A single-cell atlas of the peripheral immune response in patients with severe COVID-19

Aaron J. Wilk[1,2,5], **Arjun Rustagi**[3,5], **Nancy Q. Zhao**[2,5], **Jonasel Roque**[3], **Giovanny J. Martínez-Colón**[3], **Julia L. McKechnie**[2], **Geoffrey T. Ivison**[2], **Thanmayi Ranganath**[3], **Rosemary Vergara**[3], **Taylor Hollis**[3], **Laura J. Simpson**[3], **Philip Grant**[3], **Aruna Subramanian**[3], **Angela J. Rogers**[3,✉], **Catherine A. Blish**[1,3,4,✉]

[1]Stanford Medical Scientist Training Program, Stanford University School of Medicine, Stanford, CA, USA

[2]Stanford Immunology Program, Stanford University School of Medicine, Stanford, CA, USA

[3]Department of Medicine, Stanford University School of Medicine, Stanford, CA, USA

[4]Chan Zuckerberg Biohub, San Francisco, CA, USA

[5]These authors contributed equally: Aaron J. Wilk, Arjun Rustagi, Nancy Q. Zhao

Abstract

There is an urgent need to better understand the pathophysiology of Coronavirus disease 2019 (COVID-19), the global pandemic caused by SARS-CoV-2, which has infected more than three million people worldwide1. Approximately 20% of patients with COVID-19 develop severe

disease and 5% of patients require intensive care2. Severe disease has been associated with changes in peripheral immune activity, including increased levels of pro-inflammatory cytokines[3,4] that may be produced by a subset of inflammatory monocytes[5,6], lymphopenia[7,8] and T cell exhaustion[9,10]. To elucidate pathways in peripheral immune cells that might lead to immunopathology or protective immunity in severe COVID-19, we applied single-cell RNA sequencing (scRNA-seq) to profile peripheral blood mononuclear cells (PBMCs) from seven patients hospitalized for COVID-19, four of whom had acute respiratory distress syndrome, and six healthy controls. We identify reconfiguration of peripheral immune cell phenotype in COVID-19, including a heterogeneous interferon-stimulated gene signature, HLA class II downregulation and a developing neutrophil population that appears closely related to plasmablasts appearing in patients with acute respiratory failure requiring mechanical ventilation. Importantly, we found that peripheral monocytes and lymphocytes do not express substantial amounts of pro-inflammatory cytokines. Collectively, we provide a cell atlas of the peripheral immune response to severe COVID-19.

To profile the peripheral immune response to severe COVID-19, we performed Seq-Well-based[11,12] massively parallel single-cell RNA sequencing (scRNA-seq) on eight peripheral blood samples from seven hospitalized patients with polymerase chain reaction with reverse transcription (RT-PCR)-confirmed SARS-CoV-2 infection and six healthy controls. The demographics and clinical features of these patients are listed in Fig. 1a. The seven patients profiled were male, aged 20 to >80 years. We collected samples between 2 and 16 days following symptom onset; healthy controls were asymptomatic, four male and two female, and aged 30–50 years (Fig. 1a and Extended Data Fig. 1). Four of eight COVID-19 samples were collected from ventilated patients who were diagnosed with acute respiratory distress syndrome (ARDS; Fig. 1a). One patient (C1) was sampled twice: at nine days post-symptom onset while only requiring supplemental oxygen and at 11 days post-symptom onset following intubation. Three patients received azithromycin, which has potential immunomodulatory effects[13], at some point prior to sampling (Fig. 1a). Five patients received remdesivir in the hospital, four prior to sampling.

We sequenced 44,721 cells with an average of 3,194 cells per sample (Supplementary Table 1). We created a cells-by-genes expression matrix and performed dimensionality reduction by uniform manifold approximation and projection (UMAP) and graph-based clustering, which identified 30 clusters (Fig. 1b,c). We calculated each cluster's most highly differentially expressed (DE) genes to manually annotate clusters with their respective cellular identities (Fig. 1b,c, Supplementary Table 2 and Methods). Dimensionality reduction indicated substantial phenotypic differences between patients with COVID-19 and controls, predominantly in monocytes, T cells and natural killer (NK) cells (Fig. 1b,c).

We next quantified COVID-19-driven changes in the cell type proportions. Several innate immune cell subsets were depleted in patients with COVID-19, including γδ T cells, plasmacytoid dendritic cells (pDCs), conventional dendritic cells (DCs), CD16+ monocytes and NK cells, with the latter three cell types only significantly depleted in samples from patients with ARDS (Fig. 1d). These trends were not explained by the time post-fever or post-symptom onset (Extended Data Fig. 2). We also noted increased plasmablast

proportions in patients with COVID-19; these levels were most elevated in patients with ARDS (Fig. 1d), suggesting that more severe cases may be associated with a more robust humoral immune response, similar to previous reports[14,15]. Peripheral plasmablasts from patients with COVID-19 did not appear to share particular immunoglobulin V genes (Extended Data Fig. 3a).

Finally, a novel cell population that we annotated as 'developing neutrophils' was significantly increased only in patients with ARDS (Fig. 1d). These cells express several genes encoding neutrophil granule proteins (for example, *ELANE*, *LTF* and *MMP8*; see Fig. 4 and Supplementary Table 2 and 3)[16], but do not express genes encoding canonical neutrophil markers such as *FCGR3B* and *CXCR2* (Supplementary Table 3), and occupy a similar space as plasmablasts rather than canonical neutrophils in the UMAP embedding (Fig. 1c). In addition, they encompassed cells expressing *CEACAM8*, *ELANE* and *LYZ*, similar to recently described neutrophil progenitors[17,18], suggesting that these cells represent neutrophils at various developmental stages.

We next analyzed monocytes with more granularity, as this cellular compartment appeared to be most strongly remodeled in patients with COVID-19 (Fig. 1b,c). Dimensionality reduction of monocytes alone indicated a strong phenotypic shift in CD14+ monocytes and a depletion of CD16+ monocytes (Fig. 2a,b). We first examined expression of genes encoding inflammatory cytokines previously reported to be produced by circulating monocytes in COVID-19[5,6]. Notably, we did not identify substantial expression of pro-inflammatory cytokine genes *TNF*, *IL6*, *IL1B*, *CCL3*, *CCL4* or *CXCL2* by peripheral monocytes (Fig. 2c), suggesting that peripheral monocytes do not contribute to the putative cytokine storm in COVID-19.

To determine genes driving phenotypic remodeling in COVID-19 samples, we identified DE genes, pathways and upstream regulators by comparing cells of each COVID-19 sample to cells of all healthy controls (Fig. 2d and Supplementary Tables 4–24). Eight genes encoding HLA class II molecules were downregulated in at least six COVID-19 samples relative to healthy controls (Fig. 2d), concordant with other studies[19,20]. Scoring of individual cells by expression of all HLA class II-encoding genes revealed that this downregulation was significant in all patients with COVID-19, but potentially more prominent in ventilation-dependent patients (Fig. 2e,f and Supplementary Table 25). HLA class II downregulation is reflected in differentially regulated gene pathways including reduction of crosstalk between dendritic cells and natural killer cells (Fig. 2g and Supplementary Table 11). HLA class II downregulation was also noted in B cells (Extended Data Fig. 3b,c and Supplementary Table 10), and the extent of downregulation tended to be greater in older patients (Extended Data Fig. 4). Non-classical HLA class I genes *HLA-E* and *HLA-F* were also downregulated to a lesser degree and in fewer samples, while canonical HLA class I genes *HLA-A*, *HLA-B* and *HLA-C* were not consistently up- or downregulated (Fig. 2f).

Additionally, 32 interferon (IFN)-stimulated genes (ISGs) were upregulated by CD14+ monocytes in at least one COVID-19 sample, but this IFN signature was not uniform across all COVID-19 samples (Fig. 2d and Supplementary Table 4). Analysis of upstream regulators in CD14+ monocytes revealed an absence of predicted IFN and IFN regulatory

factor (IRF) activities in donors C2, C3 and C7 relative to the remaining COVID-19 donors (Fig. 2h). Similar patterns were observed in other cellular compartments (Extended Data Figs. 5 and 6 and Supplementary Tables 18–24). To analyze this orthogonally, we scored individual CD14$^+$ monocytes in the dataset by their expression of known ISGs and again saw minimal appreciable ISG signatures in donors C2, C3 and C7 (Fig. 2i and Supplementary Table 25). The differential ISG signature was not explained by ventilation or ARDS (Fig. 2h,i), but a higher ISG score trended towards a positive correlation with age and a negative correlation with time–distance from fever onset (Fig. 2j).

We next analyzed T and NK lymphocytes in COVID-19 samples. UMAP embedding of T and NK cells identified substantial differences in cellular phenotypes of CD4$^+$ T, CD8$^+$ T and NK cells (Fig. 3a,b). We found that CD56^{dim} NK cells, generally thought to contribute to antiviral host defense through cell-mediated cytotoxicity[21,22], were depleted primarily in ventilator-dependent patients, whereas CD56^{bright} NK cells, which are considered robust producers of IFN-γ and tumor necrosis factor α[23], were significantly depleted in all COVID-19 samples (Fig. 3c). Additionally, we identified a cluster of proliferative lymphocytes cells that appeared to be increased in most patients with COVID-19 (Fig. 3c). As SARS-CoV-2 infection has been associated with cytotoxic lymphocyte exhaustion[10], we profiled the expression of genes encoding canonical exhaustion markers by T and NK cells. However, there was no significant evidence of CD8$^+$ T cell exhaustion in patients with COVID-19 and, although exhaustion markers appeared elevated among CD4$^+$ T cells, these changes were not significant (Extended Data Fig. 7 and Supplementary Table 25). NK cells from most patients with COVID-19 appeared exhausted based on expression of *LAG3*, *PDCD1* and *HAVCR2* (Fig. 3d). Similar to our observations in peripheral monocytes, we did not detect substantial expression of pro-inflammatory cytokine genes by T or NK cells (Fig. 3e and Extended Data Figs. 5 and 8); this again indicates that transcription of pro-inflammatory cytokines by peripheral leukocytes is unlikely to be a major contributor to the putative cytokine storm in COVID-19.

We next calculated T and NK cell DE genes from each sample from a patient with COVID-19 relative to healthy controls, and used these genes to identify enriched gene pathways and upstream regulators. NK cells displayed a remarkably heterogeneous response between patients with COVID-19 (Fig. 3f and Supplementary Table 7). The most frequently downregulated genes included *FCGR3A*, *AHNAK* and *FGFBP2*, which are associated with peripheral NK cell maturity[24]. The most commonly upregulated genes included ISGs and NK cell activation genes like *PLEK* and *CD38*[25,26]. We observed similar heterogeneity of DE genes in CD4$^+$ and CD8$^+$ T cells, where the most commonly upregulated genes were ISGs (Extended Data Fig. 5 and Supplementary Tables 8 and 9).

Analysis of predicted upstream regulators indicated a strong IFN-driven response that was starkly absent from half of the profiled COVID-19 samples in both NK cells, CD4$^+$ and CD8$^+$ T cells (Fig. 3g, Extended Data Figs. 5 and 6 and Supplementary Tables 21–23). Given the importance of the IFN response and recent reports that this response is diminished during COVID-19[27,28], we evaluated ISG upregulation in each cell type to determine if ISGs were coordinately expressed across all cell types or between individuals (Fig. 3h). Although some ISGs were upregulated by most donors in a given cell type (for example, *IFI27* in

CD14$^+$ monocytes), generally ISG upregulation was not uniform within cell types or between subjects (Fig. 3h). In addition, we identified very few cytokines whose upregulation was consistent between most patients with COVID-19 (Fig. 3i). These results collectively indicate heterogeneous peripheral immune activation in COVID-19.

We next analyzed the phenotypes of plasmablasts and developing neutrophils, which appeared to be phenotypically related by dimensionality reduction (Fig. 1c). Indeed, when embedding only these cell types, developing neutrophils appeared to project linearly from plasmablasts, suggestive of a continuum of cellular phenotype between the two cell types (Fig. 4a). Cellular complexity (the number of genes sequenced per cell divided by the unique molecular identifiers (UMIs) per cell) was not higher in developing neutrophils, making it unlikely that these cells were multiplets (Extended Data Fig. 9). These cells are also unlikely to represent granulocytes that have phagocytosed B cells, a feature of hemophagocytic lymphohistiocytosis (HLH), which can be triggered by severe acute infections, because these patients did not have clinical characteristics of HLH.

