Article

COVID-19 tissue atlases reveal SARS-CoV-2 pathology and cellular targets

COVID-19, which is caused by SARS-CoV-2, can result in acute respiratory distress syndrome and multiple organ failure[1-4], but little is known about its pathophysiology. Here we generated single-cell atlases of 24 lung, 16 kidney, 16 liver and 19 heart autopsy tissue samples and spatial atlases of 14 lung samples from donors who died of COVID-19. Integrated computational analysis uncovered substantial remodelling in the lung epithelial, immune and stromal compartments, with evidence of multiple paths of failed tissue regeneration, including defective alveolar type 2 differentiation and expansion of fibroblasts and putative *TP63*[+] intrapulmonary basal-like progenitor cells. Viral RNAs were enriched in mononuclear phagocytic and endothelial lung cells, which induced specific host programs. Spatial analysis in lung distinguished inflammatory host responses in lung regions with and without viral RNA. Analysis of the other tissue atlases showed transcriptional alterations in multiple cell types in heart tissue from donors with COVID-19, and mapped cell types and genes implicated with disease severity based on COVID-19 genome-wide association studies. Our foundational dataset elucidates the biological effect of severe SARS-CoV-2 infection across the body, a key step towards new treatments.

The host response to severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection ranges from asymptomatic infection to severe coronavirus disease 2019 (COVID-19) and death. The leading cause of mortality is acute lung injury and acute respiratory distress syndrome, or direct complications with multiple organ failure[1-4]. Clinical deterioration in acute illness leads to ineffective viral clearance and collateral tissue damage[1-5]. Severe COVID-19 is also accompanied by an inappropriate pro-inflammatory host immune response and a diminished antiviral interferon response[6-8].

Many molecular and cellular questions related to COVID-19 pathophysiology remain unanswered, including how cell composition and gene programs shift, which cells are infected, and how associated genetic loci drive disease. Autopsies are crucial to understanding severe COVID-19 pathophysiology[9-12], but comprehensive genomic studies are challenged by long post-mortem intervals (PMIs).

Here, we developed a large cross-body COVID-19 autopsy biobank of 420 autopsy specimens, spanning 11 organs, and used it to generate a single-cell atlas of lung, kidney, liver and heart associated with COVID-19 and a lung spatial atlas, in a subset of 14–18 donors per organ. Our atlases provide crucial insights into the pathogenesis of severe COVID-19.

A COVID-19 autopsy cohort and biobank

We assembled an autopsy cohort of 20 male and 12 female donors, of various ages (>30–>89 years), racial/ethnic backgrounds, intermittent mandatory ventilation (IMV) periods (0–30 days) and days from symptom start to death (Fig. 1a, Supplementary Table 1). A biobank was created with a subset of 17 donors. From most donors, we collected at least lung, heart and liver tissue (Fig. 1a, Extended Data Fig. 1a, Supplementary Methods), preserving specimens for single-cell and spatial analysis. We optimized single-cell and single-nucleus RNA sequencing (sc/snRNA-Seq) protocols for Biosafety Level 3 and NanoString GeoMx workflows to spatially profile RNA from different tissue compartments by cell composition or viral RNA (Supplementary Methods).

COVID-19 cell atlases

We generated sc/snRNA-Seq atlases of lung ($n = 16$ donors, $k = 106,792$ cells/nuclei, $m = 24$ specimens; donors D1–8, 10–17), heart ($n = 18$, $k = 40,880$, $m = 19$; D1–8, 10–11, 14–17, 27–28, 31–32), liver ($n = 15$, $k = 47,001$, $m = 16$; D1–7, 10–17) and kidney ($n = 16$, $k = 33,872$, $m = 16$; D4–8, 10–12, 14–15, 17, 25–26, 28–30). Although initial tests showed some differences in proportions of cell types between snRNA-Seq and scRNA-Seq, snRNA-Seq performed better overall[13] (Extended Data Fig. 1b–d and data not shown) and was used for the remaining samples.

We developed a computational pipeline (Fig. 1b) to tackle unique technical challenges. We used CellBender remove-background[14] to remove ambient RNA, which enhanced cell distinction and marker specificity (Extended Data Fig. 1e–h; Supplementary Methods), we rapidly quality-controlled, pre-processed and batch-corrected the data with cloud-based Cumulus[15] (Extended Data Fig. 2a–g, Supplementary Methods) and we automatically annotated cells and nuclei by transferring labels from previous atlases (Fig. 2a, Extended Data Fig. 2h, Supplementary Methods). We refined these labels with manual annotation of subclusters in each main lineage (Fig. 2b, Extended Data Fig. 2i–n, Supplementary Methods). The automated annotation approach allowed us to compare against other data resources (without clustering or batch correction), while the manual approach enabled us to refine cell identity assignments with detailed domain knowledge.

A cell census of the COVID-19 lung

Automatic annotation defined 28 subsets of parenchymal, endothelial and immune cells (Fig. 2a, Supplementary Table 2, Supplementary

A list of authors and their affiliations appears online. ✉e-mail: ivlachos@bidmc.harvard.edu; shalek@mit.edu; avillani@mgh.harvard.edu; orit.r.rosen@gmail.com; aviv.regev.sc@gmail.com

Article

Methods), with further manual annotation within subgroupings (Fig. 2b, Extended Data Figs. 2, 4, Supplementary Methods). Deconvolution of bulk RNA-Seq from the same samples largely agreed (Extended Data Fig. 3a, b, Supplementary Methods), and our two annotation strategies had 94% agreement (Extended Data Fig. 3c–e).

Among immune cells we distinguished six cell myeloid subsets: $CD14^{high}CD16^{high}$ inflammatory monocytes with antimicrobial properties and five macrophage subsets (Extended Data Figs. 2j, 4b) that were enriched for scavenger receptors, toll-like receptor ligands, inflammatory transcriptional regulators or metabolism genes; four B cell and plasma cell subsets: $BLIMP1^{high}$ plasma cells[16,17], $BLIMP1^{intermediate}$ plasma cells, B cells and $JCHAIN$-expressing plasmablasts (Extended Data Figs. 2k, 4b); five T and natural killer (NK) cell subsets: two $CD4^+$ subsets: regulatory T cells (T_{reg} cells) and a metabolically active subset; one $CD8^+$ subset; and two T or NK cell subsets (Extended Data Figs. 2l, 4b), including one with cytotoxic effector genes. The dearth of neutrophils (Fig. 2a, 419 cells) is likely due to freezing or limitations of droplet-based sc/snRNA-Seq[13].

We identified seven endothelial cell (EC) subsets[18,19] (Extended Data Figs. 2m, 4b): arterial, venous and lymphatic, capillary aerocytes, capillary EC-1 and capillary EC-2 and a mixed subset (Supplementary Methods), and three stromal subsets: fibroblasts, proliferative fibroblasts and myofibroblasts[19] (Extended Data Fig. 2n, Supplementary Table 3).

There were eight epithelial subsets, including club/secretory cells, AT1 cells, AT2 cells, and proliferative AT2 cells (Fig. 2b). One subset corresponded to a previously described AT2 to AT1 transitional cell state ($KRT8^+$ pre-alveolar type 1 transitional cell state (PATS); PATS/ADI/DATP)[20–22] (Fig. 2b).

Changes in lung cell composition

In comparison with normal lung from a matching region (Fig. 2c, Supplementary Methods), numbers of AT2 cells were significantly decreased (false discovery rate (FDR) = 2.8×10^{-15}, Dirichlet multinomial regression; Supplementary Methods), possibly reflecting virally induced cell death[23–25]. Numbers of dendritic cells (FDR = 0.004), macrophages (FDR = 3.6×10^{-10}), NK cells (FDR = 0.018), fibroblasts (FDR = 0.013), lymphatic endothelial cells (FDR = 0.00058) and vascular endothelial cells (FDR = 0.00011) all increased.

Cell proportions varied between donors (Extended Data Fig. 5a, b). Whereas variation was not significantly correlated with PMI, age or sex, IMV was positively correlated with epithelial cell fraction (FDR = 0.007; Spearman ρ = 0.765) and negatively correlated with T and NK cell fraction (FDR = 0.041; ρ = −0.62). Fewer days on a ventilator may indicate a rapidly deteriorating condition. This is corroborated by the nominally significant positive correlation between epithelial cell fraction and days from symptom start to death (ρ = 0.671, P = 0.004, but FDR = 0.053).

Induced programs in epithelial cells

There were widespread, cell-type-specific transcriptional changes in lung cell types associated with COVID-19 (Extended Data Fig. 5c, Supplementary Methods), most notably in $CD16^+$ monocytes (1,580 upregulated genes), lymphatic endothelial (578), vascular endothelial (317), AT2 (309) and AT1 (307) cells. Within AT2 cells, there was higher expression (P < 0.0004) of genes associated with host viral response (Fig. 2d), including those for programmed cell death ($STAT1$), inflammation and adaptive immune response (Supplementary Table 4). Lung surfactant genes were downregulated, consistent with reports from in vitro studies[21].

Failed paths for AT1 cell regeneration

The PATS program signature was increased in COVID-19 pneumocytes (P < 2.2×10^{-16}, one-sided Mann–Whitney U test) (Fig. 2e, Extended Data
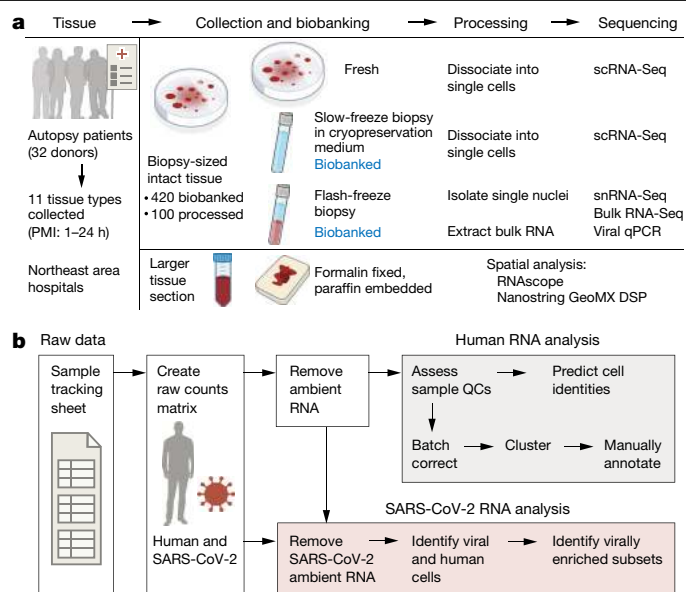

Fig. 1 | Experimental and computational pipeline for a COVID-19 autopsy atlas. a, Sample processing pipeline. Up to 11 tissue types were collected from 32 donors. **b**, sc/snRNA-Seq analysis pipeline. QC, quality control.

Fig. 5d). This progenitor program is induced during lung injury[20–22] (for example, idiopathic pulmonary fibrosis), consistent with fibrosis in severe COVID-19[26,27]. These studies also highlight fibroblast expansion, which we also observed (Fig. 2c).

A subset of PATS program cells, distinct from $KRT5^+TP63^+$ airway basal cells, expressed canonical ($KRT8/CLDN4/CDKN1A$) and non-canonical ($KRT5/TP63/KRT17$) PATS markers (Fig. 2f, Extended Data Fig. 5d, Supplementary Table 3). These may be $TP63^+$ intrapulmonary basal-like progenitor (IPBLP) cells, which were identified in H1N1 influenza mouse models[28] and act as an emergency cellular reserve for severely damaged alveoli[29]. The putative IPBLP cells express interferon virus defence and progenitor cell differentiation genes (Supplementary Table 3). Thus, multiple emergency pathways for alveolar cell regeneration are activated in lung (Fig. 2g, Discussion).

Changed cell composition with viral load

To determine viral load and associated host responses, we analysed donor- and cell-type-specific distribution of SARS-CoV-2 reads (Fig. 3a, b, Extended Data Fig. 6a–d, Supplementary Methods). Reads spanned the entire SARS-CoV-2 genome, with bias towards positive-sense alignments. A few cells had reads aligning to all viral segments, including the negative strand (Extended Data Fig. 6e), potentially indicating productive infection. Virus detection was not technically driven (Extended Data Fig. 6f–i), and inter-donor variation was consistent with SARS-CoV-2 qRT–PCR on bulk RNA (Extended Data Fig. 6j–l, Supplementary Methods). Viral load was negatively correlated with days from symptom start to death (Fig. 3c), as previously reported[30,31]. Bulk RNA-Seq yielded nine unique complete viral genomes from nine donors with high viral loads (Extended Data Fig. 6m, Supplementary Methods); all genomes carried the D614G allele. We identified no other common respiratory viral co-infections (Extended Data Fig. 6n). Total viral burden per sample (including ambient RNA; Supplementary Methods) positively correlated with proportions of mast cells, specific macrophage subsets, venular endothelial cells and capillary aerocyte endothelial cells (Extended Data Fig. 6o–u).

Genes upregulated in biopsy samples with high versus low or no viral load (Supplementary Methods) included viral response and innate immune processes ($\log_2$(fold change) > 1.4, Wald test,
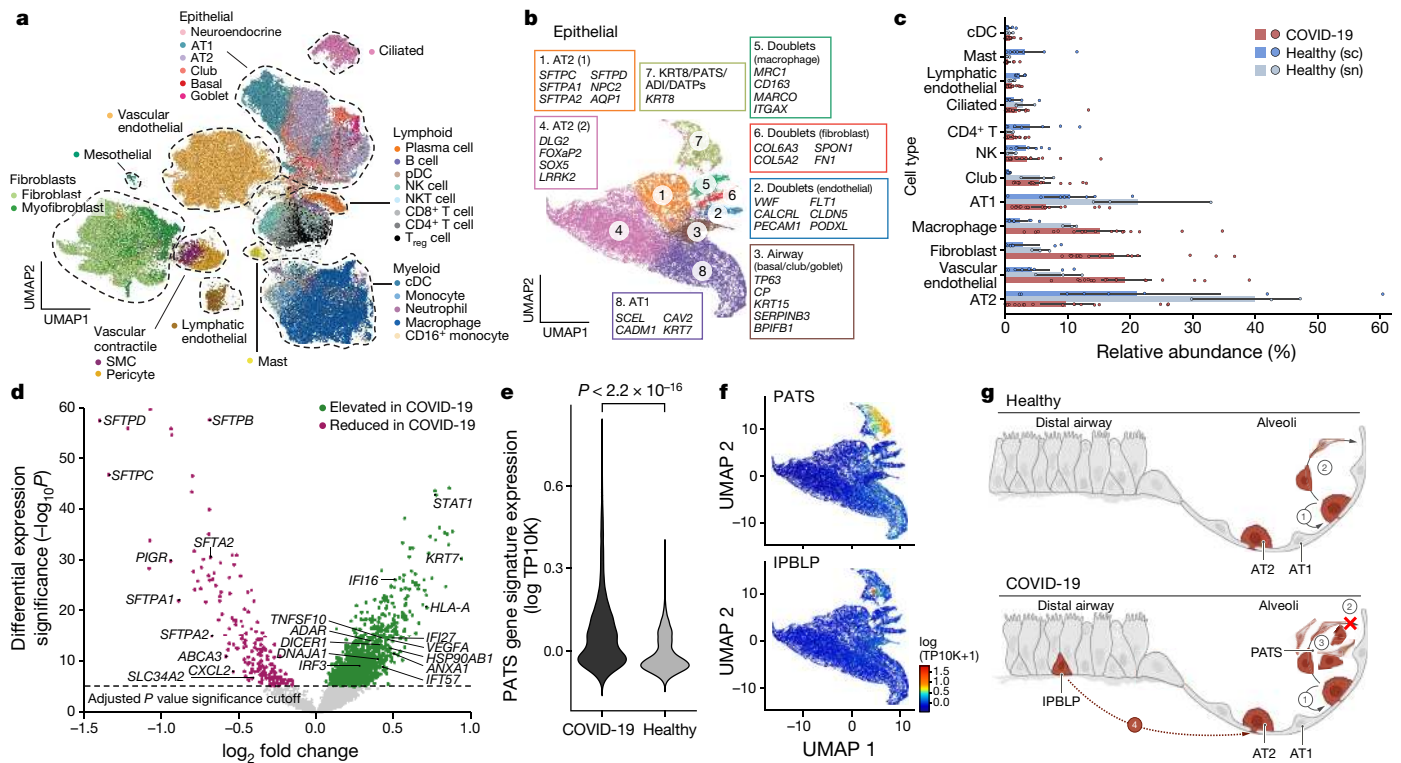
Fig. 2 | A single-cell and single-nucleus atlas of COVID-19 lung. a, Automatic prediction identifies 28 cell subsets across compartments. UMAP embedding of 106,792 harmonized sc/snRNA-Seq profiles (dots) from 24 tissue samples of 16 lung donors with COVID-19, coloured by automatic annotations (legend). **b**, Epithelial cell subsets. UMAP embedding of 21,661 epithelial cell or nucleus profiles, coloured by manual annotations, with highly expressed marker genes (boxes). **c**, **d**, Cell composition and expression differences between COVID-19 and healthy lung. **c**, Cell proportions (x axis: mean, bar; 95% confidence intervals, line) in each automatically annotated subset (y axis) in COVID-19 snRNA-Seq (red, $n = 16$), healthy snRNA-Seq (grey, $n = 3$) and healthy scRNA-Seq ($n = 8$, blue). Cell types shown have a COVID-19 versus healthy snRNA-Seq false discovery rate (FDR) of <0.05 (Dirichlet multinomial regression). **d**, Significance ($-\log_{10}(P)$, y axis) versus magnitude ($\log_2$(fold change), x axis) of differential expression of each gene (dots; horizontal dashed line, FDR < 0.05)

between COVID-19 and healthy lung from a total of 2,000 AT2 cells and 14 studies (two-sided test; Supplementary Methods). **e**, **f**, An increased pre-alveolar type 1 transitional cell state (PATS)[20–22] program in pneumocytes in COVID-19 versus healthy lung. **e**, Distribution of PATS signature scores (y axis) for 17,655 cells from COVID-19 and 24,000 cells from healthy lung pneumocytes (x axis). $P < 2.2 \times 10^{-16}$ (one-sided Mann–Whitney U test). **f**, UMAP embedding of 21,661 epithelial cell profiles (dots) coloured by signature level (colour legend, lower right) for the PATS (top) or intrapulmonary basal-like progenitor (IPBLP) cell (bottom) programs. **g**, Model of epithelial cell regeneration in healthy and COVID-19 lung. In healthy alveoli (top), AT2 cells self-renew (1) and differentiate into AT1 cells (2). In COVID-19 alveoli (bottom), AT2 cell self-renewal (1) and AT1 differentiation (2) are inhibited, resulting in PATS accumulation (3) and recruitment of airway-derived IPBLP cells to alveoli (4).

