell types
no-colocation
Combination of triplets of ABI and immune cell types
h
r=10 r=20
ABI_b-DC ADJ
CD206hi mac
CD206mid mac
INTM
r(μm)
i
ABI_a
ABI_b-DC ADJ
CD206hi mac
ADV
j
ABI_a
CD103+CD4 T cells
CD206mid mac
INTM
10 – ABI_b-DC ADJ; 1 – CD206hi mac; 8 – ABI_a; 22 – CD206mid mac;
12 – CD103+ CD4 T cells

Figure 3

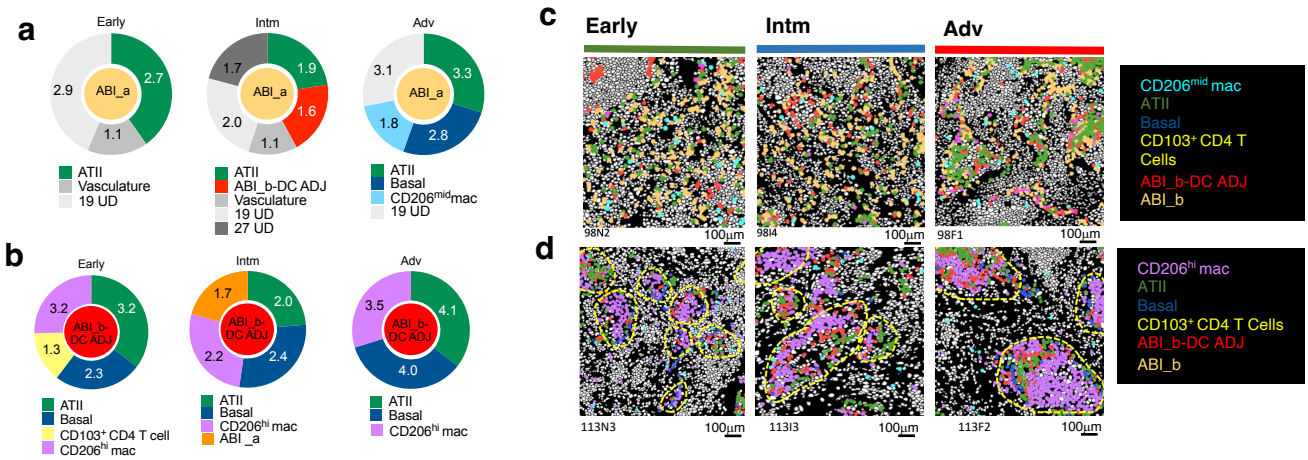
a
Early
ABI_a
2.7
1.1
2.9
ATII
Vasculature
19 UD
Intm
ABI_a
1.9
1.6
1.1
2.0
1.7
ATII
ABI_b-DC ADJ
Vasculature
19 UD
27 UD
Adv
ABI_a
3.3
2.8
1.8
3.1
ATII
Basal
CD206mid mac
19 UD
b
Early
ABI_b-DC ADJ
3.2
2.3
1.3
3.2
ATII
Basal
CD103+ CD4 T cell
CD206hi mac
Intm
ABI_b-DC ADJ
2.0
2.4
2.2
1.7
ATII
Basal
CD206hi mac
ABI _a
Adv
ABI_b-DC ADJ
4.1
4.0
3.5
ATII
Basal
CD206hi mac
c
Early
Intm
Adv
98N2
98I4
98F1
100µm
CD206mid mac
ATII
Basal
CD103+ CD4 T Cells
ABI_b-DC ADJ
ABI_b
d
113N3
113I3
113F2
100µm
CD206hi mac
ATII
Basal
CD103+ CD4 T Cells
ABI_b-DC ADJ
ABI_b