To analyze if there was any transition between the two cell types, we performed a cellular trajectory analysis by RNA velocity[29,30]. Surprisingly, this analysis demonstrated that the linear continuum of cellular phenotype represented a differentiation bridge from plasmablasts to developing neutrophils (Fig. 4a); this spectrum of plasmablast-to-neutrophil phenotype was observed in all patients with ARDS and appeared unrelated to the transcriptional dynamics of canonical neutrophils (Extended Data Fig. 10). The cells along this differentiation bridge had lost expression of genes encoding canonical plasmablast markers CD27, CD38 and TNFRSF17 and instead sequentially acquired expression of genes encoding primary (*DEF3A*, *ELANE* and *MPO*), secondary (*CHI3DL1*, *LCN2* and *LTF*) and tertiary (*MMP8*, *MMP9* and *CAMP*) neutrophil granule proteins, similar to canonical neutrophil development (Fig. 4b). Recovery of inferred latent time, which is based solely on a cell's transcriptional dynamics, also suggested a continuum from plasmablasts to developing neutrophils (Fig. 4c). Although cells at the beginning of this continuum are defined by expression of *Ig* genes, neutrophil markers like *CSF3R* and *MNDA* (encoding myeloid nuclear differentiation antigen) are upregulated as latent time progresses (Fig. 4d).

A lymphocyte-to-granulocyte differentiation process is not without precedent. Similar transitions have been described from B cells to macrophages or granulocytes, and the C/enhancer binding protein (EBP) transcription factor family has been implicated in controlling this transdifferentiation[31,32]. Two C/EBP family members, *CEBPE* and *CEBPD*, both known drivers of myeloid and granulocyte cell fates[33,34], are selectively expressed by the two clusters of cells along the differentiation bridge (Fig. 4e,f); the transition from *CEBPE* to *CEBPD* recapitulates neutrophil development in mice[35]. Collectively, we observe a developing neutrophil population that may be characteristic of ARDS in severe COVID-19 infection; our data suggest that these cells may derive from plasmablasts, but they may also represent developing neutrophils derived from emergency granulopoiesis[36].

There are several limitations of our study. Our sample size is small, only peripheral blood was evaluated and patients varied in the timing of their clinical presentation, which could influence their transcriptional landscapes. A subset of the patients were treated with the

antibiotic azithromycin, which has known immunomodulatory activity[13], while another subset were treated with the antiviral remdesivir, which targets the viral RNA-dependent RNA polymerase[37,38] and is not known to have direct immunomodulatory effects. Future studies are needed to further define the origins and phenotypes of the developing neutrophil population observed in the setting of ARDS at both the transcriptional and phenotypic level. Such studies will optimally require freshly isolated whole-blood samples from patients who are severely ill with COVID-19 as granulocytes typically fail to survive cryopreservation.

Overall, we used single-cell transcriptomics to characterize peripheral immune responses in severe COVID-19. We observed marked changes in the immune cell composition and phenotype in SARS-CoV-2 infection and immunological features of severe COVID-19 in patients with ARDS. This work represents a resource for understanding peripheral immunity in severe COVID-19 and presents new directions for the study of COVID-19 immunology and therapeutic development.

Methods

Subjects and specimen collection

We collected blood from seven patients enrolled in the Stanford University ICU Biobank study from March–April 2020 after written informed consent from patients or their surrogates (Stanford IRB approval #28205). Eligibility criteria included age 18 years and admission to Stanford Hospital with a positive SARS-CoV-2 nasopharyngeal swab by RT-PCR. Patients admitted to the wards or ICU were included, and the majority were co-enrolled in ongoing COVID-19 treatment trials at Stanford. Screening of new admissions via an electronic medical records review of all subjects was performed by the study coordinator (J.R.), research fellow (A.R.), COVID-19 clinical consultants (P.G. and A.S.) and the study principal investigator (A.J.R.), and was done every day with a goal enrollment within less than 48 h of admission to the hospital. Patients were phenotyped for ARDS using the Berlin criteria (acute onset of hypoxemic respiratory failure with a PaO_2/FIO_2 ratio (i.e., the ratio of the partial pressure of arterial oxygen to the percentage of inspired oxygen) of <300 on at least 5 cm of positive end-expiratory pressure, bilateral infiltrates on chest X-ray)[41]. To protect the identity of the COVID-19 subjects, ages are reported as ranges. For controls, blood was collected from six asymptomatic adult donors as part of the Profiling Healthy Immunity study after written informed consent (Stanford IRB approval #26571). All donors were asked for consent for genetic research.

Blood draws from patients occurred in concert with usual care to avoid unnecessary personal protective equipment usage. For both patients with COVID-19 and healthy controls, blood was collected into heparin tubes (Becton, Dickinson and Co.) and PBMCs were isolated by density gradient centrifugation using Ficoll-Paque Plus medium (GE Healthcare) and washed with Ca/Mg-free PBS. Blood was processed within 4 h of collection for all samples, and within 1 h for most. Samples from patients with COVID-19 and healthy controls were processed side by side to avoid variation from processing.

scRNA sequencing by Seq-Well

The Seq-Well platform for scRNA-seq was utilized as described previously[11,12,42]. Immediately after Ficoll separation, 50,000 PBMCs were resuspended in RPMI + 10% FCS at a concentration of 75,000 cells per ml. A 200-µl volume of this cell suspension (15,000 cells) was then loaded onto Seq-Well arrays pre-loaded with mRNA capture beads (ChemGenes). Following four washes with Dulbecco's phosphate-buffered saline (DPBS) to remove serum, the arrays were sealed with a polycarbonate membrane (pore size of 0.01 µm) for 30 min at 37 °C and then frozen at −80 °C for no less than 24 h and no more than 14 days to allow batching of samples processed at irregular hours. Next, arrays were thawed, cells lysed, transcripts hybridized to the mRNA capture beads, and beads recovered from the arrays and pooled for downstream processing. Immediately after bead recovery, mRNA transcripts were reverse-transcribed using Maxima H-RT (Thermo Fisher EPO0753) in a template-switching-based rapid amplification of cDNA ends (RACE) reaction, excess unhybridized bead-conjugated oligonucleotides were removed with exonuclease I (NEB M0293L) and second-strand synthesis was performed with Klenow fragment (NEB M0212L) to enhance transcript recovery in the event of failed template switching[42]. Whole transcriptome amplification (WTA) was performed with KAPA HiFi PCR Mastermix (Kapa Biosystems KK2602) using ~6,000 beads per 50-µl reaction volume. Resulting libraries were then pooled in sets of six (~36,000 beads per pool) and products purified by Agencourt AMPure XP beads (Beckman Coulter, A63881) with a 0.6× volume wash followed by a 0.8× volume wash. The quality and concentration of WTA products were determined using an Agilent Fragment Analyzer (Stanford Functional Genomics Facility), with a mean product size of >800 bp and a non-existent primer peak indicating successful preparation. Library preparation was performed with a Nextera XT DNA library preparation kit (Illumina FC-131−1096) with 1 ng of pooled library using dual-index primers. Tagmented and amplified libraries were again purified by Agencourt AMPure XP beads with a 0.6× volume wash followed by a 1.0× volume wash, and quality and concentration were determined by fragment analysis. Libraries between 400 and 1,000 bp with no primer peaks were considered successful and pooled for sequencing. Sequencing was performed on a NovaSeq S2 instrument (Illumina; Chan Zuckerberg Biohub). The read structure was paired-end with read 1 beginning from a custom read 1 primer[11] containing a 12-bp cell barcode and an 8-bp UMI, and with read 2 containing 50 bp of mRNA sequence.

Alignment and quality control of sequencing data

Sequencing reads were aligned and count matrices assembled using STAR[43] and dropEst[44], respectively. Briefly, the mRNA reads in read 2 demultiplexed FASTQ files were tagged with the cell barcode and UMI for the corresponding read in the read 1 FASTQ file using the dropTag function of dropEst. Next, reads were aligned with STAR using the GRCh37 (hg19) human reference genome, which included the complete genome sequences for all SARS-CoV-2 strains sequenced from California before 24 March 2020 (10 SARS-CoV-2 sequences). No SARS-CoV-2 reads were aligned from these samples using this strategy, even when the outFilterMultimapNmax behavioral option of STAR was increased from 10 (default) to 20 to accommodate potential multiple-mapping SARS-CoV-2 reads. Count matrices were built from the resulting BAM files using dropEst[44]. Count matrices for intron-aligned reads were also generated to computationally analyze cellular trajectory. Cells that

had fewer than 1,000 UMIs or greater than 15,000 UMIs, as well as cells that contained greater than 20% of reads from mitochondrial genes or rRNA genes (*RNA18S5* or *RNA28S5*), were considered low quality and removed from further analysis. To remove putative multiplets (where more than one cell may have loaded into a given well on an array), cells that expressed more than 75 genes per 100 UMIs were also filtered out. Genes that were expressed in fewer than 10 cells were removed from the final count matrix.

scRNA-seq computational pipelines and analyses

The R package Seurat was used for data scaling, transformation, clustering, dimensionality reduction, differential expression analysis and most visualization[45]. Data were scaled and transformed and variable genes identified using the SCTransform() function, and linear regression was performed to remove unwanted variation due to cellular complexity (no. of genes per cell, no. of UMIs per cell) or cell quality (% mitochondrial reads, % rRNA reads). Principal component analysis was performed using variable genes, and the first 50 principal components (PCs) were used to perform UMAP to embed the dataset into two dimensions. Next, the first 50 PCs were used to construct a shared nearest-neighbor graph (SNN; FindNeighbors()) and this SNN was used to cluster the dataset (FindClusters()) using a graph-based modularity-optimization algorithm of the Louvain method for community detection[46]. Despite upstream filtering for high-quality cells and regression on genes reflective of cell quality, two clusters were identified where 65% or 100% of the positively enriched genes were of mitochrondrial or ribosomal origin, and these clusters were removed from further analysis[47,48].

Cellular identity was determined by finding DE genes for each cluster using Seurat's implementation of the Wilcoxon rank-sum test (FindMarkers()) and comparing those markers to known cell type-specific genes from previous datasets[49–54]. Cluster annotation was confirmed using the R package SingleR[55], which compares the transcriptome of each single cell to reference datasets to determine cellular identity. Although clustering is often insufficient to separate cytotoxic T cells from NK cells[12,49], SingleR identified the majority of cells in clusters 0 and 11 (94% and 76%, respectively) as NK cells. Indeed, these two clusters were the only clusters in the dataset to be significantly enriched for both *NCAM1* and *FCGR3A* (Supplementary Table 2) and we thus annotated them as NK cells. We also observed that cluster 22, where 89% of cells were annotated as T cells by SingleR, was significantly enriched for genes encoding γδ TCR constant chains *TRGC1*, *TRGC2* and *TRDC*, and we thus annotated them as γδ T cells (Supplementary Table 2). The majority of cells in cluster 24 were labeled as common myeloid progenitors by SingleR, but this cluster also contained cells annotated as seven different lineages of hematopoietic stem cells and progenitors. Closer examination revealed that this cluster consisted of two groups of cells, one expressing *CLC* and the other expressing *CD34*, and we therefore labeled them as stem cells (SCs) and eosinophils for downstream analysis. A total of 98% of cells in cluster 27 were annotated by SingleR as myelocytes (46%), pro-myelocytes (22%), CD34⁻ pre-B cells (14%) or <q>HSC G-CSF (17%). Although these cells expressed several genes encoding for primary, secondary and tertiary neutrophil granule proteins (for example, *ELANE*, *MPO*, *LTF*, *CTSG*, *LCN2* and *MMP8*), they were distinct from cluster 25 (labeled manually and by SingleR as neutrophils) and did not express canonical neutrophil markers like *FCGR3B* and

CXCR2. As these cells demonstrated features similar to immature neutrophils and progenitors at various developmental stages[17,18], we annotated these cells as 'developing neutrophils'.