FDR-corrected $P < 0.05$; Extended Data Fig. 6v, Supplementary Table 4) and significantly overlapped with those in bulk RNA-Seq of post-mortem COVID-19 lungs in another study[32] (FDR = 3.12×10^{-6}, Kolmogorov–Smirnov test). Downregulated genes ($\log_2$(fold change) < 1.4, Wald test, FDR-corrected $P < 0.05$) were involved in surfactant metabolism dysfunction and lamellar bodies (secretory vesicles in AT2 cells[33]).

Lung cells enriched for SARS-CoV-2 RNA

Myeloid cells were the cell category most enriched for SARS-CoV-2 RNA (158 cells after correction for ambient RNA, FDR < 0.013; Fig. 3a, Extended Data Fig. 6w–y, Supplementary Methods), with particular enrichment in $CD14^{high}CD16^{high}$ inflammatory monocytes (FDR < 0.005) and $LDB2^{high}OSMR^{high}YAP1^{high}$ macrophages (FDR < 0.02; Extended Data Figs. 6x, 7a, b), although enrichment scores in individual donors varied. There was elevated, but non-significantly enriched, viral RNA in endothelial cells, with the capillary EC-2 (cluster 3, FDR < 0.017) and lymphatic endothelial cells (cluster 7, FDR < 0.006) enriched compared with other endothelial subsets (Fig. 3a, Extended Data Figs. 6w, y, 7c, d). There were also SARS-CoV-2 RNA⁺ cells among mast cells, and B and plasma cells, and viral RNA reads in multiple other cell types (Fig. 3a, Extended Data Fig. 6w). Notably, SARS-CoV-2 RNA⁺ cells did not

co-express the entry factors *ACE2* and *TMPRSS2*, or other hypothesized entry cofactors (Fig. 3b, Extended Data Fig. 7e–h).

Immune programs in SARS-CoV-2 RNA⁺ cells

SARS-CoV-2 RNA⁺ cells had distinct transcriptional programs compared with SARS-CoV-2 RNA⁻ counterparts, with differentially expressed genes (FDR < 0.05; Supplementary Methods) in epithelial and myeloid cells, including $PPARG^{high}CD15L^{high}$ macrophages and $CD14^{high}CD16^{high}$ inflammatory monocytes (Supplementary Table 5). Genes upregulated in epithelial SARS-CoV-2 RNA⁺ cells were enriched for TNF, AP1 and chemokine and cytokine signalling, SARS-CoV-2-driven cell responses in vitro[32], and keratinization pathways, which may reflect an injury response (Extended Data Fig. 7i). Genes upregulated in myeloid SARS-CoV-2 RNA⁺ cells were those associated with chemokine and cytokine signalling, and responses to interferon, TNF, intracellular pathogens and viruses (Fig. 3d, Extended Data Fig. 7j–m, Supplementary Table 5), as previously described[34,35]. Cytokines and viral host response genes were upregulated in both $CD14^{high}CD16^{high}$ inflammatory monocytes and $PPARG^{high}CD15L^{high}$ macrophages (Extended Data Fig. 7m, Supplementary Table 5), including *CXCL10* and *CXCL11*, which were upregulated in nasopharyngeal swabs[35] and bronchoalveolar lavages[34].
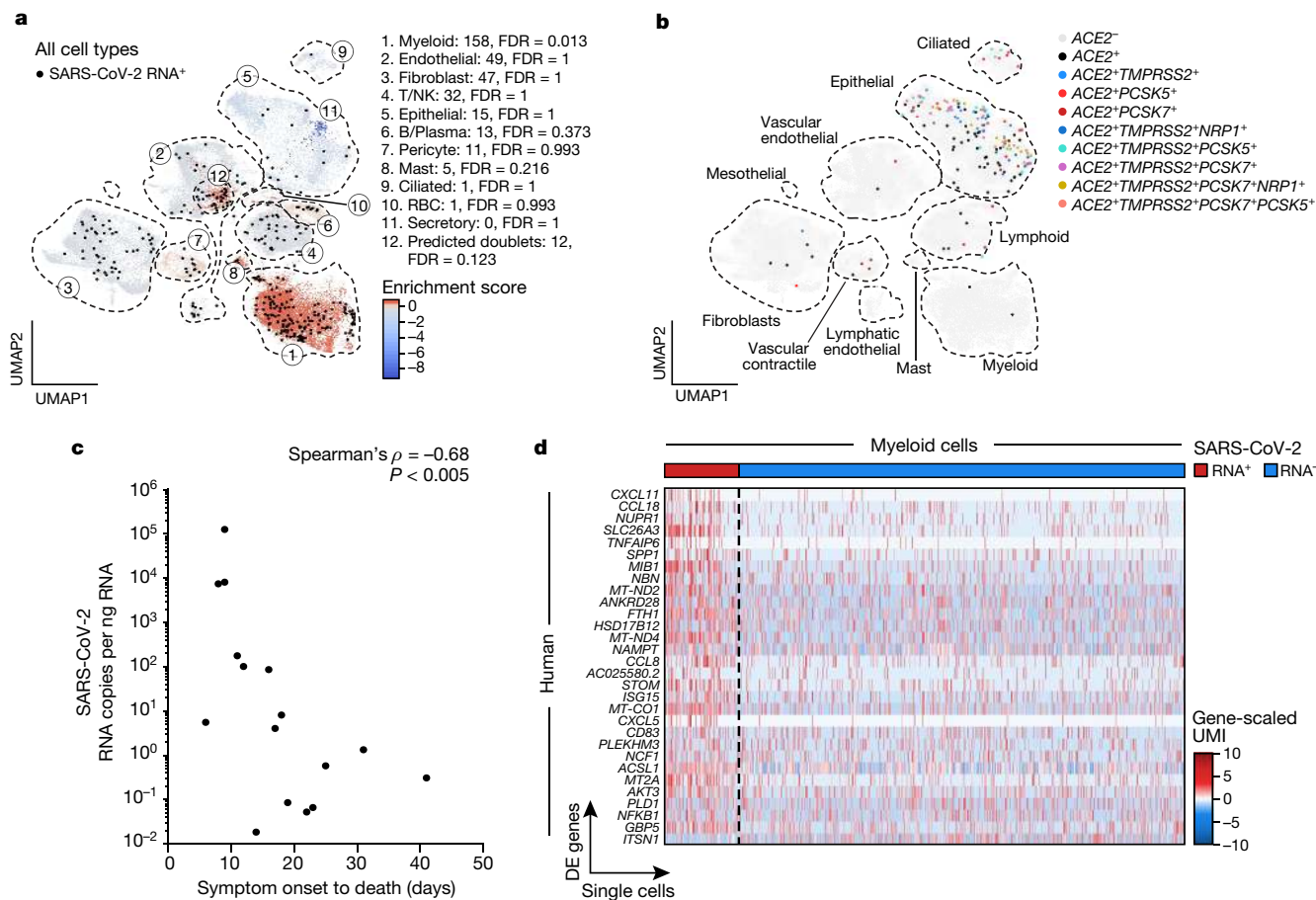
Fig. 3 | SARS-CoV-2 RNA⁺ single cells are enriched for phagocytic and endothelial cells. a, **b**, Many SARS-CoV-2 RNA⁺ single cells do not express known SARS-CoV-2 entry factors. UMAP embedding of all 106,792 lung cells or nuclei (as in Fig. 2a), showing either **a**, only the 40,581 cells from seven donors containing any SARS-CoV-2 RNA⁺ cell, coloured by viral enrichment score (Supplementary Methods, red: stronger enrichment) and by SARS-CoV-2 RNA⁺ cells (black points), and marked by annotation and FDR of enrichment (legend) or **b**, all 106,792 cells/nuclei, coloured by expression of SARS-CoV-2 entry factors (co-expression combinations with at least 10 cells are shown). Dashed lines, major cell types. **c**, Reduction in SARS-CoV-2 RNA with prolonged symptom onset to death interval (Spearman $\rho = -0.68$, $P < 0.005$, two-sided test). Symptom onset to death (x axis, days) and lung SARS-CoV-2 copies per nanogram input RNA (y axis) for each donor ($n = 16$). **d**, Expression changes in SARS-CoV-2 RNA⁺ myeloid cells. Significantly differentially expressed (DE) host genes (log-normalized and scaled digital gene expression, rows; cutoff: FDR < 0.05 and $\log_2$(fold change) > 0.5) across SARS-CoV-2 RNA⁺ ($n = 158$) and SARS-CoV-2 RNA⁻ myeloid cells ($n = 790$) (columns).

A spatial atlas of COVID-19 lung

To provide tissue context, we used Nanostring GeoMx Digital Spatial Profiling (DSP) for transcriptomic profiling from regions of interest (Supplementary Methods) in 14 donors, including three deceased healthy donors ('healthy') (Extended Data Fig. 1a). Regions of interest spanned a range of anatomical structures and viral abundance on the basis of SARS-CoV-2 RNA hybridization signals; when possible, we segmented them to PanCK⁺ and PanCK⁻, and inflamed and normal-appearing alveoli areas of illumination (AOIs) to capture RNA (Fig. 4a, Extended Data Figs. 8a, 9a, Supplementary Methods). We acquired high-quality profiles (Extended Data Fig. 8b) from matched AOIs on the basis of distance to morphological landmarks (Supplementary Methods). SARS-CoV-2 RNA expression varied by donor, with elevated levels in four donors (Extended Data Fig. 8c, d, Supplementary Methods), consistent with viral qRT–PCR and sc/snRNA-Seq. Given the good agreement between a targeted 1,811-gene panel and a whole transcriptome (WTA) panel (18,335 genes) (Extended Data Fig. 8e–g, Supplementary Table 6), we focused our analyses on WTA data. For D8–12, 18–24, we contrasted donors with COVID-19 and healthy donors and COVID-19 epithelial and non-epithelial AOIs; for D13–17, we focused on distinct anatomical regions and inflamed versus normal-appearing regions within donors.

Inflammatory activation in alveoli

Deconvolution of major cell type composition (Fig. 4b, Extended Data Fig. 8h, Supplementary Table 7, 8, Supplementary Methods) showed inferred AT1 and AT2 cells dominating the PanCK⁺ compartments and greater cellular diversity in the PanCK⁻ compartment. COVID-19 PanCK⁻ AOIs had increased fibroblast and myofibroblast scores compared with controls, in line with parallel spatial studies[36,37].

Comparing COVID-19 alveolar AOIs with control lungs from deceased healthy donors, there was upregulation of IFNα and IFNγ response genes and oxidative phosphorylation pathways (Fig. 4c, Extended Data Fig. 8i–k, Supplementary Table 6), similar to bulk RNA-Seq of highly infected tissue (*IFIT1*, *IFIT3*, *IDO1*, *GZMB*, *LAG3*, *NKG7* and *PRF1*) and to SARS-CoV-2⁺ myeloid cells (*TNFAIP6*, *CXCL11*, *CCL8*, *ISG1* and *GBP5*) and consistent with PANoptosis in a COVID-19 model[38]. Conversely, TNF, IL2–STAT5 and TGFβ signalling as well as apical junction and hypoxia were downregulated. Decreased TNF signalling expression in PanCK⁺ alveoli contrasts with its increase in SARS-CoV-2⁺ epithelial cells in snRNA-Seq and with reported[38] synergy between TNF and IFNγ in mouse models of COVID-19.

Comparison of inflamed and normal-appearing AOIs within the same alveolar biopsy samples of COVID-19 lungs (Extended Data Fig. 9, Supplementary Table 9, D13–D17), showed that upregulated genes were
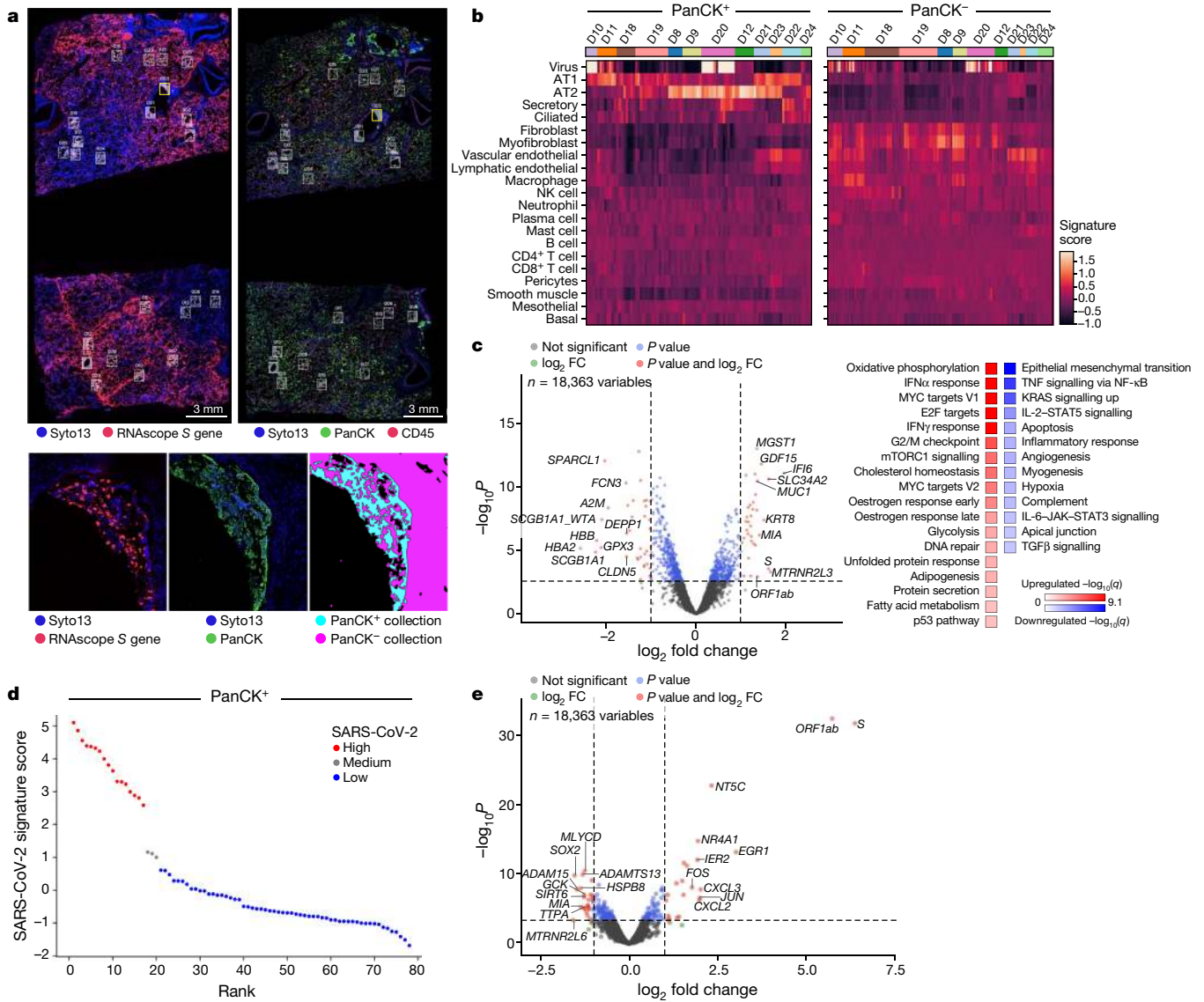
Fig. 4 | Composition and expression differences between COVID-19 and healthy lungs and between infected and uninfected regions within COVID-19 lungs. a, Example of analysed regions. Top: RNAscope (left) and immunofluorescent staining (right) of donor D20 with collection regions of interest (ROIs) and matched areas in white rectangles. Bottom: one ROI (yellow rectangle) from each scan (left and middle) and the segmented collection areas of illumination (AOIs) (right). **b**, Cell composition differences between PanCK$^+$ and PanCK$^-$ alveolar AOIs and between AOIs from COVID-19 ($n = 9,190$ AOIs) and healthy (D22–24, 38 AOIs) lungs. Expression scores (colour bar) of cell type signatures (rows) in PanCK$^+$ (left) and PanCK$^-$ (right) alveolar AOIs (columns) in whole transcriptome (WTA) data from different donors (top colour bar). **c**, Differential gene expression in COVID-19 versus healthy lung. Left: significance ($-\log_{10}(P)$, y axis) and magnitude ($\log_2$(fold change), x axis) of differential expression of each gene (dots) in WTA data between PanCK$^+$ alveoli AOIs from COVID-19 ($n = 78$) and healthy ($n = 18$) lung. Right: significance ($-\log_{10}(q)$) of enrichment (permutation test) of different pathways (rows). **d**, **e**, Changes in gene expression in SARS-CoV-2 high versus low AOIs within COVID-19 lungs in WTA data. **d**, SARS-CoV-2 high and low alveolar AOIs. PanCK$^+$ alveolar AOIs (dots) rank ordered by their SARS-CoV-2 signature score (y axis) in WTA data, and partitioned to high (red), medium (grey) and low (blue) SARS-CoV-2 AOIs. **e**, Significance ($-\log_{10}(P)$, y axis) and magnitude ($\log_2$(fold change), x axis) of differential expression of each gene (dots) in WTA data between SARS-CoV-2 high and low AOIs for PanCK$^+$ alveoli (AOIs: 17 high, 3 medium, 58 low). Horizontal dashed line, FDR = 0.05; vertical dashed lines, |$\log_2$(fold change)| = 2. FC, fold change. The top 10 differentially expressed genes by fold change are marked.

enriched for innate immune and inflammatory pathways[39,40], including neutrophil degranulation (FDR = 5.2×10^{-17}) and IFNγ (FDR = 3.4×10^{-15}) and interleukin (FDR = 1.4×10^{-13}) signalling. TNF pathway expression was elevated in inflamed tissue, albeit not significantly (FDR = 0.097). Claudins and tight junction pathways were downregulated, corroborating a disrupted alveolar barrier, as in influenza[41,42]. Cilium assembly genes were enriched when comparing bronchial epithelial AOIs and matched normal-appearing alveoli (Extended Data Fig. 9d, Supplementary Table 9).

