iScience, Volume *24*

Supplemental information

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

Kang Jin, Eric E. Bardes, Alexis Mitelpunkt, Jake Y. Wang, Surbhi Bhatnagar, Soma Sengupta, Daniel Pomeranz Krummel, Marc E. Rothenberg, and Bruce J. Aronow

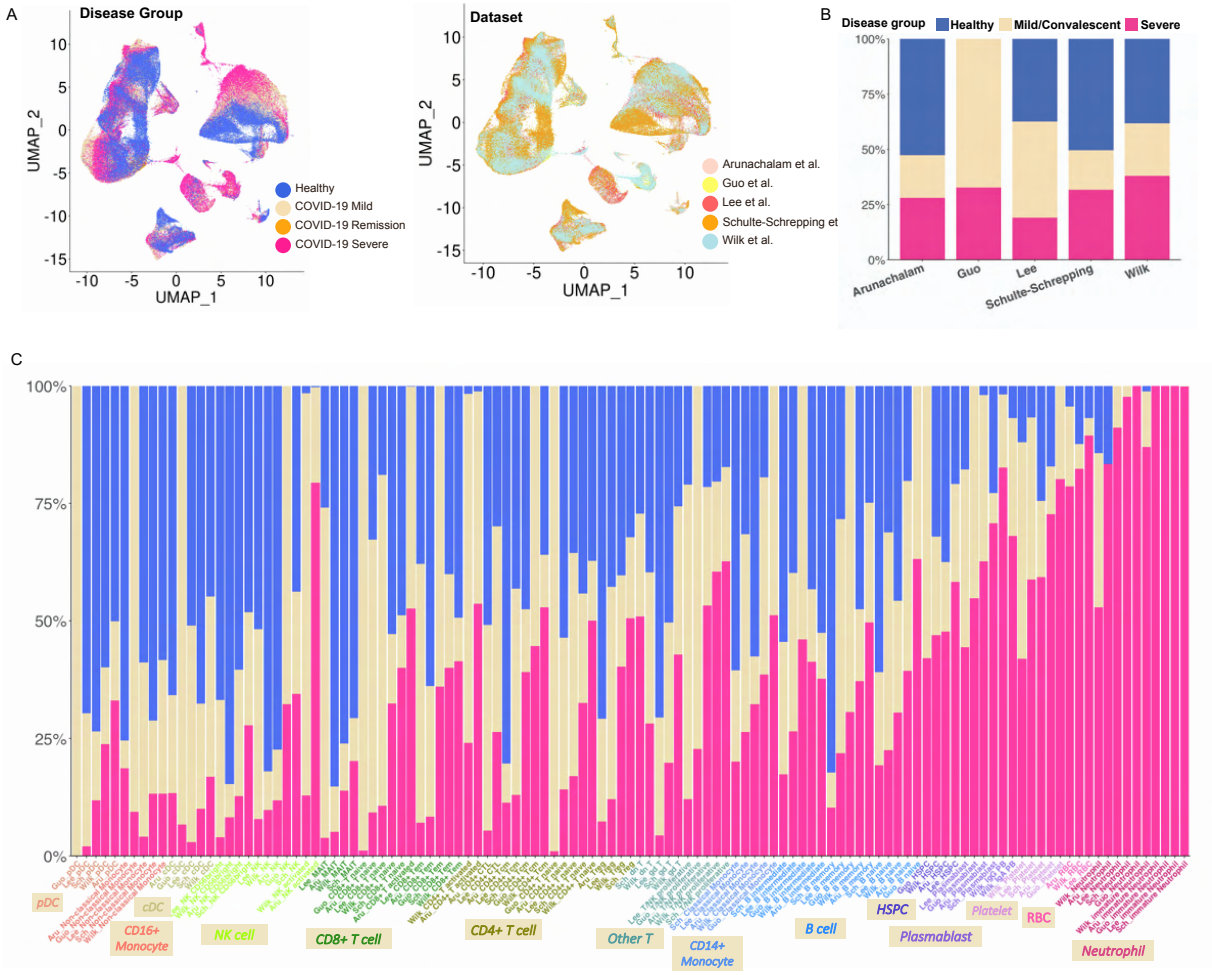

Figure S1. Cell distribution and abundance in the integrated COVID-19 PBMC data, relative to Figure 2. (A) Distributions of COVID-19 conditions (Left) and data sources (Right) for the integrated PBMC data are shown on the same UMAP of Figure 2A. (B) Bar plot depicts distributions of disease conditions in 5 individual PBMC single-cell datasets. Percentages of 3 disease conditions in each dataset is shown on y axis. (C) The integrated bar plot shows percentages of 3 disease conditions in each cell type per dataset. Dataset abbreviations and cell types were concatenated to show disease distributions of specific cell types in the selected datasets. These labels are colored by their cell type designations and ordered by the ascending percentages of COVID-19 conditions.

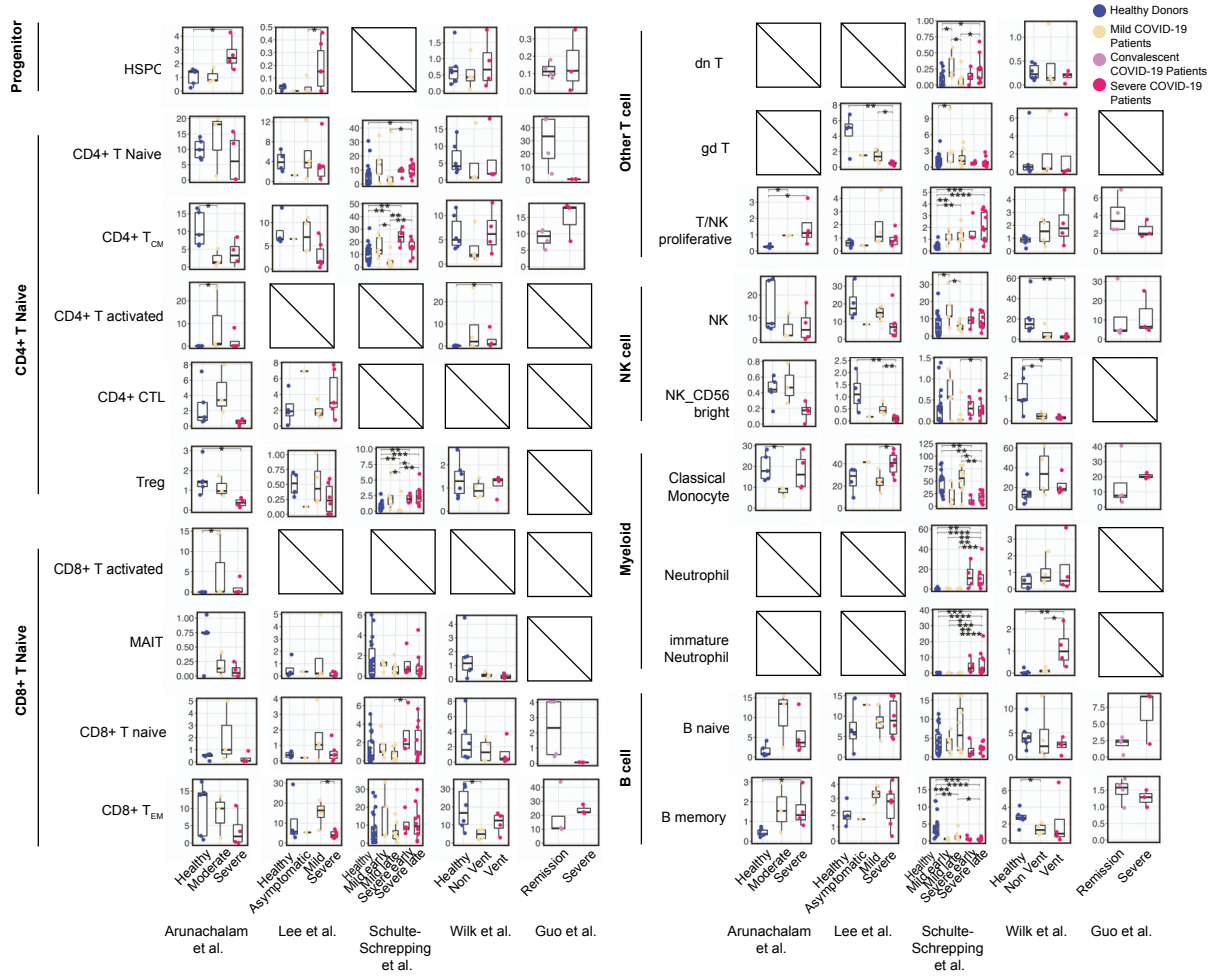

Figure S2. Dynamic changes of cell type abundances in five COVID-19 PBMC datasets, relative to Figure 2. Relative abundances and differences of major cell types in each single cell dataset are shown and compared to controls per each disease condition, per each single-cell dataset. Box plots of all cell types in PBMC are shown except for the 5 highlighted cell types shown in Figure 2B. Statistical methods are the same with Figure 2B.

