ScRNA-seq Expression of *IFI27* and *APOC2* Identifies Four Alveolar Macrophage Superclusters in Healthy BALF

Xin Li[1], Fred W. Kolling[2], Daniel Aridgides[3], Diane Mellinger[3], Alix Ashare[1,3*], and Claudia V. Jakubzick[1*]

[1]Department of Microbiology and Immunology, Dartmouth Geisel School of Medicine, Hanover, NH [2]Department of Biomedical Data Science, Dartmouth Geisel School of Medicine, Hanover, NH [3]Department of Medicine, Dartmouth Hitchcock Medical Center, Lebanon, NH [*]Co-senior authors

Correspondences:

Claudia Jakubzick, PhD
Geisel School of Medicine at Dartmouth
Department of Microbiology and Immunology
626W Borwell
One Medical Center Drive
Lebanon, NH 03756
Phone: 303-898-3791
claudia.jakubzick@dartmouth.edu

Alix Ashare, MD, PhD
Geisel School of Medicine at Dartmouth
Dartmouth Hitchcock Medical Center
Department of Medicine
Department of Microbiology and Immunology
One Medical Center Drive
Lebanon, NH 03756
alix.ashare@hitchcock.org

Running title: ScRNA-seq Airspace Macrophage Heterogeneity

Keywords: Alveolar macrophages, monocytes, heterogeneity, scRNA-seq, supercluster, interstitial macrophages

1

Summary Blurb

There are at least 14 AM subtypes; their frequency, along with other immune cells, are highly conserved across individuals suggesting a specific niche exists for each leukocyte population.

Abstract

Alveolar macrophages (AMs) reside on the luminal surface of the airways and alveoli, ensuring proper gas exchange by ingesting cellular debris and pathogens, and regulating inflammatory responses. Therefore, understanding the heterogeneity and diverse roles played by AMs, interstitial macrophages (IMs), and recruited monocytes is critical for treating airway diseases. We performed single-cell RNA sequencing on 113,213 bronchoalveolar lavage cells from four healthy and three uninflamed cystic fibrosis subjects and identified FOLR2$^+$SELENOP$^+$ and SPP1$^+$PLA2G7$^+$ IMs, monocyte subtypes, and dendritic cell 1 (DC1), DC2, migDCs, plasmacytoid DCs, lymphocytes, epithelial cells, and four AM superclusters (families) based on the expression of *IFI27* and *APOC2* genes. These 4 AM families have at least eight distinct functional members (subclusters) named after their differentially expressed gene(s): IGF1, CCL18, CXCL5, Cholesterol, Chemokine, Metallothionein, Interferon and small-cluster AMs. Interestingly, the Chemokine cluster further divides with each subcluster selectively expressing a unique combination of chemokines. One of the most striking observations, besides the heterogeneity, is the conservation of AM family members in relatively equal ratio across all AM superclusters and individuals. Transcriptional data and TotalSeq technology were used to investigate cell surface markers that distinguish resident AMs from recruited monocytes. Lastly, other AM datasets were projected onto our dataset. Similar AM superclusters and functional subclusters were observed, along with changes in AM subclusters in individuals infected with COVID-19. Overall, functional specializations of the AM subclusters suggest that there are highly regulated AM niches with defined programming states, highlighting a clear division of labor.

Graphical Abstract
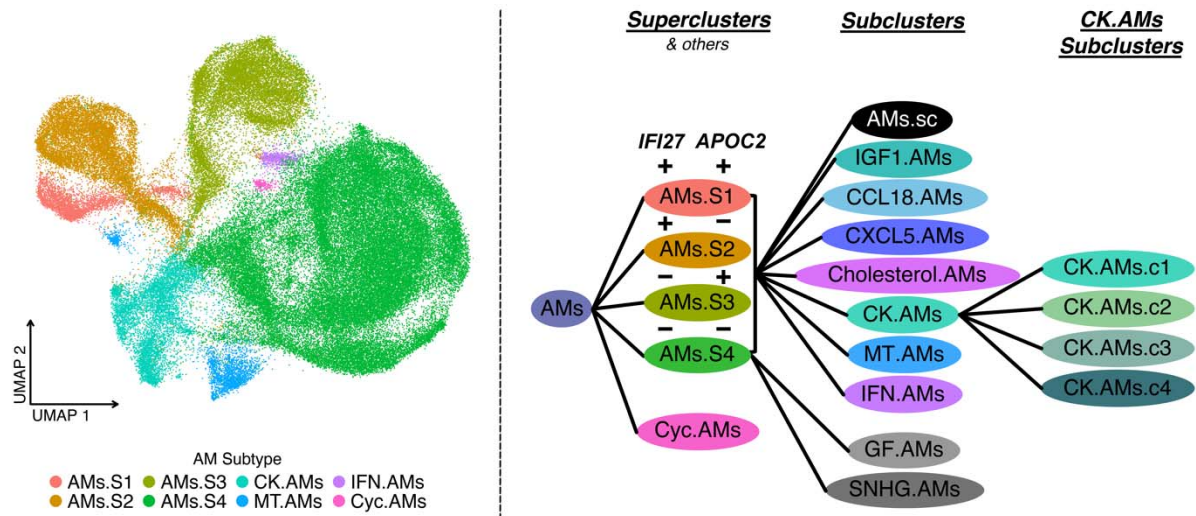
Introduction

Alveolar macrophages (AMs) are the primary phagocytes within the airspace of the lungs. In response to inflammatory stimuli, AMs secrete cytokines and chemokines to initiate the immune response. Subsequently, AMs and recruited monocytes contribute to the resolution of the inflammatory response to prevent excess damage (Fujii et al., 2021; Hetzel et al., 2021; Janssen et al., 2011; Mould et al., 2017). In addition, they play a major role in maintaining lipid homeostasis in the lung, which is essential for adequate gas exchange and alveolar epithelial integrity.

AMs may hold keys to understanding, preventing, and curing several diseases. They derive from prenatal monocytes, self-renew, and have shown remarkable plasticity by altering their transcriptome in response to the environmental changes (Aegerter et al., 2020; Arafa et al., 2022; Jakubzick et al., 2013; Janssen et al., 2011; Misharin et al., 2017; Mould et al., 2019; Yona et al., 2013). Their transcriptional profile is altered in various inflammatory lung diseases, including asthma (Fricker and Gibson, 2017; Hetzel et al., 2021), chronic obstructive pulmonary disease (O'Beirne et al., 2020), cystic fibrosis (CF) (McFarland and Rosenberg), and cancer (Bonfield, 2015; Bruscia and Bonfield, 2016; Deriy et al., 2009; Hey et al., 2021; Li et al., 2021; María Casanova-Acebes, 2021). Therefore, enhancing our current understanding of the normal cellular composition and functional diversity of AM subtypes will help characterize disease-related transcriptional changes more precisely and assess targeting specifically a given AM subtype to restore immune balance.

Single-cell RNA sequencing (scRNA-seq) technology is critical for accomplishing these goals. AMs have traditionally been thought to be a uniform population of cells that can be activated by different disease states (Mould et al., 2019). However, recent scRNA-seq studies have revealed a rich diversity in AMs in bronchoalveolar lavage fluid (BALF) from healthy subjects, with multiple subpopulations that have yet to be characterized in other diseases (Mould et al., 2021).

Our study was designed to investigate the immune cell populations of BALF and potential differences in individuals with mild CF compared to healthy controls (HC).

6

Interestingly, unbiased analysis of our scRNA-seq dataset identified four AM superclusters based on the expression of interferon alpha inducible protein 27 *(IFI27)* and apolipoprotein C2 (*APOC2)* genes, which has not been previously described. Each supercluster contains four shared subclusters named after their differentially expressed genes (DEGs): IGF1.AMs, CCL18.AMs, CXCL5.AMs, and Cholesterol.AMs. Beyond the four AM superclusters, we found four additional clusters, which can categorize into the AM superclusters: CK.AMs, MT.AMs, IFN.AMs, Cyc.AMs. The most striking AM subcluster was the CK.AMs (Chemokine cluster), which further divides into four subclusters with each selectively expressing a distinct combination of chemokines. Hence, there appears to be a division of labor in leukocyte recruitment. Moreover, the Chemokine subclusters suggest that there are distinct transcription factors (based on our regulon analysis) that regulate their differentiation and programming. This pattern is analogous to what is found in T cell biology, where distinct transcription factors regulate the cell differentiation pathway to produce a defined combination of cytokines known as Th1, Th2, Th17, etc. Overall, our dataset opens many new areas for further investigation, including addressing how resident macrophage and monocyte subtypes are altered during disease and what factors regulate their programming.

Results

BAL cell populations are conserved across healthy and mild CF subjects

To assess the cellular composition in the non-diseased airspace versus that with CF, four HCs and three subjects with mild CF underwent BAL (donors, Table 1). All three CF subjects had two copies of the F508del *CFTR* mutation. Study subjects were closely age-matched. CF subjects were selected for this study based on well-preserved lung function and general good health. To investigate the impact of aberrant CFTR function on the immune cell composition of the lung, we enrolled CF subjects without significant lung inflammation and who were not on CFTR modulator treatment. Two subjects had been on CFTR modulator treatment previously but stopped more than 12 months prior to enrollment due to side effects, while the third subject declined treatment with a CFTR modulator due to overall good health. The CF subjects in this study were on a stable medication regimen and had no exacerbations within the past 24 months.

BALF samples were spun, washed, and immediately loaded onto the 10X platform for scRNA-seq. Sequenced cells were processed, normalized, and integrated using the Seurat package (details in Methods). Uniform manifold approximation and projection (UMAP) of all seven samples illustrates a similar cluster distribution across all subjects (Figure 1A). Using a curated gene list, we identified twelve major cell types (Complete list of genes: Table S1 and Figure 1B-1D) (Brown et al., 2019; Cai et al., 2020; Collin and Bigley, 2018; Collin et al., 2013; Gautier et al., 2012; Gibbings et al., 2015; Gibbings et al., 2017; Jin et al., 2019; Leach et al., 2020; Mould et al., 2021; Schupp et al., 2020; Villani et al., 2017). As anticipated, the most abundant cell type is the alveolar macrophage, followed by monocytes, cycling myeloid cells, dendritic cell (DC) 2, FOLR2 and SPP1 interstitial macrophages, lymphocytes, epithelial cells, migratory DCs, DC1, plasmacytoid DCs, and cycling lymphoid cells (Figure 1C-1D). Next, MetaNeighbor analysis was used to confirm the classification and replicability of the identified clusters within our dataset and with two other recently published datasets, Mould et al., and Schupp et al. (Crow et al., 2018; Mould et al., 2021; Schupp et al., 2020). Within our dataset, samples derived from HC and CF subjects demonstrated a strong correlation

across all cell types (Figure 1E). Similarly, our HC sample macrophages, monocytes, and epithelial cells strongly correlates with healthy cell types from Mould et al., as do our CF sample cell types with CF sputum cell types (Figure 1F-1G). All in all, curated genes and robust alignment of our cells with previous studies support the identification of twelve cell types in healthy and CF BALF.

Four AM superclusters and shared subclusters co-exist in the airspace of the lungs

To define AM heterogeneity, we isolated and re-clustered the AM population. Unbiased clustering defined 8 AM clusters. Some clusters were previously described (Gardai et al.), such as the Metallothionein, Chemokine, IFN-reacting, and Cycling AM clusters. (Figure 2A-2D). In addition, we identified 4 novel AM superclusters based on the differential expression of two genes: *IFI27* and *APOC2* (Figure 2B-2E): *IFI27*$^+$*APOC*$^+$ AMs.S1, *IFI27*$^+$*APOC*$^-$ AMs.S2, *IFI27*$^-$*APOC*$^+$ AMs.S3, and *IFI27*$^-$*APOC*$^-$ AMs.S4. Besides *IFI27* and *APOC2*, the four AM superclusters share the same top DEGs *IFI6*, *CTSC*, *RPS4X*, *RBP4*, *FBP1*, and *GSN*, which suggest anti-microbial and metabolic function (i.e., common macrophage phenotypes) (Figure 2D).

Hierarchically, each AM supercluster contains 4 shared subclusters. The subclusters were named after their top DEG(s): IGF1, CCL18, CXCL5, and Cholesterol (based on the expression of cholesterol-biosynthesis-related genes) (Figure 3A-3C). AM Supercluster 4 (AMs.S4) contained two additional subclusters, GF and SNHG. Cells in the GF subcluster express the epithelial growth factor *AREG* and *TGFb* superfamily ligands, including, *GDF15*. This is consistent with observations made in mice, where Areg-secreting AMs contribute to tissue restoration after injury (Minutti et al., 2019; Xu et al., 2016). Whereas cells in SNHG expressed genes encoding small nucleolar RNA host genes (SNHGs) and genes encoding nuclear transcriptional regulators and transcription factors. The last subcluster referred to as the small-subcluster within the supercluster, noting a small 's' after AMs (For example, AMs.s# to differentiate them from their parental superclusters, AMs.S#) had hereditary DEGs, similar to their

9

corresponding superclusters. This small-subcluster was the supercluster minus all other subclusters (Figure 3D).

In addition to the AM superclusters, four other AM clusters can be defined based on their expression of Metallothioneins (MT.AMs), Chemokines (CK.AMs), IFN-related genes (IFN.AMs), and Cycling genes (CK.AMs) (Figure 2B). The chemokine cluster can be further divided into 4 subpopulations, which we denote as CK.AMs.c1-4. The top 6 DEGs and curated genes illustrate the unique chemokine expression pattern within each subcluster (Figure 4D). CK.AMs highly express *CCL18*, *CCL23,* and chemokine receptor *CXCR4* in CK.AMs.c1; *CCL3*, *CCL4*, *CXCL8,* and *CCL20* in CK.AMs.c2; *CXCL9*, *CXCL10* and *CXCL11* (IFN-gamma inducible chemokines) in CK.AMs.c3, and *CXCL3* in CK.AM.c4 (Figure 4B-4D). The expression of unique combinations of chemokines in each CK.AMs suggest that different subtypes of AMs are programmed to chemoattract distinct leukocytes into the airspace, depending on the inflammatory stimuli they encounter. Another scRNA-seq study described chemokine-expressing macrophages in human lung tumors: CXCL5-, CXCL12- and CXCL9/CXCL10/CXCL11-expressing macrophage subsets (Zilionis et al., 2019). Although these macrophages were not AMs, different macrophage subtypes secrete distinct combinations of chemokines.

Next, given the number of AM subtypes identified, we performed Gene Ontology (GO) enrichment analysis to infer functionality based on their DEGs (Figure 4E). Predictably, the Chemokine AM subclusters contained GO terms regarding cell interaction and response to stimuli, together with other enriched GO terms consistent with the top DEGs in each AM subtype. For instance, the top GO terms for Metallothionein AMs pertain to cellular responses to metal ions (Figure 4E and Figure S1). Likewise, the top 4 enriched terms for Cholesterol AMs contained "*cholesterol biosynthetic process*" and "*regulation of cholesterol biosynthetic process*" (Figure S1).