Comparison of SARS-CoV-2 high and low AOIs (Fig. 4d, e, Extended Data Fig. 8l, m, Supplementary Methods) revealed induction of the viral *ORF1ab* and *S* genes and upregulation of chemokine genes (*CXCL2* and *CXCL3*) and immediate early genes in the PanCK$^+$ compartment, consistent with snRNA-Seq (Supplementary Table 9, Extended Data Fig. 7i). *NT5C*, which encodes a nucleotidase with a preference for 5′-dNTPs, is consistently upregulated in SARS-CoV-2-high AOIs (Fig. 4e, Extended Data Fig. 8m, Supplementary Table 9). This gene is not known to have a role in lung injury and should be further studied.

Article

COVID-19 effect on heart, kidney and liver

We next profiled liver, heart and kidney by snRNA-Seq with automated and manual annotation of parenchymal, endothelial and immune cells (Supplementary Methods, Extended Data Figs. 10, 11). Although other studies have reported viral reads in COVID-19 non-lung tissues[43], we detected very few viral RNA reads in all three tissues, most of which could not be assigned to nuclei (Extended Data Fig. 11l); this absence was confirmed by NanoString DSP and RNAscope (data not shown).

Focusing on heart, both cell composition and gene programs changed between COVID-19 and healthy heart. There was a significant reduction in the proportion of cardiomyocytes and pericytes, and an increase in vascular endothelial cells (Extended Data Fig. 11e). Genes upregulated (FDR < 0.01) in cardiomyocytes, pericytes or fibroblasts (Extended Data Fig. 11g–i, Supplementary Table 10) included *PLCG2*, the cardiac role of which is unknown but which was induced in all major heart cell subtypes (Extended Data Fig. 11j), and *AFDN*, which is upregulated in endothelial cells (Extended Data Fig. 11k), and which encodes a junction adherens complex component[44] that is necessary for endothelial barrier function. Upregulated pathways include oxidative-stress-induced apoptosis in pericytes, cell adhesion and immune pathways in cardiomyocytes, and cell differentiation processes in fibroblasts (Supplementary Table 10).

COVID-19 cell types related through GWAS

Finally, we aimed to identify genes and cell types associated with COVID-19 risk by integrating our atlas data with genome-wide association studies (GWAS)[45] for common[46] variants associated with COVID-19 (Supplementary Methods). Among 26 genes proximal to six COVID-19 GWAS regions (Supplementary Table 11, Supplementary Methods), 14 genes had higher average expression in the lung (*P* < 0.05, *t*-test; Extended Data Fig. 12a–d), 21 had significant (FDR < 0.05) expression specificity in at least one lung cell type, including *FOXP4* (chromosome (chr.) 6, AT1 and AT2 cells), and *CCR1* and *CCRL2* (chr. 3, macrophages) (Extended Data Fig. 12e, Supplementary Table 11), and 18 were differentially expressed (FDR < 0.05) in COVID-19 compared with healthy lung (for example, *SLC6A20* in goblet cells, *CCR5* in CD8+ T cells and T_reg cells, and *CCR1* in macrophage and CD16+ monocytes (Extended Data Fig. 12f, Supplementary Table 11).

We related heritability from GWAS of COVID-19 severity traits to either cell type programs (genes enriched in a cell type in each tissue) or disease progression programs (genes differentially expressed between COVID-19 and controls in a cell type) in each tissue using sc-linker[47] (Supplementary Methods). AT2 (4.8× heritability enrichment, *P* = 0.04), CD8+ T cells (4.4×, *P* = 0.009) and ciliated cell programs in the lung, proximal convoluted tubule and connecting tubule programs in kidney, and cholangiocyte programs in liver attained nominal (but not Bonferroni-corrected) significance (Extended Data Fig. 12g, h, Supplementary Table 11). Of all disease progression programs, only the club cell program (single-cell level model) had nominally significant heritability enrichment (10.5×, *P* = 0.04 for severe COVID-19) (Extended Fig. 12g, Supplementary Table 11).

The highest number of driving genes was observed for lung AT2 cells and spanned several loci, hinting at a polygenic architecture linking AT2 cells with severe COVID-19 (Supplementary Methods, Supplementary Table 11). Implicated GWAS proximity genes include *OAS3* in lung AT2 and club cells, and *SLC4A7* in lung CD8+ T cells (Supplementary Table 11), as well as genes at unresolved significantly associated GWAS loci (Extended Data Fig. 12i), such as *FYCO1* (AT2, ciliated, club; chr. 3p), *NFKBIZ* (AT2; chr. 3q) and *DPP9* (AT2; chr. 19) (Supplementary Table 11).

Discussion

We built a biobank of severe COVID-19 autopsy tissue and atlases of COVID-19 lung, heart, liver and kidney (Extended Data Fig. 12j), complementing a sister lung atlas[48].

Among the changes in lung cell composition in COVID-19, is a reduction in AT2 cells and the presence of PATS and IPBLP-like cells, indicating that multiple regenerative strategies are invoked to re-establish alveolar epithelial cells lost to infection. A serial failure of epithelial progenitors to regenerate at a sufficient rate, first by secretory progenitor cells in the nasal passages and large and small airways, followed by alveolar AT2 cells, PATS and IPBLP cells, may eventually lead to lung failure.

Viral RNA in the lung varied, was negatively correlated with time from symptom start to death, and was primarily detected in myeloid and endothelial cells (as in nonhuman primates[49]); spatial analysis supports high virus levels at the earlier stages of infection[36,37,50]. Epithelial cells were not enriched in high viral RNA samples or in SARS-CoV-2+ cells, consistent with their excessive death. Cell-associated SARS-CoV-2 unique molecular identifiers may represent a mixture of replicating virus, immune cell engulfment and virions or virally infected cells attached to the cell surface. We did not detect viral RNA in the heart, liver or kidney, but observed other changes, including broad upregulation of *PLCG2*, a target of Bruton's tyrosine kinase (BTK), in the heart[51].

Combining our profiles with GWAS of COVID-19, we related specific cell types to heritable risk, especially AT2, ciliated and CD8+ T cells and macrophages, as well as genes in multi-gene regions underlying the association. This analysis can improve as GWAS grows and atlases expand.

Our study was limited by a modest number of donors without pre-selection of features, the terminal time point, limited distinction between viral RNA and true infection, and technical confounders such as PMIs. Nevertheless, our methods would enable studies in diverse diseased or damaged tissues. Future meta-analyses will further enhance its power and provide crucial resources for the community studying host–SARS-CoV-2 biology.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-021-03570-8.

1. Guan, W.-J. et al. Clinical characteristics of coronavirus disease 2019 in China. *N. Engl. J. Med.* **382**, 1708–1720 (2020).
2. Puelles, V. G. et al. Multiorgan and renal tropism of SARS-CoV-2. *N. Engl. J. Med.* **383**, 590–592 (2020).
3. Huang, C. et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* **395**, 497–506 (2020).
4. Xu, Z. et al. Pathological findings of COVID-19 associated with acute respiratory distress syndrome. *Lancet Respir. Med.* **8**, 420–422 (2020).
5. Varga, Z. et al. Endothelial cell infection and endotheliitis in COVID-19. *Lancet* **395**, 1417–1418 (2020).
6. Chen, G. et al. Clinical and immunological features of severe and moderate coronavirus disease 2019. *J. Clin. Invest.* **130**, 2620–2629 (2020).
7. Qin, C. et al. Dysregulation of immune response in patients with coronavirus 2019 (COVID-19) in Wuhan, China. *Clin. Infect. Dis.* **71**, 762–768 (2020).
8. Hadjadj, J. et al. Impaired type I interferon activity and exacerbated inflammatory responses in severe COVID-19 patients. *Science* **369**, 718–724 (2020).
9. Bian, X.-W. et al. Autopsy of COVID-19 patients in China. *Natl. Sci. Rev.* **7**, 1414–1418 (2020).
10. Menter, T. et al. Postmortem examination of COVID-19 patients reveals diffuse alveolar damage with severe capillary congestion and variegated findings in lungs and other organs suggesting vascular dysfunction. *Histopathology* **77**, 198–209 (2020).
11. Wichmann, D. et al. Autopsy findings and venous thromboembolism in patients with COVID-19: a prospective cohort study. *Ann. Intern. Med.* **173**, 268–277 (2020).
12. Bösmüller, H. et al. The evolution of pulmonary pathology in fatal COVID-19 disease: an autopsy study with clinical correlation. *Virchows Arch.* **477**, 349–357 (2020).
13. Slyper, M. et al. A single-cell and single-nucleus RNA-Seq toolbox for fresh and frozen human tumors. *Nat. Med.* **26**, 792 –802 (2020).
14. Fleming, S. J., Marioni, J. C. & Babadi, M. CellBender remove-background: a deep generative model for unsupervised removal of background noise from scRNA-seq datasets. Preprint at https://doi.org/10.1101/791699 (2019).
15. Li, B. et al. Cumulus provides cloud-based data analysis for large-scale single-cell and single-nucleus RNA-seq. *Nat. Methods* **17**, 793–798 (2020).
16. Shaffer, A. L. et al. Blimp-1 orchestrates plasma cell differentiation by extinguishing the mature B cell gene expression program. *Immunity* **17**, 51–62 (2002).
17. Martins, G. & Calame, K. Regulation and functions of Blimp-1 in T and B lymphocytes. *Annu. Rev. Immunol.* **26**, 133–169 (2008).

18. Schupp, J. C. et al. Integrated single cell atlas of endothelial cells of the human lung. Preprint at https://doi.org/10.1101/2020.10.21.347914 (2020).
19. Travaglini, K. J. et al. A molecular cell atlas of the human lung from single-cell RNA sequencing. *Nature* **587**, 619–625 (2020).
20. Strunz, M. et al. Alveolar regeneration through a Krt8+ transitional stem cell state that persists in human lung fibrosis. *Nat. Commun.* **11**, 3559 (2020).
21. Kobayashi, Y. et al. Persistence of a regeneration-associated, transitional alveolar epithelial cell state in pulmonary fibrosis. *Nat. Cell Biol.* **22**, 934–946 (2020).
22. Choi, J. et al. Inflammatory signals induce AT2 cell-derived damage-associated transient progenitors that mediate alveolar regeneration. *Cell Stem Cell* **27**, 366–382.e7 (2020).
23. 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).
24. Sungnak, 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).
25. Muus, C. et al. Single-cell meta-analysis of SARS-CoV-2 entry genes across tissues and demographics. *Nat. Med.* **27**, 546–559 (2021).
26. Xu, J. et al. SARS-CoV-2 induces transcriptional signatures in human lung epithelial cells that promote lung fibrosis. *Respir. Res.* **21**, 182 (2020).
27. Grillo, F., Barisione, E., Ball, L., Mastracci, L. & Fiocca, R. Lung fibrosis: an undervalued finding in COVID-19 pathological series. *Lancet Infect. Dis.* **21**, e72 (2021).
28. Vaughan, A. E. et al. Lineage-negative progenitors mobilize to regenerate lung epithelium after major injury. *Nature* **517**, 621–625 (2015).
29. Fernanda de Mello Costa, M., Weiner, A. I. & Vaughan, A. E. Basal-like progenitor cells: a review of dysplastic alveolar regeneration and remodeling in lung repair. *Stem Cell Reports* **15**, 1015–1025 (2020).
30. Wölfel, R. et al. Virological assessment of hospitalized patients with COVID-2019. *Nature* **581**, 465–469 (2020).
31. Walsh, K. A. et al. SARS-CoV-2 detection, viral load and infectivity over the course of an infection. *J. Infect.* **81**, 357–371 (2020).
32. Blanco-Melo, D. et al. Imbalanced host response to SARS-CoV-2 drives development of COVID-19. *Cell* **181**, 1036–1045.e9 (2020).
33. Johnson, N. F. Release of lamellar bodies from alveolar type 2 cells. *Thorax* **35**, 192–197 (1980).
34. Grant, R. A. et al. Circuits between infected macrophages and T cells in SARS-CoV-2 pneumonia. *Nature* **590**, 635–641 (2021).
35. Butler, D. et al. Shotgun transcriptome, spatial omics, and isothermal profiling of SARS-CoV-2 infection reveals unique host responses, viral diversification, and drug interactions. *Nat. Commun.* **12**, 1660 (2021).
36. Park, J. et al. Systemic tissue and cellular disruption from SARS-CoV-2 infection revealed in COVID-19 autopsies and spatial omics tissue maps. Preprint at https://doi.org/10.1101/2021.03.08.434433 (2021).
37. Rendeiro, A. F. et al. The spatio-temporal landscape of lung pathology in SARS-CoV-2 infection. Preprint at https://doi.org/10.1101/2020.10.26.20219584 (2020).
38. Karki, R. et al. Synergism of TNF-α and IFN-γ triggers inflammatory cell death, tissue damage, and mortality in SARS-CoV-2 infection and cytokine shock syndromes. *Cell* **184**, 149–168.e17 (2021).
39. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).
40. Mootha, V. K. et al. PGC-1α-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat. Genet.* **34**, 267–273 (2003).
41. van de Sandt, C. E. et al. Human CD8+ T cells damage noninfected epithelial cells during influenza virus infection in vitro. *Am. J. Respir. Cell Mol. Biol.* **57**, 536–546 (2017).
42. Short, K. R. et al. Influenza virus damages the alveolar barrier by disrupting epithelial cell tight junctions. *Eur. Respir. J.* **47**, 954–966 (2016).
43. Lemieux, J. E. et al. Phylogenetic analysis of SARS-CoV-2 in Boston highlights the impact of superspreading events. *Science* **371**, eabe3261 (2021).
44. Yu, H. H. & Zallen, J. A. Abl and Canoe/Afadin mediate mechanotransduction at tricellular junctions. *Science* **370**, eaba5528 (2020).
45. COVID-19 Host Genetics Initiative. The COVID-19 Host Genetics Initiative, a global initiative to elucidate the role of host genetic factors in susceptibility and severity of the SARS-CoV-2 virus pandemic. *Eur. J. Hum. Genet.* **28**, 715–718 (2020).
46. Severe Covid-19 GWAS Group. Genomewide association study of severe Covid-19 with respiratory failure. *N. Engl. J. Med.* **383**, 1522–1534 (2020).
47. Jagadeesh, K. A. et al. Identifying disease-critical cell types and cellular processes across the human body by integration of single-cell profiles and human genetics. Preprint at https://doi.org/10.1101/2021.03.19.436212 (2021).
48. Melms, J. C. et al. A molecular single-cell lung atlas of lethal COVID-19. *Nature* https://doi.org/10.1038/s41586-021-03569-1 (2021).
49. Speranza, E. et al. SARS-CoV-2 infection dynamics in lungs of African green monkeys. Preprint at https://doi.org/10.1101/2020.08.20.258087 (2020).
50. Desai, N. et al. Temporal and spatial heterogeneity of host response to SARS-CoV-2 pulmonary infection. Preprint at https://doi.org/10.1101/2020.07.30.20165241 (2020).
51. Liu, T.-M. et al. Hypermorphic mutation of phospholipase C, γ2 acquired in ibrutinib-resistant CLL confers BTK independency upon B-cell receptor activation. *Blood* **126**, 61–68 (2015).

