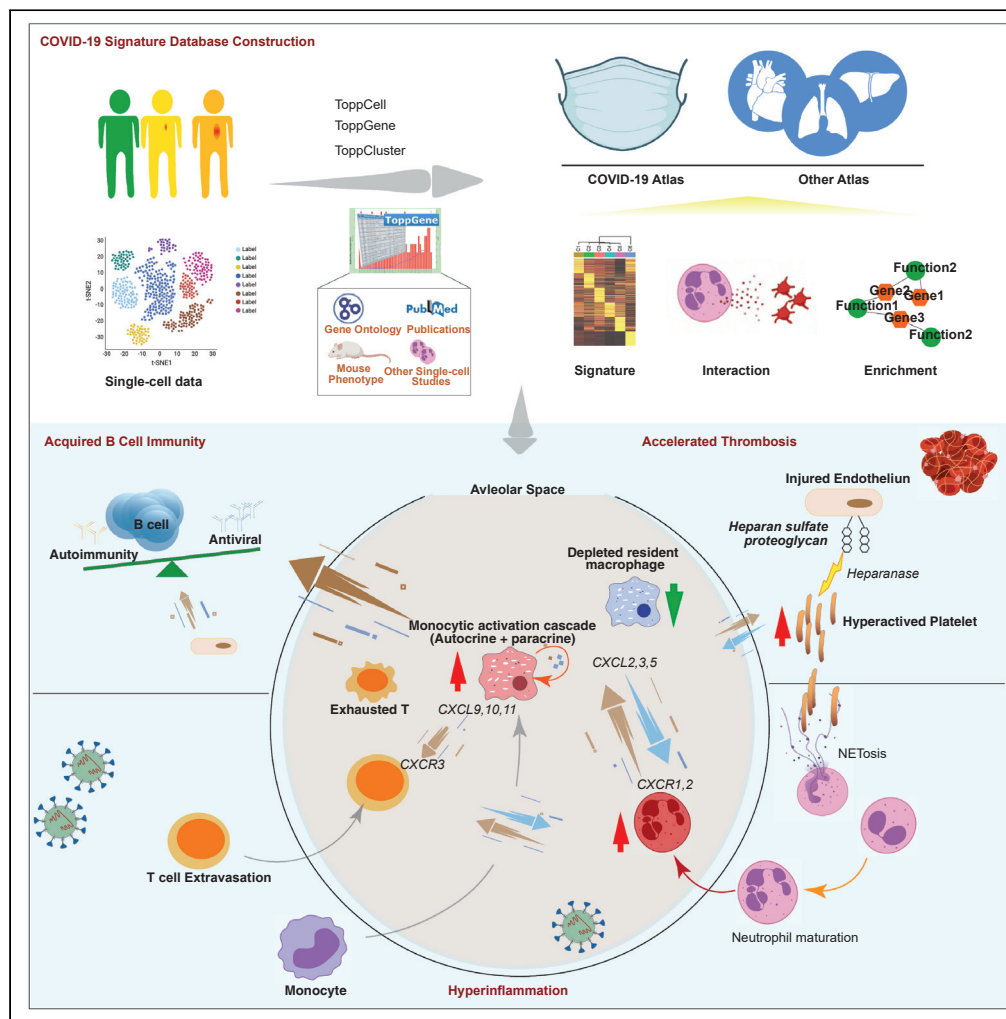## Article

An interactive single cell web portal identifies gene and cell networks in COVID-19 host responses

Kang Jin, Eric E. Bardes, Alexis Mitelpunkt, ..., Daniel Pomeranz Krummel, Marc E. Rothenberg, Bruce J. Aronow

bruce.aronow@cchmc.org

Highlights

Topp-toolkit creates the first COVID-19 immune signature atlas

Monocytic cell subtypes form a signaling cascade capable of creating the cytokine storm of severe COVID-19

Activated platelets display dysregulated expression of adhesion, activation, and coagulation genes with upregulated heparanase

Autoantibody generation outside of germinal centers is promoted by multilineage multicellular activator network of autoimmunity-associated genes

Article

An interactive single cell web portal identifies gene and cell networks in COVID-19 host responses

Kang Jin,[1,2] Eric E. Bardes,[1] Alexis Mitelpunkt,[1,3,4] Jake Y. Wang,[1] Surbhi Bhatnagar,[1,5] Soma Sengupta,[6] Daniel Pomeranz Krummel,[6] Marc E. Rothenberg,[7] and Bruce J. Aronow[1,5,8,9,*]

SUMMARY

Numerous studies have provided single-cell transcriptome profiles of host responses to SARS-CoV-2 infection. Critically lacking however is a data mine that allows users to compare and explore cell profiles to gain insights and develop new hypotheses. To accomplish this, we harmonized datasets from COVID-19 and other control condition blood, bronchoalveolar lavage, and tissue samples, and derived a compendium of gene signature modules per cell type, subtype, clinical condition, and compartment. We demonstrate approaches to interacting with, exploring, and functional evaluating these modules via a new interactive web portal ToppCell (http://toppcell.cchmc.org/). As examples, we develop three hypotheses: (1) alternatively-differentiated monocyte-derived macrophages form a multicelllar signaling cascade that drives T cell recruitment and activation; (2) COVID-19-generated platelet subtypes exhibit dramatically altered potential to adhere, coagulate, and thrombose; and (3) extrafollicular B maturation is driven by a multilineage cell activation network that expresses an ensemble of genes strongly associated with risk for developing post-viral autoimmunity.

INTRODUCTION

COVID-19 clinical outcomes are variable. The poorer outcomes due to this infection are highly associated with immunological and inflammatory responses to SARS-Cov-2 infection (Shi et al., 2020; Tay et al., 2020) and many recent single cell expression profiling studies have characterized patterns of immunoinflammatory responses among individuals, mostly during acute infection phases. Different studies have revealed a spectrum of responses that range from lymphopenia (Cao, 2020; Wang et al., 2020), cytokine storms (Mehta et al., 2020; Pedersen and Ho, 2020), differential interferon responses (Blanco-Melo et al., 2020; Hadjadj et al., 2020) and emergency myelopoiesis (Schulte-Schrepping et al., 2020; Silvin et al., 2020). However, a variety of obstacles limit the ability of the research and medical communities to explore and compare these studies to pursue additional questions and gain additional insights that could improve our understanding of cell type specific responses to SARS-Cov-2 infection and their impact on clinical outcome.

Although many studies have focused on the peripheral blood mononuclear cells (PBMC) (Arunachalam et al., 2020; Guo et al., 2020; Lee et al., 2020; Schulte-Schrepping et al., 2020; Wilk et al., 2020) because of ease of procurement, other studies have profiled airway locations via bronchoalveolar lavage (BAL) (Grant et al., 2021; Liao et al., 2020), nasopharyngeal swabs, and bronchial brushes (Chua et al., 2020). Additional sampling sites that could also be infected or affected have also been approached in autopsy-derived materials from the central nervous system (Heming et al., 2021; Yang et al., 2021), and other sites (Delorey et al., 2021). Moreover, as major COVID-19 consortiums working on the collection and integration of each of their individual studies and interpreting important features of these individual datasets as downloadable datasets or browsable versions, such as single cell portal (https://singlecell.broadinstitute.org/single_cell/covid19) and COVID-19 Cell Atlas (https://www.covid19cellatlas.org/), using these data beyond markers, cell types, and individual signatures is either not possible or not accomplishable across-datasets. Thus, a well-organized and systematic study of immune cells across tissues for in-depth biological explorations is an unmet need for a deeper understanding of the underlying basis of the breadth of COVID-19 host defense and pathobiology.

Here we harmonized and analyzed eight high quality publicly available single-cell RNA-seq datasets from COVID-19 and immunologically-related studies that in total covered more than 480,000 cells isolated from

[1]Division of Biomedical Informatics, Cincinnati Children's Hospital Medical Center, Cincinnati, OH 45229, USA

[2]Department of Biomedical Informatics, University of Cincinnati, Cincinnati, OH 45229, USA

[3]Pediatric Rehabilitation, Dana-Dwek Children's Hospital, Tel Aviv Medical Center, Tel Aviv, 6423906, Israel

[4]Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, 6997801, Israel

[5]Department of Electrical Engineering and Computer Science, University of Cincinnati, Cincinnati, OH 45221, USA

[6]Department of Neurology and Rehabilitation Medicine, University of Cincinnati College of Medicine, Cincinnati, OH 45267, USA

[7]Division of Allergy and Immunology, Department of Pediatrics, Cincinnati Children's Hospital Medical Center, University of Cincinnati, Cincinnati, OH 45229, USA

[8]Department of Pediatrics, University of Cincinnati School of Medicine, Cincinnati, OH 45256, USA

[9]Lead contact

*Correspondence: bruce.aronow@cchmc.org

https://doi.org/10.1016/j.isci.2021.103115

A

Data Collection

Data Source	Subject	Sample	Data Availability
Liao et al.	4 healthy donors 3 moderate COVID-19 patients 6 severe COVID-19 patients	BAL	GSE145926
Wilk et al.	6 healthy donors 7 COVID-19 patients (1 in floor and 6 in ICU)	PBMC	GSE150728
Bharat et al.	2 healthy donors 3 severe COVID-19 donors	Post-mortem/ Explanted Lung Biopsy	GSE158127
Grant et al.	1 healthy donor 5 severe COVID-19 patients	BAL	GSE155249
Lee et al.	4 healthy donors 8 COVID-19 patients (asymptomatic, mild, severe) 5 severe influenza patients	PBMC	GSE149689
Arunachalam et al.	5 healthy donors 3 moderate COVID-19 patients 4 severe COVID-19 patients	PBMC	GSE155673
Schulte-Schrepping et al.	21 healthy donors 8 mild COVID-19 patients 10 severe COVID-19 patients	PBMC	EGAS00001004571
Guo et al.	2 severe COVID-19 patients (severe and remission stages)	PBMC	GSE150861
Reyes et al.	36 healthy donors 29 sepsis patients	PBMC	SCP548, SCP550
Schafflick et al.	4 healthy donors 4 multiple sclerosis patients	PBMC	GSE138266

* More details in *toppcell.cchmc.com* and Table.S1

B
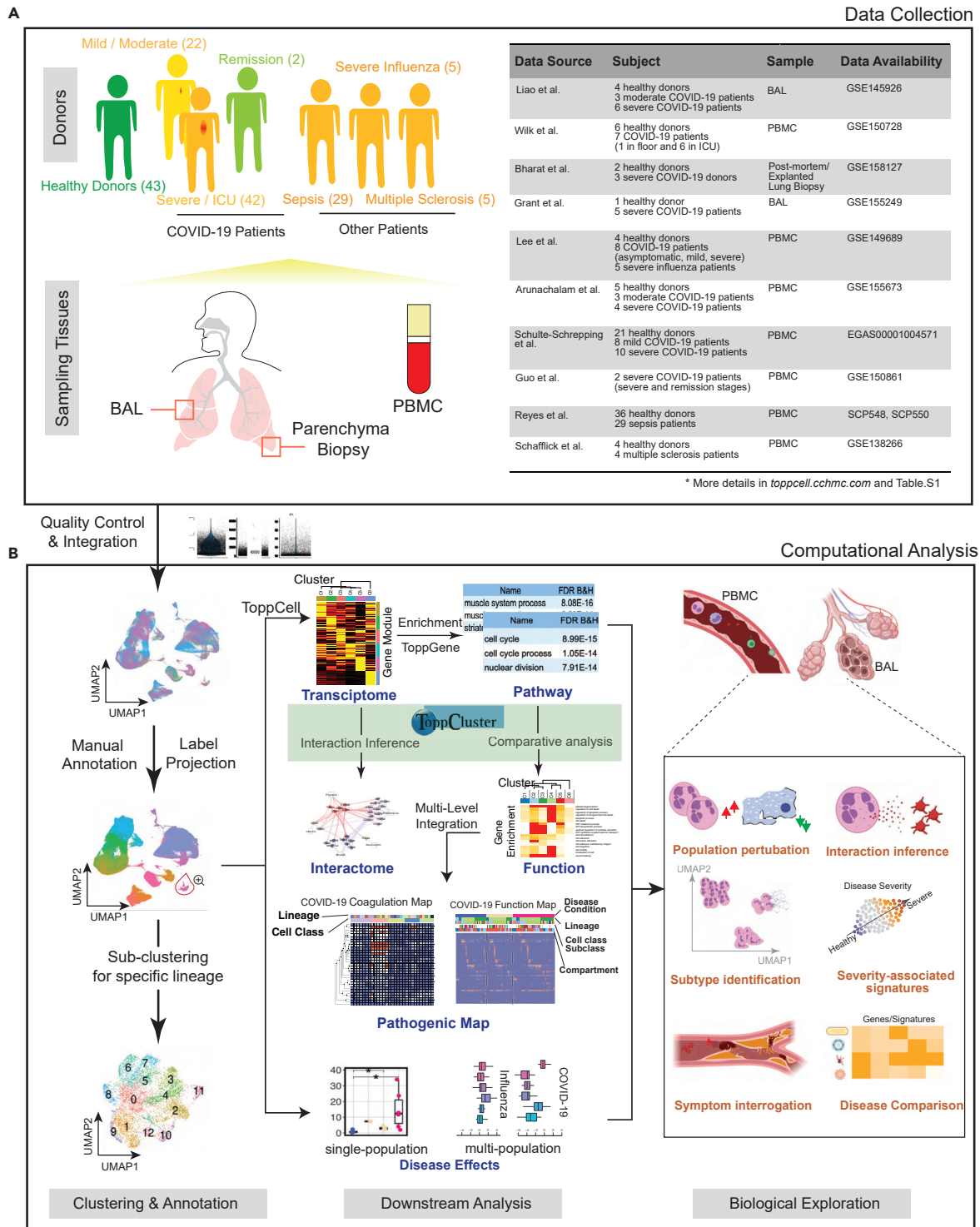
Computational Analysis

Figure 1. Creating a COVID-19 signature atlas

(A) Representative aggregation of multiple single-cell RNA-sequencing datasets from COVID-19 and related studies. The present study is derived from a total of 231,800 peripheral blood mononuclear cells (PBMCs), 101,800 bronchoalveolar lavage (BAL) cells and 146,361 lung parenchyma cells from 43 healthy; 22 mild, 42 severe, and 2 convalescent patients. Data was collated from eight public datasets (right).