Lastly, to understand the transcription factors and co-factors driving each AM subtype, we performed single-cell level regulon analysis using the SCENIC (Single-Cell rEgulatory Network Inference and Clustering) package in R and its Python

implementation pySCENIC (Aibar et al., 2017). Co-expressed genes were used to identify and score regulons enriched in each cluster. For instance, the proinflammatory regulator of the cytokine-induced apoptotic pathway, *BACH2*, is the top regulon identified for CK.AMs.c2, while the *IRF7* regulon is enriched in IFN-reacting AMs (Figure 4F and Figure S1). This analysis offers yet another vantage point for understanding the distinct AM subtypes and corroborates their defined DEGs and GO terms.

Like AM Chemokine clusters, monocytes also display unique chemokine-expressing clusters

We performed unbiased clustering for monocytes, which outlined 8 distinct populations (Figure 5A-5C). The largest cluster is denoted as conventional monocytes (Conv.Mono) that express classical myeloid cell markers shared with the other seven clusters (Figure 5D). An additional six populations are defined as functional monocytes based on the expression of specific genes, including FCN1.Mono, CCL2.Mono, CXCL5.Mono, CXCL9.Mono, CCL13.Mono, and VEGFA.Mono (Figure 5D). To note, we also identified an intermediate monocyte cluster (Int.Mono) with distinct DEGs that encompasses cells transitioning from Conv.Mono to other functional monocyte phenotypes. Thus, monocytes appear to display a division of labor in orchestrating and recruiting leukocytes into the airspace, similar to what is observed for chemokine-expressing AMs.

IL10 is selectively secreted by two IM clusters

Other than AMs and monocytes, we observed several unexpected cell type clusters in the BALF, including IMs and DCs. The IMs in our dataset form two distinct clusters with distinct gene expression signatures, which we refer to as FOLR2.IMs and SPP1.IMs (Figure 1D and Figure 5D-5E). Genes such as *LGMN*, *MARCKS* (Figure 5E), *TMEM37,* and *MERTK* (data not shown) are specifically expressed by IMs, serving as potential markers to distinguish this population from other macrophage/monocyte cell types. It is worth mentioning that these two IM clusters are the main producer of IL10, suggesting their immunoregulatory role in the airspace.

11

Next, we performed gene ontology and regulon analysis on the monocyte and IM clusters (Figure S2). The most notable GO term was for FCN1.Mono, whose genes were enriched for bacterial defense mechanisms suggesting the importance of this monocyte subtype during pulmonary bacterial infections (Figure S2A). The GO analysis also illustrated the exogenous antigen presentation function in FOLR2.IMs and the inflammatory response genes in SPP1.IMs (Figure S2A).

Pseudotime analysis suggests monocytes differentiate into AMs and IMs

ScRNA-seq captures all cell types at a given time point, producing a series of expression profiles across cells as they transition from one state to another. To define the transitional processes of macrophage development, we performed pseudotime analysis using all the monocytes, AMs, and IMs in our dataset (Cao et al., 2019; Qiu et al., 2017a; Qiu et al., 2017b; Trapnell et al., 2014). The trajectory analysis revealed starting points of origin in the monocyte cluster (purple region and white open dots, Figure 6A), revealing two directions of differentiation: one toward IMs and the other toward cycling AMs and AMs.S4 (Figure 6A-6B). AMs.S4 appears to have its own starting branch point leading to all other AM clusters (Figure 6A-6B). This suggests that AMs.S4 may contribute to all other AM subtypes in the airspace, being the largest supercluster that does not express *IFI27* or *APOC2*.

To pinpoint the genes that change as a function of pseudotime trajectory, DEG analysis was applied. As monocytes differentiate into clusters of AMs and IMs, different genes are up-regulated sequentially, likely as a cause or consequence of changes in their cell state over time (Figure 6C). Genes for cell proliferation and myeloid cell function were among the top DEGs (Figure 6C). For example, *ISG15* is highly expressed during the later progression of pseudotime and is mainly up-regulated in IFN-reacting AMs. Whereas another among the top 10 DEGs, *MT2A*, change along pseudotime and associate with Metallothionein AMs. Interestingly, AM superclusters AMs.S1 and AMs.S2 seem to evolve much later than their counterparts AMs.S3 and AMs.S4, even though they share 4 subclusters and most gene expression, except *IFI27* and *APOC2* (Figure 6C).

Although scRNA-seq has revolutionized our knowledge of macrophage heterogeneity, there are instances where investigators need to isolate bulk populations to perform either functional or morphological analyses. Therefore, flow cytometric antibodies that can be used to sort and enrich bulk AMs from recruited monocytes are needed. Using flow cytometry, we demonstrate that, in addition to high side-scatter for AMs and low side-scatter for monocytes, investigators can use antibodies against CD43 (SPN) and CD169 to define AMs and CD93, CD36, and CD14 to identify monocytes (Figure 6D) (Bharat et al., 2016; Desch et al., 2015; Gibbings and Jakubzick, 2018; Yu et al., 2015; Yu and Tighe, 2018). ScRNA-seq and TotalSeq (for single-cell level protein expression assessment) also support the use of CD169 and CD43 for AMs compared to CD93, CD36, and CD14 for monocytes (Figure 6E and Figure S3). Thus, several forms of technology validate that the higher expression of CD43 and CD169 in BAL cells define AMs over recruited monocytes.

Bivariate expression of *IFI27* and *APOC2* also exist in other AM subtypes

Although our unbiased AM analysis in Figure 2B resulted in 8 clusters, four AM superclusters, and four other clusters (i.e., Chemokine, Metallothionein, Interferon, and Cycling AMs), we questioned whether the four other non-supercluster AMs express all four combinations of *IFI27* and *APOC2* (Figure 7A). Indeed, we observed a quadrant distribution of *IFI27* and *APOC2* for Chemokine, Metallothionein, and Interferon AMs (Figure 7B). This suggests that non-supercluster AMs clusters are a part of the four AM superclusters. And the reason this was not observed in the analysis in Figure 2B is because of the overpowering expression of other functional genes in the non-supercluster AMs. Based on this observation, we hypothesize that discrete compartments in the lung dictate the expression of *IFI27* and *APOC2,* and within these compartments, a family of AMs containing all the AM subtypes, exist to ensure proper physiological functions (Figure 7C).

Subsequently, we examined whether AM superclusters occur in aggregated HC or CF, and in each individual sample. Unbiased analysis of each sample, or combined samples, demonstrates the formation of four unique AM superclusters (Figure S4).

13

Thus, the robust recurrence of AM superclusters in each sample further validates our discovery and analysis.

AM Superclusters are conserved across different cohorts of publicly available datasets

Lastly, we examined other AM datasets to ensure that our findings are not unique to our dataset. Six distinct datasets were selected based on the tissue, processing procedures, health status and cell numbers (Figure 8A) (Mould et al., 2021)(Schupp, 2020 #865) (Dominguez Conde et al., 2022; Liao et al., 2020) (Vieira Braga et al., 2019) (Mayr et al., 2021). Within the samples, only AMs were selected and re-clustered for UMAP visualization. Alignment of the top DEGs from AM clusters was performed (Figure 8B). All datasets contained the known AM subtypes that expressed chemokine, metallothionein, interferon and cycling genes. However, the superclusters defining *IFI27* and *APOC2* expression were only observed in the datasets of Liao et al., and Conde et al., (Dominguez Conde et al., 2022; Liao et al., 2020). After accounting for sample size and sequencing depth, we hypothesize that the difference between the presence of AM superclusters from datasets that have it versus not might be based on processing procedures. Those that form superclusters are freshly loaded samples that did not lose APOC2 expression. It appears that the freezing and thawing BAL or lung samples will result in the loss of APOC2 mRNA expression and why superclusters will not form. Finally, using the Seurat "MapQuery" function, the datasets from other investigators were projected onto our UMAP visualization used in Figures 2-4 (Figure 8C-D, Figure S6). The overall distribution was assessed. As anticipated, due to the expression of APOC2, same tissue sample, and processing procedure, Liao et al., dataset displayed the best overlay (Liao et al., 2020). Interestingly, compared to healthy BAL, Liao et al., COVID-19 infected BAL demonstrated significantly more CK.AMs and IFN.AMs (Liao et al., 2020), suggesting the role of AMs in the development of uncontrolled inflammation, cytokine storm, and an imbalanced hyperactive immune system during SARS-CoV-2 infection.

14

Discussion

AMs have multiple functional roles, including clearing surfactant components, cellular debris, inhaled particulates, and pathogens. Therefore, it is not surprising that in the airspace, there would be multiple AMs and monocyte subtypes with different specializations. ScRNA-seq technology has revolutionized our understanding of AMs, and we are only beginning to appreciate the depth and complexity of airway macrophages and monocytes in healthy and diseased lungs. Here, we used scRNA-seq to show that AMs form 4 superclusters based on the expression of two genes, *IFI27* and *APOC2*, whether integrated or independent. *IFI27* is a protein-coding gene regulated by interferon and estradiol (Rasmussen et al., 1993). It has been reported that IFI27 protein plays diverse roles in anti-tumor immunity, including suppressing the transcription of PD-L1 and promoting epithelial-mesenchymal transition (Deng et al., 2020; Li et al., 2015). In the lung, *IFI27* has also been suggested as a biomarker to differentiate influenza and bacterial infections (Tang et al., 2017). *APOC2* encodes a lipid-binding protein belonging to the apolipoprotein gene family, functioning as a co-factor to activate lipoprotein lipase for the hydrolyzation of triglycerides. The unique expression pattern of *IFI27* and *APOC2* generates four AM superclusters. Each supercluster contains subclusters defined by characteristic DEGs, including those for cytokines and chemokines. In addition to AM superclusters, we observed four other AM clusters. Of particular interest is the Chemokine AM cluster, which divides further into four subclusters. Each of these chemokine clusters expressed a combination of chemokines that selectively chemoattract different leukocytes into the environment. Strikingly, there appears to be a defined niche for each AM subtype given the tightly regulated frequency of each AM subtype across all samples.

Why do the superclusters and subclusters exist? Given the limited literature on the functionality of *IFI27* and *APOC2*, it is hard to speculate on the contribution of these two genes. In addition to the hypothesis that each AM supercluster exists in a discrete compartment of the lung, AM superclusters might regulate their environment in a tissue-specific niche and functionally specific manner. It would be scientifically and clinically

15

interesting to know which CK.AM subclusters regulate the leukocyte environment in the airspace during and after infection with respiratory viruses like SARS-CoV-2.

Chemokines are chemoattractant cytokines, and this type of cytokine profiling resembles what has been observed for decades in T cell biology. Each T cell type, such as Th1 and Th2, selectively produces a characteristic set of cytokines. It is known that the differentiation of T cells is driven by distinct master regulators, transcription factors such as T-bet and Gata-3. Likewise, each AM profiled in this study outlined a distinct set of transcription factors observed in our regulon analysis. Therefore, an important future goal would be to investigate which transcription factors regulate the individual Chemokine AM subtypes and beyond. Overall, the findings presented here open the door to identify new therapeutic targets for inflammatory lung disease while significantly enhancing our understanding of the complexity that exists with airspace macrophages.

Other immune cell types that are not normally thought to reside in airway space were captured in our BALF analysis: distinct DC types, monocytes and IMs. IMs are mainly investigated in mouse lungs, and multiple subtypes of IMs have been reported (Bedoret et al., 2009; Chakarov et al., 2019; Gibbings et al., 2017; Schyns et al., 2019; Ural et al., 2020). Among those subtypes, the well-described *FOLR2* positive IMs also exists in the human BAL cell population, suggesting homology across species. Another IM population denoted as SPP1.IMs due to the high expression of SPP1 (Leach et al., 2020). This IM subtype seems to be consistent with previously reported *SPP1/MERTK*-expressing macrophages in idiopathic pulmonary fibrosis (Morse et al., 2019). We also report several common markers for both human IM populations, *LGMN* and *MARCKS*, observed in a previous study (Reyfman et al., 2019).

This study was originally designed to investigate the cellular and transcriptional differences in cells extracted from BALF between a cohort of CF subjects with normal lung function and minimal lung inflammation compared to healthy controls in order to investigate the impact of aberrant CFTR function. Many lines of evidence point to impaired innate immune cell function in the CF lung (Bessich et al., 2013; Chen et al., 2018; Di et al., 2006; Gardai et al.; Hazlett et al., 2020; Murphy et al., 2010; Paemka et

al., 2017; Ruytinx et al., 2018), however, we observed some minor differences between CF and HC BALF cells (Figure S5), and it is unclear whether these differences are true across all CF patients. What is striking is that even though all three CF subjects had two copies of the F508del *CFTR* mutation, when the CF lung is in a healthy-like state, the macrophage, monocyte, IM, and DC heterogeneity is highly conserved, and its niche appropriately filled. All in all, the CF subjects in our study had preserved lung function and limited inflammation by design. Future studies could include subjects with more significant lung disease to determine the impact of inflammatory milieu on CF immune cell gene expression. Additionally, comparisons of CF subjects with and without lung disease could help identify factors in airspace immune cells that predispose one to disease progression. Overall, several fundamental questions require further investigation, including the AM cluster location, function, transcription factor(s) that regulate their programming state, and whether similar macrophage superclusters exist in other organs or species.

Materials and Methods

Human Subjects

Four healthy control (HC) subjects and three subjects with CF underwent research bronchoscopy as previously described (Aridgides et al., 2019; Hogan et al., 2016). Briefly, after local anesthesia to the posterior pharynx and intravenous conscious sedation, a flexible fiberoptic bronchoscope was inserted transorally and passed through the vocal cords into the trachea. The bronchoscope was sequentially wedged into tertiary bronchi in the right upper lobe (RUL), right middle lobe (RML), and right lower lobe (RLL), saline was instilled, and BALF was collected. An aliquot of 10 mL of BALF from the RML was used for the studies outlined here.

Sample Processing

Sample processing and library construction were performed immediately after lavage. 1X phosphate-buffered saline (PBS) containing 0.5 mM EDTA and 0.1% bovine serum albumin (BSA) was added to the BALF. Samples were filtered through 100-micron cell strainers and centrifugated at 300 x g for five minutes at 4 °C, and supernatants were discarded. TotalSeq antibodies were added at concentrations recommended by the manufacturer and stained for 30 min on ice (Stoeckius et al., 2017), details in online supplement. After staining, chilled Hank's balanced salt solution (HBSS) supplemented with 0.5% BSA was used to wash the sample. Samples were again filtered through 100-micron cell strainers and spun as above. Supernatants were aspirated, and pellets were resuspended with HBSS plus 0.5% BSA at an approximate concentration of 7.5×10^5 cells/mL. Cell quality and viability were assessed with a Cellometer K2 (Nexcelom Bioscience, Lawrence, MA). All samples had viability >80%. Single cells were then processed using the Chromium Next GEM Single Cell 3' Platform (10X Genomics, Pleasanton, CA). Approximately 30,000 cells were loaded on each channel with an average recovery rate of 24,000 cells. Libraries were sequenced on NextSeq 500/550 (Illumina, San Diego, CA) with an average sequencing depth of 50,000 reads/cell.