Article

Toni M. Delorey[1,64], Carly G. K. Ziegler[2,3,4,5,6,7,64], Graham Heimberg[1,64],
Rachelly Normand[2,8,9,10,11,64], Yiming Yang[1,8,64], Åsa Segerstolpe[1,64], Domenic Abbondanza[1,64],
Stephen J. Fleming[12,13,64], Ayshwarya Subramanian[1,64], Daniel T. Montoro[2,64],
Karthik A. Jagadeesh[1,64], Kushal K. Dey[14,64], Pritha Sen[2,8,15,16,64], Michal Slyper[1,64],
Yered H. Pita-Juárez[2,10,17,18,19,64], Devan Phillips[1,64], Jana Biermann[20,21,64],
Zohar Bloom-Ackermann[22], Nikolaos Barkas[12], Andrea Ganna[23,24], James Gomez[22],
Johannes C. Melms[20,21], Igor Katsyv[25], Erica Normandin[2,10], Pourya Naderi[10,17,18],
Yury V. Popov[10,26,27], Siddharth S. Raju[2,28,29], Sebastian Niezen[10,26,27], Linus T.-Y. Tsai[2,10,26,30,31],
Katherine J. Siddle[2,32], Malika Sud[1], Victoria M. Tran[22], Shamsudheen K. Vellarikkal[2,33],
Yiping Wang[20,21], Liat Amir-Zilberstein[1], Deepak S. Atri[2,33], Joseph Beechem[34],
Olga R. Brook[35], Jonathan Chen[2,36], Prajan Divakar[34], Phylicia Dorceus[1], Jesse M. Engreitz[2,37],
Adam Essene[26,30,31], Donna M. Fitzgerald[38], Robin Fropf[34], Steven Gazal[39], Joshua Gould[1,12],
John Grzyb[40], Tyler Harvey[1], Jonathan Hecht[10,17], Tyler Hether[34], Judit Jané-Valbuena[1],
Michael Leney-Greene[2], Hui Ma[1,8], Cristin McCabe[1], Daniel E. McLoughlin[38], Eric M. Miller[34],
Christoph Muus[2,41], Mari Niemi[23], Robert Padera[40,42,43], Liuliu Pan[34], Deepti Pant[26,30,31],
Carmel Pe'er[1], Jenna Pfiffner-Borges[1], Christopher J. Pinto[16,38], Jacob Plaisted[40],
Jason Reeves[34], Marty Ross[34], Melissa Rudy[2], Erroll H. Rueckert[34], Michelle Siciliano[40],
Alexander Sturm[22], Ellen Todres[1], Avinash Waghray[44,45], Sarah Warren[34], Shuting Zhang[22],
Daniel R. Zollinger[34], Lisa Cosimi[46], Rajat M. Gupta[2,33], Nir Hacohen[2,9,47], Hanina Hibshoosh[25],
Winston Hide[10,17,18,19], Alkes L. Price[14], Jayaraj Rajagopal[38], Purushothama Rao Tata[48],
Stefan Riedel[10,17], Gyongyi Szabo[2,10,26], Timothy L. Tickle[1,12], Patrick T. Ellinor[49,65],
Deborah Hung[22,50,51,65], Pardis C. Sabeti[2,32,52,53,54,65], Richard Novak[55,65], Robert Rogers[26,56,65],
Donald E. Ingber[41,55,57,65], Z. Gordon Jiang[10,26,27,65], Dejan Juric[16,38,65], Mehrtash Babadi[12,13,65],
Samouil L. Farhi[1,65], Benjamin Izar[20,21,58,59,65], James R. Stone[36,65],
Ioannis S. Vlachos[2,10,17,18,19,65]✉, Isaac H. Solomon[40,65], Orr Ashenberg[1,65],
Caroline B. M. Porter[1,65], Bo Li[1,8,16,65], Alex K. Shalek[2,3,4,5,6,7,10,44,60,61,62,65]✉,
Alexandra-Chloé Villani[2,8,9,16,65]✉, Orit Rozenblatt-Rosen[1,63,65]✉ & Aviv Regev[1,5,53,63,65]✉

[1]Klarman Cell Observatory, Broad Institute of MIT and Harvard, Cambridge, MA, USA. [2]Broad Institute of MIT and Harvard, Cambridge, MA, USA. [3]Program in Health Sciences & Technology, Harvard Medical School and Massachusetts Institute of Technology, Boston, MA, USA. [4]Institute for Medical Engineering & Science, Massachusetts Institute of Technology, Cambridge, MA, USA. [5]Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology, Cambridge, MA, USA. [6]Ragon Institute of MGH, MIT, and Harvard, Cambridge, MA, USA. [7]Harvard Graduate Program in Biophysics, Harvard University, Cambridge, MA, USA. [8]Center for Immunology and Inflammatory Diseases, Department of Medicine, Massachusetts General Hospital, Boston, MA, USA. [9]Center for Cancer Research, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA. [10]Harvard Medical School, Boston, MA, USA. [11]Massachusetts Institute of Technology, Cambridge, MA, USA. [12]Data Sciences Platform, Broad Institute of MIT and Harvard, Cambridge, MA, USA. [13]Precision Cardiology Laboratory, Broad Institute of MIT and Harvard, Cambridge, MA, USA. [14]Department of Epidemiology, Harvard T. H. Chan School of Public Health, Harvard University, Boston, MA, USA. [15]Division of Infectious Diseases, Department of Medicine, Massachusetts General Hospital, Boston, MA, USA. [16]Department of Medicine, Harvard Medical School, Boston, MA, USA. [17]Department of Pathology, Beth Israel Deaconess Medical Center, Boston, MA, USA. [18]Harvard Medical School Initiative for RNA Medicine, Boston, MA, USA. [19]Cancer Research Institute, Beth Israel Deaconess Medical Center, Boston, MA, USA. [20]Department of Medicine, Division of Hematology/Oncology, Columbia University Irving Medical Center, New York, NY, USA. [21]Columbia Center for Translational Immunology, New York, NY, USA. [22]Infectious Disease and Microbiome Program, Broad Institute of MIT and Harvard, Cambridge, MA, USA. [23]Institute for Molecular Medicine Finland, Helsinki, Finland. [24]Analytical & Translational Genetics Unit, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA. [25]Department of Pathology and Cell Biology, Columbia University Irving Medical Center, New York, NY, USA. [26]Department of Medicine, Beth Israel Deaconess Medical Center, Boston, MA, USA. [27]Division of Gastroenterology, Hepatology and Nutrition, Department of Medicine, Beth Israel Deaconess Medical Center, Boston, MA, USA. [28]Department of Systems Biology, Harvard Medical School, Boston, MA, USA. [29]FAS Center for Systems Biology, Department of Organismic and Evolutionary Biology, Cambridge, MA, USA. [30]Division of Endocrinology, Diabetes, and Metabolism, Beth Israel Deaconess Medical Center, Boston, MA, USA. [31]Boston Nutrition and Obesity Research Center Functional Genomics and Bioinformatics Core Boston, Boston, MA, USA. [32]Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA, USA. [33]Divisions of Cardiovascular Medicine and Genetics, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA. [34]NanoString Technologies Inc, Seattle, WA, USA. [35]Department of Radiology, Beth Israel Deaconess Medical Center, Boston, MA, USA. [36]Department of Pathology, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA. [37]Department of Genetics and BASE Initiative, Stanford University School of Medicine, Stanford, CA, USA. [38]Massachusetts General Hospital Cancer Center, Department of Medicine, Massachusetts General Hospital, Boston, MA, USA. [39]Center for Genetic Epidemiology, Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA. [40]Department of Pathology, Brigham and Women's Hospital, Boston, MA, USA. [41]John A. Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge, MA, USA. [42]Harvard-MIT Division of Health Sciences and Technology, Cambridge, MA, USA. [43]Department of Pathology, Harvard Medical School, Boston, MA, USA. [44]Harvard Stem Cell Institute, Cambridge, MA, USA. [45]Center for Regenerative Medicine, Massachusetts General Hospital, Boston, MA, USA. [46]Infectious Diseases Division, Department of Medicine, Brigham and Women's Hospital, Boston, MA, USA. [47]Department of Medicine, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA. [48]Department of Cell Biology, Duke University School of Medicine, Durham, NC, USA. [49]Cardiovascular Disease Initiative, Broad Institute of MIT and Harvard, Cambridge, MA, USA. [50]Department of Genetics, Harvard Medical School, Boston, MA, USA. [51]Department of Molecular Biology and Center for Computational and Integrative Biology, Massachusetts General Hospital, Boston, MA, USA. [52]Department of Immunology and Infectious Diseases, Harvard T. H. Chan School of Public Health, Harvard University, Boston, MA, USA. [53]Howard Hughes Medical Institute, Chevy Chase, MD, USA. [54]Massachusetts Consortium on Pathogen Readiness, Boston, MA, USA. [55]Wyss Institute for Biologically Inspired Engineering, Harvard University, Boston, MA, USA. [56]Massachusetts General Hospital, Boston, MA, USA. [57]Vascular Biology Program and Department of Surgery, Boston Children's Hospital, Harvard Medical School, Boston, MA, USA. [58]Herbert Irving Comprehensive Cancer Center, Columbia University Irving Medical Center, New York, NY, USA. [59]Program for Mathematical Genomics, Columbia University Irving Medical Center, New York, NY, USA. [60]Program in Computational & Systems Biology, Massachusetts Institute of Technology, Cambridge, MA, USA. [61]Program in Immunology, Harvard Medical School, Boston, MA, USA. [62]Department of Chemistry, Massachusetts Institute of Technology, Cambridge, MA, USA. [63]Present address: Genentech, South San Francisco, CA, USA. [64]These authors contributed equally: Toni M. Delorey, Carly G. K. Ziegler, Graham Heimberg, Rachelly Normand, Yiming Yang, Åsa Segerstolpe, Domenic Abbondanza, Stephen J. Fleming, Ayshwarya Subramanian, Daniel T. Montoro, Karthik A. Jagadeesh, Kushal K. Dey, Pritha Sen, Michal Slyper, Yered H. Pita-Juárez, Devan Phillips, Jana Biermann. [65]These authors jointly supervised this work: Patrick T. Ellinor, Deborah Hung, Pardis C. Sabeti, Richard Novak, Robert Rogers, Donald E. Ingber, Z. Gordon Jiang, Dejan Juric, Mehrtash Babadi, Samouil L. Farhi, Benjamin Izar, James R. Stone, Ioannis S. Vlachos, Isaac H. Solomon, Orr Ashenberg, Caroline B. M. Porter, Bo Li, Alex K. Shalek, Alexandra-Chloé Villani, Orit Rozenblatt-Rosen, Aviv Regev.

Reporting summary

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

Code availability

All samples were initially processed using Cumulus (https://github.com/klarman-cell-observatory/cumulus), which we ran on the Terra Cloud platform (https://app.terra.bio/). Code for all other analyses is available on GitHub (https://github.com/klarman-cell-observatory/covid19-autopsy).

Data availability

Processed sequencing data (sc/snRNA-Seq and bulk) are available in the Gene Expression Omnibus (GEO, https://www.ncbi.nlm.nih.gov/geo/) under accession number GSE171668 and raw human sequencing data are available in the controlled access repository DUOS (https://duos.broadinstitute.org/), under dataset IDs DUOS-000126, DUOS-000127, DUOS-000128 and DUOS-000129. Viral genome assemblies and short-read sequencing data are publicly available in NCBI's GenBank and SRA databases, respectively, under BioProject PRJNA720544. GenBank accessions for SARS-CoV-2 genomes are MW885875–MW885883. Data for other tissues in the biobank will be released as they are acquired. Further queries can be addressed to: covid19-autopsy-northeast@broadinstitute.org .

The processed data are available on the Single Cell Portal: Lung, https://singlecell.broadinstitute.org/single_cell/study/SCP1052/.

Heart, https://singlecell.broadinstitute.org/single_cell/study/SCP1216/.

Kidney, https://singlecell.broadinstitute.org/single_cell/study/SCP1214/.

Liver, https://singlecell.broadinstitute.org/single_cell/study/SCP1213/.

Nanostring GeoMx raw and normalized count matrices are available in GEO under accession number GSE163530. Raw images will be available upon request. Source data are provided with this paper.

52. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15 (2018).
53. Korsunsky, I. et al. Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* **16**, 1289–1296 (2019).
54. Schiller, H. B. et al. The Human Lung Cell Atlas: a high-resolution reference map of the human lung in health and disease. *Am. J. Respir. Cell Mol. Biol.* **61**, 31–41 (2019).
55. Brill, B., Amir, A. & Heller, R. Testing for differential abundance in compositional counts data, with application to microbiome studies. Preprint at https://arxiv.org/abs/1904.08937 (2019).
56. Ravindra, N. G. et al. Single-cell longitudinal analysis of SARS-CoV-2 infection in human airway epithelium. Preprint at https://doi.org/10.1101/2020.05.06.081695 (2020).

Author contributions A.K.S., A.-C.V., I.S.V., Z.G.J., O.R.-R. and A.R. conceived and led the study. These authors contributed equally as co-second authors: Z.B.-A., N.B., A.G., J.G., J.C.M., I.K., E.N., P.N., Y.V.P., S.S.R., S.N., L.T.-Y.T., K.J.S., M.Su., V.M.T., S.K.V., Y.W. T.M.D., C.G.K.Z., Å.S., D.A. designed protocols and carried out experiments together with D.P., Z.B.-A., V.M.T., A.S., S.Z., J.G., J.H., E.N., M.su., C.M., L.T., A.E., D.Pa., L.P., J.C.M., L.A.-Z. C.B.M.P., C.G.K.Z., O.A., R.N., G.H., K.J., K.S., B.L., Y.Y., S.F., A.S., P.N., Y.P.J., P.N., T.He., J.R., W.H., I.S.V., T.M.D. and J.B. designed and performed computational analysis. M.B., N.B., J.G., Y.W., R.G., S.S.R., H.M., P.S., A.W., C.M., M.L.-G., T.H., D.T.M., S.W., D.R.Z., E.R., M.R., E.M., R.F., P.Di., A.G., C.Pe. and M.N. provided input to and assisted with computational analysis. K.J., K.D., A.G., J.E., S.G., A.R. and A.P. contributed methods and performed integrated analysis for GWAS. P.R.T., D.T.M., Z.G.J., Y.P., G.S., S.N., S.R. and J.R. provided clinical and biological expertise. D.F., D.J., D.E.M., C.P., S.K.V., E.K. and J.S. provided clinical expertise, performed sample acquisition and/or administrative coordination at MGH. J.H., R.R., R.N., O.R.B., Z.G.J., Y.P. and D.I. provided clinical expertise, performed sample acquisition and/or administrative coordination at BIDMC. I.H.S., D.A., L.C., J.G., R.P., M.Si., provided clinical expertise, performed sample acquisition and/or administrative coordination at BWH. I.G., H.H. and B.I. provided clinical expertise, performed sample acquisition and/or administrative coordination at CUIMC/NYP. P.D., D.P., J.J-V. and J.B. helped with sample coordination and sample receipt at the Broad Institute. N.B and S.R. performed bulk RNA-Seq deconvolution analysis. E.N., M.R. and K.S. performed viral qPCR, whole-genome sequencing and phylogenetic analyses. M.S. provided input for sc/snRNA-Seq experiments and protocols. J.J.-V., E.T., O.R.R. and A.-C.V. managed the study and tissue acquisition. T.L.T. contributed computational expertise and advice. T.M.D., C.B.M.P., C.G.K.Z., G.H., R.N., K.J., O.A., B.L., Z.G.J., I.S.V, Y.Y., S.F., A.S., D.T.M., A.K.S., A.-C.V., O.R.-R. and A.R. wrote the manuscript, with input from all authors. D.H., P.C.S., N.H. and P.T.E. supervised research.

Article

Extended Data Fig. 1 | See next page for caption.

Extended Data Fig. 1 | A COVID-19 autopsy cohort, data quality and ambient RNA removal for a single-cell/nucleus lung atlas. a, COVID-19 cohort overview. IMV, intermittent mandatory ventilation days; PMI, post-mortem interval; S/s, time from symptom start to death in days. **b**–**d**, Comparison of cell composition by scRNA-Seq and snRNA-Seq in matched samples. Proportion of cells (*x* axis) of each type (colour code) in sc/snRNA-Seq samples from the same three donors (D3, D8, D12). **e**–**h**, Cellbender remove-background on a single sample (D1). **e**, CellBender improves cell clustering and expression specificity by removing ambient RNA and empty (non-cell) droplets. UMAP plot of snRNA-Seq profiles (dots) either before (left) or after (right) CellBender processing, coloured by clusters, with CellBender-determined empty droplets in black ($k = 2{,}508$ droplets removed, $k = 10{,}687$ cells remaining). **f**, **g**, CellBender improves specificity of individual genes and cell type signatures. UMAP embedding of single nucleus profiles before CellBender (left) and after CellBender (right) processing, coloured by expression of the surfactant protein gene *SFTPA1* (**f**) or signature score (SCANPY[52] score_genes function, colour bar) for gene sets specific to lung AT2 (**g**) cells. Colour bar saturation chosen to emphasize low expression. **h**, Improved specificity of surfactant gene expression with CellBender (same sample). Expression level (log (average unique molecular identifier (UMI) count per cell), colour) and percentage of cells with non-zero expression (dot size) of surfactant genes (columns) across cell clusters (rows) before (left) and after (right) CellBender processing. Also shown, for comparison, are the results of an alternative method, DecontX (middle).

Article
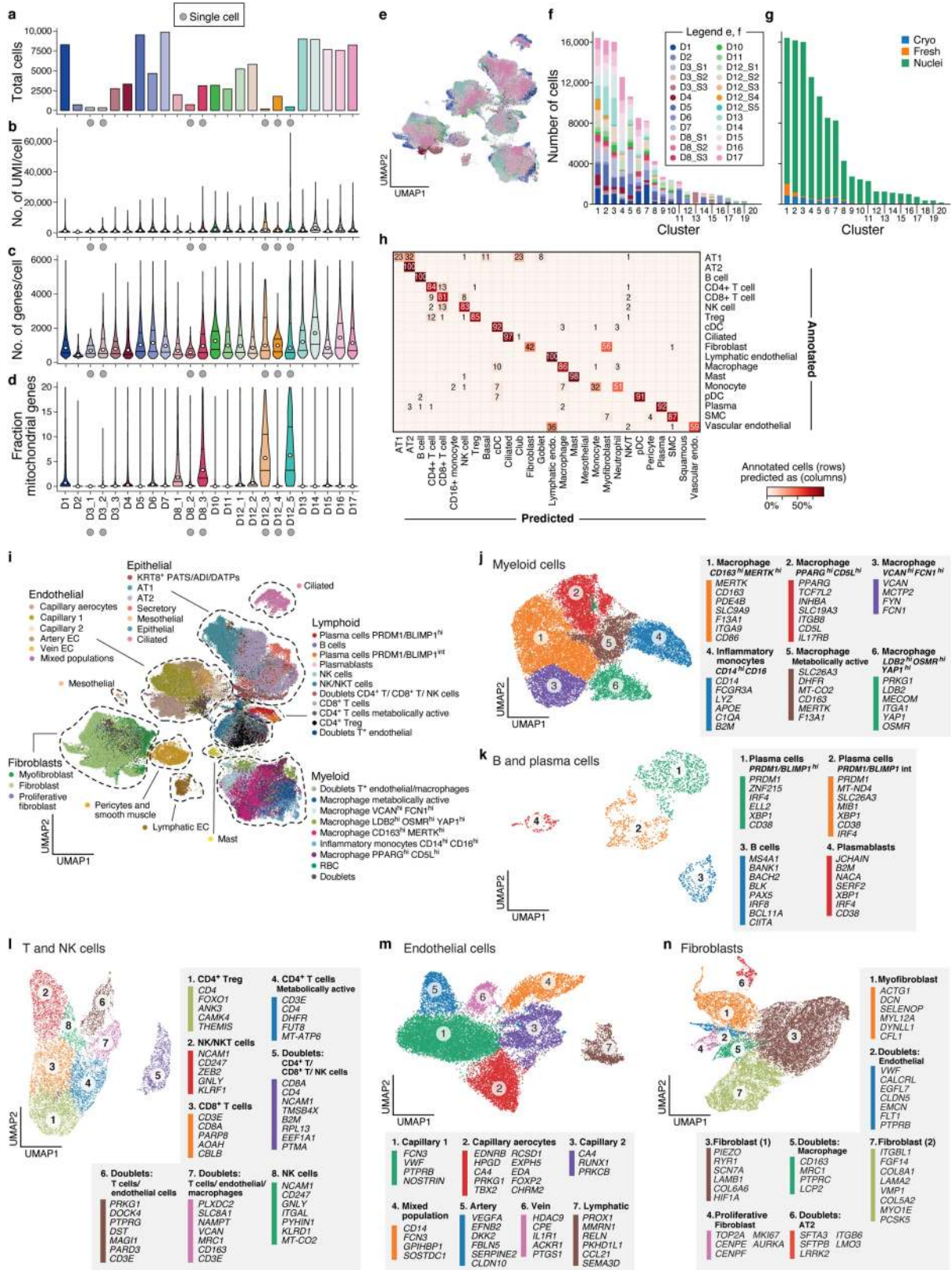
Extended Data Fig. 2 | See next page for caption.