(B) Data analysis pipeline of the study using Topp-toolkit. It includes three phases: (1) clustering and annotation; (2) downstream analysis using Topp-toolkit; (3) biological exploration. Output includes the evaluation of abundance of cell populations, cell type (cluster) specific gene modules, functional associations

Figure 1. *Continued*

of disease-associated cell classes and clusters, inference of cell-cell interactions, as well as comparative analysis across diseases, including influenza, sepsis, and multiple sclerosis. Additional newer datasets not included in this manuscript are present and will continue to be added to ToppCell (http://toppcell.cchmc.org). See also Table S1.

peripheral blood, bronchial alveolar lavage and lung parenchyma samples, and assembled an integrated COVID-19 atlas (https://toppcell.cchmc.org/). We established a framework for deriving, characterizing, and establishing reference gene expression signatures from these harmonized datasets using modular and hierarchical approaches based on signatures per class, subclass, and signaling/activation and clinical status per each sample group. Leveraging these gene expression signature modules, we demonstrate data mining approaches that allow for the identification of a series of fundamental disease processes: (1) an intercellular monocytic activation cascade capable of mediating the emergence of hyperinflammatory monocyte-derived alveolar macrophages in severe COVID-19 patients; (2) the generation of several alternatively differentiated platelet subtypes with dramatically different expression of sets of genes associated with critical platelet tasks capable of altering vascular and tissue responses to infectious agents; and (3) a multilineage and multi cell type cooperative signaling network with the potential to drive extrafollicular B maturation at a lesion site, but do so with high risk for the development of B cell-associated immunity. Additionally, immune hallmarks of COVID-19 patients were compared with other immune-mediated diseases using single-cell data from patients with influenza, sepsis, or multiple sclerosis. Consistent and varied compositional and gene patterns were identified across these implicating striking COVID-19 effects in some individuals.

RESULTS

Creating the first COVID-19 signature atlas using ToppCell portal

To have a comprehensive coverage of cells, we collated single-cell data of COVID-19 patients from eight public datasets, which in total contains 231,800 PBMCs, 101,800 BAL cells, and 146,361 lung parenchyma cells from donors: 43 healthy; 22 mild; 42 severe; and 2 convalescent patients (Figure 1A, Table S1).

To assemble an integrated atlas of human cell responses to COVID-19, we sought to harmonize metadata encompassing clinical information, sampling compartments, and cell and gene expression module designations. Doing so provides a rich framework for detecting perturbations of cell repertoire and differentiative state adaptations. We first integrated single cell RNA-seq data in Seurat (Stuart et al., 2019) and annotated cell types using canonical markers (Table S2). Further annotations of B cell and T cell subtypes were completed using the reference-based labeling tool Azimuth (Hao et al., 2021). Sub-clustering was applied for some cell types, such as neutrophils and platelets, to interrogate finer resolutions of disease-specific sub-populations (Figure 1B). Using the ToppCell toolkit (https://toppcell.cchmc.org/), we created an atlas of more than 3,000 hierarchically organized gene modules corresponding to differentially expressed genes (DEGs) derived from all cell classes and sub-classes among all compartments and clinical subgroups such as disease severity levels (Table S1). These modules were then used to infer gene networks within celltypes and subtypes as well as cell-cell interactions that could be further combined for functional comparative enrichment, interactions, and fuzzy network AI-based analyses using ToppCluster (Kaimal et al., 2010) and ToppGene (Chen et al., 2009) (Figure 1B), such as differential functional enrichments of sub-clusters of platelets. Integration of ToppCluster output of cells from multiple compartments and disease conditions built pathogenic heatmaps and networks, highlighted by the coagulation map of COVID-19 (Figure S12). In addition, perturbation of cell abundance was evaluated either in one cell population, or in multiple populations across diseases. Taken together, we investigated cell abundance changes, severity-associated signatures, mechanisms of COVID-19 specific symptoms and unique features of COVID-19 as an immune-mediated disease (Figure 1B).

Dynamic changes and balance of COVID-19 immune repository in blood and lung

After the aforementioned cell annotation procedure, we identified 28 and 24 distinct cell types in PBMC and BAL, respectively (Figures 2A and 2C; Table S2). Shifts of Uniform Manifold Approximation and Projection (UMAP) of cell type distributions were observed in both compartments of mild and severe patients (Figures 2A, 2C, S1A and 3A). In PBMC, conventional dendritic cells (cDC), plasmacytoid dendritic cells (pDC), and non-classical monocytes displayed a prominent reduction in severe patients (Figures 2B and S1C), consistent with prior reports (He et al., 2020; Laing et al., 2020; Wilk et al., 2020). In contrast, severe patients demonstrated dramatic expansion of neutrophils, especially immature stages (Figures S1C and

A PBMC Cell Class

B
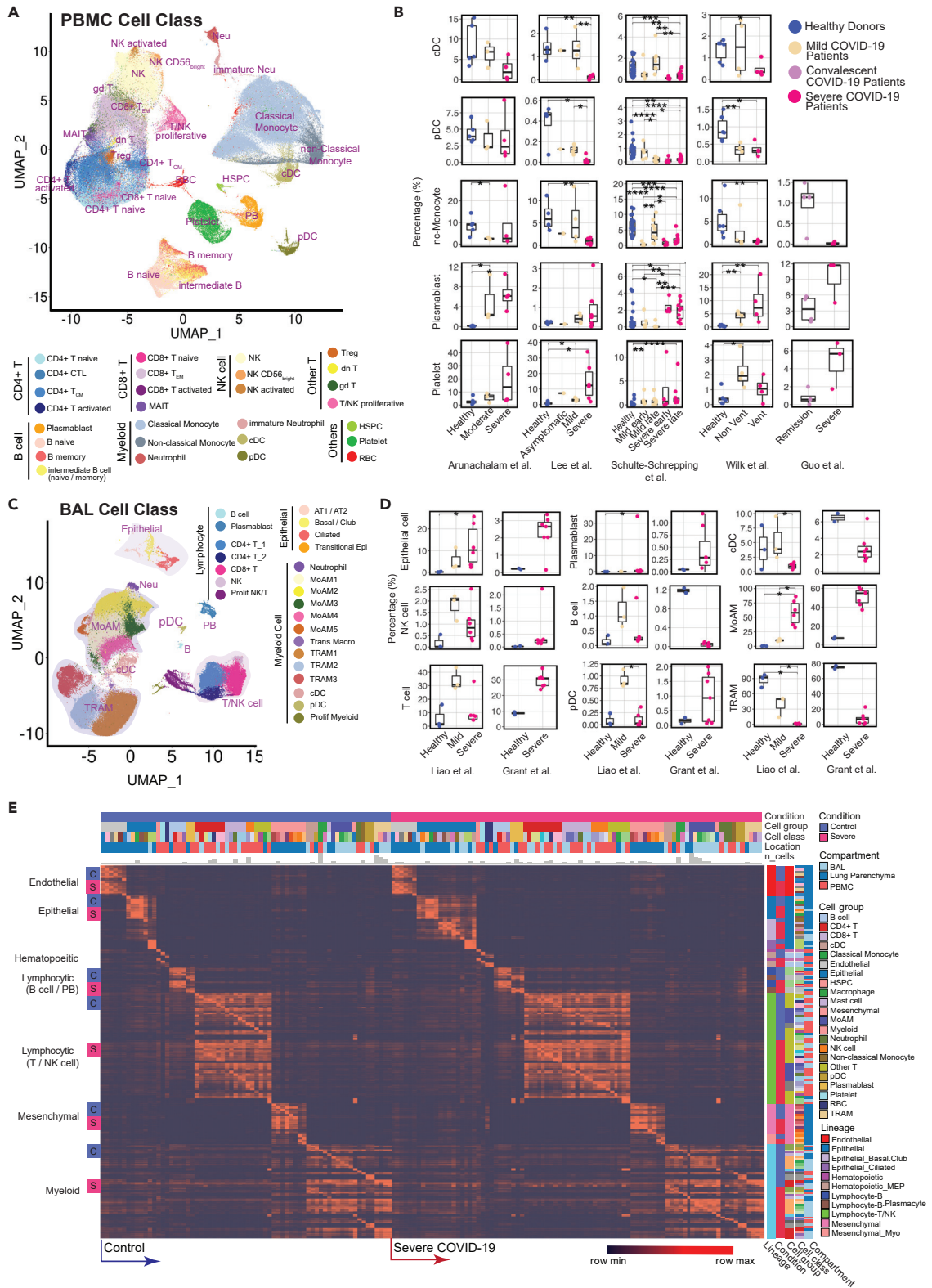
C BAL Cell Class

D

E

Figure 2. Modularized representation of cell type specific gene signatures and dynamic changes of cell abundance

(A) Uniform Manifold Approximation and Projection (UMAP) of 28 distinct cell types identified in the integrated peripheral blood mononuclear cell (PBMC) data.

(B) Comparative analysis of cell abundance effects of COVID-19. Reproducible multi-study data present high impact effects on 5 cell types in PBMC. Percentages of selected cell types in each sample are shown (where Vent: Ventilated patients; Non Vent: Non-ventilated patients). Significance between two conditions was measured by the Mann–Whitney rank-sum test (Wilcoxon, paired = False), which was also used in following significance tests of cell abundance changes in this study. *: $p \leq 0.05$; **: $p \leq 0.01$; ***: $p \leq 0.001$; ****: $p \leq 0.0001$. Boxplot figures: the lower and upper hinges correspond to the 25th and 75th percentiles; the upper whisker extends from the hinge to the largest value no further than 1.5 × inter-quartile range (IQR); the lower whisker extends from the hinge to the smallest value at most 1.5 × IQR of the hinge. The line within the box corresponds to the median.

(C) UMAP of 24 distinct cell types identified in the integrated BAL data.

(D) Dynamic changes of cell abundances for cell types in two bronchoalveolar lavage (BAL) single-cell datasets. Statistical methods are same with (B).

(E) ToppCell allows for gene signatures to be hierarchically organized by lineage, cell type, subtype, and disease condition. The global heatmap shows gene modules with top 50 upregulated genes (student t test) for each cell type in a specific disease condition and compartment. Gene modules from control donors and severe COVID-19 patients were included in the figure. See also Figures S1–S4 and Table S2.

S2). Integration with evoked pathways in the following analysis implicated that neutrophil expansion was likely the consequence of emergency myelopoiesis (Wilk et al., 2021). In addition, a general down-regulation of T cell and NK cell was observed, consistent with lymphopenia reported in clinical practices (Pedersen and Ho, 2020; Terpos et al., 2020) (Figures S1C and S2). However, the trend of T cell subtypes varies across studies and individuals, apart from proliferative T cells which have a dramatic increase in mild and severe patients (Figure S2). Notably, plasmablasts substantially increased in COVID-19 patients, and especially so in severe patients, suggesting upregulated antibody production (De Biasi et al., 2020) (Figures 2B and S1C). Expansion of platelets is another significant change observed in severe patients, possibly leading to immunothrombosis in the lung, which could be closely associated with the severity of the disease (Middleton et al., 2020; Nicolai et al., 2020) (Figures 2B and S1C).

In samples obtained from patients' lungs, we observed the depletion of FABP4^{high} tissue-resident alveolar macrophages (TRAM) and dramatic expansion of FCN1^{high} monocyte-derived alveolar macrophages (MoAM) in severe patients (Figures 2C, 2D, and S3D). Mild patients exhibited a moderate reduction of tissue-resident macrophages, but no evidence of aggregation of monocyte-derived macrophages (Figures 2C, 2D, S3A, and S3D). Dynamic changes of these two subtypes suggest increased tissue chemoattraction (Merad and Martin, 2020) and potential damage of patients' lungs (McGonagle et al., 2020). In addition, neutrophils were only identified in severe patients in the integrated BAL data (Figures 2C and S3A), which might be related with neutrophil extracellular traps (NETs) in the lung (Barnes et al., 2020). However, more samples are required to draw a solid conclusion. We also noted that conventional dendritic cells decreased in the severe patients, which is consistent with the trend of the counterpart in PBMC data. Opposite to the change in PBMC, an expansion of plasmacytoid dendritic cells is observed in both mild and severe patients (Figure 2D). Other cell types, including T cell and NK cell in the BAL, also have converse changes of their counterparts in PBMC, which could be attracted by lung macrophages or epithelial cells after infection or damages (Chua et al., 2020) (Figures 2D and S3D). These changes were consistently observed in lung parenchyma samples from severe COVID-19 patients (Figure S4). With cells well-annotated in the integrated COVID-19 atlas, we drew a global heatmap for cells in both blood and lung using ToppCell gene modules (top 50 DEG in each module) of all identified cell classes. While there was conservation of gene patterns involved in healthy donors and severe COVID-19 patients, there were substantial differences most notably in myeloid cells (Figure 2E). Such hierarchically ordered ToppCell gene modules were broadly used in visualization, large-scale comparisons and fine-resolution investigations in the following analyses.

Myeloid cell atlas: functionally distinct neutrophils at different levels of maturation and derailed macrophages in the lung

