












# Chronic lung diseases are associated with gene expression programs favoring SARS-CoV-2 entry and severity

Linh T. Bui <sup>1,87</sup>, Nichelle I. Winters<sup>2,87</sup>, Mei-I Chung<sup>1</sup>, Chitra Joseph <sup>3</sup>, Austin J. Gutierrez <sup>1</sup>, Arun C. Habermann<sup>2</sup>, Taylor S. Adams<sup>4</sup>, Jonas C. Schupp <sup>4</sup>, Sergio Poli <sup>5</sup>, Lance M. Peter<sup>1</sup>, Chase J. Taylor<sup>2</sup>, Jessica B. Blackburn<sup>2</sup>, Bradley W. Richmond <sup>2,6</sup>, Andrew G. Nicholson<sup>7,8</sup>, Doris Rassl<sup>9</sup>, William A. Wallace <sup>10,11</sup>, Ivan O. Rosas<sup>12</sup>, R. Gisli Jenkins <sup>3</sup>, Naftali Kaminski <sup>4</sup>, Jonathan A. Kropski<sup>2,6,13,88</sup>, Nicholas E. Banovich<sup>1,88</sup> , and the Human Cell Atlas Lung Biological Network\*

Patients with chronic lung disease (CLD) have an increased risk for severe coronavirus disease-19 (COVID-19) and poor outcomes. Here, we analyze the transcriptomes of 611,398 single cells isolated from healthy and CLD lungs to identify molecular characteristics of lung cells that may account for worse COVID-19 outcomes in patients with chronic lung diseases. We observe a similar cellular distribution and relative expression of SARS-CoV-2 entry factors in control and CLD lungs. CLD AT2 cells express higher levels of genes linked directly to the efficiency of viral replication and the innate immune response. Additionally, we identify basal differences in inflammatory gene expression programs that highlight how CLD alters the inflammatory microenvironment encountered upon viral exposure to the peripheral lung. Our study indicates that CLD is accompanied by changes in cell-type-specific gene expression programs that prime the lung epithelium for and influence the innate and adaptive immune responses to SARS-CoV-2 infection.

<sup>1</sup>Translational Genomics Research Institute, Phoenix, AZ, USA. <sup>2</sup>Division of Allergy, Pulmonary and Critical Care Medicine, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, USA. <sup>3</sup>Respiratory Medicine NIHR Biomedical Research Centre, University of Nottingham, Nottingham, UK. <sup>4</sup>Section of Pulmonary, Critical Care and Sleep Medicine, Yale School of Medicine, New Haven, CT, USA. <sup>5</sup>Division of Pulmonary and Critical Care Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA. <sup>6</sup>Department of Veterans Affairs Medical Center, Nashville, TN, USA. <sup>7</sup>National Heart and Lung Institute, Imperial College, London, UK. <sup>8</sup>Department of Histopathology, Royal Brompton and Harefield NHS Foundation Trust, London, UK. <sup>9</sup>Pathology Research, Royal Papworth Hospital NHS Foundation Trust, Cambridge, UK. <sup>10</sup>Department of Pathology, Royal Infirmary of Edinburgh, Edinburgh, UK. <sup>11</sup>Division of Pathology, Edinburgh University Medical School, Edinburgh, UK. <sup>12</sup>Pulmonary, Critical Care and Sleep Medicine, Department of Medicine, Baylor College of Medicine, Houston, TX, USA. <sup>13</sup>Department of Cell and Developmental Biology, Vanderbilt University, Nashville, TN, USA. <sup>87</sup>These authors contributed equally: Linh T. Bui, Nichelle I. Winters. <sup>88</sup>These authors jointly supervised this work: Jonathan A. Kropski, Nicholas E. Banovich. \*A list of authors and their affiliations appears at the end of the paper. email: [nbanovich@tgen.org](mailto:nbanovich@tgen.org)

In December 2019, a respiratory disease associated with a novel coronavirus emerged in Wuhan, China<sup>1–3</sup>. The syndrome, now called COVID-19, was caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and has since rapidly spread worldwide<sup>4</sup>. As of May 18, 2021, a total of over 163 million confirmed COVID-19 cases and more than 3.3 million deaths have been reported around the globe<sup>5</sup>.

The clinical manifestations of SARS-CoV-2 infection range from asymptomatic to fulminant cases of acute respiratory distress syndrome (ARDS) and life-threatening multi-system organ failure. Development of ARDS in patients with SARS-CoV-2 dramatically increases the risk of ICU admission and death<sup>6–12</sup>. Risk factors for severe SARS-CoV-2 include age, smoking status, ethnicity and male sex<sup>13–15</sup>. Baseline comorbidities including hypertension, diabetes and obesity, increase SARS-CoV-2 susceptibility and severity<sup>1,10,16–19</sup>. In addition, chronic lung disease (CLD) has been identified as a risk factor for hospitalization and mortality in patients with COVID-19<sup>20–27</sup>. Patients with chronic obstructive pulmonary disease (COPD) and interstitial lung disease (ILDs), especially Idiopathic Pulmonary Fibrosis (IPF), have a significantly higher COVID-19 mortality rate compared to patients without chronic lung disease<sup>28</sup>. However, the molecular mechanisms underlying the increased risk of SARS-CoV-2 severity and mortality in patients with pre-existing lung diseases are not well understood.

In this work, we performed an integrated analysis of four lung single cell RNA-sequencing (scRNA-seq) datasets<sup>29–32</sup> in addition to unpublished data, together including 78 control and 132 CLD samples ( $n = 31$  COPD, 82 IPF and 19 other ILDs), to investigate the molecular basis of SARS-CoV-2 severity and mortality risk in CLD patients. We found that CLD is associated with baseline changes in cell-type specific expression of genes related to viral replication and the immune response, as well as evidence of immune exhaustion and altered inflammatory gene expression. Together, these data provide a molecular framework underlying the increased risk of SARS-CoV-2 severity and poor outcomes in patients with certain pre-existing CLD.

## Results

**Integrated analysis of lung single cell RNA sequencing datasets.** To determine why COVID-19 patients with CLD have a higher risk of severe infection and poorer outcomes, we performed an integrated analysis on four published scRNA-seq lung datasets: Northwestern (biopsy deemed representative of explanted lung), Pittsburgh and VUMC/TGen (biopsy of apical and basal region of explanted lung) and Yale/BWH (longitudinal biopsy through explanted lung) (Supplementary Table 1a), in addition to previously unpublished samples (VUMC/TGen). We analyzed the transcriptomes from 611,398 single cells derived from healthy donors (78 samples), COPD (31 samples), IPF (82 samples) and Non-IPF ILD (Other ILD, 19 samples) (Supplementary Table 1b, 2). Using published cell type specific markers<sup>31,32</sup>, we identified 32 distinct cell types in the dataset (Supplementary Fig. 1). Overall, we observed similar cell type proportions between the different datasets and diagnosis groups (Supplementary Table 3, Supplementary Fig. 2), with the exception of high AT2 cell numbers in the Northwestern dataset, as expected due to the protocol favoring isolation of AT2 and macrophages (personal communication). Despite the variation in sample collection and processing at different research institutes, the similarity in cell type composition per dataset indicated the compatibility of samples and that no major sampling bias would confound an integrated analysis.

**Expression profile of SARS-CoV-2 associated receptors and factors in the diseased lung.** SARS-CoV-2 utilizes the host *ACE2*,

and other putative factors such as *BSG*, *NRP1* and *HSPA5*, as entry receptors and *TMPRSS2*, *CTSL* or *FURIN* as priming proteases to facilitate cellular entry<sup>33–40</sup>. Consistent with prior reports analyzing normal lung tissue<sup>33,34,41</sup>, *ACE2* and *TMPRSS2* are expressed predominantly in epithelial cell types (Fig. 1a), while other putative SARS-CoV-2 entry receptors (*BSG*, *NRP1*, *HSPA5*) and priming proteases (*CTSL*, *FURIN*) have substantially more widespread expression in nearly all cell types (Supplementary Fig. 3). The total number and proportion of *ACE2* + cells are highest in pericytes, type 2 alveolar cells (AT2) and secretory cells, while *TMPRSS2* is widely expressed in all epithelial cell types. There were no significant differences in the proportion of *ACE2* + cells in any cell-type in CLD versus control groups (Fig. 1b). The proportion of *TMPRSS2* + AT2 cells is decreased in IPF lungs while *TMPRSS2* + AT1 and Transitional AT2 cells are higher in all CLD samples; and *TMPRSS2* + *SCGB3A2* + *SCGB1A1* + club cells are in significantly higher numbers in COPD patients compared to controls (Fig. 1c). The putative entry factor *NRP1* is expressed in more pDCs, myofibroblasts and HAS1 high fibroblasts in CLD samples compared to control (Supplementary Fig. 3).

Next, we compared the number of double positive cells, i.e., cells co-expressing a receptor and priming protease, in control and CLD samples. A notable fraction of cells co-expresses all established and putative entry receptors (*ACE2*, *BSG*, *NRP1*, *HSPA5*) and proteases (*TMPRSS2*, *CTSL*, *FURIN*); AT2 cells comprised nearly half of all such cells (43.3%) (Fig. 1d, Supplementary Fig. 4). While the percentage of cells co-expressing *ACE2* and priming proteases (*TMPRSS2*, *CTSL*, *FURIN*) was similar across disease subtypes, there was a significantly higher number of cells co-expressing *ACE2* and *FURIN* in the COPD AT2 and Transitional AT2 cells (Fig. 1e, Supplementary Fig. 4). We detected significant differences in the number of cells co-expressing *BSG*, *NRP1*, *HSPA5* with a priming protease in CLD samples in multiple cell types (Fig. 1f, Supplementary Fig. 4).

To examine whether CLD patients express higher levels of SARS-CoV-2 receptors and priming proteases, we performed differential expression analysis of those genes in the CLD versus control samples. The two major SARS-CoV-2 cellular entry factors, *ACE2* and *TMPRSS2*, have similar expression profiles in the disease and control samples. *ACE2* expression is relatively low in all cell types and there were no significant differences in *ACE2* expression in CLD groups compared to control (Supplementary Fig. 5). The putative alternative receptor *NRP1*, recently confirmed as another host entry factor for SARS-CoV-2<sup>35</sup>, is slightly up-regulated in the COPD macrophages, but down-regulated in both IPF and Other-ILD macrophages (Supplementary Fig. 6). *TMPRSS2* expression is high in AT1, AT2, Transitional AT2, PNEC/ionocytes and club cells (Fig. 2b, Supplementary Fig. 5) and is slightly upregulated in the AT2 COPD samples ( $\log_2FC = 0.28$ ,  $q$  value 0.04) (Supplementary Dataset 3), in contrast to a recent publication demonstrating decreased *TMPRSS2* expression in severe COPD<sup>42</sup>. Two alternative priming proteases (*CTSL* and *FURIN*) are expressed at low level and show no significant differences in expression between control and disease samples (Fig. 2b, Supplementary Fig. 5). However, the SARS-CoV-2 entry gene score (calculated on the average expression levels of all SARS-CoV-2 entry factors over a random control gene set) is significantly increased in the CLD samples in many epithelial cell types, including AT1, AT2, Basal, Club cells, and *KRT5*/*KRT17* + cells, an ECM-producing epithelial cell type enriched in the fibrotic lung<sup>31,32</sup> (Fig. 2c, Supplementary Fig. 8a, b). Together, these data suggest CLD is associated with modest changes in expression of established SARS-CoV-2 entry factors, and alternative mechanisms are likely