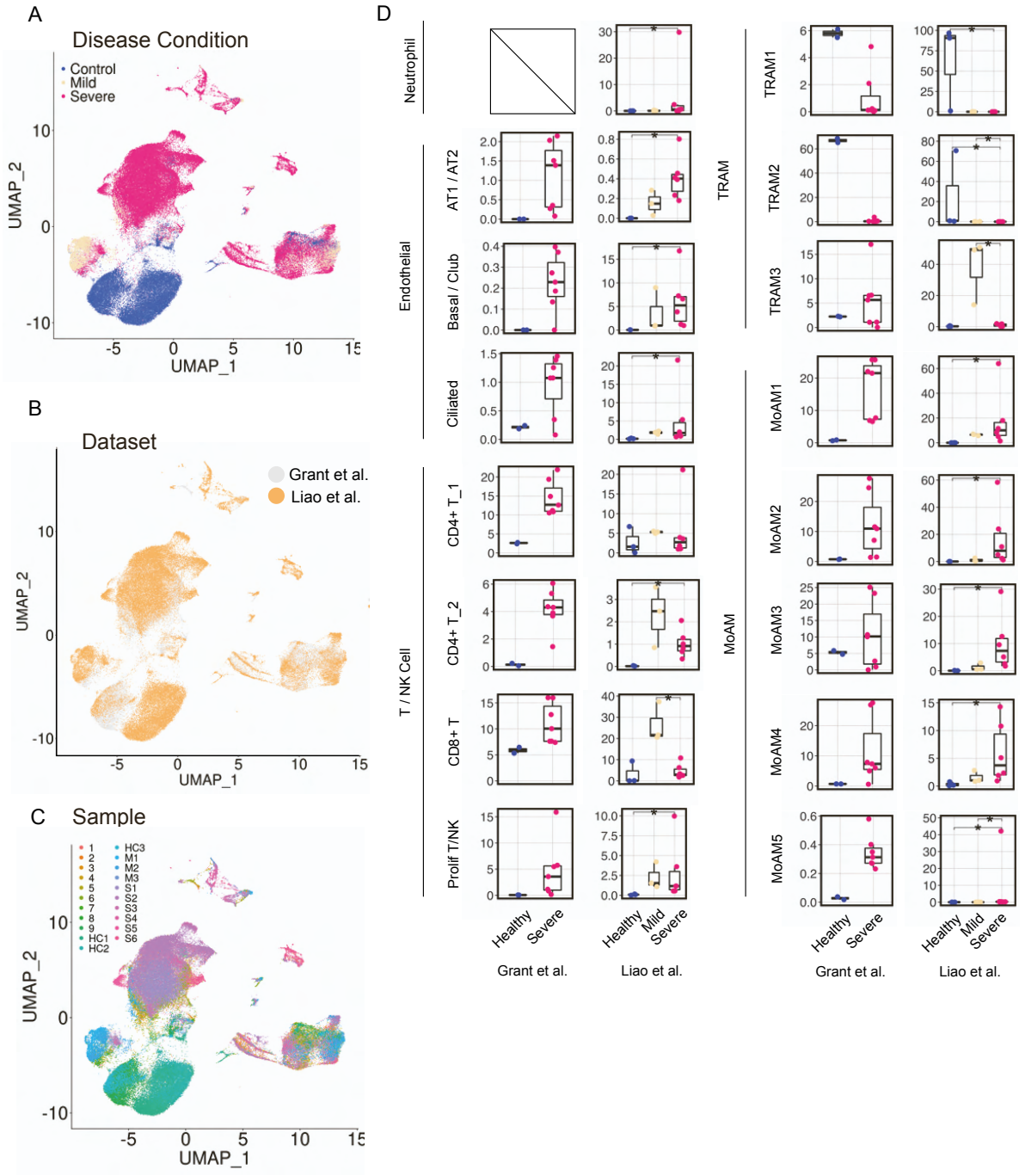

Figure S3. Cell distributions and dynamic changes in the integrated COVID-19 BAL data, relative to Figure 2. (A-C) Distributions of disease conditions (A), data sources (B) and samples (C) are shown on the same UMAP of Figure 2C. (D) Box plots depict dynamic changes of cell types across COVID-19 conditions in BAL that are not covered in Figure 2D. Statistical methods are the same with Figure 2B.

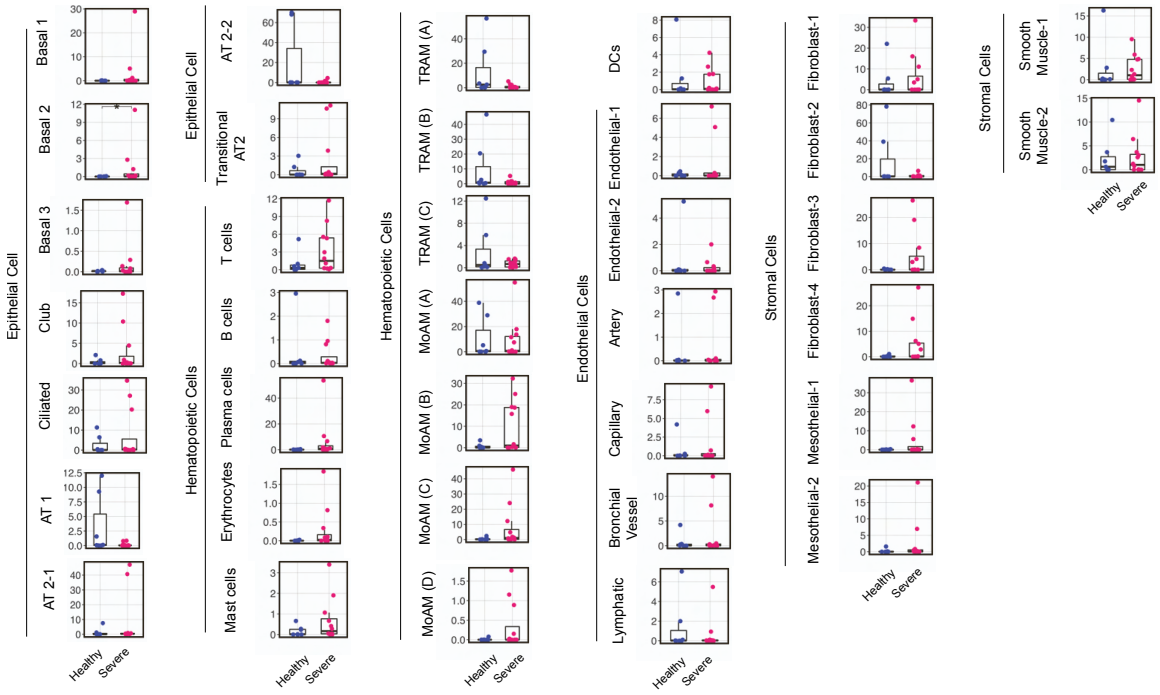

Figure S4. Cell type abundance changes in COVID-19 lung parenchyma dataset, relative to Figure 2. Box plots depict percentages of cell types in control samples and severe COVID-19 samples. We used cell type clusters identified in the original publication but modified cell naming of macrophage subtypes to distinguish monocyte derived macrophage subtypes present in BAL fluid samples. Statistical methods are the same with Figure 2B.