Dysregulated myeloid cells have been reported as an important marker of severe COVID-19 patients (Schulte-Schrepping et al., 2020; Silvin et al., 2020). To gain a deeper and comprehensive understanding of these cells, we applied the sub-clustering strategy on the integrated data of key cell types, such as neutrophils and macrophages, and then generated gene modules for comparative functional analysis and interactome inference. We successfully identified 5 neutrophil sub-clusters after the integration of PBMC and BAL data, including 3 FCGR3B+ mature sub-clusters and 2 FCGR3B- immature sub-clusters (Figures 3A and S5B; Table S3). They're mainly from severe patients and their gene modules were generated and subjected to comparative functional enrichment using ToppCell and ToppCluster (Figures 3C, 3D and S5A). We identified proliferative neutrophils (referred to as pro-neutrophils and Neu4) and MMP8^{high} precursor immature
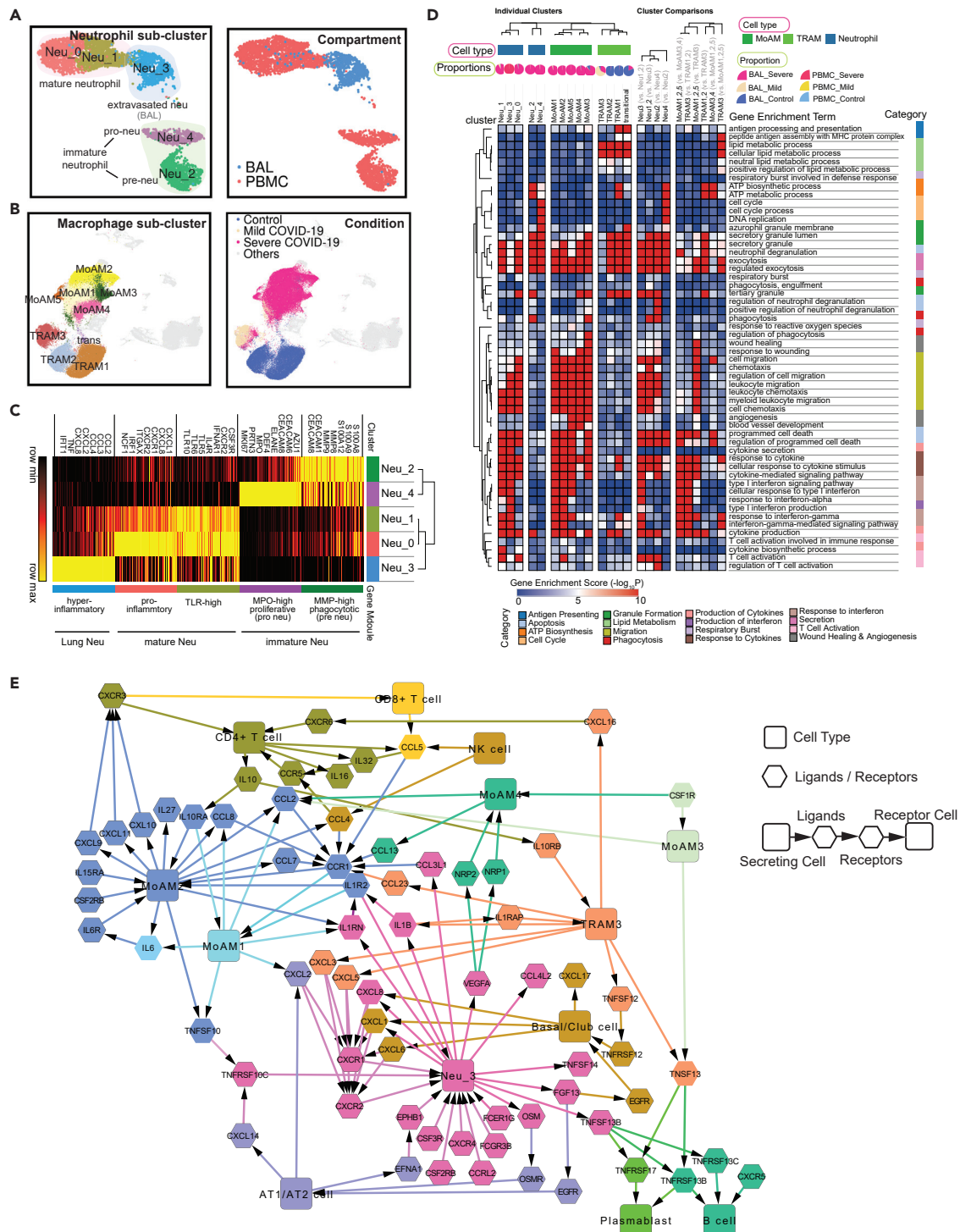
Figure 3. Functional analysis of compartment-specific immature and subtype-differentiated neutrophils and monocytic macrophages in COVID-19 patients

(A) Five sub-clusters and three cell groups were identified after the integration of neutrophils in peripheral blood mononuclear cells (PBMC) and bronchoalveolar lavage (BAL) (Left). The distribution of compartments is shown on the right.

(B) Sub-clusters (Left) and COVID-19 conditions (Right) of monocyte-derived macrophages and tissue-resident macrophages were identified after integration of BAL datasets.

Figure 3. *Continued*

(C) Heatmap of gene modules from ToppCell with top 200 upregulated genes for each neutrophil sub-cluster. Important neutrophil-associated genes and inferred roles of sub-clusters were shown on two sides.

(D) Heatmap of associations between subclusters of neutrophils and macrophages and myeloid-cell-associated pathways (Gene Ontology). Gene modules with 200 upregulated genes for sub-clusters were used for enrichment in ToppCluster. Additionally, enrichment of top 200 differentially expressed genes (DEGs) for comparisons in Figures S5D and S6B were appended on the right. Gene enrichment scores, defined as $-\log_{10}$(adjusted p value), were calculated as the strength of associations. Pie charts showed the proportions of COVID-19 conditions in each cluster.

(E) Gene interaction network in the BAL of severe patients. Highly expressed ligands and receptors of each cell type in the BAL of severe patients were selected based on Figure S8. Among them, genes with unique and distinct expression patterns in each cell type were chosen, for example, CCL17 for cDCs and CXCR1 for neutrophils. Interaction was inferred using both CellChat database and embedded cell interaction database in ToppCell. The molecular interaction between cell types is represented as a flow of arrows in such a way: secreting cells → ligands → receptors → receptor cells, where rectangles represent various cell types and hexagons represent ligands and receptors. The color for a gene is consistent with the cell type with the highest expression level. See also Figures S5–S11 and Table S3.

neutrophils (referred to as pre-neutrophils and Neu2) (Figures 3A and S5B) consistent with prior studies (Schulte-Schrepping et al., 2020). Although immune response genes and pathways were barely activated in the immature neutrophils, they displayed upregulation of granule formation pathways and NETosis-associated proteins, including ELANE, DEFA4, and MPO, especially in Neu4 (Schulte-Schrepping et al., 2020; Wilk et al., 2021) (Figures 3C, S5B, and S5C). Upregulated myeloid leukocyte mediated immunity in Neu2 suggests involvement of this cell type in antiviral function (Figure S5D). Yet, the absence of cytokine and interferon response pathways suggests the lack of mature immune responses (Figure 3D). Notably, compared to mature neutrophils (Neu0 and Neu1) in the blood, the extravasated hyperinflammatory sub-cluster (Neu3) from BAL of severe patients shows extraordinarily high expression of interferon-stimulated genes, as well as prominent upregulation of productions and responses to cytokines and interferons (Figures 3C, 3D, S5B and S5D).

MoAM and TRAM were two main macrophage types in the BAL (Figure 2C); both are known to have distinct roles in immune responses in the lung (Liao et al., 2020). As described above, five sub-clusters among the expanded COVID-19 patient-specific MoAM (Figure 3B; Table S3) were found, where the loss of HLA class II genes and elevation of interferon-stimulated genes (ISGs) were consistently observed (Figures 3F and S6A). Relative to MoAM3,4, MoAM1,2,5 displayed an upregulation of interferon responses and cytokine production (Figures 3D and S6B; Table S3), indicating their pro-inflammatory characteristics. Notably, MoAM5 shows dramatic upregulation of IL-6 secretion and cytokine receptor binding activities (Figures S7A–S7D). However, cells in this sub-cluster were mainly from one severe patient (Figure S3C). We still need more data to fully understand such dramatic upregulation of IL-6 secretion in some severe patients. Similar to MoAM, we also identified two distinct groups of TRAM in BAL (Figures 3B and S6B), including quiescent TRAM (TRAM1 and TRAM2) and activated TRAM (TRAM3). The quiescent group was mainly from healthy donors with enriched pathways of ATP metabolism (Figure 3D), while the activated group from mild and severe patients displays upregulation of ISGs and cytokine signaling pathways (Figure S6B; Table S3). However, the magnitude of activation and inflammatory responses in TRAM3 is smaller than MoAM1,2,5. Not surprisingly, stronger antigen processing and presentation activities were observed in TRAM3 relative to MoAM1,2,5 (Figures 3D and S6B; Table S3). Collectively, we concluded that tissue-resident macrophages were greatly depleted in severe patients as the front-line innate immune responders in the lung. Pro-inflammatory monocyte-derived macrophages infiltrate into the lung, leading to the cytokine storm and damage of the lung. Large amounts of infiltration of MoAM were not observed in mild COVID-19 patients, probably due to the controlled infection, which could explain milder lung damages in those patients.

To develop insights into molecular and cellular interaction networks active in the lung microenvironment of COVID-19 patients (Garvin et al., 2020), we focused on signaling ligands, receptors and pathways using ToppCell and CellChat (Figures 3E, S8A and S8B). Notably, basal cells, MoAMs, neutrophils and T cells all contributed to the cytokine, chemokine, and interleukin signaling networks. Strikingly, severe patient specific MoAM2 shows the broadest upregulation of signaling ligands, including CCL2, CCL3, CCL7, CCL8, CXCL9, CXCL9, CXCL10, CXCL11, IL6, IL15, and IL27, suggesting its role as a signaling network hub that is distinct from the other major signaling ligand-expressing cells of BAL such as epithelial and other myeloid cell types such as TRAM3 and proliferating myeloid cells (Figure S8A). Among the MoAM2 top signaling molecules, attractants CXCL8, CXCL9, and CXCL10 are known to target CXCR3 on T cells, suggesting their role is to stimulate migration of T cells to the epithelial interface and into BAL fluid (Figure 3E) (Liao et al., 2020). In addition, many of MoAM2's ligands have the potential to cause autocrine signaling activation via IL6-IL6R, IL1RN-ILR2, CCL7-CCR1, CCL2-CCR1, and CCL4-CCR1,

Figure 4. COVID-19 driven reprogramming of platelets leads to drastically altered expression of genes associated with platelet adhesion, activation, coagulation and thrombosis

(A and B) Uniform Manifold Approximation and Projections (UMAPs) show distributions of sub-clusters (A) and COVID-19 conditions (B) of platelets after the integration of PBMC datasets.

(C) Severity-associated coagulation genes were selected and shown on the heatmap, with disease and sub-cluster specific gene patterns identified and labeled. Their functional associations with coagulation pathways were retrieved from ToppGene and shown on the right.

Figure 4. *Continued*

(D) Functional and phenotypical associations of coagulation-association genes in each gene pattern from (C). Associations were retrieved from ToppGene enrichment output and eight enrichment categories are represented by rectangles with unique colors. Thicken colored edges represent associations of highlighted functions, such as cell-matrix binding and regulation of coagulation. Hypothesized pathway cascade that potentially drives thromboembolism in COVID-19 patients is drawn in red dotted line. See also Figures S12, S13 and Table S4.

indicating its active roles in self-stimulation and development, which further amplify the attraction and migration of T cells and other immune cells. Notably, CCR1 was also expressed in activated TRAM3, but with a lower level. Although IL6 expression level is relatively low compared to other ligands in BAL data, substantial expression of IL6R was observed in MoAMs. The CCL and CXCL signaling pathways of neutrophils are less strong than MoAMs (Figure S8B), but they displayed high expression levels of CXCR1 and CXCR2, which binds with a large number of the chemokines from MoAM and epithelial cells (Figure 3E). In addition, neutrophils exhibit an extraordinarily high level of IL1B, which could potentially in turn activate macrophages (Figures S8A and S8B). TRAM3 also displayed a unique pattern of signaling molecules, with a substantial level of CCL23 which could potentially attract MoAM by the interaction with CCR1. Secretion of CXCL3 and CXCL5 in TRAM3 toward CXCR2 could be a potential chemoattraction pathway for neutrophils. In turn, neutrophils could activate TRAM3 by secreting IL1B, which binds with IL1RAP. In addition, CD4+ T cells could also activate TRAM3 by IL10-IL10RB interaction (Figures 3E, S8A, and S8B).

In addition to substantial alterations of neutrophils and macrophages, the upregulation of ISGs was observed in classical monocytes of both mild and severe patients (cMono4, cMono3), whereas the reduction of the MHC class II cell surface receptor HLA-DR genes was only observed in severe patients (cMono3) (Figure S9). In cDCs, polarization of interleukin secretion was observed in mild-patient and severe-patient specific clusters (Figure S10F). Collectively, dynamic changes of marker genes, transcriptional profiles, signaling molecules, and biological activities reveal the heterogeneity of myeloid cell sub-clusters across disease severity (Figure S11C). Pro-inflammatory gene expression was found in all major myeloid cell types, including cMono4, DC1, and DC9 in PBMC and Neu3, DC10, MoAM1, MoAM2, and MoAM5 in BAL of COVID-19 patients. The reduction of MHC class II (HLA-II) genes is a common feature of classical monocytes and macrophages in COVID-19 patients and implies impaired capacity to activate T cell adaptive immunity.

COVID-19 coagulation and immunothrombosis map

Individuals severely affected during acute phase COVID-19 infection, and in particular those with significantly elevated risk of death, frequently demonstrate striking dysregulation of coagulation and thrombosis characterized by hypercoagulability and microvascular thrombosis (endothelial aggregations of platelets and fibrin) and highly elevated D-dimer levels. Yet, COVID-19 does not lead to wide scale consumption of fibrinogen and clotting factors (Iba et al., 2020a; Levi et al., 2020; Middleton et al., 2020; Nicolai et al., 2020; Rapkiewicz et al., 2020). At present, we lack a molecular or cellular explanation of the underlying basis of this pathobiology (Aid et al., 2020; Middleton et al., 2020). To evaluate candidate effectors of this pathobiology, we used a list of genes associated with abnormal thrombosis from mouse and human gene mutation phenotypes and identified parenchymal lung sample endothelial cells and platelets in PBMC as cell types highly enriched with respect to genes responsible for the regulation of hemostasis (Figure S12). Because platelet counts were greatly elevated in severe versus mild individuals, we further examined platelet gene expression signatures and cell type differentiation and identified six distinct platelet sub-clusters shared across all datasets after data integration (Figures 4A and 4B). Severe-patient-specific PLT0 is an interesting sub-cluster with elevated integrin genes, including ITGA2B, ITGB1, ITGB3, and ITGB5, as well as thrombosis-related genes, such as SELP, HPSE, ANO6, and PF4V1. Antibodies against the latter are associated with thrombosis including adverse reactions to recent COVID-19 vaccine ChAdOx1 nCoV-19 (Schultz et al., 2021). In addition, upregulated pathways of hemostasis, wound healing and blood coagulation were also observed in PLT0 (Figure S13A; Table S4). Importantly, PLT2 is an inflammatory sub-cluster with an upregulation of ISGs and interferon signaling pathways, whereas PLT4 is highlighted by upregulated post-transcriptional RNA splicing activities (Figure S13A and S13C).