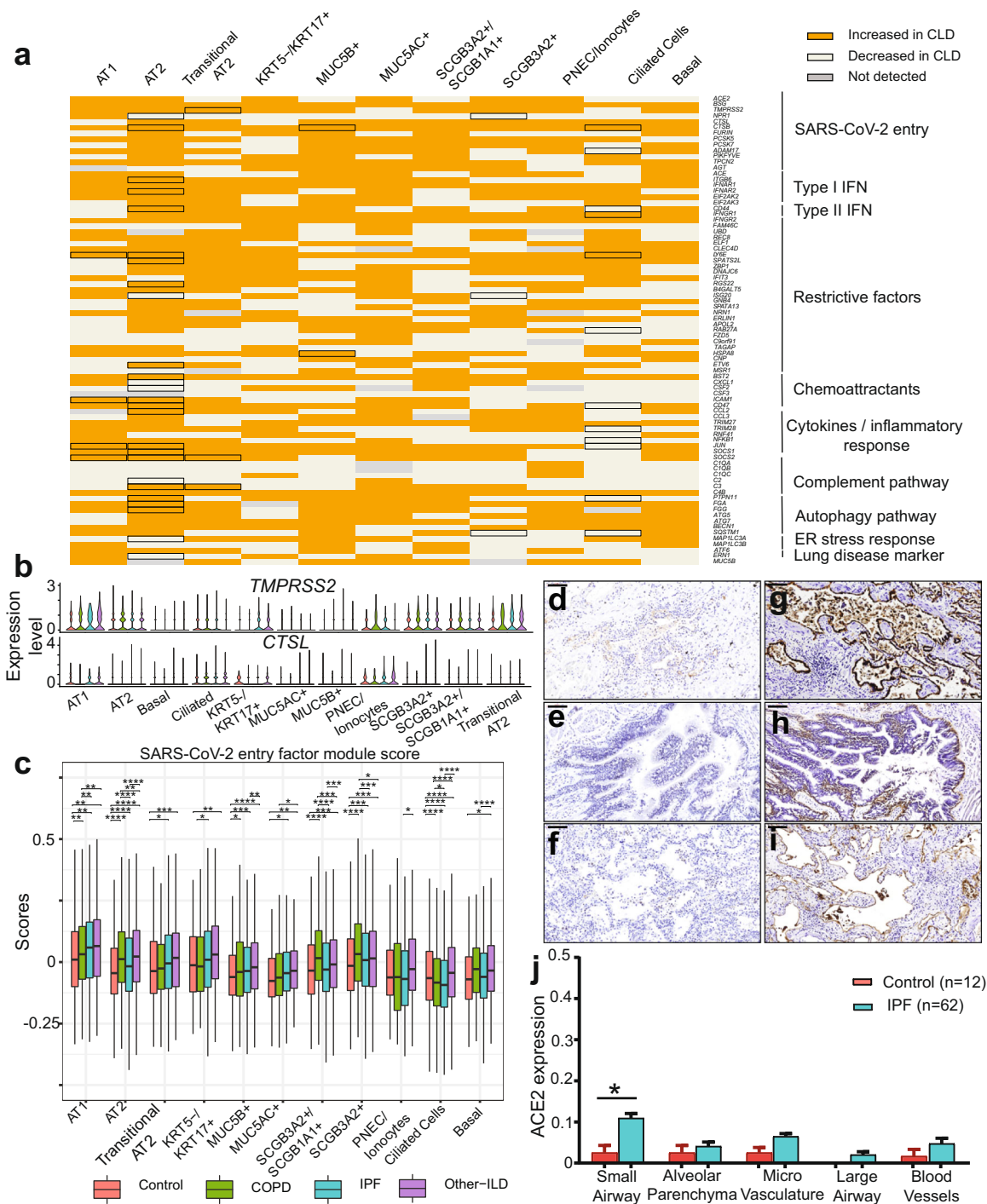


**Fig. 1** Percentage of cells expressing SARS-CoV-2 receptor genes in lung cell types in different diagnosis subgroups. **a** Percentage of cells expressing *ACE2* and *TMPRSS2* in all cell types. Numbers are the total number of *ACE2* + or *TMPRSS2* + cells in each cell type in the dataset. **b, c** Percentage of cells expressing *ACE2* (**b**) and *TMPRSS2* (**c**) in each diagnosis group in the epithelial cell types. **d** Venn diagram shows overlapping of cells co-expressing the proposed receptors (*ACE2*, *BSG* and *NRP1*) and the protease *TMPRSS2*. **e, f** Percentage of cells co-expressing receptors and *TMPRSS2* split by cell type and diagnosis group. Plots were generated with mean values of percentage of cells per individual samples, and data are presented as mean values  $\pm$  SEM. Significant differences between diagnosis groups were calculated using Tukey\_HSD test,  $p$  value  $< 0.05$ ; \* $p$  value  $< 0.01$ ; \*\* $p$  value  $< 0.001$ ; \*\*\*\* $p$  value  $< 0.0001$ ; \*\*\*\*.

additionally responsible for observed differences in outcome severity.

**Dysregulation of viral infection and innate immune response genes in disease epithelial cells.** Given the relatively modest differences in SARS-CoV-2 entry factors in epithelial cells between CLD and control lungs (Fig. 2b, Supplementary Figs. 5, 6), we hypothesized that rather than greatly increased cellular

susceptibility to SARS-CoV-2 infection, patients with CLD are predisposed to severe lung injury due to underlying differences in epithelial gene expression in key pathways mediating the antiviral response. Focusing on the epithelial cell population (a total of 143,114 cells), we selectively examined genes that have been demonstrated in the SARS, MERS and rapidly expanding COVID-19 literature to impact viral pathogenesis. We noted that many of these genes are significantly dysregulated (Bonferroni adjusted  $p$ -value,  $FDR, \leq 0.1$ ) in several epithelial cell types



**Fig. 2** Expression profile of SARS-CoV-2 mediators and response genes in the epithelial cell population. **a** Binary heatmap representing a manually curated list of genes associated with SARS-CoV-2. Orange elements indicate genes with increased expression and white elements indicate genes with decreased expression in CLD samples; Not detected: gene expression was not detected in either of the two tested populations (CLD vs. Control). Differential expressed genes (DEGs) between CLD and control samples ( $FDR \leq 0.1$ ) are outlined in black. **b** Violin plot depicts gene expression level in CLD and control of the two SARS-CoV-2 proteases *TMPRSS2* and *CTSL*. **c** SARS-CoV-2 entry module score in different cell types, SARS-CoV-2 mediators included *ACE2*, *BSG* (*CD147*), *NPR1*, *HSPA5* (*GRP78*), *TMPRSS2*, *CTSL*, *ADAM17*, *FURIN*. The outliers were removed in this plot, please see Supplementary Fig. 8a with outliers included. Boxes: interquartile range, lower and upper hinges correspond to the first and third quartiles, upper and lower whisker extends from the hinge to the largest values or smallest values of 1.5 x interquartile range; \* $p$  value < 0.05, \*\* $p$  value < 0.01, \*\*\* $p$  value < 0.001, \*\*\*\* $p$ -value < 0.0001, Tukey\_HSD post-hoc test. ACE2 and ITGB6 protein expression in IPF lung sections. IPF lung sections stained for ACE2: **d** small airway, **e** large airway and **f** lung parenchyma. IPF lung sections stained for  $\alpha$ 6: **g** small airway, **h** large airway and **i** lung parenchyma. **j** Semi-quantitative evaluation of ACE2 scoring among control ( $n = 12$  for each tissue) and IPF ( $n = 62$  for each tissue) sections (both the percentage of staining and staining intensity of ACE2 expression; 0-Negative; 1-0- $\leq 10\%$ ; 2-11- $\leq 25\%$ ; 3- $\leq 26\%$ ), data are presented as mean values  $\pm$  SEM. Significant differences between IPF and control were calculated using Tukey HSD test,  $p$  value < 0.05 \*. Scale bar = 100  $\mu$ m. For **d-j**: a total of 12 normal lung samples and 62 IPF samples were used.



(Fig. 2a). Included are genes thought to directly impact viral replication (*TMPRSS2*, *NRP1*, *CTSB*), interferon stimulated genes (ISGs) thought to be involved in restricting viral entry and replication (*LY6E*, *SPATS2L*)<sup>43</sup>, and key regulators of the host viral response including cytokine and inflammatory response genes (IFN type I and type II receptors, *SOCS1/2*, *CCL2*, *CD47*) (Fig. 2a, Supplementary Fig. 8c, d). In addition, the complement pathway gene *C3*, an important component of the innate immune response and previously found to be elevated in SARS patients<sup>44</sup>, and autophagy (*FGG*, *FGA*, *PTPN11*) genes are also significantly dysregulated in many CLD epithelial cells; these pathways are important for propagating viral infection and the host response<sup>45–47</sup>. Among the epithelial cell types, AT2 cells have the largest number of significantly dysregulated genes in CLD compared to control samples (Fig. 2a). These data suggest that there are basal differences in the expression profiles of genes regulating viral infection and the immune response in diseased epithelial cells, in particular in AT2 cells, and that this epithelial “priming” may contribute to COVID-19 severity and poor outcomes.

**Elevated ACE2 protein expression level in the small airways in IPF lungs.** To further study the expression of the major SARS-CoV-2 entry factor *ACE2* in the fibrotic lungs, we examined protein levels of *ACE2* in different lung regions using the anti-*ACE2* ab108252 antibody (Supplementary Fig. 9a). In agreement with the transcript quantification above and previous immunohistology analysis<sup>48</sup>, we detected overall low expression level of *ACE2* across all tissue types in both IPF (Fig. 2d–f) and control lung sections (Supplementary Fig. 9b, c). Semi-quantitative evaluation of *ACE2* expression scoring showed elevated *ACE2* expression in all IPF sections compared to control, reaching statistical significance in the IPF small airway sections (Fig. 2j), suggesting that while overall *ACE2* expression is low, there is a regional concentration of *ACE2* + cells within the distal IPF lung that may promote a more severe localized viral response. Upregulation of the epithelial integrin alpha-V beta-6 ( $\alpha v\beta 6$ ) plays an important role in enhanced fibrosis in response to lung injury<sup>49</sup>, and enhances TGF $\beta$  activation which can suppress type I interferon responses from alveolar macrophages increasing susceptibility to viral infection<sup>50</sup>. We detected a significant increase of  $\alpha v\beta 6$  integrin expression in all lung sections isolated from IPF patients (Fig. 2g–i). While there was additional positive staining in the peripheral lung,  $\alpha v\beta 6$  expression is highest in the AT2 epithelial cells in the IPF samples compared to overall low expression level in the normal lung sections (Supplementary Fig. 9b, c), mirroring the expression data of *ITGB6* described below (Fig. 3b).

