A Web Portal and Workbench for Biological Dissection of Single Cell 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 P. Krummel[6], Marc E. Rothenberg[7], Bruce J. Aronow[1,8*]

[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

[*]Corresponding author. Email: bruce.aronow@cchmc.org (B.A.)

Abstract

Numerous studies have provided single-cell transcriptome profiles of host responses to SARS-CoV-2 infection. Critically lacking however is a reusable datamine to allow users to compare and explore these data for insight, inference, and hypothesis generation. To accomplish this, we harmonized datasets from blood, bronchoalveolar lavage and tissue samples from COVID-19 and other control conditions and derived a compendium of gene signature modules per cell type, subtype, clinical condition and compartment. We demonstrate approaches for exploring and evaluating their significance via a new interactive web portal (ToppCell). As examples, we develop three hypotheses: (1) a multicellular signaling cascade among alternatively differentiated monocyte-derived macrophages whose tasks include T cell recruitment and activation; (2) novel platelet subtypes with drastically modulated expression of genes responsible for adhesion, coagulation and thrombosis; (3) a multilineage cell activator network able to drive extrafollicular B maturation via an ensemble of genes extensively associated with risk for developing autoimmunity.

Teaser

Implicating COVID-19 Gene and Cell Networks Responsible for Inflammation, Thromboembolism and Autoimmune Pathobiology.

Short title

Data mining COVID-19 Gene and Cell Networks

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 (*1*, *2*) 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 (*3*, *4*), cytokine storms (*5*, *6*), differential interferon responses (*7*, *8*) and emergency myelopoiesis (*9*, *10*). 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.

Whereas many studies have focused on the peripheral blood mononuclear cells (PBMC) (*9*, *11–14*) due to ease of procurement, other studies have profiled airway locations via bronchoalveolar lavage (BAL) (*15*, *16*), nasopharyngeal swabs, and bronchial brushes (*17*). Additional sampling sites that could also be infected or affected have also been approached in autopsy-derived materials from the central nervous system (*18*, *19*), and other sites (*20*). 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 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 datamining 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 (Fig. 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 (*21*) 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 (*22*). Sub-clustering was applied for some cell types, such as neutrophils and platelets, to interrogate finer resolutions of disease-specific sub-populations (Fig. 1B). Using the ToppCell toolkit (https://toppcell.cchmc.org/), we created over 3,000 hierarchical gene modules of the most significant differentially expressed genes (DEGs) for all cell classes and sub-clusters across compartments and disease severity (table S1). These modules were then used to infer cell-cell interactions as well as upregulated pathways, which were further combined for functional comparative analysis in a specific cell manner in ToppCluster (*23*) (Fig. 1B), such as sub-clusters

of platelets. Integration of ToppCluster output of cells from multiple compartments and disease conditions built pathogenic maps, highlighted by the coagulation map of COVID-19 (fig. 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 (Fig. 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 (Fig. 2, A and C, table S2). Shifts of Uniform Manifold Approximation and Projection (UMAP) of cell type distributions were observed in both compartments of mild and severe patients (Fig. 2, A and C, fig. S1A and fig. S3A). In PBMC, conventional dendritic cells (cDC), plasmacytoid dendritic cells (pDC) and non-classical monocytes displayed a prominent reduction in severe patients (Fig. 2B and fig. S1C), consistent with prior reports (*11*, *24*, *25*). In contrast, severe patients demonstrated dramatic expansion of neutrophils, especially immature stages (fig. S1C and fig. S2). Integration with evoked pathways in the following analysis implicated that neutrophil expansion was likely the consequence of emergency myelopoiesis (*26*). Additionally, a general down-regulation of T cell and NK cell was observed, consistent with lymphopenia reported in clinical practices (*5*, *27*) (fig. S1C and fig. 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 (fig. S2). Notably, plasmablasts substantially increased in COVID-19 patients, and especially so in severe patients, suggesting upregulated antibody production (*28*) (Fig. 2B and fig. 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 (*29*, *30*) (Fig. 2B and fig. 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 (Fig. 2, C and D and fig. S3D). Mild patients exhibited a moderate reduction of tissue-resident macrophages, but no evidence of aggregation of monocyte-derived macrophages (Fig. 2, C and D, fig. S3, A and D). Dynamic changes of these two subtypes suggest increased tissue chemoattraction (*31*) and potential damage of patients' lungs (*32*). In

50 addition, neutrophils were only identified in severe patients in the integrated BAL data (Fig. 2C
51 and fig. S3A), which might be related with neutrophil extracellular traps (NETs) in the lung (*33*).
52 However, more samples are required to draw a solid conclusion. We also noted conventional
53 dendritic cells decreased in the severe patients, which is consistent with the trend of the
54 counterpart in PBMC data. Opposite to the change in PBMC, an expansion of plasmacytoid
55 dendritic cells is observed in both mild and severe patients (Fig. 2D). Other cell types, including T
56 cell and NK cell in the BAL, also have converse changes of their counterparts in PBMC, which
57 could be attracted by lung macrophages or epithelial cells after infection or damages (*17*) (Fig. 2D
58 and fig. S3D). These changes were consistently observed in lung parenchyma samples from
59 severe COVID-19 patients (fig. S4). With cells well-annotated in the integrated COVID-19 atlas,
60 we drew a global heatmap for cells in both blood and lung using ToppCell gene modules (top 50
61 DEG in each module) of all identified cell classes. While there was conservation of gene patterns
62 involved in healthy donors and severe COVID-19 patients, there were substantial differences
63 most notably in myeloid cells (Fig. 2E). Such hierarchically ordered ToppCell gene modules were
64 broadly used in visualization, large-scale comparisons and fine-resolution investigations in the
65 following analyses.
66

67 **Myeloid Cell Atlas: Functionally Distinct Neutrophils at Different Levels of Maturation and**
68 **Derailed Macrophages in the Lung**

69 Dysregulated myeloid cells have been reported as an important marker of severe COVID-19
70 patients (*9*, *10*). In order to gain a deeper and comprehensive understanding of these cells, we
71 applied the sub-clustering strategy on the integrated data of key cell types, such as neutrophils and
72 macrophages, and then generated gene modules for comparative functional analysis and
73 interactome inference. We successfully identified 5 neutrophil sub-clusters after the integration of
74 PBMC and BAL data, including 3 FCGR3B+ mature sub-clusters and 2 FCGR3B- immature sub-
75 clusters (Fig. 3A and fig. S5B, table S3). They're mainly from severe patients and their gene
76 modules were generated and subjected to comparative functional enrichment using ToppCell and
77 ToppCluster (Fig. 3, C and D, fig. S5A). We identified proliferative neutrophils (referred to as
78 pro-neutrophils and Neu4) and MMP8^{high} precursor immature neutrophils (referred to as pre-
79 neutrophils and Neu2) (Fig. 3A and fig. S5B) consistent with prior studies (*9*). While immune
80 response genes and pathways were barely activated in the immature neutrophils, they displayed
81 upregulation of granule formation pathways and NETosis-associated proteins, including ELANE,
82 DEFA4 and MPO, especially in Neu4 (*9*, *26*) (Fig. 3C, fig. S5, B and C). Upregulated myeloid

leukocyte mediated immunity in Neu2 suggests involvement of this cell type in anti-viral function (Fig. S5D). Yet, the absence of cytokine and interferon response pathways suggests the lack of mature immune responses (Fig. 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 (Fig. 3, C and D, fig S5, B and D).

MoAM and TRAM were two main macrophage types in the BAL (Fig. 2C); both are known to have distinct roles in immune responses in the lung (*15*). As described above, five sub-clusters among the expanded COVID-19 patient-specific MoAM (Fig. 3B, table S3) were found, where the loss of HLA class II genes and elevation of interferon-stimulated genes (ISGs) were consistently observed (Fig. 3F and fig. S6A). Relative to MoAM3,4, MoAM1,2,5 displayed an upregulation of interferon responses and cytokine production (Fig. 3D, fig. S6B and table S3), indicating their pro-inflammatory characteristics. Notably, MoAM5 shows dramatic upregulation of IL-6 secretion and cytokine receptor binding activities (Fig. S7, A to D). However, cells in this sub-cluster were mainly from one severe patient (fig. 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 (Fig. 3B and fig. 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 (Fig. 3D), while the activated group from mild and severe patients displays upregulation of ISGs and cytokine signaling pathways (fig. S6B and 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 (Fig. 3D, fig. S6B and 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 an understanding of the interaction network in the lung microenvironment of severe COVID-19 patients, we focused on signaling ligands, receptors and pathways using ToppCell and CellChat (Fig. 3E and fig. S8, A and B). 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 (fig. 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 (Fig. 3E) (*15*). 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, 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 (fig. 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 (Fig. 3E). In addition, neutrophils exhibit an extraordinarily high level of IL1B, which could potentially in turn activate macrophages (fig. S8, A and B). 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 towards CXCR2 could be a potential chemoattraction pathway for neutrophils. In turn, neutrophils could activate TRAM3 by secreting IL1B, which binds with IL1RAP. Additionally, CD4+ T cells could also activate TRAM3 by IL10-IL10RB interaction (Fig. 3E and fig. S8, A and B).

In addition to neutrophils and macrophages, the upregulation of ISGs was observed in classical monocytes of both mild and severe patients (cMono3, cMono4), while the reduction of the MHC class II cell surface receptor HLA-DR genes was only observed in severe patients (cMono4) (fig. S9). In cDCs, polarization of interleukin secretion was observed in mild-patient and severe-patient specific clusters (fig. 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 (fig. S11C). Pro-inflammatory gene expression was found in all major myeloid cell types, including cMono4, DC1, DC9 in PBMC and Neu3, DC10, MoAM1, MoAM2, 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 thromboses (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 (*29*, *30*, *34–36*). At present, we lack a molecular or cellular explanation of the underlying basis of this pathobiology (*29*, *37*). 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 (fig. 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 (Fig. 4, A and B). Severe-patient-specific PLT0 is an interesting sub-cluster with elevated integrin genes, including ITGA2B, ITGB1, ITGB3, 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 (*38*). In addition, upregulated pathways of hemostasis, wound healing and blood coagulation were also observed in PLT0 (Fig. S13A and table S4). Importantly, PLT2 is an inflammatory sub-cluster with an upregulation of ISGs and interferon signaling pathways, while PLT4 is highlighted by upregulated post-transcriptional RNA splicing activities (Fig. S13, A and C).

Severity-associated gene patterns were also identified by selecting coagulation-associated genes modules (Fig. 4C and 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) (*39*), Anoctamin-6 (ANO6) (*40*), and selectin P (SELP) (*41*) (Fig. 4, C and D). Heparanase is an endoglycosidase that cleaves heparan sulfate constituents, a major component of anti-coagulation glycocalyx on the surface of vascular endothelium (*42*, *43*). Upregulated heparanase was related to upregulation of cell-matrix adhesion and coagulation (Fig. 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 (*44*). In addition, other upregulated genes involved in coagulation-associated activities were also observed, including wound healing, fibrinolysis, platelet aggregation and activation (Fig. 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 thrombocytopaenic purpura (ITP) is now recognized as a known disease complication (*45–49*). 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 (Fig. 5, A and B and fig. S14, A and B). Several COVID-19 specific sub-clusters were identified in B cells, such as ISG^{high} activated B cells (cluster 7) (Fig. 5A). Importantly, activated B cells showed dramatic upregulation of interferon signaling pathways and cytokine productions (fig. 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 (Fig. 5B). In contrast, non-dividing plasmablasts displayed upregulation of immunoglobulin genes (IGHA1, IGHA2, IGKC), B cell markers (CD79A) (*50*), interleukin receptors (IL2RG) and type II HLA complex (HLA-DOB) (Fig. 5C and 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 (fig. S14, E and F). Antibody production activities were upregulated in non-dividing plasmablasts based on gene enrichment analysis (Fig. S15A and table S5). Collectively, we inferred that non-dividing plasmablasts had definite immunoglobulin isotypes and were actively involved in immune responses towards COVID infection, while developing plasmablasts were less mature but highly proliferative to replenish the repertoire of plasma cells.

Since 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 (Fig. 5D). First, gene modules of B cells and other cells in severe patients were collected and subjected to ToppGene for enrichment analysis. Then we queried autoimmunity-associated terms in the enriched output and identified associated genes. After that, we retrieved interaction pairs using ToppCluster and CellChat database (*51*). 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, autoimmune encephalomyelitis (*52–56*). Candidate cytokine and chemokine ligand genes were 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 (Fig. 5E and table S5). These interaction pairs have been linked with auto-immunity (*57*, *58*). In addition, we analyzed single-cell studies (*59*, *60*) 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 (fig. S15, C and D). 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 (Fig. 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 (fig. S16 and fig. 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 (Fig. 6A and 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 (Fig. 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, while 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 (Fig. 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 Fig. 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 (*12*), sepsis (*61*) and multiple sclerosis (*62*). 404,125 cells were included after the integration of PBMC single-cell datasets (Fig. 7A and fig. 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 (Fig. 7B). However, the reduction of non-classical monocytes is more significant in severe COVID-19 patients than severe influenza or mild COVID-19 patients (Fig. 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 (Fig. 7B).

In addition to dynamic changes of cell ratios, we also investigated the regulation of immune mediator genes across various diseases (Fig. 7C and table S7). IL-6 is an important factor of cytokine storms in COVID-19 (*63*). 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 (Fig. 7C). Specific ligands, including CXCL2, CXCL3, CCL20 were upregulated in both severe COVID-19 patients and severe influenza patients. CCR4 and IL2RA is 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 network 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 procoagulant endoglycosidase that cleaves anti-coagulation heparan sulfate constituents on endothelial cells and potentially causes thrombotic vascular damages. Additionally, heparanase-cleaved heparan sulphate (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 (*64*), further contributing to the hyperinflammatory environment in COVID-19 patients. Since heparanase is recognized as a hallmark in tumor progression and metastasis (*65*), 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 (*66*). Approaches to combatting NETs could a potential anticoagulation treatment (*67*).

We propose a signaling network which potentially shapes the differentiation of B cells towards 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 naïve 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 (*68*). 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 naïve 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, TNFRSF13) are known to be genetically associated with rheumatoid arthritis, lupus, and risk of developing autoimmune disease in humans and mouse models. The development of different patterns of autoimmunity may be the main hallmark of "Long Haul" Covid disease and could explain why some individuals develop different autoantibodies and suffer different forms of clinical consequences depending on which antigens drive the B-cell maturation. 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, while 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 (*69*). 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 hypotheses 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 statistical methods that extend the current ToppCell / ToppGene Suite based on fuzzy measure similarity, Page-Rank, and cell-cell signaling approaches.

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 to the accuracy of future studies. Additionally, 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.

Materials and Methods

Experimental design and 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 Fig. 1A and table S1. Lung biopsy samples were taken from the explanted lung or post-mortem lungs of COVID-19 patients (*70*). 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 (*26*), 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 (*14*).

We also collected PBMC single-cell RNA-seq data from 29 sepsis patients (*61*) and 4 multiple sclerosis (*62*) patients for comparative analysis of immune-mediated diseases (Fig. 1A and 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. After quality control, we finally harvested 483,765 high-quality cells from 8 studies (table S1). 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 (*71*).

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. Most of them used the default normalization approach in the Seurat or Scanpy pipeline. We transferred them to $\log_2$(CPM+1) to make data consistently normalized. We also prepared corresponding raw count files for data integration.

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. 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 (*11*) 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 (*22*). 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 cell and CD8+ Central Memory T cell were found by Azimuth but removed later because of low scores. 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 in a customized hierarchical order. In our study, we hierarchically annotated cells with multiple layers, including compartments, disease conditions, lineages, cell classes and sub-clusters. 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. With hierarchically ordered cell

annotations, we calculated their DEGs in a hierarchical way as well. 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

Abundant gene modules were generated with ToppCell. After that, we used ToppGene (https://toppgene.cchmc.org/) for gene enrichment analysis. Genes in each gene module were sent to ToppGene platform as input for enrichment in different domains. GO-Molecular Function, GO-Biological Process and GO-Cellular Component and Mouse Phenotype were usually used for enrichment. 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 (fig. 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 (fig. 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 (72)* 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.

Statistics Analysis of Cell Proportion Changes in Different Disease Stages

Cell proportion differences between disease groups for specific types and subtypes (Fig. 2 and fig. S2 to S4) shown on box plots were measured by Mann-Whitney test (Wilcoxon, paired=False). Significance between two disease conditions were shown on the top.

To investigate the dynamic changes of cell proportions across various immune-mediated diseases, we followed the approach in recent literature (*12*) (Fig. 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.

References

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

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

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

4. F. Wang, J. Nie, H. Wang, Q. Zhao, Y. Xiong, L. Deng, S. Song, Z. Ma, P. Mo, Y. Zhang, Characteristics of Peripheral Lymphocyte Subset Alteration in COVID-19 Pneumonia. *J. Infect. Dis.* **221**, 1762–1769 (2020).

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

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

7. D. Blanco-Melo, B. E. Nilsson-Payant, W.-C. Liu, S. Uhl, D. Hoagland, R. Møller, T. X. Jordan, K. Oishi, M. Panis, D. Sachs, T. T. Wang, R. E. Schwartz, J. K. Lim, R. A. Albrecht, B. R. tenOever, Imbalanced Host Response to SARS-CoV-2 Drives Development of COVID-19. *Cell*. **181**, 1036–1045.e9 (2020).

8. J. Hadjadj, N. Yatim, L. Barnabei, A. Corneau, J. Boussier, N. Smith, H. Péré, B. Charbit, V. Bondet, C. Chenevier-Gobeaux, P. Breillat, N. Carlier, R. Gauzit, C. Morbieu, F. Pène, N. Marin, N. Roche, T.-A. Szwebel, S. H. Merkling, J.-M. Treluyer, D. Veyer, L. Mouthon, C. Blanc, P.-L. Tharaux, F. Rozenberg, A. Fischer, D. Duffy, F. Rieux-Laucat, S. Kernéis, B. Terrier, Impaired type I interferon activity and inflammatory responses in severe COVID-19 patients. *Science*. **369**, 718–724 (2020).

9. J. Schulte-Schrepping, N. Reusch, D. Paclik, K. Baßler, S. Schlickeiser, B. Zhang, B. Krämer, T. Krammer, S. Brumhard, L. Bonaguro, E. De Domenico, D. Wendisch, M. Grasshoff, T. S. Kapellos, M. Beckstette, T. Pecht, A. Saglam, O. Dietrich, H. E. Mei, A. R. Schulz, C. Conrad, D. Kunkel, E. Vafadarnejad, C.-J. Xu, A. Horne, M. Herbert, A. Drews, C. Thibeault, M. Pfeiffer, S. Hippenstiel, A. Hocke, H. Müller-Redetzky, K.-M. Heim, F. Machleidt, A. Uhrig, L. Bosquillon de Jarcy, L. Jürgens, M. Stegemann, C. R. Glösenkamp, H.-D. Volk, C. Goffinet, M. Landthaler, E. Wyler, P. Georg, M. Schneider, C. Dang-Heine, N. Neuwinger, K. Kappert, R. Tauber, V. Corman, J. Raabe, K. M. Kaiser, M. T. Vinh, G. Rieke, C. Meisel, T. Ulas, M. Becker, R. Geffers, M. Witzenrath, C. Drosten, N. Suttorp, C. von Kalle, F. Kurth, K. Händler, J. L. Schultze, A. C. Aschenbrenner, Y. Li, J. Nattermann, B. Sawitzki, A.-E. Saliba, L. E. Sander, Deutsche COVID-19 OMICS Initiative (DeCOI), Severe COVID-19 Is Marked by a Dysregulated Myeloid Cell Compartment. *Cell*. **182**, 1419–1440.e23 (2020).