Extended Data Fig. 2 | Quality control and annotation in the COVID-19 lung cell atlas. a–**d**, Quality-control metrics for 24 lung samples (*n* = 16 donors). Number of cells or nuclei (**a**, *y* axis) and distributions (median and first and third quartiles) of number of UMIs per cell or nucleus (**b**, *y* axis), number of genes per cell/nucleus (**c**, *y* axis) and fraction of mitochondrial genes per cell/nucleus (**d**, *y* axis) across the samples (*x* axis) in the lung atlas. ScRNA-Seq samples are labelled by a grey circle. **e**–**g**, Cross-sample integration corrects batch effects. **e**, UMAP (as in Fig. 2a) of 106,792 sc/snRNA-Seq profiles after Harmony[53] correction (Supplementary Methods) coloured by sample ID. **f**, **g**, Donors and processing protocols across clusters. Number of cells (*y* axis) from different donors (**f**) or processing protocols (**g**) in each Leiden cluster (*x* axis). **h**, Cross validation of automatic annotation. Percentage of cells (colour bar) annotated in a class by Schiller et al.[54] that we predict for each class (columns). **i**, Identification of main lineage annotations by manual annotation. UMAP of 106,792 sc/snRNA-Seq profiles after Harmony[53] correction (as in Fig. 2a) coloured by manual annotation done in subclustering of each lineage. Dashed lines: chosen compartments for subclustering. **j**–**n**, Refined annotation of cell subsets within lineages. UMAP embeddings of each selected cell lineage with cells coloured by manually annotated subclusters. Colour legends highlight highly expressed marker genes for select subsets. **j**, myeloid cells (*k* = 24,417 cells/nuclei); **k**, B and plasma cells (*k* = 1,693); **l**, T and NK cells (*k* = 9,950); **m**, endothelial cells (*k* = 20,366); and **n**, fibroblast (*k* = 20,925).
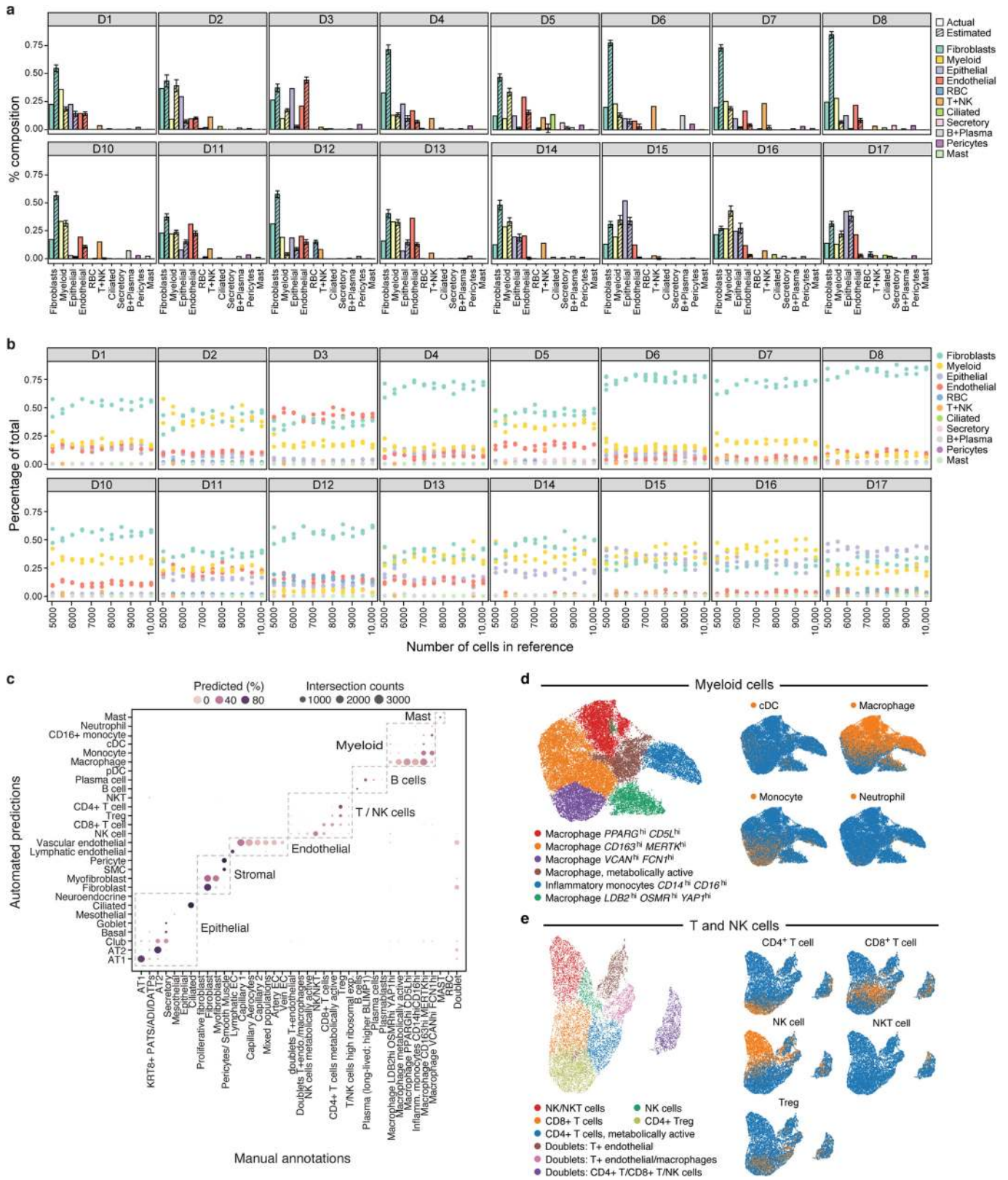
Extended Data Fig. 3 | See next page for caption.

Extended Data Fig. 3 | Bulk RNA-Seq deconvolution and comparison of automatic and manual annotations in the COVID-19 lung cell atlas.
a, **b**, Deconvolution of bulk RNA-Seq libraries from adjacent lung tissue.
a, Mean proportion (*y* axis, error bars = s.d. estimates from bulk RNA-Seq deconvolution (hatched bars; from MuSiC[55]) and from sc/snRNA-Seq (filled bars) for each of 11 cell subsets (*x* axis) in each of 16 bulk RNA-Seq lung samples (panels) from 10 random samples of 10,000 cells each. **b**, Robustness of cell proportion estimates to the number of single cells sampled for the reference data. Mean proportion (*y* axis, from MuSiC) estimates for each of 11 cell subsets (colour dots) in each of 16 bulk RNA-Seq lung samples (panels) when using three independent samples of 1,000–10,000 cells from the single-cell reference (*x* axis). **c**–**e**, Agreement between automated and manual annotations. **c**. High consistency between automatic and manual annotations. The proportion (colour intensity) and number (dot size) of cells with a given predicted annotation (rows) in each manual annotation category (columns). **d**, **e**, UMAP embedding of myeloid (*k* = 24,417 cells or nuclei) (**d**) and T and NK (*k* = 9,950 cells); (**e**) cell profiles coloured by manually annotated subclusters (left) or automated predictions (right).
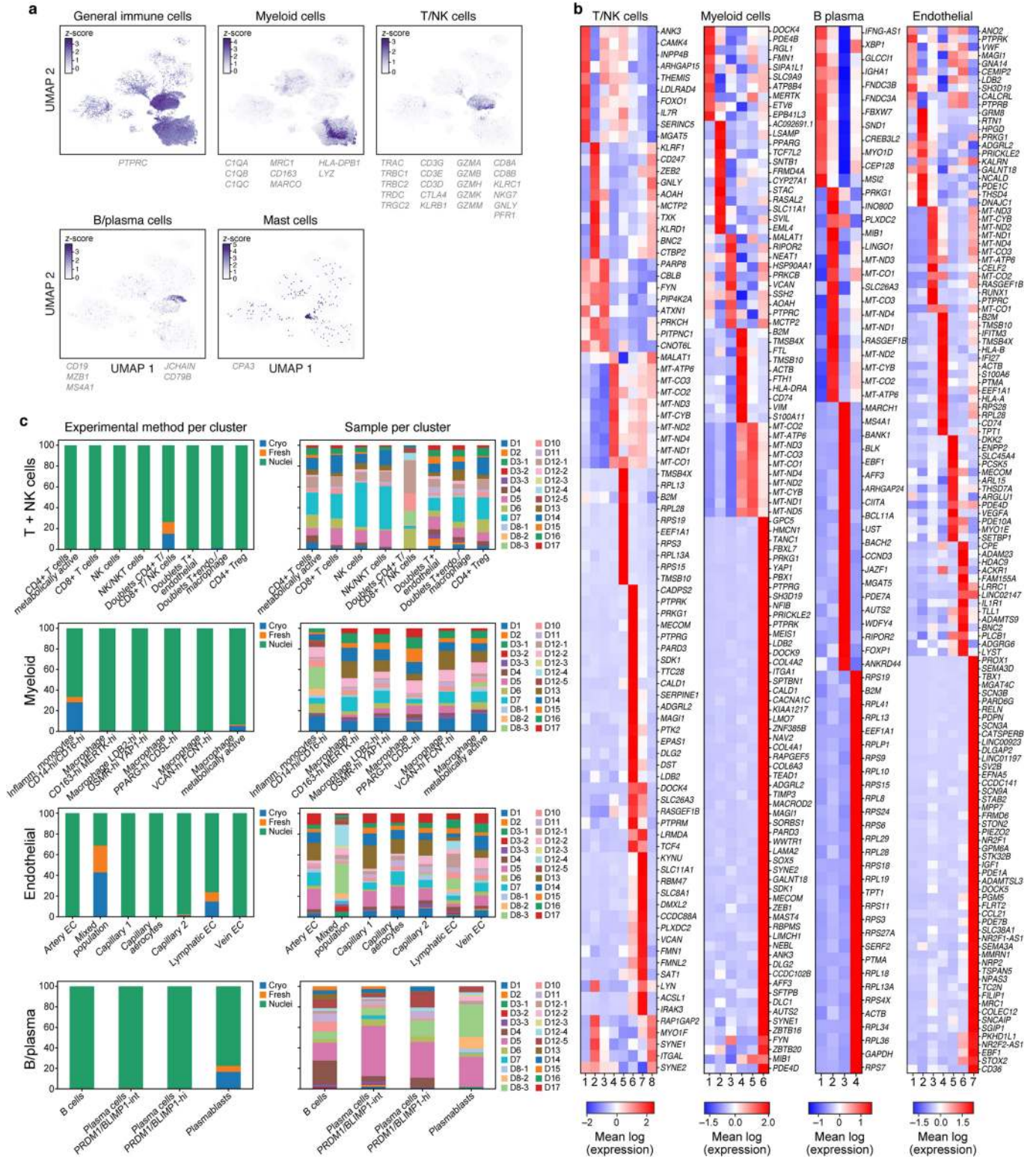
Extended Data Fig. 4 | Manual annotation in the COVID-19 lung cell atlas.
a, **b**, Identification of main immune lineage annotations. **a**, UMAP of 106,792 sc/
snRNA-Seq profiles after Harmony correction (as in Fig. 2a) coloured by
expression of genes (colour bar, genes listed below) used to separate immune
cell sub-lineages (Supplementary Methods). **b**, Differentially expressed (DE)
genes between subclusters within each lineage. Expression (colour bar) of
genes (rows) that are differentially expressed (Supplementary Methods) across
the subclusters (columns) within each compartment. DE genes shown are a

union of the following: (i) top 10 DE genes between clusters, (ii) DE genes above
an AUC of 0.8 and 0.75 for B/Plasma cells, (iii) pseudo-bulk DE genes above a
log(fold change) threshold (thresholds: endothelial = 4.2, T/NK = 3, myeloid = 4,
B/plasma = 2) (label on top). **c**, Batch correction within lineage. Fraction of
cells/nuclei (y axis) from different processing protocols (left) or different
donors (right, n = 17) in each subcluster (x axis) after batch correction with
Harmony[53] within each lineage.

a Frequencies of myeloid subsets across donors

Legend:
- Macrophage LDB2[hi] OSMR[hi] YAP1[hi]
- Macrophage metabolically active
- Macrophage PPARG[hi] CD5L[hi]
- Inflammatory monocytes CD14[hi]CD16[hi]
- Macrophage CD163[hi] MERTK[hi]
- Macrophage VCAN[hi] FCN1[hi]
- MAST
- RBC

b Cellular compartmental compositions

Legend:
- Epithelial
- Stromal
- Endothelial
- Myeloid
- Lymphoid

c 1580 genes, increased expression

587, 317, 309, 307, 274, 254, 234, 169, 167, 127, 122, 114, 95, 87, 71, 44, 43, 38

d KRT8/PATS — IPBLPs / Other — TP63 — KRT5 — KRT17 — KRT8

Airway cells — Basal cells / Other — TP63 — KRT5 — KRT8

Extended Data Fig. 5 | Cell-intrinsic programs and epithelial regenerative cell states in the COVID-19 lung cell atlas. a, **b**, Differences in cell composition across donors. Percentage of cells (*y* axis) from each myeloid subset (legend) in each donor (*x* axis). **b**, Percentage of cells (*y* axis) from each main lineage (legend) in each donor (*x* axis), rank ordered by proportion of epithelial cells (blue). **c**, Myeloid, endothelial and pneumocyte cells show substantial changes in cell intrinsic expression profiles in COVID-19 lung. log₂(fold change) (*y* axis) between COVID-19 and healthy lung for each elevated gene (dot) in each cell