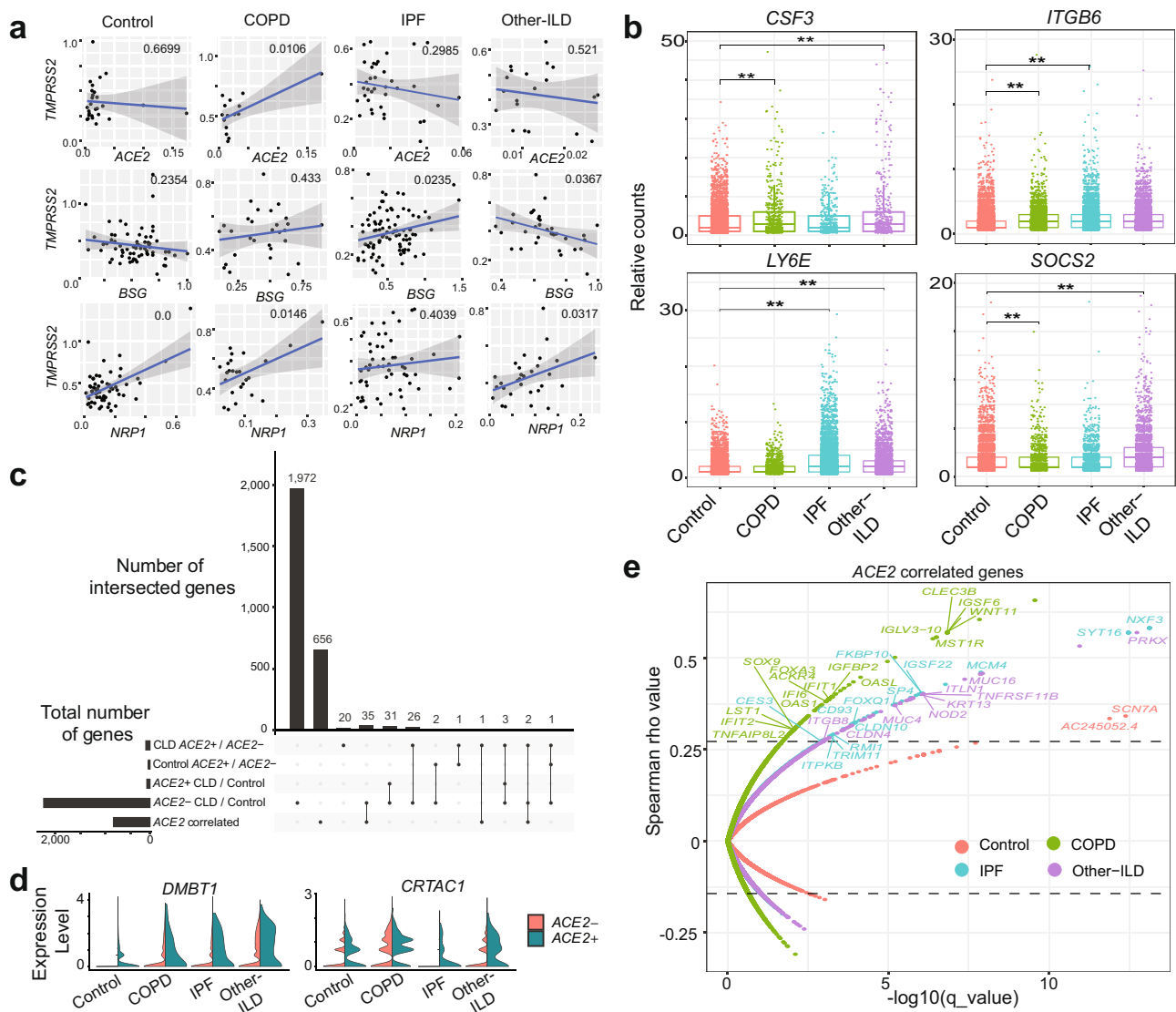
**CLD specific ACE2+ transcriptional profiles in AT2 cells.** In the distal lung, AT2 cells have been proposed to be the primary targets of SARS-CoV-2<sup>34,41,51</sup> and comprise the initial micro-environment the virus encounters. Thus, we examined the gene expression profile of CLD AT2 cells in more detail. As described above, AT2 cells in all diagnosis subgroups have significantly higher SARS-CoV-2 entry gene scores than control cells (Fig. 2c). In addition, CLD AT2 cells express higher levels of many genes related to viral infection and innate immune responses than any epithelial cell type (Fig. 2a). COPD and Other-ILD, but not IPF, AT2 cells express higher levels of *CSF3*, an important cytokine in the regulation of granulocytes, and the suppressor of cytokine signaling-2 (*SOCS2*) (Fig. 3b). The epithelial integrin *ITGB6*, involved in wound healing and pathogenic fibrosis<sup>52</sup>, is upregulated in COPD and IPF AT2 cells; the ISG lymphocyte antigen 6 complex (*LY6E*), known to restrict SARS-CoV-2 entry<sup>43,53</sup>, is upregulated in the IPF and Other-ILD AT2 cells (Fig. 3b, Supplementary Fig. 8e). Gene correlation analysis showed strong

positive correlation between *TMPRSS2* and *ACE2*, *NRP1* in COPD AT2 cells (Fig. 3a, Supplementary Fig. 10). *NRP1* expression is also positively correlated with the protease *FURIN* in the AT2 cells isolated from IPF samples (Supplementary Fig. 10c).

Since *ACE2* is the best-established SARS-CoV-2 entry factor, and AT2 cells accounted for 54.63% of all *ACE2* + epithelial cells, we focused on the transcriptional profile of *ACE2* + AT2 cells. All of the 34 differentially expressed genes ( $FDR \leq 0.1$ ) in *ACE2* + AT2 cells between CLD and control overlapped with the *ACE2*- cells CLD vs. control analysis (Fig. 3c), suggesting that these genes reflected the disease state and were not related to *ACE2* expression. However, when we performed the same differential expression analysis on *ACE2* + vs. *ACE2*- CLD cells, we identified 20 unique genes that were dysregulated in CLD *ACE2* + cells (Fig. 3c, Supplementary Table 4). Among these 20 genes, the tumor suppressor *DMBT1*, a glycoprotein that has been shown previously to be highly expressed in *ACE2* + AT2 cells<sup>54</sup> and can bind to SARS-CoV-2 spike proteins<sup>55</sup>, and the cartilage acidic protein 1 (*CRTAC1*), previously known to be downregulated significantly in COVID-19 patients with severe infection<sup>56</sup>, were upregulated in *ACE2* + compared to *ACE2*- AT2 CLD cells (Fig. 3d).

Next, we sought to identify *ACE2* correlated genes in the *ACE2* + AT2 cells in different disease groups; thus, identifying the immediate cellular environment SARS-CoV-2 encounters upon infecting a host. We performed Spearman correlation analysis with Benjamini–Hochberg adjusted *p* values and identified distinct gene profiles significantly correlated with *ACE2* for each disease group (Fig. 3c, e). There were only two *ACE2* correlated genes in the Control samples with a cutoff of 99th percentile Spearman rho values and *q* value less than 0.03, none of those genes are associated with the immune response. In the disease samples, we identified 706 genes (COPD: 330 genes, IPF: 108 genes and Other-ILD: 268 genes) with significant correlation to *ACE2* (99th percentile rho values, *q* value less than 0.03) (Supplementary Dataset 4). *ACE2* correlated genes are involved in various cellular processes, including viral processes (Supplementary Table 5). Many *ACE2*-correlated genes in the disease samples are associated with the innate and antiviral immune response. In the COPD samples, genes with strong correlation coefficients with *ACE2* include several interferon-induced genes (*IFI6*, *IFIT1*, *IFIT2*), a modulator of innate immune function (*OAS1*), the chemokine receptor *ACKR4*, a gene associated with West Nile viral infection (*OASL*), and the ECM regulated transcription factor *SOX9*. In the IPF samples, *ACE2* expression is strongly correlated with the nuclear factor *NXF3*, the transcription factor *SP4*, the antiviral factor *TRIM11*, and the Forkhead Box Q1 (*FOXQ1*). In other ILD diseases (non-IPF related), the integrin *ITGB8*, a member of the TNF receptor family (*TNFRSF11B*), an important component of the immune response system (*NOD2*) and an innate immune pathway component (*ITLN1*) are among the genes with high correlation with *ACE2*. The transcription factor *FOXQ1* was identified among the 20 unique transcription factors specific for SARS-CoV-2 in a recent in silico study<sup>57</sup>, while *OAS1* was among the top 50 genes with a significant correlation coefficient with *ACE2* in a previous study<sup>33</sup>. The presence of immune-associated genes in these gene correlation profiles suggests that in patients with CLD, *ACE2* + AT2 cells are conditioned and primed to express these genes to cope with viral infection.

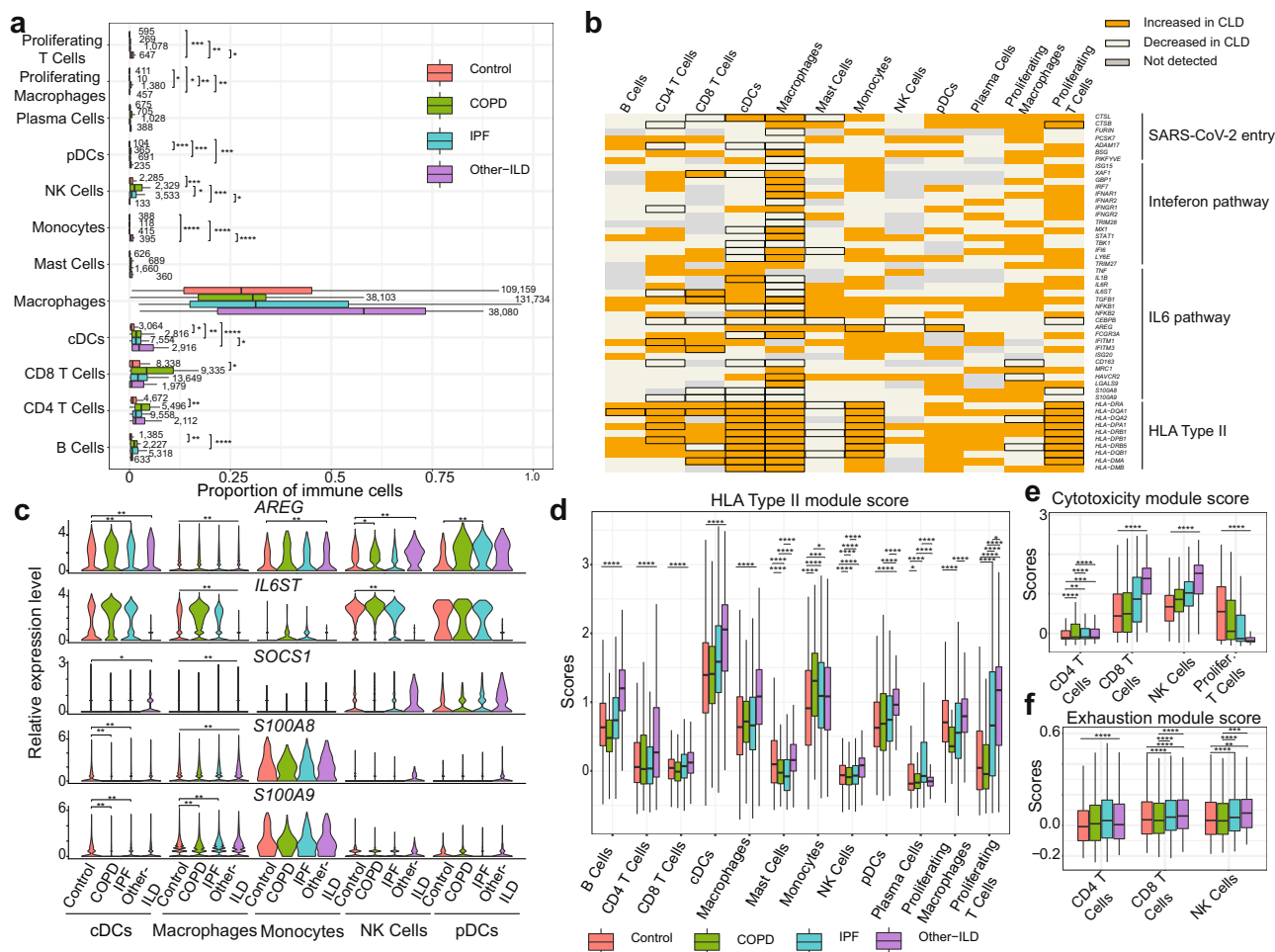
**Baseline differences in inflammatory response programs in chronic lung disease.** Recent publications have suggested that immune dysregulation, including sustained cytokine production and hyper-inflammation, is associated with SARS-CoV-2 severity<sup>58–61</sup>. We performed an in-depth examination of the