10. A. Silvin, N. Chapuis, G. Dunsmore, A.-G. Goubet, A. Dubuisson, L. Derosa, C. Almire, C. Hénon, O. Kosmider, N. Droin, P. Rameau, C. Catelain, A. Alfaro, C. Dussiau, C. Friedrich, E. Sourdeau, N. Marin, T.-A. Szwebel, D. Cantin, L. Mouthon, D. Borderie, M. Deloger, D. Bredel, S. Mouraud, D. Drubay, M. Andrieu, A.-S. Lhonneur, V. Saada, A. Stoclin, C. Willekens, F. Pommeret, F. Griscelli, L. G. Ng, Z. Zhang, P. Bost, I. Amit, F. Barlesi, A. Marabelle, F. Pène, B. Gachot, F. André, L. Zitvogel, F. Ginhoux, M. Fontenay, E. Solary, Elevated Calprotectin and Abnormal Myeloid Cell Subsets Discriminate Severe from Mild COVID-19. *Cell*. **182** (2020), pp. 1401–1418.e18.

11. A. J. Wilk, A. Rustagi, N. Q. Zhao, J. Roque, G. J. Martínez-Colón, J. L. McKechnie, G. T. Ivison, T. Ranganath, R. Vergara, T. Hollis, L. J. Simpson, P. Grant, A. Subramanian, A. J. Rogers, C. A. Blish, A single-cell atlas of the peripheral immune response in patients with severe COVID-19. *Nature Medicine*. **26** (2020), pp. 1070–1076.

12. J. S. Lee, S. Park, H. W. Jeong, J. Y. Ahn, S. J. Choi, H. Lee, B. Choi, S. K. Nam, M. Sa, J.-S. Kwon, S. J. Jeong, H. K. Lee, S. H. Park, S.-H. Park, J. Y. Choi, S.-H. Kim, I. Jung, E.-C. Shin, Immunophenotyping of COVID-19 and influenza highlights the role of type I interferons in development of severe COVID-19. *Sci Immunol*. **5** (2020), doi:10.1126/sciimmunol.abd1554.

13. P. S. Arunachalam, F. Wimmers, C. K. P. Mok, R. A. P. M. Perera, M. Scott, T. Hagan, N. Sigal, Y. Feng, L. Bristow, O. Tak-Yin Tsang, D. Wagh, J. Coller, K. L. Pellegrini, D. Kazmin, G. Alaaeddine, W. S. Leung, J. M. C. Chan, T. S. H. Chik, C. Y. C. Choi, C. Huerta, M. Paine McCullough, H. Lv, E. Anderson, S. Edupuganti, A. A. Upadhyay, S. E. Bosinger, H. T. Maecker, P. Khatri, N. Rouphael, M. Peiris, B. Pulendran, Systems biological assessment of immunity to mild versus severe COVID-19 infection in humans. *Science*. **369**, 1210–1220 (2020).

14. C. Guo, B. Li, H. Ma, X. Wang, P. Cai, Q. Yu, L. Zhu, L. Jin, C. Jiang, J. Fang, Q. Liu, D. Zong, W. Zhang, Y. Lu, K. Li, X. Gao, B. Fu, L. Liu, X. Ma, J. Weng, H. Wei, T. Jin, J. Lin, K. Qu, Single-cell analysis of two severe COVID-19 patients reveals a monocyte-associated and tocilizumab-responding cytokine storm. *Nat. Commun.* **11**, 3924 (2020).

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

16. R. A. Grant, L. Morales-Nebreda, N. S. Markov, Alveolitis in severe SARS-CoV-2 pneumonia is driven by self-sustaining circuits between infected alveolar macrophages and T cells. *bioRxiv* (2020) (available at https://www.biorxiv.org/content/10.1101/2020.08.05.238188v1.abstract).

17. R. L. Chua, S. Lukassen, S. Trump, B. P. Hennig, D. Wendisch, F. Pott, O. Debnath, L. Thürmann, F. Kurth, M. T. Völker, J. Kazmierski, B. Timmermann, S. Twardziok, S. Schneider, F. Machleidt, H. Müller-Redetzky, M. Maier, A. Krannich, S. Schmidt, F. Balzer, J. Liebig, J. Loske, N. Suttorp, J. Eils, N. Ishaque, U. G. Liebert, C. von Kalle, A. Hocke, M. Witzenrath, C. Goffinet, C. Drosten, S. Laudi, I. Lehmann, C. Conrad, L.-E. Sander, R. Eils, COVID-19 severity correlates with airway epithelium–immune cell interactions identified by single-cell analysis. *Nat. Biotechnol.* **38**, 970–979 (2020).

18. A. C. Yang, F. Kern, P. M. Losada, C. A. Maat, G. Schmartz, Broad transcriptional dysregulation of brain and choroid plexus cell types with COVID-19. *bioRxiv* (2020) (available at https://www.biorxiv.org/content/10.1101/2020.10.22.349415v1.abstract).

19. M. Heming, X. Li, S. Räuber, A. K. Mausberg, A.-L. Börsch, M. Hartlehnert, A. Singhal, I.-N. Lu, M. Fleischer, F. Szepanowski, O. Witzke, T. Brenner, U. Dittmer, N. Yosef, C. Kleinschnitz, H. Wiendl, M. Stettner, G. Meyer Zu Hörste, Neurological Manifestations of COVID-19 Feature T Cell Exhaustion and Dedifferentiated Monocytes in Cerebrospinal Fluid. *Immunity*. **54**, 164–175.e6 (2021).

20. T. M. Delorey, C. G. K. Ziegler, G. Heimberg, R. Normand, Y. Yang, A. Segerstolpe, D. Abbondanza, S. J. Fleming, A. Subramanian, D. T. Montoro, K. A. Jagadeesh, K. K. Dey, P. Sen, M. Slyper, Y. H. Pita-Juárez, D. Phillips, Z. Bloom-Ackerman, N. Barkas, A. Ganna, J. Gomez, E. Normandin, P. Naderi, Y. V. Popov, S. S. Raju, S. Niezen, L. T.-Y. Tsai, K. J. Siddle, M. Sud, V. M. Tran, S. K. Vellarikkal, L. Amir-Zilberstein, D. S. Atri, J. Beechem, O. R. Brook, J. Chen, P. Divakar, P. Dorceus, J. M. Engreitz, A. Essene, D. M. Fitzgerald, R. Fropf, S. Gazal, J. Gould, J. Grzyb, T. Harvey, J. Hecht, T. Hether, J. Jane-Valbuena, M. Leney-Greene, H. Ma, C. McCabe, D. E. McLoughlin, E. M. Miller, C. Muus, M. Niemi, R. Padera, L. Pan, D. Pant, C. Pe'er, J. Pfiffner-

Borges, C. J. Pinto, J. Plaisted, J. Reeves, M. Ross, M. Rudy, E. H. Rueckert, M. Siciliano, A. Sturm, E. Todres, A. Waghray, S. Warren, S. Zhang, D. R. Zollinger, L. Cosimi, R. M. Gupta, N. Hacohen, W. Hide, A. L. Price, J. Rajagopal, P. R. Tata, S. Riedel, G. Szabo, T. L. Tickle, D. Hung, P. C. Sabeti, R. Novak, R. Rogers, D. E. Ingber, Z. Gordon Jiang, D. Juric, M. Babadi, S. L. Farhi, J. R. Stone, I. S. Vlachos, I. H. Solomon, O. Ashenberg, C. B. M. Porter, B. Li, A. K. Shalek, A.-C. Villani, O. Rozenblatt-Rosen, A. Regev, A single-cell and spatial atlas of autopsy tissues reveals pathology and cellular targets of SARS-CoV-2. *bioRxiv* (2021), doi:10.1101/2021.02.25.430130.

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

22. Y. Hao, S. Hao, E. Andersen-Nissen, W. M. Mauck, S. Zheng, A. Butler, M. J. Lee, A. J. Wilk, C. Darby, M. Zagar, P. Hoffman, M. Stoeckius, E. Papalexi, E. P. Mimitou, J. Jain, A. Srivastava, T. Stuart, L. B. Fleming, B. Yeung, A. J. Rogers, J. M. McElrath, C. A. Blish, R. Gottardo, P. Smibert, R. Satija, Integrated analysis of multimodal single-cell data. *Cold Spring Harbor Laboratory* (2020), p. 2020.10.12.335331.

23. V. Kaimal, E. E. Bardes, S. C. Tabar, A. G. Jegga, B. J. Aronow, ToppCluster: a multiple gene list feature analyzer for comparative enrichment clustering and network-based dissection of biological systems. *Nucleic Acids Res.* **38**, W96–102 (2010).

24. A. G. Laing, A. Lorenc, I. Del Molino Del Barrio, A. Das, M. Fish, L. Monin, M. Muñoz-Ruiz, D. R. McKenzie, T. S. Hayday, I. Francos-Quijorna, S. Kamdar, M. Joseph, D. Davies, R. Davis, A. Jennings, I. Zlatareva, P. Vantourout, Y. Wu, V. Sofra, F. Cano, M. Greco, E. Theodoridis, J. D. Freedman, S. Gee, J. N. E. Chan, S. Ryan, E. Bugallo-Blanco, P. Peterson, K. Kisand, L. Haljasmägi, L. Chadli, P. Moingeon, L. Martinez, B. Merrick, K. Bisnauthsing, K. Brooks, M. A. A. Ibrahim, J. Mason, F. Lopez Gomez, K. Babalola, S. Abdul-Jawad, J. Cason, C. Mant, J. Seow, C. Graham, K. J. Doores, F. Di Rosa, J. Edgeworth, M. Shankar-Hari, A. C. Hayday, Author Correction: A dynamic COVID-19 immune signature includes associations with poor prognosis. *Nat. Med.* **26**, 1951 (2020).

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

26. A. J. Wilk, M. J. Lee, B. Wei, B. Parks, R. Pi, G. J. Martínez-Colón, T. Ranganath, N. Q. Zhao, S. Taylor, W. Becker, S. C.-19 Biobank, D. Jimenez-Morales, A. L. Blomkalns, R. O'Hara, E. A. Ashley, K. C. Nadeau, S. Yang, S. Holmes, M. Rabinovitch, A. J. Rogers, W. J. Greenleaf, C. A. Blish, Multi-omic profiling reveals widespread dysregulation of innate immunity and hematopoiesis

in COVID-19. *Cold Spring Harbor Laboratory* (2020), p. 2020.12.18.423363.

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

28. S. De Biasi, D. Lo Tartaro, M. Meschiari, L. Gibellini, C. Bellinazzi, R. Borella, L. Fidanza, M. Mattioli, A. Paolini, L. Gozzi, D. Jaacoub, M. Faltoni, S. Volpi, J. Milić, M. Sita, M. Sarti, C. Pucillo, M. Girardis, G. Guaraldi, C. Mussini, A. Cossarizza, Expansion of plasmablasts and loss of memory B cells in peripheral blood from COVID-19 patients with pneumonia. *Eur. J. Immunol.* **50**, 1283–1294 (2020).

29. E. A. Middleton, X.-Y. He, F. Denorme, R. A. Campbell, D. Ng, S. P. Salvatore, M. Mostyka, A. Baxter-Stoltzfus, A. C. Borczuk, M. Loda, M. J. Cody, B. K. Manne, I. Portier, E. S. Harris, A. C. Petrey, E. J. Beswick, A. F. Caulin, A. Iovino, L. M. Abegglen, A. S. Weyrich, M. T. Rondina, M. Egeblad, J. D. Schiffman, C. C. Yost, Neutrophil extracellular traps contribute to immunothrombosis in COVID-19 acute respiratory distress syndrome. *Blood*. **136**, 1169–1179 (2020).

30. L. Nicolai, A. Leunig, S. Brambs, R. Kaiser, T. Weinberger, M. Weigand, M. Muenchhoff, J. C. Hellmuth, S. Ledderose, H. Schulz, C. Scherer, M. Rudelius, M. Zoller, D. Höchter, O. Keppler, D. Teupser, B. Zwißler, M. von Bergwelt-Baildon, S. Kääb, S. Massberg, K. Pekayvaz, K. Stark, Immunothrombotic Dysregulation in COVID-19 Pneumonia Is Associated With Respiratory Failure and Coagulopathy. *Circulation*. **142**, 1176–1189 (2020).

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

32. D. McGonagle, K. Sharif, A. O'Regan, C. Bridgewood, The Role of Cytokines including Interleukin-6 in COVID-19 induced Pneumonia and Macrophage Activation Syndrome-Like Disease. *Autoimmun. Rev.* **19**, 102537 (2020).

33. B. J. Barnes, J. M. Adrover, A. Baxter-Stoltzfus, A. Borczuk, J. Cools-Lartigue, J. M. Crawford, J. Daßler-Plenker, P. Guerci, C. Huynh, J. S. Knight, M. Loda, M. R. Looney, F. McAllister, R. Rayes, S. Renaud, S. Rousseau, S. Salvatore, R. E. Schwartz, J. D. Spicer, C. C. Yost, A. Weber, Y. Zuo, M. Egeblad, Targeting potential drivers of COVID-19: Neutrophil extracellular traps. *J. Exp. Med.* **217** (2020), doi:10.1084/jem.20200652.

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

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

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

37. M. Aid, K. Busman-Sahay, S. J. Vidal, Z. Maliga, S. Bondoc, C. Starke, M. Terry, C. A. Jacobson, L. Wrijil, S. Ducat, O. R. Brook, A. D. Miller, M. Porto, K. L. Pellegrini, M. Pino, T. N. Hoang, A. Chandrashekar, S. Patel, K. Stephenson, S. E. Bosinger, H. Andersen, M. G. Lewis, J. L. Hecht, P. K. Sorger, A. J. Martinot, J. D. Estes, D. H. Barouch, Vascular Disease and Thrombosis in SARS-CoV-2-Infected Rhesus Macaques. *Cell*. **183**, 1354–1366.e13 (2020).

38. N. H. Schultz, I. H. Sørvoll, A. E. Michelsen, L. A. Munthe, F. Lund-Johansen, M. T. Ahlen, M. Wiedmann, A.-H. Aamodt, T. H. Skattør, G. E. Tjønnfjord, P. A. Holme, Thrombosis and Thrombocytopenia after ChAdOx1 nCoV-19 Vaccination. *N. Engl. J. Med.* (2021), doi:10.1056/NEJMoa2104882.

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

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

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

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

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

44. J. W. M. Heemskerk, N. J. A. Mattheij, J M E, Platelet-based coagulation: different populations, different functions. *Journal of Thrombosis and Haemostasis*. **11** (2013), pp. 2–16.

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

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

47. C. N. Gruber, R. S. Patel, R. Trachtman, L. Lepow, F. Amanat, F. Krammer, K. M. Wilson, K. Onel, D. Geanon, K. Tuballes, M. Patel, K. Mouskas, T. O'Donnell, E. Merritt, N. W. Simons, V. Barcessat, D. M. Del Valle, S. Udondem, G. Kang, S. Gangadharan, G. Ofori-Amanfo, U. Laserson, A. Rahman, S. Kim-Schulze, A. W. Charney, S. Gnjatic, B. D. Gelb, M. Merad, D. Bogunovic, Mapping Systemic Inflammation and Antibody Responses in Multisystem Inflammatory Syndrome in Children (MIS-C). *Cell*. **183**, 982–995.e14 (2020).

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

49. Y. Zhou, T. Han, J. Chen, C. Hou, L. Hua, S. He, Y. Guo, S. Zhang, Y. Wang, J. Yuan, C. Zhao, J. Zhang, Q. Jia, X. Zuo, J. Li, L. Wang, Q. Cao, E. Jia, Clinical and autoimmune characteristics of severe and critical cases of COVID-19. *Clin. Transl. Sci.* **13**, 1077–1086 (2020).

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

51. S. Jin, C. F. Guerrero-Juarez, L. Zhang, I. Chang, P. Myung, M. V. Plikus, Q. Nie, Inference and analysis of cell-cell communication using CellChat. *Cold Spring Harbor Laboratory* (2020), p. 2020.07.21.214387.

52. G. Aust, D. Sittig, L. Becherer, U. Anderegg, A. Schütz, P. Lamesch, E. Schmücking, 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 (2004).

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

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

55. T. Kuwabara, F. Ishikawa, T. Yasuda, K. Aritomi, H. Nakano, Y. Tanaka, Y. Okada, M. Lipp, T. Kakiuchi, CCR 7 Ligands Are Required for Development of Experimental Autoimmune Encephalomyelitis through Generating IL-23-Dependent Th17 Cells. *The Journal of Immunology*. **183**, 2513–2521 (2009).

56. K. Hirota, H. Yoshitomi, M. Hashimoto, S. Maeda, S. Teradaira, N. Sugimoto, T. Yamaguchi, T. Nomura, H. Ito, T. Nakamura, N. Sakaguchi, S. Sakaguchi, 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 (2007).

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

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

59. F. Zhang, K. Wei, K. Slowikowski, C. Y. Fonseka, D. A. Rao, S. Kelly, S. M. Goodman, D. Tabechian, L. B. Hughes, K. Salomon-Escoto, G. F. M. Watts, A. H. Jonsson, J. Rangel-Moreno, N. Meednu, C. Rozo, W. Apruzzese, T. M. Eisenhaure, D. J. Lieb, D. L. Boyle, A. M. Mandelin 2nd, Accelerating Medicines Partnership Rheumatoid Arthritis and Systemic Lupus Erythematosus (AMP RA/SLE) Consortium, B. F. Boyce, E. DiCarlo, E. M. Gravallese, P. K. Gregersen, L. Moreland, G. S. Firestein, N. Hacohen, C. Nusbaum, J. A. Lederer, H. Perlman, C. Pitzalis, A. Filer, V. M. Holers, V. P. Bykerk, L. T. Donlin, J. H. Anolik, M. B. Brenner, S. Raychaudhuri, Defining inflammatory cell states in rheumatoid arthritis joint synovial tissues by integrating single-cell transcriptomics and mass cytometry. *Nat. Immunol.* **20**, 928–942 (2019).

60. A. Arazi, D. A. Rao, C. C. Berthier, A. Davidson, Y. Liu, P. J. Hoover, A. Chicoine, T. M. Eisenhaure, A. H. Jonsson, S. Li, D. J. Lieb, F. Zhang, K. Slowikowski, E. P. Browne, A. Noma, D. Sutherby, S. Steelman, D. E. Smilek, P. Tosta, W. Apruzzese, E. Massarotti, M. Dall'Era, M. Park, D. L. Kamen, R. A. Furie, F. Payan-Schober, W. F. Pendergraft 3rd, E. A. McInnis, J. P. Buyon, M. A. Petri, C. Putterman, K. C. Kalunian, E. S. Woodle, J. A. Lederer, D. A. Hildeman, C. Nusbaum, S. Raychaudhuri, M. Kretzler, J. H. Anolik, M. B. Brenner, D. Wofsy, N. Hacohen, B. Diamond, Accelerating Medicines Partnership in SLE network, The immune cell landscape in kidneys of patients with lupus nephritis. *Nat. Immunol.* **20**, 902–914 (2019).

61. M. Reyes, M. R. Filbin, R. P. Bhattacharyya, K. Billman, T. Eisenhaure, D. T. Hung, B. D. Levy, R. M. Baron, P. C. Blainey, M. B. Goldberg, N. Hacohen, An immune-cell signature of bacterial

sepsis. *Nat. Med.* **26**, 333–340 (2020).

62. D. Schafflick, C. A. Xu, M. Hartlehnert, M. Cole, A. Schulte-Mecklenbeck, T. Lautwein, J. Wolbert, M. Heming, S. G. Meuth, T. Kuhlmann, C. C. Gross, H. Wiendl, N. Yosef, G. M. zu Horste, Integrated single cell analysis of blood and cerebrospinal fluid leukocytes in multiple sclerosis. *Nature Communications*. **11** (2020), , doi:10.1038/s41467-019-14118-w.