Severity-associated gene patterns were also identified by selecting coagulation-associated genes modules (Figure 4C; Table S4), indicating distinct coagulation activities across platelets. Apart from pan-platelet genes, we found dramatic upregulation of genes involved in platelet activation, fibrinogen binding and blood coagulation in platelets of severe COVID-19, including procoagulant heparanase (HPSE) (Osterholm et al., 2013), Anoctamin-6 (ANO6) (Swieringa et al., 2018), and selectin P (SELP) (Sparkenbaugh and
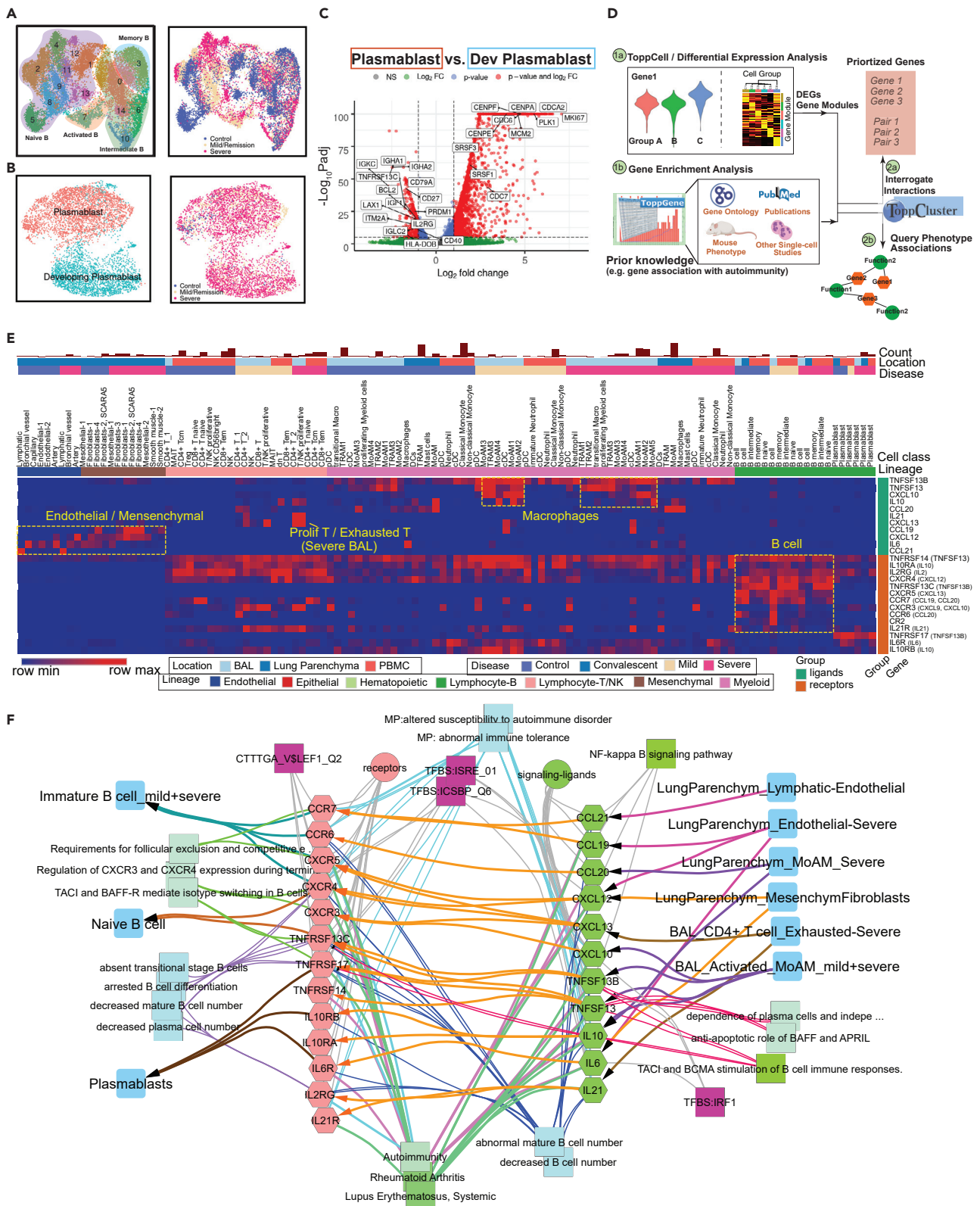
Figure 5. Implicating a multi-lineage cell network capable of driving extrafollicular B cell maturation and the emergence of humoral autoimmunity in COVID-19 patients

(A) Uniform Manifold Approximation and Projections (UMAPs) of sub-clusters (Left) and COVID-19 conditions (Right) of B cells after integration of peripheral blood mononuclear cells (PBMC) and bronchoalveolar lavage (BAL) datasets.

(B) UMAPs of subtypes (Left) and COVID-19 conditions (Right) of plasmablasts after integration of PBMC and BAL datasets.

(C) Volcano plot depicts differentially expressed genes between plasmablasts and developing plasmablasts. Student t-tests were applied and p values were adjusted by the Benjamini-Hochberg procedure. Thresholds of adjusted p values and log fold changes in the volcano plot are 1.0×10^{-6} and 1.0, respectively.

(D) Workflow of discovering and prioritizing candidate genes related to a disease-specific phenotype with limited understanding.

(E) The heatmap shows the normalized expression levels of candidate ligands and receptors for COVID-19 autoimmunity in multiple compartments in healthy donors and COVID-19 patients. Binding ligands of receptor genes were shown in parentheses on the right. Hot spots of expression are highlighted.

(F) Network analysis of autoimmunity-associated gene expression by COVID-19 cell types. Prior knowledge associated gene associations include GWAS, OMIM, mouse knockout phenotype, and additional recent manuscripts were selected from ToppGene enrichment results of differentially expressed ligands and receptors and shown on the network. Orange arrows present the interaction directions from ligands (green) to receptors (pink) on B cells. Annotations for these genes, including single-cell co-expression (blue), mouse phenotype (light blue), transcription factor binding site (purple) and signaling pathways (green) are shown. See also Figures S14, S15 and Table S5.

Pawlinski, 2013) (Figures 4C and 4D). Heparanase is an endoglycosidase that cleaves heparan sulfate constituents, a major component of anti-coagulation glycocalyx on the surface of vascular endothelium (Edovitsky et al., 2006; Iba et al., 2020b). Upregulated heparanase was related to upregulation of cell-matrix adhesion and coagulation (Figure 4D). Thrombotic vascular damages could be caused by the degradation function of heparinase enriched in platelets of severe patients. Elevation of ANO6 is known to trigger phospholipid scrambling in platelets, resulting in phosphatidylserine exposure which is essential for activation of the clotting system (Heemskerk et al., 2013). In addition, other upregulated genes involved in coagulation-associated activities were also observed, including wound healing, fibrinolysis, platelet aggregation and activation (Figure 4D), which likely collectively contribute to the clotting issue of severe COVID-19 patients.

Emergence of developing plasmablasts and B cell association with autoimmunity

Autoimmune disorders in COVID-19 patients such as immune thrombocytopenic purpura (ITP) is now recognized as a known disease complication (Ehrenfeld et al., 2020; Fujii et al., 2020; Gruber et al., 2020; Rodríguez et al., 2020; Zhou et al., 2020). However, little is known about the molecular and cellular mechanism behind it. To examine this further, we integrated B cells and plasmablasts from both PBMC and BAL and conducted systematic analysis (Figures 5A, 5B, S14A, and S14B). Several COVID-19 specific sub-clusters were identified in B cells, such as ISG^high activated B cells (cluster 7) (Figure 5A). Importantly, activated B cells showed dramatic upregulation of interferon signaling pathways and cytokine productions (Figure S15A), indicating its anti-virus characteristics. Notably, plasmablasts were mainly observed in severe COVID-19 patients, where a group of proliferative cells was identified and labeled as developing plasmablasts (Figure 5B). In contrast, non-dividing plasmablasts displayed upregulation of immunoglobulin genes (IGHA1, IGHA2, and IGKC), B cell markers (CD79A) (Mason et al., 1995), interleukin receptors (IL2RG) and type II HLA complex (HLA-DOB) (Figure 5C; Table S5). In addition, non-dividing plasmablasts showed unique isotypes of immunoglobulin (Ig) in sub-regions of UMAP, whereas developing plasmablasts displayed obscure Ig types (Figure S14E and S14F). Antibody production activities were upregulated in non-dividing plasmablasts based on gene enrichment analysis (Figure S15A; Table S5). Collectively, we inferred that non-dividing plasmablasts had definite immunoglobulin isotypes and were actively involved in immune responses toward COVID infection, whereas developing plasmablasts were less mature but highly proliferative to replenish the repertoire of plasma cells.

Because there are few clues of gene associations of autoimmunity in COVID-19, we brought up a hypothesis-driven, prior knowledge-based approach to discover and prioritize genes for the specific phenotype (Figure 5D). First, gene modules of B cells and other cells in severe patients were collected and subjected to ToppGene (Chen et al., 2009) for enrichment analysis. Then we queried autoimmunity-associated terms in the enriched output and identified associated genes. After that, we retrieved interaction pairs using the ToppCluster and CellChat database (Jin et al., 2021). In the end, we identified genes that are not only involved in autoimmunity, but have a mediator role in the immune signaling network. Using this approach, we observed several candidate pairs of genes, including TNFSF13B-TNSRSF13, IL10-IL10RA, IL21-IL21RA, IL6-IL6R, CXCL13-CXCR5, CXCL12-CXCR4, CCL21-CCR7, CCL19-CCR7, and CCL20-CCR6 in severe patients, which were enriched for autoimmune diseases, such as autoimmune thyroid diseases, lupus nephritis, and autoimmune encephalomyelitis (Aust et al., 2004; Hirota et al., 2007; Kuwabara et al., 2009; Lee et al., 2010; Steinmetz et al., 2008). Candidate cytokine and chemokine ligand genes were
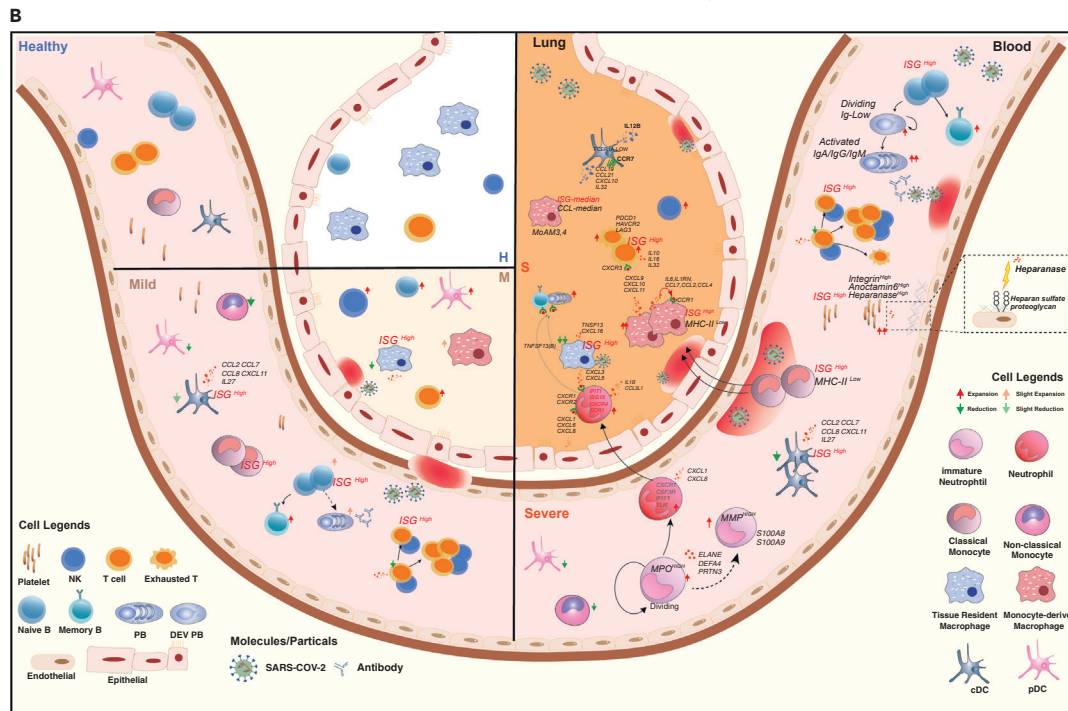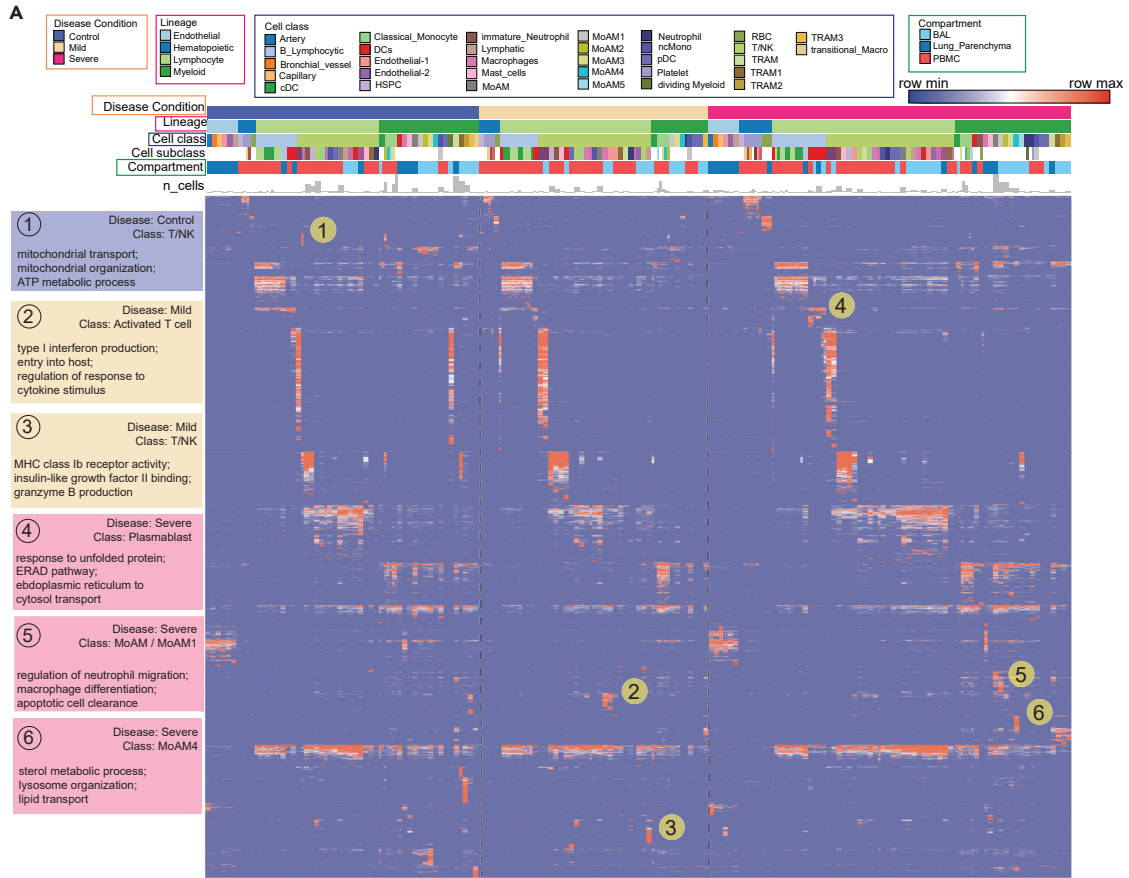
Figure 6. Comparative analysis of cell type specific gene signatures associated with lineage, class, subclass, compartment, and disease state in the COVID-19 atlas

(A) Enrichment scores of gene modules for all cell types across different compartments and COVID-19 conditions were generated by ToppCluster and shown on the heatmap. ToppCluster enriched functions from Gene Ontology, Human Phenotype, Mouse Phenotype, Pathway and Interaction databases were used to generate a feature matrix (cell types by features) and were hierarchically clustered. Hot spots of the disease-specific enrichments were highlighted and details were shown on the left. More details can be found in STAR Methods.