**Fig. 3** CLD AT2 cells exhibit baseline differences in gene expression profile coping with viral infection. **a** Significant gene expression correlation in AT2 cells between *TMPPRSS2* and *ACE2*, *BSG* (*CD147*) and *NRP1* in COPD and IPF samples, each dot represents the average expression level of the genes of interest per sample, pairwise gene correlation analysis was done using a fitting linear model and *p* value was calculated using Anova. **b** Boxplot shows differences in gene expression of selected SARS-CoV-2 response genes in the AT2 cell types among different diagnosis groups, Boxes: interquartile range, lower and upper hinges correspond to the first and third quartiles, upper and lower whisker extends from the hinge to the largest values or smallest values of  $1.5 \times$  interquartile range; \*\**p* value-adj  $\leq 0.05$  (negative binomial test, corrected for Age, Ethnicity, Smoking\_status and Dataset). **c** Upset plot shows shared differential expression genes (DEGs) between different comparisons: *ACE2*<sup>-</sup> CLD vs. Control, *ACE2*<sup>+</sup> CLD vs. Control, CLD *ACE2*<sup>+</sup> vs. *ACE2*<sup>-</sup>, Control *ACE2*<sup>+</sup> vs. *ACE2*<sup>-</sup> and *ACE2* correlated genes in the AT2 cells. **d** Upregulation of two genes uniquely differentially expressed in the CLD *ACE2*<sup>+</sup> vs. *ACE2*<sup>-</sup>. **e** Spearman gene correlation analysis identified genes correlated with *ACE2* expression in AT2 *ACE2*<sup>+</sup> cells in different diagnosis groups, *p*-value was adjusted using Benjamini-Hochberg corrections, dashed lines indicate the 99th percentile of Spearman rho values.

immune cell population to determine whether preexisting immune dysregulation in chronic lung disease patients could contribute to SARS-CoV-2 severity and mortality. We analyzed a total of 421,059 cells from 12 immune cell types (Supplementary Table 3, Supplementary Fig. 1) and found significant increases in the proportion of CD4 T Cells, CD8 T Cells, cDCs and NK cells in the disease groups, most notably in COPD samples (Fig. 4a). Similar to Fig. 2a, we examined the expression of SARS-CoV-2 and cellular immune response genes in the CLD immune cells. Several genes related to SARS-CoV-2 entry (*CTSL*, *CTSB*, *ADAM17*) and components of the Interferon and IL6 pathways are significantly dysregulated in the CLD Macrophages and cDCs (Fig. 4b). Moreover, many immune cells isolated from CLD samples showed elevated levels of genes in the major

histocompatibility complex (MHC) class II genes (HLA type II genes) (Fig. 4b). HLA type II gene module score increased across all disease groups but especially in the Other-ILD samples, compared to controls (Fig. 4d). Type I IFN response (IFN $\alpha$  score) is slightly elevated in the diseased macrophages and pDCs (Supplementary Fig. 11a). IL6-associated tocilizumab responsive genes (IL6 score) are expressed at a higher level in the disease groups IPF and Other-ILD, but lower in the COPD samples (Supplementary Fig. 11b). Previous studies demonstrated elevated exhaustion levels in CD8 T cells in severely affected COVID-19 patients<sup>62,63</sup>. All CLD T cells have higher expression levels of cytotoxicity and exhaustion genes compared to controls (Fig. 4e, f). These perturbations in the T Cell population of CLD lungs may diminish the host immune response to viral infection,



**Fig. 4 Analysis of SARS-CoV-2 candidate immune response genes in immune cells.** **a** Quantification of cell types as a percent of all immune cells in control and diseased lungs, numbers represent the total numbers of the cell type per individual samples, data are presented as mean values  $\pm$  SEM. **b** Binary heatmap representing a manually curated list of genes associated with SARS-CoV-2. Orange elements indicate genes with increased expression and white elements indicate genes with decreased expression; Not detected: gene expression was not detected in either of the two tested populations (CLD vs. Control); DEGs with  $FDR \leq 0.1$  are outlined in black. **c** Differential expression analysis for SARS-CoV-2 immune candidate genes in cDCs, Macrophages and Monocytes. \* $p$ -adjusted value  $< 0.1$ , \*\* $p$ -adjusted value  $< 0.05$ ,  $p$ -adjusted value was Bonferroni adjusted from Seurat FindMarkers differential expression analysis using a negative binomial test and corrected for Age, Ethnicity, Smoking\_status and Dataset. **d** Compared to the healthy control samples, HLA type II score is higher in all disease groups (especially Other-ILD). In the T cell population, cytotoxicity scores (**e**) and exhaustion scores (**f**) are higher in the disease samples than in control samples. In **a**, **d**, **e**, and **f**: Boxes: interquartile range, lower and upper hinges correspond to the first and third quartiles, upper and lower whisker extends from the hinge to the largest values or smallest values of 1.5 x interquartile range; Tukey\_HSD post-hoc test: \* $p$  value  $< 0.05$ , \*\* $p$  value  $< 0.01$ , \*\*\* $p$  value  $< 0.001$ , \*\*\*\* $p$  value  $< 0.0001$ . See Supplementary Fig. 10 for plots with outliers included for **d-f**.

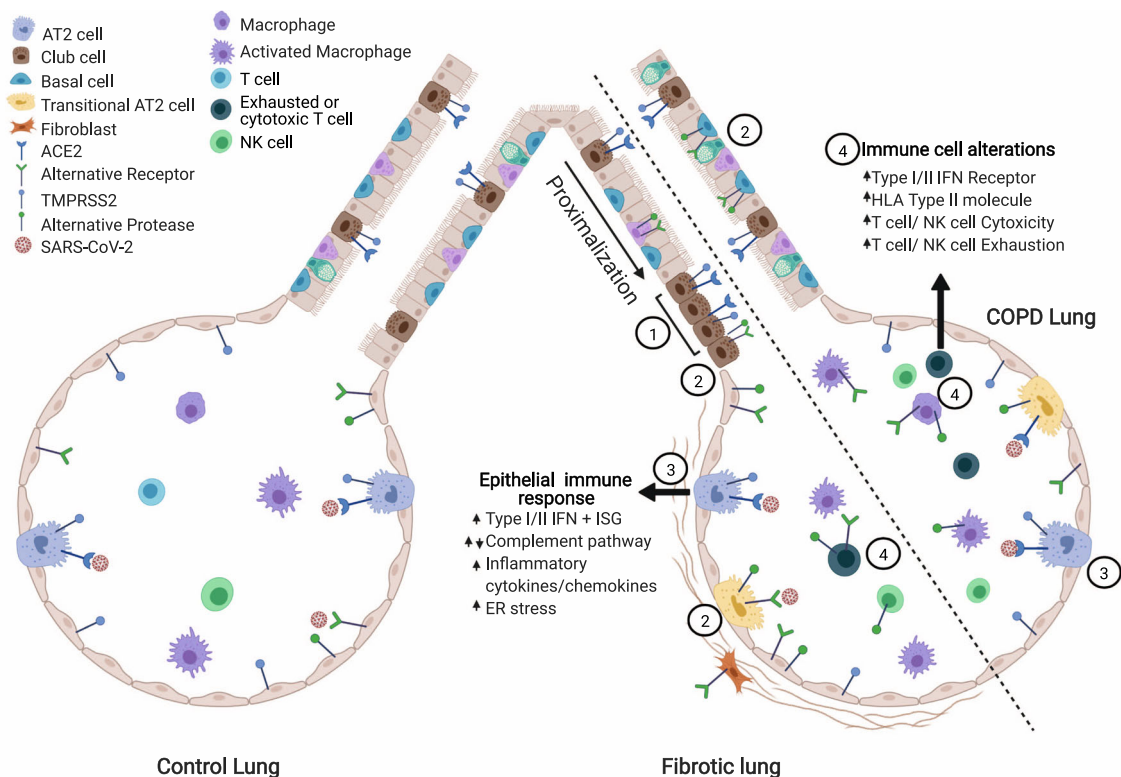
leading to a higher risk of severe disease and poor outcomes in response to SARS-CoV-2 infection.

To further investigate differences in immune cell type-specific gene expression profiles, we examined expression levels of genes associated with viral infection in disease versus control samples. Amphiregulin (*AREG*), a ligand for epidermal growth factor receptor, is known to have essential roles in wound repair and inflammation resolution; furthermore, upregulation of *AREG* is associated with viral infections of the lung<sup>64</sup>. In COVID-19 patients, *AREG* is significantly upregulated in PBMCs<sup>65</sup>, monocytes, CD4 T Cells, NK cells, neutrophils, and DCs<sup>61</sup>, suggesting that upregulation of *AREG* may be an attempt to ameliorate the severe injury induced by SARS-CoV-2 infection. We observed reduced expression of *AREG* in the cDCs and macrophages, but not in the monocytes, in the CLD samples (Fig. 4c, Supplementary Dataset 3, Supplementary Fig. 12). *SOCS1*, a suppressor of cytokine signaling, was shown to reduce the type I IFN antiviral response in bronchial epithelial cells after influenza infection<sup>66,67</sup>.

Expression of the *S100A8/A9*, members of the *S100* family, and the IL6 co-receptor *IL6ST* was elevated in COVID-19 patients<sup>68-70</sup>. In our study, *S100A8/A9* expression is lower in the disease samples in cDCs, macrophages and monocytes while *SOCS1* expression is elevated in Other-ILD samples in NK Cells and pDCs (Fig. 4c, Supplementary Dataset 3). *IL6ST* expression level is elevated significantly in COPD and IPF but reduced dramatically in Other-ILD samples in macrophages ( $\log_2FC = -2.75$ ,  $q$  value =  $1.63e-61$ ) (Fig. 4c, Supplementary Dataset 3). These basal differences in inflammatory gene expression programs highlight how chronic lung disease alters the inflammatory microenvironment encountered upon viral exposure to the peripheral lung.

**Discussion**

The COVID-19 pandemic, caused by the SARS-CoV-2 virus, has affected tens of millions of individuals around the globe in just



**Fig. 5 Model of alterations in the diseased lung related to SARS-CoV2 pathogenesis.** (1) In the IPF lung, there is a proximalization of the distal airway. *ACE2* + epithelial cells cluster in the small airways though total *ACE2* + cell numbers are similar to control. (2) The viral entry score (accounting for all described putative receptors and proteases) is increased in diseased lungs. (3) Diseased epithelial cells have alterations in key SARS-CoV-2 response genes/pathways. (4) In the CLD lung, there is increased expression of cytotoxicity and exhaustion genes in immune cell populations and alterations in viral response pathways (interferon, antigen presentation). Figure created in Biorender.com.

the first nine months of 2020. Patients with CLD have an increased risk for severe SARS-CoV-2 infection: COPD patients have a five-fold increased risk of severe COVID-19<sup>23,24,71–73</sup> and ILD patients have up to a four-fold increased odds of death from COVID-19<sup>28,74</sup>. Here, we performed an integrated transcriptomic analysis of scRNA-seq data from healthy and CLD patients to identify potential molecular causative factors determining SARS-CoV-2 severity. To summarize the results (Fig. 5): (1) *ACE2* and *TMPRSS2* are expressed predominantly in epithelial cells and there are no significant differences in the number of *ACE2* + cells in all cell types in disease compared to control samples; (2) a viral entry score including multiple entry factors is increased in cells isolated from diseased lungs; (3) CLD epithelial cells, especially AT2 cells, exhibit pre-existing dysregulation of genes involved in viral infection and the immune response; (4) *ACE2* protein levels are elevated in the IPF small airway sections; (5) the CLD *ACE2* + cells differentially express genes related to SARS-CoV-2 infection compared to CLD *ACE2*- AT2 cells; (6) a unique *ACE2* correlated gene profile for each diagnosis group included antiviral and immune regulatory genes; (7) there are baseline differences in the cellular immune population in disease compared to control samples.