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

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

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

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

67. C. Thålin, Y. Hisada, S. Lundström, N. Mackman, H. Wallén, Neutrophil Extracellular Traps: Villains and Targets in Arterial, Venous, and Cancer-Associated Thrombosis. *Arterioscler. Thromb. Vasc. Biol.* **39**, 1724–1738 (2019).

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

69. S. Spoerl, A. N. Kremer, M. Aigner, N. Eisenhauer, P. Koch, L. Meretuk, P. Löffler, M. Tenbusch, C. Maier, K. Überla, L. Heinzerling, B. Frey, G. Lutzny-Geier, T. H. Winkler, G. Krönke, M. Vetter, H. Bruns, M. F. Neurath, A. Mackensen, A. E. Kremer, S. Völkl, Upregulation of CCR4 in activated CD8+ T cells indicates enhanced lung homing in patients with severe acute SARS-CoV-2 infection. *Eur. J. Immunol.* (2021), doi:10.1002/eji.202049135.

70. A. Bharat, M. Querrey, N. S. Markov, S. Kim, C. Kurihara, R. Garza-Castillon, A. Manerikar, A. Shilatifard, R. Tomic, Y. Politanska, H. Abdala-Valencia, A. V. Yeldandi, J. W. Lomasney, A. V. Misharin, G. R. S. Budinger, Lung transplantation for patients with severe COVID-19. *Sci. Transl. Med.* **12** (2020), doi:10.1126/scitranslmed.abe4282.

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

72. K. Blighe, S. Rana, M. Lewis, EnhancedVolcano: Publication-Ready Volcano Plots With Enhanced Colouring and Labeling.(2019). *R Package Version*. **1** (2018).

Acknowledgments

We thank Pablo Garcia-Nieto, Ambrose Carr and Jonah Cool and the Chan Zuckerberg Initiative for hosting the data on cellxgene. We acknowledge suggestions and help from Greta Beekhuis. We thank the support from Pediatric Cell Atlas and high performance computational cluster of CCHMC. Some figures were created using https://biorender.com. **Funding**: Funding for this study was provided by LungMap (U24 and HL148865), Digestive Health Center (P30, DK078392) and Harold C. Schott Foundation funding of the Harold C. Schott Endowed Chair, UC College of Medicine. **Author contributions:** Conceptualization, K.J. and B.A.; Methodology, K.J., B.A. and E.B.; 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.; Supervision, B.A.. **Competing interests:** Authors declare that they have no competing interests.
Data and materials availability: Public single-cell RNA-seq datasets of PBMC in COVID-19 patients are available on NCBI Gene Expression Omnibus and European Genome-phenome Archive, including GSE150728, GSE155673, GSE150861, GSE149689 and EGAS00001004571 (or 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. Gene modules of all datasets analyzed using ToppCell web portal are available on COVID-19 Atlas in ToppCell, 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 Fig. 1A and table S1. An interactive interface of integrated PBMC data and subclusters of immune cells will be public on cellxgene upon publication. Codes of preprocessing, normalization, clustering and plotting of single-cell datasets will be available on github upon publication.

Figures and Tables

Fig. 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 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).

Fig. 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. *: $p <= 0.05$; **: $p <= 0.01$; ***: $p <= 0.001$; ****: $p <= 0.0001$. (**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. (**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.

Fig. 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. (**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 fig. S5D and fig. 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 were drawn based on fig. S8. Interaction was inferred using both CellChat database and embedded cell interaction database in ToppCell.

Fig. 4. COVID-19 driven reprogramming of platelets leads to drastically altered expression of genes associated with platelet adhesion, activation, coagulation and thrombosis. (**A-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. (**D**) Functional and phenotypical associations of coagulation-association genes in each gene pattern from (B). Associations were retrieved from ToppGene enrichment. Fibrinolysis is highlighted.

Fig. 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. (**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.

Fig. 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 hierarchically clustered. Hot spots of the disease-specific enrichments were highlighted and details were shown on the left. More details can be found in 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.

Fig. 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

98 controls. Log2(ratio) was calculated to show the levels of changes. *, $p<0.05$, **, $p<0.01$, ***,
99 $p<0.001$. Statistical models can be found in the Methods. Leuk-UTI: sepsis patients that enrolled
100 into UTI with leukocytosis (blood WBC ≥ 12,000 per mm3) but no organ dysfunction. (**C**)
101 Normalized expression values of key genes involved in immune signaling and responses are
102 shown for cell types across multiple diseases. Lowly expressed genes (maximal average
103 expression level across all cell types in the heatmap is less than 0.5 after Log_2CPM normalization)
104 were removed.
105
106 **Figure S1. Cell distribution and abundance in the integrated COVID-19 PBMC data.** (**A**)
107 Distributions of COVID-19 conditions (Left) and data sources (Right) for the integrated PBMC
108 data are shown on the same UMAP of Figure 2A. (**B**) Bar plot depicts distributions of disease
109 conditions in 5 individual PBMC single-cell datasets. Percentages of 3 disease conditions in each
110 dataset is shown on y axis. (**C**) The integrated bar plot shows percentages of 3 disease conditions
111 in each cell type per dataset. Dataset abbreviations and cell types were concatenated to show
112 disease distributions of specific cell types in the selected datasets. These labels are colored by
113 their cell type designations and ordered by the ascending percentages of COVID-19 conditions.
114
115 **Figure S2. Dynamic changes of cell type abundances in five COVID-19 PBMC datasets.**
116 Relative abundances and differences of major cell types in each single cell dataset are shown and
117 compared to controls per each disease condition, per each single-cell dataset. Box plots of all cell
118 types in PBMC are shown except for the 5 highlighted cell types shown in Figure 2B. Statistical
119 methods are the same with Figure 2B.
120
121 **Figure S3. Cell distributions and dynamic changes in the integrated COVID-19 BAL data.**
122 (**A-C**) Distributions of disease conditions (A), data sources (B) and samples (C) are shown on the
123 same UMAP of Figure 2C. (**D**) Box plots depict dynamic changes of cell types across COVID-19
124 conditions in BAL that are not covered in Figure 2D. Statistical methods are the same with Figure
125 2B.
126
127 **Figure S4. Cell type abundance changes in COVID-19 lung parenchyma dataset.** Box plots
128 depict percentages of cell types in control samples and severe COVID-19 samples. We used cell
129 type clusters identified in the original publication but modified cell naming of macrophage

subtypes to distinguish monocyte derived macrophage subtypes present in BAL fluid samples. Statistical methods are the same with Figure 2B.

Figure S5. Sub-cluster-specific genes of neutrophils of COVID-19 patients. (**A**) Distribution of disease conditions (Left) and data sources (Right) for the integrated neutrophil data on the same UMAP of Figure 3A. (**B**) UMAPs of neutrophil sub-cluster-associated genes from Figure 3C. Normalized expression values for each gene were used. (**C**) Normalized expression values of neutrophil-associated genes and other important immune signatures are shown for 5 neutrophil sub-clusters. Lowly expressed genes (genes with maximal average expression level across all neutrophil sub-clusters less than 0.5 after Log_2CPM normalization) were removed from the gene pool of cytokines, chemokines, ISGs, interleukins, interferons, corresponding receptors and MHC-II. (**D**) The volcano plot depicts differentially expressed genes between circulating mature neutrophils (Neu0,1) and extravasated neutrophils (Neu3) (Left); as well as DEGs between pro-neutrophils (Neu4) and pre-neutrophils (Neu2) (Right). Statistical methods are the same with Figure 5C. Representative enriched biological processes (Gene Ontology) are shown in the bottom.

Figure S6. Macrophage-related signatures in the integrated BAL data. (**A**) Normalized expression values of myeloid-cell-associated genes and other important immune signatures are shown for 9 macrophage sub-clusters. Lowly expressed genes (genes with maximal average expression level across all macrophage sub-clusters less than 0.5 after Log_2CPM normalization) were removed from the gene pool of MHC-II, cytokines, chemokines, ISGs, interleukins, interferons and their receptors. (**B**) Volcano plots were drawn for DEGs of MoAM3,4 versus MoAM1,2,5 (Left) and TRAM 1,2 versus TRAM3 (Middle) and TRAM3 versus MoAM1,2,5 (Right). Statistical methods are the same with Figure 5C. (**C**) Normalized expression values were shown on the same UMAP of Figure 3B for important genes including macrophage signatures, ISGs, interferons, receptors and MHC-II.

Figure S7. A uniquely-activated monocyte-derived cell type (MoAM5) exhibits a broad signature of cytokines, chemokines, and interleukins including IL6. (**A**) Normalized expression values of IL6 on the same reference UMAP of integrated BAL data as Figure 3B. (**B**) Scale expression levels of IL6 for each macrophage sub-cluster on the violin plot. (**C**) Heatmap of expression levels of pan-MoAM signatures and MoAM5-specifc signatures in all myeloid cells in

both PBMC and BAL. (**D**) Network of functional and phenotypic associated pan-MoAM signatures and MoAM5-specific signatures from (C). Associations were retrieved from ToppGene enrichment results. IL6 is highlighted in the network. As a caveat, the MoAM5 subtype represented a small fraction among the BAL MoAM subtypes and the majority of these cells were observed in a single severely-affected individual.

Figure S8. Cell type and cell subtype-specific divisions of cytokine, chemokine, and interleukin signaling pathways in BAL of severe COVID-19 patients. (**A**) Heatmap of expression patterns of ligands and receptors in cytokine, chemokine, interleukin, CSF and TNFSF signaling pathways across cell types of BAL in severe patients. Average normalized expression values were shown and lowly expressed ligands or receptors (maximal normalized expression value for a row in the heatmap < 0.5) were removed. To reduce bias, MoAM5 was removed because cells in the cluster were mainly from one patient. Cell types that have less than 5% cells from severe patients were removed, including TRAM1 and TRAM2. Neutrophils are highlighted in the heatmap. (**B**) Interaction network of BAL cells in severe patients using CellChat. CCL, CXCL and IL1 signaling pathways were shown. The width of edges represents the strength of interactions and the size of nodes represents the abundance of cell types.

Figure S9. Characteristics of sub-clusters of classical monocytes in the integrated COVID-19 PBMC data. (**A**) UMAPs of 4 sub-clusters (Left) and COVID-19 conditions (Right) of classical monocytes are shown. Grey dots are other myeloid cells in the UMAP of integrated PBMC myeloid data. (**B**) UMAPs of normalized expression values of specific signatures for classical monocyte sub-clusters. (**C**) Normalized expression values of monocyte-associated genes and other important immune signatures are shown for 4 classical monocyte sub-clusters. (**D**) Gene modules of classical monocyte sub-clusters, as well as other myeloid cell types in the integrated PBMC myeloid data. Representative genes in each module are shown on the left. ToppGene enrichment results for classical monocyte sub-clusters are shown on the right. Columns are clustered using hierarchical clustering. (**E**) Similarity matrix of myeloid cell types using genes in (D). Pearson correlation was used to evaluate similarity. (**F**) Dot plot of MHC-II, ISGs, interleukin genes and cell cycle genes for each myeloid cell type. Scale values were used.

Figure S10. Features of conventional dendritic cell sub-clusters and polarized signaling genes. (**A**) UMAPs of 13 sub-clusters (Left) and sources (Right) of conventional dendritic cells

after data integration. (**B**) Normalized expression values of sub-cluster-specific genes on the UMAP. (**C**) Normalized expression values of cDC-associated genes and other important immune signatures are shown for 13 cDC sub-clusters. (**D**) Gene modules of cDC sub-clusters with 200 most significantly upregulated genes in each module. Representative genes are shown on the left. Gene enrichment results of some modules from ToppGene are shown on the right. (**E**) Similarity matrix of sub-clusters using genes in (D). Pearson correlation was used for similarity scores and hierarchical clustering was applied for rows and columns. (**F**) The heatmap shows the clustering of signaling genes, including cytokines, chemokines, interleukins and their receptors. Red boxes highlight severe patients associated sub-clusters and their upregulated genes. Green boxes highlight mild patients-associated sub-clusters and their upregulated genes.

Figure S11. Landscape of myeloid cells in the integrated PBMC and BAL data. (**A-B**) UMAPs of myeloid cells in integrated PBMC (A) and BAL (B) data. Cell types which were further clustered are highlighted in different colors. (**C**) The heatmap shows associations between subclusters of myeloid cells and myeloid-cell-associated pathways, such as antigen presenting, T cell activation, phagocytosis etc. 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 sub-cluster.

Figure S12. Gene expression signatures of cell types and subtypes activated by COVID-19 are extensively associated with coagulation, hemostasis, and thrombosis-associated pathways, functions, and knockout phenotypes. (**A**) Functional association heatmap of gene signatures from COVID-19 cell types demonstrates differential enrichment for pathways associated with coagulation, vascular permeability, complement, extravasation, platelet activation and aggregation, response to wounding, as shown. Gene modules of cell types and sub-clusters that participate in these pathways were used to calculate enrichment scores. (**B**) Network of upregulated genes in coagulation/thrombosis-associated pathways (A) shows the potential gene-gene interactions in immunothrombosis of COVID-19 patients. CellChat and ToppCell/ToppGene protein-protein ligand receptor and cell adhesion interaction databases were used to find interaction pairs among upregulated genes. (**C**) A new network derived from (B) shows integrin-associated interactions between platelets and other cells.

Figure S13. Emergence of platelet subtypes suggestive of functionally significant alternative roles in in hemostasis, coagulation, wound response, and neutrophil recruitment and activation. (**A**) The heatmap shows ToppCell gene modules of 6 platelet sub-clusters in COVID-19 PBMC. Each gene module contains 200 most significant genes for each sub-cluster and important genes are shown on the left. Gene enrichment analysis was conducted using ToppGene and top enrichment results from biological processes (Gene Ontology) are shown on the right. (**B**) Dot plot of integrin and other platelet-associated genes. Scale values are shown on the figure. (**C**) Heatmap of associations between subclusters of platelets and platelet-associated pathways (Gene Ontology). Gene enrichment scores, defined as $-\log_{10}$(adjusted p value), were calculated and shown.

Figure S14. Consistent emergence of a series of early and maturing B cells and plasmablasts in BAL fluid and PBMC across multiple datasets. (**A-B**) UMAPs of B cells (A) and plasmablasts (B) from multiple datasets. (**C-D**) UMAP of normalized expression values of immunoglobulin genes (C) and ISGs (D) for B cells. (**E-F**) UMAP of normalized expression values of immunoglobulin genes (E) and sub-cluster associated genes, such as cell cycle genes and B cell markers (F) for plasmablasts. (**G**) Gene modules of B cell sub-clusters and plasmablast subtypes with 200 most significant genes in each module. Hierarchical clustering was applied for columns. (**H**) Three representative enriched biological processes (Gene Ontology) are shown for these two subtypes using DEGs of plasmablasts in Figure 5C.

Figure S15. Gene Enrichment analysis of B cell subtypes and autoimmune-associated signatures. (**A**) Heatmap shows gene enrichment scores of B-cell-associated pathways for each B cell sub-cluster and plasmablast subtype. (**B**) Pathway and function association network of upregulated genes in B cells of BAL in mild COVID-19 patients. (**C-D**) Heatmaps show normalized expression levels of autoimmune-associated ligands and receptors (Figure 5E) in lupus nephritis (C) and rheumatoid arthritis (D).

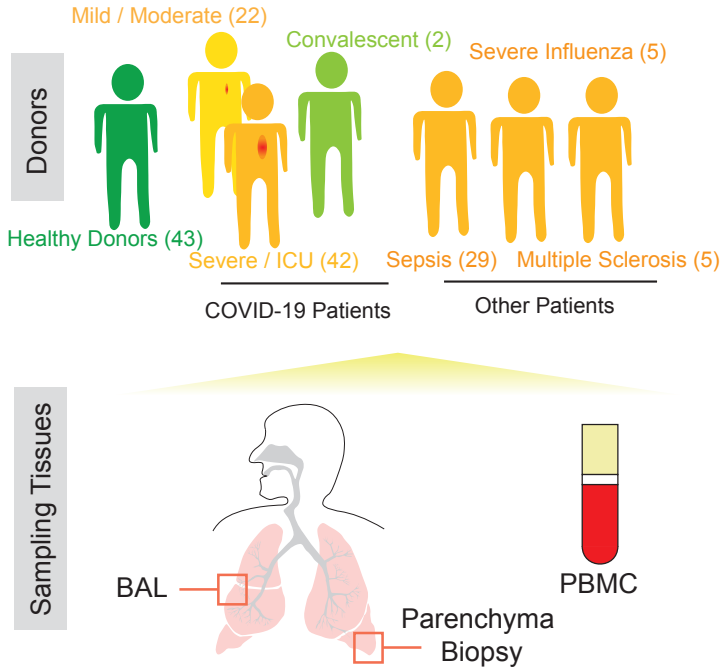
Figure S16. Distinct subtypes of T cells and NK cells in COVID-19 BAL data. (**A-C**) UMAPs of subtypes (A), COVID-19 conditions (B) and data sources (C) of T cells and NK cells in the integrated BAL data. (D-E) UMAPs of normalized expression values of exhausted T cell markers (**D**) and ISGs (**E**).

Figure S17. Various T cell and NK cell subtypes in the integrated PBMC data. (**A-B**) UMAPs of T cell and NK cell subtypes (A) and COVID-19 conditions (B) after integration of T cells in 5 PBMC single-cell datasets. (**C**) Dot plot shows T cell and NK cell subtype associated genes for each subtype per disease condition. Labels of cell types of healthy donors, mild patients and severe patients are colored by blue, yellow and red. Scaled expression values are shown using a color scheme.

Figure S18. Various cell types in immune-mediated diseases. (**A, C, E**) Distributions of cell types identified in influenza (A), sepsis (C) and multiple sclerosis (E) patients were shown on UMAPs. (**B, D, F**) Distributions of disease conditions in influenza (B), sepsis (D) and multiple sclerosis (F) patients were shown on UMAPs.