Gene pathway and upstream regulator analysis was performed with Ingenuity Pathway Analysis (IPA; Qiagen). The parent Seurat object was divided into individual objects consisting of cells from a particular cellular compartment (for example, CD4+ T cells, NK cells, CD16+ monocytes and so on). Next, DE genes between the cells from each sample from a patient with COVID-19 and the cells from all healthy controls were calculated by FindMarkers() and cell quality-associated markers were removed. The average log(fold-change) of each DE gene calculated by FindMarkers() was supplied to IPA. To construct heatmaps of DE genes, genes were filtered to have a two-sided P value < 0.05 by Seurat's implementation of the Wilcoxon rank-sum test, and the average log(fold-change) of each DE gene plotted. To construct heatmaps of canonical pathways and predicted upstream regulators, genes were filtered to have a right-sided P value < 0.05 by IPA's implementation of Fisher's exact test, and the z-score of each pathway or regulator was plotted. Analysis of cellular trajectory by RNA velocity was performed using the package scVelo using dynamical modeling[30]. For all dot plots, average expression was calculated as the mean of $e^x - 1$, where x is the SCT-transformed count of each gene in cells of a given identity class. Dot plots with hierarchical clustering were generated using FlexDotPlot[56].
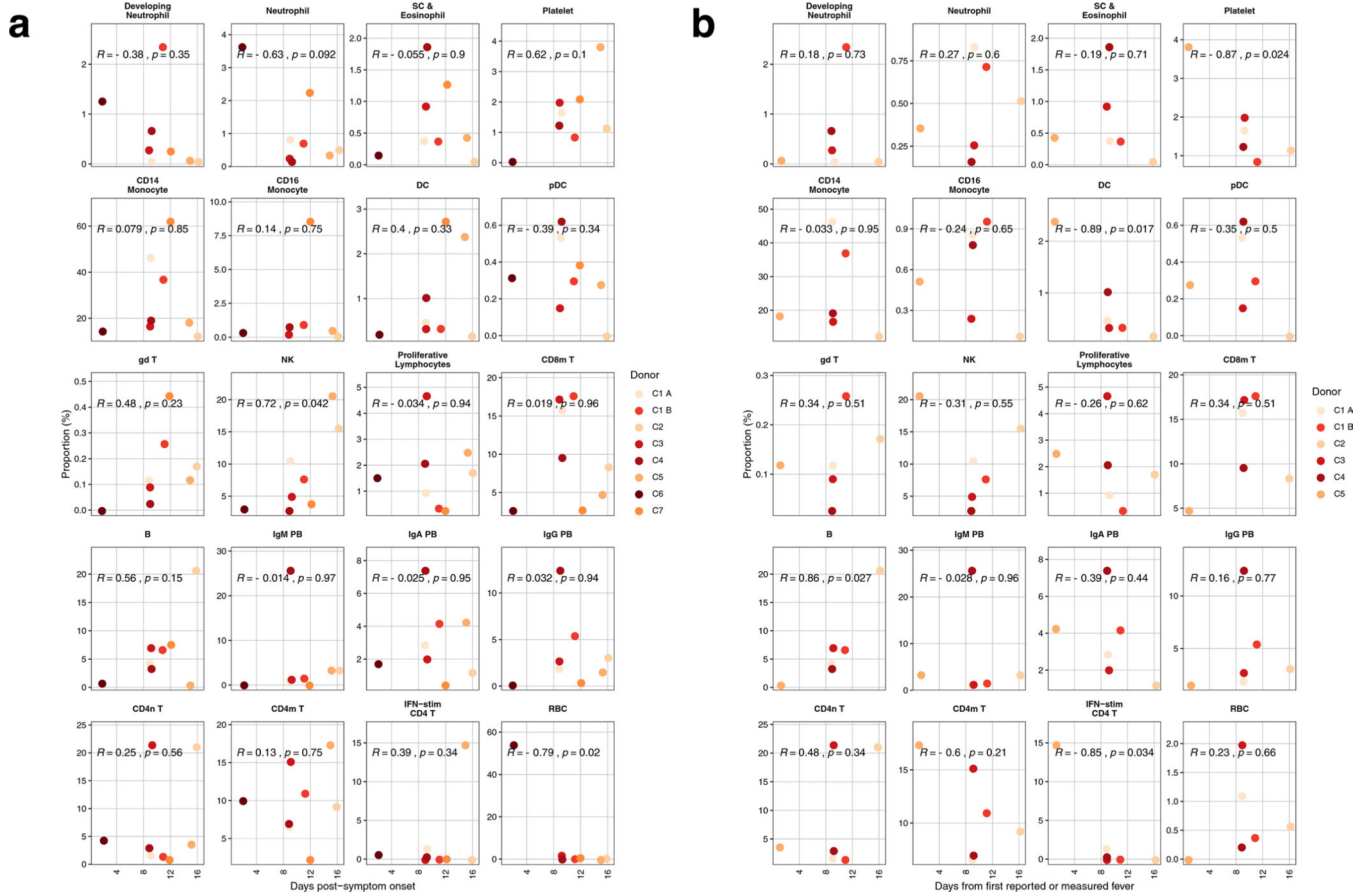
Extended Data

a

Donor	Age	Sex	Race
H1	49	F	White
H2	49	M	White
H3	36	F	White
H4	49	M	White
H5	48	M	Asian
H6	37	M	White

b

| Race	# of COVID-19 donors
Asian	2
Black or African American	1
Hispanic/Latino	2
White	2
Total	**7**

Extended Data Fig. 1 |. Demographic characteristics of all analyzed donors.
a, Age, sex, and race of $n = 6$ profiled healthy donors. **b**, Races represented by $n = 7$ patients with COVID-19 in this study.

Extended Data Fig. 2 |. Proportions of PBMC cell subsets do not correlate with days post-symptom onset or days to fever onset.

Proportions of each cell type in each sample colored by donor of origin. The *x*-axis corresponds to **a**, the days post-symptom onset (*n* = 8 COVID-19 samples) or **b**, the days from first reported or measured fever (*n* = 6 COVID-19 samples from patients who had experienced fever). Shown are the Pearson correlation coefficient and exact two-sided p-values for each scatter plot.
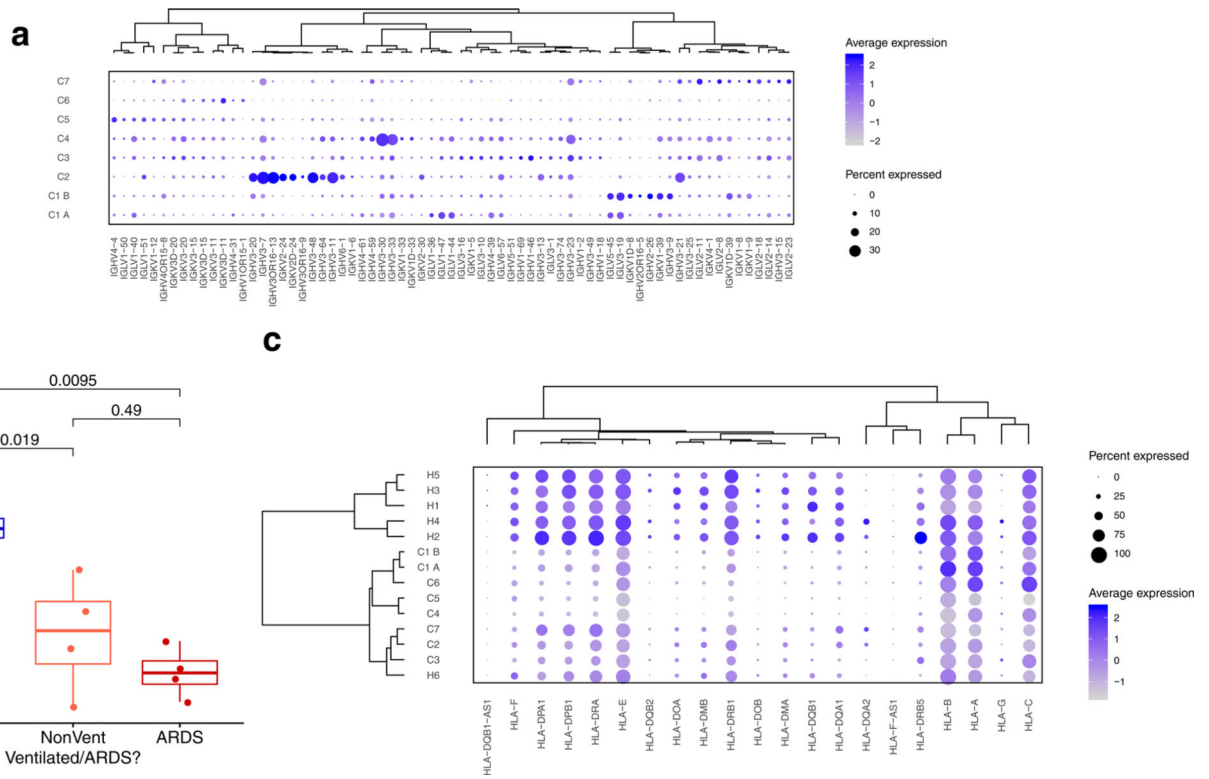
Extended Data Fig. 3 |. Heterogeneity of Ig V gene usage and HLA class II downregulation in B cells of COVID-19 patients.

a, Dot plot depicting average and percent expression of all detected *Ig* genes by $n = 8$ COVID-19 samples. Number of cells: C1 A – 182, C1 B – 618, C2 – 123, C3 – 351, C4 – 1,577, C5 – 196, C6 – 32, C7 – 29. **b**, Boxplot depicting the mean HLA class II expression score for all B cells of each sample in the dataset, colored by ventilation/ARDS status (healthy controls are colored in blue, $n = 6$; non-ventilated COVID-19 patients in orange, $n = 4$; ventilated COVID-19 patients in red, $n = 4$). Shown are exact two-sided p values by the Wilcoxon rank-sum test. Boxplot features correspond to: minimum whisker = 25th percentile - 1.5 * inter-quartile range (IQR) or the lowest value within; minimum box = 25th percentile; center = median; maximum box = 75th percentile; maximum whisker = 75th percentile + 1.5 * IQR or greatest value within. **c**, Dot plot showing percent and average expression of all detected *HLA* genes by B cells for $n = 8$ COVID-19 samples and $n = 6$ healthy controls. Number of cells: C1 A – 351, C1 B – 1,020, C2 – 495, C3 – 883, C4 – 1,835, C5 – 246, C6 – 77, C7 – 298, H1 – 190, H2 – 299, H3 – 298, H4 – 588, H5 – 511, H6 – 200.
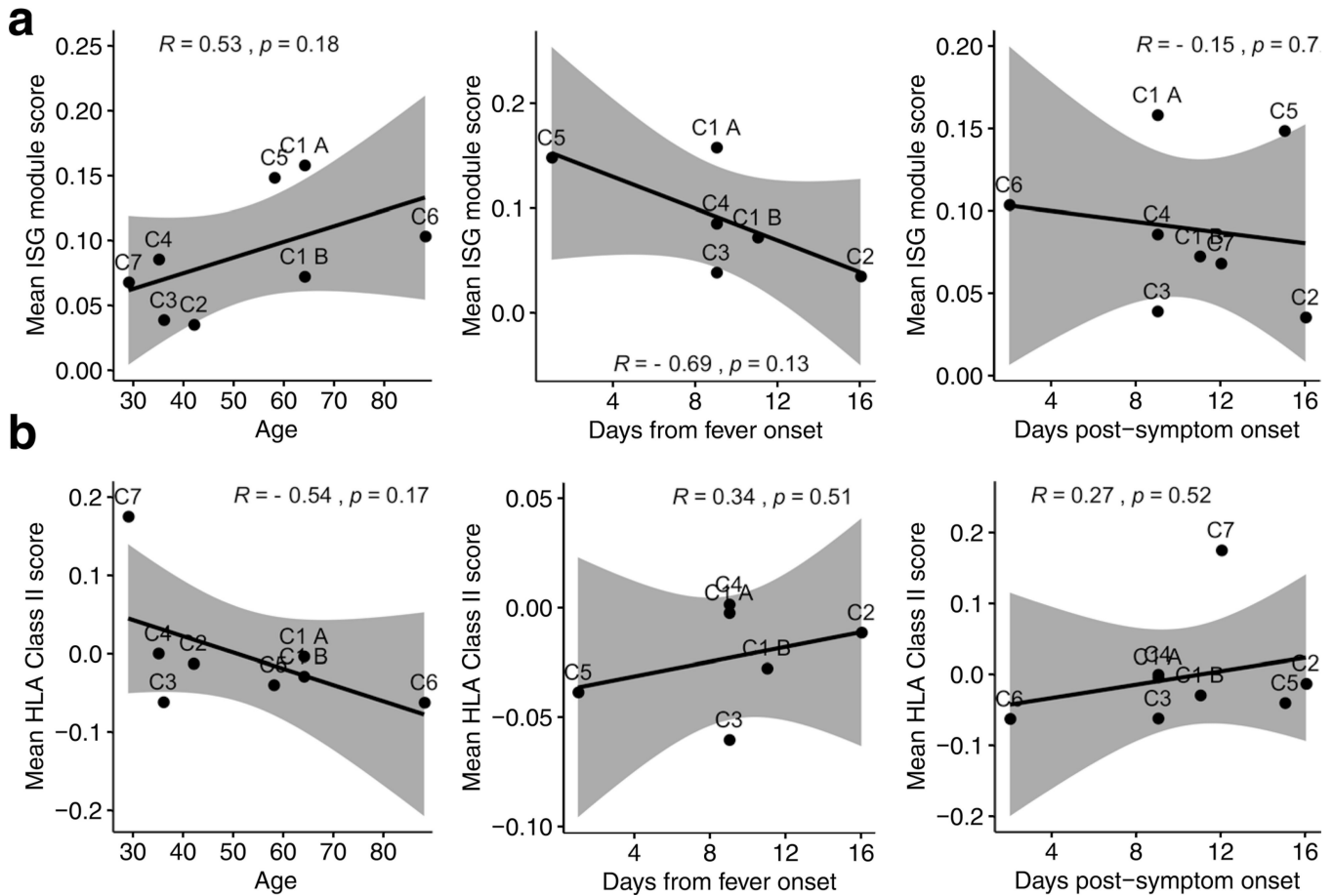
Extended Data Fig. 4 |. Trends between ISG or HLA class II expression scores and clinical parameters in total PBMCs.