Data Preparation

Raw sequencing reads were demultiplexed, mapped to the GRCh38 human reference genome, and gene-expression matrices were generated using CellRanger v6.1 (10X Genomics, Pleasanton, CA). The following analyses were conducted in R 4.1 (Team, 2014) and Python 3.6. Seurat package v4.0 was used for downstream data analyses (Hao et al., 2021), and figures were produced using the package ggplot2. Following a standard workflow, the gene-expression matrix was filtered to discard cells with less than 200 genes, as well as genes that were expressed in less than 3 cells. Samples then underwent quality control to remove cells with either too many or too few expressed genes (average around 2000 and 7000) and cells with too many mtRNA (average around 10%), resulting in a total of 113,213 cells. Then, "SCTransform" was applied with the "glmGamPoi" method to normalize gene expression data (Ahlmann-Eltze and Huber, 2021; Stuart et al., 2019), and a CLR transformation was applied to normalize protein data within each cell. After individual preparation, all the samples were introduced into a combined Seurat object via "FindIntegrationAnchors" and "IntegrateData" functions (Stuart et al., 2019). Then scaled values of variable genes were then subject to principal component analysis (PCA) for linear dimension reduction. A shared nearest neighbor network was created based on Euclidean distances between cells in multidimensional PC space (the first 50 PC were used) and a fixed number of neighbors per cell (30 neighbors). This was used to generate a two-dimensional Uniform Manifold Approximation and Projection (UMAP) for visualization as a harmonized atlas to dissect the cell populations in BALF of healthy and CF donors.

TotalSeq Antibodies

TotalSeq antibodies were added at concentrations recommended by the manufacturer and stained for 30 min on ice (Stoeckius et al., 2017). CD88 (B1046), CD1c (B0160), CD1a (B0402), CD64 (B0162), CD14 (B0081), CD93 (B0446), CD197 (B0148), XCR1 (B0208), CD163 (B0358), CD36 (B0407), CD206 (B0205) HLA-DR (B0159), CD169 (B0206) TotalSeq B (BioLegend, San Diego, CA). TotalSeq antibodies were added to the last four samples (HC3, HC4, CF2, CF3).

Differentially Expressed Genes

Differentially expressed genes (DEGs) were calculated with "FindAllMarkers" function of Seurat in R 4.1 to study the different expression profiles in different cell types and up-regulated genes in HC or CF cohort (Hao et al., 2021; Team, 2014). The "data" matrices of "SCT" assay were used, and the minimal log fold change was set to 0.25. Only genes that were detected in more than 25% of cells in either of the two populations were used to compute the DEGs with the Wilcoxon rank-sum test. Markers were identified as genes exhibiting significant upregulation when compared against all other clusters and defined by having a Bonferroni adjusted p-value < 0.05. The DEGs are ranked by the adjusted p--value to select the top DEGs for downstream analysis.

Cell Type Identification

To identify cell types, "FindClusters" function with the Leiden algorithm with the various resolution from 0.5 to 2.0 in the Seurat package were used for clustering (Hao et al., 2021). "FindAllMarkers" function was then applied. The top DEGs of individual clusters were examined for well-studied marker genes across literatures and the clusters were then annotated for the most likely identity. For re-clustered AMs and Mono subtypes, "FindClusters" was performed with different resolution until the right resolution is reached so that each cluster has unique gene expression pattern.

Cross-sample Cell Type Replicability

To compare cross-sample and cross-dataset transcriptomes, we performed MetaNeighbor (Crow et al., 2018) analysis between individual samples from our dataset and publicly available datasets addressing similar questions, Mould et al. and Schupp et al. (Mould et al., 2021; Schupp et al., 2020). Datasets were downloaded from UCSC Cell Browser (Kent et al., 2002) (https://cells.ucsc.edu/?ds=healthy-bal; https://cells.ucsc.edu/?ds=human-sputum) and gene expression omnibus (GEO; GSE151928; GSE145360). SummarizedExperiment objects were constructed using combined "data" matrices of SCT assays and the "variableGenes" function of

20

MetaNeighbor package was used to identify the highly variable genes as input of similarity assessment (Crow et al., 2018). The cell–cell similarities were represented by the mean area under the receiver operator characteristic curve (AUROC) scores, keeping the original cell type from every dataset. The correlation matrices were then passed to ComplexHeatmap package and clustered with "spearman" method (Gu et al., 2016). Heatmap splits were applied to highlight the focus of the study.

Gene Ontology Analysis

The gene set functional analysis was conducted with TopGO (Marini and Binder, 2019) (Fisher's exact test) to get a general overview of the more represented functional categories with the DEGs in each cluster. Background genes of gene enrichment assays were defined as all the detected genes in datasets and the biological process was selected as the target of ontology analysis. The DEGs of different cell types were used as the enrichment input. Top 5 gene ontology (GO) terms ranked by P value (Rantakari et al.) for every cell type was selected and compared with other cell types. Overlaps of those GO terms were removed from the final visualization.

Regulon Analysis

All the cell types were downsampled to maximal 1000 cells. TF activity was analyzed using SCENIC package of R 4.1 and pySCENIC of Python 3.8 for the cell types with downsampled count matrices as the input (Aibar et al., 2017; Huynh-Thu et al., 2010; Van de Sande et al., 2020). The regulons and TF activity (area under the curve) for each cell were calculated with motif collection version v9 as well as hg38__refseq r80__10kb_up_and_down_tss.mc9nr.featherdatabases from the cisTarget (https://resources.aertslab.org/cistarget/). The transcription factor regulons prediction was performed using the pySCENIC default parameters (Aibar et al., 2017; Huynh-Thu et al., 2010; Van de Sande et al., 2020). The resulting AUC scores per each cell and adjacency matrices were used for downstream quantification of regulon specificity score. Top 5 transcription factor ranked by specificity score for every cell type was selected and compared with other cell types. Overlaps of those transcription factors were removed from the final visualization.

21

Pseudotime Analysis

The package Monocle3 was performed to analyze single-cell trajectories (Cao et al., 2019; Qiu et al., 2017a; Qiu et al., 2017b; Trapnell et al., 2014). Pseudotime values were determined for all myeloid cells except DCs, which have a different ontogeny. All the cell types were downsampled to maximal 1000 cells. Then Monocle3 was used to learn the sequence of gene expression changes each cell must go through as part of a dynamic biological process and then arrange each cell in the trajectory (Cao et al., 2019; Qiu et al., 2017a; Qiu et al., 2017b; Trapnell et al., 2014). Two nodes of monocytes population were set as the start allowing the visualization of the differentiation process of monocytes to macrophages. To investigate the dynamic gene expression between monocytes and macrophages, DEGs over the pseudotime were calculated by the "graph_test" function. By ranking the P value, top 100 pseudotime DEGs were identified. By further downsampling the cell types to maximal 100 cells, the expression levels of the top 100 pseudotime DEGs were then visualized and the top 10 were labeled.

Cross-sample Reference Mapping

Publicly available AM/monocyte datasets were downloaded from UCSC Cell Browser (Kent et al., 2002) (https://cells.ucsc.edu/?ds=covid19-balf; https://cells.ucsc.edu/?ds=healthy-bal; https://cells.ucsc.edu/?ds=pan-immune+myeloid; https://cells.ucsc.edu/?ds=teichmann-asthma+lung-atlas+lung-atlas-others; https://cells.ucsc.edu/?ds=human-sputum), gene expression omnibus (GEO; GSE151928; GSE145360) and GitHub (https://github.com/theislab/2020_Mayr). To investigate the newly identified AM/monocyte subtypes in publicly available datasets, "FindTransferAnchors" and "TransferData" functions in the Seurat package were used to infer the cell types in the aforementioned query datasets (PMID: 31178118). The query cells were classified into different AM/monocyte subtypes based on their expression profile similarity with our dataset. Prediction scores were also returned to evaluate the reference prediction. Moreover, "MapQuery" function was applied to

22

project the query datasets onto the our/reference UMAP structure for a more straightforward visualization (PMID: 31178118).

Flow Cytometry

BAL cells were stained with the following monoclonal Abs: anti human CD45 BUV805 clone Hl30, anti-human CD14 V500, clone M□P9, anti-human HLA-DR APC-Cy7, clone L243, anti-human CD206 PerCp-Cy5.5, clone 15-2, anti-human CD36 APC, clone AC106, anti-human CD169 FITC and anti-human CD93 PE-Cy7. The viability dye DAPI (#D9542, Sigma, Milwaukee, WI) was added immediately before each sample acquisition on a BD Symphony A3 analyzer (BD Biosciences, San Diego, CA). Data were analyzed using FlowJo (Tree Star, Ashland, OR). Antigen-specific antibodies and isotype controls were obtained from BioLegend (San Diego, CA), eBioscience (Ebioscience, AUT) and BD Biosciences (San Diego, CA).

Data Availability

The sequencing raw data and processed data used in this paper are available in Gene Expression Omnibus, accession code GSE193782. The processed data is also available for online visualization at https://ams-supercluster.cells.ucsc.edu (visualization for Figure 2-4 can be viewed with the "AMs Subset UMAP" layout).

Study Approval

All subjects provided written informed consent, and this study was approved by the Institutional Review Board of Dartmouth-Hitchcock Medical Center (protocol #22781).

Acknowledgement

Author Contributions

XL, AA, and CJ prepared the manuscript; XL, FK and CJ performed the experiments; DA and AA enrolled all subjects, collected samples, and curated clinical data; DM processed clinical samples; XL performed the scRNA-seq and other bioinformatics analyses; all authors provided intellectual input, critical feedback, discussed results, and designed experiments.

Conflict of interest

All authors declare that there are no competing interests.

Figure Legends

Figure 1 - 12 BAL cell types identified in healthy control (HC) and CF samples.

(A) ScRNA-seq data shows integrated UMAP with HC and CF, along with individual sample distribution. (B) UMAP demonstrating the twelve major cell types: alveolar macrophages (AMs), myeloid and lymphoid cycling cells (Cyc.Mye and Cyc.Lym), FOLR2 and SPP1 interstitial macrophages, monocytes (Mono), Dendritic cell 1 (DC1), Dendritic cell 2 (DC2), Migratory DC (Mig.DC), plasmacytoid DCs (pDCs), lymphocytes (Lym), and epithelial cells (Epi). (C) Percentage bar graph outlines the distribution of individual cell types for each subject. (D) Dot plot shows the expression of curated genes in each individual cell type. (E) Correlation heat map of cell types against all samples in in-house dataset (Li). The lower quadrant highlights the corresponding cell types. (F) Correlation heat map of HC cell types in in-house and Mould dataset. The lower quadrant highlights the corresponding cell types. (G) Correlation heat map of HC and CF cell types in in-house and Schupp dataset. The lower quadrant highlights the corresponding cell types.

Figure 2 - Unbiased analysis of scRNA-seq data illustrates four AM superclusters based on *IFI27* and *APOC2* expression.

(A) UMAP of re-clustered AMs shows sample source distribution. (B) UMAP shows distinct AM clusters: AM Supercluster 1 (AMs.S1), AMs.S2, AMs.S3 AMs.S4, Metallothionein (MT.AMs), Chemokine (CK.AMs), Interferon (IFN.AMs) and Cycling AM (Cyc.AMs) clusters. (C) Percentage bar graph outlines the distribution of individual AM clusters for each subject. (D) Dot plot shows the expression of the top 6 DEGs in individual AM clusters, duplicates removed. For example, AMs.S4's top 6 DEGs are already present in the first three AM superclusters; therefore, AMs.S4 is seen on the Y axis but not the X axis. (E-F) Feature plot shows the expression of *IFI27* and *APOC2*, which define the four AM superclusters.

Figure 3 - AM superclusters contain 4 shared subclusters and a small-supercluster.

26

(A) UMAP illustrates the shared AM subclusters. (B) Percentage bar graph outlines the distribution of individual AM subclusters for each subject. (C) Featured plots show four representative genes for each shared AM subcluster. (D) Dot plot illustrates the expression of the top 6 DEGs in each individual AM subcluster, duplicates removed.

Figure 4 - Chemokine AM subclusters express unique combinations of chemokines.

(A) UMAP illustrates distinct chemokine-expressing AM subclusters: CK.AMs.c1, CK.AMs.c2, CK.AMs.c3, and CK.AMs.c4. (B) Percentage bar graph outlines the distribution of individual chemokine-expressing AM subclusters for each subject. (C) Featured plots show four representative gene expression for individual chemokine-expressing AM subcluster. (D) Dot plot illustrates the expression of the top 6 DEGs (duplicates removed) and curated chemokine genes in individual chemokine-expressing AM subcluster. (E) Heat map shows the top 5 gene ontology terms (biological process) for each AM subtype. Top 1 term is labeled. (F) Heat map shows the top 5 regulons for each AM subtype. Top 1 regulon is labeled.

Figure 5 - BAL monocytes and IMs contain heterogeneous clusters.

(A) UMAP of monocytes shows sample source distribution. (B) UMAP shows 8 distinct monocyte clusters: conventional monocyte cluster (Conv.Mono) and intermediate monocyte cluster (Int.Mono) and others. (C) Percentage bar graph outlines the distribution of individual monocyte clusters for each subject. (D) Dot plot shows the expression of the top 6 DEGs in individual monocytes clusters, duplicates removed. (E) Featured plots show the expression of different DEGs of 2 IM clusters.

Figure 6 - Pseudotime analysis of monocytes and macrophages in BALF.

(A) Pseudotime plot shows the evolvement of monocytes to macrophages with the root as white dots and the principal points as solid black dots. (B) UMAP shows distinct monocyte and macrophage clusters included for Pseudotime analysis. (C) Heat map shows the top 100 pseudotime DEGs across the monocyte-to-macrophage branch. Top 10 DEGs are labelled. (D) Flow cytometry plot illustrates the high expression of CD43

and CD169 on AMs and CD93, CD36 and CD14 on monocytes. (E) Featured plots show representative gene expression and TotalSeq detection in AMs and monocytes.

Figure 7 - Quadrant distribution of AM subtypes for the bivariate expression of AM subtypes for the of *IFI27* and *APOC2* and summary of proposed AM subtype classification.

(A) AM subtype model supported by direct data. (B) Density plot showing the expression of *IFI27* and *APOC2* for Chemokine, Metallothionein, Interferon, and Cycling AMs. (C) Alternative AM subtype model proposed in the results section and discussion.

Figure 8 - AM superclusters are present in other publicly available datasets.