(B) Summarizing predicted functions and interplay of immune cells in COVID-19 blood and lung. Aforementioned key observations in this study were shown in peripheral blood mononuclear cells (PBMC) and bronchoalveolar lavage (BAL) in healthy donors, mild and severe COVID-19 patients, including changes of cell abundance, specific marker genes, upregulated secretion, cell development and cell-cell interactions. See also Table S6.

expressed in various cell types in PBMC and BAL, including IL21 and CXCL13 from exhausted T cells of BAL, CXCL12 from mesenchymal cells, IL6, and CCL21 from endothelial cells, CCL19 from cDC and CCL20, TNFSF13B, and TNFSF13 from lung macrophages (Figure 5E; Table S5). These interaction pairs have been linked with auto-immunity (Klimatcheva et al., 2015; Wong et al., 2010). In addition, we analyzed single-cell studies (Arazi et al., 2019; Zhang et al., 2019) of rheumatoid arthritis and lupus nephritis patients and found that high expression levels of the candidate receptors in B cells and ligands in other cells were also observed, such as CXCL13 in helper T cells and CXCR5 in B cells in both studies (Figures S15C and S15D). However, more evidence is still required to infer the association between these interactions and autoimmunity in COVID-19 patients. Supported by the evidence above, we drew a network for potential mediator interactions of B cells and their associations with autoimmune disorders, where linkages with diseases, such as rheumatoid arthritis, systemic lupus erythematosus, were highlighted, as well as linkages with mouse phenotypes, such as abnormal immune tolerance and increased susceptibility to autoimmune disorder (Figure 5F). As a caveat, although using prior knowledge to prioritize gene and cell-associated functions and interactions may introduce biases, such approaches also have the potential to highlight key similarities and differences between different disease causes and clinical responses and improve our understanding of the molecular and cellular mechanisms at work.

Functional map and immune cell interplay landscape in COVID-19

As above, where highly significant enrichments of unique functions and pathways could be identified in the subtypes of multiple cell classes, such as neutrophils, platelets and B cells, we sought to get a more holistic understanding of COVID-19 specific cell class and subclass-level signatures, including T cell subtypes (Figures S16 and S17), we built an integrative functional map of all cell types in three compartments across multiple disease conditions using a highly integrated gene module set (Figure 6A; Table S6). All enriched functional associations in ToppCluster for gene modules of cell types and sub-clusters were depicted. They were grouped by disease conditions and compartments to show heterogeneity of cellular functions in different circumstances.

In the heatmap (Figure 6A), most enrichments were consistently observed across cells of healthy donors and COVID-19 patients. However, some unique patterns were also identified. For example, T cells, and NK cells in healthy donors show enrichments of mitochondrial transport and ATP metabolic process, whereas activated T cells in mild patients show upregulation of type I interferon production and cytokine signaling. Enrichments of macrophage differentiation and neutrophil migration regulation were uniquely found in MoAM1 in severe patients (Figure 6A). The function map provides a high-level approach to investigate functional variations of cells across disease conditions and compartments. The predicted interplay of immune cells across multiple compartments and disease conditions is displayed in Figure 6B. Cell proportion changes, sub-cluster specific signatures, and cell-cell interaction are also depicted.

Similarity and heterogeneity between COVID-19 and other immune-mediated diseases

To further analyze COVID-19 specific immune signatures, we compared immune cells from COVID-19 patients with cells in other immune-mediated diseases, including severe influenza (Lee et al., 2020), sepsis (Reyes et al., 2020), and multiple sclerosis (Schafflick et al., 2020). 404,125 cells were included after the integration of PBMC single-cell datasets (Figures 7A and S18; Table S7). Dynamic changes of cell abundance were compared in diseases versus healthy donors. Similar to COVID-19 patients, severe influenza patients also exhibited the reduction of non-classical monocytes, pDC, cDC, and CD4+ TCM, but the effect of the former two types was smaller in magnitude (Figure 7B). However, the reduction of non-classical monocytes is more significant in severe COVID-19 patients than severe influenza or mild COVID-19 patients (Figure 7B). Notably, NK cell reduction is associated with COVID-19 severity, whereas T cell depletion is a more dramatic perturbation in severe influenza. Within these comparisons, the expansion of plasmablasts is consistently observed, whereas the accumulation of platelets is unique to SARS-CoV-2 and in particular, to severe COVID-19 clinical status (Figure 7B).
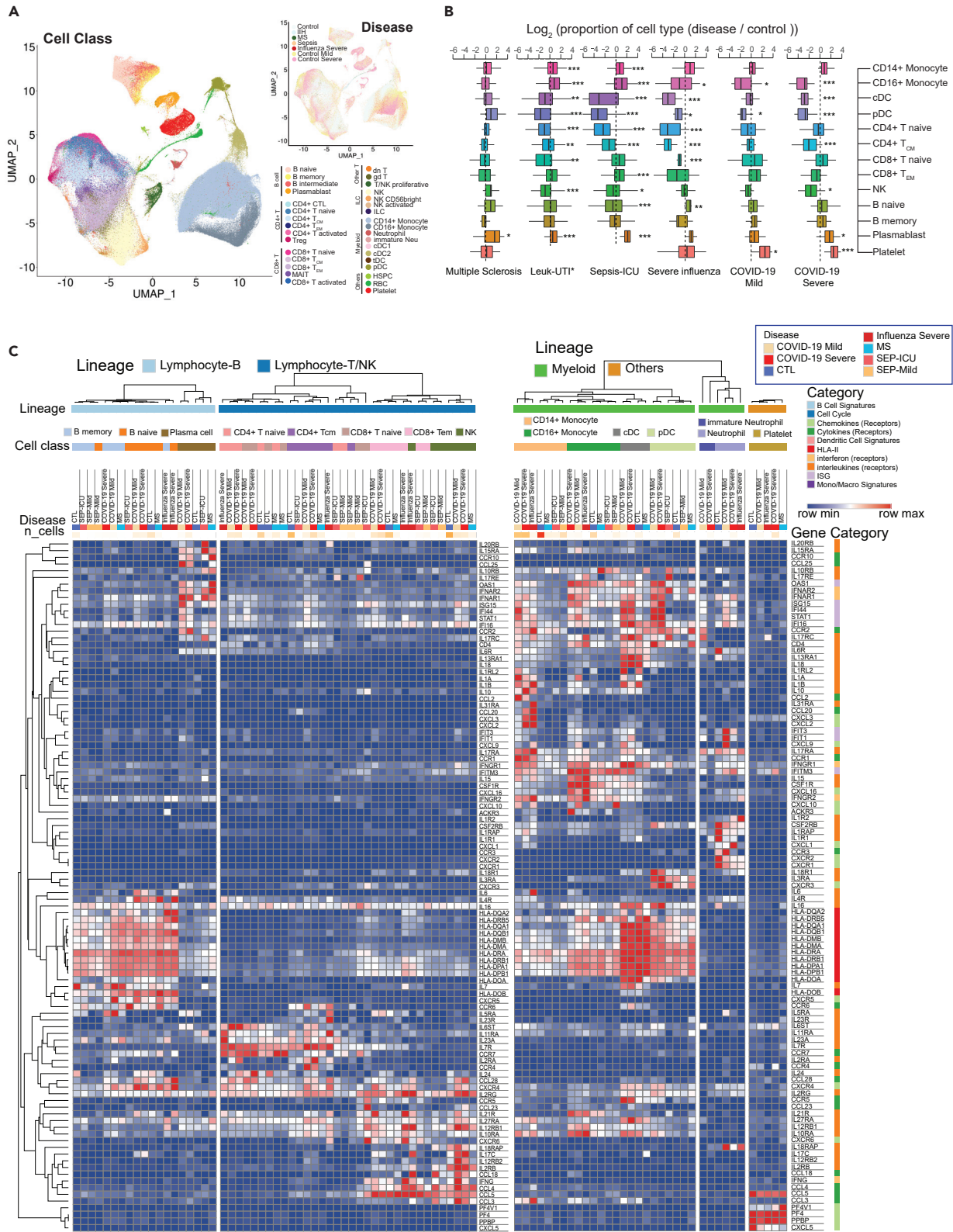
Figure 7. Comparative analysis of differentially-expressed immunoregulatory genes between COVID-19 and other immune-mediated diseases

(A) Uniform Manifold Approximation and Projection (UMAP) shows the distributions of cell types (Left) and diseases (Top right) after the integration of datasets in multiple studies. MS: multiple sclerosis; IIH: idiopathic intracranial hypertension. IIH patients were recruited as controls in the multiple sclerosis study.

(B) Dynamic changes of immune cell types in different immune-mediated diseases compared to healthy controls. Log2(ratio) was calculated to show the levels of changes. *, p < 0.05, **, p < 0.01, ***, p < 0.001. Statistical models can be found in the STAR Methods. Leuk-UTI: sepsis patients that enrolled into UTI with leukocytosis (blood WBC ≥12,000 per mm3) but no organ dysfunction. Box features are same with Figure 2B.

(C) Normalized expression values of key genes involved in immune signaling and responses are shown for cell types across multiple diseases. Lowly expressed genes (maximal average expression level across all cell types in the heatmap is less than 0.5 after Log_2CPM normalization) were removed. See also Figure S18 and Table S7.

In addition to dynamic changes of cell ratios, we also investigated the regulation of immune mediator genes across various diseases (Figure 7C; Table S7). IL-6 is an important factor of cytokine storms in COVID-19 (Zhao, 2020). As shown in the heatmap, naive B cells are the main sources of IL-6 in COVID-19 patients while CD14+ monocytes show the highest expression levels in severe influenza patients (Figure 7C). Specific ligands, including CXCL2, CXCL3, and CCL20 were upregulated in both severe COVID-19 patients and severe influenza patients. CCR4 and IL2RA are uniquely high in CD4+ T cells of COVID-19 patients. Interestingly, most PBMC myeloid cell types displayed upregulated levels of interferon-stimulated genes in both COVID-19 and influenza, especially in COVID-19, where highest levels of ISGs in CD14+ Monocytes, cDC and pDC were observed.

DISCUSSION

In this work, we have constructed an innovative immune signature atlas of the blood and lung of COVID-19 patients using the integrated single cell RNA-sequencing data and Topp-toolkit. By virtue of systemic analysis of large sample size from multiple sampling sites, consistent immunopathology-associated changes of cell abundance and transcriptional profiles were observed in the circulating and lung immune repertoire of COVID-19 patients. The established single cell atlas and the provided public portal (https://toppcell. cchmc.org/) enables the query of candidate molecules and pathways in each of these processes.

Leveraging this approach, we identified three major candidate mechanisms capable of driving COVID-19 severity: (1) a cascade-like network of proinflammatory autocrine and paracrine ligand receptor interactions among subtypes of differentiating mononuclear, lymphoid, as well as other cell types; (2) the production of emergency platelets whose gene expression signatures implicate significantly elevated potential for adhesion, thrombosis, attenuated fibrinolysis, and potential to enhance the release of heparin-bound cytokines as well as further influence the activation of neutrophils causing further inflammatory cell recruitment and neutrophil netosis; and (3) the extrafollicular activation of naive and immature B cells via a multilineage network that includes monocytic subtypes and exhausted T cells of cytokines and interleukins with the potential to generate local antigen specific response to virus infected targets and collateral autoimmunity. More details will be discussed below.

We identified dramatically expanded macrophages which were marked by the loss of HLA class II genes and upregulation of interferon-stimulated genes. It implicates a key role for these activated macrophages involved in signaling networks and less so in activation of adaptive T cell immunity. Among them, MoAM2 displayed hyperinflammatory responses and extraordinary high levels of signaling molecules, which are involved in both autocrine (e.g., IL-6, CCL2, CCL4, and CCL8) and paracrine (e.g., CXCL2, CXCL9, CXCL10, and CXCL11) signaling pathways. The former pathway contributed to the self-stimulation and development, which amplified the paracrine pathway for T cell and neutrophil chemoattraction. The latter two cell types in turn activated MoAMs with cytokines genes (CCL5, IL10 of T cells and IL1B of neutrophils, respectively). Based on the intercellular and multifactor complexity of the signaling cascade we have outlined, to effectively control a malignant inflammatory cascade, it may be essential to consider simultaneously targeting multiple nodes of this network of cytokines and interleukins. In addition, HLA-DR^{low} monocytes, likely reflecting dysfunctional cells, were observed in severe infection. This, along with evidence of emergency myelopoiesis with immature circulating neutrophils into the circulation was detected in severe COVID-19. These neutrophils had transcriptional programs suggestive of dysfunction and immunosuppression not seen in patients with mild COVID-19. As such, we have presented evidence for the contribution of defective monocyte activation and dysregulated myelopoiesis to severe COVID.

Platelet expansion is uniquely observed in COVID-19 versus other immune-mediated diseases. Strikingly, these activated platelets were highlighted with abnormal thrombosis and upregulated heparanase, a pro-coagulant endoglycosidase that cleaves anti-coagulation heparan sulfate constituents on endothelial cells and potentially causes thrombotic vascular damages. In addition, heparanase-cleaved heparan sulfate (HS) fragments were capable of stimulating the release of pro-inflammatory cytokines, such as IL1B, IL6, IL8, IL10, and TNF through the TLR-4 pathway in PBMC (Goodall et al., 2014), further contributing to the hyper-inflammatory environment in COVID-19 patients. Because heparanase is recognized as a hallmark in tumor progression and metastasis (Jayatilleke and Hulett, 2020), we hypothesize COVID-19 infection could be associated with higher occurrence of lung tumor metastasis. However, more data is required to support it. Pro-neutrophil secreted proteins (e.g., ELANE, DEF4) of neutrophil extracellular trap (NET), which have been reported to be associated with higher risk of morbid thrombotic events (Zuo et al., 2021). Approaches for combating NETs could be a potential anticoagulation treatment (Thålin et al., 2019).

We propose a signaling network which potentially shapes the differentiation of B cells toward the formation of autoantibodies. Proliferation and activation of inflammatory myeloid cells and the formation of exhausted CD4+ T helper around an area of direct or indirect viral tissue injury leads to the production of a set of interleukins and cytokines known to have both direct cell activating and maturing effects on naive and immature B cells. Previous report had revealed the exaggerated extrafollicular B cell response, which is part of a mechanism that stimulates somatic mutation and maturation of B cells to produce plasma cells with specificity for antigens present in the vicinity of tissue damage sites (Farris and Guthridge, 2020). In the absence of macrophages or dendritic cells to restrict self vs non-self, the presence of IL-10, IL-21, CXCL13 CXC10, IL-6, and others acting on receptors present in naive and immature B cells leads to the selection and maturation of self-reactive maturation of B cells clones with formation of autoantibodies. Many of these COVID-19-activated genes (e.g., CXCL13, CCL19, CCL20, and TNFRSF13) are known to be genetically associated with rheumatoid arthritis, lupus, and risk of developing auto-immune disease in humans and mouse models. The development of different patterns of autoimmunity may be a hallmark of "Long Haul" Covid disease and could explain why some individuals develop different autoanti-bodies and suffer different forms of clinical consequences depending on which antigens drive the B-cell matu-ration. Thus, an additional prediction that could be made based on these findings and our network model is that among individuals treated with corticosteroids at the time these auto-immunogenic processes are activated, there should be a protective effect and lower likelihood of developing post acute sequela of Covid.