a, b, Scatter plots depicting the correlation between (**a**) the mean ISG expression score or (**b**) the mean HLA class II expression score and patient age, days from fever onset, or days post-symptom onset (Supplementary Table 25). For all plots, only $n = 8$ COVID-19 patient samples are included. For plots depicting correlations with days from fever onset, $n = 6$ as the two COVID-19 patients who never reported fever (C6 and C7) are excluded. Pearson correlation coefficients, exact two-sided p-values, and the 95% confidence interval are shown for each comparison.
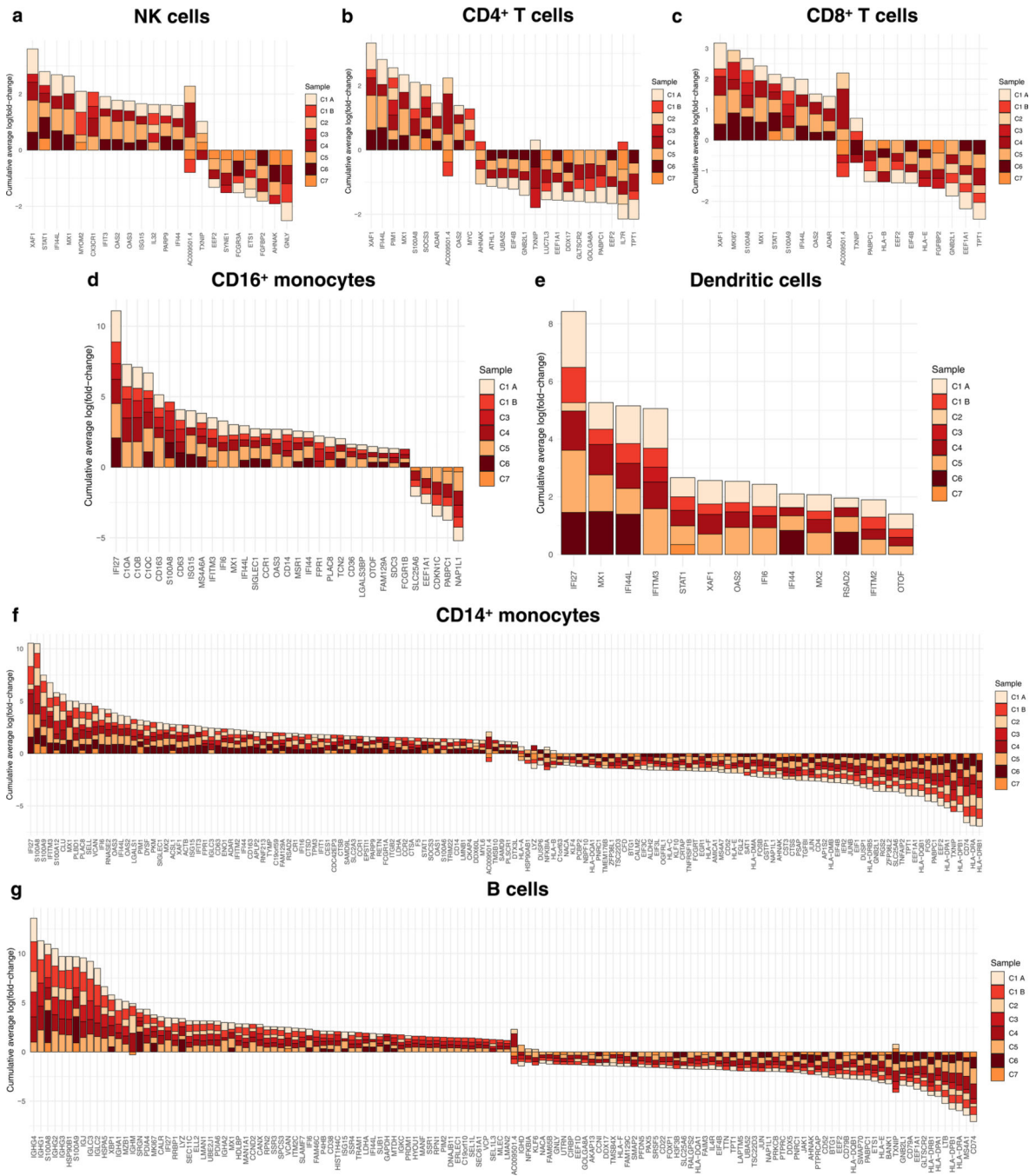
Extended Data Fig. 5 |. Consensus of DE genes in each major cell type.
Each bar plot represents genes differentially expressed in at least half (4) of $n = 8$ COVID-19 samples relative to all healthy controls. DE genes were considered those with an average |log(fold-change)| $> \pm 0.25$ and a two-sided p value < 0.05 as determined by Seurat's implementation of the Wilcoxon rank-sum test. Each bar represents the cumulative log(fold-change) of each DE gene, colored by the contributions of individual COVID-19 samples. DE genes are shown for **a**, NK cells, **b**, CD4[+] T cells, **c**, CD8[+] T cells, **d**, CD16[+] monocytes, **e**, Dendritic cells, **f**, CD14[+] monocytes, and **g**, B cells. There were no DE genes in at least four
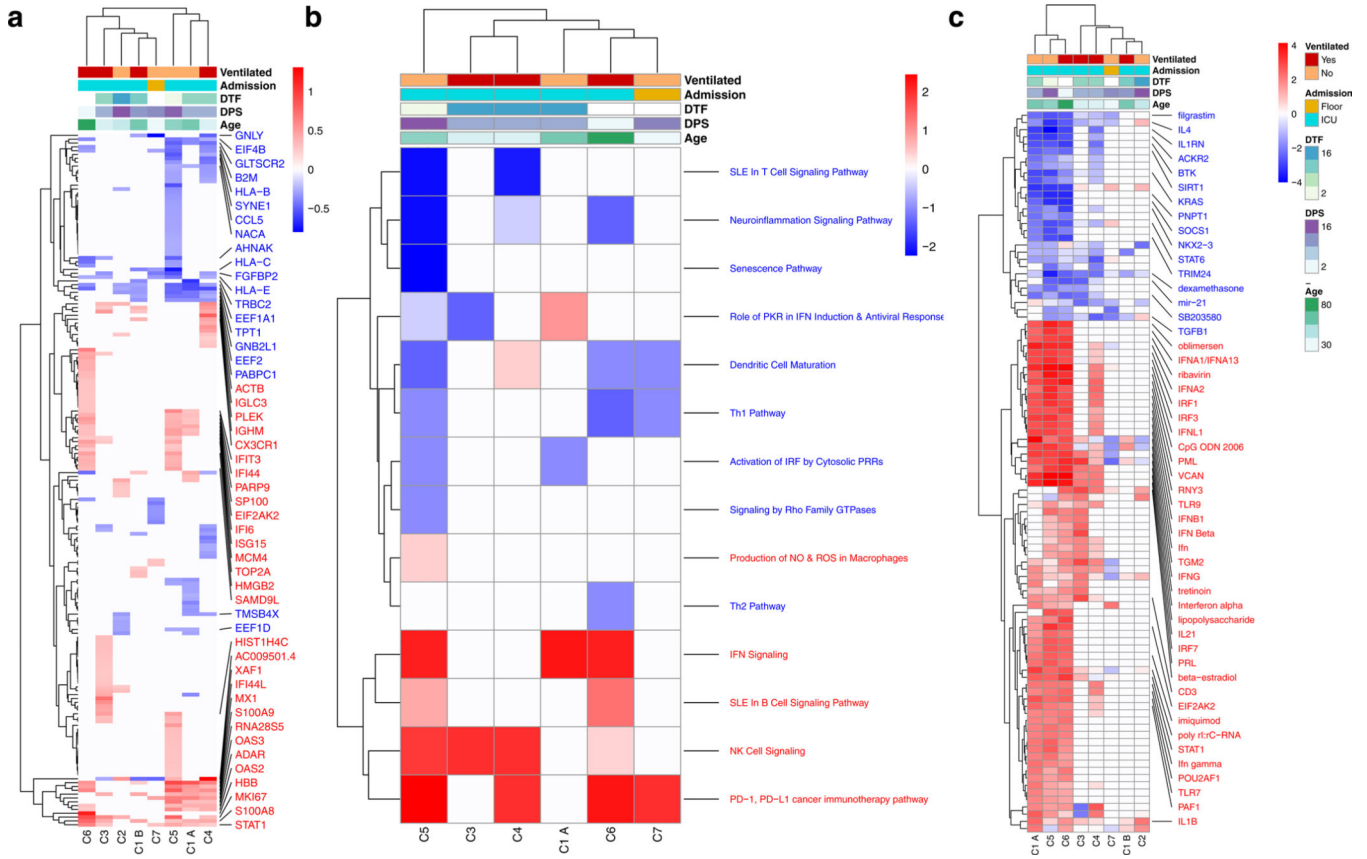
COVID-19 samples for γδ T cells. Total number of cells per donor: C1 A – 3,222, C1 B – 4,889, C2 – 1,695, C3 – 6,206, C4 – 3,559, C5 – 2,391, C6 – 794, C7 – 3,145; cells from all healthy controls ($n = 16,231$ cells) were used to generate comparisons with each COVID-19 sample.

Extended Data Fig. 6 |. Differentially regulated genes, canonical pathways, and upstream regulators in CD8$^+$ T cells.

a, Heatmap of DE genes for each COVID-19 sample colored by average log(fold-change). DE genes were calculated by comparing gene expression of individual COVID-19 samples with gene expression of all healthy controls using Seurat's implementation of the Wilcoxon rank-sum test. Only DE genes with a two-sided p value<0.05 adjusted for multiple comparisons by Bonferroni's correction are shown. These DE genes were used to identify **b**, enriched canonical pathways and **c**, upstream regulators, both colored by z-score, using Ingenuity Pathway Analysis (IPA). The (**a**) 50 genes, (**b**) 14 pathways, or (**c**) 50 regulators with the highest absolute average log(fold-change) or z-score across all donors are labeled. Genes with a net positive average log(fold-change) or z-score are labeled in red; genes with a net negative average log(fold-change) or z-score are labeled in blue. All (**b**) pathways and (**c**) upstream regulators are statistically significant by IPA's implementation of Fisher's exact test at a right-sided p < 0.05. Number of cells: C1 A – 563, C1 B – 904, C2 – 176, C3 – 1,437, C4 – 432, C5 – 183, C6 – 80, C7 – 102; cells from all healthy controls (*n* = 2,885 cells) were used to generate comparisons with each COVID-19 sample.
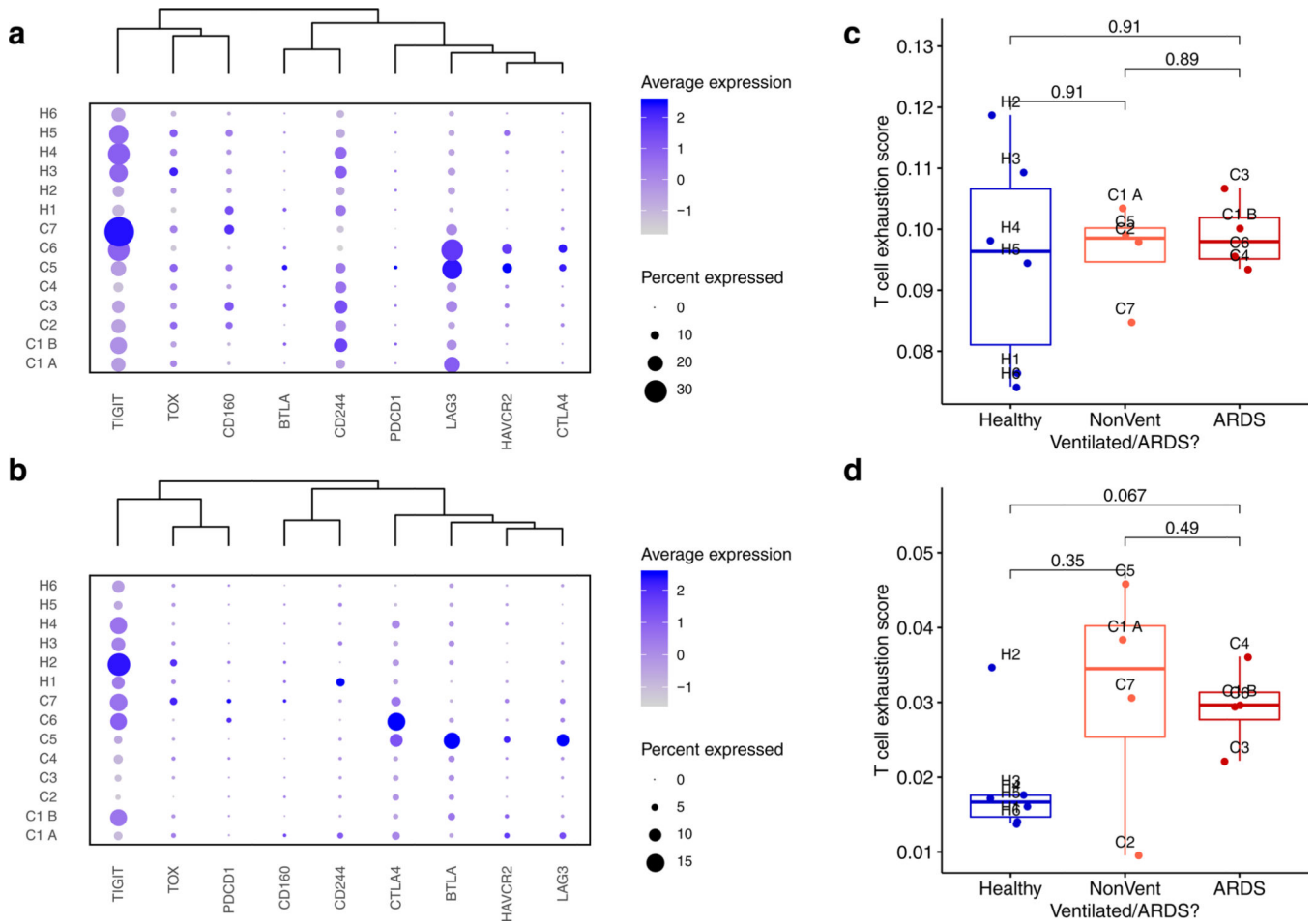
Extended Data Fig. 7 |. CD8+ and CD4+ T cells from COVID-19 patients do not consistently express higher levels of exhaustion markers.