(A) AM summary table of publicly available datasets that were compared to our dataset. (B) Featured plots show the expression of 7 selected AM cluster markers (*IFI27, APOC2, CCL4, CXCL10, CXCL3, MT1G, ISG15*) in represented datasets. Only AMs were selected and re-clustered for UMAP visualization; (C) Schematic representation where reference mapping with Seurat package allows for the projection of publicly available datasets onto our AM supercluster-annotated dataset from Figures 2; (D) The reference mapping projection of the represented dataset; (E) Percentage bar graph outlines the distribution of individual AM clusters for each health status in different datasets.

Figure S1 - Complete gene ontology and regulon analysis heat map for AM subtypes.

(A) Heat map showing the top 5 gene ontology terms (biological process, duplicates removed) for each AM subtype. (B) Heat map showing the top 5 regulons (duplicates removed) for each AM subtype.

Figure S2 - Complete gene ontology and regulon analysis heat map for monocyte subtypes.

(A) Heat map showing the top 5 gene ontology terms (biological process, duplicates removed) for each monocyte subtype. (B) Heat map showing the top 5 regulons (duplicates removed) for each monocyte subtype.

Figure S3 - Comparison of selected RNA expression and protein detection for AMs and monocytes.

(A) Radar plots illustrate the higher expression/detection levels of *SPN* mRNA and CD169 protein in AMs compared to monocytes, but not for *SIGLEC1* mRNA, the mRNA coding for CD169 protein. (B) Radar plots illustrate the higher mRNA expression and protein detection levels of CD14, CD36, CD93 in monocytes compared to AMs.

Figure S4 - Four AM superclusters also exist in individual subject.

(A) AMs from the integrated HC or CF samples cluster unbiasedly into four superclusters confirmed by the expression pattern of *IFI27* and *APOC2*. (B) AMs from the individual HC and CF samples cluster unbiasedly into four superclusters confirmed by the expression pattern of *IFI27* and *APOC2*.

Figure S5 - Gene expression analysis between HC and CF samples.

(A) Volcano plot demonstrates the DEGs in either HC or CF samples. Genes with a P value smaller than 0.05 and log2 fold change larger than 25% are shown. Top 15 DEGs are labeled.   (B) Dot Chart shows the top 3 DEGs in either HC or CF samples for the twelve major cell types.   (C) Dot Chart shows the top 3 DEGs in either HC or CF

samples for individual AM subtypes. (D) Dot Chart shows the top 3 DEGs in either HC or CF samples for individual monocyte subtypes.

Figure S6 - AM subtypes and monocyte subtypes also exist in other publicly available datasets.

(A) The reference mapping projection of represented datasets to the AMs in our dataset. (B) The reference mapping projection of represented datasets to the monocytes in our dataset. (C) Monocyte summary table with details of compared publicly available datasets present in this figure. (D) Percentage bar graph outlines the distribution of individual monocyte clusters for each health status in different datasets. (E) Heat map showing the prediction scores of AMs in Liao's study mapping to the AM subtypes in our dataset. (F) Heat map showing the prediction scores of monocytes in Mould's study mapping to the monocyte subtypes in our dataset.

Table 1 - Demographic characteristics for recruited subjects.

Table S1 - Curated markers for major cell types.

Table S2 - DEGs of major cell types demonstrated in Figure 1.

Table S3 - DEGs of AM supercluster cell types demonstrated in Figure 2.

Table S4 - DEGs of AM cell types demonstrated in Figure 3 and Figure 4.

Table S5 - DEGs of monocyte and IM cell types demonstrated in Figure 5.

Table S6 - DEGs of all cells between CF and HC demonstrated in Figure S5A.

Table S7 - DEGs of major cell types between CF and HC demonstrated in Figure S5B.

Table S8 - DEGs of AM cell types between CF and HC demonstrated in Figure S5C.

Table S9 - DEGs of monocyte cell types between CF and HC demonstrated in Figure S5D.

31

References

Aegerter, H., J. Kulikauskaite, S. Crotta, H. Patel, G. Kelly, E.M. Hessel, M. Mack, S. Beinke, and A. Wack. 2020. Influenza-induced monocyte-derived alveolar macrophages confer prolonged antibacterial protection. *Nat Immunol* 21:145-157.

Ahlmann-Eltze, C., and W. Huber. 2021. glmGamPoi: fitting Gamma-Poisson generalized linear models on single cell count data. *Bioinformatics* 36:5701-5702.

Aibar, S., C.B. Gonzalez-Blas, T. Moerman, V.A. Huynh-Thu, H. Imrichova, G. Hulselmans, F. Rambow, J.C. Marine, P. Geurts, J. Aerts, J. van den Oord, Z.K. Atak, J. Wouters, and S. Aerts. 2017. SCENIC: single-cell regulatory network inference and clustering. *Nat Methods* 14:1083-1086.

Arafa, E.I., A.T. Shenoy, K.A. Barker, N.S. Etesami, I.M. Martin, C. Lyon De Ana, E. Na, C.V. Odom, W.N. Goltry, F.T. Korkmaz, A.K. Wooten, A.C. Belkina, A. Guillon, E.C. Forsberg, M.R. Jones, L.J. Quinton, and J.P. Mizgerd. 2022. Recruitment and training of alveolar macrophages after pneumococcal pneumonia. *JCI Insight* 7:

Aridgides, D.S., D.L. Mellinger, D.A. Armstrong, H.F. Hazlett, J.A. Dessaint, T.H. Hampton, G.T. Atkins, J.L. Carroll, and A. Ashare. 2019. Functional and metabolic impairment in cigarette smoke-exposed macrophages is tied to oxidative stress. *Sci Rep* 9:9624.

Bedoret, D., H. Wallemacq, T. Marichal, C. Desmet, F. Quesada Calvo, E. Henry, R. Closset, B. Dewals, C. Thielen, P. Gustin, L. de Leval, N. Van Rooijen, A. Le Moine, A. Vanderplasschen, D. Cataldo, P.V. Drion, M. Moser, P. Lekeux, and F. Bureau. 2009. Lung interstitial macrophages alter dendritic cell functions to prevent airway allergy in mice. *J Clin Invest* 119:3723-3738.

Bessich, J.L., A.B. Nymon, L.A. Moulton, D. Dorman, and A. Ashare. 2013. Low levels of insulin-like growth factor-1 contribute to alveolar macrophage dysfunction in cystic fibrosis. *J Immunol* 191:378-385.

Bharat, A., S.M. Bhorade, L. Morales-Nebreda, A.C. McQuattie-Pimentel, S. Soberanes, K. Ridge, M.M. DeCamp, K.K. Mestan, H. Perlman, G.R. Budinger, and A.V. Misharin. 2016. Flow Cytometry Reveals Similarities Between Lung Macrophages in Humans and Mice. *Am J Respir Cell Mol Biol* 54:147-149.

Bonfield, T.L. 2015. Macrophage Dysfunction in Cystic Fibrosis: A Therapeutic Target to Enhance Self-Immunity. *Am J Respir Crit Care Med* 192:1406-1407.

Brown, C.C., H. Gudjonson, Y. Pritykin, D. Deep, V.P. Lavallee, A. Mendoza, R. Fromme, L. Mazutis, C. Ariyan, C. Leslie, D. Pe'er, and A.Y. Rudensky. 2019. Transcriptional Basis of Mouse and Human Dendritic Cell Heterogeneity. *Cell* 179:846-863 e824.

Bruscia, E.M., and T.L. Bonfield. 2016. Cystic Fibrosis Lung Immunity: The Role of the Macrophage. *J Innate Immun* 8:550-563.

Cai, Y., Y. Dai, Y. Wang, Q. Yang, J. Guo, C. Wei, W. Chen, H. Huang, J. Zhu, C. Zhang, W. Zheng, Z. Wen, H. Liu, M. Zhang, S. Xing, Q. Jin, C.G. Feng, and X. Chen. 2020. Single-cell transcriptomics of blood reveals a natural killer cell subset depletion in tuberculosis. *EBioMedicine* 53:102686.

Cao, J., M. Spielmann, X. Qiu, X. Huang, D.M. Ibrahim, A.J. Hill, F. Zhang, S. Mundlos, L. Christiansen, F.J. Steemers, C. Trapnell, and J. Shendure. 2019. The single-cell transcriptional landscape of mammalian organogenesis. *Nature* 566:496-502.

Chakarov, S., H.Y. Lim, L. Tan, S.Y. Lim, P. See, J. Lum, X.M. Zhang, S. Foo, S. Nakamizo, K. Duan, W.T. Kong, R. Gentek, A. Balachander, D. Carbajo, C. Bleriot, B. Malleret, J.K.C. Tam, S. Baig, M. Shabeer, S.E.S. Toh, A. Schlitzer, A. Larbi, T. Marichal, B. Malissen,

J. Chen, M. Poidinger, K. Kabashima, M. Bajenoff, L.G. Ng, V. Angeli, and F. Ginhoux. 2019. Two distinct interstitial macrophage populations coexist across tissues in specific subtissular niches. *Science* 363:

Chen, Y., D.A. Armstrong, L.A. Salas, H.F. Hazlett, A.B. Nymon, J.A. Dessaint, D.S. Aridgides, D.L. Mellinger, X. Liu, B.C. Christensen, and A. Ashare. 2018. Genome-wide DNA methylation profiling shows a distinct epigenetic signature associated with lung macrophages in cystic fibrosis. *Clin Epigenetics* 10:152.

Collin, M., and V. Bigley. 2018. Human dendritic cell subsets: an update. *Immunology* 154:3-20.

Collin, M., N. McGovern, and M. Haniffa. 2013. Human dendritic cell subsets. *Immunology* 140:22-30.

Crow, M., A. Paul, S. Ballouz, Z.J. Huang, and J. Gillis. 2018. Characterizing the replicability of cell types defined by single cell RNA-sequencing data using MetaNeighbor. *Nat Commun* 9:884.

Deng, R., C. Zuo, Y. Li, B. Xue, Z. Xun, Y. Guo, X. Wang, Y. Xu, R. Tian, S. Chen, Q. Liu, J. Chen, J. Wang, X. Huang, H. Li, M. Guo, X. Wang, M. Yang, Z. Wu, J. Wang, J. Ma, J. Hu, G. Li, S. Tang, Z. Tu, H. Ji, and H. Zhu. 2020. The innate immune effector ISG12a promotes cancer immunity by suppressing the canonical Wnt/beta-catenin signaling pathway. *Cell Mol Immunol* 17:1163-1179.

Deriy, L.V., E.A. Gomez, G. Zhang, D.W. Beacham, J.A. Hopson, A.J. Gallan, P.D. Shevchenko, V.P. Bindokas, and D.J. Nelson. 2009. Disease-causing mutations in the cystic fibrosis transmembrane conductance regulator determine the functional responses of alveolar macrophages. *J Biol Chem* 284:35926-35938.

Desch, A.N., S.L. Gibbings, R. Goyal, R. Kolde, J. Bednarek, T. Bruno, J.E. Slansky, J. Jacobelli, R. Mason, Y. Ito, E. Messier, G.J. Randolph, M. Prabagar, S.M. Atif, E. Segura, R.J. Xavier, D.L. Bratton, W.J. Janssen, P.M. Henson, and C.V. Jakubzick. 2015. Flow Cytometric Analysis of Mononuclear Phagocytes in Non-diseased Human Lung and Lung-draining Lymph Nodes. *Am J Respir Crit Care Med*

Di, A., M.E. Brown, L.V. Deriy, C. Li, F.L. Szeto, Y. Chen, P. Huang, J. Tong, A.P. Naren, V. Bindokas, H.C. Palfrey, and D.J. Nelson. 2006. CFTR regulates phagosome acidification in macrophages and alters bactericidal activity. *Nat Cell Biol* 8:933-944.

Dominguez Conde, C., C. Xu, L.B. Jarvis, D.B. Rainbow, S.B. Wells, T. Gomes, S.K. Howlett, O. Suchanek, K. Polanski, H.W. King, L. Mamanova, N. Huang, P.A. Szabo, L. Richardson, L. Bolt, E.S. Fasouli, K.T. Mahbubani, M. Prete, L. Tuck, N. Richoz, Z.K. Tuong, L. Campos, H.S. Mousa, E.J. Needham, S. Pritchard, T. Li, R. Elmentaite, J. Park, E. Rahmani, D. Chen, D.K. Menon, O.A. Bayraktar, L.K. James, K.B. Meyer, N. Yosef, M.R. Clatworthy, P.A. Sims, D.L. Farber, K. Saeb-Parsy, J.L. Jones, and S.A. Teichmann. 2022. Cross-tissue immune cell analysis reveals tissue-specific features in humans. *Science* 376:eabl5197.

Fricker, M., and P.G. Gibson. 2017. Macrophage dysfunction in the pathogenesis and treatment of asthma. *Eur Respir J* 50:

Fujii, W., T.S. Kapellos, K. Bassler, K. Handler, L. Holsten, R. Knoll, S. Warnat-Herresthal, M. Oestreich, E.R. Hinkley, J. Hasenauer, C. Pizarro, C. Thiele, A.C. Aschenbrenner, T. Ulas, D. Skowasch, and J.L. Schultze. 2021. Alveolar macrophage transcriptomic profiling in COPD shows major lipid metabolism changes. *ERJ Open Res* 7:

Gardai, S.J., K.A. McPhillips, S.C. Frasch, W.J. Janssen, A. Starefeldt, J.E. Murphy-Ullrich, D.L. Bratton, P.A. Oldenborg, M. Michalak, and P.M. Henson. 2005. Cell-surface
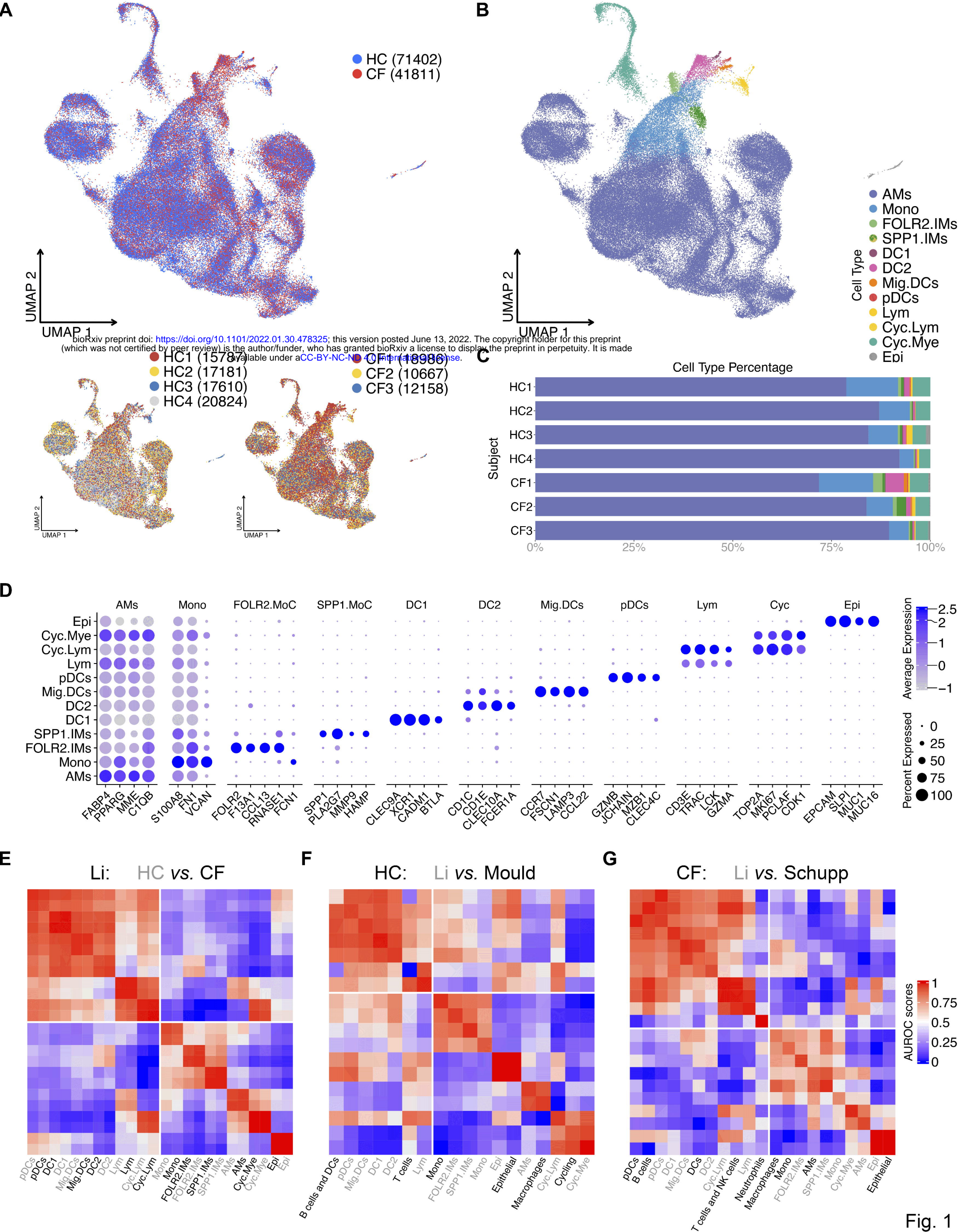
calreticulin initiates clearance of viable or apoptotic cells through trans-activation of LRP on the phagocyte. *Cell* 123:321-334.

Gautier, E.L., T. Shay, J. Miller, M. Greter, C. Jakubzick, S. Ivanov, J. Helft, A. Chow, K.G. Elpek, S. Gordonov, A.R. Mazloom, A. Ma'ayan, W.J. Chua, T.H. Hansen, S.J. Turley, M. Merad, G.J. Randolph, and C. Immunological Genome. 2012. Gene-expression profiles and transcriptional regulatory pathways that underlie the identity and diversity of mouse tissue macrophages. *Nat Immunol* 13:1118-1128.

Gibbings, S.L., R. Goyal, A.N. Desch, S.M. Leach, M. Prabagar, S.M. Atif, D.L. Bratton, W. Janssen, and C.V. Jakubzick. 2015. Transcriptome analysis highlights the conserved difference between embryonic and postnatal-derived alveolar macrophages. *Blood* 126:1357-1366.

Gibbings, S.L., and C.V. Jakubzick. 2018. A Consistent Method to Identify and Isolate Mononuclear Phagocytes from Human Lung and Lymph Nodes. *Methods Mol Biol* 1799:381-395.

Gibbings, S.L., S.M. Thomas, S.M. Atif, A.L. McCubbrey, A.N. Desch, T. Danhorn, S.M. Leach, D.L. Bratton, P.M. Henson, W.J. Janssen, and C.V. Jakubzick. 2017. Three Unique Interstitial Macrophages in the Murine Lung at Steady State. *Am J Respir Cell Mol Biol* 57:66-76.

Gu, Z., R. Eils, and M. Schlesner. 2016. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* 32:2847-2849.

Hao, Y., S. Hao, E. Andersen-Nissen, W.M. Mauck, 3rd, S. Zheng, A. Butler, M.J. Lee, A.J. Wilk, C. Darby, M. Zager, P. Hoffman, M. Stoeckius, E. Papalexi, E.P. Mimitou, J. Jain, A. Srivastava, T. Stuart, L.M. Fleming, B. Yeung, A.J. Rogers, J.M. McElrath, C.A. Blish, R. Gottardo, P. Smibert, and R. Satija. 2021. Integrated analysis of multimodal single-cell data. *Cell* 184:3573-3587 e3529.

Hazlett, H.F., T.H. Hampton, D.S. Aridgides, D.A. Armstrong, J.A. Dessaint, D.L. Mellinger, A.B. Nymon, and A. Ashare. 2020. Altered iron metabolism in cystic fibrosis macrophages: the impact of CFTR modulators and implications for Pseudomonas aeruginosa survival. *Sci Rep* 10:10935.

Hetzel, M., M. Ackermann, and N. Lachmann. 2021. Beyond "Big Eaters": The Versatile Role of Alveolar Macrophages in Health and Disease. *Int J Mol Sci* 22:

Hey, J., M. Paulsen, R. Toth, D. Weichenhan, S. Butz, J. Schatterny, R. Liebers, P. Lutsik, C. Plass, and M.A. Mall. 2021. Epigenetic reprogramming of airway macrophages promotes polarization and inflammation in muco-obstructive lung disease. *Nat Commun* 12:6520.

Hogan, D.A., S.D. Willger, E.L. Dolben, T.H. Hampton, B.A. Stanton, H.G. Morrison, M.L. Sogin, J. Czum, and A. Ashare. 2016. Analysis of Lung Microbiota in Bronchoalveolar Lavage, Protected Brush and Sputum Samples from Subjects with Mild-To-Moderate Cystic Fibrosis Lung Disease. *PLoS One* 11:e0149998.

Huynh-Thu, V.A., A. Irrthum, L. Wehenkel, and P. Geurts. 2010. Inferring regulatory networks from expression data using tree-based methods. *PLoS One* 5:

Jakubzick, C., E.L. Gautier, S.L. Gibbings, D.K. Sojka, A. Schlitzer, T.E. Johnson, S. Ivanov, Q. Duan, S. Bala, T. Condon, N. van Rooijen, J.R. Grainger, Y. Belkaid, A. Ma'ayan, D.W. Riches, W.M. Yokoyama, F. Ginhoux, P.M. Henson, and G.J. Randolph. 2013. Minimal differentiation of classical monocytes as they survey steady-state tissues and transport antigen to lymph nodes. *Immunity* 39:599-610.
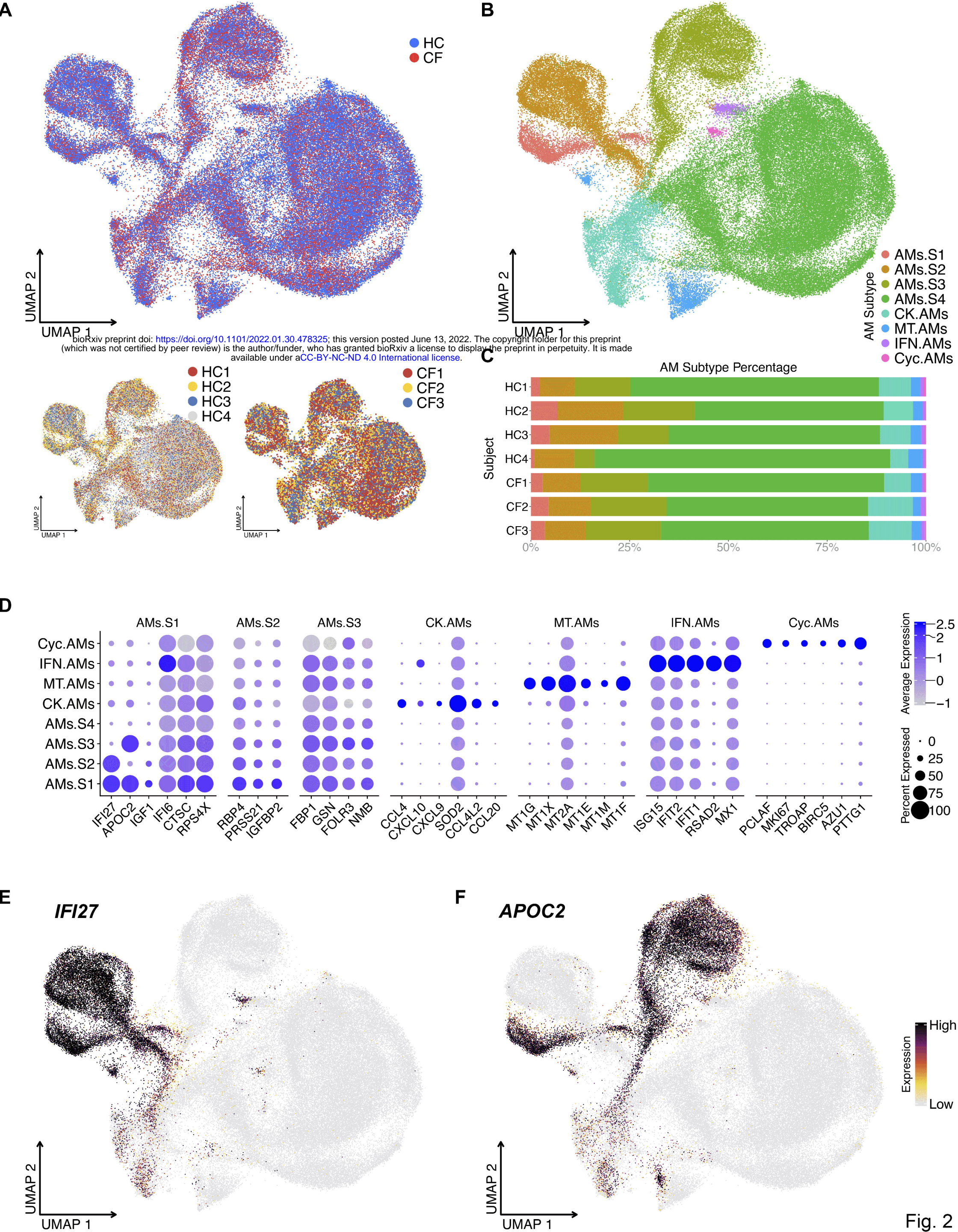
Janssen, W.J., L. Barthel, A. Muldrow, R.E. Oberley-Deegan, M.T. Kearns, C. Jakubzick, and P.M. Henson. 2011. Fas Determines Differential Fates of Resident and Recruited Macrophages During Resolution of Acute Lung Injury. *Am J Respir Crit Care Med*

Jin, X., L. Meng, Z. Yin, H. Yu, L. Zhang, W. Liang, S. Wang, G. Liu, and L. Zhang. 2019. Characterization of dendritic cell subtypes in human cord blood by single-cell sequencing. *Biophysics Reports* 5:199-208.

Kent, W.J., C.W. Sugnet, T.S. Furey, K.M. Roskin, T.H. Pringle, A.M. Zahler, and D. Haussler. 2002. The human genome browser at UCSC. *Genome Res* 12:996-1006.

Leach, S.M., S.L. Gibbings, A.D. Tewari, S.M. Atif, B. Vestal, T. Danhorn, W.J. Janssen, T.D. Wager, and C.V. Jakubzick. 2020. Human and Mouse Transcriptome Profiling Identifies Cross-Species Homology in Pulmonary and Lymph Node Mononuclear Phagocytes. *Cell Rep* 33:108337.

Li, S., Y. Xie, W. Zhang, J. Gao, M. Wang, G. Zheng, X. Yin, H. Xia, and X. Tao. 2015. Interferon alpha-inducible protein 27 promotes epithelial-mesenchymal transition and induces ovarian tumorigenicity and stemness. *J Surg Res* 193:255-264.

Li, X., K. Rawat, and C.V. Jakubzick. 2021. Targeting resident macrophages in cancer. *Nat Immunol* 22:1078-1079.

Liao, M., Y. Liu, J. Yuan, Y. Wen, G. Xu, J. Zhao, L. Cheng, J. Li, X. Wang, F. Wang, L. Liu, I. Amit, S. Zhang, and Z. Zhang. 2020. Single-cell landscape of bronchoalveolar immune cells in patients with COVID-19. *Nat Med* 26:842-844.

María Casanova-Acebes, E.D., Andrew Leader, Jessica LeBerichel, Jovan Nikolic, Blanca M. Morales, Markus Brown, Christie Chang, Leanna Troncoso, Steven T. Chen, Ana Sastre-Perona, Matthew D. Park, Alexandra Tabachnikova, Maxime Dhainaut, Pauline Hamon, Barbara Maier, Catherine M. Sawai, Esperanza Agulló-Pascual, Markus Schober, Brian Brown, Boris Reizis, Thomas Marron, Ephraim Kenigsberg, Christine Moussion, Philippe Benaroch, Julio A. Aguirre-Ghiso, and Miriam Merad. 2021. Tissue-resident macrophages provide a pro-tumorigenic niche to early NSCLC cells. *Nature*

Marini, F., and H. Binder. 2019. pcaExplorer: an R/Bioconductor package for interacting with RNA-seq principal components. *BMC Bioinformatics* 20:331.

Mayr, C.H., L.M. Simon, G. Leuschner, M. Ansari, J. Schniering, P.E. Geyer, I. Angelidis, M. Strunz, P. Singh, N. Kneidinger, F. Reichenberger, E. Silbernagel, S. Bohm, H. Adler, M. Lindner, B. Maurer, A. Hilgendorff, A. Prasse, J. Behr, M. Mann, O. Eickelberg, F.J. Theis, and H.B. Schiller. 2021. Integrative analysis of cell state changes in lung fibrosis with peripheral protein biomarkers. *EMBO Mol Med* 13:e12871.

McFarland, H.I., and A.S. Rosenberg. 2009. Skin allograft rejection. *Current protocols in immunology / edited by John E. Coligan ... [et al.]* Chapter 4:Unit 4 4.

Minutti, C.M., R.V. Modak, F. Macdonald, F. Li, D.J. Smyth, D.A. Dorward, N. Blair, C. Husovsky, A. Muir, E. Giampazolias, R. Dobie, R.M. Maizels, T.J. Kendall, D.W. Griggs, M. Kopf, N.C. Henderson, and D.M. Zaiss. 2019. A Macrophage-Pericyte Axis Directs Tissue Restoration via Amphiregulin-Induced Transforming Growth Factor Beta Activation. *Immunity* 50:645-654 e646.