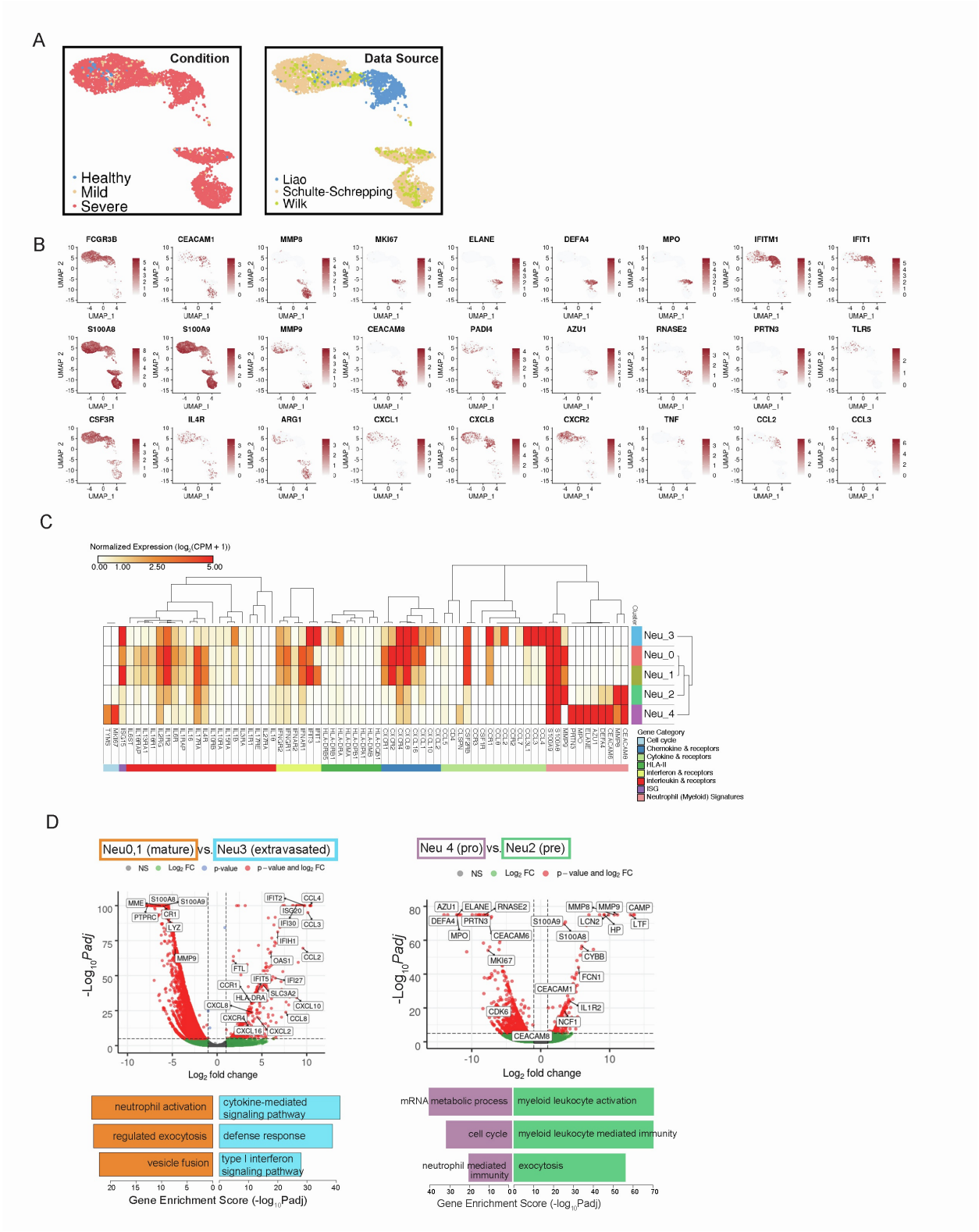

Figure S5. Sub-cluster-specific genes of neutrophils of COVID-19 patients, relative to Figure 3. (A) Distribution of disease conditions (Left) and data sources (Right) for the integrated

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.

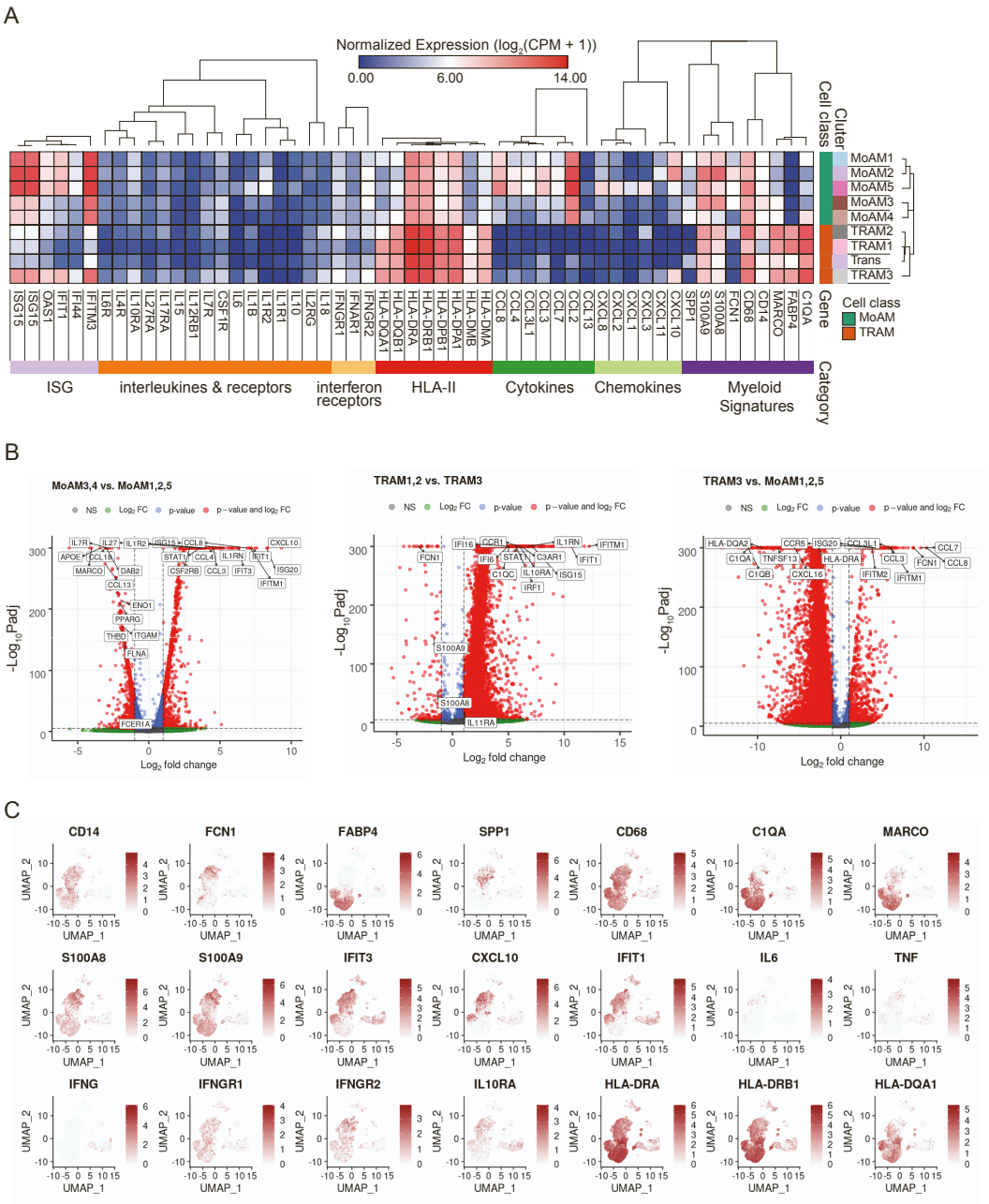

Figure S6. Macrophage-related signatures in the integrated BAL data, relative to Figure 3. (A) Normalized expression values of myeloid-cell-associated genes and other important immune signatures are shown for 9 macrophage sub-clusters. Lowly expressed genes (genes with maximal average expression level 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.