Consistent and varied compositional changes and gene patterns of immune cells were identified in COVID-19, influenza and sepsis. Expansion of plasmablasts, as well as the reduction of non-classical monocytes, are more significant changes in severe COVID-19 patients, whereas the depletion of T cells is more dramatic in severe influenza patients. The accumulation is a unique immune hallmark of COVID-19 within the selected diseases, which contributes to the coagulation abnormalities and thrombosis, a key cause of fatality in COVID-19 patients. Different signaling gene patterns were identified across immune-mediated diseases, with CCR4 only highly expressed in CD4+ T cells of COVID-19 patients, which might be related with extravasation of these cells (Spoerl et al., 2021). Upregulated interferon-stimulated genes of myeloid cells in PBMC revealed the inflammatory environment of COVID-19.

Collectively, using the COVID-19 single cell atlas data exploration environment, we have illustrated is that researchers are now enabled to systematically explore, learn, and formulate new hypotheses within and between compartments, cell types, and biological processes, and provided access to these reprocessed datasets through a suite of explorative and evaluative tools. Moreover, we have shown different hypothe-ses can be developed and explored using the approaches that we have outlined and the database that we have provided. Certainly additional critical information will also be obtained using approaches that include *in situ* spatial, temporal data as well as those of viral products and viral and inflammatory-process affected complexes. Next steps for improving its ability to be mined more deeply will be based on additional sta-tistical methods that extend the current ToppCell/ToppGene Suite based on fuzzy measure similarity, Page-Rank, and cell-cell signaling approaches.

Limitations of study

There are several limitations in our study. Different studies used various standards of COVID-19 severity definition. To generalize conclusions, we simplified disease conditions into several universal groups. Prospectively, a standardized definition of disease stages will assist in the accuracy of future studies. In ad-ditionn, the timing of sample collection was not considered as a variable in this study, rather disease stages

were used to consolidate data across samples. We lack follow-up data of patients with sequela, which will be helpful for understanding the long-haul effects of the disease.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- METHOD DETAILS
  - Single-cell RNA-seq data source
  - Data preprocessing and normalization
  - Integration of PBMC datasets and BAL datasets using reciprocal PCA in seurat
  - Cell annotations using canonical markers after unsupervised clustering
  - Cell annotations using Azimuth
  - Sub-clustering for specific cell types
  - Generation of ToppCell gene modules
  - Gene enrichment analysis using ToppGene
  - Generation of functional association heatmap using ToppCluster
  - Cell interaction inference in immunothrombosis activities and cytokine signaling pathways
  - Generation of volcano plots
  - Construction of COVID-19 functional enrichment map
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2021.103115.

AUTHOR CONTRIBUTIONS

Conceptualization, K.J. and B.A.; Methodology, K.J., B.A., E.B., A.M. and J.Y.W; Investigation, K.J. and B.A.; Writing – Original Draft, K.J.; Writing – Review & Editing, K.J., B.A., M.E.R., A.M., D.A.P.K, S.S.G. and S.B.; Funding Acquisition, B.A.; Resources, K.J.; Data Curation, K.J. and E.B.; Visualization, E.B., B.A..; Supervision, B.A.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

Aid, M., Busman-Sahay, K., Vidal, S.J., Maliga, Z., Bondoc, S., Starke, C., Terry, M., Jacobson, C.A., Wrijil, L., Ducat, S., et al. (2020). Vascular disease and thrombosis in SARS-CoV-2-infected rhesus macaques. Cell 183, 1354–1366.e13.

Arazi, A., Rao, D.A., Berthier, C.C., Davidson, A., Liu, Y., Hoover, P.J., Chicoine, A., Eisenhaure, T.M., Jonsson, A.H., Li, S., et al. (2019). The immune cell landscape in kidneys of patients with lupus nephritis. Nat. Immunol. 20, 902–914.

Arunachalam, P.S., Wimmers, F., Mok, C.K.P., Perera, R.A.P.M., Scott, M., Hagan, T., Sigal, N., Feng, Y., Bristow, L., Tak-Yin Tsang, O., et al. (2020). Systems biological assessment of immunity to mild versus severe COVID-19 infection in humans. Science 369, 1210–1220.

Aust, G., Sittig, D., Becherer, L., Anderegg, U., Schütz, A., Lamesch, P., and Schmücking, E. (2004). The role of CXCR5 and its ligand CXCL13 in the compartmentalization of lymphocytes in thyroids affected by autoimmune thyroid diseases. Eur. J. Endocrinol. 150, 225–234.

Barnes, B.J., Adrover, J.M., Baxter-Stoltzfus, A., Borczuk, A., Cools-Lartigue, J., Crawford, J.M., Daßler-Plenker, J., Guerci, P., Huynh, C., Knight, J.S., et al. (2020). Targeting potential drivers of COVID-19: neutrophil extracellular traps. J. Exp. Med. 217, e20200652.

Bharat, A., Querrey, M., Markov, N.S., Kim, S., Kurihara, C., Garza-Castillon, R., Manerikar, A., Shilatifard, A., Tomic, R., Politanska, Y., et al. (2020). Lung transplantation for patients with severe COVID-19. Sci. Transl. Med. 12, eabe4282.

Blanco-Melo, D., Nilsson-Payant, B.E., Liu, W.-C., Uhl, S., Hoagland, D., Møller, R., Jordan, T.X., Oishi, K., Panis, M., Sachs, D., et al. (2020). Imbalanced host response to SARS-CoV-2 drives development of COVID-19. Cell 181, 1036–1045.e9.

Blighe, K., Rana, S., and Lewis, M. (2021). EnhancedVolcano: publication-ready volcano plots with enhanced colouring and labeling. R package version 1.10.0. https://github.com/kevinblighe/EnhancedVolcano. https://doi.org/10.18129/B9.bioc.EnhancedVolcano.

Cao, X. (2020). COVID-19: immunopathology and its implications for therapy. Nat. Rev. Immunol. 20, 269–270.

Chen, J., Bardes, E.E., Aronow, B.J., and Jegga, A.G. (2009). ToppGene suite for gene list enrichment analysis and candidate gene prioritization. Nucleic Acids Res. 37, W305–W311.

Chua, R.L., Lukassen, S., Trump, S., Hennig, B.P., Wendisch, D., Pott, F., Debnath, O., Thürmann, L., Kurth, F., Völker, M.T., et al. (2020). COVID-19 severity correlates with airway epithelium–immune cell interactions identified by single-cell analysis. Nat. Biotechnol. 38, 970–979.

De Biasi, S., Lo Tartaro, D., Meschiari, M., Gibellini, L., Bellinazzi, C., Borella, R., Fidanza, L., Mattioli, M., Paolini, A., Gozzi, L., et al. (2020). Expansion of plasmablasts and loss of memory B cells in peripheral blood from COVID-19 patients with pneumonia. Eur. J. Immunol. 50, 1283–1294.

Delorey, T.M., Ziegler, C.G.K., Heimberg, G., Normand, R., Yang, Y., Segerstolpe, Å., Abbondanza, D., Fleming, S.J., Subramanian, A., Montoro, D.T., et al. (2021). COVID-19 tissue atlases reveal SARS-CoV-2 pathology and cellular targets. Nature 595, 107–113.

DePasquale, E.A.K., Schnell, D.J., Van Camp, P.-J., Valiente-Alandí, Í., Blaxall, B.C., Grimes, H.L., Singh, H., and Salomonis, N. (2019). DoubletDecon: deconvoluting doublets from single-cell RNA-sequencing data. Cell Rep 29, 1718–1727.e8.

Ding, J., Adiconis, X., Simmons, S.K., Kowalczyk, M.S., Hession, C.C., Marjanovic, N.D., Hughes, T.K., Wadsworth, M.H., Burks, T., Nguyen, L.T., et al. (2020). Systematic comparison of single-cell and single-nucleus RNA-sequencing methods. Nat. Biotechnol. 38, 737–746.

Edovitsky, E., Lerner, I., Zcharia, E., Peretz, T., Vlodavsky, I., and Elkin, M. (2006). Role of endothelial heparanase in delayed-type hypersensitivity. Blood 107, 3609–3616.

Ehrenfeld, M., Tincani, A., Andreoli, L., Cattalini, M., Greenbaum, A., Kanduc, D., Alijotas-Reig, J., Zinserling, V., Semenova, N., Amital, H., et al. (2020). Covid-19 and autoimmunity. Autoimmun. Rev. 19, 102597.

Farris, A.D., and Guthridge, J.M. (2020). Overlapping B cell pathways in severe COVID-19 and lupus. Nat. Immunol. 21, 1478–1480.

Fujii, H., Tsuji, T., Yuba, T., Tanaka, S., Suga, Y., Matsuyama, A., Omura, A., Shiotsu, S., Takumi, C., Ono, S., et al. (2020). High levels of anti-SSA/Ro antibodies in COVID-19 patients with severe respiratory failure: a case-based review. Clin. Rheumatol. 39, 3171–3175.

Garvin, M.R., Alvarez, C., Miller, J.I., Prates, E.T., Walker, A.M., Amos, B.K., Mast, A.E., Justice, A., Aronow, B., and Jacobson, D. (2020). A mechanistic model and therapeutic interventions for COVID-19 involving a RAS-mediated bradykinin storm. Elife 9, e59177.

Goodall, K.J., Poon, I.K.H., Phipps, S., and Hulett, M.D. (2014). Soluble heparan sulfate fragments generated by heparanase trigger the release of pro-inflammatory cytokines through TLR-4. PLoS One 9, e109596.

Grant, R.A., Morales-Nebreda, L., Markov, N.S., Swaminathan, S., Querrey, M., Guzman, E.R., Abbott, D.A., Donnelly, H.K., Donayre, A., Goldberg, I.A., et al. (2021). Circuits between infected macrophages and T cells in SARS-CoV-2 pneumonia. Nature 590, 635–641.

Gruber, C.N., Patel, R.S., Trachtman, R., Lepow, L., Amanat, F., Krammer, F., Wilson, K.M., Onel, K., Geanon, D., Tuballes, K., et al. (2020). Mapping systemic inflammation and antibody responses in multisystem inflammatory syndrome in children (MIS-C). Cell 183, 982–995.e14.

Guo, C., Li, B., Ma, H., Wang, X., Cai, P., Yu, Q., Zhu, L., Jin, L., Jiang, C., Fang, J., et al. (2020). Single-cell analysis of two severe COVID-19 patients reveals a monocyte-associated and tocilizumab-responding cytokine storm. Nat. Commun. 11, 3924.

Hadjadj, J., Yatim, N., Barnabei, L., Corneau, A., Boussier, J., Smith, N., Péré, H., Charbit, B., Bondet, V., Chenevier-Gobeaux, C., et al. (2020). Impaired type I interferon activity and inflammatory responses in severe COVID-19 patients. Science 369, 718–724.

Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W.M., 3rd, Zheng, S., Butler, A., Lee, M.J., Wilk, A.J., Darby, C., Zager, M., et al. (2021). Integrated analysis of multimodal single-cell data. Cell 184, 3573–3587.e29.

He, R., Lu, Z., Zhang, L., Fan, T., Xiong, R., Shen, X., Feng, H., Meng, H., Lin, W., Jiang, W., et al. (2020). The clinical course and its correlated immune status in COVID-19 pneumonia. J. Clin. Virol. 127, 104361.

Heemskerk, J.W.M., Mattheij, N.J.A., and E., J.M. (2013). Platelet-based coagulation: different populations, different functions. J. Thromb. Haemost. 11, 2–16.

Heming, M., Li, X., Räuber, S., Mausberg, A.K., Börsch, A.-L., Hartlehnert, M., Singhal, A., Lu, I.-N., Fleischer, M., Szepanowski, F., et al. (2021). Neurological manifestations of COVID-19 feature T cell exhaustion and dedifferentiated monocytes in cerebrospinal fluid. Immunity 54, 164–175.e6.

Hirota, K., Yoshitomi, H., Hashimoto, M., Maeda, S., Teradaira, S., Sugimoto, N., Yamaguchi, T., Nomura, T., Ito, H., Nakamura, T., et al. (2007). Preferential recruitment of CCR6-expressing Th17 cells to inflamed joints via CCL20 in

rheumatoid arthritis and its animal model. J. Exp. Med. *204*, 2803–2812.

Iba, T., Levy, J.H., Connors, J.M., Warkentin, T.E., Thachil, J., and Levi, M. (2020a). The unique characteristics of COVID-19 coagulopathy. Crit. Care *24*, 360.

Iba, T., Connors, J.M., and Levy, J.H. (2020b). The coagulopathy, endotheliopathy, and vasculitis of COVID-19. Inflamm. Res. *69*, 1181–1189.

Jayatilleke, K.M., and Hulett, M.D. (2020). Heparanase and the hallmarks of cancer. J. Transl. Med. *18*, 453.

Jin, S., Guerrero-Juarez, C.F., Zhang, L., Chang, I., Ramos, R., Kuan, C.-H., Myung, P., Plikus, M.V., and Nie, Q. (2021). Inference and analysis of cell–cell communication using CellChat. Nat. Commun. *12*, 1088.

Kaimal, V., Bardes, E.E., Tabar, S.C., Jegga, A.G., and Aronow, B.J. (2010). ToppCluster: a multiple gene list feature analyzer for comparative enrichment clustering and network-based dissection of biological systems. Nucleic Acids Res. *38*, W96–W102.

Klimatcheva, E., Pandina, T., Reilly, C., Torno, S., Bussler, H., Scrivens, M., Jonason, A., Mallow, C., Doherty, M., Paris, M., et al. (2015). CXCL13 antibody for the treatment of autoimmune disorders. BMC Immunol. *16*, 6.

Kuwabara, T., Ishikawa, F., Yasuda, T., Aritomi, K., Nakano, H., Tanaka, Y., Okada, Y., Lipp, M., and Kakiuchi, T. (2009). CCR 7 ligands are required for development of experimental autoimmune encephalomyelitis through generating IL-23-dependent Th17 cells. J. Immunol. *183*, 2513–2521.

Laing, A.G., Lorenc, A., Del Molino Del Barrio, I., Das, A., Fish, M., Monin, L., Muñoz-Ruiz, M., McKenzie, D.R., Hayday, T.S., Francos-Quijorna, I., et al. (2020). Author Correction: a dynamic COVID-19 immune signature includes associations with poor prognosis. Nat. Med. *26*, 1951.