Figure 4

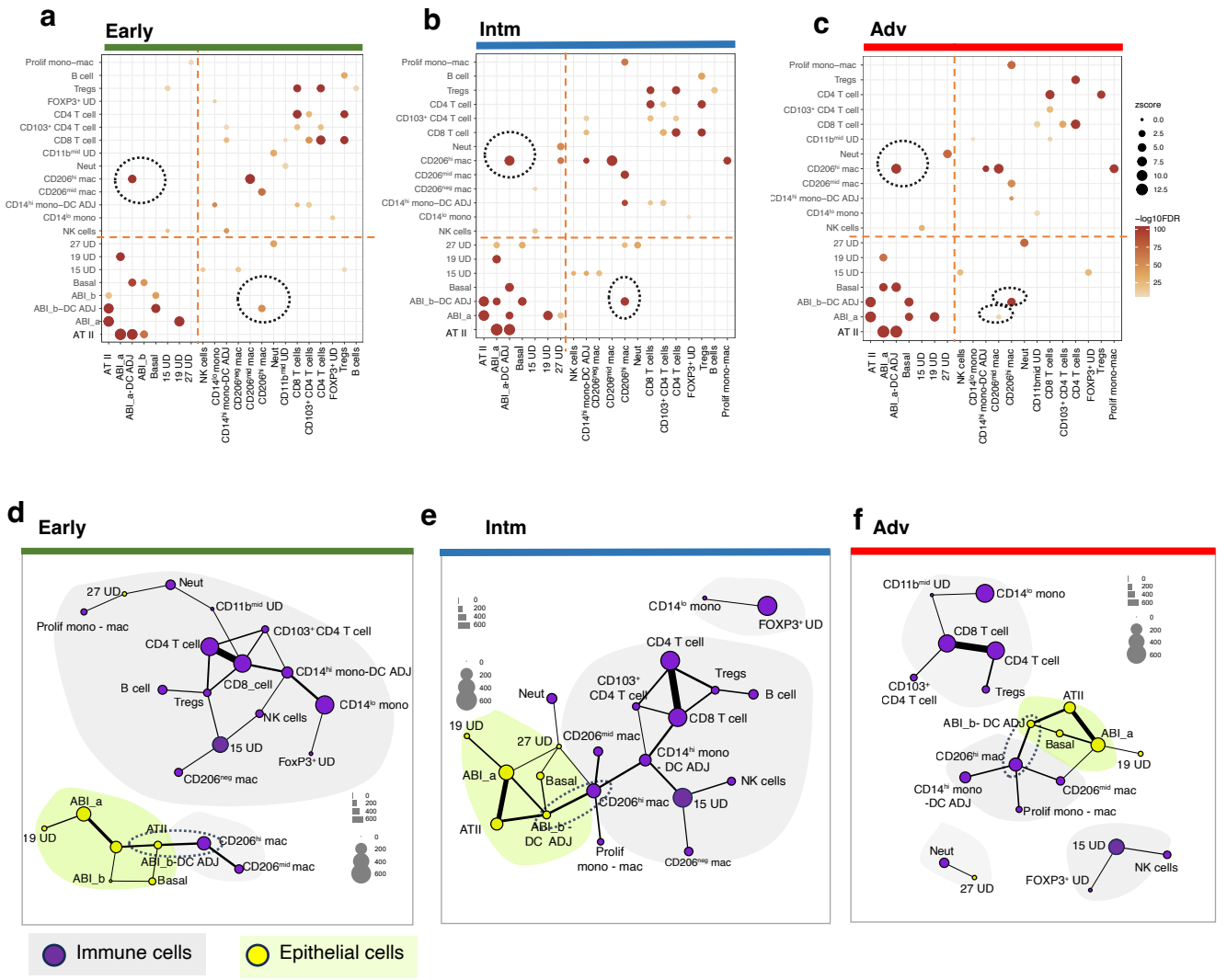
a
Early
b
Intm
c
Adv
zscore
−log10FDR
d
Early
e
Intm
f
Adv
Immune cells
Epithelial cells