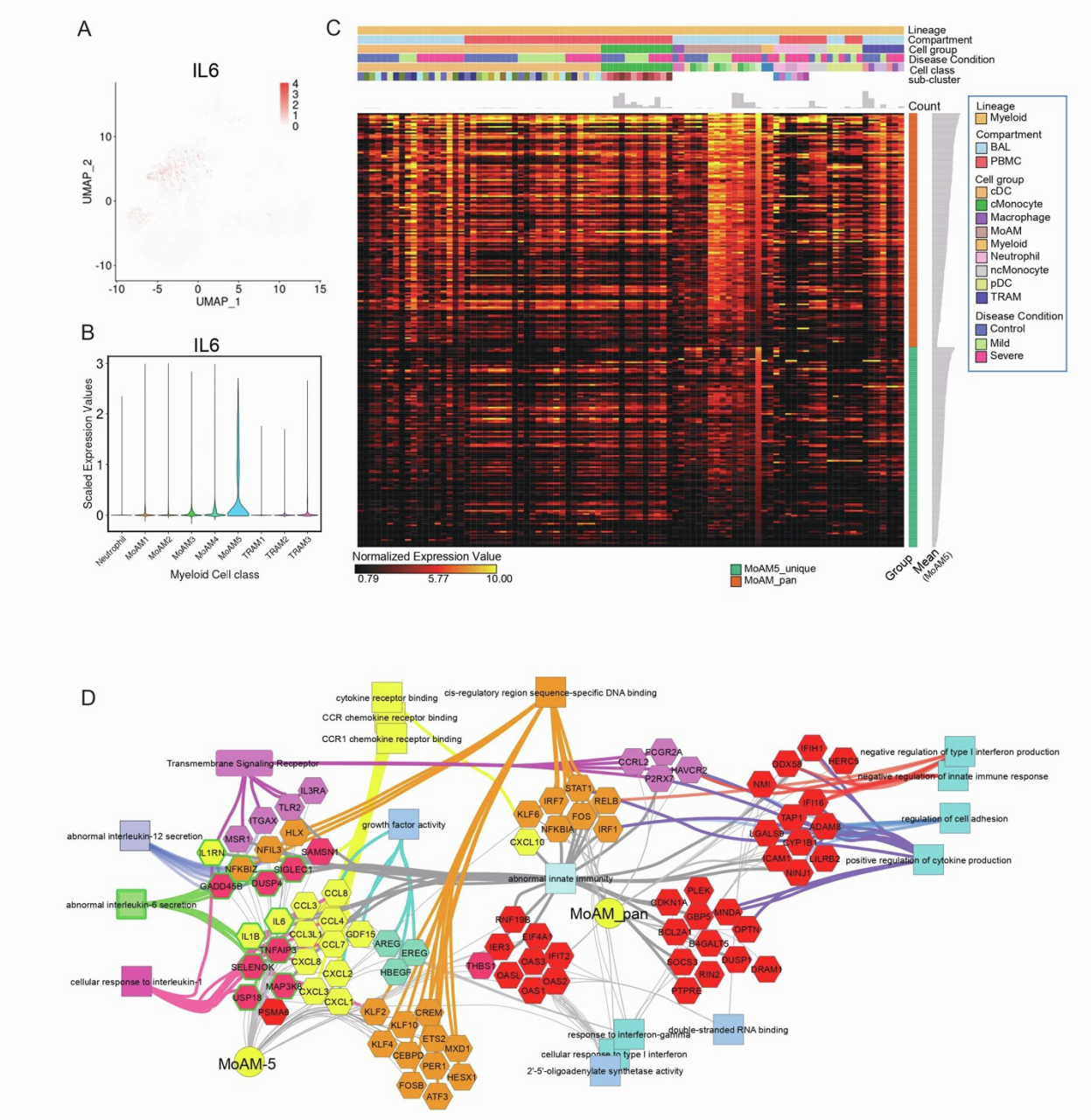

Figure S7. A uniquely-activated monocyte-derived cell type (MoAM5) exhibits a broad signature of cytokines, chemokines, and interleukins including IL6, relative to Figure 3. (A) Normalized expression values of IL6 on the same reference UMAP of integrated BAL data as Figure 3B. (B) Scale expression levels 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.

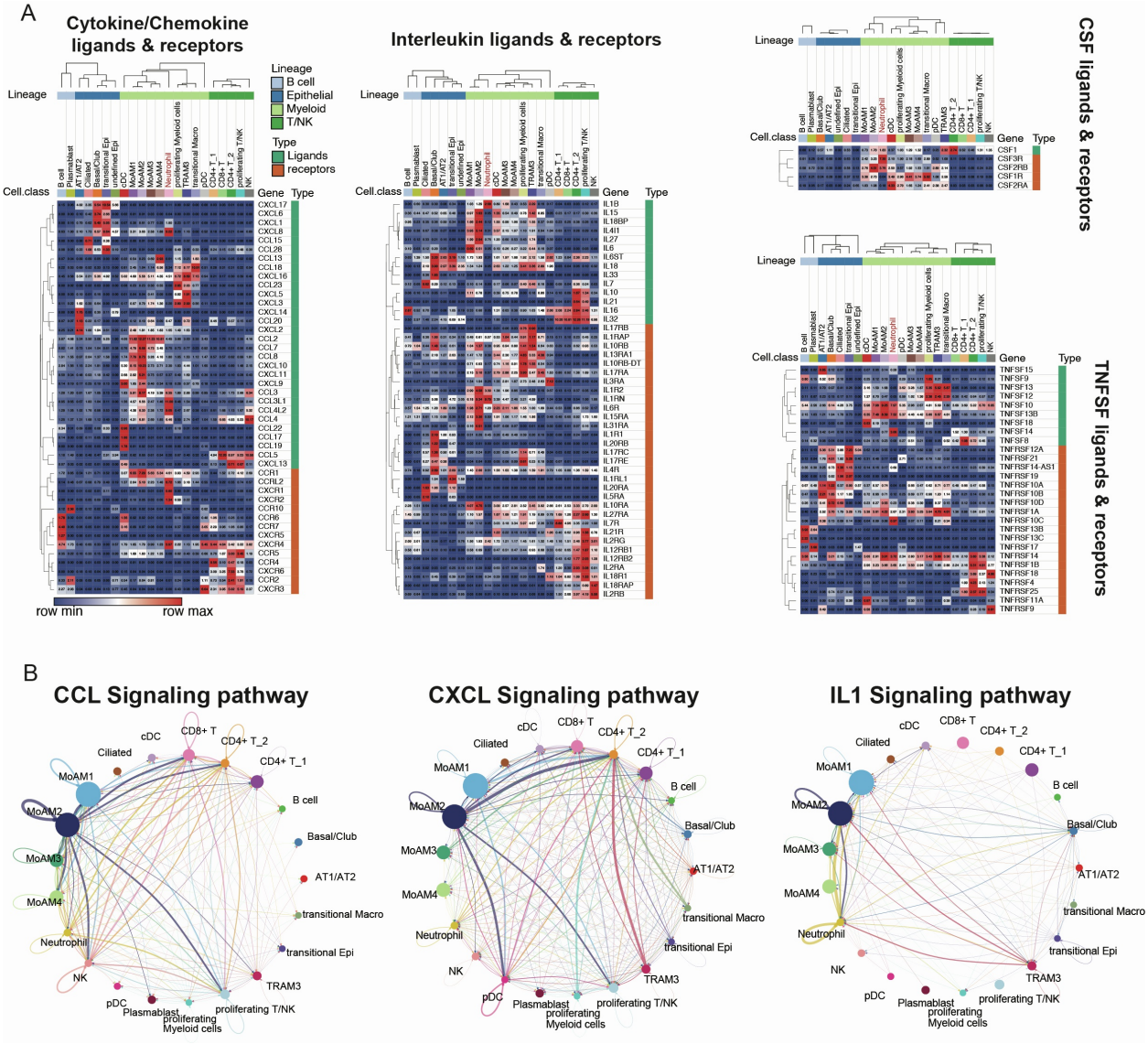

Figure S8. Cell type and cell subtype-specific divisions of cytokine, chemokine, and interleukin signaling pathways in BAL of severe COVID-19 patients, relative to Figure 3. (A) Heatmap of expression patterns of ligands and receptors in cytokine, chemokine, interleukin, CSF and TNFSF signaling pathways 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.