a, **b**, Dot plot depicting the percent and average expression of canonical genes associated with T cell exhaustion by (**a**) CD8+ and (**b**) CD4+ T cells from $n = 8$ COVID-19 samples and $n = 6$ healthy controls. **c-d**, Boxplot showing the mean T cell exhaustion of module score[57] (see Supplementary Table 25) of CD8+ T cells (**c**) or CD4+ T cells (**d**) from each sample, colored by healthy donors ($n = 6$, blue), non-ventilated COVID-19 patients ($n = 4$, orange), or ventilated COVID-19 patients ($n = 4$, red). Shown are exact two-sided p values by Wilcoxon rank-sum test. Boxplot features correspond to: minimum whisker = 25th percentile - 1.5 * inter-quartile range (IQR) or the lowest value within; minimum box = 25th percentile; center = median; maximum box = 75th percentile; maximum whisker = 75th percentile + 1.5 * IQR or greatest value within.
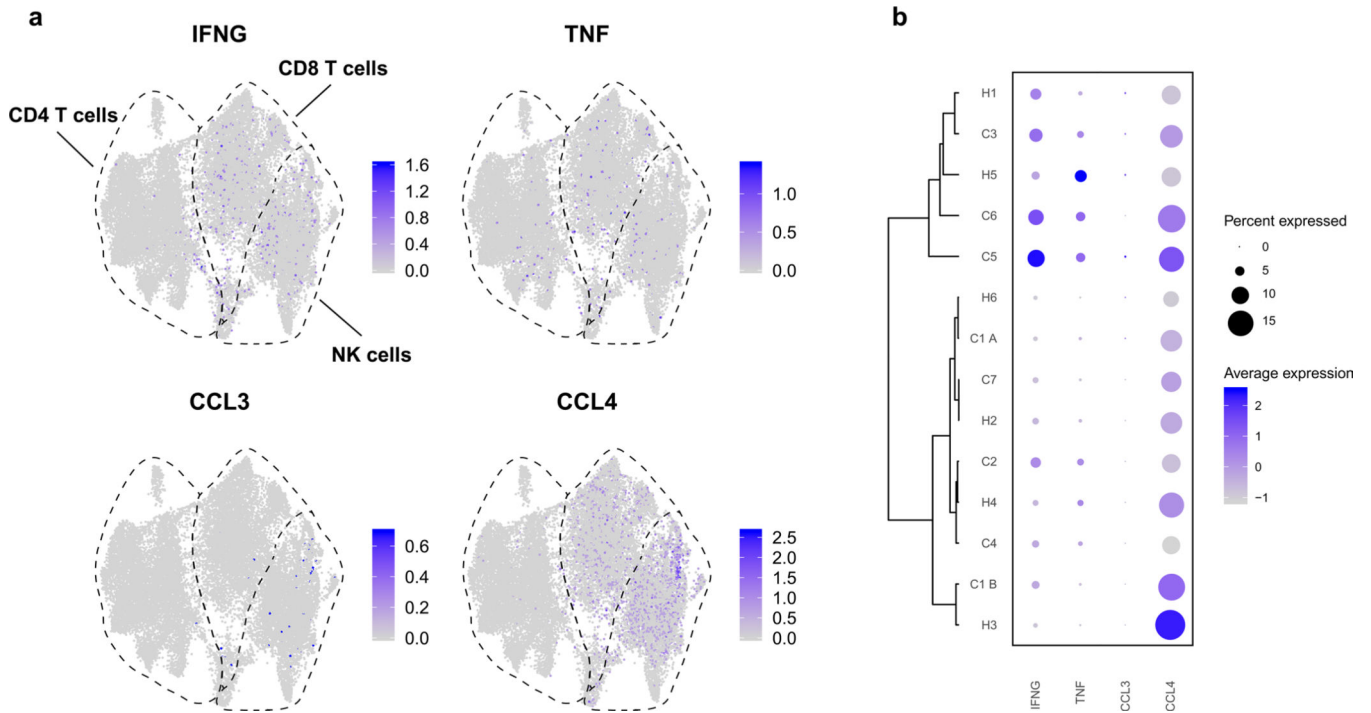
Extended Data Fig. 8 |. Expression of pro-inflammatory cytokines is not a feature of circulating cytotoxic lymphocytes in COVID-19.

a, UMAP embedding of CD4$^+$ T, CD8$^+$ T, and NK cells colored by expression of canonical pro-inflammatory cytokines *IFNG, TNF, CCL3*, and *CCL4*. n = 22,016 cells are plotted from n = 14 biologically independent samples. Dotted lines correspond to the cell type boundaries identified in Fig. 3b. **b**, Dot plot depicting average and percent expression of canonical cytotoxic lymphocyte pro-inflammatory cytokines by CD8$^+$ T cells from n = 8 COVID-19 samples and n = 6 healthy controls.
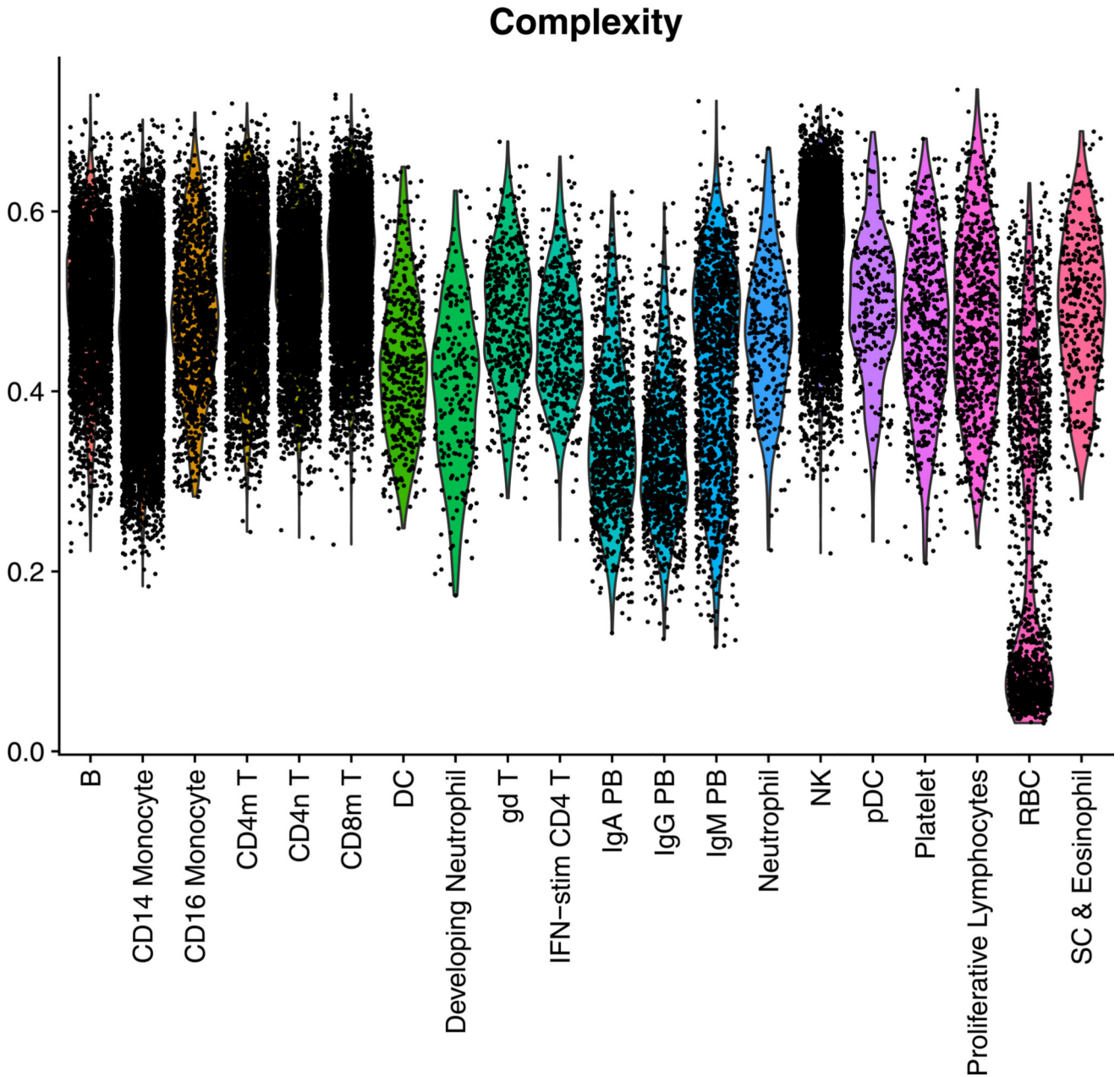
Complexity

Extended Data Fig. 9 |. Developing neutrophils likely do not represent multiplets.
Complexity, defined as the number of genes detected per cell divided by the number of
UMIs in that cell, for each cell in the dataset grouped by annotated cell type. Complexity of
all $n = 44,721$ cells from $n = 8$ COVID-19 samples and $n = 6$ healthy controls is depicted.
Each violin plot is trimmed at the maximum and minimum value for each cell type.
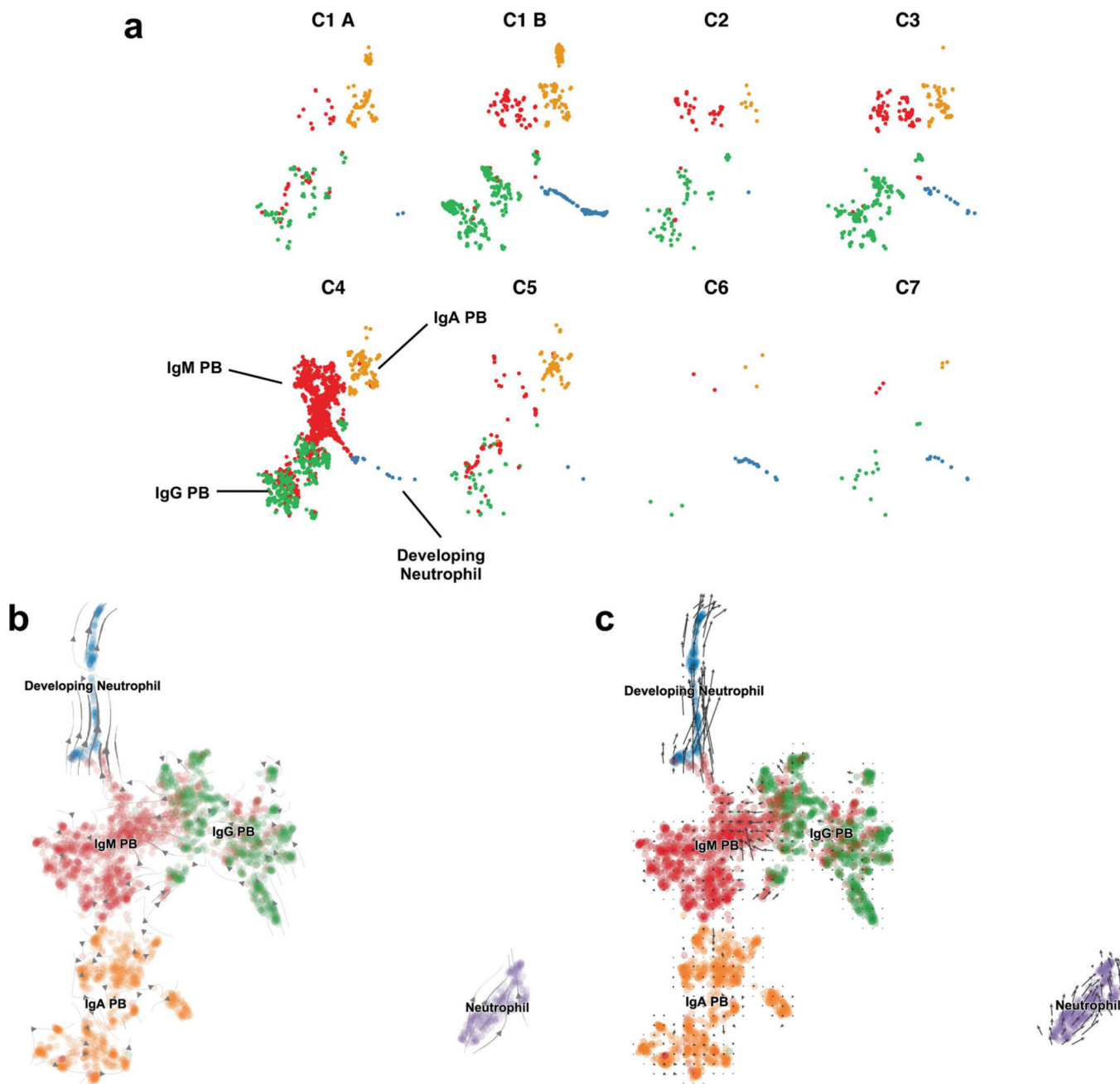
Extended Data Fig. 10 |. The spectrum of plasmablast-to-granulocyte phenotype is present in patients experiencing ARDS and this putative transdifferentiation is likely unrelated to transcriptional dynamics of canonical neutrophils.