Figure 5

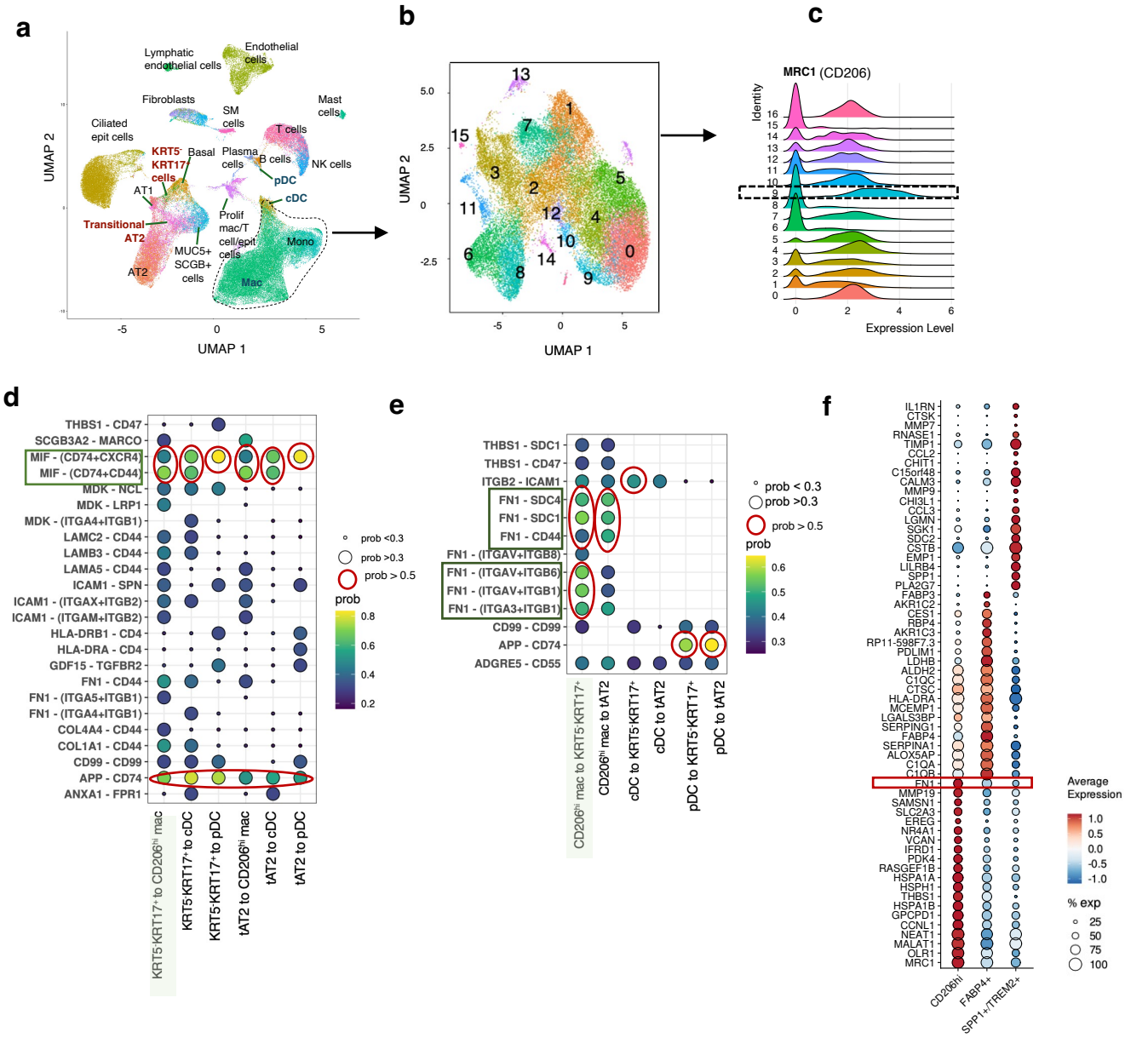
a
Lymphatic endothelial cells
Endothelial cells
Fibroblasts
SM cells
Mast cells
Ciliated epit cells
T cells
KRT5- KRT17+ cells
Basal
Plasma cells
B cells
NK cells
pDC
cDC
AT1
Transitional AT2
Prolif mac/T cell/epit cells
Mono
MUC5+ SCGB+ cells
AT2
Mac
UMAP 2
UMAP 1
b
UMAP 2
UMAP 1
c
MRC1 (CD206)
Identity
Expression Level
d
THBS1 - CD47
SCGB3A2 - MARCO
MIF - (CD74+CXCR4)
MIF - (CD74+CD44)
MDK - NCL
MDK - LRP1
MDK - (ITGA4+ITGB1)
LAMC2 - CD44
LAMB3 - CD44
LAMA5 - CD44
ICAM1 - SPN
ICAM1 - (ITGAX+ITGB2)
ICAM1 - (ITGAM+ITGB2)
HLA-DRB1 - CD4
HLA-DRA - CD4
GDF15 - TGFBR2
FN1 - CD44
FN1 - (ITGA5+ITGB1)
FN1 - (ITGA4+ITGB1)
COL4A4 - CD44
COL1A1 - CD44
CD99 - CD99
APP - CD74
ANXA1 - FPR1
KRT5-KRT17+ to CD206hi mac
KRT5-KRT17+ to cDC
KRT5-KRT17+ to pDC
tAT2 to CD206hi mac
tAT2 to cDC
tAT2 to pDC
prob <0.3
prob >0.3
prob > 0.5
prob
e
THBS1 - SDC1
THBS1 - CD47
ITGB2 - ICAM1
FN1 - SDC4
FN1 - SDC1
FN1 - CD44
FN1 - (ITGAV+ITGB8)
FN1 - (ITGAV+ITGB6)
FN1 - (ITGAV+ITGB1)
FN1 - (ITGA3+ITGB1)
CD99 - CD99
APP - CD74