Lee, H.-T., Shiao, Y.-M., Wu, T.-H., Chen, W.-S., Hsu, Y.-H., Tsai, S.-F., and Tsai, C.-Y. (2010). Serum BLC/CXCL13 concentrations and renal expression of CXCL13/CXCR5 in patients with systemic lupus erythematosus and lupus nephritis. J. Rheumatol. *37*, 45–52.

Lee, J.S., Park, S., Jeong, H.W., Ahn, J.Y., Choi, S.J., Lee, H., Choi, B., Nam, S.K., Sa, M., Kwon, J.-S., et al. (2020). Immunophenotyping of COVID-19 and influenza highlights the role of type I interferons in development of severe COVID-19. Sci. Immunol. *5*, eabd1554.

Levi, M., Thachil, J., Iba, T., and Levy, J.H. (2020). Coagulation abnormalities and thrombosis in patients with COVID-19. Lancet Haematol. *7*, e438–e440.

Liao, M., Liu, Y., Yuan, J., Wen, Y., Xu, G., Zhao, J., Cheng, L., Li, J., Wang, X., Wang, F., et al. (2020). Single-cell landscape of bronchoalveolar immune cells in patients with COVID-19. Nat. Med. *26*, 842–844.

Mason, D.Y., Cordell, J.L., Brown, M.H., Borst, J., Jones, M., Pulford, K., Jaffe, E., Ralfkiaer, E., Dallenbach, F., and Stein, H. (1995). CD79a: a novel marker for B-cell neoplasms in routinely processed tissue samples. Blood *86*, 1453–1459.

McGonagle, D., Sharif, K., O'Regan, A., and Bridgewood, C. (2020). The role of cytokines including interleukin-6 in COVID-19 induced pneumonia and macrophage activation syndrome-like disease. Autoimmun. Rev. *19*, 102537.

Mehta, P., McAuley, D.F., Brown, M., Sanchez, E., Tattersall, R.S., and Manson, J.J.; HLH Across Speciality Collaboration, UK (2020). COVID-19: consider cytokine storm syndromes and immunosuppression. Lancet *395*, 1033–1034.

Merad, M., and Martin, J.C. (2020). Author Correction: pathological inflammation in patients with COVID-19: a key role for monocytes and macrophages. Nat. Rev. Immunol. *20*, 448.

Middleton, E.A., He, X.-Y., Denorme, F., Campbell, R.A., Ng, D., Salvatore, S.P., Mostyka, M., Baxter-Stoltzfus, A., Borczuk, A.C., Loda, M., et al. (2020). Neutrophil extracellular traps contribute to immunothrombosis in COVID-19 acute respiratory distress syndrome. Blood *136*, 1169–1179.

Nicolai, L., Leunig, A., Brambs, S., Kaiser, R., Weinberger, T., Weigand, M., Muenchhoff, M., Hellmuth, J.C., Ledderose, S., Schulz, H., et al. (2020). Immunothrombotic dysregulation in COVID-19 pneumonia is associated with respiratory failure and coagulopathy. Circulation *142*, 1176–1189.

Osterholm, C., Folkersen, L., Lengquist, M., Pontén, F., Renné, T., Li, J., and Hedin, U. (2013). Increased expression of heparanase in symptomatic carotid atherosclerosis. Atherosclerosis *226*, 67–73.

Pedersen, S.F., and Ho, Y.-C. (2020). SARS-CoV-2: a storm is raging. J. Clin. Invest. *130*, 2202–2205.

Rapkiewicz, A.V., Mai, X., Carsons, S.E., Pittaluga, S., Kleiner, D.E., Berger, J.S., Thomas, S., Adler, N.M., Charytan, D.M., Gasmi, B., et al. (2020). Megakaryocytes and platelet-fibrin thrombi characterize multi-organ thrombosis at autopsy in COVID-19: a case series. EClinicalMedicine *24*, 100434.

Reyes, M., Filbin, M.R., Bhattacharyya, R.P., Billman, K., Eisenhaure, T., Hung, D.T., Levy, B.D., Baron, R.M., Blainey, P.C., Goldberg, M.B., et al. (2020). An immune-cell signature of bacterial sepsis. Nat. Med. *26*, 333–340.

Rodríguez, Y., Novelli, L., Rojas, M., De Santis, M., Acosta-Ampudia, Y., Monsalve, D.M., Ramírez-Santana, C., Costanzo, A., Ridgway, W.M., Ansari, A.A., et al. (2020). Autoinflammatory and autoimmune conditions at the crossroad of COVID-19. J. Autoimmun. *114*, 102506.

Schafflick, D., Xu, C.A., Hartlehnert, M., Cole, M., Schulte-Mecklenbeck, A., Lautwein, T., Wolbert, J., Heming, M., Meuth, S.G., Kuhlmann, T., et al. (2020). Integrated single cell analysis of blood and cerebrospinal fluid leukocytes in multiple sclerosis. Nat. Commun. *11*, 247.

Schulte-Schrepping, J., Reusch, N., Paclik, D., Baßler, K., Schlickeiser, S., Zhang, B., Krämer, B., Krammer, T., Brumhard, S., Bonaguro, L., et al. (2020). Severe COVID-19 is marked by a dysregulated myeloid cell compartment. Cell *182*, 1419–1440.e23.

Schultz, N.H., Sørvoll, I.H., Michelsen, A.E., Munthe, L.A., Lund-Johansen, F., Ahlen, M.T., Wiedmann, M., Aamodt, A.-H., Skattør, T.H., Tjønnfjord, G.E., et al. (2021). Thrombosis and thrombocytopenia after ChAdOx1 nCoV-19 vaccination. N. Engl. J. Med. *384*, 2124–2130.

Shi, Y., Wang, Y., Shao, C., Huang, J., Gan, J., Huang, X., Bucci, E., Piacentini, M., Ippolito, G., and Melino, G. (2020). COVID-19 infection: the perspectives on immune responses. Cell Death Differ *27*, 1451–1454.

Silvin, A., Chapuis, N., Dunsmore, G., Goubet, A.-G., Dubuisson, A., Derosa, L., Almire, C., Hénon, C., Kosmider, O., Droin, N., et al. (2020). Elevated calprotectin and abnormal myeloid cell subsets discriminate severe from mild COVID-19. Cell *182*, 1401–1418.e18.

Sparkenbaugh, E., and Pawlinski, R. (2013). Interplay between coagulation and vascular inflammation in sickle cell disease. Br. J. Haematol. *162*, 3–14.

Spoerl, S., Kremer, A.N., Aigner, M., Eisenhauer, N., Koch, P., Meretuk, L., Löffler, P., Tenbusch, M., Maier, C., Überla, K., et al. (2021). Upregulation of CCR4 in activated CD8+ T cells indicates enhanced lung homing in patients with severe acute SARS-CoV-2 infection. Eur. J. Immunol. *51*, 1436–1448.

Steinmetz, O.M., Velden, J., Kneissler, U., Marx, M., Klein, A., Helmchen, U., Stahl, R.A.K., and Panzer, U. (2008). Analysis and classification of B-cell infiltrates in lupus and ANCA-associated nephritis. Kidney Int. *74*, 448–457.

Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., 3rd, Hao, Y., Stoeckius, M., Smibert, P., and Satija, R. (2019). Comprehensive integration of single-cell data. Cell *177*, 1888–1902.e21.

Swieringa, F., Spronk, H.M.H., Heemskerk, J.W.M., and van der Meijden, P.E.J. (2018). Integrating platelet and coagulation activation in fibrin clot formation. Res. Pract. Thromb. Haemost. *2*, 450–460.

Tay, M.Z., Poh, C.M., Rénia, L., MacAry, P.A., and Ng, L.F.P. (2020). The trinity of COVID-19: immunity, inflammation and intervention. Nat. Rev. Immunol. *20*, 363–374.

Terpos, E., Ntanasis-Stathopoulos, I., Elalamy, I., Kastritis, E., Sergentanis, T.N., Politou, M., Psaltopoulou, T., Gerotziafas, G., and Dimopoulos, M.A. (2020). Hematological findings and complications of COVID-19. Am. J. Hematol. *95*, 834–847.

Thålin, C., Hisada, Y., Lundström, S., Mackman, N., and Wallén, H. (2019). Neutrophil extracellular traps: villains and targets in arterial, venous, and cancer-associated thrombosis. Arterioscler. Thromb. Vasc. Biol. *39*, 1724–1738.

Wang, F., Nie, J., Wang, H., Zhao, Q., Xiong, Y., Deng, L., Song, S., Ma, Z., Mo, P., and Zhang, Y. (2020). Characteristics of peripheral lymphocyte subset alteration in COVID-19 pneumonia. J. Infect. Dis. *221*, 1762–1769.

Wilk, A.J., Rustagi, A., Zhao, N.Q., Roque, J., Martínez-Colón, G.J., McKechnie, J.L., Ivison, G.T., Ranganath, T., Vergara, R., Hollis, T., et al. (2020). A single-cell atlas of the peripheral immune response in patients with severe COVID-19. Nat. Med. *26*, 1070–1076.

Wilk, A.J., Lee, M.J., Wei, B., Parks, B., Pi, R., Martínez-Colón, G.J., Ranganath, T., Zhao, N.Q., Taylor, S., Becker, W., et al. (2021). Multi-omic profiling reveals widespread dysregulation of innate immunity and hematopoiesis in COVID-19. J. Exp. Med. *218*, e20210582.

Wolf, F.A., Angerer, P., and Theis, F.J. (2018). SCANPY: large-scale single-cell gene expression data analysis. Genome Biol. *19*, 15.

Wong, C.K., Wong, P.T.Y., Tam, L.S., Li, E.K., Chen, D.P., and Lam, C.W.K. (2010). Elevated production of B cell chemokine CXCL13 is correlated with systemic lupus erythematosus disease activity. J. Clin. Immunol. *30*, 45–52.

Yang, A.C., Kern, F., Losada, P.M., Agam, M.R., Maat, C.A., Schmartz, G.P., Fehlmann, T., Stein, J.A., Schaum, N., Lee, D.P., et al. (2021). Dysregulation of brain and choroid plexus cell types in severe COVID-19. Nature *595*, 565–571.

Zhang, F., Wei, K., Slowikowski, K., Fonseka, C.Y., Rao, D.A., Kelly, S., Goodman, S.M., Tabechian, D., Hughes, L.B., Salomon-Escoto, K., et al. (2019). Defining inflammatory cell states in rheumatoid arthritis joint synovial tissues by integrating single-cell transcriptomics and mass cytometry. Nat. Immunol. *20*, 928–942.

Zhao, M. (2020). Cytokine storm and immunomodulatory therapy in COVID-19: role of chloroquine and anti-IL-6 monoclonal antibodies. Int. J. Antimicrob. Agents *55*, 105982.

Zhou, Y., Han, T., Chen, J., Hou, C., Hua, L., He, S., Guo, Y., Zhang, S., Wang, Y., Yuan, J., et al. (2020). Clinical and autoimmune characteristics of severe and critical cases of COVID-19. Clin. Transl. Sci. *13*, 1077–1086.

Zuo, Y., Zuo, M., Yalavarthi, S., Gockman, K., Madison, J.A., Shi, H., Woodard, W., Lezak, S.P., Lugogo, N.L., Knight, J.S., et al. (2021). Neutrophil extracellular traps and thrombosis in COVID-19. J. Thromb. Thrombolysis *51*, 446–453.

CellPress
OPEN ACCESS

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
scRNA-seq for COVID-19 PBMC	Arunachalam et al., 2020	GSE155673
scRNA-seq for COVID-19 PBMC	Guo et al., 2020	GSE150861
scRNA-seq for COVID-19 PBMC	Lee et al., 2020	GSE149689
scRNA-seq for COVID-19 PBMC	Schulte-Schrepping et al., 2020	EGAS00001004571
scRNA-seq for COVID-19 PBMC	Wilk et al., 2020	GSE150728
scRNA-seq for COVID-19 BAL	Liao et al., 2020	GSE145926
scRNA-seq for COVID-19 BAL	Grant et al., 2021	GSE155249
scRNA-seq for COVID-19 Lung Biopsy	Bharat et al., 2020	GSE158127
scRNA-seq for Sepsis PBMC	Reyes et al., 2020	SCP548, SCP550
scRNA-seq for Multiple-Sclerosis PBMC	Schafflick et al., 2020	GSE138266
Integrated COVID-19 scRNA-seq data	This paper	Mendeley Data: https://doi.org/10.17632/vrxdg7mm6x.1
Integrated COVID-19 scRNA-seq interface	This paper	cellxgene: https://cellxgene.cziscience.com/collections/b9fc3d70-5a72-4479-a046-c2cc1ab19efc
Software and Algorithms		
Seurat (v4.0)	Hao et al., 2021	https://satijalab.org/seurat/
R (v3.6.1)	N/A	https://www.r-project.org/
Scanpy (v1.7.2)	Wolf et al., 2018	https://scanpy-tutorials.readthedocs.io/en/latest/index.html
Python (v3.7.0)	N/A	https://www.python.org/
CellChat (v1.0.0)	Jin et al., 2021	http://www.cellchat.org/
ToppGene	Chen et al., 2009	https://toppgene.cchmc.org/
ToppCluster	Kaimal et al., 2010	https://toppcluster.cchmc.org/
ToppCell	this paper	https://toppcell.cchmc.org/
EnhancedVolcano (v1.4.0)	Blighe et al., 2021	https://bioconductor.org/packages/release/bioc/html/EnhancedVolcano.html
Cellxgene	N/A	https://github.com/chanzuckerberg/cellxgene

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the Lead Contact, Bruce Aronow (bruce.aronow@cchmc.org).

Materials availability

This study did not generate new unique reagents.

Data and code availability

● Public single-cell RNA-seq datasets of PBMC in COVID-19 patients are available on NCBI Gene Expression Omnibus (GEO) and European Genome-phenome Archive, including GSE150728, GSE155673, GSE150861, GSE149689 and EGAS00001004571 (or Schulte-Schrepping_2020_COVID19_10x_PBMC under FastGenomics). BAL single-cell RNA-seq datasets of COVID-19 patients are available on GSE145926 and GSE155249. Lung Parenchyma single-cell RNA-seq data are available on GSE158127. Single-cell

RNA-seq data of sepsis patients are available on the Single Cell Portal SCP548 and SCP550. Data of multiple sclerosis patients are available on GSE128266. Data of severe influenza patients are available on GSE149689. Accession numbers are also listed in the key resources table. Original and processed data can be downloaded on Mendeley Data (https://doi.org/10.17632/vrxdg7mm6x.1). Gene modules of all datasets analyzed using ToppCell web portal are available on COVID-19 Atlas in ToppCell (https://toppcell.cchmc.org), including gene modules from either a single dataset or an integrated dataset. Gene modules from the integration of specific cell types, such as B cells and neutrophils are also listed in ToppCell. More details are listed in Figure1A and Table S1. An interactive interface of integrated PBMC data and subclusters of immune cells will be public on cellxgene (https://cellxgene.cziscience.com/collections/b9fc3d70-5a72-4479-a046-c2cc1ab19efc).