Misharin, A.V., L. Morales-Nebreda, P.A. Reyfman, C.M. Cuda, J.M. Walter, A.C. McQuattie-Pimentel, C.I. Chen, K.R. Anekalla, N. Joshi, K.J.N. Williams, H. Abdala-Valencia, T.J. Yacoub, M. Chi, S. Chiu, F.J. Gonzalez-Gonzalez, K. Gates, A.P. Lam, T.T. Nicholson, P.J. Homan, S. Soberanes, S. Dominguez, V.K. Morgan, R. Saber, A. Shaffer, M. Hinchcliff, S.A. Marshall, A. Bharat, S. Berdnikovs, S.M. Bhorade, E.T. Bartom, R.I.
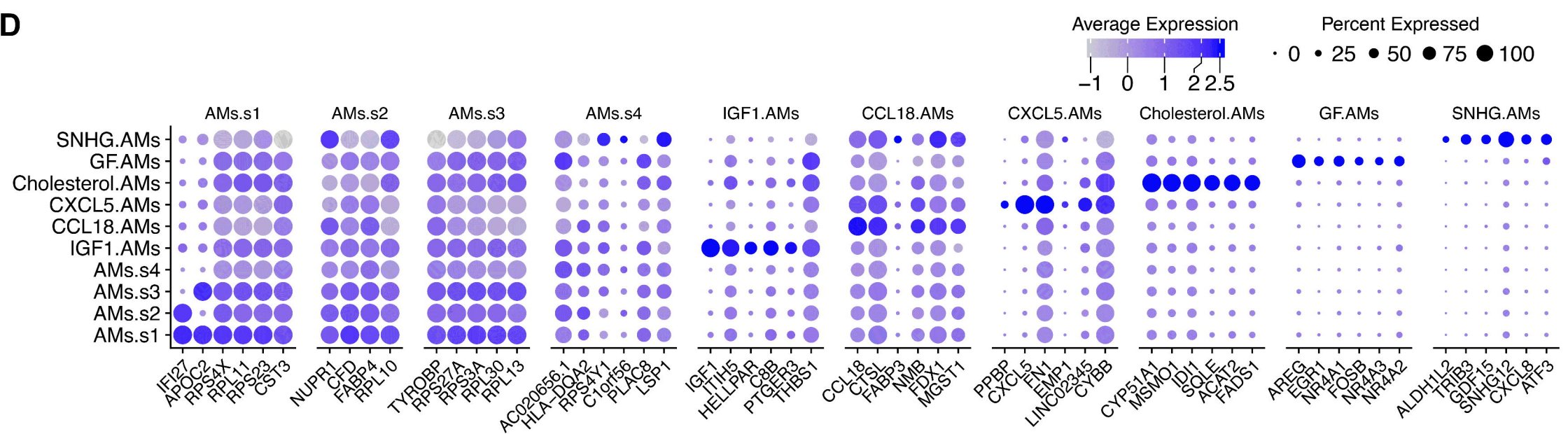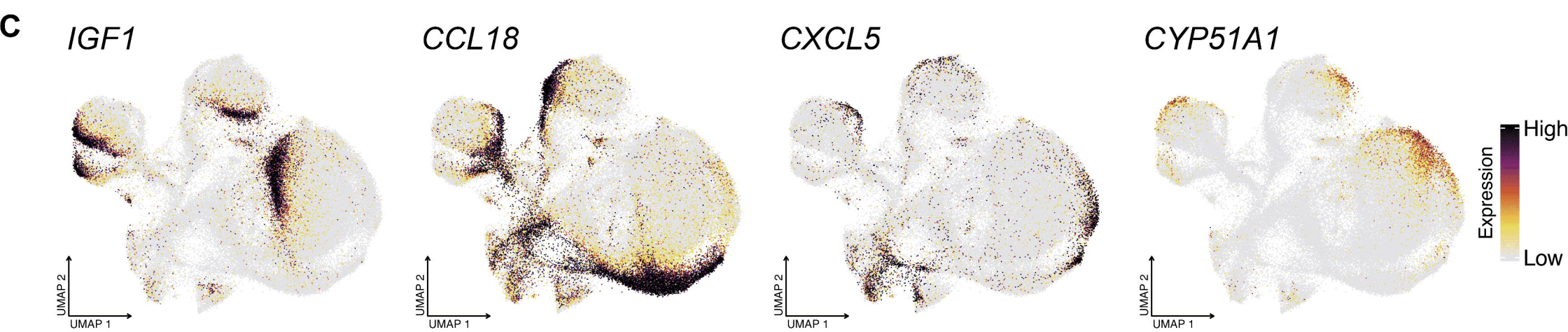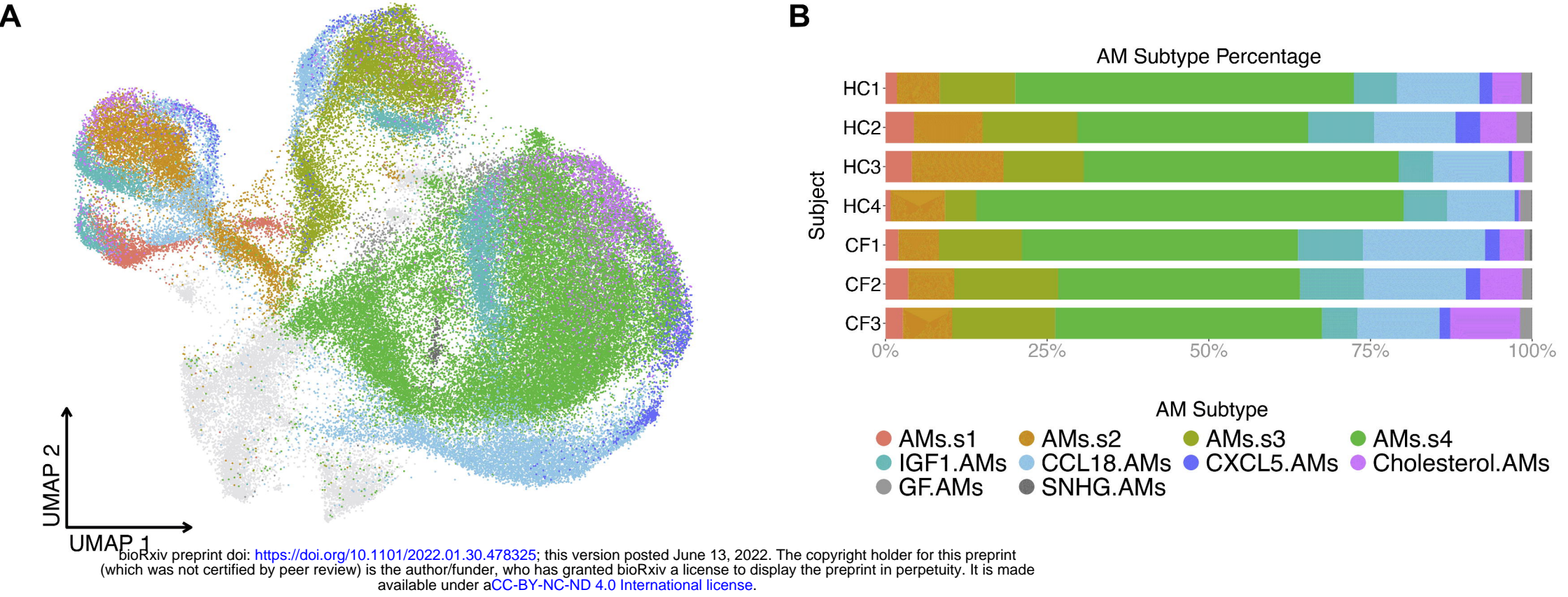
Morimoto, W.E. Balch, J.I. Sznajder, N.S. Chandel, G.M. Mutlu, M. Jain, C.J. Gottardi, B.D. Singer, K.M. Ridge, N. Bagheri, A. Shilatifard, G.R.S. Budinger, and H. Perlman. 2017. Monocyte-derived alveolar macrophages drive lung fibrosis and persist in the lung over the life span. *J Exp Med* 214:2387-2404.

Morse, C., T. Tabib, J. Sembrat, K.L. Buschur, H.T. Bittar, E. Valenzi, Y. Jiang, D.J. Kass, K. Gibson, W. Chen, A. Mora, P.V. Benos, M. Rojas, and R. Lafyatis. 2019. Proliferating SPP1/MERTK-expressing macrophages in idiopathic pulmonary fibrosis. *Eur Respir J* 54:

Mould, K.J., L. Barthel, M.P. Mohning, S.M. Thomas, A.L. McCubbrey, T. Danhorn, S.M. Leach, T.E. Fingerlin, B.P. O'Connor, J.A. Reisz, A. D'Alessandro, D.L. Bratton, C.V. Jakubzick, and W.J. Janssen. 2017. Cell Origin Dictates Programming of Resident versus Recruited Macrophages during Acute Lung Injury. *Am J Respir Cell Mol Biol* 57:294-306.

Mould, K.J., N.D. Jackson, P.M. Henson, M. Seibold, and W.J. Janssen. 2019. Single cell RNA sequencing identifies unique inflammatory airspace macrophage subsets. *JCI Insight* 4:

Mould, K.J., C.M. Moore, S.A. McManus, A.L. McCubbrey, J.D. McClendon, C.L. Griesmer, P.M. Henson, and W.J. Janssen. 2021. Airspace Macrophages and Monocytes Exist in Transcriptionally Distinct Subsets in Healthy Adults. *Am J Respir Crit Care Med* 203:946-956.

Murphy, B.S., H.M. Bush, V. Sundareshan, C. Davis, J. Hagadone, T.J. Cory, H. Hoy, D. Hayes, Jr., M.I. Anstead, and D.J. Feola. 2010. Characterization of macrophage activation states in patients with cystic fibrosis. *J Cyst Fibros* 9:314-322.

O'Beirne, S.L., S.A. Kikkers, C. Oromendia, J. Salit, M.R. Rostmai, K.V. Ballman, R.J. Kaner, R.G. Crystal, and S.M. Cloonan. 2020. Alveolar Macrophage Immunometabolism and Lung Function Impairment in Smoking and Chronic Obstructive Pulmonary Disease. *Am J Respir Crit Care Med* 201:735-739.

Paemka, L., B.N. McCullagh, M.H. Abou Alaiwa, D.A. Stoltz, Q. Dong, C.O. Randak, R.D. Gray, and P.B. McCray, Jr. 2017. Monocyte derived macrophages from CF pigs exhibit increased inflammatory responses at birth. *J Cyst Fibros* 16:471-474.

Qiu, X., A. Hill, J. Packer, D. Lin, Y.A. Ma, and C. Trapnell. 2017a. Single-cell mRNA quantification and differential analysis with Census. *Nat Methods* 14:309-315.

Qiu, X., Q. Mao, Y. Tang, L. Wang, R. Chawla, H.A. Pliner, and C. Trapnell. 2017b. Reversed graph embedding resolves complex single-cell trajectories. *Nat Methods* 14:979-982.

Rantakari, P., N. Jappinen, E. Lokka, E. Mokkala, H. Gerke, E. Peuhu, J. Ivaska, K. Elima, K. Auvinen, and M. Salmi. 2016. Fetal liver endothelium regulates the seeding of tissue-resident macrophages. *Nature* 538:392-396.

Rasmussen, U.B., C. Wolf, M.G. Mattei, M.P. Chenard, J.P. Bellocq, P. Chambon, M.C. Rio, and P. Basset. 1993. Identification of a new interferon-alpha-inducible gene (p27) on human chromosome 14q32 and its expression in breast carcinoma. *Cancer Res* 53:4096-4101.

Reyfman, P.A., J.M. Walter, N. Joshi, K.R. Anekalla, A.C. McQuattie-Pimentel, S. Chiu, R. Fernandez, M. Akbarpour, C.I. Chen, Z. Ren, R. Verma, H. Abdala-Valencia, K. Nam, M. Chi, S. Han, F.J. Gonzalez-Gonzalez, S. Soberanes, S. Watanabe, K.J.N. Williams, A.S. Flozak, T.T. Nicholson, V.K. Morgan, D.R. Winter, M. Hinchcliff, C.L. Hrusch, R.D. Guzy, C.A. Bonham, A.I. Sperling, R. Bag, R.B. Hamanaka, G.M. Mutlu, A.V. Yeldandi, S.A. Marshall, A. Shilatifard, L.A.N. Amaral, H. Perlman, J.I. Sznajder, A.C.
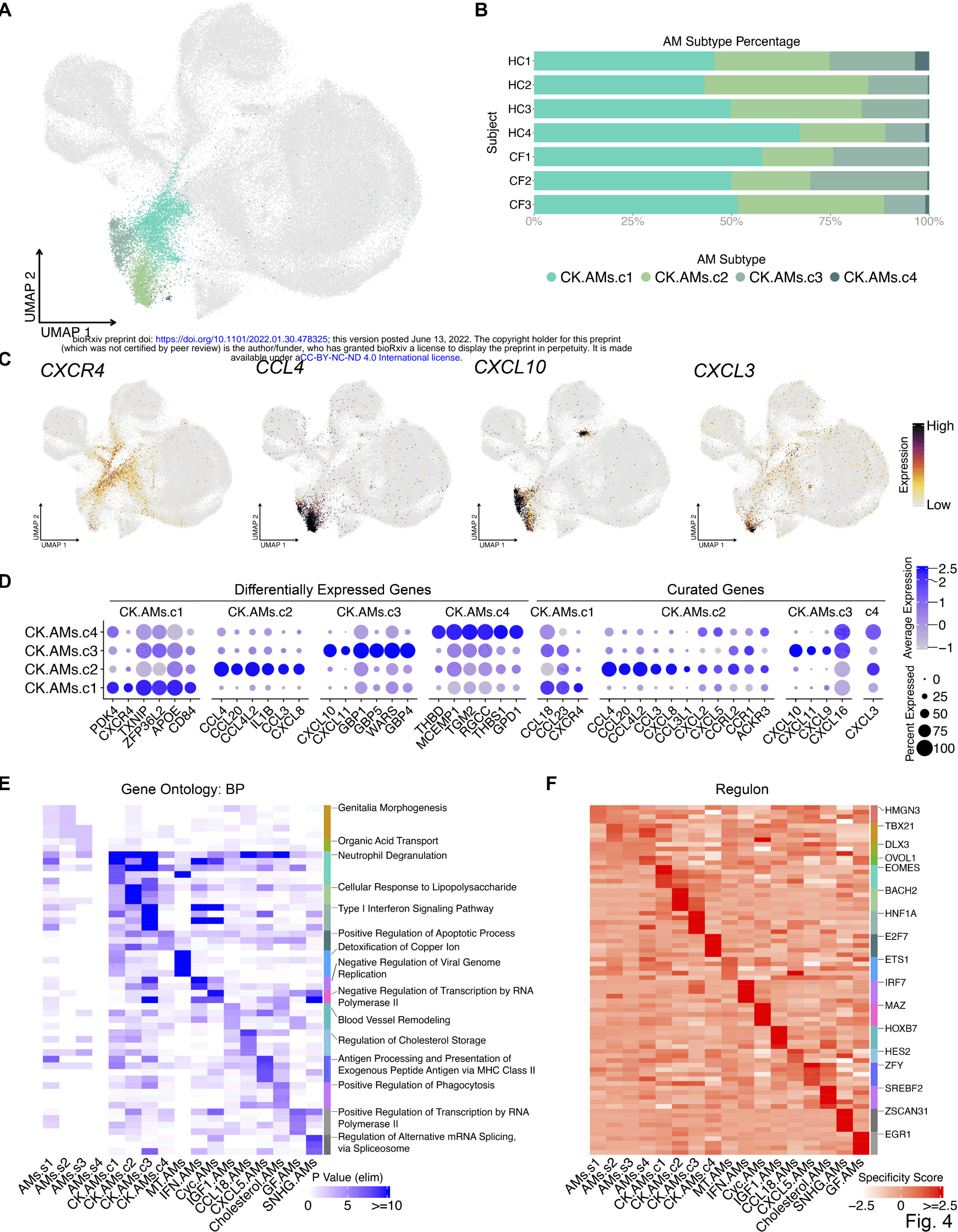
Argento, C.T. Gillespie, J. Dematte, M. Jain, B.D. Singer, K.M. Ridge, A.P. Lam, A. Bharat, S.M. Bhorade, C.J. Gottardi, G.R.S. Budinger, and A.V. Misharin. 2019. Single-Cell Transcriptomic Analysis of Human Lung Provides Insights into the Pathobiology of Pulmonary Fibrosis. *Am J Respir Crit Care Med* 199:1517-1536.