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

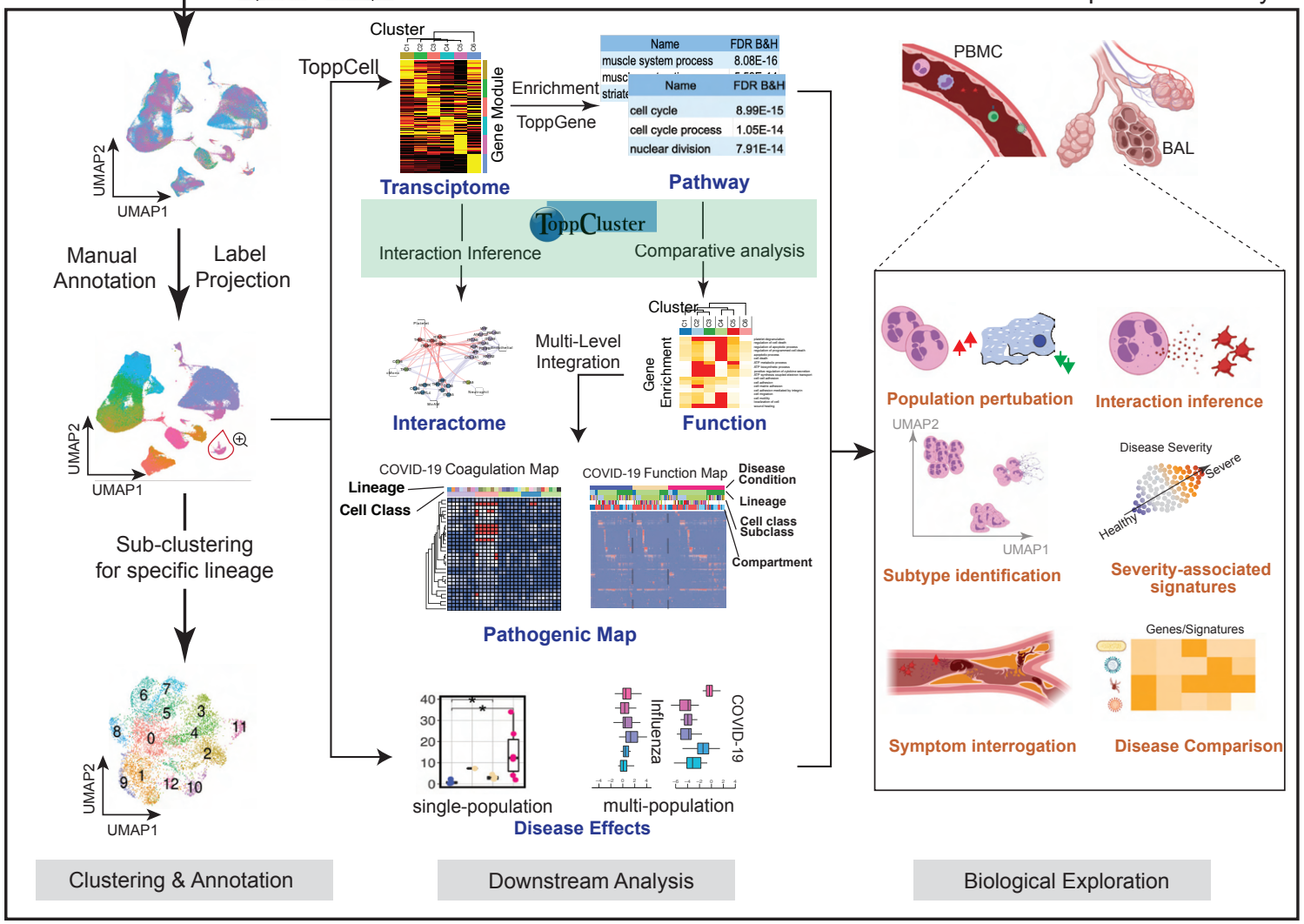


Fig. 1. Creating a COVID-19 Signature Atlas

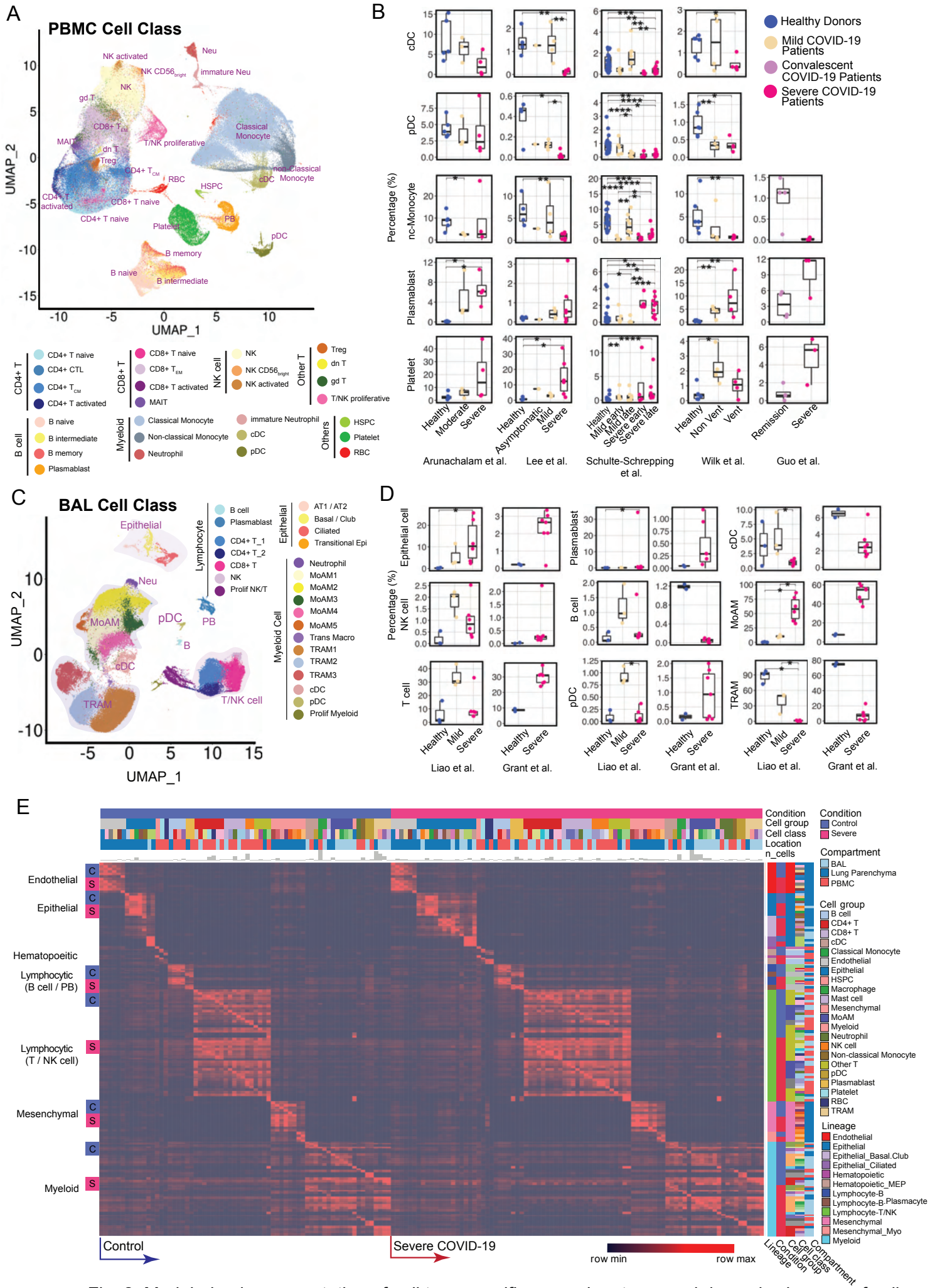


Fig. 2. Modularized representation of cell type specific gene signatures and dynamic changes of cell

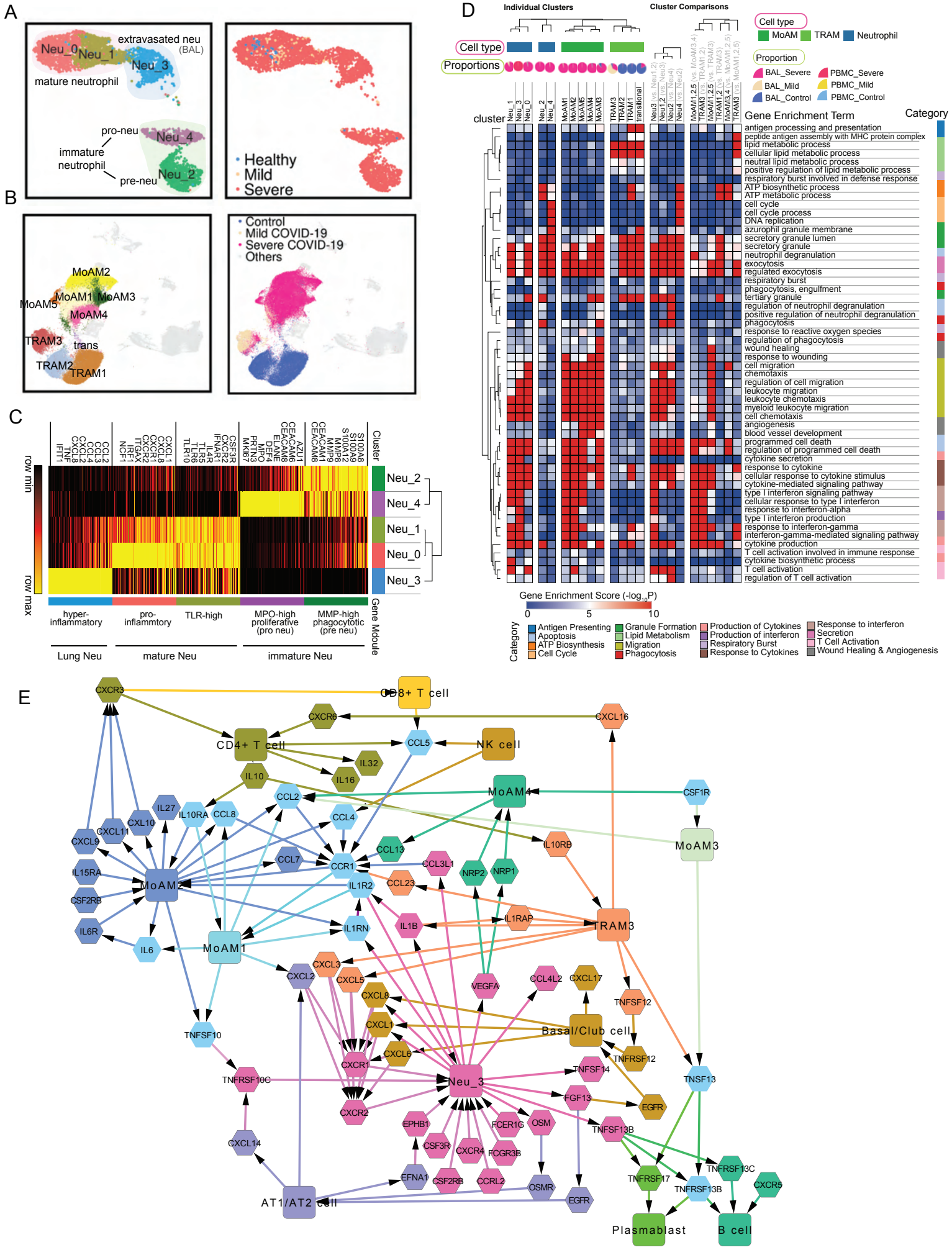


Fig. 3. Functional analysis of compartment-specific immature and subtype-differentiated neutrophils and monocytic macrophages in COVID-19 patients

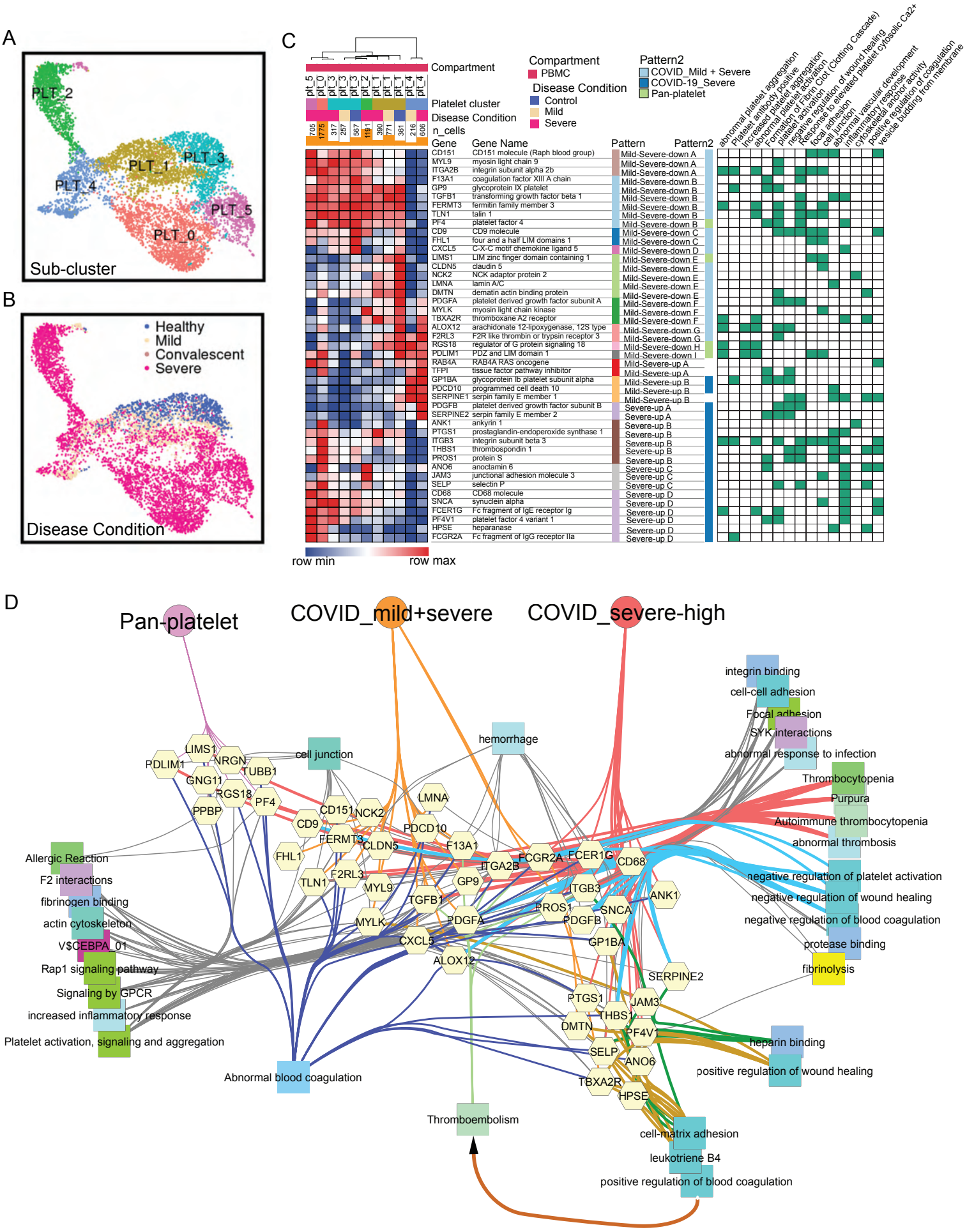


Fig. 4. COVID-19 driven reprogramming of platelets leads to drastically altered expression of genes associated with platelet adhesion, activation, coagulation and thrombosis

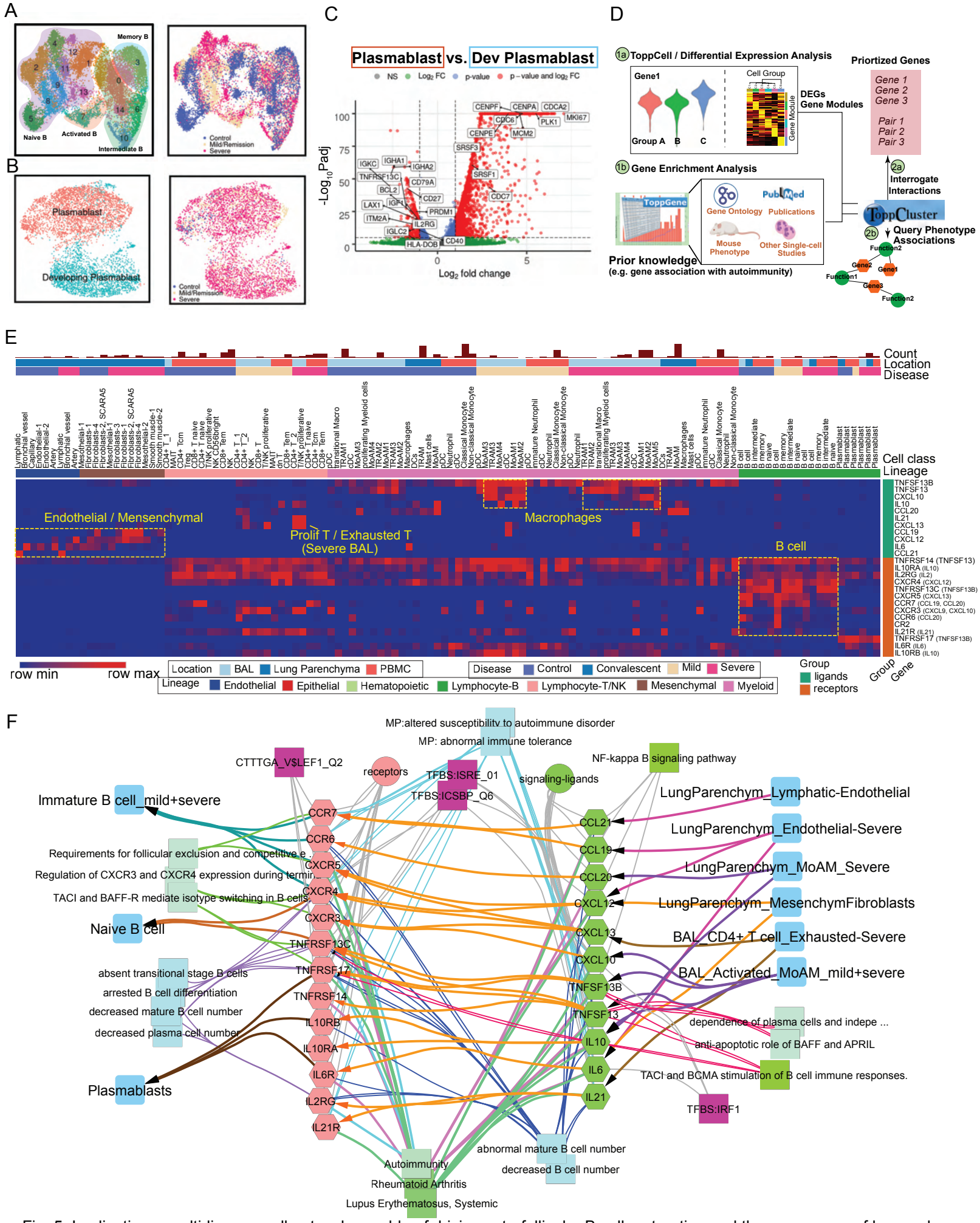
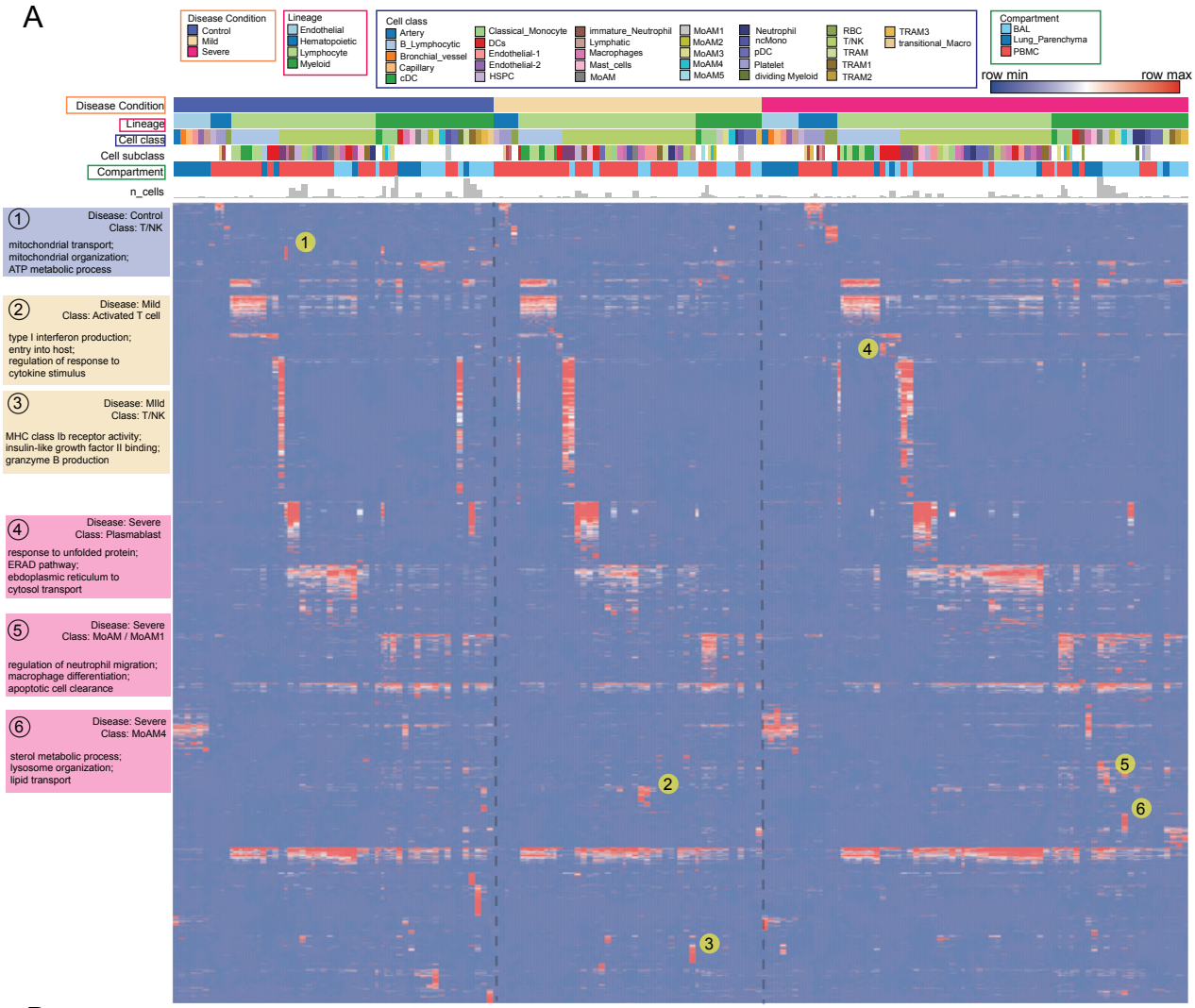


Fig. 5. Implicating a multi-lineage cell network capable of driving extrafollicular B cell maturation and the emergence of humoral autoimmunity in COVID-19 patients.