subset (*x* axis, by automatic annotation). Black bars, number of genes with significantly increased expression (adjusted *P* < 7.5 × 10⁻⁶). Computed using a single cell-based differential expression model applied to a meta-differential expression analysis between COVID-19 and healthy samples across 14 studies (see Supplementary Methods). **d**, PATS and IBPLP cells in COVID-19 lung. UMAP embeddings of 1,550 *KRT8*⁺ PATS-expressing cells (top) or of 1,394 airway epithelial cells (bottom) coloured by IPBLP cells or basal cells (orange, leftmost panels) or characteristic markers (purple, remaining panels).
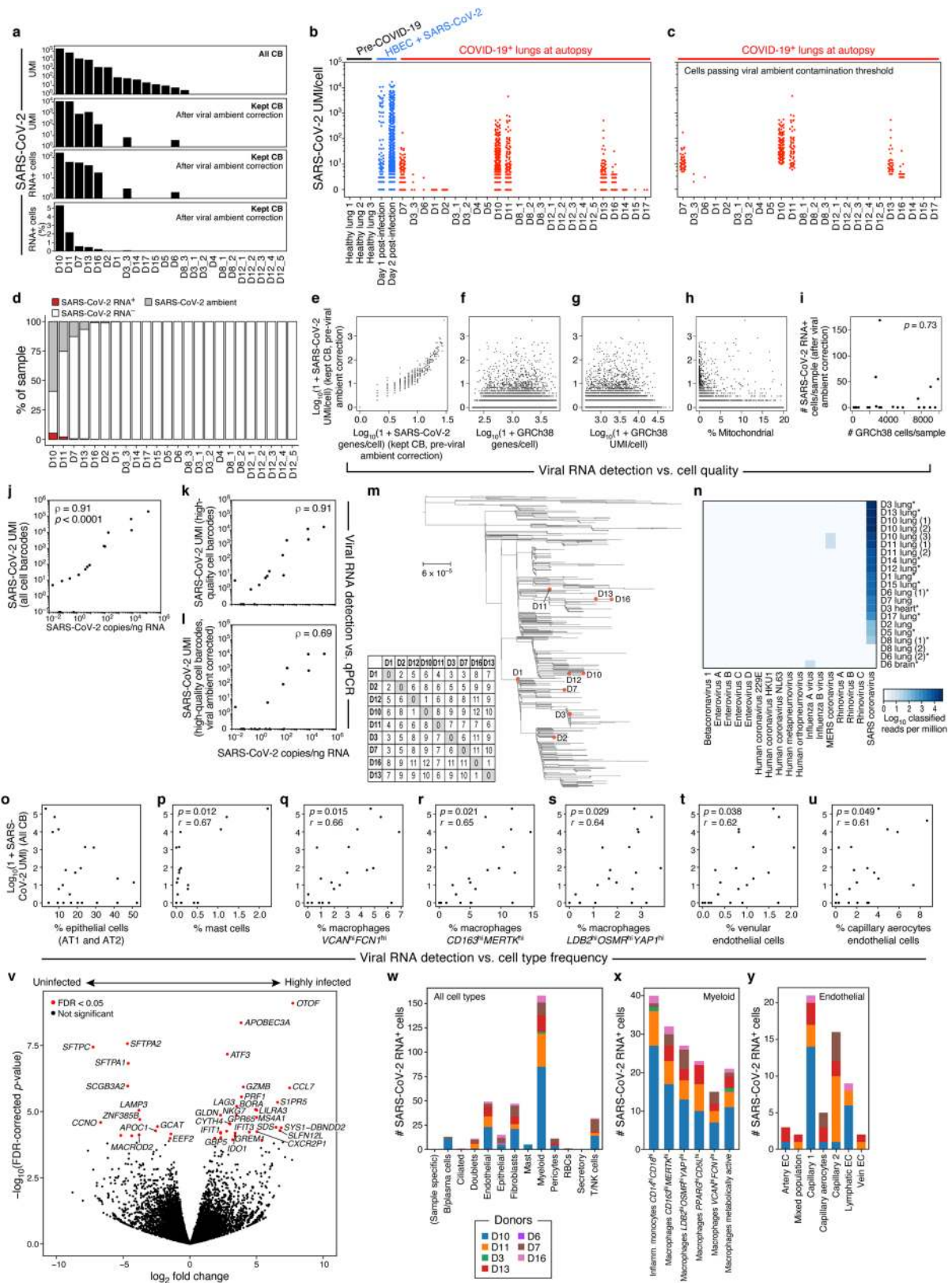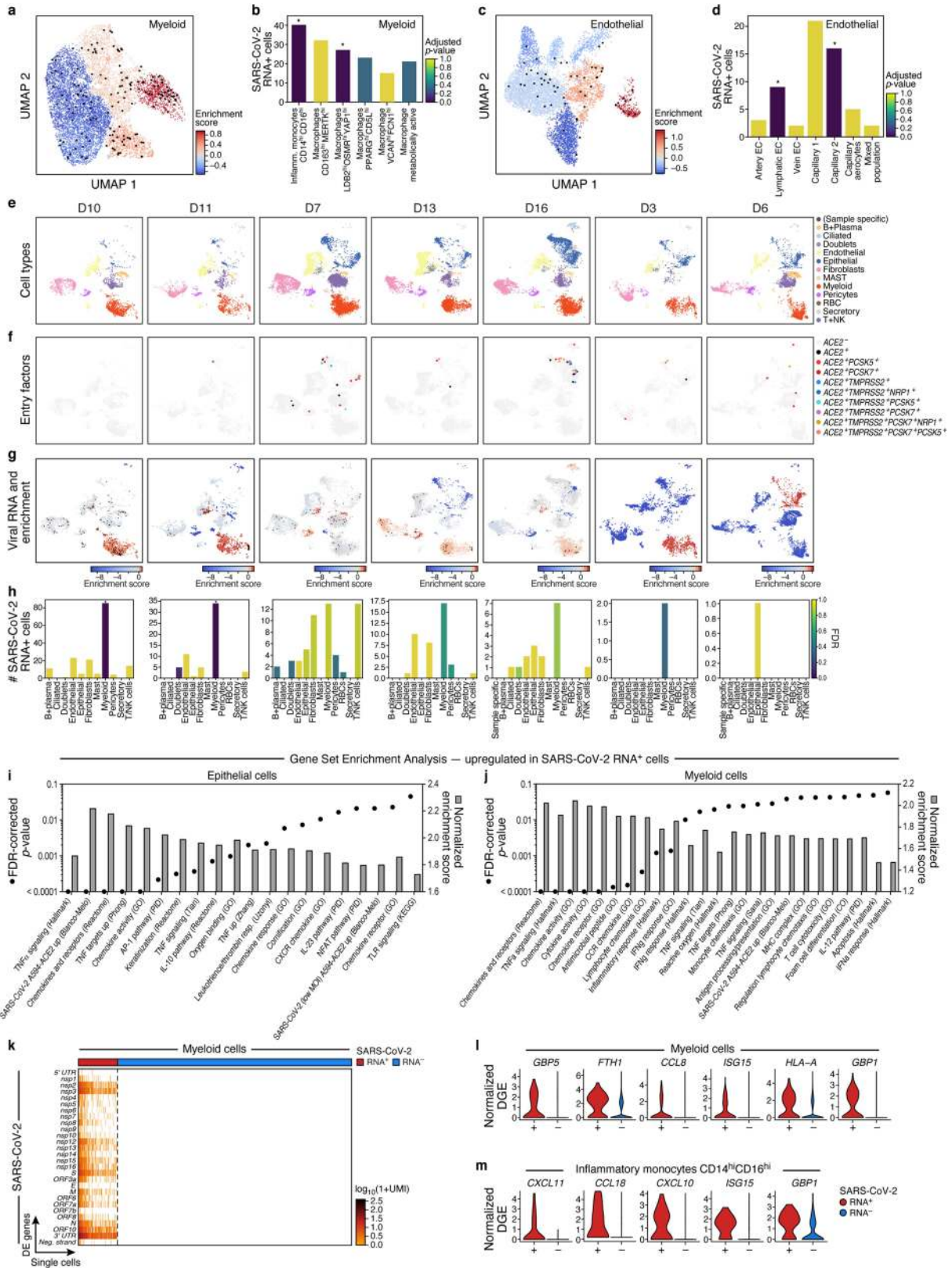
Extended Data Fig. 6 | See next page for caption.

Extended Data Fig. 6 | SARS-CoV-2-RNA+ cells distinguished by sc/snRNA-Seq. a, Detection of SARS-CoV-2 UMIs from sc/snRNA-Seq data. SARS-CoV-2 UMIs from all cell barcodes (top) and after ambient correction (second from top). Number (second from bottom) and percentage (bottom) of SARS-CoV-2 RNA+ cells after ambient correction ($m = 24$ specimens). **b, c**, Effect of ambient RNA on SARS-CoV-2 RNA+ detection. Number of SARS-CoV-2 aligning UMI per cell barcode (CB) (y axis) in healthy lung (**b**, black), in vitro SARS-CoV-2 infected human bronchial epithelial cells (HBEC)[56] (**b**, blue) or lung samples from COVID-19 donors at autopsy either with CB with high-quality capture of human mRNA (**b**, red) or after removal of cells whose viral alignments were attributed to ambient contamination (**c**, Supplementary Methods). **d**, Variation in SARS-CoV-RNA+ cells across donors. Percentage of cells (y axis) assigned as SARS-CoV-2 RNA− (white), SARS-CoV-2 RNA+ (red) or SARS-CoV-2 ambient (grey, Supplementary Methods) across the donors (x axis), sorted by proportion of SARS-CoV-RNA+ cells. **e–i**, Viral RNA detection does not correlate with cell quality metrics. **e–h**, Number of SARS-CoV-2 UMIs (before ambient viral correction) for each cell (y axis) versus either number of SARS-CoV-2 genes for that cell (**e**, x axis), number of human (GRCh38) genes per cell (**f**, x axis), number of human (GRCh38) UMI per cell (**g**, x axis) or percentage of human (GRCh38) mitochondrial UMIs per cell (**h**, x axis). **i**, Number of retained high-quality cells (x axis) and number of SARS-CoV-2 RNA+ cells (y axis) in each sample (dots) after correction for ambient viral reads. Pearson's $r = 0.07$, two-sided $P = 0.73$. **j–l**, Agreement in viral RNA detection between qPCR and sn/scRNA-Seq. Number of SARS-CoV-2 copies measured by CDC N1 qPCR on bulk RNA extracted from matched tissue samples (x axis) and the number of SARS-CoV-2

aligning UMI (y axis) for each sample (dot) from all reads (**j**, $P < 0.0001$, two-sided), all reads from high-quality cell barcodes (**k**, $P < 0.0001$), and after viral ambient RNA correction (**l**, $P = 0.0042$). Spearman's ρ reported, two-sided test. **m**, Genetic diversity of SARS-CoV-2. Maximum likelihood phylogenetic tree of 772 SARS-CoV-2 genomes from cases in Massachusetts between January and May 2020. Orange points, donors in this cohort. **n**, Specificity of SARS-CoV-2 infection. $\log_{10}(1+$ reads) in each donor (columns) assigned to different viruses (rows) by metagenomic classification using Kraken2 from bulk RNA-Seq. Asterisks denote targeted capture. **o–u**, Relation between SARS-CoV-2 RNA and different cell types. Number of SARS-CoV-2 aligning UMIs in each (including all CB) and the proportion of epithelial (**o**), mast (**p**), macrophage *VCAN^high^FCN1^high^* (**q**), macrophages *CD163^high^MERTK^high^* (**r**), macrophages *LDB2^high^OSMR^high^YAP1^high^* (**s**), venular endothelial (**t**) or capillary aerocytes (**u**) cells in these samples (x axes). Pearson's r denoted in the upper left corner with significance after Bonferroni correction (P). **v**, Effect of viral load on bulk RNA profiles. Significance ($-\log_{10}(P)$, y axis) and magnitude ($\log_2$(fold change), x axis) of differential expression of each gene (dots) between three donors with highest viral load and six donors with lowest or undetectable viral load profiled by bulk RNA-Seq. Red points, FDR < 0.05. **w–y**, Distribution of SARS-CoV-2 RNA+ cells across cell types and subsets. Number of SARS-CoV-2 RNA+ cells (y axis) from each donor (colour) across major categories (**w**, x axis), myeloid subsets (**x**, inflammatory monocytes: 40 cells, five donors; *LDB2^high^OSMR^high^YAP1^high^* macrophages: 27 cells, five donors; x axis), or endothelial subsets (**y**, capillary endothelial cells: 16 cells, four donors; lymphatic endothelial cells: nine cells, three donors; x axis).

Extended Data Fig. 7 | See next page for caption.

Extended Data Fig. 7 | Donor-specific enrichment of SARS-CoV-2 RNA⁺ cells and host responses to viral RNA. a–**d**, SARS-CoV-2 RNA⁺ cells are enriched in specific lineages and subtypes. **a**, **c**, UMAP embeddings of either myeloid cells (**a**), or endothelial cells (**c**) from seven donors containing any SARS-CoV-2 RNA⁺ cell, and coloured by viral enrichment score (colour bar; red, stronger enrichment) and by SARS-CoV-2 RNA⁺ cells (black points). **b**, **d**, Number of SARS-CoV-2 RNA⁺ cells (y axis) per cell type/subset (x axis) in myeloid (**b**) or endothelial (**d**) subsets. Bar colour, FDR (dark blue, higher significance, Supplementary Methods; *FDR < 0.05). **e**–**h**, Variation across donors. **e**–**g**, UMAP embeddings of sc/snRNA-Seq profiles from each of seven donors containing any SARS-CoV-2 RNA⁺ cell (columns), coloured by major cell categories (**e**), expression of SARS-CoV-2 entry factors (**f**) or SARS-CoV-2 RNA enrichment per cluster (**g**, red/blue colour bar; red, high enrichment; black

points, SARS-CoV-2 RNA⁺ cells). **h**, Number of SARS-CoV-2 RNA⁺ cells (y axis) across major cell types (x axis) from each of seven donors containing any SARS-CoV-2 RNA⁺ cell (columns). Bar colour, FDR (dark blue, higher significance). *FDR < 0.05. **i**, **j**, Normalized enrichment score (bars, right y axis) and significance (points, FDR, left y axis) (by GSEA[39,40], Supplementary Methods) of different functional gene sets (x axis) in genes upregulated in SARS-CoV-2 RNA⁺ epithelial (**i**) or myeloid (**j**) cells. **k**, Expression of SARS-CoV-2 genomic features (log-normalized UMI counts; rows) across SARS-CoV-2 RNA⁺ (k = 158 cells) and SARS-CoV-2 RNA⁻ (k = 790) myeloid cells (columns). **l**, **m**, Distribution of normalized expression levels (y axis) for select significantly differentially expressed genes between SARS-CoV-2 RNA⁻ and SARS-CoV-2 RNA⁺ cells from all myeloid cells or $CD14^{high}CD16^{high}$ inflammatory monocytes. DGE, differential gene expression.
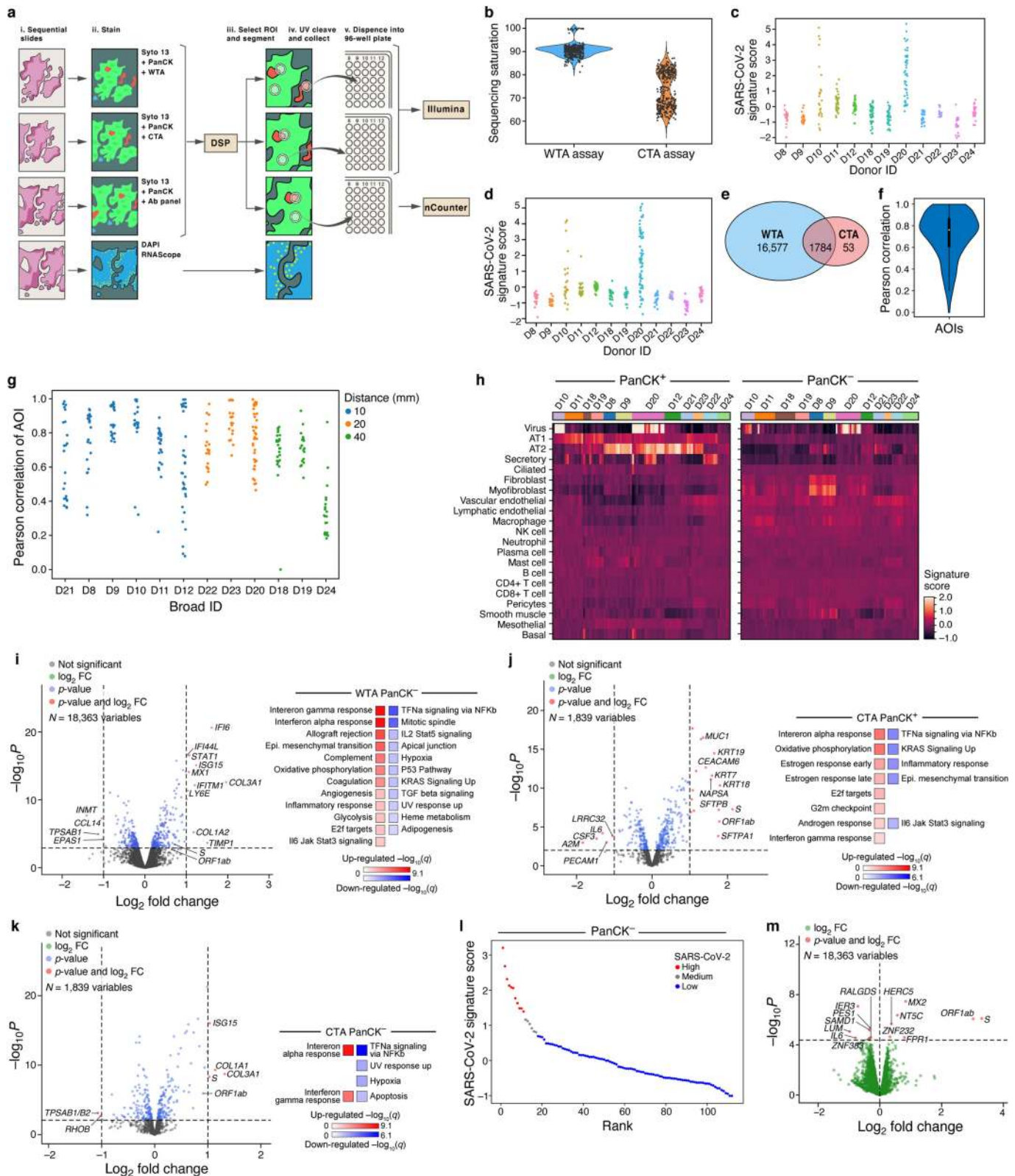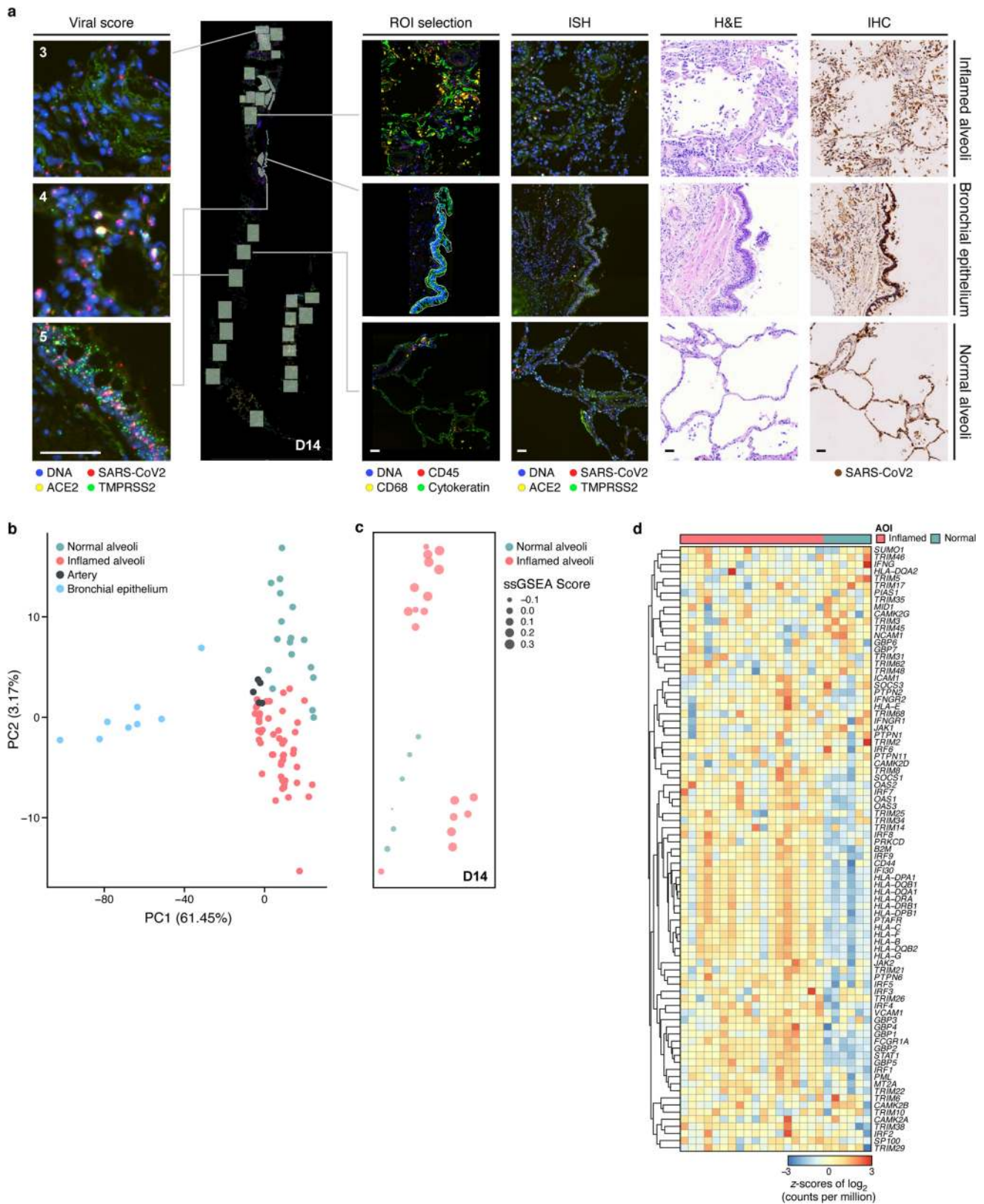
Extended Data Fig. 8 | See next page for caption.