- Codes of preprocessing, normalization, clustering and plotting of single-cell datasets are available on github: https://github.com/KANG-BIOINFO/COVID-19-Atlas.

- Any additional information required to reanalyze the data reported in the paper is available from the lead contact upon request.

METHOD DETAILS

Single-cell RNA-seq data source

To have a comprehensive understanding of immune cells in different repertoires, we collected 8 public COVID-19 single-cell RNA-seq datasets of multiple compartments, including peripheral blood mononuclear cells, bronchoalveolar lavage and lung biopsy, which in total covered over 43 healthy donors, 22 mild/moderate, 42 severe and 2 convalescent COVID-19 patients. More details can be found in Figure 1A and Table S1 and Data Availability. Lung biopsy samples were taken from the explanted lung or postmortem lungs of COVID-19 patients (Bharat et al., 2020). Various criteria were used in these publications to describe COVID-19 severity. For example, we found asymptomatic, mild, moderate and floor COVID-19 patients under the definition of non-severe COVID-19 patients in our data sources. A recent paper used the WHO score of COVID-19 severity to categorize disease conditions of patients (Wilk et al., 2021), which is a more standardized and robust approach for the description of disease stages. However, in order to address the issue of missing information for disease stratification and to simplify the comparison, we grouped disease conditions into three groups, including healthy donors, mild COVID-19 patients and severe COVID-19 patients. Convalescent patients were excluded in some of our analysis for simplification. Sequencing data of healthy donors in Guo et al. was excluded since it was not from the same institute (Guo et al., 2020).

We also collected PBMC single-cell RNA-seq data from 29 sepsis patients (Reyes et al., 2020) and 4 multiple sclerosis (Schafflick et al., 2020) patients for comparative analysis of immune-mediated diseases (Figure 1A; Table S1). Data sources can be found in Data Availability.

Data preprocessing and normalization

For datasets with raw UMI counts, we first removed cells with less than 300 detected genes or less than 600 UMI counts. Then cells with more than 15% counts of mitochondrial genes were filtered out. Genes expressed in less than 5 cells were removed.

To remove doublets and low-quality cells which might pass the filtering (Ding et al., 2020), we further interrogated cell qualities in individual datasets. Distributions of UMI counts or number of genes per cell were used as criteria of clusters with low sequencing depth, which is one of important features of low-quality cells. Doublets usually express marker genes of multiple cell types (DePasquale et al., 2019), which is an important evidence that we used for doublet cluster detection. Before we integrated cells, we first investigated cell qualities and filter low-quality cells in each individual dataset. Original data without quality control can be found in Data Availability.

Cluster 14 in Arunchalam et al.'s data was removed due to its low-sequencing depth; Cluster 13 in Guo et al.'s data and cluster 9 in Schulte-Schrepping et al.'s data were removed for the same reason; the bottom region of cluster 2 on the UMAP, which is close to platelets but expresses high levels of T cell markers, was defined as T cell – platelet doublets and removed using cellxgene.

For some datasets that only provide processed and normalized *h5ad* or *rds* files, we checked their preprocessing procedures in the original publications and confirmed that stringent quality control procedures were used.

After quality control, we finally harvested 483,765 high-quality cells from 8 studies (Table S1). Both original and processed data can be found in Data Availability. We normalized the total UMI counts per gene to 1 million (CPM) and applied log_2(CPM+1) transformation for heatmap visualization and downstream differential gene expression analysis. Steps above were done in Scanpy (Wolf et al., 2018).

Integration of PBMC datasets and BAL datasets using reciprocal PCA in seurat

We input raw count files of 5 preprocessed PBMC datasets into Seurat and created a list of Seurat objects. Reciprocal PCA procedure (https://satijalab.org/seurat/v3.2/integration.html#reciprocal-pca) was used for data integration. First, normalization and variable feature detection were applied for each dataset in the list. Then we used *SelectIntegrationFeatures* to select features for downstream integration. Next, we scaled data and ran the principal component analysis with selected features using *ScaleData* and *RunPCA*. Then we found integration anchors and integrated data using *FindIntegrationAchnors* and *IntegrateData*. RPCA was used as the reduction method. After integration, we scaled data and ran PCA on integrated expression values. UMAP was generated using the top 30 reduced dimensions with *RunUMAP*. The same approach was also used in BAL data integration and multi-disease integration. We also used it for the integration of specific cell types across multiple datasets, for example, the integration of neutrophils from PBMC and BAL datasets. Compared with standard workflow and SCTransform (https://satijalab.org/seurat/v3.2/integration.html) in Seurat, we found Reciprocal PCA is much less computation-intensive and time-consuming, making the integration of multiple large single-cell datasets feasible.

Cell annotations using canonical markers after unsupervised clustering

Cell annotations were assigned in each dataset and then mapped to the integrated data. For some datasets without available cell annotations, we first used unsupervised clustering in Scanpy. Detailed steps include (1) detecting top 3,000 highly variable genes using *pp.highly_variable_genes*; (2) scaling each gene to unit variance on highly variable genes using *pp.scale*; (3) running PCA using *arpack* approach in *tl.pca*; (4) finding neighbors using *pp.neighbors*; (5) running leiden clustering with resolution of 1 using *tl.leiden* (resolutions were determined swiftly based on the size and complexity of data). More details can be found in the code (point to it). For datasets with available annotations, we checked their validity and corrected wrong annotations. For example, hematopoietic stem and progenitor cells (HSPC) were mistakenly annotated as "SC&Eosinophil" in the original paper (Wilk et al., 2020) and were corrected in our annotation.

After unsupervised clustering, well recognized immune cell markers were used to annotate clusters, including CD4+ T cell markers such as TRAC, CD3D, CD3E, CD3G, CD4; CD8+ T cell markers such as CD8A, CD8B, NKG7; NK cell markers such as NKG7, GNLY, KLRD1; B cell markers such as CD19, MS4A1, CD79A; plasmablast markers such as MZB1, XBP1; monocyte markers such as S100A8, S100A9, CST3, CD14; conventional dendritic cell markers such as XCR1, plasmacytoid dendritic cell markers such as TCF4; megakaryocyte/platelet marker PPBP; red blood cell markers HBA1, HBA2; HSPC marker CD34. Exhaustion-associated markers, including PDCD1, HAVCR2, CTLA4 and LAG3 were used to identify exhausted T cells.

Additionally, other markers were used for annotations of lung-specific cells, including AGER, MSLN for AT1 cells; SFTPC, SFTPB for AT2 cells; SCGB3A2, SCGB1A1 for Club cells; TPPP3, FOXJ1 for Ciliated cells; KRT5 for Basal cells; CFTR for Ionocytes; FABP4, CD68 for tissue-resident macrophages; FCN1 for monocyte-derived macrophages, TPSB2 for Mast cells. More details can be found in Table S2.

Cell annotations using Azimuth

To better annotate T cells in our study, we applied Azimuth (https://satijalab.org/azimuth/), a tool for reference-based single-cell analysis developed in Seurat version 4.0 (Hao et al., 2021). High-quality PBMC single-cell data in Azimuth was used as the reference for label projection. After removing annotations with low prediction scores or low mapping scores, we got a collection of well-annotated T cell subtypes, including CD4+ Cytotoxic T cell, CD4+ Naive T cell, CD4+ Central Memory T cell, CD8+ Naive T cell, CD8+ Effector Memory cell, gamma-delta T cell, double-negative T cell. CD4+ Effector Memory T (CD4+ T_{EM}) cell and

CD8+ Central Memory T (CD8+ T_{CM}) cell were found by Azimuth. However, we manually re-annotated them into other T cell subtypes, such as CD4+ T_{CM} and CD8+ T_{EM} cells, based on expression levels of marker genes. There're several reasons for the re-annotation: (1) There's no clear cluster or boundary for these two predicted cell types on reduced dimension (UMAP), which indicates that they may have no distinct transcriptomic pattern compared with other T cell subtypes; (2) Many marker genes didn't show significant expression patterns in these two cell types. For example, CD4+ T_{EM} markes from Azimuth, such as TMSB10 and ITGB1, didn't show higher expression levels in Azimuth-predicted CD4+ T_{EM} cells in Arunchalam et al.'s data; (3) Although developers and researchers of Azimuth spent a lot of time annotating cells, the existence of CD4 T_{EM} and CD8 T_{CM} in single-cell data was not widely observed by other groups. In addition, we didn't use CITE-seq data like the reference dataset of Azimuth. Thus, the power of detecting fine resolution of CD4 and CD8 T cell subtypes might be relatively low.

Apart from annotations of T cell subtypes, we also found CD56-bright NK cell, intermediate B cell and memory B cell using Azimuth.

Sub-clustering for specific cell types

Sub-clustering was used for the discovery of subtypes or distinct stages of a specific cell type. In our work, we applied sub-cluster for various immune cell types, including classical monocytes, neutrophils, conventional dendritic, B cells and platelets. First, all cells in the specific cell type were integrated using the same procedure as PBMC data integration. Then Louvain clustering (resolution = 0.5, except for sub-clustering of classical monocytes where resolution = 0.3) was applied to detect sub-clusters of those cells. Importantly, neutrophils, cDCs and B cells were retrieved from both PBMC and BAL, whereas classical monocytes and platelets were only retrieved from PBMC.

Generation of ToppCell gene modules

ToppCell (https://toppcell.cchmc.org/) was designed to parallelly analyze transcriptional profiles of single-cell datasets by organizing differential expressed gene modules hierarchically as a function of sample types and compartments, clinical subgroups, and cell lineage, class, and subclass designations that emerge from post-hoc cell type evaluations. All the cells were grouped into specific hierarchical categories. For example, "PBMC_severe COVID-19_myeloid cells_classical-monocytes_cMono1" represents cells belonging to cMono1 (a sub-cluster of classical monocytes) in PBMC of severe COVID-19 patients. Within hierarchically ordered cell annotations, we calculated corresponding DEGs in a hierarchical way as well and consistently. We defined customized ranges for comparisons and applied t-test based on normalized expression values. More details can be seen on ToppCell website. Usually, the top 200 most differentially genes in each comparison were picked up as the gene modules for the selected cell group, which are the starting point of downstream analysis, including gene enrichment in ToppGene and interaction inference in ToppCluster. All gene modules in our study were curated in COVID-19 Atlas (https://toppcell.cchmc.org/biosystems/go/index3/COVID-19 Atlas) and ImmuneMap (https://toppcell.cchmc.org/biosystems/go/index3/ImmuneMap) on the ToppCell website.

Gene enrichment analysis using ToppGene

Gene modules generated with ToppCell were individually analyzed using the ToppGene/ToppFun, and comparatively analyzed using ToppCluster. Principle annotation categorizations used were GO-Molecular Function, GO-Biological Process, GO-Cellular Component, Mouse and Human Mutation Phenotypes, as well as ImmGen reference signatures, MSigDB Pathways, NCBI Pub2Gene and Protein Interactions, and DisGeNet Curated sources. P values of enrichment results were adjusted using the Benjamini-Hochberg procedure.

Generation of functional association heatmap using ToppCluster

Genes in gene modules of selected cell types or sub-clusters were sent to ToppCluster (https://toppcluster.cchmc.org/). Then multi-group functional enrichment was drawn for input gene modules and $-\log_{10}$(adjusted p-value) was used as the gene enrichment score to represent the strength of association between gene modules and pathways. Scores greater than 10 were trimmed to 10. Pathways from Gene Ontologies, including Molecular Functions, Biological Process and Cellular Component in the option list were used for the enrichment of gene modules in myeloid cells, B cells and platelets. In order to gain a broader knowledge of immunothrombosis-related pathways, "Pathway" and "Mouse Phenotype'' in the

option list were also selected for enrichment. Morpheus was used for visualization of the heatmap (https://software.broadinstitute.org/morpheus/).

Cell interaction inference in immunothrombosis activities and cytokine signaling pathways

CellChat was used to infer the signaling network in the BAL of severe patients (Figure S8B). All 3 categories of interactions were used in the database *CellChatDB.human*. Over-expressed ligands or receptors in each cell type were first identified for further identification of over-expressed interaction pairs. Then cytokine, chemokine and IL signaling probability between multiple cell types was inferred using *computeCommunProb* and *computeCommunProbPathway*.

ToppCell was used to infer interactions in immunothrombosis. We first selected genes related to coagulation or immunothrombosis pathways from subtypes of endothelial cells, platelets, neutrophils, classical monocytes and monocyte-derived macrophages by filtering the output of ToppCluster (Figure S12A). Then we used CellChatDB as the knowledge base to find the subset of genes participating in cell-cell interaction, including genes involved in signaling via secretion, cell-cell contact and extracellular matrix interaction. These genes in each cluster were sent to ToppCluster to infer the interaction network using protein-protein interactions (PPI) between those genes.

Generation of volcano plots

We first calculated differential expressed genes using *tl.rank_genes_groups* in Scanpy. Adjusted p values and log fold changes in the output were used as the input of volcano plots. R package *EnhancedVolcano* (Blighe et al., 2021) was used to draw figures.

Construction of COVID-19 functional enrichment map

In order to characterize functional properties of cell types and subtypes observed in BAL, PBMC, and lung parenchymal samples from control, mild, and severe COVID-19 patient samples, we used the library of gene expression signatures ("Gene Module Report" from ToppCell) as an input to the ToppCluster enrichment analyzer web server (Kaimal et al., 2010). Using categories of *Gene Ontology*, *Human Phenotype*, *Mouse Phenotype*, *Pathway* and *Protein Interaction,* a matrix was constructed using minus log P enrichment values for each celltype gene list and then all cells and enriched features could be clustered and ordered based on their shared or distinct properties that could then be associated with lineage, cell subclass, tissue compartment, and disease state.

QUANTIFICATION AND STATISTICAL ANALYSIS

Cell proportion differences between disease groups for specific types and subtypes (Figures 2, S2–S4) shown on box plots were measured by Mann-Whitney test (Wilcoxon, paired = False). Significance between two disease conditions were shown on the top of figures. More details can be found in figure legends.

To investigate the dynamic changes of cell proportions across various immune-mediated diseases, we followed the approach in recent literature (Lee et al., 2020) (Figure 7B). For each disease condition, we computed the relative ratio of each cell type in individual disease samples divided by individual healthy samples. Log_2 transformed values were shown in the box plot. Then we calculated relative ratios of each cell type between all sample pairs of healthy donors as a control. To compute the significance, we used a two-sided Kolmogorov-Smirnov (KS) test using relative ratios in diseases and those values in healthy donors. More details can be found in figure legends.