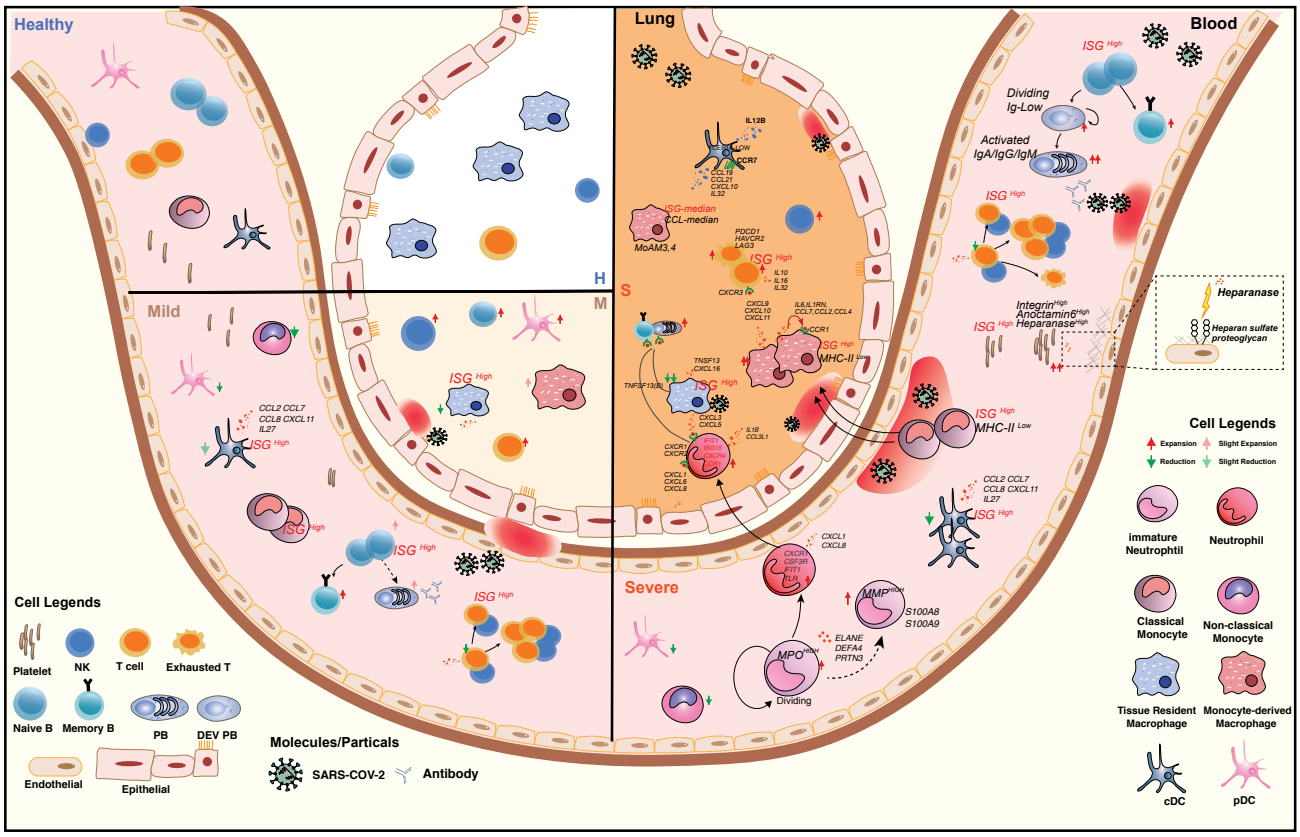


Fig. 6. Comparative analysis of cell type specific gene signatures associated with lineage, class, subclass, compartment, and disease state in the COVID-19 atlas

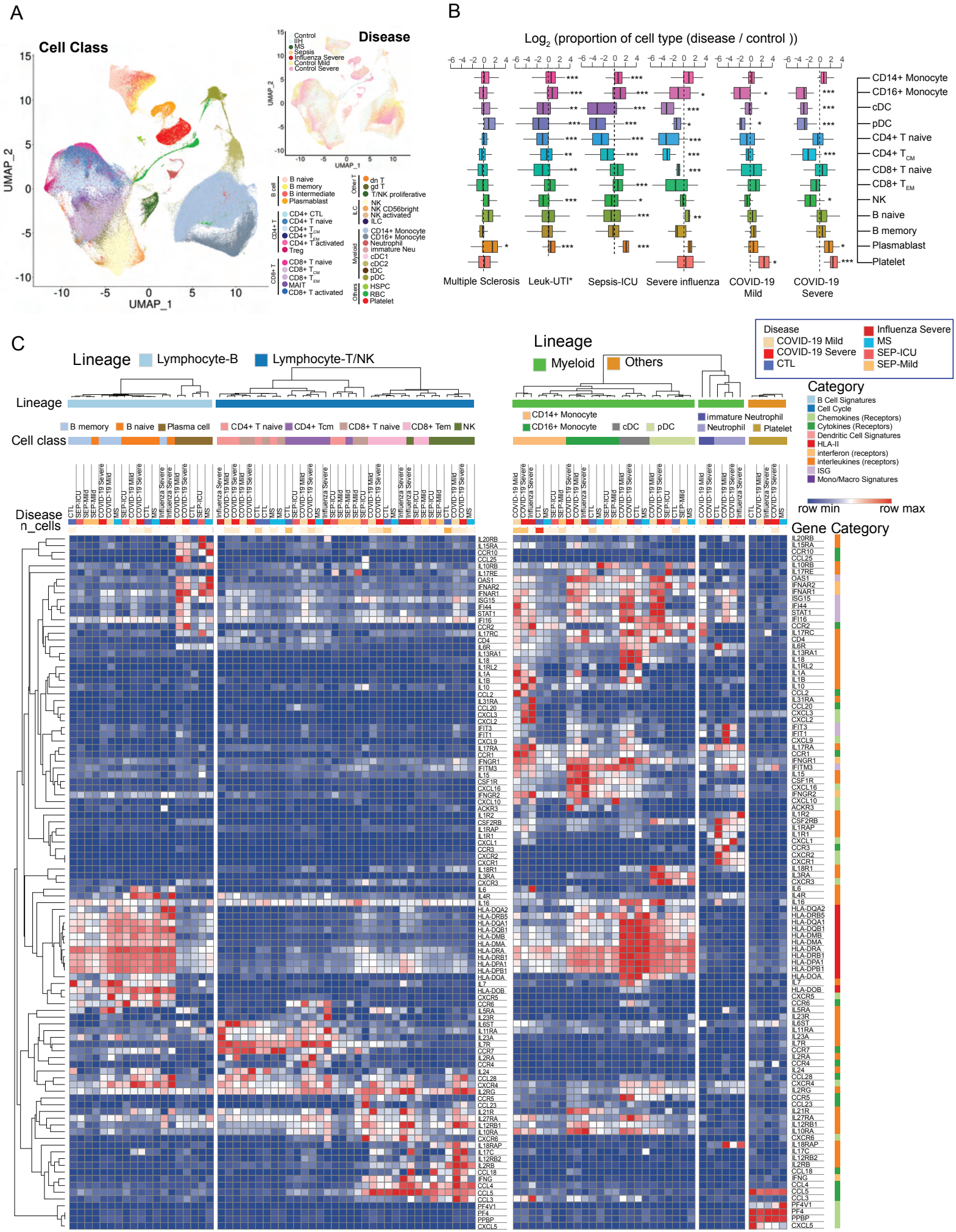


Fig. 7. Comparative analysis of differentially-expressed immunoregulatory genes between COVID-19 and other immune-mediated diseases.

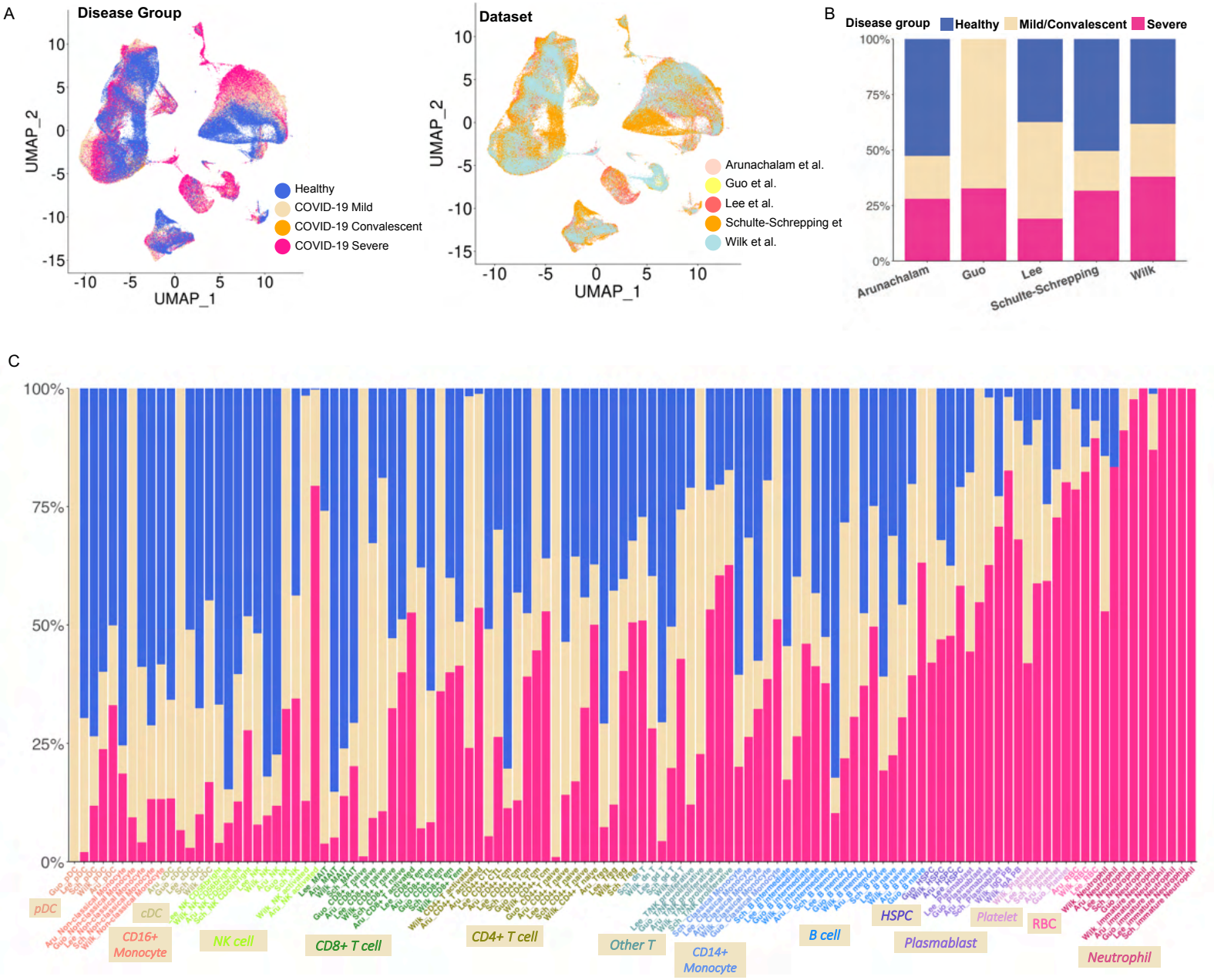


Figure S1. Cell distribution and abundance in the integrated COVID-19 PBMC data, relative to Figure 2.

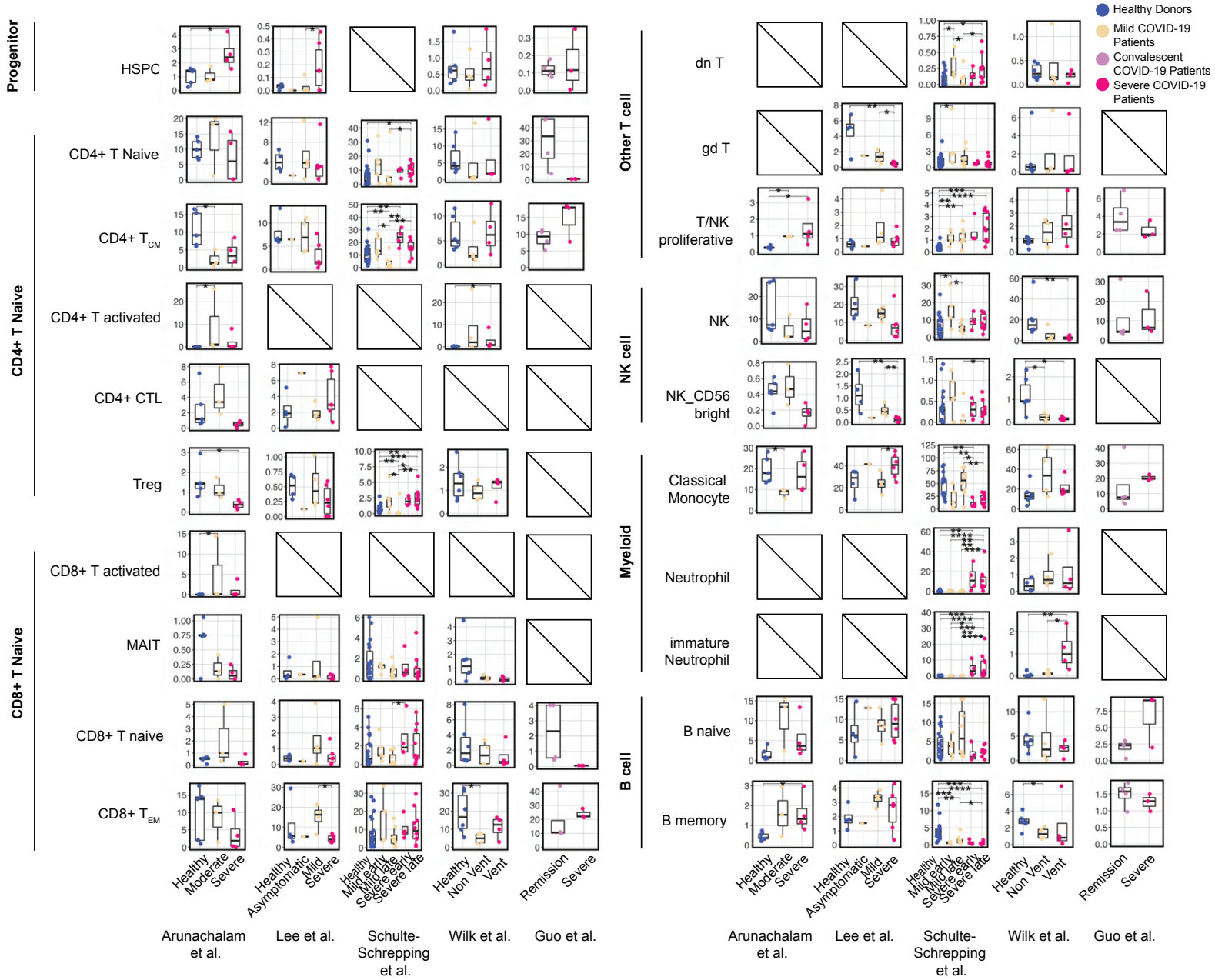


Figure S2. Dynamic changes of cell type abundance in the integrated PBMC data, relative to Figure 2.

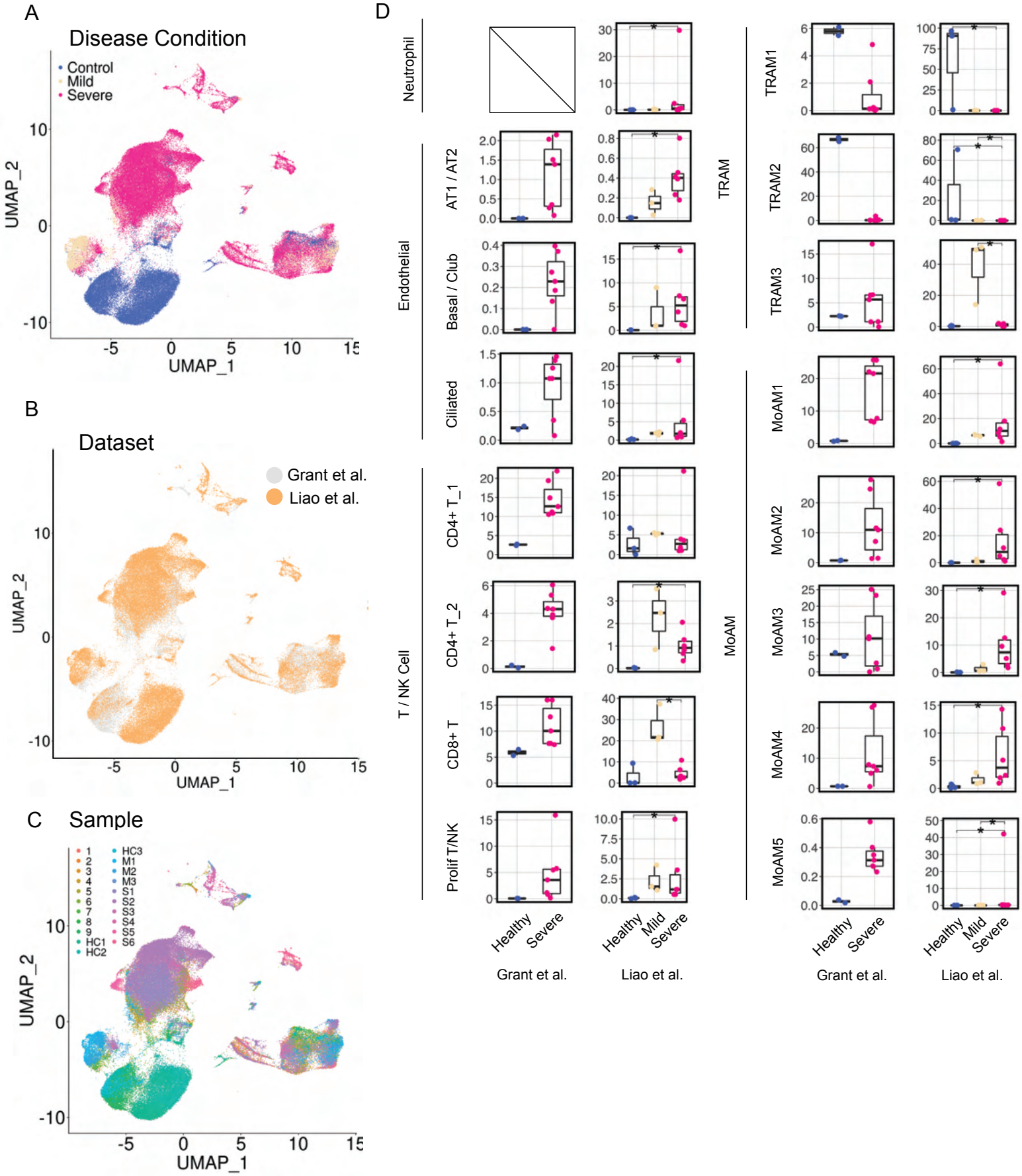


Figure S3. Cell distributions and dynamic changes in the integrated COVID-19 BAL data, relative to Figure 2.