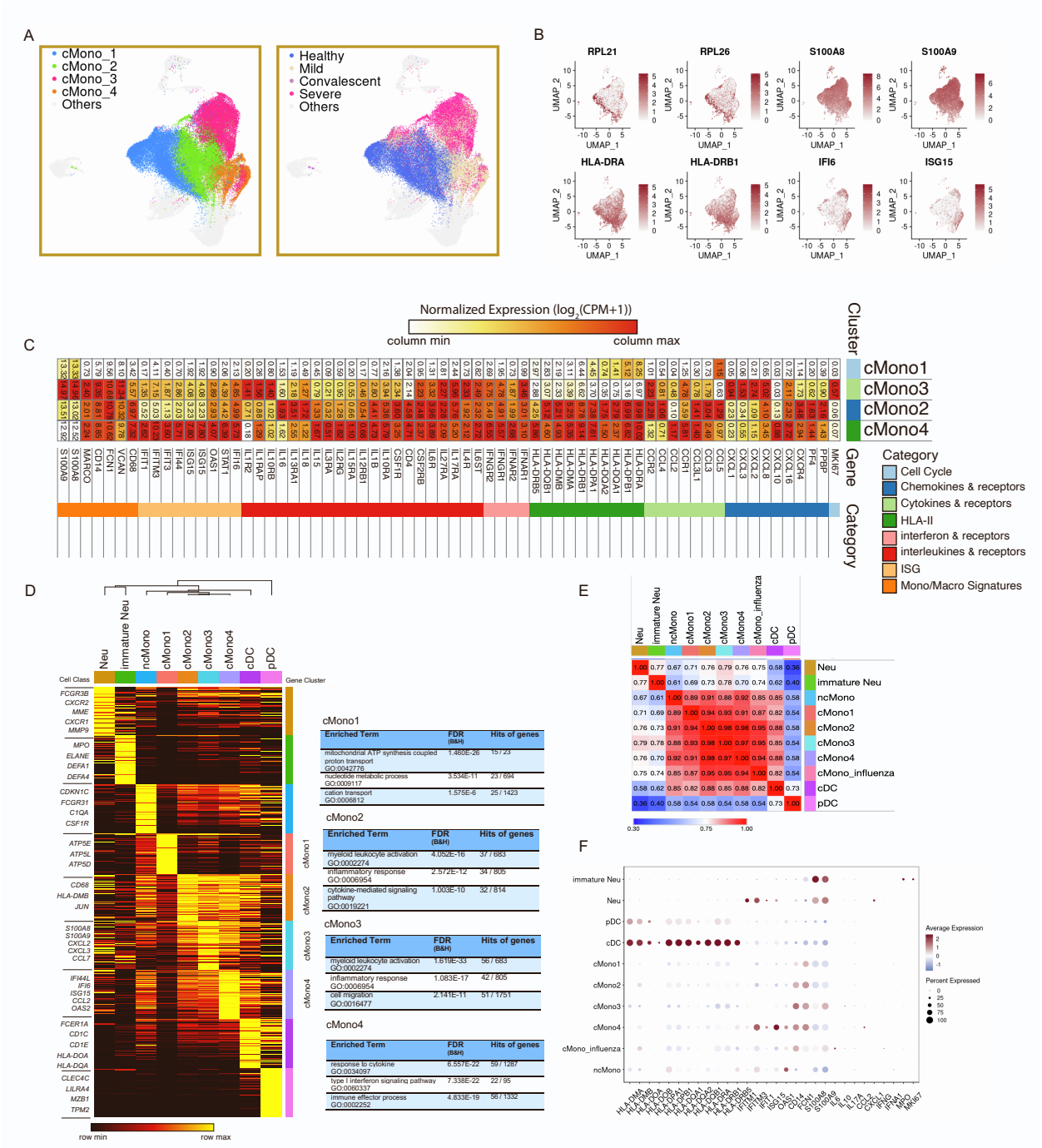

Figure S9. Characteristics of sub-clusters of classical monocytes in the integrated COVID-19 PBMC data, relative to Figure 3. (A) UMAPs of 4 sub-clusters (Left) and COVID-19 conditions (Right) of classical monocytes are shown. Grey dots are other myeloid cells in the UMAP of integrated PBMC myeloid data. (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.

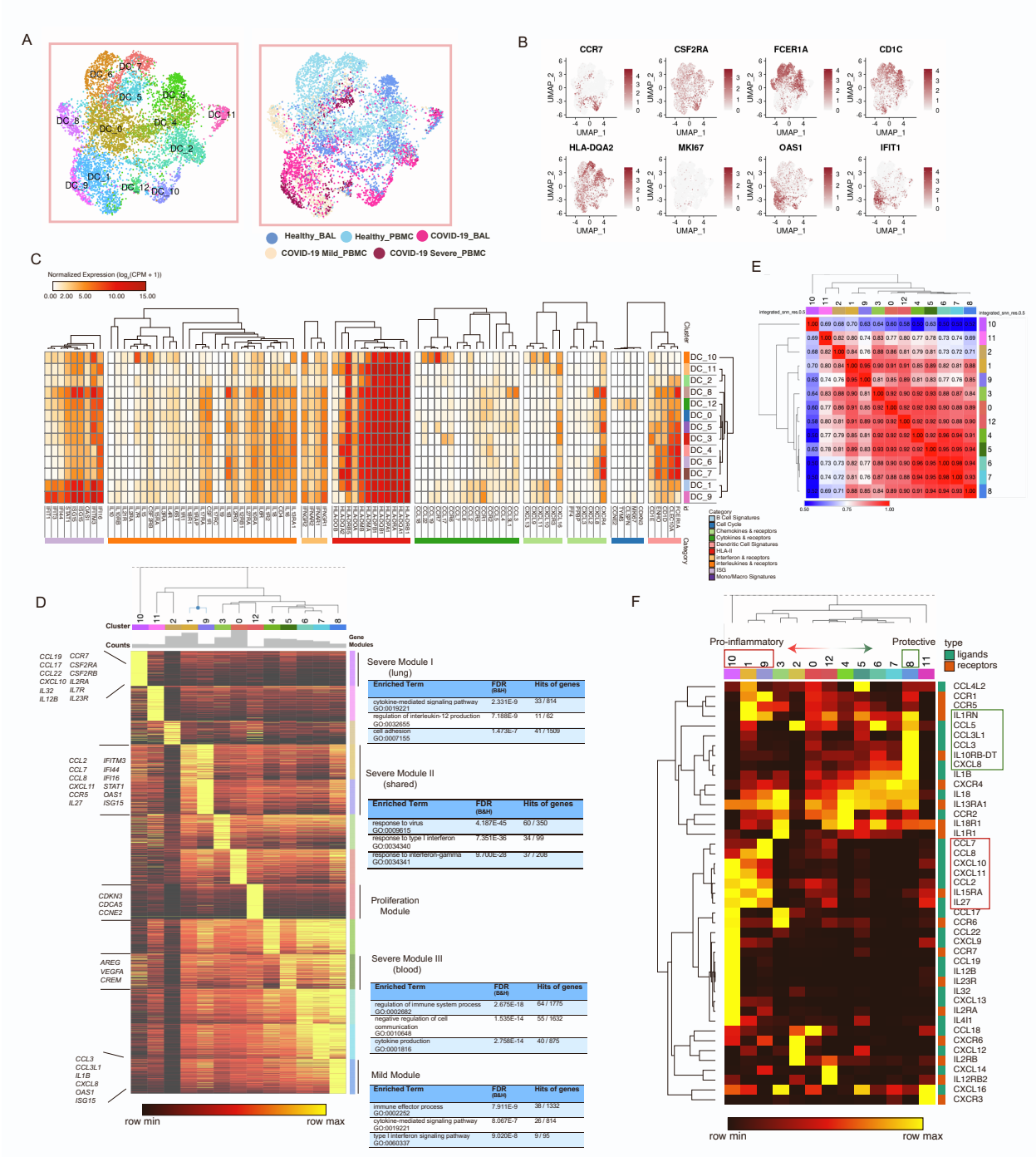

Figure S10. Features of conventional dendritic cell sub-clusters and polarized signaling genes, relative to Figure 3. (A) UMAPs of 13 sub-clusters (Left) and sources (Right) of conventional dendritic cells after data integration. (B) Normalized expression values of sub-cluster-specific genes on the UMAP. (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.