a, UMAP embedding of plasmablasts and activated granulocytes faceted by COVID-19 sample. $n = 3,187$ cells are plotted from $n = 14$ biologically independent samples. **b-c**, UMAP embeddings of plasmablasts, activated granulocytes, and canonical neutrophils overlaid with RNA velocity stream (**b**) or velocity grid (**c**). $n = 3,911$ cells are plotted from $n = 14$ biologically independent samples.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.
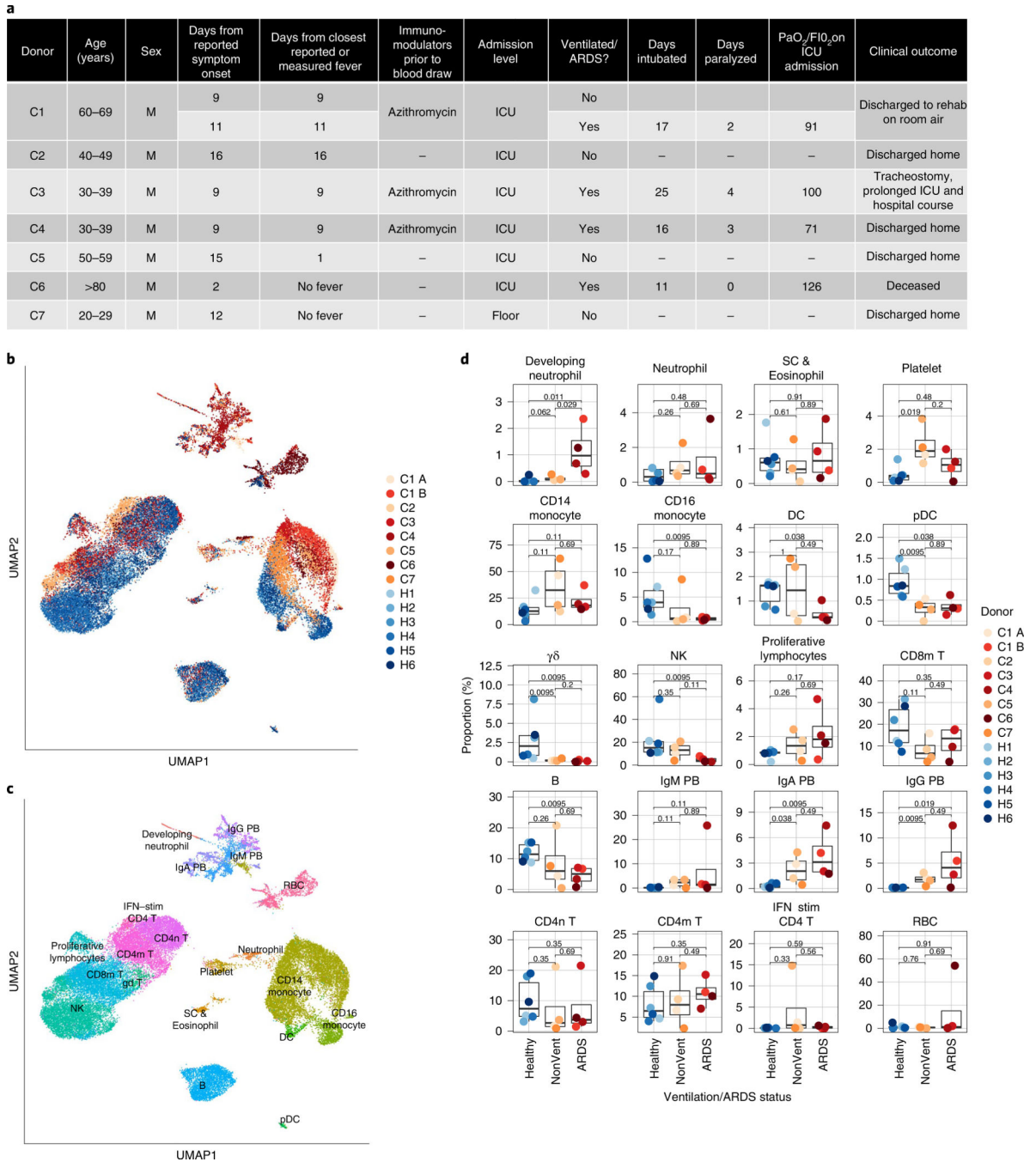
Acknowledgements

References

1. Dong E, Du H & Gardner L An interactive web-based dashboard to track COVID-19 in real time. Lancet Infect. Dis 20, 533–534 (2020). [PubMed: 32087114]

2. Wu Z & McGoogan JM Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China: summary of a report of 72314 cases from the Chinese Center for Disease Control and Prevention. JAMA 323, 1239–1242 (2020).

3. Huang C et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet 395, 497–506 (2020). [PubMed: 31986264]

4. Chen G et al. Clinical and immunologic features in severe and moderate Coronavirus Disease 2019. J. Clin. Invest (in the press).

5. Zhou Y et al. Pathogenic T cells and inflammatory monocytes incite inflammatory storm in severe COVID-19 patients. Natl Sci. Rev (in the press).

6. Guo C et al. Tocilizumab treatment in severe COVID-19 patients attenuates the inflammatory storm incited by monocyte centric immune interactions revealed by single-cell analysis. Preprint at bioRxiv 10.1101/2020.04.08.029769 (2020).

7. Guan W-J et al. Clinical characteristics of coronavirus disease 2019 in China. N. Engl. J. Med 382, 1708–1720 (2020). [PubMed: 32109013]

8. Qin C et al. Dysregulation of immune response in patients with COVID-19 in Wuhan, China. Clin. Infect. Dis (in the press).

9. Zheng H-Y et al. Elevated exhaustion levels and reduced functional diversity of T cells in peripheral blood may predict severe progression in COVID-19 patients. Cell. Mol. Immunol (in the press).

10. Diao B et al. Reduction and functional exhaustion of T cells in patients with coronavirus disease 2019 (COVID-19). Front. Immunol 11, 827 (2020). [PubMed: 32425950]

11. Gierahn TM et al. Seq-Well: portable, low-cost RNA sequencing of single cells at high throughput. Nat. Methods 14, 395–398 (2017). [PubMed: 28192419]

12. Kazer SW et al. Integrated single-cell analysis of multicellular immune dynamics during hyperacute HIV-1 infection. Nat. Med 26, 511–518 (2020). [PubMed: 32251406]

13. Zimmermann P, Ziesenitz VC, Curtis N & Ritz N The immunomodulatory effects of macrolides—a systematic review of the underlying mechanisms. Front. Immunol 9, 302 (2018). [PubMed: 29593707]

14. Zhao J et al. Antibody responses to SARS-CoV-2 in patients of novel coronavirus disease 2019. Clin. Infect. Dis (in the press).

15. Zhang B et al. Immune phenotyping based on neutrophil-to-lymphocyte ratio and IgG predicts disease severity and outcome for patients with COVID-19. Infect. Dis (in the press).

16. Cassatella MA, Östberg NK, Tamassia N & Soehnlein O Biological roles of neutrophil-derived granule proteins and cytokines. Trends Immunol. 40, 648–664 (2019). [PubMed: 31155315]

17. Zhu YP et al. Identification of an early unipotent neutrophil progenitor with pro-tumoral activity in mouse and human bone marrow. Cell Rep. 24, 2329–2341.e8 (2018). [PubMed: 30157427]

18. Xie X et al. Single-cell transcriptome profiling reveals neutrophil heterogeneity and orchestrated maturation during homeostasis and bacterial infection. Preprint at bioRxiv https://www.biorxiv.org/content/10.1101/792200v1 (2019).

19. Ong EZ et al. A dynamic immune response shapes COVID-19 progression. Cell Host Microbe (in the press).

20. Giamarellos-Bourboulis EJ et al. Complex immune dysregulation in COVID-19 patients with severe respiratory failure. Cell Host Microbe (2020).

21. Caligiuri MA Human natural killer cells. Blood 112, 461–469 (2008). [PubMed: 18650461]

22. Wilk AJ & Blish CA Diversification of human NK cells: lessons from deep profiling. J. Leukoc. Biol 103, 629–641 (2018). [PubMed: 29350874]

23. Poli A et al. CD56bright natural killer (NK) cells: an important NK cell subset. Immunology 126, 458–465 (2009). [PubMed: 19278419]

24. Hydes T et al. IL-12 and IL-15 induce the expression of CXCR6 and CD49a on peripheral natural killer cells. Immun. Inflamm. Dis 6, 34–46 (2018). [PubMed: 28952190]

25. Smith SL et al. Diversity of peripheral blood human NK cells identified by single-cell RNA sequencing. Blood Adv. 4, 1388–1406 (2020). [PubMed: 32271902]

26. Gars ML et al. Pregnancy-induced alterations in NK cell phenotype and function. Front. Immunol. 10, 2469 (2019). [PubMed: 31708922]

27. Bianco Mello D et al. Imbalanced host response to SARS-CoV-2 drives development of COVID-19. Cell S0092–8674, 30489–3048 (2020).

28. Hadjadj J et al. Impaired type I interferon activity and exacerbated inflammatory responses in severe Covid-19 patients. Preprint at https://www.medrxiv.org/content/10.1101/2020.04.19.20068015v1 (2020).

29. La Manno G et al. RNA velocity of single cells. Nature 560, 494–498 (2018). [PubMed: 30089906]

30. Bergen V, Lange M, Peidli S, Wolf FA & Theis FJ Generalizing RNA velocity to transient cell states through dynamical modeling. Preprint at bioRxiv https://www.biorxiv.org/content/10.1101/820936v1 (2019).

31. Xie H, Ye M, Feng R & Graf T Stepwise reprogramming of B cells into macrophages. Cell 117, 663–676 (2004). [PubMed: 15163413]

32. Cirovic B et al. C/EBP-induced transdifferentiation reveals granulocyte-macrophage precursor-like plasticity of B cells. Stem Cell Rep. 8, 346–359 (2017).

33. Lekstrom-Himes JA The role of C/EBP(epsilon) in the terminal stages of granulocyte differentiation. Stem Cells 19, 125–133 (2001). [PubMed: 11239167]

34. Balamurugan K & Sterneck E The many faces of C/EBPδ and their relevance for inflammation and cancer. Int. J. Biol. Sci 9, 917–933 (2013). [PubMed: 24155666]

35. Evrard M et al. Developmental analysis of bone marrow neutrophils reveals populations specialized in expansion, trafficking and effector functions. Immunity 48, 364–379.e8 (2018). [PubMed: 29466759]

36. Yvan-Charvet L & Ng LG Granulopoiesis and neutrophil homeostasis: a metabolic, daily balancing act. Trends Immunol. 40, 598–612 (2019). [PubMed: 31256783]

37. Gordon CJ et al. Remdesivir is a direct-acting antiviral that inhibits RNA-dependent RNA polymerase from severe acute respiratory syndrome coronavirus 2 with high potency. J. Biol. Chem 295, 6785–6797 (2020). [PubMed: 32284326]

38. Siegel D et al. Discoveryand synthesis of a phosphoramidate prodrug of a pyrrolo[2,1-f][triazin-4-amino]adenine C-nucleoside (GS-5734) for the treatment of Ebola and emerging viruses. J. Med. Chem 60, 1648–1661 (2017). [PubMed: 28124907]

39. Schmidt T et al. CD66b overexpression and homotypic aggregation of human peripheral blood neutrophils after activation by a gram-positive stimulus. J. Leukoc. Biol 91, 791–802 (2012). [PubMed: 22319104]

40. Ericson JA et al. Gene expression during the generation and activation of mouse neutrophils: implication of novel functional and regulatory pathways. PLoS ONE 9, e108553 (2014). [PubMed: 25279834]

41. ARDS Definition, TaskForce et al. Acute respiratory distress syndrome: the Berlin Definition. JAMA 307, 2526–2533 (2012). [PubMed: 22797452]

42. Hughes TK et al. Highly efficient, massively-parallel single-cell RNA-seq reveals cellular states and molecular features of human skin pathology. Preprint at bioRxiv https://www.biorxiv.org/content/10.1101/689273v1 (2019).