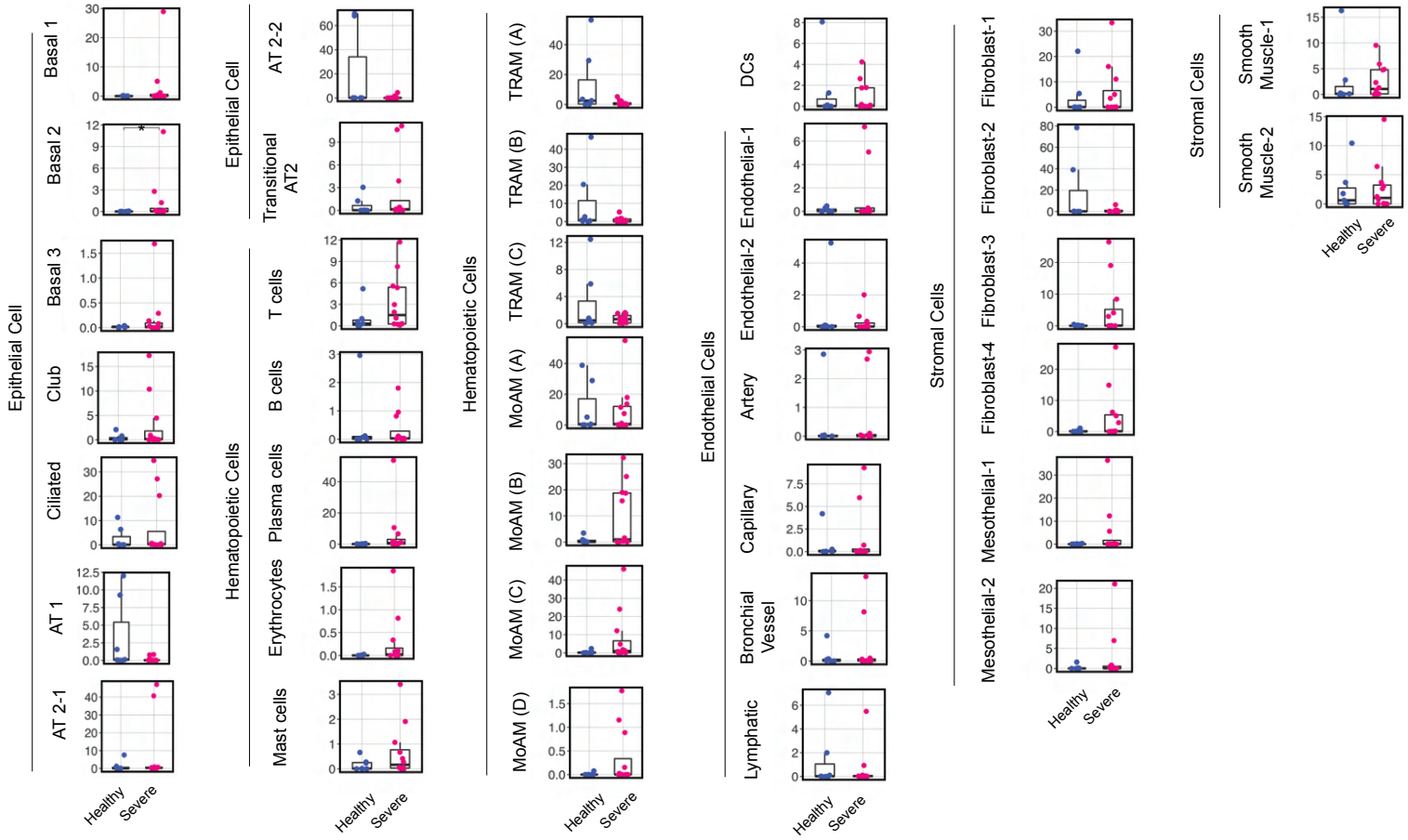


Figure S4. Dynamic changes of cell types in the COVID-19 lung parenchyma dataset, relative to Figure 2.

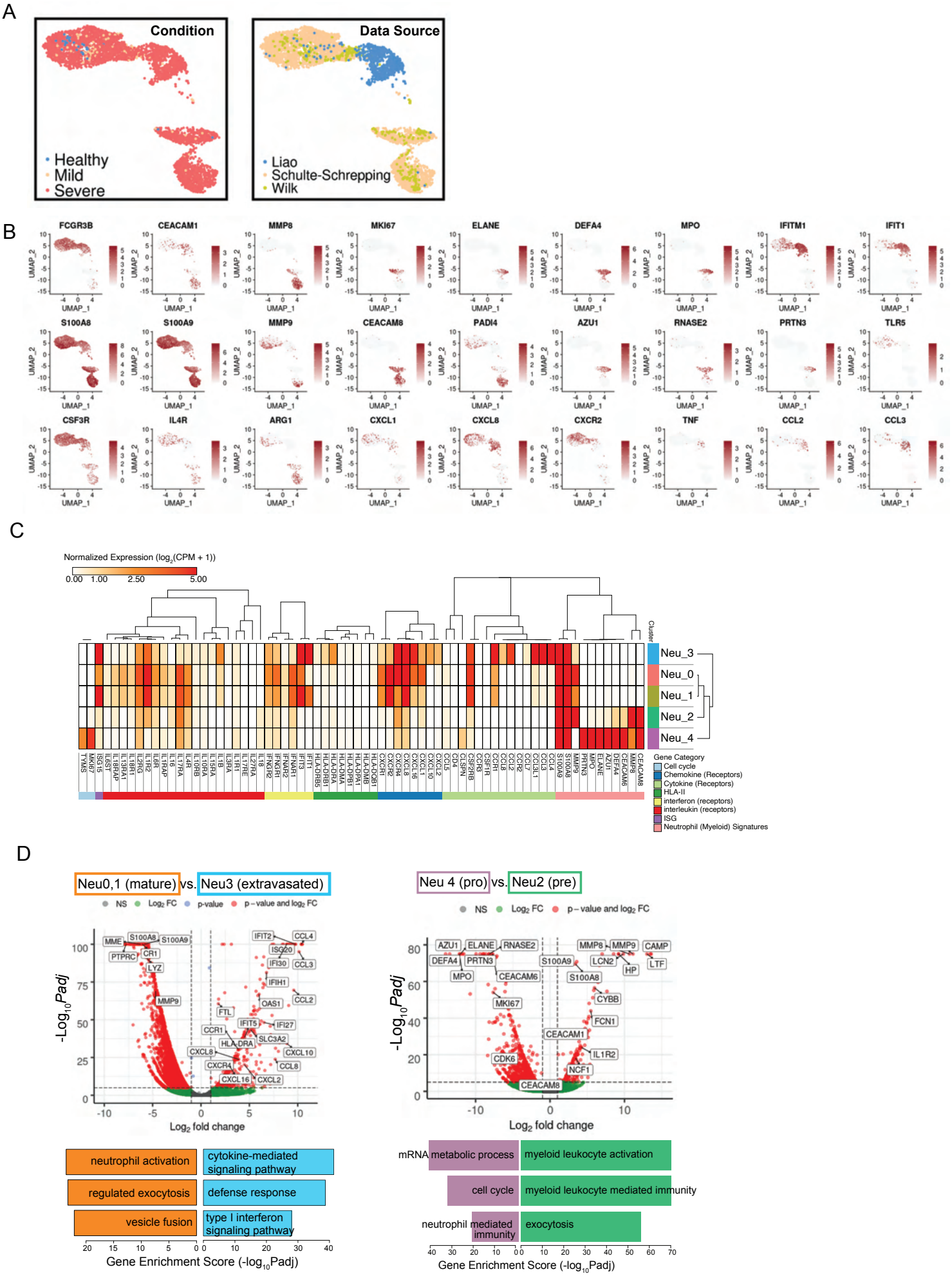


Figure S5. Sub-cluster-specific genes of neutrophils of COVID-19 patients, relative to Figure 3.

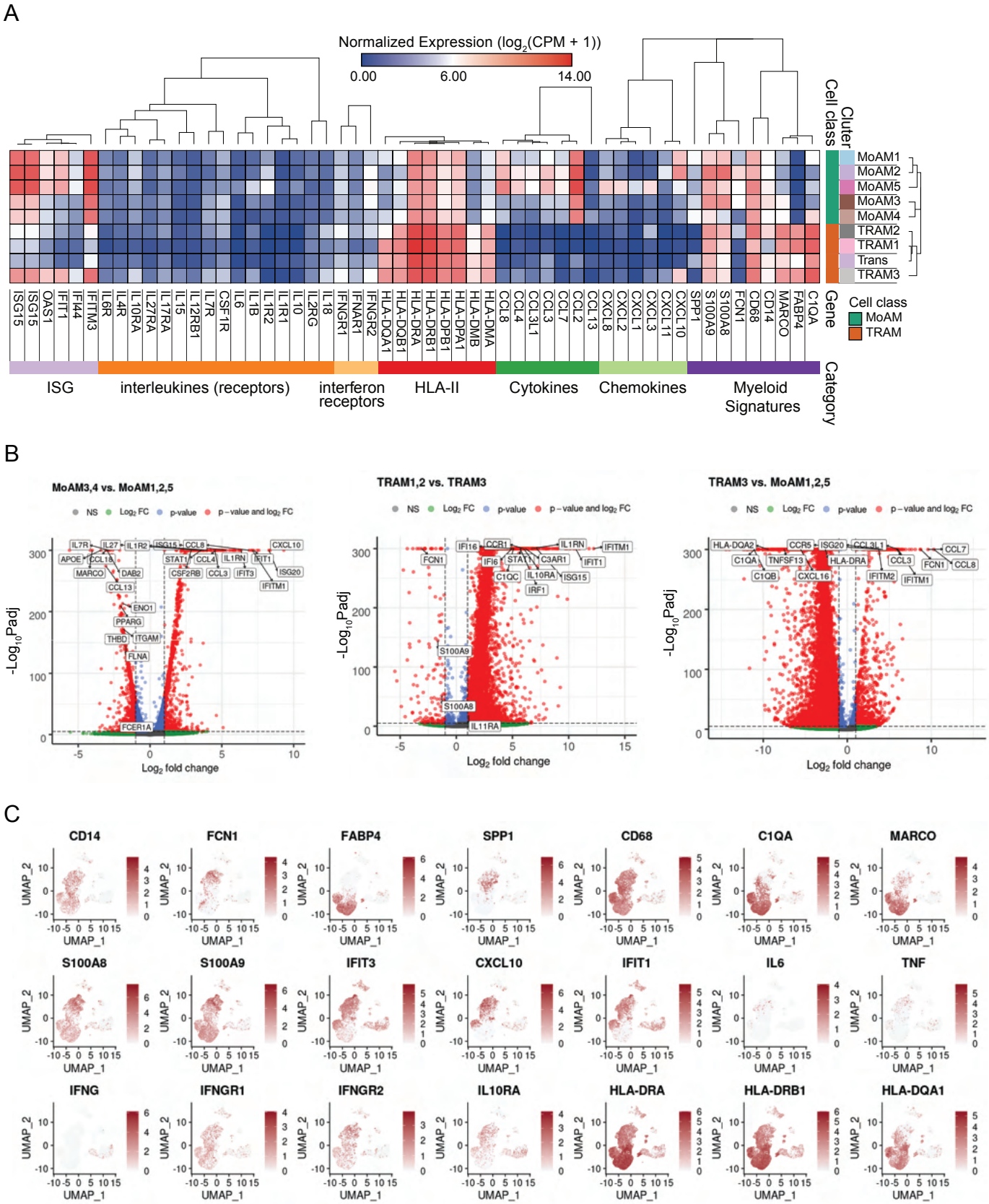


Figure S6. Macrophage-related signatures in the integrated BAL data, relative to Figure 3.

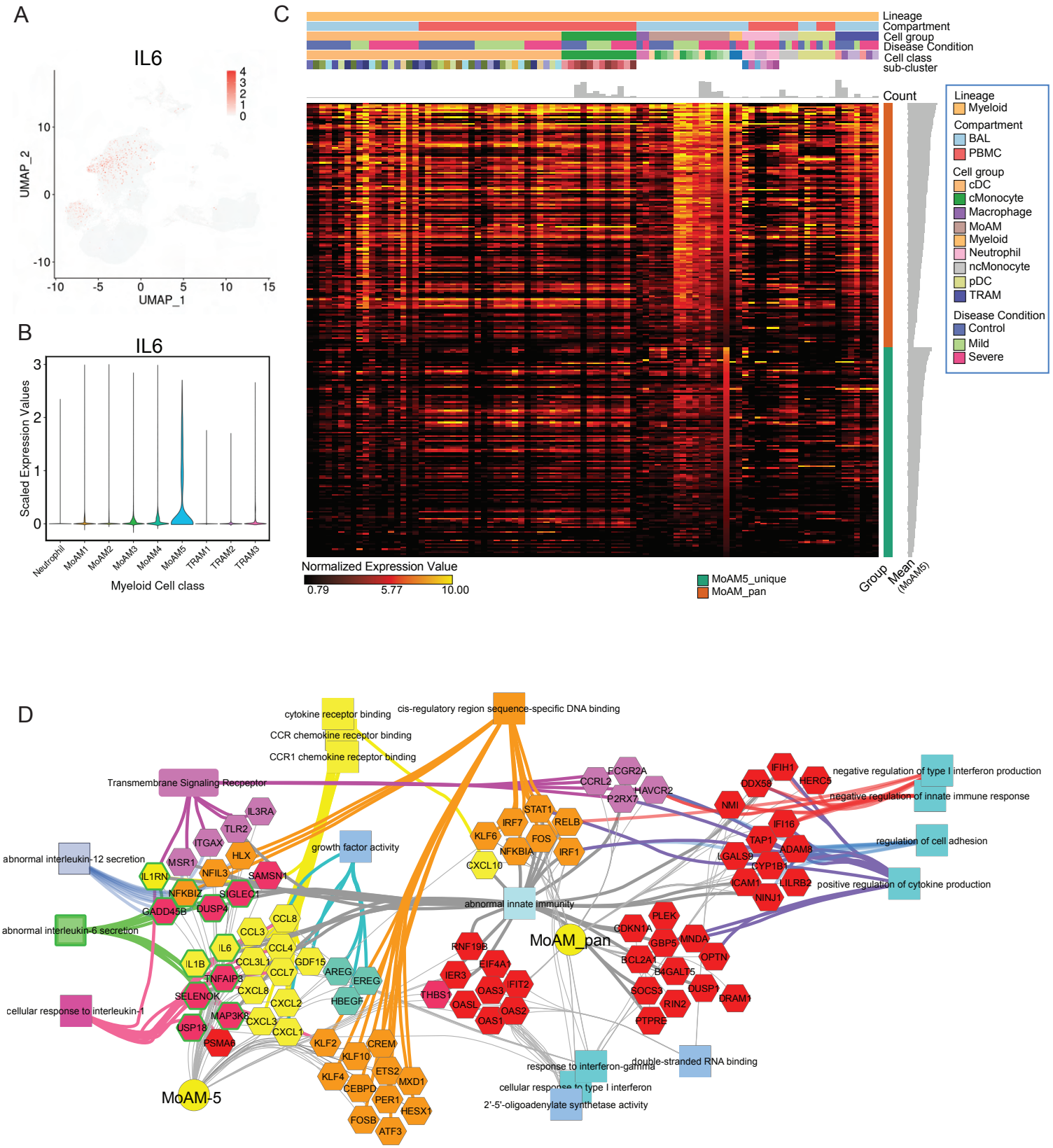


Figure S7. A uniquely-activated monocyte-derived cell type (MoAM5) exhibits a broad signature of cytokines, chemokines, and interleukins including IL6, relative to Figure 3.

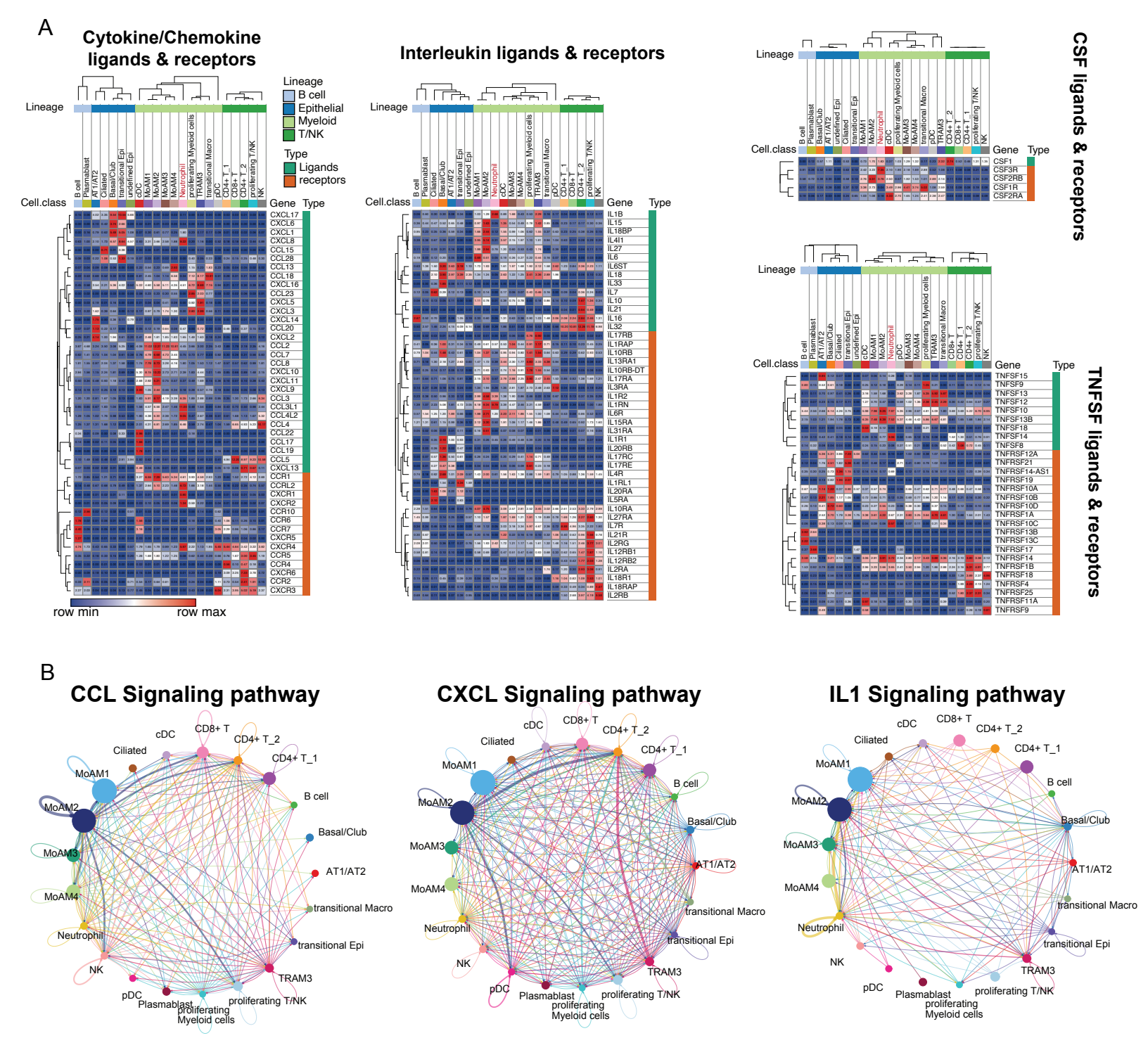


Figure S8. Cell type and cell subtype-specific divisions of cytokine, chemokine, and interleukin signaling pathways in BAL of severe COVID-19 patients, relative to Figure 3.