Similar to other coronaviruses, SARS-CoV-2 utilizes cellular receptors (*ACE2* and putatively, *BSG*, *NRP1* and *HSPA5* gene products) and priming proteases (*TMPRSS2*, *CTSL*, *FURIN*), for viral entry. These factors are expressed predominantly in the upper and lower airways, with *ACE2* being expressed highly in nasal goblet and ciliated cells and in a subset of AT2 cells and the absorptive enterocytes in the gut<sup>33,34,41,48,51</sup>. We observed a similar expression pattern of *ACE2* in our dataset, with AT2 cells having the highest number *ACE2* + cells. To our knowledge,

publications investigating baseline expression of these SARS-CoV-2 entry factors in lung disease have been limited to asthma and COPD with variable results. For example, studies in asthma patients showed elevated expression of *ACE2*, *TMPRSS2*, and *FURIN* in patients with severe but not mild-moderate asthma<sup>75,76</sup>. Leung et al. performed bulk RNAseq and immunohistochemical staining on bronchial epithelial cells and showed significantly elevated expression levels of *ACE2* and *ACE2* protein in the small airways of COPD patients compared to control<sup>77</sup>. Another study on bronchoscopically isolated tissue showed no relationship between disease status (mild to moderate asthma or COPD) on the expression levels of all SARS-CoV-2 entry factors<sup>42</sup>. Our study utilized scRNAseq technology to study gene expression at a very granular level and did not identify increased *ACE2* expression at the single-cell level in CLD, including COPD. However, in the IPF lung, there was a regional concentration of *ACE2* + cells in the small airways upon immunohistochemical examination (Fig. 3a, g), similar to the findings of Leung et al.<sup>77</sup>. While the overall frequency of *ACE2* + cells and the *ACE2* expression level may be low, changes in the proportional cellular makeup of the diseased lung epithelium may lead to a proportionate increase in *ACE2* + “infectable” cells in the distal lung. Importantly, IPF lungs exhibit abnormal expansion of epithelial cell programs, specifically the presence of more proximal specific cell types in the distal lungs<sup>31,32</sup>. Thus, our data along with previously published studies together suggest that while overall differences in *ACE2* expression and other entry factors may be minimal in CLD, the localization of susceptible cells in the distal lung may promote disease pathogenesis and severity. However, it seems clear that viral entry alone cannot explain the variation in disease severity between patients with and without CLD. In a



COVID-19 autopsy study, Dolorey *et al.* found that in these patients who succumbed to complications of infection, SARS-CoV-2 RNA + cells in the lung were largely in the myeloid lineage and did not overlap with entry factor expression<sup>78</sup>. Thus, once the infection has become established and significant cellular injury has taken place, viral entry factor expression may no longer be essential to continued propagation of injury.

A balanced immune response is crucial to viral clearance and avoidance of excessive injury to the host, as evidenced by poor outcomes related both to immunosuppression as well as hyperinflammation in COVID-19 patients<sup>79</sup>. COPD patients with severe COVID-19 had elevated serum levels of various inflammatory cytokines including IL-2R, IL-6, IL-8, IL-10, and TNF- $\alpha$  suggesting there may be global alterations in the immune response<sup>27</sup>. We observed that COPD AT2 cells expressed elevated levels of immune-response related genes (*CSF3*, *ITGB6*, *SOCS2*). G-CSF (encoded by *CSF3*) is found at high levels in patients with severe COVID-19 and thought to play a role in the hyperinflammatory syndrome while *SOCS2* is part of a negative feedback system that regulates the response to cytokines<sup>80,81</sup>. *ACE2* correlated genes in this cell population were enriched for regulators of the immune response (Fig. 2e), with several of these genes found to be upregulated in alveolosphere cultures infected with SARS-CoV-2<sup>36</sup>. In addition to alterations in the cytokine microenvironment, changes in cellular immune populations were also identified in the COPD samples, dysregulation of several genes in inflammatory pathways (*AREG*, *IL6ST*, *S100A8/A9*, *SOCS1*), and high levels of cytotoxic and exhaustion-related genes in CD4 and CD8 T Cells from COPD lungs. Expression of cytotoxic and exhaustion genes was increased compared to controls but similar in IPF and Other-ILD immune cell types. Together, our data suggest that the immune microenvironment at both the molecular and cellular level in the fibrotic and COPD lung is dysregulated and may promote severe infection and poor outcomes in COVID-19.

One limitation of our study is that we focus mainly on the peripheral regions of the lungs, and do not analyze cells in the upper airways or trachea. It is possible that there are significant differences in SARS-CoV-2 entry gene expression between disease and control samples in the more proximal regions of the lungs. Our study is also limited to the expression profiles of patients with CLD without SARS-CoV-2 infection, as samples from patients who are both infected with SARS-CoV-2 and have chronic lung disease are difficult to collect at present. In addition, scRNA-seq is inherently limited to analyses of gene expression which does not always correlate with protein levels. The net effect of many changes in gene expression levels is also difficult to predict and requires determination in experimental models. For example, our analysis demonstrated increased expression of viral entry restriction factors concurrently with an increased viral entry gene score in diseased epithelial cells. Furthermore, scRNA-seq cannot determine spatial relationships which would preclude analysis of cellular behaviors influenced by neighboring cells such as the priming of the viral spike protein by adjacent protease positive cells<sup>82</sup>. Nevertheless, given the inherent limitations in studying human biology, our study highlights crucial areas for future research into the pathogenesis of COVID-19 in patients with CLD including the dysregulation of genes related to viral replication and the innate immune response in epithelial cells, and basal differences in inflammatory cell gene expression programs.

## Methods

**scRNA-seq samples.** scRNA-seq data were obtained from published data with samples in the “VUMC/TGen” dataset from Habermann *et al.*<sup>32</sup> (GEO accession GSE135893), samples in the “Yale/BWH” dataset came from Adams *et al.*<sup>31</sup> (GEO

accession number GSE136831), samples in the “Pittsburgh” dataset from Morse *et al.*<sup>30</sup> (GEO accession GSE128033) and samples in the “Northwestern” dataset from Reyfman *et al.*<sup>29</sup> (GEO accession GSE122960) (Supplementary Tables 1, 2). For specific IRB review of each dataset, please refer to the original paper cited here. In addition, there are 39 unpublished scRNA-seq samples in the “VUMC/TGen” dataset that were collected under Vanderbilt IRB #s 060165, 171657 and Western IRB # 20181836.

**scRNA-seq data processing.** Seurat v4.0 package<sup>83</sup> was used to process the scRNA-seq data. Specifically, for the Pittsburgh, Northwestern datasets and 39 unpublished samples from the VUMC/TGen, the Cell Ranger (10X Genomics) output files were read into Seurat using the function *Read10X*, the remaining datasets were already in Seurat format and were loaded using the function *readRDS*. To eliminate low-quality/dying cells or empty droplets, we removed any cells containing fewer than 1000 genes or more than 25% mitochondrial genes. Due to the large size of the joint dataset, we performed *SCTransform*<sup>84</sup> for normalization and scaling of each dataset separately, split into four major cell populations using known markers: *EPCAM* + (Epithelial), *PECAM1* + *PTPRC* - (Endothelial), *PTPRC* + (Immune) and *EPCAM*- *PECAM*- *PTPRC*- (Mesenchymal). Each population from the four datasets was then merged together to generate four merged Seurat objects (Endothelial, Epithelial, Immune and Mesenchymal). Next, each object was SCTransformed with “dataset” being used as *batch\_var* to eliminate batch effects between datasets. Cell clustering was performed using the Seurat function *FindNeighbors* and *FindClusters* and cell type annotation was manually performed on each object using known cell-type specific markers (Supplementary Fig. 132). For each cell population, cell type annotation was performed at four levels, ranging from the most general to more granular annotation. Cells expressing more than one cell type specific marker were identified as doublets. After removing doublet cells, all four population datasets were merged to generate the final ILD object containing a total of 611,398 cells and 32 distinct cell types (Supplementary Table 3, Supplementary Fig. 1).

**Integrated analysis of joint dataset.** To calculate the percentage of single positive or double positive cells for *ACE2* and other cofactors, we counted the number of cells with >0 transcripts of corresponding genes. For double positive, cells have >0 transcripts of both genes of interest.

To assess the expression profile of SARS-CoV-2 mediators (*ACE2*, *BSG*, *NRP1*, *HSPA5*), the corresponding proteases (*TMPRSS2*, *CTSL*, *FURIN*) and other candidate genes involved in SARS-CoV-2 infection in different chronic lung disease subset (COPD, IPF or Other-ILD), we ran the function *FindMarkers* in Seurat package using the negative binomial test. Using Seurat function *CellCycleScoring*, we calculated the cell cycle state across all epithelial cells to ensure the cell cycle is not a factor contributing to the differential expression analysis (Supplementary Fig. 7). To account for batch effects, we used the parameter “*latent\_vars*” to incorporate the four variables (Age, Ethnicity, Smoking status and Dataset) into the negative binomial model. For the binary heatmap, the differential expression analysis was performed between the Disease (including all chronic disease subset) and Control samples. Then, log<sub>2</sub>-fold-change was converted into 0 (downregulated in the disease samples) or 1 (upregulated in the disease samples) regardless of the Bonferroni adjusted *p* values. Heatmaps were generated from the adjusted log<sub>2</sub>FC values using the *heatmap.2* function of the *gplots* R package<sup>85</sup>. For the boxplots, count numbers of selected genes were plotted using the *geom\_boxplot* and *geom\_jitter* function of the *ggplot2* R package<sup>86</sup>.

**Gene module score.** To calculate the combined expression of genes, we used the *AddModuleScore* in Seurat v3.1.5 package. SARS-CoV-2 entry gene scores were calculated on SARS-CoV-2 receptors and mediators: *ACE2*, *BSG* (*CD147*), *NRP1*, *HSPA5*(*GRP78*), *TMPRSS2*, *CTSL*, *FURIN* and *ADAM17*. Viral entry restriction ISGs: *LY6E*, *CLEC4D*, *UBD*, *ELF1*, *FAM46C*, *REC8*<sup>43</sup>. Viral replication inhibition ISGs: *SPATS2L*, *ZBP1*, *DNAJC6*, *IFIT3*, *RGS22*, *IFIT1*, *IFIT5*, *B4GALT5*<sup>43</sup>. HLA type II score includes *HLA-DRA*, *HLA-DQA1*, *HLA-DQA2*, *HLA-DPA1*, *HLA-DRB1*, *HLA-DPB1*, *HLA-DQB2*, *HLA-DRB5*, *HLA-DQB1*, *HLA-DMA*, *HLA-DMB*. IFN score includes *ISG15*, *IFI44*, *IFI27*, *CXCL10*, *RSAD2*, *IFIT1*, *IFI44L*, *CCL8*, *XAF1*, *GBP1*, *IRF7*, *CEACAM1*. IL6 scores were calculated on six tocilizumab responsive genes: *ARID5A*, *SOCS3*, *PIM1*, *BCL3*, *BATF*, *MYC* that are associated with the IL-6 pathway<sup>61</sup>. Cytotoxicity associated genes include *PRF1*, *GZMH*, *IFNG*, *NGK7*, *KLRG1*, *PRF1*<sup>61</sup> and *GNL1*, *GZMB*, *GZMK*<sup>87</sup>. Exhaustion gene set: *LAG3*, *TIGIT*, *PDCD1*, *CTLA4*, *HAVCR2*, *TOX*<sup>63</sup>, and *PRDM1*, *MAF*<sup>61</sup>. Significant differences between different groups were calculated using the Tukey\_HSD statistic test in the R package *rstatix* with a confidence level of 0.95 (Supplementary Dataset 2).

**Gene correlation analysis.** To identify genes that correlate with *ACE2* in the AT2 *ACE2* + cells, we performed Spearman correlation coefficient analysis on the log-transformed and normalized data using the function *cor.test* in the R stats v3.6.1 package with Benjamini-Hochberg corrections for *p*-adjusted values. Gene ontology analysis for the significant correlated genes (*p* value  $\leq$  0.03 and 99th percentile rho) was performed with the Bioconductor R package TopGO version 2.42 and the Bioconductor annotation data package *org.Hs.eg.db* version 3.12.0;

Kolmogorov–Smirnov (KS) statistic method with TopGO default algorithm weight01 test was used for GO term enrichment test<sup>88</sup>, and Benjamini Hochberg (BH) adjusted *p* values (*q* values or FDR) were computed using the R function *p.adj*.