43. Dobin A et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29, 15–21 (2013). [PubMed: 23104886]

44. Petukhov V et al. dropEst: pipeline for accurate estimation of molecular counts in droplet-based single-cell RNA-seq experiments. Genome Biol. 19, 78 (2018). [PubMed: 29921301]

45. Butler A, Hoffman P, Smibert P, Papalexi E & Satija R Integrating single-cell transcriptomic data across different conditions, technologies and species. Nat. Biotechnol. 36, 411–420 (2018). [PubMed: 29608179]

46. Waltman L & van Eck NJ A smart local moving algorithm for large-scale modularity-based community detection. Eur. Phys. J. B 86, 471 (2013).

47. Carter RA et al. A single-cell transcriptional atlas of the developing murine cerebellum. Curr. Biol 28, 2910–2920.e2 (2018). [PubMed: 30220501]

48. Freytag S, Tian L, Lönnstedt I, Ng M & Bahlo M Comparison of clustering tools in R for medium-sized 10x Genomics single-cell RNA-sequencing data. F1000Res. 7, 1297 (2018). [PubMed: 30228881]

49. Gutierrez-Arcelus M et al. Lymphocyte innateness defined by transcriptional states reflects a balance between proliferation and effector functions. Nat. Commun 10, 687 (2019). [PubMed: 30737409]

50. Villani A-C et al. Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes and progenitors. Science 356, eaah4573 (2017). [PubMed: 28428369]

51. Kang HM et al. Multiplexed droplet single-cell RNA-sequencing using natural genetic variation. Nat. Biotechnol 36, 89–94 (2018). [PubMed: 29227470]

52. Nimmerjahn F & Ravetch JV Fcγ receptors as regulators of immune responses. Nat. Rev. Immunol 8, 34–47 (2008). [PubMed: 18064051]

53. Nimmerjahn F & Ravetch JV Fcγ receptors: old friends and new family members. Immunity 24, 19–28 (2006). [PubMed: 16413920]

54. Palmer C, Diehn M, Alizadeh AA & Brown PO Cell-type specific gene expression profiles of leukocytes in human peripheral blood. BMC Genomics 7, 115 (2006). [PubMed: 16704732]

55. Aran D et al. Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. Nat. Immunol 20, 163–172 (2019). [PubMed: 30643263]

56. Leonard S, Rolland A, Tarte K, Chalmel F & Lardenois A FlexDotPlot: a universal and modular dot plot visualization tool for complex multifaceted data. Preprint at bioRxiv https://www.biorxiv.org/content/10.1101/2020.04.03.023655v1 (2020).

57. Miller BC et al. Subsets of exhausted CD8[+] T cells differentially mediate tumor control and respond to checkpoint blockade. Nat. Immunol 20, 326–336 (2019). [PubMed: 30778252]

a

Donor	Age (years)	Sex	Days from reported symptom onset	Days from closest reported or measured fever	Immuno-modulators prior to blood draw	Admission level	Ventilated/ ARDS?	Days intubated	Days paralyzed	PaO₂/FI0₂ on ICU admission	Clinical outcome
C1	60–69	M	9	9	Azithromycin	ICU	No				Discharged to rehab on room air
			11	11			Yes	17	2	91	
C2	40–49	M	16	16	–	ICU	No	–	–	–	Discharged home
C3	30–39	M	9	9	Azithromycin	ICU	Yes	25	4	100	Tracheostomy, prolonged ICU and hospital course
C4	30–39	M	9	9	Azithromycin	ICU	Yes	16	3	71	Discharged home
C5	50–59	M	15	1	–	ICU	No	–	–	–	Discharged home
C6	>80	M	2	No fever	–	ICU	Yes	11	0	126	Deceased
C7	20–29	M	12	No fever	–	Floor	No	–	–	–	Discharged home

Fig. 1 |. Expansion of plasmablasts and depletion of multiple innate immune cell subsets in the periphery of patients with COVID-19.
a, Demographics, sample characteristics and disease course of patients with COVID-19. **b**, UMAP dimensionality reduction embedding of peripheral blood mononuclear cells (PBMCs) from all profiled samples ($n = 44,721$ cells) colored by donor of origin. IDs of patients with COVID-19 ($n = 7$) begin with 'C' and are colored in shades of orange (patients who were not ventilated at the time of draw) or red (patients with ARDS who were ventilated at the time of draw) and those of healthy donors begin with 'H' ($n = 6$) and are colored in blues. **c**, UMAP embedding of the entire dataset colored by orthogonally

generated clusters labeled by manual cell type annotation. **d**, Proportions of each cell type in each sample colored by donor of origin. The *x* axes correspond to the ventilation or ARDS status of each patient. Shown are exact two-sided *P* values by the Wilcoxon rank-sum test. *n* = 6, *n* = 4 and *n* = 4 biologically independent samples for Healthy, NonVent and ARDS, respectively. Boxplot features: minimum whisker, 25th percentile − 1.5 × inter-quartile range (IQR) or the lowest value within; minimum box, 25th percentile; center, median; maximum box, 75th percentile; maximum whisker, 75th percentile + 1.5 × IQR or greatest value within.
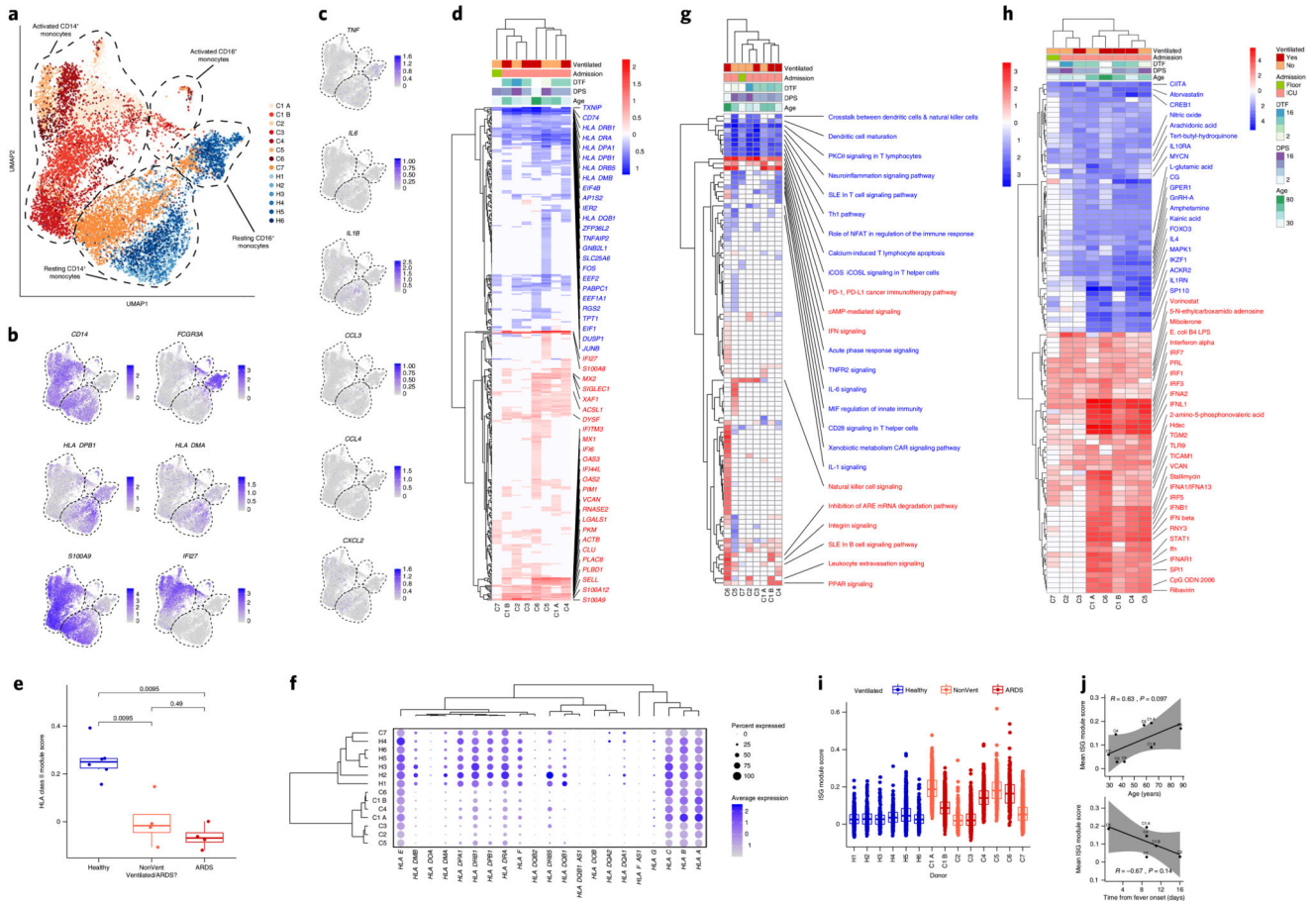
Fig. 2 |. Robust HLA class II downregulation and type I interferon-driven inflammatory signatures in monocytes are characteristics of SARS-CoV-2 infection.
**a**, UMAP embedding of all monocytes colored by sample of origin. $n = 10,339$ cells are plotted from $n = 14$ biologically independent samples. **b**, UMAP embedding of monocytes colored by *CD14* and *FCGR3A* (encoding CD16a, to distinguish between CD14+ and CD16+ monocytes), *HLA-DPB1* and *HLA-DMA* (illustrating HLA class II downregulation in patients with COVID-19) and *S100A9* and *IFI27* (demonstrating canonical inflammatory signatures in patients with COVID-19). **c**, UMAP embedding of monocytes colored by genes encoding pro-inflammatory cytokines previously reported to be produced by circulating monocytes in severe COVID-19[6], namely *TNF*, *IL6*, *IL1B*, *CCL3*, *CCL4* and *CXCL2*. **d,g,h**, Heatmaps of DE genes (**d**), differentially regulated canonical pathways (**g**) and differentially regulated predicted upstream regulators (**h**) between CD14+ monocytes of each COVID-19 sample compared to CD14+ monocytes of all healthy controls. The heatmap in **d** is colored by average log(fold-change), while heatmaps in **g** and **h** are colored by *z*-score. All displayed genes, pathways and regulators are statistically significant at the *P* < 0.05 confidence level by Seurat's implementation of the Wilcoxon rank-sum test (two-sided, adjusted for multiple comparisons using Bonferroni's correction, in **d**) or Ingenuity Pathway Analysis (IPA) implementation of the Fisher exact test (right-tailed, in **g** and **h**). The 50 genes (**d**), 25 pathways (**g**) or 50 regulators (**h**) with the highest absolute average log(fold-