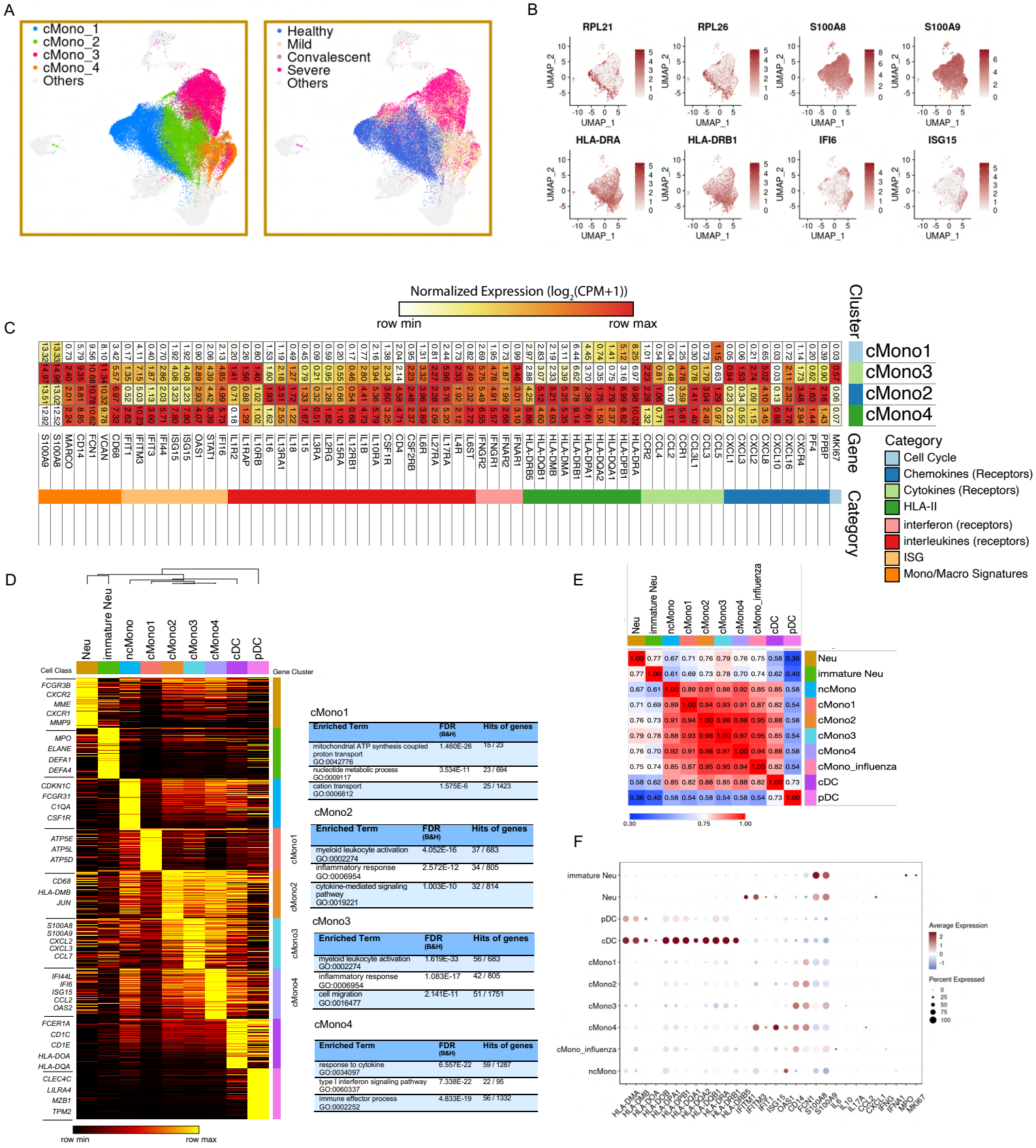


Figure S9. Characteristics of sub-clusters of classical monocytes in the integrated COVID-19 PBMC data, relative to Figure 3.

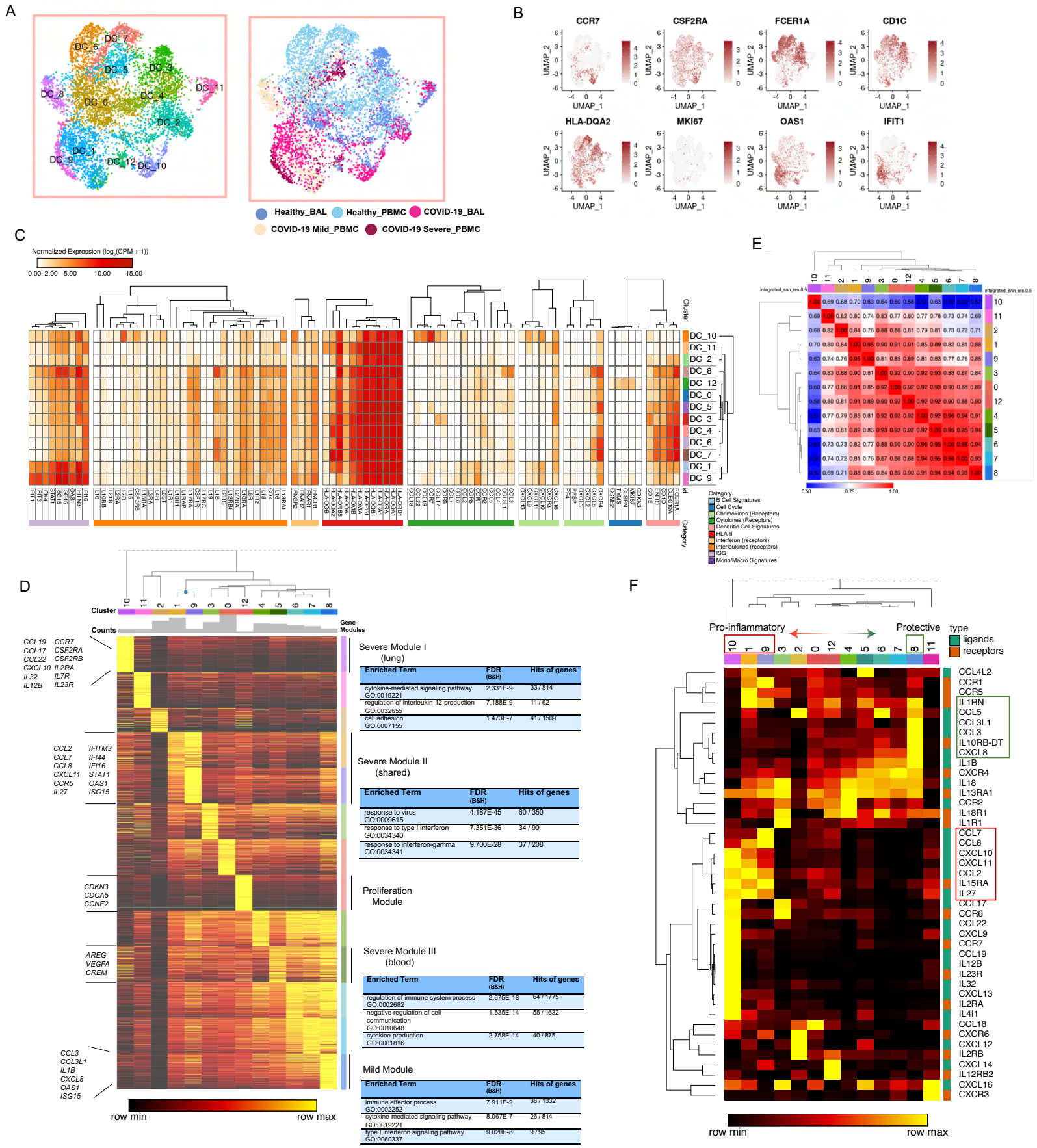


Figure S10. Features of conventional dendritic cell sub-clusters and polarized signaling genes, relative to Figure 3.

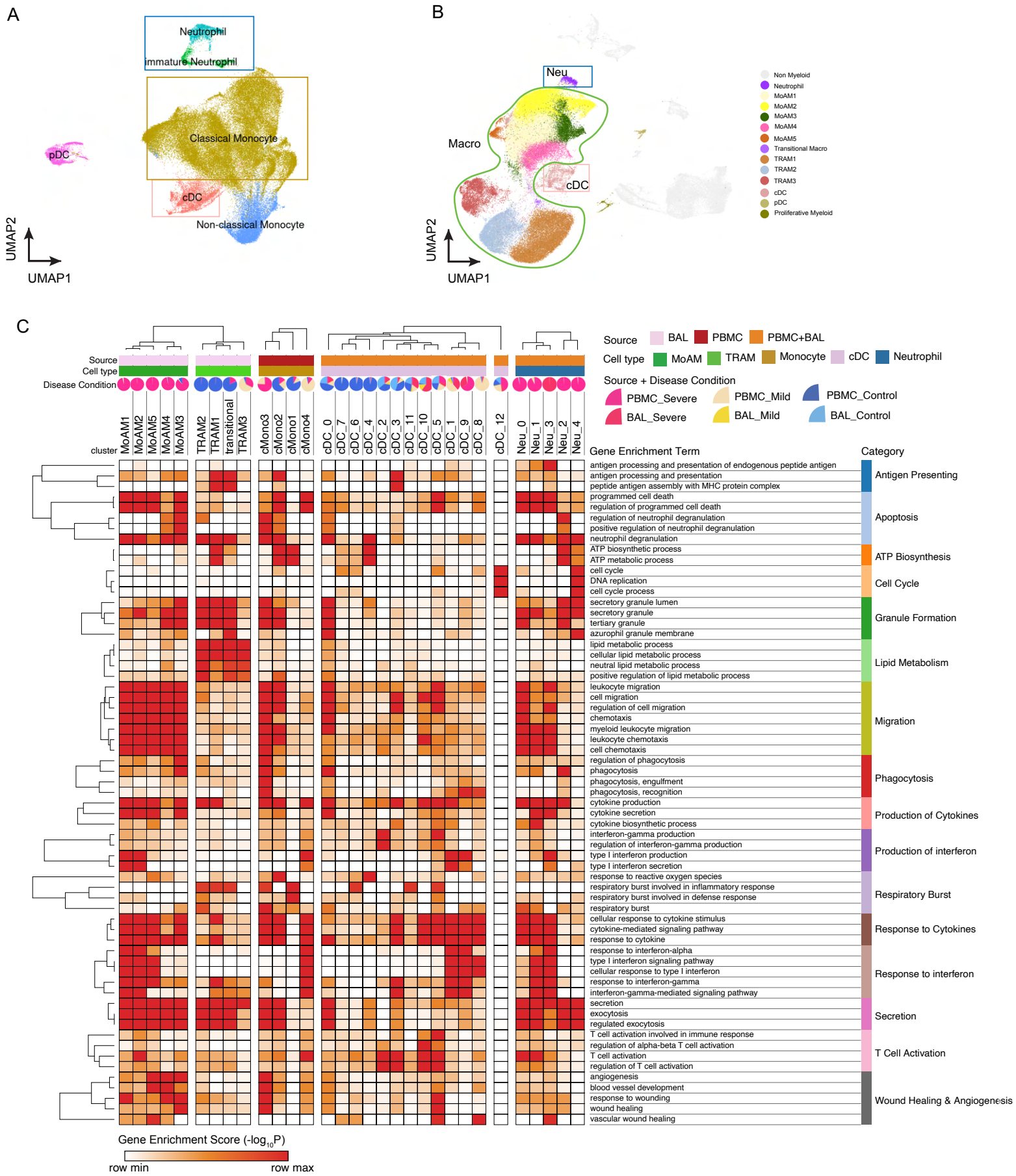


Figure S11. Landscape of myeloid cells in the integrated PBMC and BAL data, relative to Figure 3.

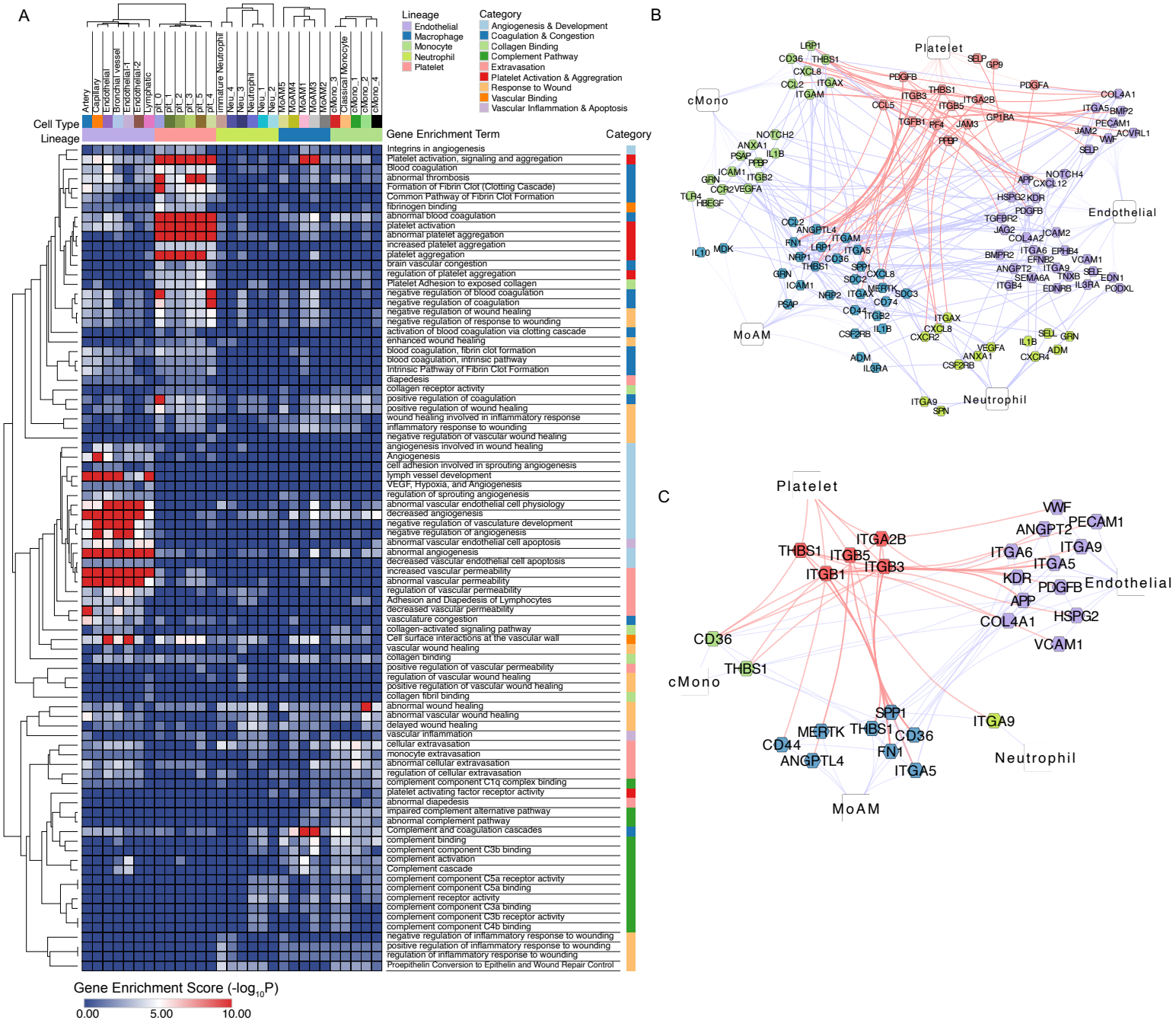


Figure S12. Gene expression signatures of cell types and subtypes activated by COVID-19 are extensively associated with coagulation, hemostasis, and thrombosis-associated pathways, functions, and knockout phenotypes, relative to Figure 4.

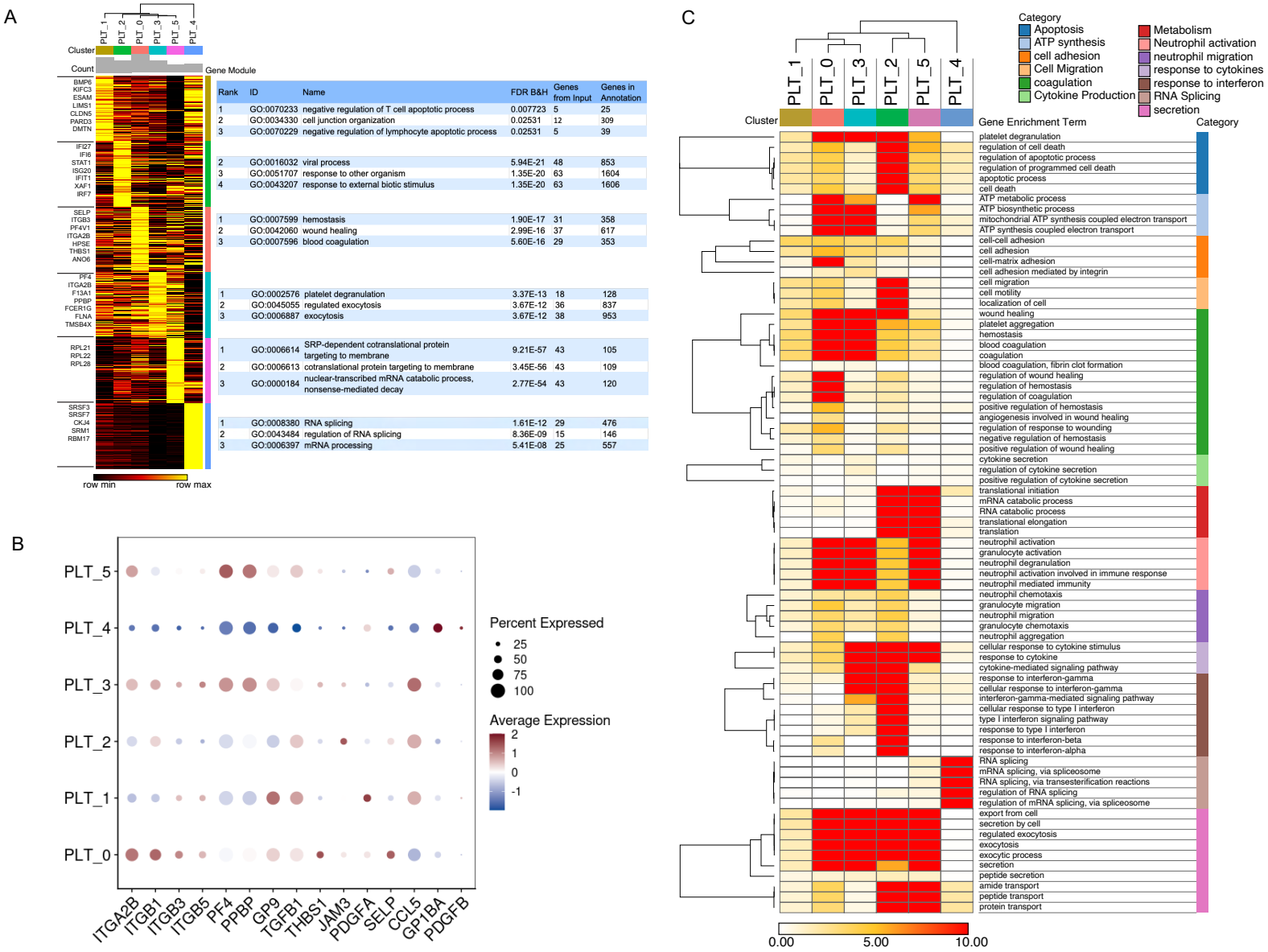


Figure S13. Emergence of platelet subtypes suggestive of functionally significant alternative roles in in hemostasis, coagulation, wound response, and neutrophil recruitment and activation, relative to Figure 4.