**Immunohistochemistry of ACE2 and anti- $\alpha$ v $\beta$ 6 integrin.** Formalin-fixed paraffin-embedded histological sections of human lung were cut at 5-microns and dewaxed in xylene prior to rehydration in decreasing concentrations of ethanol. The tissue samples were obtained after informed consent and local ethics approval (South East Scotland SAHSC Bioresource-reference number 06/S1101/41; Brompton Node samples—reference number 15/SC/0101; Papworth Node Samples; non-diseased controls- reference number (Q)GM030404 and Nottingham BRC samples- reference number 08/H0407/1). IHC staining was performed using the Novocastra Novolink™ Polymer Detection Systems kit (Code: RE7280-K, Leica, Biosystems, Newcastle, UK) as previously described<sup>89</sup>. Heat-induced citrate antigen retrieval (pH 6.0) and pepsin antigen retrieval was performed for Rabbit monoclonal ACE2 (ab108252, EPR4435(2) Abcam, UK) and the anti- $\alpha$ v $\beta$ 6 integrin antibody (6.2A1; Biogen, Cambridge, MA, USA), respectively. Rabbit monoclonal ACE2 (1:400) and anti- $\alpha$ v $\beta$ 6 integrin (1:3000) was diluted in Leica antibody diluent (RE AR9352, Leica, Biosystems, UK) and incubated with the sections overnight at 4 °C. Novolink DAB substrate buffer plus was used as the chromogen and the slides were counterstained with Novolink haematoxylin for 6 min, dehydrated and cover slipped. A negative control without the application of the primary antibody, and was also used to ensure staining was only related to the presence of the antibody.

The immunohistochemically stained slides were scanned using a ScanScope XT Slide Scanner (Leica Aperio Technologies, Vista, CA, USA) under 20 $\times$  objective magnification (0.5  $\mu$ m resolution) using Panoramic Viewer (3DHISTECH Ltd Budapest, Hungary) slide viewing software. Both the percentage of staining and staining intensity of ACE2 expression in lung sections were individually assessed. For ACE2 quantification, the following scoring system of five high-power fields at X40. per tissue section were used:

The coding was performed prior to scoring/analysis as: 0- Negative; 1- 0– $\leq$ 10%; 2- 11– $\leq$ 25%; 3-  $\leq$ 26%. Statistical analyses were completed using GraphPad Prism 7.0 (GraphPad Software, San Diego, CA, USA). One-way analysis of variance was used for comparison of more than two datasets and significant differences between diagnosis groups were calculated using the Tukey HSD test.

**ACE2 western blot.** Cell protein was isolated using Cell Lysis Buffer (Cell signalling, USA) supplemented with protease inhibitor cocktail (Sigma, USA) and the quantification performed using BCA Protein Assay Kit (ThermoFisher Scientific, UK). Western blotting was performed using 4–12%, pre-cast Bis-Tris gradient gels (ThermoFisher Scientific, UK) and 25  $\mu$ g of protein was loaded per lane. Immunoblots were incubated with anti-ACE2 (ab108252; Rabbit monoclonal-Abcam-EPR4435(2)—1:500 dilution of stock antibody) diluted in 5% skim milk/Tris buffered saline with 0.1% Tween-20 for overnight at 4 °C. A loading control of GAPDH was also used to demonstrate protein loading (ab8245; Mouse monoclonal-anti-GAPDH antibody [6C5] at 1:10000 dilution of stock antibody). Following day immunoblots were incubated with an anti-mouse-HRP and anti-rabbit-HRP conjugated secondary antibodies (Dako, USA) at 1:2500 for 1 hr at room temperature. Visualization was performed with Clarity Max™ ECL Substrate (Biorad, UK) on a Licor C-DiGit. For more information on the anti-ACE2 antibody, please refer to the manufacturer's datasheet here: <https://www.abcam.com/ace2-antibody-epr44352-ab108252.html>. Two replicates were performed for the western blot.

**Statistical analysis.** Tukey Honest Significant Difference (Tukey\_HSD) statistical test from the R package *rstatix* with a confidence level of 0.95 was used to test statistical dependence of cells expressing the SARS-CoV-2 mediators among chronic disease subsets. Tukey\_HSD test was also used to test significant difference in gene expression module score, quantification of cell types and the ACE2 protein expression quantification. Significant differences in gene expression were the Bonferroni adjusted *p*-values calculated from the FindMarkers function between Control and Disease groups (COPD, IPF, Other ILD) using the fitted negative binomial model and latent\_vars parameters as described above. Significance in gene correlation analysis between ACE2 and other SARS-CoV-2 entry factors (Fig. 3a and Supplementary Fig. 10) was calculated using Anova.

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

## Data availability

The majority of the data used in this manuscript are publicly available from published papers: GEO accession “GSE135893”<sup>32</sup>, GEO accession “GSE136831”<sup>31</sup>, GEO accession “GSE128033”<sup>30</sup> and GEO accession “GSE122960”<sup>29</sup>. The unpublished data from VUMC/TGen (39 samples) are included in the supplementary data (Supplementary Dataset 1) as a count matrix format containing all the genes being used in the manuscript. All other relevant data supporting the key findings of this study are available within the article and its Supplementary Information files or from the corresponding author upon reasonable

request. Source data are provided with this paper. A reporting summary for this article is available as a Supplementary Information file. Source data are provided with this paper.

## Code availability

The code for genomic analyses in this paper is available at [https://github.com/tgen/banovichlab/Disease\\_lung\\_COVID19\\_2020/](https://github.com/tgen/banovichlab/Disease_lung_COVID19_2020/).

Received: 1 December 2020; Accepted: 4 June 2021;

Published online: 14 July 2021

## References

- Huang, C. et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* **395**, 497–506 (2020).
- Zhu, N. et al. A Novel Coronavirus from Patients with Pneumonia in China, 2019. *N. Engl. J. Med.* **382**, 727–733 (2020).
- Wang, C., Horby, P. W., Hayden, F. G. & Gao, G. F. A novel coronavirus outbreak of global health concern. *Lancet* **395**, 470–473 (2020).
- Zhou, P. et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **579**, 270–273 (2020).
- Home - Johns Hopkins Coronavirus resource center. <https://coronavirus.jhu.edu/>.
- Marini, J. J. & Gattinoni, L. Management of COVID-19 Respiratory Distress. *JAMA* (2020) <https://doi.org/10.1001/jama.2020.6825>.
- Murthy, S., Gomersall, C. D. & Fowler, R. A. Care for critically ill patients with COVID-19. *JAMA* (2020) <https://doi.org/10.1001/jama.2020.3633>.
- Bhatraju, P. K. et al. Covid-19 in critically ill patients in the seattle region—case series. *N. Engl. J. Med.* **382**, 2012–2022 (2020).
- Wu, C. et al. Risk factors associated with acute respiratory distress syndrome and death in patients with Coronavirus disease 2019 pneumonia in Wuhan, China. *JAMA Intern. Med.* (2020) <https://doi.org/10.1001/jamainternmed.2020.0994>.
- Zhou, F. et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *Lancet* **395**, 1054–1062 (2020).
- Fan, E. et al. COVID-19-associated acute respiratory distress syndrome: is a different approach to management warranted? *Lancet Respir. Med.* **8**, 816–821 (2020).
- Gattinoni, L. et al. COVID-19 does not lead to a “typical” acute respiratory distress syndrome. *Am. J. Respir. Crit. Care Med.* **201**, 1299–1300 (2020).
- Zheng, Z. et al. Risk factors of critical & mortal COVID-19 cases: a systematic literature review and meta-analysis. *J. Infect.* **81**, e16–e25 (2020).
- Williamson, E. J. et al. Factors associated with COVID-19-related death using OpenSAFELY. *Nature* **584**, 430–436 (2020).
- Chen, L. et al. Risk factors for death in 1859 subjects with COVID-19. *Leukemia* **34**, 2173–2183 (2020).
- Wortham, J. M. Characteristics of persons who died with COVID-19—United States, February 12–May 18, 2020. *Morb. Mortal. Wkly. Rep.* **69**, 923–929 (2020).
- Dowd, J. B. et al. Demographic science aids in understanding the spread and fatality rates of COVID-19. *Proc. Natl Acad. Sci. U.S.A.* **117**, 9696–9698 (2020).
- Richardson, S. et al. Presenting characteristics, comorbidities, and outcomes among 5700 patients hospitalized with COVID-19 in the New York City Area. *JAMA* (2020) <https://doi.org/10.1001/jama.2020.6775>.
- Feng, Y. et al. COVID-19 with different severities: a multicenter study of clinical features. *Am. J. Respir. Crit. Care Med.* **201**, 1380–1388 (2020).
- CDC. Coronavirus Disease 2019 (COVID-19). <https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/people-with-medical-conditions.html> (2020).
- Southern, B. D. Patients with interstitial lung disease and pulmonary sarcoidosis are at high risk for severe illness related to COVID-19. *Cleve. Clin. J. Med.* <https://doi.org/10.3949/ccjm.87a.ccc026> (2020).
- George, P. M., Wells, A. U. & Jenkins, R. G. Pulmonary fibrosis and COVID-19: the potential role for antifibrotic therapy. *Lancet Respir. Med.* **8**, 807–815 (2020).
- Lippi, G. & Henry, B. M. Chronic obstructive pulmonary disease is associated with severe coronavirus disease 2019 (COVID-19). *Respir. Med.* **167**, 105941 (2020).
- Alqahtani, J. S. et al. Prevalence, severity and mortality associated with COPD and smoking in patients with COVID-19: a rapid systematic review and meta-analysis. *PLoS One* **15**, e0233147 (2020).
- Halpin, D. M. G., Faner, R., Sibila, O., Badia, J. R. & Agusti, A. Do chronic respiratory diseases or their treatment affect the risk of SARS-CoV-2 infection? *Lancet Respir. Med.* **8**, 436–438 (2020).