Ruytinx, P., P. Proost, J. Van Damme, and S. Struyf. 2018. Chemokine-Induced Macrophage Polarization in Inflammatory Conditions. *Front Immunol* 9:1930.

Schupp, J.C., S. Khanal, J.L. Gomez, M. Sauler, T.S. Adams, G.L. Chupp, X. Yan, S. Poli, Y. Zhao, R.R. Montgomery, I.O. Rosas, C.S. Dela Cruz, E.M. Bruscia, M.E. Egan, N. Kaminski, and C.J. Britto. 2020. Single-Cell Transcriptional Archetypes of Airway Inflammation in Cystic Fibrosis. *Am J Respir Crit Care Med* 202:1419-1429.

Schyns, J., Q. Bai, C. Ruscitti, C. Radermecker, S. De Schepper, S. Chakarov, F. Farnir, D. Pirottin, F. Ginhoux, G. Boeckxstaens, F. Bureau, and T. Marichal. 2019. Non-classical tissue monocytes and two functionally distinct populations of interstitial macrophages populate the mouse lung. *Nat Commun* 10:3964.

Stoeckius, M., C. Hafemeister, W. Stephenson, B. Houck-Loomis, P.K. Chattopadhyay, H. Swerdlow, R. Satija, and P. Smibert. 2017. Simultaneous epitope and transcriptome measurement in single cells. *Nat Methods* 14:865-868.

Stuart, T., A. Butler, P. Hoffman, C. Hafemeister, E. Papalexi, W.M. Mauck, 3rd, Y. Hao, M. Stoeckius, P. Smibert, and R. Satija. 2019. Comprehensive Integration of Single-Cell Data. *Cell* 177:1888-1902 e1821.

Tang, B.M., M. Shojaei, G.P. Parnell, S. Huang, M. Nalos, S. Teoh, K. O'Connor, S. Schibeci, A.L. Phu, A. Kumar, J. Ho, A.F.A. Meyers, Y. Keynan, T. Ball, A. Pisipati, A. Kumar, E. Moore, D. Eisen, K. Lai, M. Gillett, R. Geffers, H. Luo, F. Gul, J. Schreiber, S. Riedel, D. Booth, A. McLean, and K. Schughart. 2017. A novel immune biomarker IFI27 discriminates between influenza and bacteria in patients with suspected respiratory infection. *Eur Respir J* 49:

Team, R.C. 2014. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL

. *http://www.R-project.org*

Trapnell, C., D. Cacchiarelli, J. Grimsby, P. Pokharel, S. Li, M. Morse, N.J. Lennon, K.J. Livak, T.S. Mikkelsen, and J.L. Rinn. 2014. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat Biotechnol* 32:381-386.

Ural, B.B., S.T. Yeung, P. Damani-Yokota, J.C. Devlin, M. de Vries, P. Vera-Licona, T. Samji, C.M. Sawai, G. Jang, O.A. Perez, Q. Pham, L. Maher, P. Loke, M. Dittmann, B. Reizis, and K.M. Khanna. 2020. Identification of a nerve-associated, lung-resident interstitial macrophage subset with distinct localization and immunoregulatory properties. *Sci Immunol* 5:

Van de Sande, B., C. Flerin, K. Davie, M. De Waegeneer, G. Hulselmans, S. Aibar, R. Seurinck, W. Saelens, R. Cannoodt, Q. Rouchon, T. Verbeiren, D. De Maeyer, J. Reumers, Y. Saeys, and S. Aerts. 2020. A scalable SCENIC workflow for single-cell gene regulatory network analysis. *Nat Protoc* 15:2247-2276.

Vieira Braga, F.A., G. Kar, M. Berg, O.A. Carpaij, K. Polanski, L.M. Simon, S. Brouwer, T. Gomes, L. Hesse, J. Jiang, E.S. Fasouli, M. Efremova, R. Vento-Tormo, C. Talavera-Lopez, M.R. Jonker, K. Affleck, S. Palit, P.M. Strzelecka, H.V. Firth, K.T. Mahbubani, A. Cvejic, K.B. Meyer, K. Saeb-Parsy, M. Luinge, C.A. Brandsma, W. Timens, I. Angelidis, M. Strunz, G.H. Koppelman, A.J. van Oosterhout, H.B. Schiller, F.J. Theis,
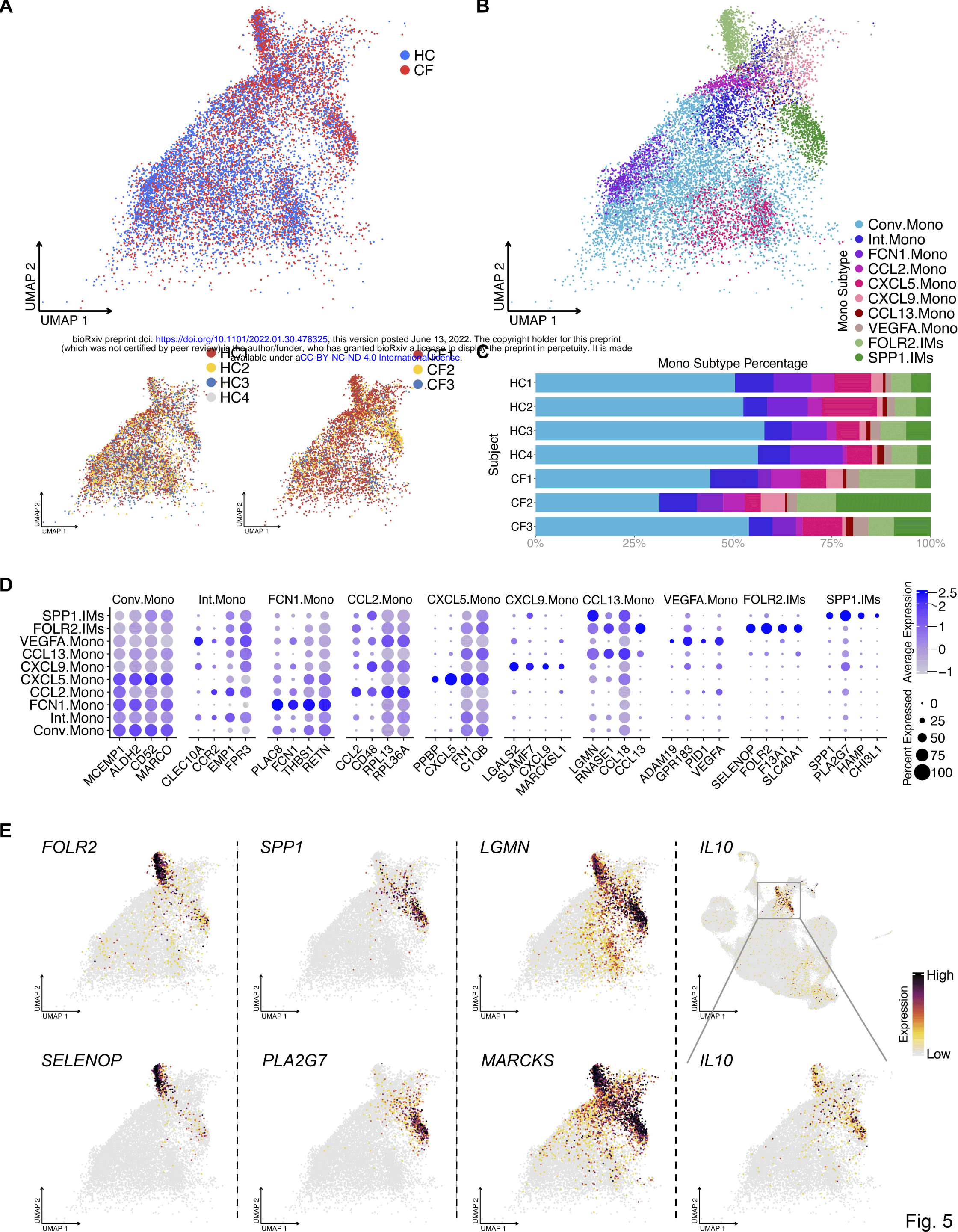
     M. van den Berge, M.C. Nawijn, and S.A. Teichmann. 2019. A cellular census of human lungs identifies novel cell states in health and in asthma. *Nat Med* 25:1153-1163.

Villani, A.C., R. Satija, G. Reynolds, S. Sarkizova, K. Shekhar, J. Fletcher, M. Griesbeck, A. Butler, S. Zheng, S. Lazo, L. Jardine, D. Dixon, E. Stephenson, E. Nilsson, I. Grundberg, D. McDonald, A. Filby, W. Li, P.L. De Jager, O. Rozenblatt-Rosen, A.A. Lane, M. Haniffa, A. Regev, and N. Hacohen. 2017. Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. *Science* 356:

Xu, Y., C. Meng, G. Liu, D. Yang, L. Fu, M. Zhang, Z. Zhang, H. Xia, S. Yao, and S. Zhang. 2016. Classically Activated Macrophages Protect against Lipopolysaccharide-induced Acute Lung Injury by Expressing Amphiregulin in Mice. *Anesthesiology* 124:1086-1099.

Yona, S., K.W. Kim, Y. Wolf, A. Mildner, D. Varol, M. Breker, D. Strauss-Ayali, S. Viukov, M. Guilliams, A. Misharin, D.A. Hume, H. Perlman, B. Malissen, E. Zelzer, and S. Jung. 2013. Fate mapping reveals origins and dynamics of monocytes and tissue macrophages under homeostasis. *Immunity* 38:79-91.

Yu, Y.A., D.F. Hotten, Y. Malakhau, E. Volker, A.J. Ghio, P.W. Noble, M. Kraft, J.W. Hollingsworth, M.D. Gunn, and R.M. Tighe. 2015. Flow Cytometric Analysis of Myeloid Cells in Human Blood, Bronchoalveolar Lavage, and Lung Tissues. *Am J Respir Cell Mol Biol*

Yu, Y.A., and R.M. Tighe. 2018. Isolation and Characterization of Human Lung Myeloid Cells. *Methods Mol Biol* 1809:111-119.