change) or z-score across all donors are labeled. Genes with a net positive average log(fold-change) or z-score are labeled in red; genes with a net negative average log(fold-change) or z-score are labeled in blue. DPS, days post-symptom onset; DTF, days from first reported or measured fever. **e**, Boxplot showing the mean HLA class II module score of CD14$^+$ monocytes from each sample, colored by healthy donors (blue), non-ventilated patients with COVID-19 (orange) or ventilated patients with COVID-19 (red). Shown are exact P values by two-sided Wilcoxon rank-sum test. $n = 6$, $n = 4$ and $n = 4$ biologically independent samples for Healthy, NonVent and ARDS, respectively. **f**, Dot plot depicting percent expression and average expression of all detected *HLA* genes in CD14$^+$ monocytes by donor. **i**, Boxplot showing the *IFNA* module score of each cell, colored by healthy donors (blue), non-ventilated patients with COVID-19 (orange) or ventilated patients with COVID-19 (red). **j**, Scatter plots depicting the correlation between the mean ISG module score of CD14$^+$ monocytes in each sample and the patient age (top) and time–distance from first measured or reported fever (bottom). Shown are Pearson's r, exact two-sided P values and the 95% confidence interval. $n = 8$ (top) and $n = 6$ (bottom) independent biological samples. Number of cells for **d,f–i**: C1 A, 1,561; C1 B, 1,858; C2, 217; C3, 1,102; C4, 713; C5, 462; C6, 277; C7, 2,095; H1, 680; H2, 325; H3, 215; H4, 166; H5, 444; H6, 224. For **d,g–h**, cells from all healthy controls ($n = 2,054$ cells) are used to generate comparisons with each COVID-19 sample. For **e,i**, boxplot features: minimum whisker, 25th percentile − 1.5 × IQR or the lowest value within; minimum box, 25th percentile; center, median; maximum box, 75th percentile; maximum whisker, 75th percentile + 1.5 × IQR or greatest value within.
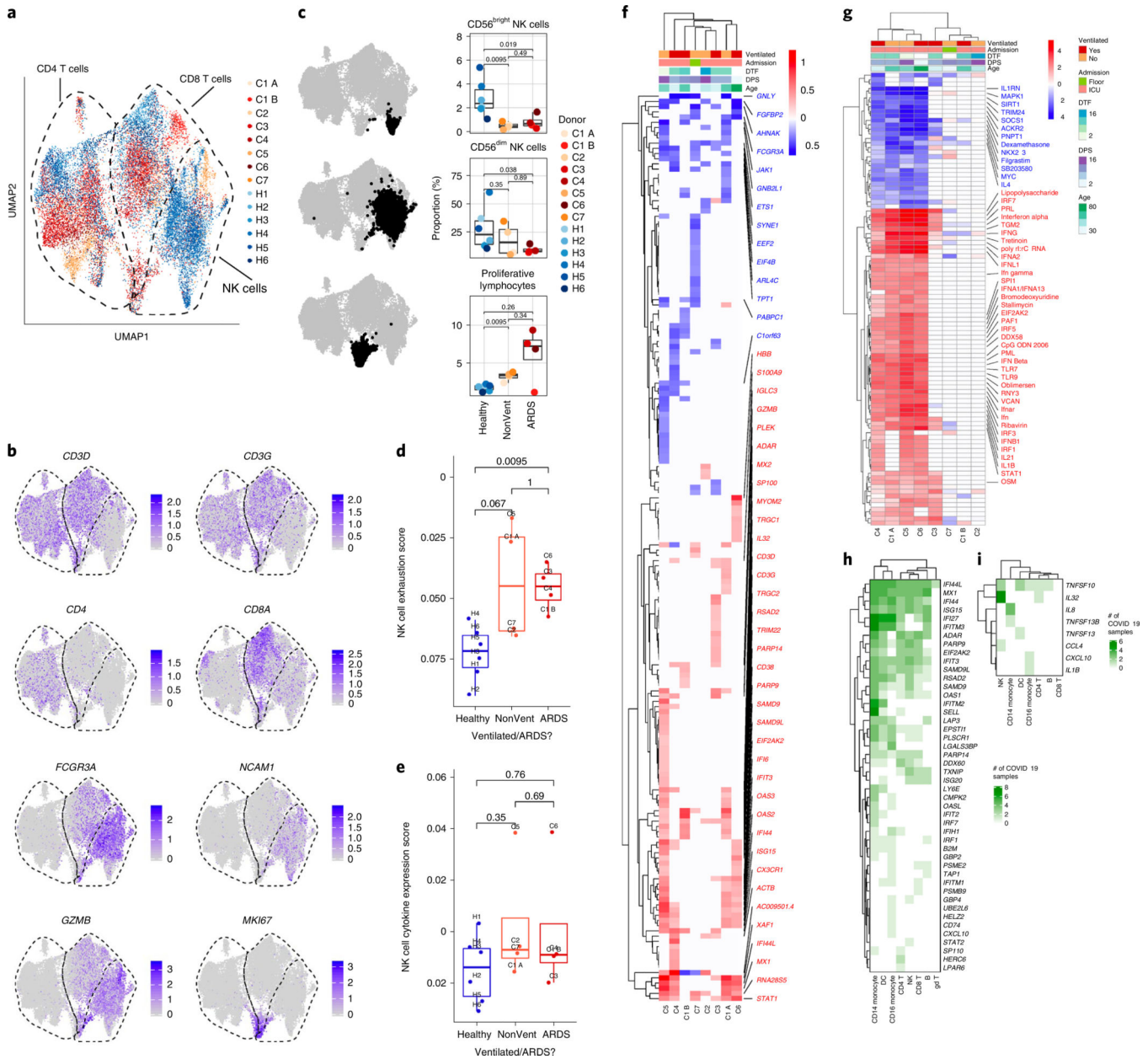
Fig. 3 |. Heterogeneous patterns of NK cell exhaustion and IFN response in COVID-19.
a, UMAP embedding of CD4$^+$ T cells, CD8$^+$ T cells and NK cells colored by sample of origin. **b**, UMAP embedding colored by lineage genes (*CD3D*, *CD3G*, *CD4*, *CD8A*, *FCGR3A* and *NCAM1*) and selected functional/phenotypic markers (*GZMB* and *MKI67*). For **a,b**, n = 22,016 cells are plotted from n = 14 biologically independent samples. **c**, Boxplots depicting proportions of CD56^{dim} NK cells, CD56^{bright} NK cells and proliferating lymphocytes among total T and NK cells by sample of origin. The cells used to calculate each proportion are highlighted in bold black in the adjacent UMAP embeddings and were identified by manually labeling clusters generated by clustering CD4$^+$ T cells, CD8$^+$ T cells and NK cells alone. Shown are exact two-sided P values from the Wilcoxon rank-sum test. n = 386 (top), n = 4,899 (middle), n = 781 (bottom) total from n = 6, n = 4 and n = 4

biologically independent samples for Healthy, NonVent and ARDS, respectively. **d**, Boxplot showing the mean expression score by only NK cells of three canonical markers of NK cell exhaustion: *LAG3*, *PDCD1* (encoding PD-1) and *HAVCR2* (encoding TIM-3). Shown are exact two-sided *P* values by Wilcoxon rank-sum test. **e**, Boxplot showing the mean expression score by only NK cells of four canonical NK cell cytokine genes (*CCL3*, *CCL4*, *IFNG* and *TNF*). Shown are exact *P* values by Wilcoxon rank-sum test. For **d,e**, $n = 6$, $n = 4$ and $n = 4$ biologically independent samples for Healthy, NonVent and ARDS, respectively. In **c**–**e**, boxplot features: minimum whisker, 25th percentile − 1.5 × IQR or the lowest value within; minimum box, 25th percentile; center, median; maximum box, 75th percentile; maximum whisker, 75th percentile + 1.5 × IQR or greatest value within. **f,g**, Heatmaps of DE genes (**f**) and differentially regulated predicted upstream regulators (**g**) between NK cells of each COVID-19 sample compared to NK cells of all healthy controls. As in Fig. 2, **f** is colored by average log(fold-change), while **g** is colored by *z*-score. All displayed genes and regulators are statistically significant at the $P < 0.05$ confidence level by Seurat's implementation of the Wilcoxon rank-sum test (two-sided, adjusted for multiple comparisons using Bonferroni's correction, **f**) or IPA's implementation of Fisher exact test (right-tailed, **g**). The 50 genes or regulators with the highest absolute average log(fold-change) or *z*-score across all donors are labeled. Genes with a net positive average log(fold-change) or *z*-score are labeled in red; genes with a net negative average log(fold-change) or *z*-score are labeled in blue. DPS, days post-symptom onset; DTF, days from first reported or measured fever. Number of cells for **f,g**: C1 A, 354; C1 B, 387; C2, 271; C3, 328; C4, 104; C5, 518; C6, 58; C7, 130; cells from all healthy controls ($n = 4{,}707$ cells) were used to generate comparisons with each COVID-19 sample. **h,i**, Heatmaps of differentially upregulated ISGs (**h**; Supplementary Table 25) and cytokines (**i**; Supplementary Table 25) in donors with COVID-19, colored by the number of COVID-19 samples in which the gene was differentially expressed relative to all healthy controls. DE genes used to construct these heatmaps are provided in Supplementary Tables 4–10. An ISG or cytokine was counted as differentially expressed if it had an average log(fold-change) > 0.25 and an adjusted two-sided *P* value < 0.05 by Seurat's implementation of the Wilcoxon rank-sum test. $n = 8$ biologically independent COVID-19 samples compared to $n = 6$ biologically independent healthy controls.
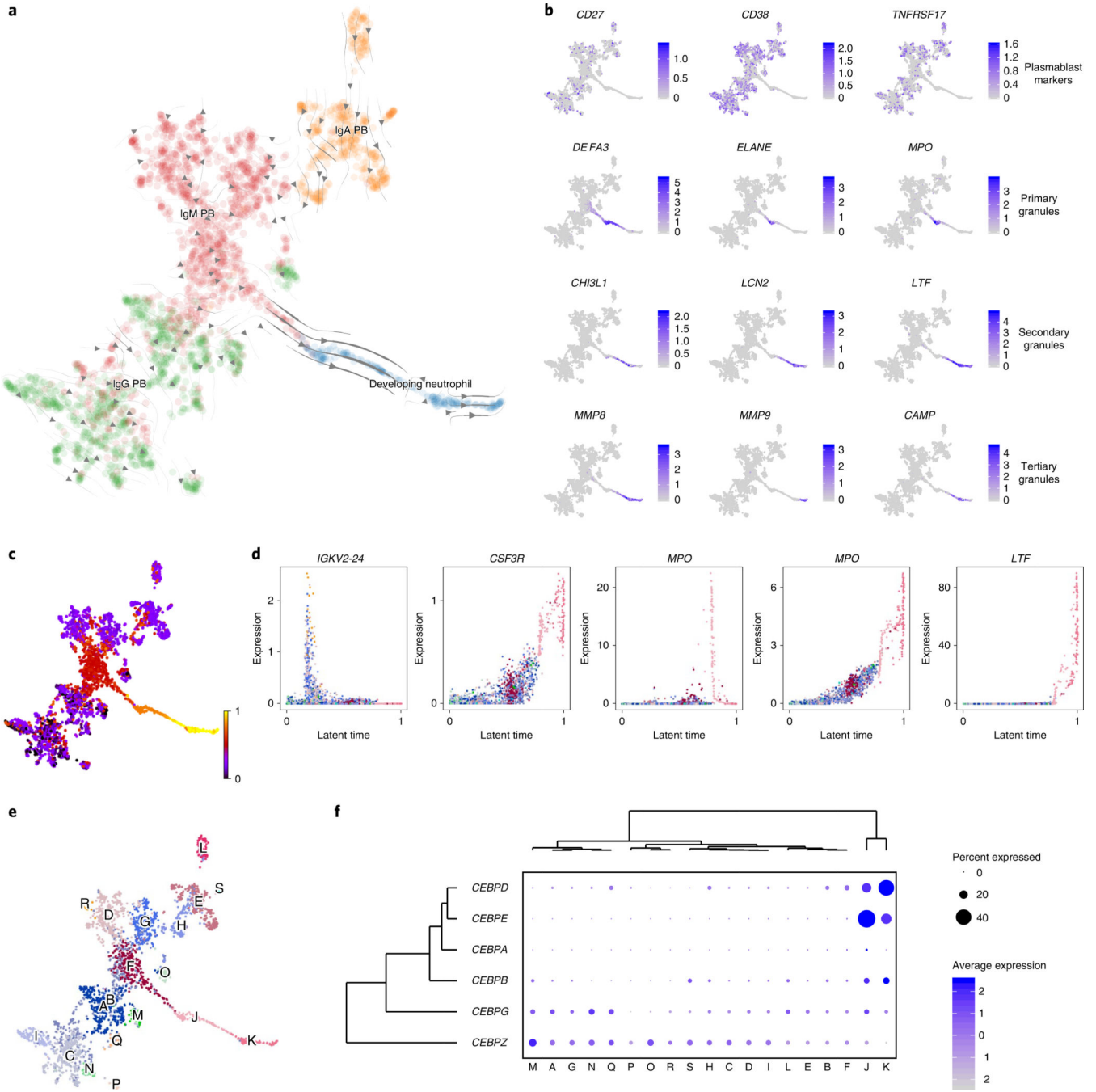
Fig. 4 |. Developing neutrophils are characteristic of patients with severe COVID-19 and may differentiate from plasmablasts.
a, UMAP embedding of plasmablasts and developing neutrophils, colored by annotated cell type and overlaid with the RNA velocity stream. **b**, UMAP embedding colored by canonical plasmablast marker genes (*CD27*, *CD38* and *TNFRSF17*) and genes encoding primary (*DEF3A*, *ELANE* and *MPO*), secondary (*CHI3DL1*, *LCN2* and *LTF*) and tertiary (*MMP8*, *MMP9* and *CAMP*) neutrophil granule proteins[35,39,40]. **c**, UMAP embedding colored by inferred latent time. **d**, Scatter plots showing expression of a selection of cluster-defining

genes across inferred latent time. **e**, UMAP embedding colored by orthogonally generated clusters. **f**, Dot plot depicting expression of *CEBP* family members in each identified cluster. For all panels, $n = 3{,}187$ cells from $n = 8$ biologically independent COVID-19 samples and $n = 6$ biologically independent healthy controls.