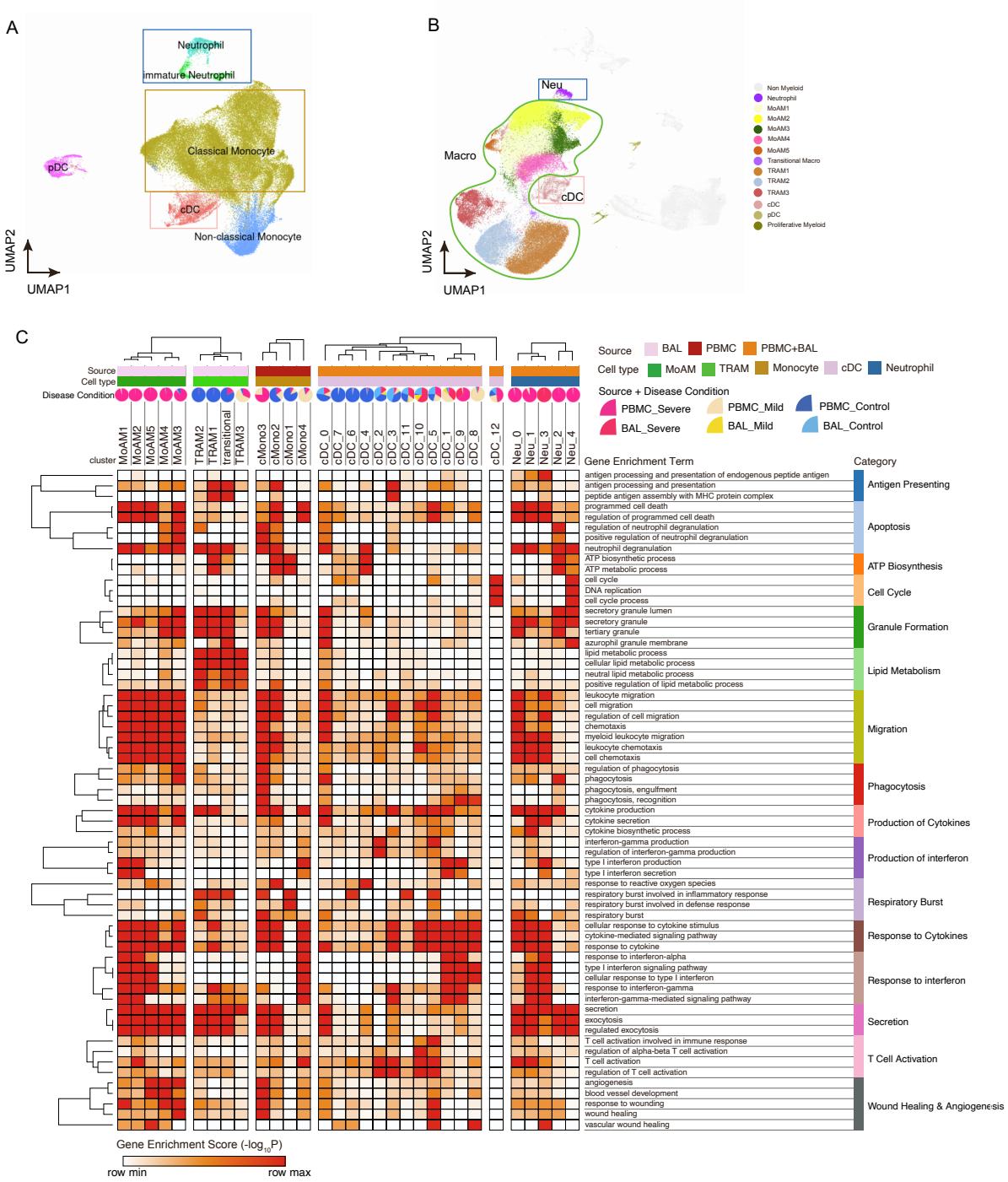

Figure S11. Landscape of myeloid cells in the integrated PBMC and BAL data, relative to Figure 3. (A-B) UMAPs of myeloid cells in integrated PBMC (A) and BAL (B) data. Cell types which were further clustered are highlighted in different colors. (C) The heatmap shows associations between subclusters of 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.

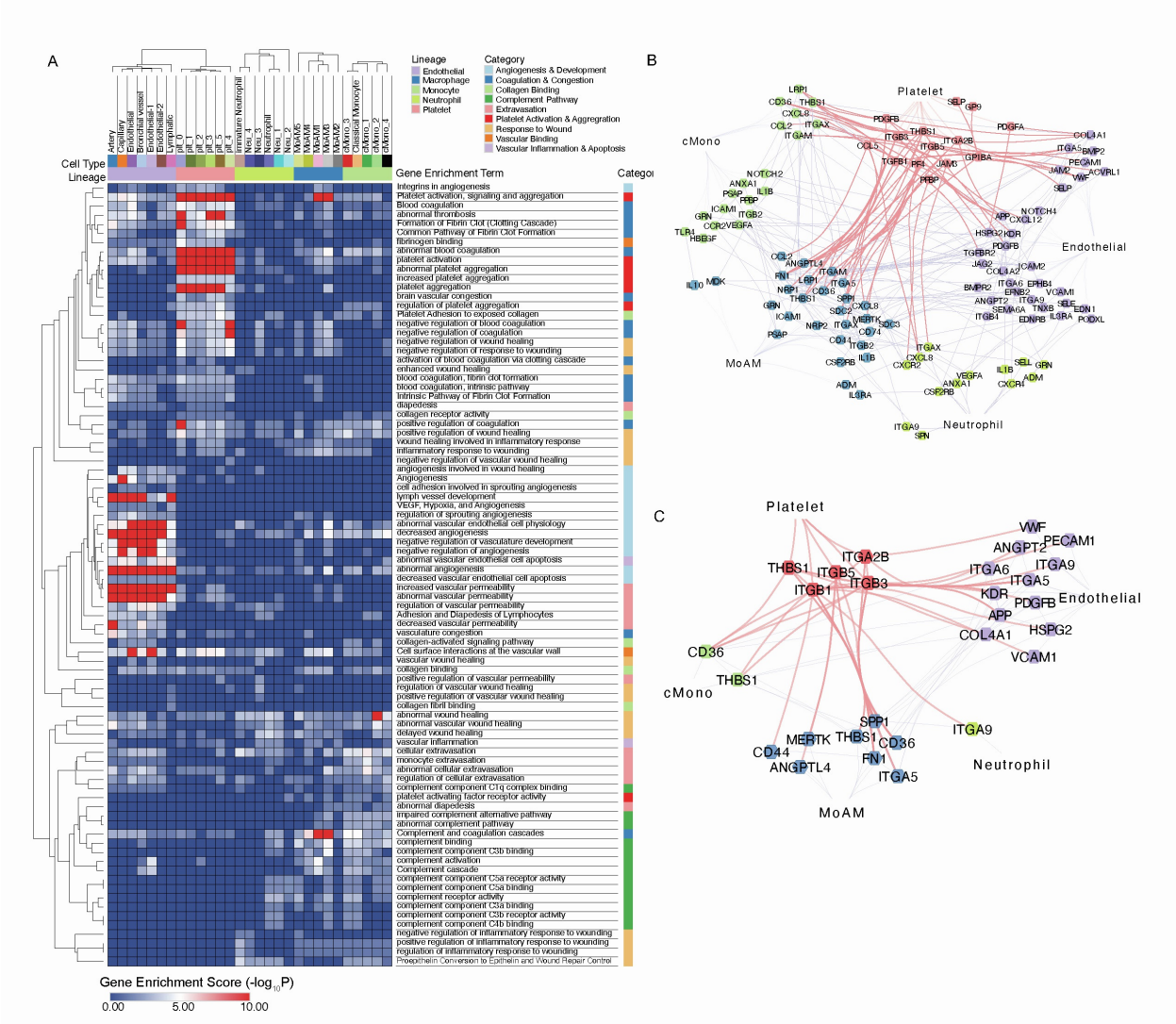

Figure S12. Gene expression signatures of cell types and subtypes activated by COVID-19 are extensively associated with coagulation, hemostasis, and thrombosis-associated pathways, functions, and knockout phenotypes, relative to Figure 4. (A) Functional association heatmap of gene signatures from COVID-19 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.

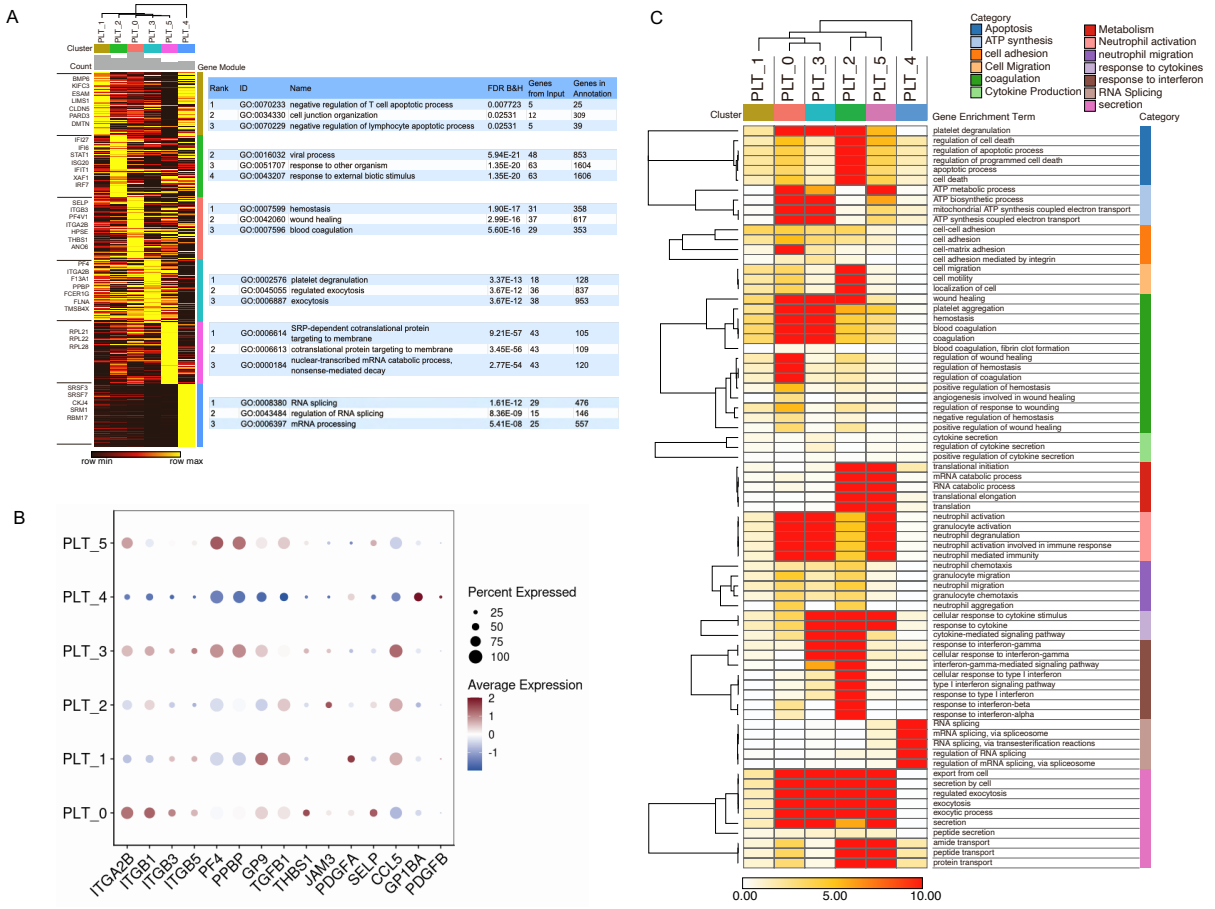

Figure S13. Emergence of platelet subtypes implicating functionally significant alternative roles in hemostasis, coagulation, wound response, and neutrophil recruitment and activation, relative to Figure 4. (A) The heatmap shows ToppCell gene modules of 6 platelet sub-clusters in COVID-19 PBMC. Each 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.