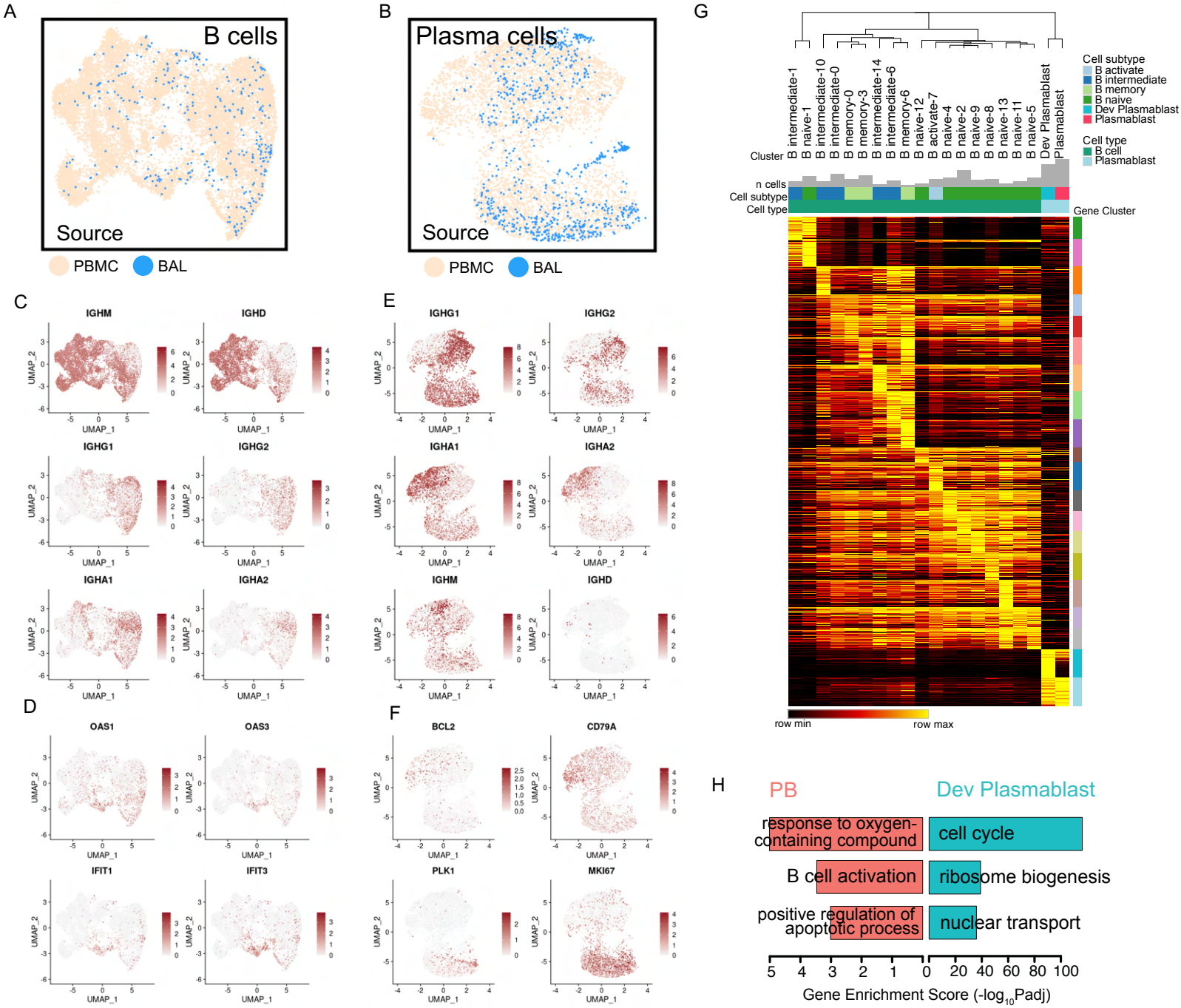


Figure S14. Consistent emergence of a series of early and maturing B cells and plasmablasts in BAL fluid and PBMC across multiple datasets, relative to Figure 5.

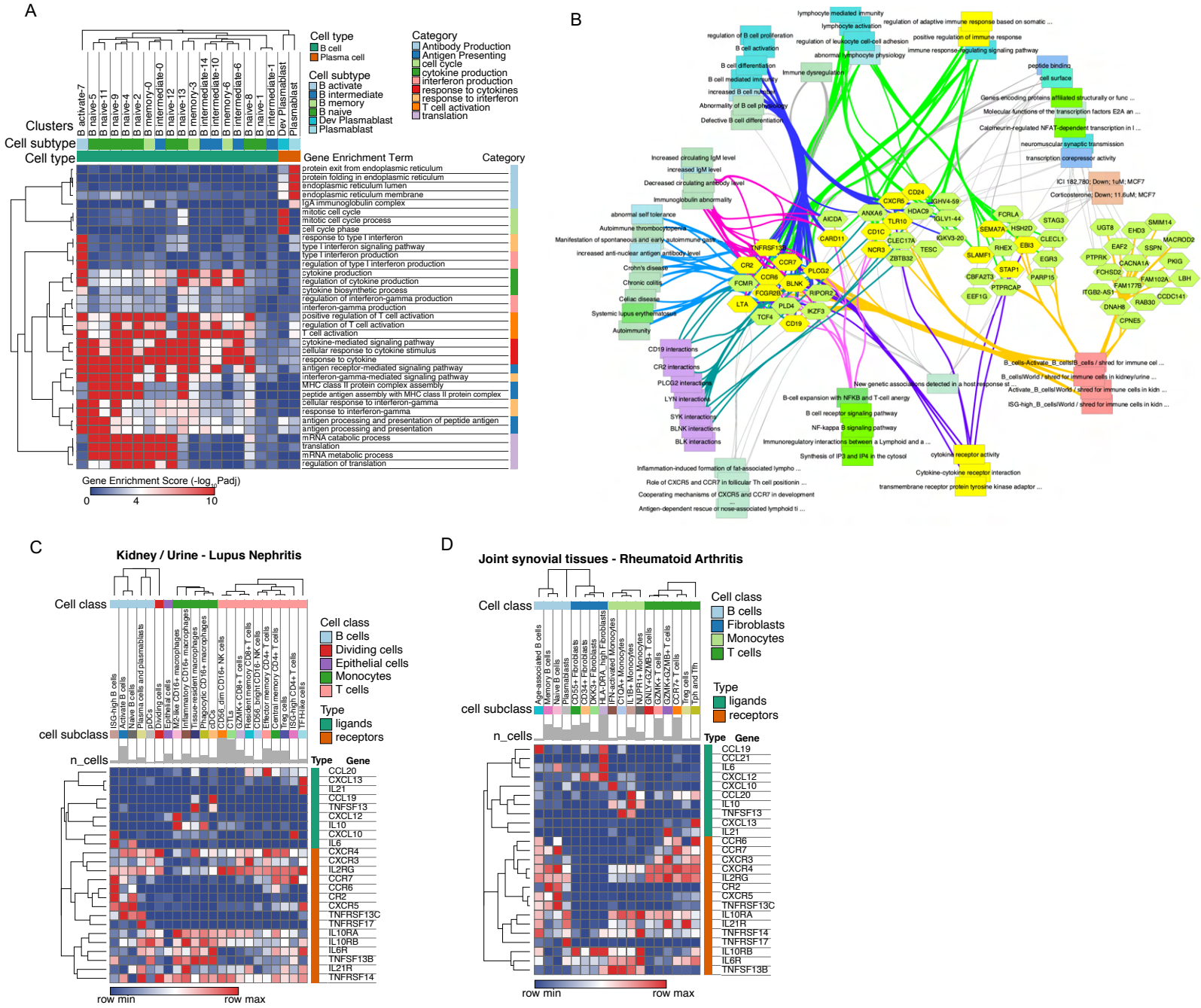


Figure S15. Gene Enrichment analysis of B cell subtypes and autoimmune-associated signatures, relative to Figure 5.

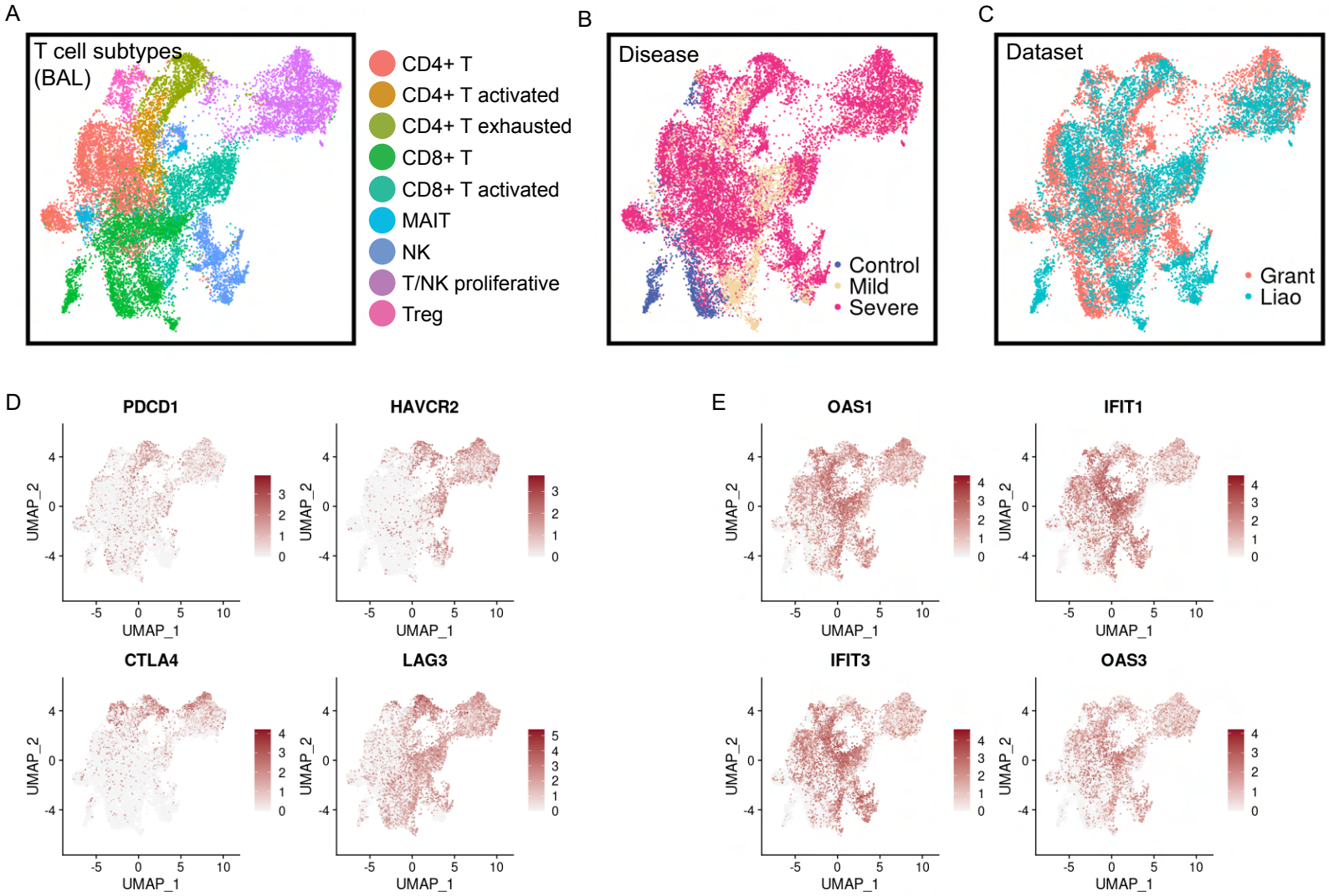


Figure S16. Distinct subtypes of T cells and NK cells in COVID-19 BAL data.

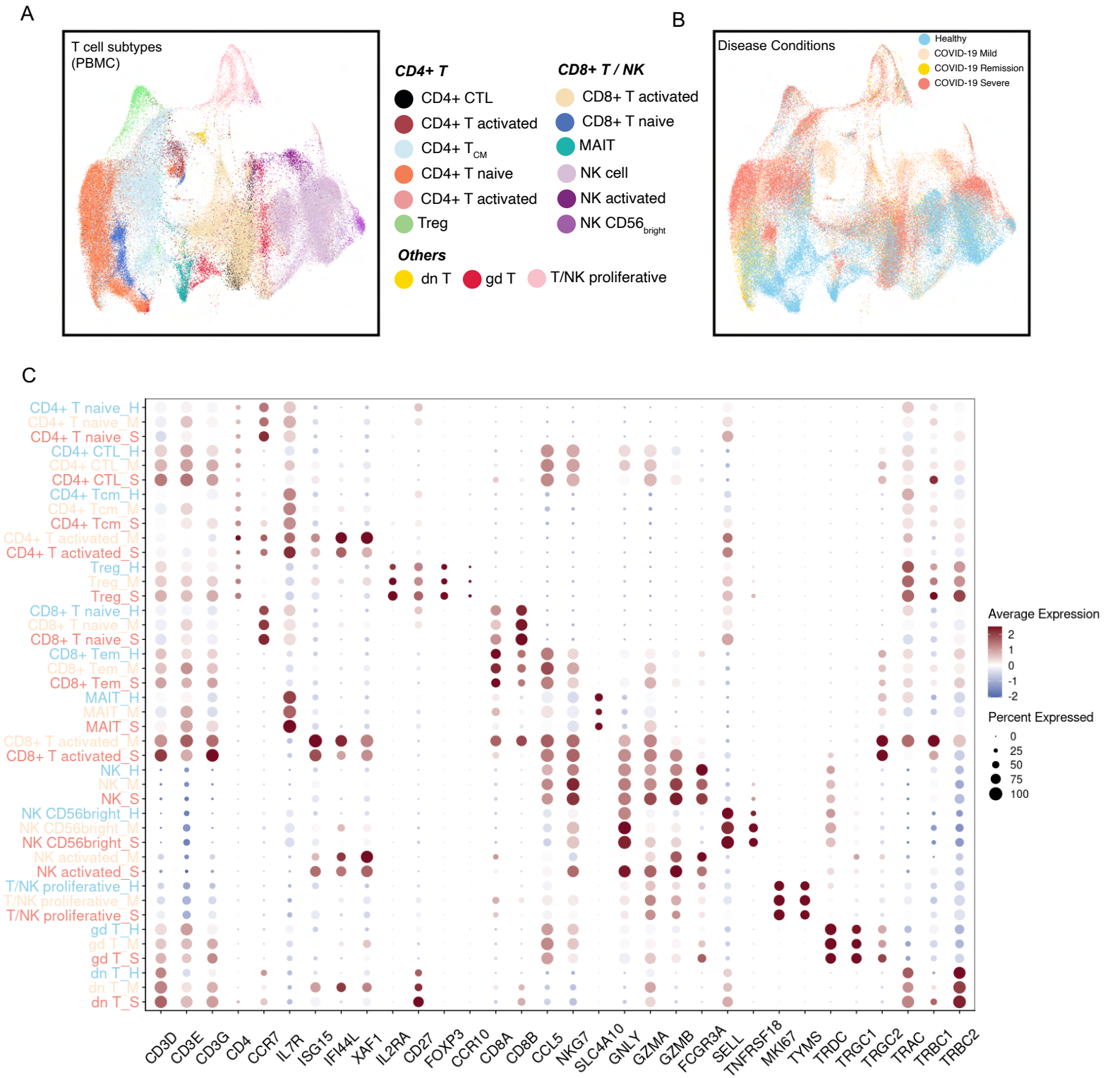


Figure S17. Various T cell and NK cell subtypes in the integrated PBMC data.

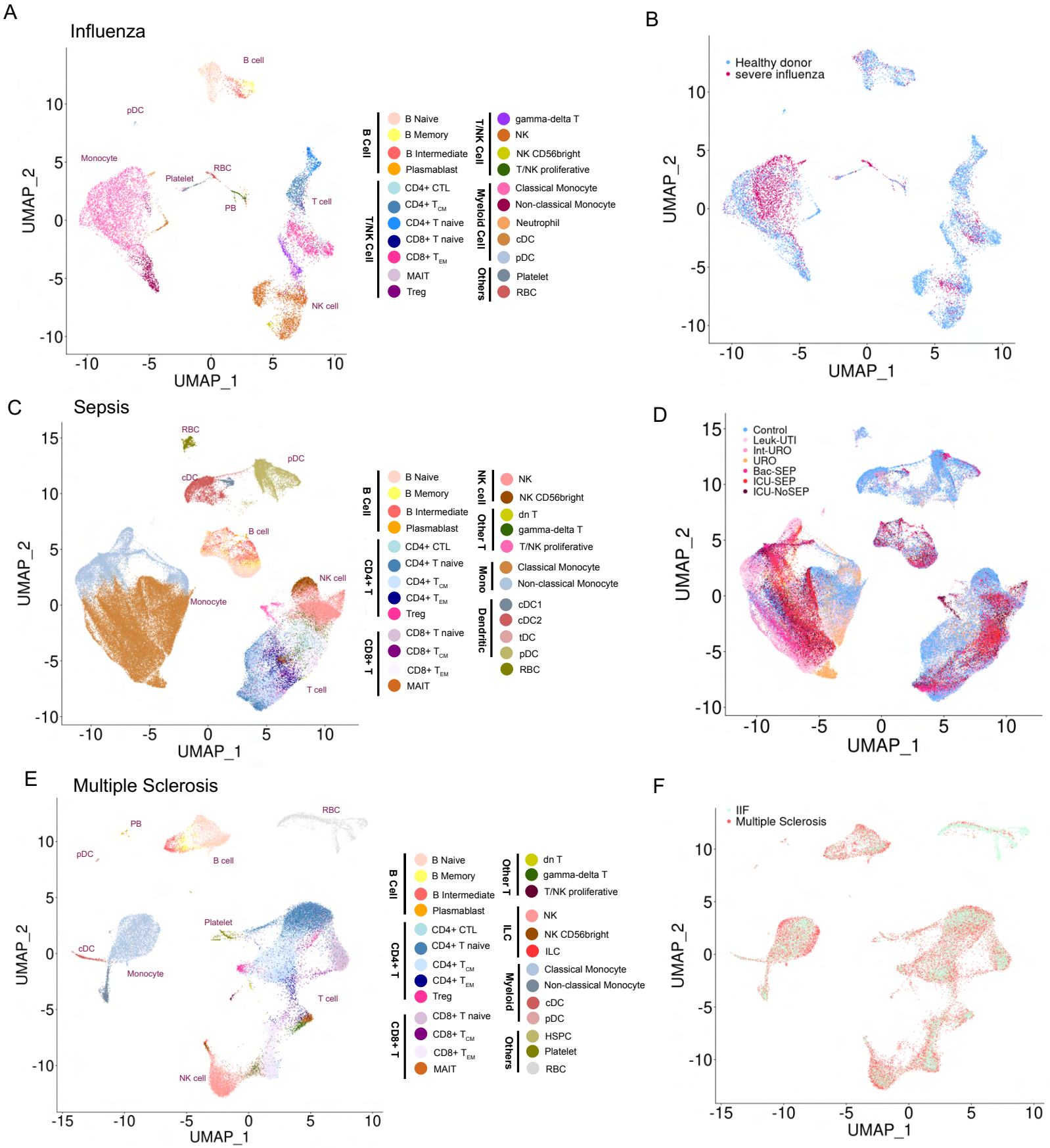


Figure S18. Various cell types in immune-mediated diseases, relative to Figure 7.