Zilionis, R., C. Engblom, C. Pfirschke, V. Savova, D. Zemmour, H.D. Saatcioglu, I. Krishnan, G. Maroni, C.V. Meyerovitz, C.M. Kerwin, S. Choi, W.G. Richards, A. De Rienzo, D.G. Tenen, R. Bueno, E. Levantini, M.J. Pittet, and A.M. Klein. 2019. Single-Cell Transcriptomics of Human and Mouse Lung Cancers Reveals Conserved Myeloid Populations across Individuals and Species. *Immunity* 50:1317-1334 e1310.
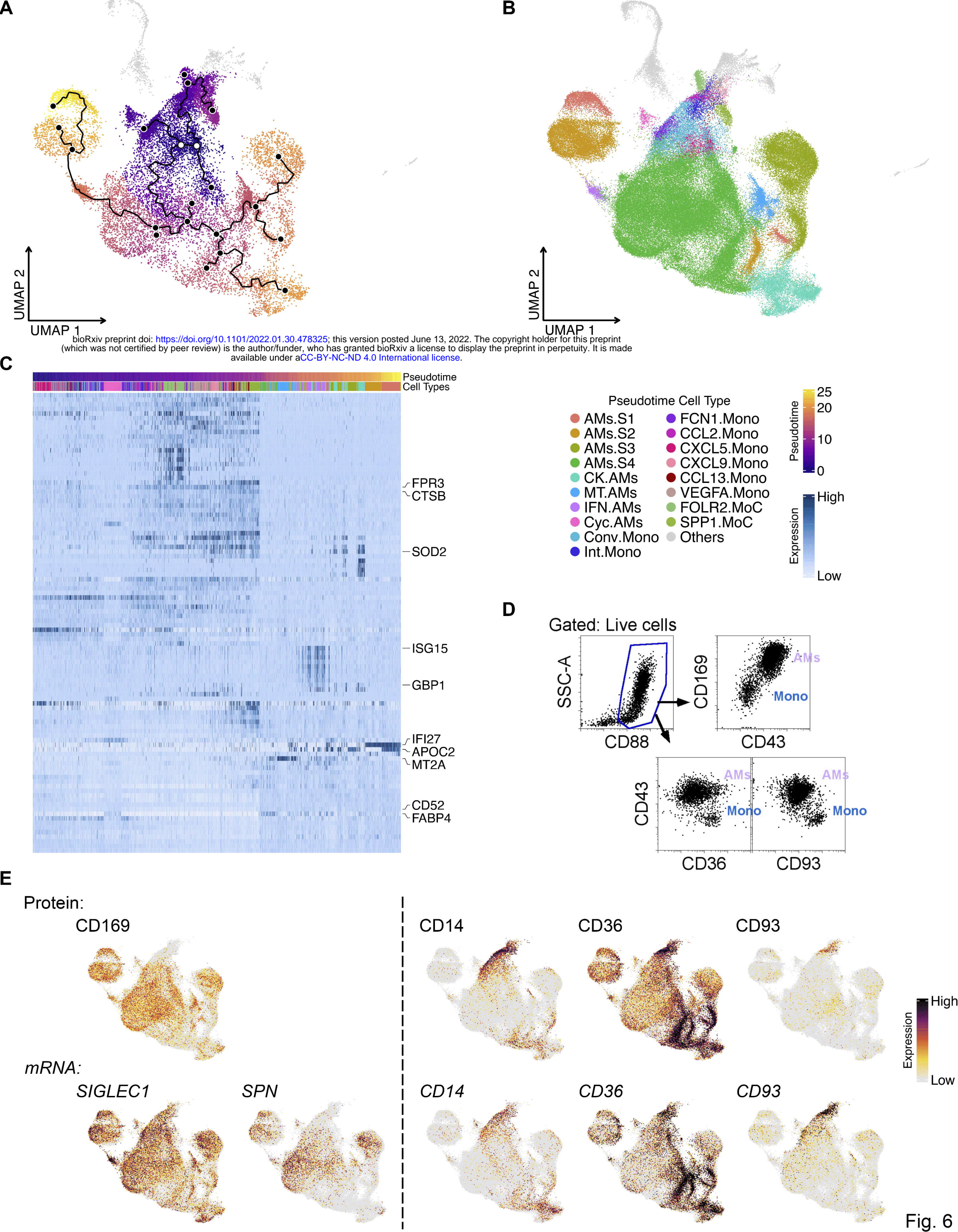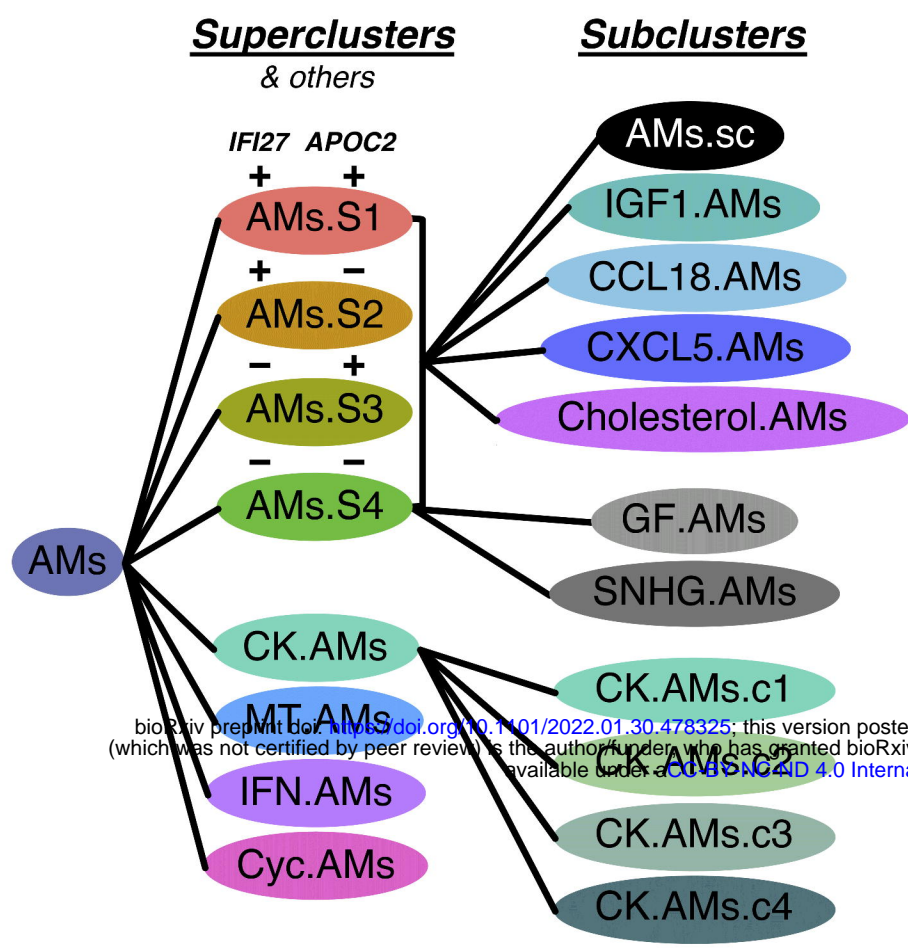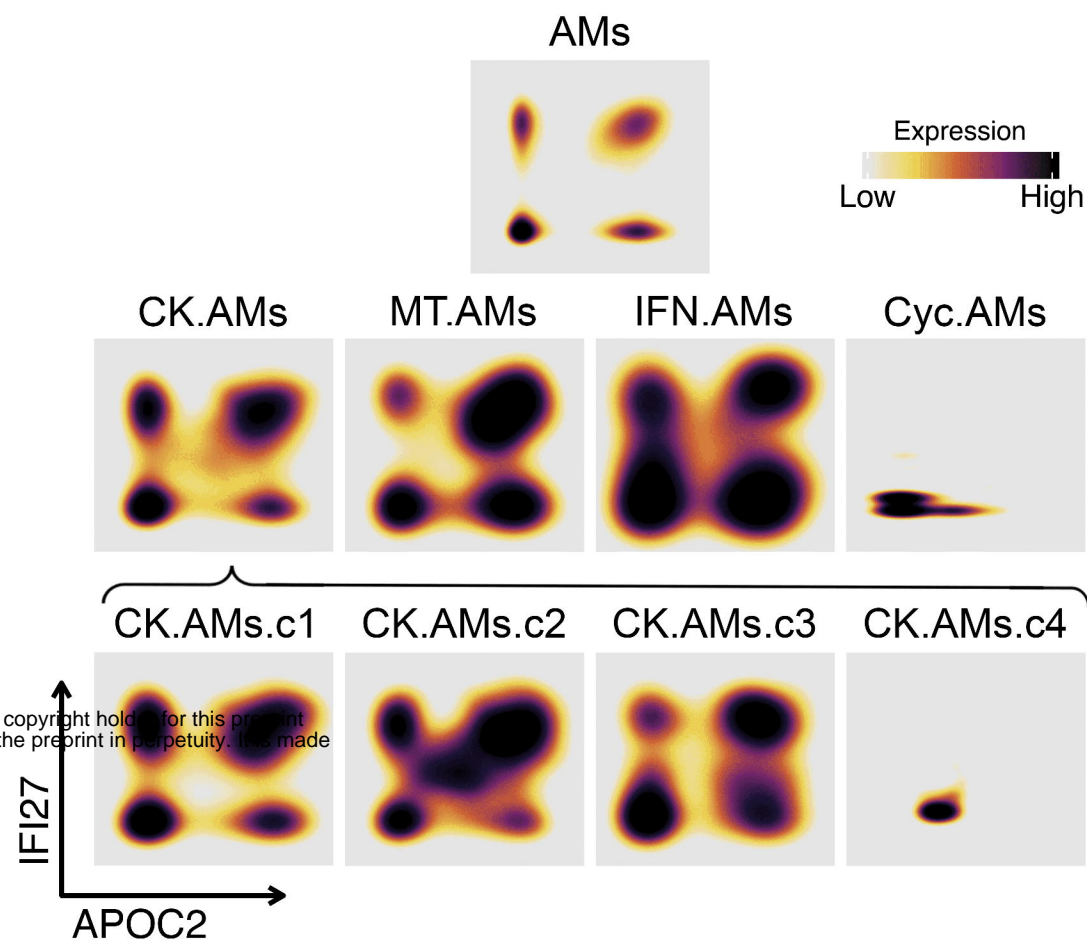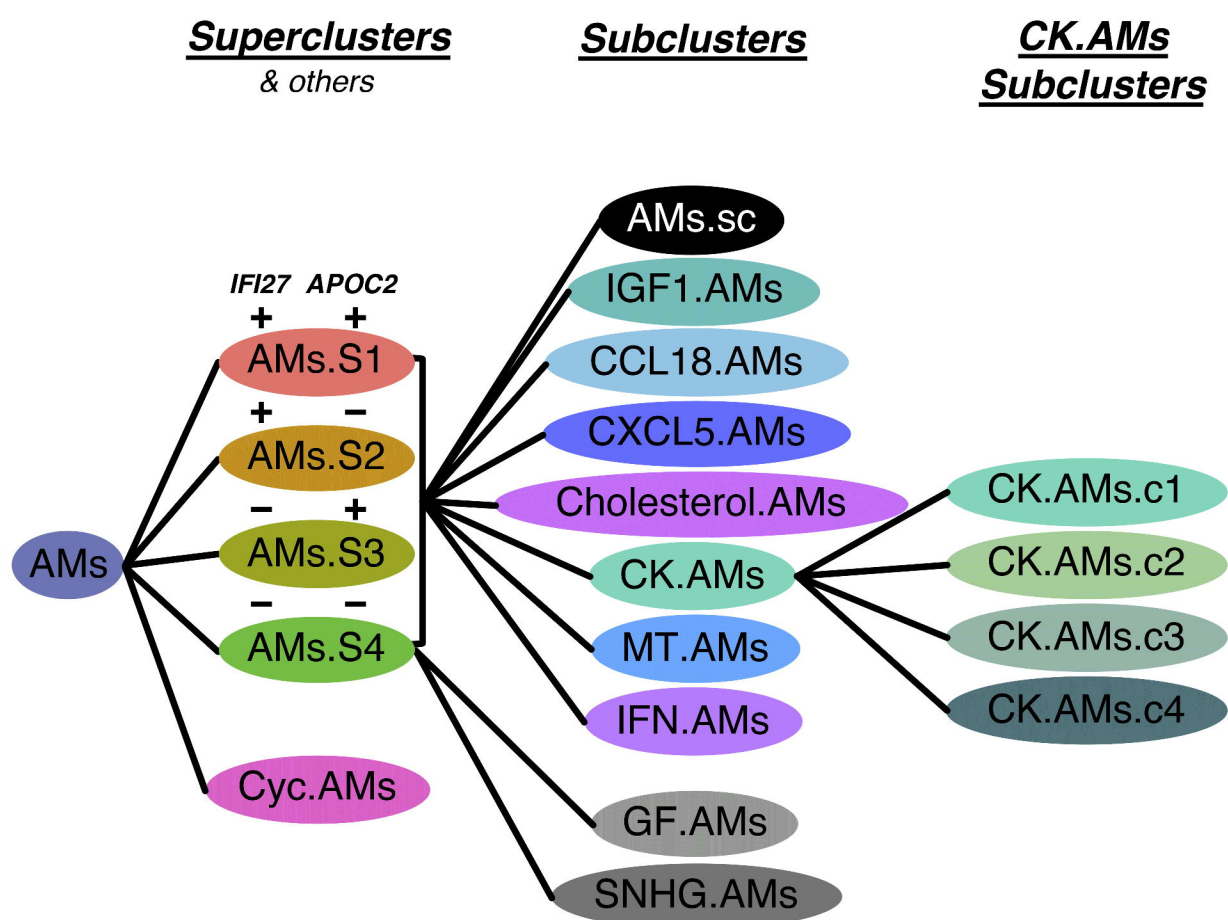
Fig. 1

Fig. 2

A

B
AM Subtype Percentage

C
IGF1   *CCL18*   *CXCL5*   *CYP51A1*

D

Fig. 3

A

UMAP 2 / UMAP 1

B

AM Subtype Percentage

Subject: HC1, HC2, HC3, HC4, CF1, CF2, CF3

AM Subtype: CK.AMs.c1, CK.AMs.c2, CK.AMs.c3, CK.AMs.c4

C

CXCR4 *CCL4* *CXCL10* *CXCL3*

Expression: High / Low

D

Differentially Expressed Genes — Curated Genes

CK.AMs.c1, CK.AMs.c2, CK.AMs.c3, CK.AMs.c4 / CK.AMs.c1, CK.AMs.c2, CK.AMs.c3, c4

Average Expression: −1 to 2.5

Percent Expressed: 0, 25, 50, 75, 100

DEG genes: PDK4, CXCR4, TXNIP, ZFP36L2, APOE, CD84 | CCL4, CCL20, CCL4L2, IL1B, CCL3, CXCL8 | CXCL10, CXCL11, GBP1, GBP5, WARS, GBP4 | THBD, MCEMP1, TGM2, RGCC, THBS1, GPD1

Curated genes: CCL18, CCL23, CXCR4 | CCL4, CCL20, CCL4L2, CCL3, CXCL8, CXCL3L1, CXCL2, CXCL5, CCRL2, CCR1, ACKR3 | CXCL10, CXCL11, CXCL9, CXCL16 | CXCL3

E

Gene Ontology: BP

- Genitalia Morphogenesis
- Organic Acid Transport
- Neutrophil Degranulation
- Cellular Response to Lipopolysaccharide
- Type I Interferon Signaling Pathway
- Positive Regulation of Apoptotic Process
- Detoxification of Copper Ion
- Negative Regulation of Viral Genome Replication
- Negative Regulation of Transcription by RNA Polymerase II
- Blood Vessel Remodeling
- Regulation of Cholesterol Storage
- Antigen Processing and Presentation of Exogenous Peptide Antigen via MHC Class II
- Positive Regulation of Phagocytosis
- Positive Regulation of Transcription by RNA Polymerase II
- Regulation of Alternative mRNA Splicing, via Spliceosome

X-axis: AMs.s1, AMs.s2, AMs.s3, AMs.s4, CK.AMs.c1, CK.AMs.c2, CK.AMs.c3, CK.AMs.c4, MT.AMs, IFN.AMs, Cyc.AMs, IGF1.AMs, CCL18.AMs, CXCL5.AMs, Cholesterol.AMs, GF.AMs

P Value (elim): 0, 5, >=10

F

Regulon

Genes: HMGN3, TBX21, DLX3, OVOL1, EOMES, BACH2, HNF1A, E2F7, ETS1, IRF7, MAZ, HOXB7, HES2, ZFY, SREBF2, ZSCAN31, EGR1

X-axis: AMs.s1, AMs.s2, AMs.s3, AMs.s4, CK.AMs.c1, CK.AMs.c2, CK.AMs.c3, CK.AMs.c4, MT.AMs, IFN.AMs, Cyc.AMs, IGF1.AMs, CCL18.AMs, CXCL5.AMs, Cholesterol.AMs, GF.AMs

Specificity Score: −2.5, 0, >=2.5

Fig. 4

Fig. 5

Fig. 6

A Direct Model

Superclusters
& others

Subclusters

B

AMs

C Proposed Model

Superclusters
& others

Subclusters

CK.AMs Subclusters

Fig. 7

Study	Tissue	Processing	Health Status	Cell No.	Reference
Li	BAL	Immediate loading within 1 hour	HC/CF	94884	
Liao	BAL	Processing within 2 hours	HC/COVID-19	49417	PMID: 32398875
Mould	BAL	Freeze-thaw	HC	38534	PMID: 33079572
Conde	Deceased lung piece	Homogenizing and digesting	HC	17238	PMID: 35549406
Braga	Airway/parenchyma (Brush/Biopsy/Deceased lung piece)	Homogenizing and digesting	HC	4650	PMID: 31209336
Mayr	Fresh lung piece	Homogenizing and digesting	HC/Endstage Fibrosis	64974	PMID: 33650774
Schupp	Sputum	Freeze-thaw	HC/CF	656	PMID: 32603604

B

C

D

Fig. 8