ADGRE5 - CD55
CD206hi mac to KRT5-KRT17+
CD206hi mac to tAT2
cDC to KRT5-KRT17+
cDC to tAT2
pDC to KRT5-KRT17+
pDC to tAT2
prob < 0.3
prob >0.3
prob > 0.5
prob
f
IL1RN
CTSK
MMP7
RNASE1
TIMP1
CCL2
CHIT1
C15orf48
CALM3
MMP9
CHI3L1
CCL3
LGMN
SGK1
SDC2
CSTB
EMP1
LILRB4
SPP1
PLA2G7
FABP3
AKR1C2
CES1
RBP4
AKR1C3
RP11-598F7.3
PDLIM1
LDHB
ALDH2
C1QC
CTSC
HLA-DRA
MCEMP1
LGALS3BP
SERPING1
FABP4
SERPINA1
ALOX5AP
C1QA
C1QB
FN1
MMP19
SAMSN1
SLC2A3
EREG
NR4A1
VCAN
IFRD1
PDK4
RASGEF1B
HSPA1A
HSPH1
THBS1
HSPA1B
GPCPD1
CCNL1
NEAT1
MALAT1
OLR1
MRC1
CD206hi
FABP4+
SPP1+/TREM2+
Average Expression
% exp
g
10 μm
CD206 FN DNA
Fn - fibronectin
h
CD206hi mac
ABIs
AT2
Fibroblast
CD103+ CD4 T cells
DCs
AT1
receptor
Anti-apoptosis
Counter-regulation of glucocorticoid activity
↑Pro-inflammatory cytokines
MIF
CD74/CXCR4
CD74/CD44
Fn
SDC1
SDC4
CD44
ITGAV+ITGB6
ITGAV+ITGB1
ITGA3+ITGB1
Augment TGF-β and Wnt signaling (SDC1)
Downregulate TGF-β (SDC4)
Augment phagocytic responses of mac
Chemoattractant for fibroblasts and mac
Prime mac (CSF1rec phosphorylation)

Figure 6