26. Leung, J. M., Niikura, M., Yang, C. W. T. & Sin, D. D. COVID-19 and COPD. *Eur. Respir. J.* **56**, 2002018 (2020).
27. Song, J. et al. Distinct effects of asthma and COPD comorbidity on disease expression and outcome in patients with COVID-19. *Allergy* (2020) <https://doi.org/10.1111/all.14517>.
28. Jenkins, G. et al. Outcome of hospitalisation for COVID-19 in patients with Interstitial Lung Disease: an international multicentre study. *Respir. Med.* <https://doi.org/10.1101/2020.07.15.20152967> (2020).
29. Reyfman, P. A. et al. Single-cell transcriptomic analysis of human lung provides insights into the pathobiology of pulmonary fibrosis. *Am. J. Respir. Crit. Care Med.* **199**, 1517–1536 (2019).
30. Morse, C. et al. Proliferating SPP1/MERTK-expressing macrophages in idiopathic pulmonary fibrosis. *Eur. Respir. J.* **54**, 1802441 (2019).
31. Adams, T. S. et al. Single-cell RNA-seq reveals ectopic and aberrant lung-resident cell populations in idiopathic pulmonary fibrosis. *Sci. Adv.* **6**, eaba1983 (2020).
32. Habermann, A. C. et al. Single-cell RNA sequencing reveals profibrotic roles of distinct epithelial and mesenchymal lineages in pulmonary fibrosis. *Sci. Adv.* **6**, eaba1972 (2020).
33. Sunjnak, W. et al. SARS-CoV-2 entry factors are highly expressed in nasal epithelial cells together with innate immune genes. *Nat. Med.* **26**, 681–687 (2020).
34. Muus, C. et al. Single-cell meta-analysis of SARS-CoV-2 entry genes across tissues and demographics. *Nat. Med.* **27**, 546–559 (2021).
35. Cantuti-Castelvetri, L. et al. Neuropilin-1 facilitates SARS-CoV-2 cell entry and infectivity. *Science* **370**, 856–860 (2020).
36. Katsura, H. et al. Human lung stem cell-based alveolospheres provide insights into SARS-CoV-2-mediated interferon responses and pneumocyte dysfunction. *Cell Stem Cell* <https://doi.org/10.1016/j.stem.2020.10.005> (2020).
37. Coutard, B. et al. The spike glycoprotein of the new coronavirus 2019-nCoV contains a furin-like cleavage site absent in CoV of the same clade. *Antivir. Res.* **176**, 104742 (2020).
38. Hoffmann, M. et al. SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. *Cell* **181**, 271–280.e8 (2020).
39. Ibrahim, I. M., Abdelmalek, D. H., Elshahat, M. E. & Elfiky, A. A. COVID-19 spike-host cell receptor GRP78 binding site prediction. *J. Infect.* **80**, 554–562 (2020).
40. Aguiar, J. A. et al. Gene expression and in situ protein profiling of candidate SARS-CoV-2 receptors in human airway epithelial cells and lung tissue. *Eur. Respir. J.* (2020) <https://doi.org/10.1183/13993003.01123-2020>.
41. Ziegler, C. G. K. et al. SARS-CoV-2 receptor ACE2 is an interferon-stimulated gene in human airway epithelial cells and is detected in specific cell subsets across tissues. *Cell* **181**, 1016–1035.e19 (2020).
42. Aliee, H. et al. Determinants of SARS-CoV-2 receptor gene expression in upper and lower airways. *medRxiv* (2020).
43. Martin-Sancho, L. et al. Functional landscape of SARS-CoV-2 cellular restriction. *Mol. Cell* (2021) <https://doi.org/10.1016/j.molcel.2021.04.008>.
44. Chen, J.-H. et al. Plasma proteome of severe acute respiratory syndrome analyzed by two-dimensional gel electrophoresis and mass spectrometry. *Proc. Natl Acad. Sci. U.S.A.* **101**, 17039–17044 (2004).
45. Yang, N. & Shen, H.-M. Targeting the endocytic pathway and autophagy process as a novel therapeutic strategy in COVID-19. *Int. J. Biol. Sci.* **16**, 1724–1731 (2020).
46. So, J.-S. Roles of endoplasmic reticulum stress in immune responses. *Mol. Cells* **41**, 705 (2018).
47. Chan, C.-P. et al. Modulation of the unfolded protein response by the severe acute respiratory syndrome coronavirus spike protein. *J. Virol.* **80**, 9279–9287 (2006).
48. Hikmet, F. et al. The protein expression profile of ACE2 in human tissues. *Mol Syst Biol.* **16**, e9610 (2020).
49. Jenkins, G. Demystifying pulmonary fibrosis. *Am. J. Physiol. Lung Cell. Mol. Physiol.* **319**, L554–L559 (2020).
50. Meliopoulou, V. A. et al. An epithelial integrin regulates the amplitude of protective lung interferon responses against multiple respiratory pathogens. *PLoS Pathog.* **12**, e1005804 (2016).
51. Singh, M., Bansal, V. & Feschotte, C. A single-cell RNA expression map of human coronavirus entry factors. *Cell Rep.* **32**, 108175 (2020).
52. Meecham, A. & Marshall, J. F. The ITGB6 gene: its role in experimental and clinical biology. *Gene X* **5**, 100023 (2020).
53. Pfaender, S. et al. LY6E impairs coronavirus fusion and confers immune control of viral disease. *Nat Microbiol* **5**, 1330–1339 (2020).
54. Han, G. & Sinjab Single-cell analysis of human lung epithelia reveals concomitant expression of the SARS-CoV-2 receptor ACE2 with multiple virus receptors and scavengers in .... *Cancers* **13**, 1250 (2021).
55. Zarei, M. et al. Potential role of glycoprotein 340 in milder SARS-CoV-2 infection in children. *Expert Rev. Anti. Infect. Ther.* **19**, 675–677 (2021).
56. Overmyer, K. A. et al. Large-scale multi-omic analysis of COVID-19 severity. *Cell Syst.* **12**, 23–40.e7 (2021).
57. Chetta, M., Rosati, A., Marzullo, L., Tarsitano, M. & Bukvic, N. A SARS-CoV-2 host infection model network based on genomic human Transcription Factors (TFs) depletion. *Heliyon* **6**, e05010 (2020).
58. Lucas, C. et al. Longitudinal analyses reveal immunological misfiring in severe COVID-19. *Nature* **584**, 463–469 (2020).
59. Valle, D. M. D. et al. An inflammatory cytokine signature predicts COVID-19 severity and survival. *Nat. Med.* (2020) <https://doi.org/10.1038/s41591-020-1051-9>.
60. Giamarellos-Bourboulis, E. J. et al. Complex immune dysregulation in COVID-19 patients with severe respiratory failure. *Cell Host Microbe* **27**, 992–1000.e3 (2020).
61. Unterman, A. et al. Single-cell omics reveals dyssynchrony of the innate and adaptive immune system in progressive COVID-19. *medRxiv* (2020).
62. Zheng, H.-Y. et al. Elevated exhaustion levels and reduced functional diversity of T cells in peripheral blood may predict severe progression in COVID-19 patients. *Cell. Mol. Immunol.* **17**, 541–543 (2020).
63. Zhang, J.-Y. et al. Single-cell landscape of immunological responses in COVID-19 patients. *Nat. Imm* **21**, 1107–1118 (2020).
64. Zaiss, D. M. W., Gause, W. C., Osborne, L. C. & Artis, D. Emerging functions of amphiregulin in orchestrating immunity, inflammation, and tissue repair. *Immunity* **42**, 216–226 (2015).
65. Xiong, Y. et al. Transcriptomic characteristics of bronchoalveolar lavage fluid and peripheral blood mononuclear cells in COVID-19 patients. *Emerg. Microbes Infect.* **9**, 761–770 (2020).
66. Pothlichet, J., Chignard, M. & Si-Tahar, M. Cutting edge: innate immune response triggered by influenza A virus is negatively regulated by SOCS1 and SOCS3 through a RIG-I/IFNAR1-dependent pathway. *J. Immunol.* **180**, 2034–2038 (2008).
67. Sallenave, J.-M. & Guillot, L. Innate immune signaling and proteolytic pathways in the resolution or exacerbation of SARS-CoV-2 in Covid-19: key therapeutic targets? *Front. Immunol.* **11**, 1229 (2020).
68. Coperchini, F., Chiovato, L., Croce, L., Magri, F. & Rotondi, M. The cytokine storm in COVID-19: an overview of the involvement of the chemokine/chemokine-receptor system. *Cytokine Growth Factor Rev.* **53**, 25–32 (2020).
69. Zhu, L. et al. Single-cell sequencing of peripheral mononuclear cells reveals distinct immune response landscapes of Covid-19 and influenza patients. *Immunity* **53**, 685–696.e3 (2020).
70. Chen, L. et al. Elevated serum levels of S100A8/A9 and HMGB1 at hospital admission are correlated with inferior clinical outcomes in COVID-19 patients. *Cell. Mol. Immunol.* **17**, 992–994 (2020).
71. Banerjee, A. et al. Estimating excess 1-year mortality associated with the COVID-19 pandemic according to underlying conditions and age: a population-based cohort study. *Lancet* **395**, 1715–1725 (2020).
72. Docherty, A. B. et al. Features of 20 133 UK patients in hospital with covid-19 using the ISARIC WHO clinical characterisation protocol: prospective observational cohort study. *BMJ* **369**, m1985 (2020).
73. Guan, W.-J. et al. Clinical characteristics of Coronavirus disease 2019 in China. *N. Engl. J. Med.* **382**, 1708–1720 (2020).
74. Esposito, A. J. et al. Increased odds of death for patients with interstitial lung disease and COVID-19: a case-control study. *Am. J. Respir. Crit. Care Med.* (2020) <https://doi.org/10.1164/rccm.202006-2441LE>.
75. Kermani, N. et al. Sputum ACE2, TMPRSS2 and FURIN gene expression in severe neutrophilic asthma. *Respir Res* **22**, 10 (2021).
76. Camiolo, M., Gauthier, M., Kaminski, N., Ray, A. & Wenzel, S. E. Expression of SARS-CoV-2 receptor ACE2 and coincident host response signature varies by asthma inflammatory phenotype. *J. Allergy Clin. Immunol.* **146**, 315–324.e7 (2020).
77. Leung, J. M. et al. ACE-2 expression in the small airway epithelia of smokers and COPD patients: implications for COVID-19. *Eur. Respir. J.* **55**, 2000688 (2020).
78. Delorey, T. M. et al. A single-cell and spatial atlas of autopsy tissues reveals pathology and cellular targets of SARS-CoV-2. *Cold Spring Harbor Lab.* 2021.02.25.430130 (2021) <https://doi.org/10.1101/2021.02.25.430130>.
79. Fung, M. & Babik, J. M. COVID-19 in immunocompromised hosts: what we know so far. *Clin. Infect. Dis.* (2020) <https://doi.org/10.1093/cid/ciaa863>.
80. Lu, C., Li, S. & Liu, Y. Role of immunosuppressive therapy in rheumatic diseases concurrent with COVID-19. *Ann. Rheum. Dis.* **79**, 737–739 (2020).
81. Yoshimura, A., Naka, T. & Kubo, M. SOCS proteins, cytokine signalling and immune regulation. *Nat. Rev. Immunol.* **7**, 454–465 (2007).
82. Wei, J. et al. Genome-wide CRISPR screens reveal host factors critical for SARS-CoV-2 infection. *Cell* **184**, 76–91.e13 (2021).
83. Stuart, T. et al. Comprehensive integration of single cell data. *Cell* **177**, 1888–1902 (2019).



84. Hafemeister, C. & Satija, R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol.* **20**, 296 (2019).
85. Warnes, G. R. et al. gplots: various R programming tools for plotting data. *R package version 2*, 1 (2009).
86. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*. (Springer, 2016).
87. Szabo, P. A. et al. Single-cell transcriptomics of human T cells reveals tissue and activation signatures in health and disease. *Nat. Commun.* **10**, 4706 (2019).
88. Alexa, A., Rahnenführer, J. & Lengauer, T. Improved scoring of functional groups from gene expression data by decorrelating GO graph structure. *Bioinformatics* **22**, 1600–1607 (2006).
89. Joseph, C. et al. Elevated MMP9 expression in breast cancer is a predictor of shorter patient survival. *Breast Cancer Res. Treat.* **182**, 267–282 (2020).

### Acknowledgements

This study was supported by the NIH/NHLBI R01HL145372 (NEB/JAK), the Department of Defense W81XWH1910415 (NEB/JAK), Doris Duke Charitable Foundation (J. A.K.), T32HL094296 (NIW, JBB), the Department of Veterans Affairs IK2BX003841 (BWR), DoD W81XWH-19-1-0131 (J.C.S.), R01HL127349 (N.K.), R01HL141852 (N.K.), U01HL145567 (N.K.), UH2 HL123886 (N.K.), and a generous gift from Three Lakes Partners to N.K. and I.O.R. The integrated data sets were funded by various sponsors as indicated in the original publications. R.G.J. is funded by and NIHR Research Professorship (RP-2017-08-ST2-014).

### Author contributions

L.T.B., N.I.W., M.-I.C., N.E.B., and J.A.K. conceived and designed the analysis. Sample collection was performed by A.C.H., T.S.A., J.C.S., S.P., L.M.P., C.J.T., J.B.B., B.W.R., A.G. N., D.R., W.A.W., R.G.J., A.J.G., L.T.B., N.E.B., and J.A.K. performed quality checks, data integration, and computational analyses. L.T.B., N.I.W., N.E.B., and J.A.K. analyzed and interpreted scRNA-seq data. C.J. performed the immunohistology and semi-quantification analysis. L.T.B., N.I.W., M.-I.C., N.E.B., and J.A.K. wrote and revised the manuscript, with significant input from I.O.R., R.G.J., N.K., and the HCA Lung Biological Network. All authors read and approved the manuscript before submission.