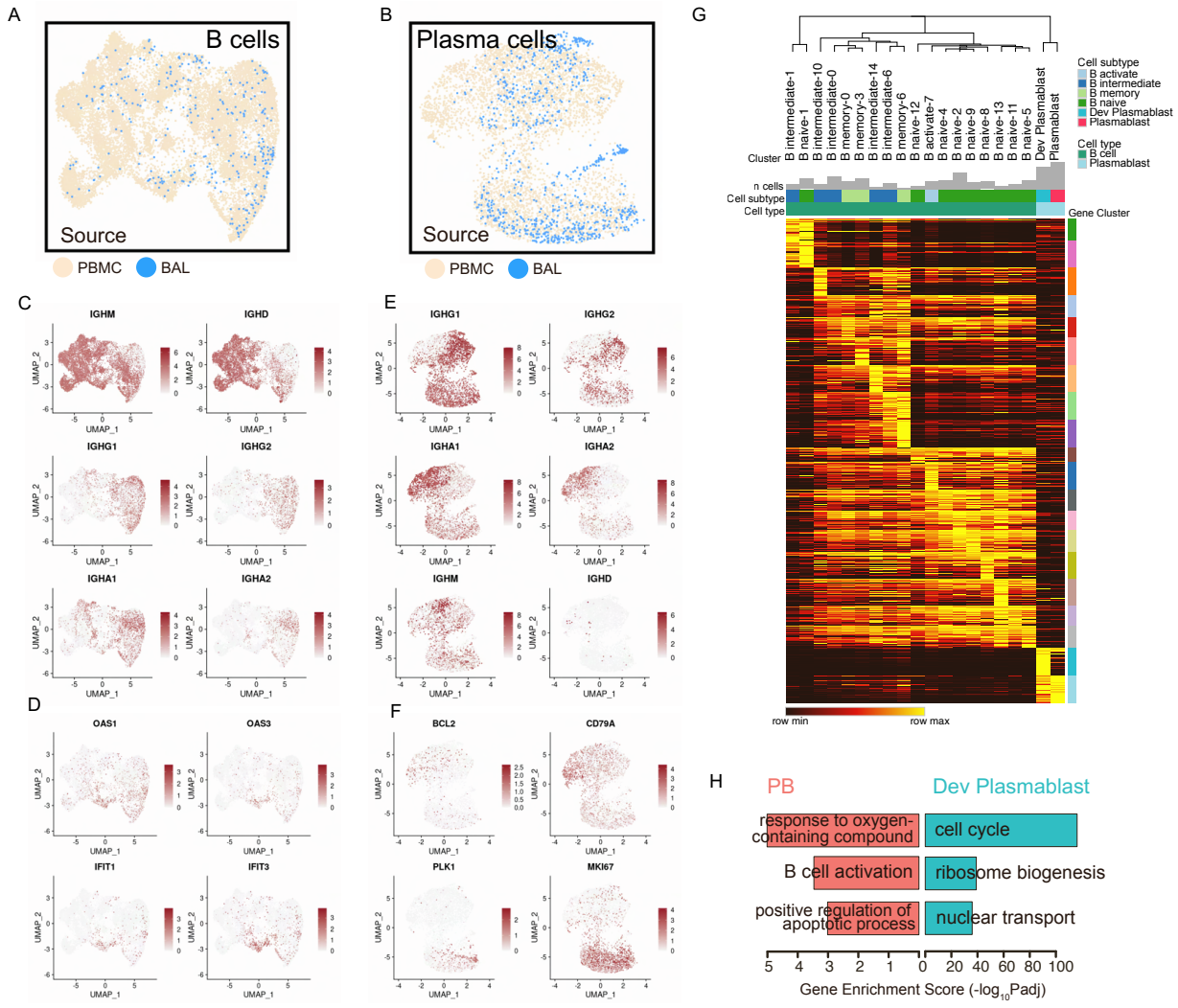

Figure S14. Consistent emergence of a series of early and maturing B cells and plasmablasts in BAL fluid and PBMC across multiple datasets, relative to Figure 5. (A-B) UMAPs of B cells (A) and plasmablasts (B) from multiple datasets. (C-D) UMAP of normalized expression values of immunoglobulin genes (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.

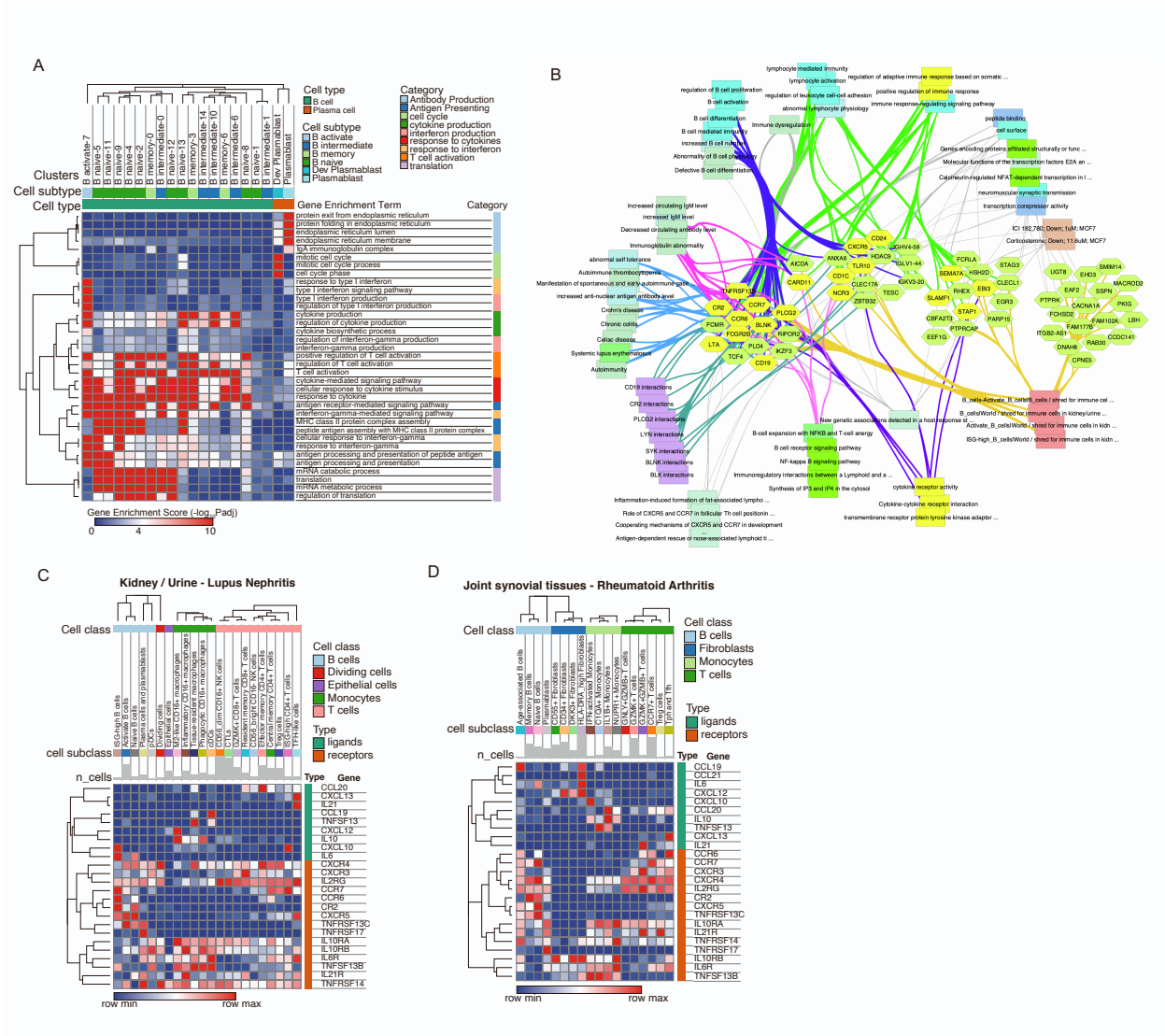

Figure S15. Gene Enrichment analysis of B cell subtypes and autoimmune-associated signatures, relative to Figure 5. (A) Heatmap shows gene enrichment scores of B-cell-associated pathways for each B cell sub-cluster and plasmablast subtype. (B) Pathway and function association network of upregulated genes 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).