Extended Data Fig. 8 | NanoString GeoMx experiment design and analysis.
a, Overview of spatial profiling experiments. **b**, Distribution of sequencing saturation (y axis, %) for WTA and CTA AOIs (x axis). **c, d**, SARS-CoV-2 signature score (y axis) for each WTA (**c**) and CTA (**d**) AOI (dots) from each donor (x axis). **e**, Overlap of WTA and CTA genes. **f, g**, Agreement between WTA and CTA. **f**, Distribution (box, interquartile range; white point, median; violin range, min–max) of Pearson correlation coefficients (y axis) between WTA and CTA profiles (for common genes across 296 AOIs). **g**, Pearson correlation coefficient (y axis) of WTA and CTA common genes for each AOI pair (dot) from each donor (x axis), sorted by distance between WTA and CTA sections (blue, 10 mm; orange, 20 mm; green, 40 mm). **h**, Cell composition differences between PanCK$^+$ and PanCK$^-$ alveolar AOIs and between AOIs from COVID-19 ($n = 9,161$ AOIs) and healthy (D22–24, 40 AOIs) lungs. Expression scores (colour bar) of cell type signatures (rows) in PanCK$^+$ (left) and PanCK$^-$ (right) alveolar AOIs (columns) in CTA data from different donors (top colour bar).

i–k, Differential gene expression in COVID-19 versus healthy lung. Left: significance ($-\log_{10}(P)$, y axis) and magnitude ($\log_2$(fold change), x axis) of differential expression of each gene (dots) in WTA for PanCK$^-$ (**i**, 112 COVID-19 versus 20 healthy), and in CTA for PanCK$^+$ (**j**, 69 COVID-19 versus 18 healthy) and PanCK$^-$ (**k**, 92 COVID-19 versus 22 healthy) alveoli. Horizontal dashed line, FDR = 0.05; vertical dashed lines, $|\log_2$(fold change)$| = 2$. Right: significance ($-\log_{10}(q)$) of enrichment (permutation test) of different pathways (rows). **l, m**, Changes in gene expression in SARS-CoV-2 high versus low AOIs within COVID-19 lungs in WTA data. **l**, PanCK$^-$ alveolar AOIs (dots) rank ordered by their SARS-CoV-2 signature score (y axis) in WTA data, and partitioned to high (red), medium (grey) and low (blue) SARS-CoV-2 AOIs. **m**, Significance ($-\log_{10}(P)$, y axis) and magnitude ($\log_2$(fold change), x axis) of differential expression of each gene (dots) in WTA data between SARS-CoV-2 high and low AOIs for PanCK$^-$ alveoli (ROIs: 11 high, six medium, 95 low). Horizontal dashed line, FDR = 0.05.

Extended Data Fig. 9 | See next page for caption.

**Extended Data Fig. 9 | GeoMx WTA DSP analysis of lung biopsy samples reveals region- and inflammation-specific expression programs. a**, Region selection. Serial sections of lung biopsy samples (five donors, D13–17; image depicts serial sections of D14) processed with GeoMx WTA DSP with four-colour staining (DNA, CD45, CD68, PanCK), RNAscope with probes against (SARS-CoV-2 *S* gene (used to derive semiquantitative viral load scores), *ACE2*, *TMPRSS2*), H&E staining and immunohistochemistry (IHC) with anti-SARS-CoV-2 S-protein. Scale bar, 100 μm. **b**–**d**, Region- and inflammation-specific expression programs. **b**, The first two principal components (PCs, *x* and *y* axes) from lung ROI gene expression profiles from donors D13–17, spanning normal-appearing alveoli (green; D14 = 6 AOIs, D15 = 2 AOIs, D16 = 5 AOIs, D17 = 4 AOIs); inflamed alveoli (magenta; D13 = 14 AOIs, D14 = 18 AOIs, D15 = 7 AOIs, D16 = 3 AOIs, D17 = 8 AOIs); bronchial epithelium (blue; D14 = 2 AOIs, D15 = 1 AOI, D16 = 2 AOIs, D17 = 3 AOIs) and arterial blood vessels (black; D13 = 2 AOIs, D15 = 3 AOIs). **c**, GSEA score (circle size, legend) of the enrichment of the IFNγ pathway in each normal-appearing (green; 6 AOIs) and inflamed (magenta; 18 AOIs) alveolar AOIs (dot) from the section of donor D14 (in **a**), placed in their respective physical coordinates on the tissue section (as in **a**). **d**, Expression (colour bar, $\log_2$(counts per million)) of IFNγ pathway genes (rows) from normal-appearing (green, $n = 6$) and inflamed (magenta, $n = 18$) alveoli AOIs (columns) from D14 lung biopsy.
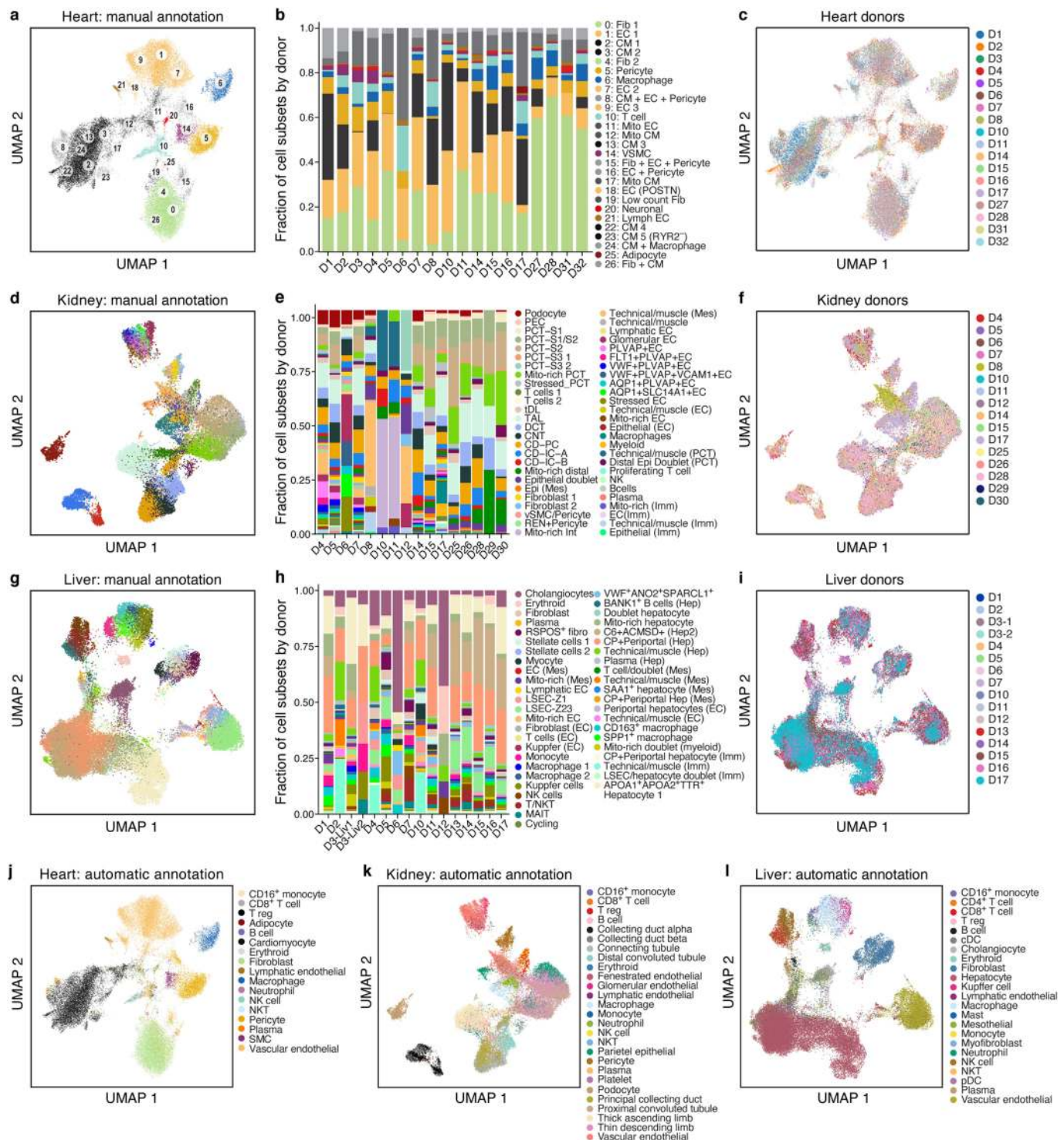
Extended Data Fig. 10 | A single-nucleus atlas of heart, kidney and liver COVID-19 tissues. a–**c**, COVID-19 heart cell atlas. UMAP embedding of 40,880 heart nuclei (dots) (*n* = 18 donors, *m* = 19 specimens) coloured by Leiden resolution 1.5 clustering with manual post hoc annotations (**a**) or donors (**c**). **b**, Proportions of cell types (*y* axis) in each sample. **d**–**f**, COVID-19 kidney cell atlas. UMAP embedding of 33,872 kidney nuclei (dots) (*n* = 16, *m* = 16) coloured by clustering with manual post hoc annotations (**d**) or donors (**f**). **e**, Proportion of cells (*y* axis) in each sample. **g**–**i**, COVID-19 liver cell atlas. **g**, **i**, UMAP embedding of 47,001 liver nuclei (dots) (*n* = 15, *m* = 16), coloured by clustering with manual post hoc annotations (**g**) or donors (**i**). **h**, Proportions of cell types (*y* axis) in each sample. **j**–**l**, Automatic annotations. UMAP embeddings, coloured by predicted cell type labels by automatic annotation for heart (**j**), kidney (**k**) and liver (**l**).
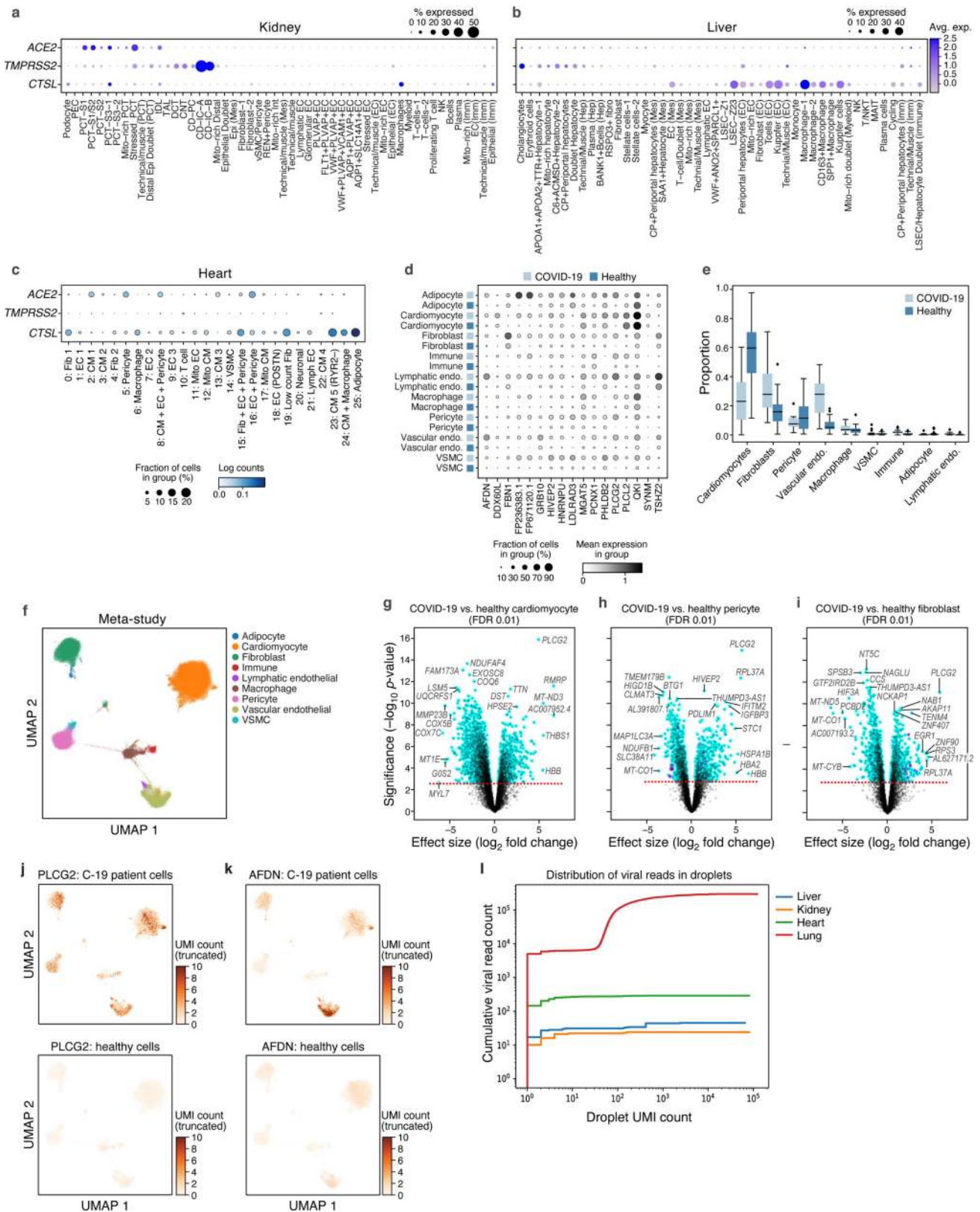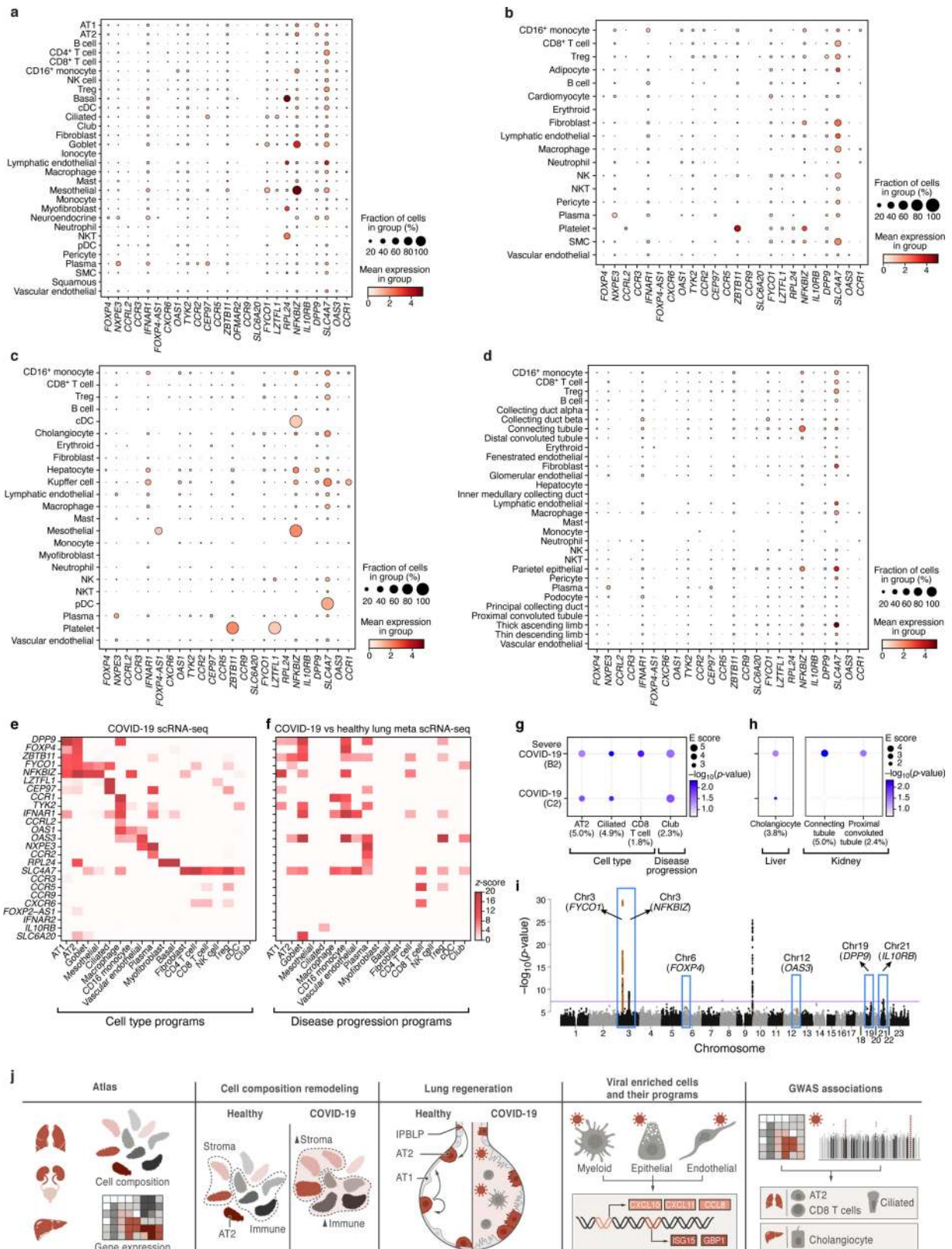
Extended Data Fig. 11 | See next page for caption.