### Competing interests

J.A.K. has received advisory board fees from Boehringer Ingelheim, Inc, Janssen Pharmaceuticals, is on the scientific advisory board of APIE Therapeutics, and has research contracts with Genentech. In the last 36 months, N.K. reported personal fees from Biogen Idec, Boehringer Ingelheim, Third Rock, Samumed, Numedii, AstraZeneca, Life Max, Teravance, RohBar, and Pliant and Equity in Pliant; collaboration with MiRagen, AstraZeneca; Grant from VeracYTE, all outside the submitted work. In addition, N.K. has

a patent for New Therapies in Pulmonary Fibrosis, and Peripheral Blood Gene Expression licensed to Biotech. A.G.N. has received advisory board fees from Boehringer Ingelheim, Galapagos, Medical Quantitative Image Analysis and personal fees for educational material from Up to Date and Boehringer Ingelheim. RGJ reports grants from AstraZeneca, grants from Biogen, personal fees from Boehringer Ingelheim, personal fees from Chiesi, personal fees Daewoong, personal fees from Galapagos, grants from Galeco, grants from GlaxoSmithKline, personal fees from Heptares, nonfinancial support from NuMedii, grants and personal fees from Pliant, personal fees from Promedior, non-financial support from Redx, personal fees from Roche, other from Action for Pulmonary Fibrosis, outside the submitted work. All other authors declare no competing interests.

### Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-021-24467-0>.

**Correspondence** and requests for materials should be addressed to N.E.B.

**Peer review information** *Nature Communications* thanks Manvendra Singh and the other anonymous reviewer(s) for their contribution to the peer review of this work.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021

## Human Cell Atlas Lung Biological Network

Alexander V. Misharin<sup>14</sup>, Alexander M. Tsankov<sup>15</sup>, Avrum Spira<sup>16,17</sup>, Pascal Barbry<sup>18</sup>, Alvis Brazma<sup>19</sup>, Christos Samakovlis<sup>20,21</sup>, Douglas P. Shepherd<sup>22</sup>, Emma L. Rawlins<sup>23,24</sup>, Fabian J. Theis<sup>25,26</sup>, Jennifer Griffonnet<sup>27</sup>, Haeock Lee<sup>28</sup>, Herbert B. Schiller<sup>29,30</sup>, Paul Hofman<sup>31,32,33</sup>, Joseph E. Powell<sup>34,35</sup>, Joachim L. Schultze<sup>36,37</sup>, Jeffrey Whitsett<sup>38</sup>, Jiyeon Choi<sup>39</sup>, Joakim Lundeberg<sup>40</sup>, Naftali Kaminski, Jonathan A. Kropski, Nicholas E. Banovich, Jose Ordovas-Montanes<sup>41,42,43,44</sup>, Jayaraj Rajagopal<sup>43</sup>, Kerstin B. Meyer<sup>45</sup>, Mark A. Krasnow<sup>46</sup>, Kourosh Saeb-Parsy<sup>47</sup>, Kun Zhang<sup>48</sup>, Robert Lafyatis<sup>49</sup>, Sylvie Leroy<sup>50,51</sup>, Muzlifah Haniffa<sup>52,53,54</sup>, Martijn C. Nawijn<sup>55</sup>, Marko Z. Nikolic<sup>56</sup>, Maarten van den Berge<sup>57,58</sup>, Malte Kuhnemund<sup>59</sup>, Charles-Hugo Marquette<sup>60,61</sup>, Michael Von Papen<sup>62</sup>, Oliver Eickelberg<sup>63</sup>, Orit Rosenblatt-Rosen<sup>64</sup>, Paul A. Reyfman<sup>65</sup>, Dana Pe'er<sup>66</sup>, Peter Horvath<sup>67,68</sup>, Purushothama Rao Tata<sup>69</sup>, Aviv Regev<sup>64,70,71</sup>, Mauricio Rojas<sup>72</sup>, Max A. Seibold<sup>73</sup>, Alex K. Shalek<sup>74,75,76</sup>, Jason R. Spence<sup>77,78,79</sup>, Sarah A. Teichmann<sup>45,80</sup>, Stephen Quake<sup>81</sup>, Thu Elizabeth Duong<sup>82</sup>, Tommaso Biancalani<sup>41</sup>, Tushar Desai<sup>83</sup>, Xin Sun<sup>84,85</sup> & Laure Emmanuelle Zaragosi<sup>86</sup>

<sup>14</sup>Division of Pulmonary and Critical Care Medicine, Northwestern University, Chicago, IL, USA. <sup>15</sup>Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA. <sup>16</sup>Department of Medicine, Boston University School of Medicine, Boston, MA, USA. <sup>17</sup>Johnson & Johnson Innovation, Cambridge, MA, USA. <sup>18</sup>Université Côte d'Azur, CNRS, IPMC, Sophia-Antipolis, France. <sup>19</sup>European Molecular Biology



Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Trust Genome Campus, Hinxton, Cambridge, UK. <sup>20</sup>SciLifeLab, Department of Molecular Biosciences, Stockholm University, Stockholm, Sweden. <sup>21</sup>Cardiopulmonary Institute, Justus Liebig University, Giessen, Germany. <sup>22</sup>Center for Biological Physics and Department of Physics, Arizona State University, Tempe, AZ, USA. <sup>23</sup>Wellcome Trust/CRUK Gurdon Institute, Cambridge, UK. <sup>24</sup>Department Physiology, Development and Neuroscience, University of Cambridge, Cambridge, UK. <sup>25</sup>Institute of Computational Biology, Helmholtz Zentrum München, Oberschleißheim, Germany. <sup>26</sup>Departments of Mathematics and Life Sciences, Technical University Munich, Munich, Germany. <sup>27</sup>Pneumology Department, Nice University-Affiliated Hospital, Nice, France. <sup>28</sup>Department of Biomedicine and Health Sciences, The Catholic University of Korea, Seoul, Korea. <sup>29</sup>Comprehensive Pneumology Center (CPC)/Institute of Lung Biology and Disease (ILBD), Helmholtz Zentrum München, Oberschleißheim, Germany. <sup>30</sup>Member of the German Center for Lung Research (DZL), Munich, Germany. <sup>31</sup>Laboratory of Clinical and Experimental Pathology, Pasteur Hospital, University Côte d'Azur, Nice, France. <sup>32</sup>Hospital-Related Biobank, Pasteur Hospital, University Côte d'Azur, Nice, France. <sup>33</sup>FHU OncoAge, Pasteur Hospital, BP69, Nice cedex 01, France. <sup>34</sup>Garvan-Weizmann Centre for Cellular Genomics, Garvan Institute of Medical Research, Sydney, NSW, Australia. <sup>35</sup>UNSW Cellular Genomics Futures Institute, University of New South Wales, Sydney, NSW, Australia. <sup>36</sup>Department for Genomics & Immunoregulation, LIMES-Institute, University of Bonn, Bonn, Germany. <sup>37</sup>PRECISE Platform for Single Cell Genomics & Epigenomics, Germany Center for Neurodegenerative Diseases and University of Bonn, Bonn, Germany. <sup>38</sup>Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA. <sup>39</sup>Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD, USA. <sup>40</sup>SciLifeLab, Department of Gene Technology, KTH Royal Institute of Technology, Stockholm, Sweden. <sup>41</sup>Broad Institute of MIT and Harvard, Cambridge, MA, USA. <sup>42</sup>Division of Gastroenterology, Hepatology, and Nutrition, Boston Children's Hospital, Boston, MA, USA. <sup>43</sup>Harvard Stem Cell Institute, Cambridge, MA, USA. <sup>44</sup>Program in Immunology, Harvard Medical School, Boston, MA, USA. <sup>45</sup>Cellular Genetics Programme, Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge, UK. <sup>46</sup>Department of Biochemistry and Wall Center for Pulmonary Vascular Disease, Stanford University, Stanford, CA, USA. <sup>47</sup>Department of Surgery, University of Cambridge and NIHR Cambridge Biomedical Research Centre, Cambridge, UK. <sup>48</sup>Department of Bioengineering, UCSD, La Jolla, CA, USA. <sup>49</sup>Division of Rheumatology, Department of Medicine, University of Pittsburgh Medical Center, Pittsburgh, PA, USA. <sup>50</sup>Department of Pulmonary Medicine and Allergology, Université Côte d'Azur, CHU de Nice, FHU OncoAge, Nice, France. <sup>51</sup>CNRS UMR 7275—Institut de Pharmacologie Moléculaire et Cellulaire, Sophia Antipolis, France. <sup>52</sup>Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge, UK. <sup>53</sup>Biosciences Institute, Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne, UK. <sup>54</sup>Department of Dermatology and NIHR Newcastle Biomedical Research Centre, Newcastle Hospitals NHS Foundation Trust, Newcastle upon Tyne, UK. <sup>55</sup>Department of Pathology and Medical Biology, University of Groningen, GRIAC Research Institute, University Medical Center Groningen, Groningen, the Netherlands. <sup>56</sup>UCL Respiratory, Division of Medicine, University College London, London, UK. <sup>57</sup>Department of Pulmonary Diseases, University Medical Center Groningen, Groningen, the Netherlands. <sup>58</sup>Groningen Research Institute for Asthma and COPD, University of Groningen, Groningen, the Netherlands. <sup>59</sup>Cartana AB, Stockholm, Sweden. <sup>60</sup>Team 4, IRCAN, FHU OncoAge, University Côte d'Azur, CNRS, INSERM, Nice CEDEX 02, France. <sup>61</sup>Department of Pneumology and Oncology, CHU Nice, FHU OncoAge, University Côte d'Azur, Nice, France. <sup>62</sup>Comma Soft AG, Bonn, Germany. <sup>63</sup>Department of Medicine, University of Pittsburgh, Pittsburgh, PA, USA. <sup>64</sup>Klarman Cell Observatory, Broad Institute of MIT and Harvard, Cambridge, MA, USA. <sup>65</sup>Division of Pulmonary and Critical Care Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL, USA. <sup>66</sup>Computational and Systems Biology Program, Sloan Kettering Institute, Memorial Sloan Kettering Cancer Center, New York, NY, USA. <sup>67</sup>Biological Research Centre of the Hungarian Academy of Sciences, Szeged, Hungary. <sup>68</sup>Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Tukholmankatu, Helsinki, Finland. <sup>69</sup>Department of Cell Biology, Regeneration Next Initiative, Duke University School of Medicine, Durham, NC, USA. <sup>70</sup>Department of Biology, Howard Hughes Medical Institute, MIT, Cambridge, MA, USA. <sup>71</sup>Genentech, South San Francisco, CA, USA. <sup>72</sup>Division of Pulmonary, Allergy and Critical Care Medicine, University of Pittsburgh, Pittsburgh, PA, USA. <sup>73</sup>Department of Pediatrics; Center for Genes, Environment, and Health, National Jewish Health, Denver, CO, USA. <sup>74</sup>Ragon Institute of MGH, MIT, and Harvard, Cambridge, MA, USA. <sup>75</sup>Institute for Medical Engineering and Science (IMES), Koch Institute for Integrative Cancer Research, Cambridge, MA, USA. <sup>76</sup>Department of Chemistry, Massachusetts Institute of Technology, Cambridge, MA, USA. <sup>77</sup>Department of Internal Medicine, Gastroenterology, University of Michigan Medical School, Ann Arbor, MI, USA. <sup>78</sup>Department of Cell and Developmental Biology, University of Michigan Medical School, Ann Arbor, MI, USA. <sup>79</sup>Department of Biomedical Engineering, University of Michigan College of Engineering, Ann Arbor, MI, USA. <sup>80</sup>Department of Physics/Cavendish Laboratory, University of Cambridge, JJ Thompson Ave, Cambridge, UK. <sup>81</sup>Chan Zuckerberg Biohub, San Francisco, CA, USA. <sup>82</sup>Department of Pediatrics, Respiratory Medicine, University of California, San Diego, CA, USA. <sup>83</sup>Department of Medicine and Institute for Stem Cell Biology and Regenerative Medicine, Stanford University School of Medicine, Stanford, CA, USA. <sup>84</sup>Department of Pediatrics, University of California, San Diego, San Diego, CA, USA. <sup>85</sup>Laboratory of Genetics, University of Wisconsin-Madison, Madison, WI, USA. <sup>86</sup>Université Côte d'Azur, CNRS, IPMC, Sophia-Antipolis, France.