Article

Extended Data Fig. 11 | Entry factors in heart, kidney and liver COVID-19 tissues and differential gene expression in heart cell atlas. a–**c**, SARS-CoV-2 entry factors are expressed in kidney, liver and heart cells. Average expression (dot colour) and fraction of expressing cells (colour, size) of SARS-CoV-2 entry factors (rows) across cell subsets (columns) in the kidney (**a**), liver (**b**) and heart (**c**). **d**–**k**, Genes and pathways differentially expressed between COVID-19 and healthy heart cells. **d**, log mean expression per cell (dot colour) and fraction of expressing cells (dot size) across cell types from healthy or COVID-19 heart (rows) for select genes (columns) that are differentially expressed between COVID-19 and healthy cells. **e**, Proportions of each cell type for COVID-19 (n = 15) and healthy (n = 28, two studies) samples (boxplots: middle line, mean; box bounds, first and third quartiles; whiskers, 1.5× the interquartile range; minima, smallest observed proportion; maxima, highest observed proportion). **f**, UMAP embedding of integrated COVID-19 and healthy snRNA-Seq profiles (dots) coloured by major cell types. Plot limited to a subset of 151,373 high-quality cells for visualization purposes. **g**–**i**, Cell-type-specific differentially expressed genes in COVID-19 versus healthy nuclei. Differential expression ($\log_2$(fold change), x axis), and associated significance ($-\log_{10}(P)$, y axis; Supplementary Methods) for each gene (dot) between COVID-19 and healthy nuclei of cardiomyocytes (**g**), pericytes (**h**) and fibroblasts (**i**). Dashed line, FDR = 0.01. **j**, **k**, UMAP embedding of the meta-analysis atlas (as in **f**) but showing only COVID-19 (top) or healthy (bottom) nuclei profiles (dots) coloured by expression of *PLCG2* (**j**) or *AFDN* (**k**). **l**, Low levels of viral UMIs in heart, liver and kidney, compared with lung. Cumulative viral read counts as a function of droplet UMI count. In lung (red) most virus-positive droplets are empty droplets (total UMI count approximately 100) with some virus-positive droplets that contain nuclei (UMI count > approximately 1,000), but in heart (green), liver (blue) and kidney (orange), most of the 'virus-positive' droplets have fewer than ten total UMI counts, indicating that these reads are not trustworthy.

Extended Data Fig. 12 | See next page for caption.

Extended Data Fig. 12 | Expression of GWAS curated genes across lung, heart, liver and kidney atlases. a–d, Mean expression (dot colour, log(TP10K + 1)) and proportion of expressing cells (dot size) for each of 26 curated GWAS implicated genes (columns) in each cell subset (rows) for lung (**a**), heart (**b**), liver (**c**) and kidney (**d**) COVID-19 autopsy atlases. Results only reported for genes with expression in at least one cell subset in the underlying tissue. Some GWAS genes have higher expression in the lung compared with the other three tissues. **e, f,** Mean expression (**e**, z-score relative to all other cell types, colour bar) or differential expression (**f**, z-score of DE analysis of expression in COVID-19 versus healthy cells of the same type) of 25 out of 26 GWAS implicated genes (rows) from six genomic loci associated with COVID-19 (based on summary statistics data from COVID-19 HGI meta analysis[45] across lung cell types (columns). *ABO* was not considered as it was not reliably recovered in scRNA-Seq data. **g, h,** Cell type and disease progression gene programs in the lung (**g**), liver and kidney (**h**) that contribute to heritability of COVID-19 severity. Magnitude (circle size, *E* score) and significance (colour, $-\log_{10}(P)$) of the enrichment of cell type programs and cell-type-specific disease programs (columns) that were significantly enriched for COVID-19 or severe COVID-19 phenotypes (rows). All results are conditional on 86 baseline-LDv2.1 model annotations. **i,** Nomination of single best candidate genes at unresolved GWAS significant loci by aggregating gene level information across program classes and cell types. Significance ($-\log_{10}(P)$, *y* axis) of GWAS association signal at locus (*x* axis). Blue boxes, significantly associated loci[45] at a genome-wide significance level (purple horizontal bar). **j,** Schematic summarizing the key findings and contributions of this study.

Corresponding author(s): Aviv Regev

Last updated by author(s): Apr 12, 2021

Reporting Summary

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

Statistics

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

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

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

Software and code

Policy information about availability of computer code

Data collection   no software was used for data collection

Data analysis
1. Terra, a Cloud platform for storing and sharing data and analysis, tools; https://app.terra.bio/
2. Cumulus/cellranger_workflow on Cumulus, a (cloud-based) framework for running cellranger mkfastq and cellranger counts; used to run cellranger counts, version run was Snapshot 10 on Terra ;Read the Docs: https://cumulus.readthedocs.io/en/latest/cellranger.html; Terra WDL: https://portal.firecloud.org/?return=terra#methods/cumulus/cellranger_workflow/10
3. SARS-CoV-2 genome , used to align viral reads. Transcriptome reference name: BetaCov/South Korea/KCDC03/2020 based on NC_045512.2 https://github.com/hyeshik/sars-cov-2-transcriptome
4. Cumulus/cumulus workflow on Cumulus, a (cloud-based) framework for high-throughput single cell and single nucleus analysis using Pegasus; used for quality control and clustering analysis on individual samples, version run was Snapshot 29 on Terra; Read the Docs: https://cumulus.readthedocs.io/en/latest/cumulus.html# ; Terra WDL: https://portal.firecloud.org/?return=terra#methods/cumulus/cumulus/29
5. CellBender remove-background, removes ambient RNA and other technical artifacts from count matrices, version 0.2.0. Read the Docs: https://cellbender.readthedocs.io; Terra WDL: cellbender/remove-background (snapshot 11) Terra WDL: https://portal.firecloud.org/#methods/cellbender/remove-background/11
6. Scanpy, Python package for scRNA-seq data handling/processing, version 1.5.1+1.5.2.dev5+ge5d246aa; https://scanpy.readthedocs.io
7. Harmony-Pytorch, Python implementation of Harmony batch correction method, version 0.1.3; https://github.com/lilab-bcb/harmony-pytorch
8. Pegasus, Python package for scRNA-seq data handling/processing and generating heatmaps for NanoString GeoMx data, version 0.17.2; 1.0.0; https://pegasus.readthedocs.io
9. DESeq2, R package for analysis differential gene expression,version 1.28.0 for bulk RNA seq analysis, version 1.30.0 for viral and spatial DE analysis http://bioconductor.org/packages/release/bioc/html/DESeq2.html
10. MuSiC, R package for estimation of cell type proportions in bulk RNA-seq data, version 0.1.1; https://github.com/xuranw/MuSiC
11. GSEA, software for analyzing gene set enrichments, version 4.1.0 (run with database available as of 11/1/2020); https://www.gsea-

msigdb.org/gsea/index.jsp

12. GeoMx NGS Pipeline (DND) Processing Nanostring GeoMx NGS data for WTA and CTA assays, version 1.0.0; https://blog.nanostring.com/geomx-online-user-manual/Content/NGS_DND/Running_DND.htm#Running3

13. Limma,R package for differential gene expression analysis for NanoString GeoMx and heart snRNA-seq data, version 3.44.3;http://bioconductor.org/packages/release/bioc/html/limma.html

14. edgeR, R package for differential gene expression analysis for NanoString GeoMx data, version 3.28.1 or higher;https://bioconductor.org/packages/release/bioc/html/edgeR.html

15. EnhancedVolcano, R package for generating volcano plots for differential genes for analysis on NanoString GeoMx data, version1.6.0; https://bioconductor.org/packages/release/bioc/html/EnhancedVolcano.html

16. fgsea, R package for gene set enrichment analysis on NanoString GeoMx and heart snRNA-seq data, version 1.14.0; http://bioconductor.org/packages/release/bioc/html/fgsea.html

18. Viral-ngs, a collection of pipelines for viral genomic analyses including genome assembly and metagenomic classification, version 2.0.21; https://viral-ngs.readthedocs.io/en/latest/; https://dockstore.org/organizations/BroadInstitute/collections/pgs

19. Scikit-learn, Python module for machine learning, version 0.23; https://scikit-learn.org/stable/

20. Statsmodels, Python module for statistical modeling version 0.12.1 https://www.statsmodels.org/stable/index.html

21. ldsc, Python module for GWAS heritability analysis. https://github.com/bulik/ldsc

22. MAGMA, C++ command line interface for gene-level GWAS analysis version 1.08b
    https://ctg.cncr.nl/software/magma

23. scCODA, statistical testing for compositional analysis for scRNA-seq data , v0.1.1.post1, https://github.com/theislab/scCODA/releases/tag/0.1.1.post1

24. adjusted_rand_score from sklearn.metrics.cluster was used to compute rand index for sub-clustering. https://scikit-learn.org/stable/modules/generated/sklearn.metrics.adjusted_rand_score.html

25. DecontX, ambient RNA removal from scRNA-seq count matrix data (part of the "celda" package) https://github.com/campbio/celda version 1.5.11

26. GSVA,R package for gene set enrichment analysis was used to estimate the ssGSEA score for the alveoli NanoString GeoMX data. https://www.bioconductor.org/packages/release/bioc/html/GSVA.html

27. Seurat R package for snRNA-seq data analysis v3.2.1

28. R packages ggplot2 v3.3.2, dplyr 0.8.0.1, reshape2 v1.4.3 and cowplot v1.1.0 for visualization

29. liger R package v0.5.0 https://github.com/welch-lab/liger (Linked Inference of Genomic Experimental Relationships)

30. RSEM for bulk RNA-seq analysis , v1.2.8. , https://deweylab.github.io/RSEM/

31. STAR for bulk RNA-seq alignment, v2.6.0c , https://github.com/alexdobin/STAR

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

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:
- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Data availability
Processed sequencing data (sc/snRNA-Seq and bulk) are available in the Gene Expression Omnibus (GEO, https://www.ncbi.nlm.nih.gov/geo/) under accession no. GSE171668 and raw human sequencing data is available in the controlled access repository DUOS (https://duos.broadinstitute.org/), under Dataset IDs DUOS-000126, DUOS-000127, DUOS-000128 and DUOS-000129. Viral genome assemblies and short-read sequencing data are publicly available on NCBI's Genbank and SRA databases, respectively, under BioProject PRJNA720544. GenBank accessions for SARS-CoV-2 genomes are MW885875-MW885883. Data for other tissues in the biobank will be released as they are acquired.

The processed data is available on the Single Cell Portal:
Lung - https://singlecell.broadinstitute.org/single_cell/study/SCP1052/
Heart - https://singlecell.broadinstitute.org/single_cell/study/SCP1216/
Kidney - https://singlecell.broadinstitute.org/single_cell/study/SCP1214/
Liver - https://singlecell.broadinstitute.org/single_cell/study/SCP1213/

Nanostring GeoMx raw and normalized count matrices are available on GEO under accession no. GSE163530. Raw images will be available upon request.

Field-specific reporting

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

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

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

Life sciences study design

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

| Sample size | We generated sc/snRNA-Seq atlases of: |

Sample size	lung (n=16 donors, k=106,792 cells/nuclei, x=23 specimens; Donors=D1-8, 10-17) heart (n=18, k=40,880, x=19 specimens, D1-8, 10-11,14-17, 27-28, 31-32) liver (n=15, k=47,001, x=16 specimens; D1-7,10-17) kidney (n=16, k= 33,872, x=16 specimens;D4-8,10-12,14-15,17,25-26,28-30) We generated spatial data on the following: lung(n= 17 donors, x= 17 samples, Donors=D8-17,22-24) heart(n=1 donor, x= 1 sample, Donor =D20) heart(n=1 donor, x= 1 sample, Donor =D20) Because these are samples from human COVID-19 autopsy donors, we collected samples from as many donors that would consent over the collection period. We did not perform any power analyses prior to this.
Data exclusions	CellBender was used to remove ambient RNA and other technical artifacts from the count matrices. Following CellBender, individual samples were processed using Cumulus, including filtering out cells/nuclei with fewer than 400 UMI, 200 genes, or greater than 20% of UMIs mapped to mitochondrial genes.
Replication	These are samples from human COVID-19 autopsy donors, so we could not replicate samples
Randomization	These are samples from human COVID-19 autopsy donors, so we could not randomize
Blinding	These are samples from human COVID-19 autopsy donors, so this was not applicable to our study

Reporting for specific materials, systems and methods

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

Materials & experimental systems

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

Methods

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

Antibodies

Antibodies used	Immune Cell Profiling Panel (Core); Nanostring Inc; GMX-PROCONCT-HICP-12, Item 121300101, Lot# 0474026 IO Drug Target Panel; GMX-PROMODNCT-HIODT-12, Item 121300102, Lot# 0474029 Immune Activation Status Panel; Nanostring Inc; GMX-PROMODNCT-HIAS-12, Item 121300103, Lot# 0474032 Immune Cell Typing Panel; Nanostring Inc; GMX-PROMODNCT-HICT-12, Item 121300104, Lot# 0474035 Cell Death Panel; Nanostring Inc; GMX-PROMOD-NCTHCD-12, Lot# 0474050 MAPK Signaling Panel; Nanostring Inc; GMX-PROMOD-NCTHMAPK-12, Lot# 0474047 PI3K/AKT Signaling Panel; Nanostring Inc; GMX-PROMOD-NCTHPI3K-12, Lot# 0474053 Covid-19 GeoMx-formatted Antibody Panel including (TMPRSS2, clone EPR3861; ACE2, clone EPR4436; Cathepsin L/V/K/H, clone EPR8011; DDX5, clone EPR7239; and SARS-CoV-2 spike glycoprotein, polyclonal) ; Abcam; ab273594, Lot# GR3347471-1 GeoMx Solid Tumor TME Morphology Kit; Nanostring Inc; GMX-PRO-MORPH-HST-12; Item 121300310 Alexa Fluor® 647 alpha-Smooth Muscle Actin Antibody, clone 1A4 ; Novus Bio; IC1420R CD68 antibody,KP1 clone from Santa Cruz (sc-20060 AF594)
Validation	Nanostring morphological and staining panels are pre-validated by the manufacturer: https://www.nanostring.com/wp-content/uploads/2020/12/GeoMx_Antibody_Validation_White_Paper-3.pdf Morphological markers were previously demonstrated in human tissue in https://doi.org/10.1101/2020.08.25.267336

Human research participants

Policy information about studies involving human research participants

Population characteristics	Extended Data Table 1 - Patient metadata table
Recruitment	For BWH: Subjects were recruited who had died with positive SARS-CoV-2 NP swab test prior to death, and were consented for autopsy to be performed at BWH less than 24 hours from the time of death. No decisions were influenced by subject age,

race/ethnicity, sex/gender, pre-mortem treatments, or co-morbidities.

For MGH: All patients at the Massachusetts General Hospital (MGH) who succumbed from SARS-CoV-2 infection, as confirmed by the qRT-PCR assays performed on nasopharyngeal swab specimens, were eligible for clinical autopsy upon consent by their healthcare proxy or next of kin. A subset of these patients were also enrolled in the MGH Rapid Autopsy Protocol if they had a history of known or suspected malignancy. Their clinical data and research specimens were collected in accordance with Dana Farber/Harvard Cancer Center Institutional Review Board-approved protocol 13-416.

For BIDMC: The COVID rapid autopsy program was active at BIDMC from April 23, 2020 through May 6, 2020. An email was sent to all physicians caring for COVID patients notifying them about the existence of the program and that participation in the research autopsy program could be offered to families of deceased patients. The decision to offer participation in the autopsy research program to the next of kin of decedents was at the discretion of their treating physicians. In total, five autopsies were performed, representing a small fraction of the patients treated at BIDMC during the initial COVID surge of Spring 2020. No efforts were made to specifically include or exclude subjects based on any demographic data or pre-existing medical condition.

For NYP: Inclusion criteria for autopsies from COVID-19 donors cared for at New York Presbyterian Hospital/Columbia University Medical Center included real-time reverse transcription polymerase chain reaction (RT-PCR) confirmed infection, consent to perform rapid autopsy and post mortem intervals <10 hours. Appropriate consent was obtained from donors or the donors' next of kin. All procedures performed on donor samples were in accordance with the ethical standards of the IRB and the Helsinki Declaration and its later amendments. Frozen control tissues were assessed by a pulmonary pathologist and represent "uninvolved" regions of biobanked tumor resections. Donor characteristics reflect the age, gender, and race representation of patients admitted to New York Presbyterian Hospital/Columbia University Medical Center with COVID-19. Control samples were selected to reflect median age distribution of COVID-19 cases included in the study and match the gender distribution.

Ethics oversight	Secondary analysis of samples at the Broad Institute was covered under Massachusetts Institute of Technology (MIT) IRB protocols 1603505962 and 1612793224, or the NHSR (not-involving-human-subjects research) protocol ORSP-3635. No subject recruitment or ascertainment was performed as part of the Broad protocol. Samples added to this protocol also underwent IRB review and approval at the institutions where the samples were originally collected. Specifically, Dana-Farber Cancer Institute approved the protocol 13-416, Partners/Massachusetts General Hospital and Brigham and Women's Hospital approved the following protocols: 2020P000804, 2020P000849, 2015P002215; Beth Israel Deaconess approved protocol 2020P000406.224. No subject recruitment or ascertainment was performed as part of the Broad protocol. All tissue specimens of lethal COVID-19 and controls collected at New York Presbyterian Hospital/Columbia University Medical Center were under IRB approved protocols (IRB-AAAT0785 and IRB-AAAB2667). Appropriate consent was obtained from patients or the patients' next of kin. All procedures performed on patient samples at New York Presbyterian Hospital/Columbia University Medical Center were in accordance with the ethical standards of the IRB and the Helsinki Declaration and its later amendments.

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