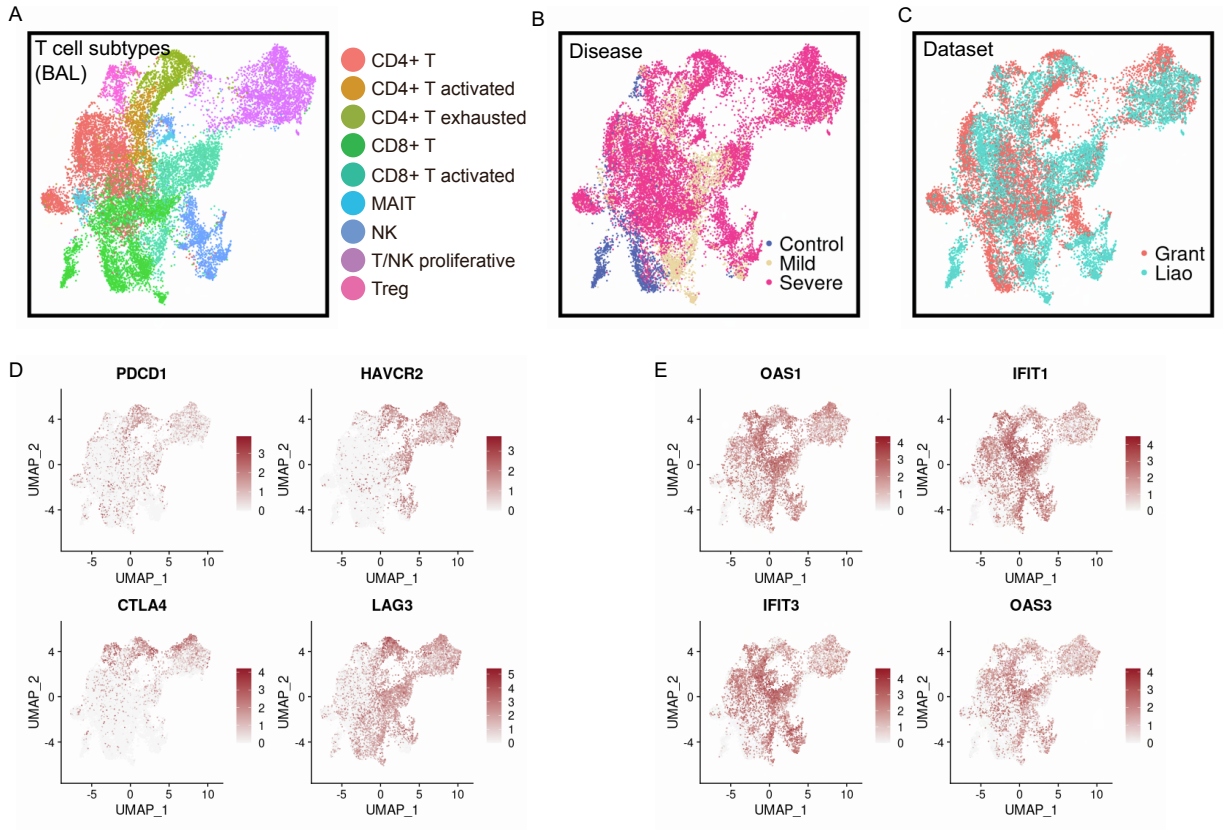

Figure S16. Distinct subtypes of T cells and NK cells in COVID-19 BAL data. (A-C) UMAPs of subtypes (A), COVID-19 conditions (B) and data sources (C) of T cells and NK cells in the integrated BAL data. (D-E) UMAPs of normalized expression values of exhausted T cell markers (D) and ISGs (E).

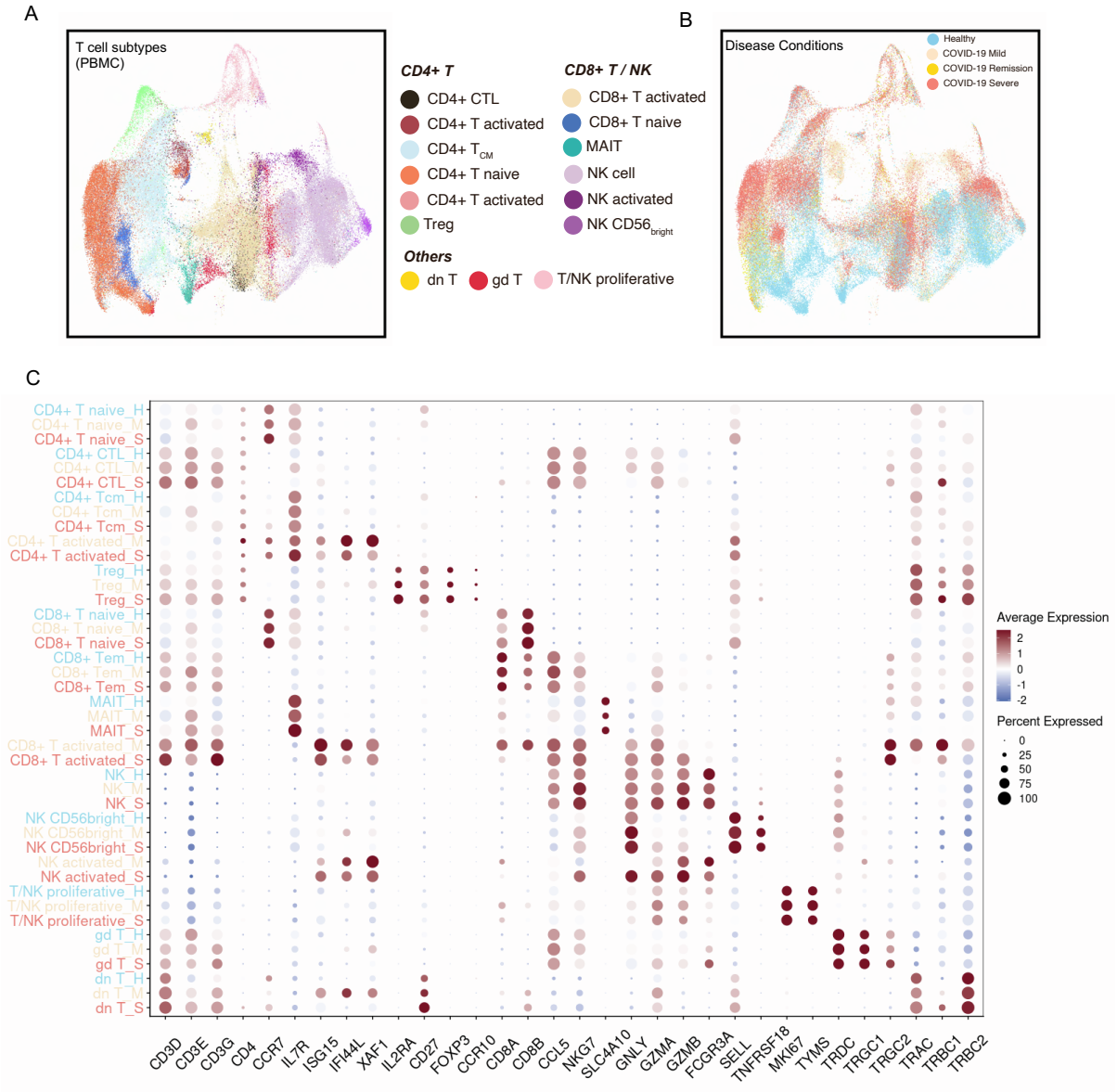

Figure S17. Various T cell and NK cell subtypes in the integrated PBMC data. (A-B) UMAPs of T cell and NK cell subtypes (A) and COVID-19 conditions (B) after integration of T cells in 5 PBMC single-cell datasets. (C) Dot plot shows T cell and NK cell subtype associated genes for each subtype per disease 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.

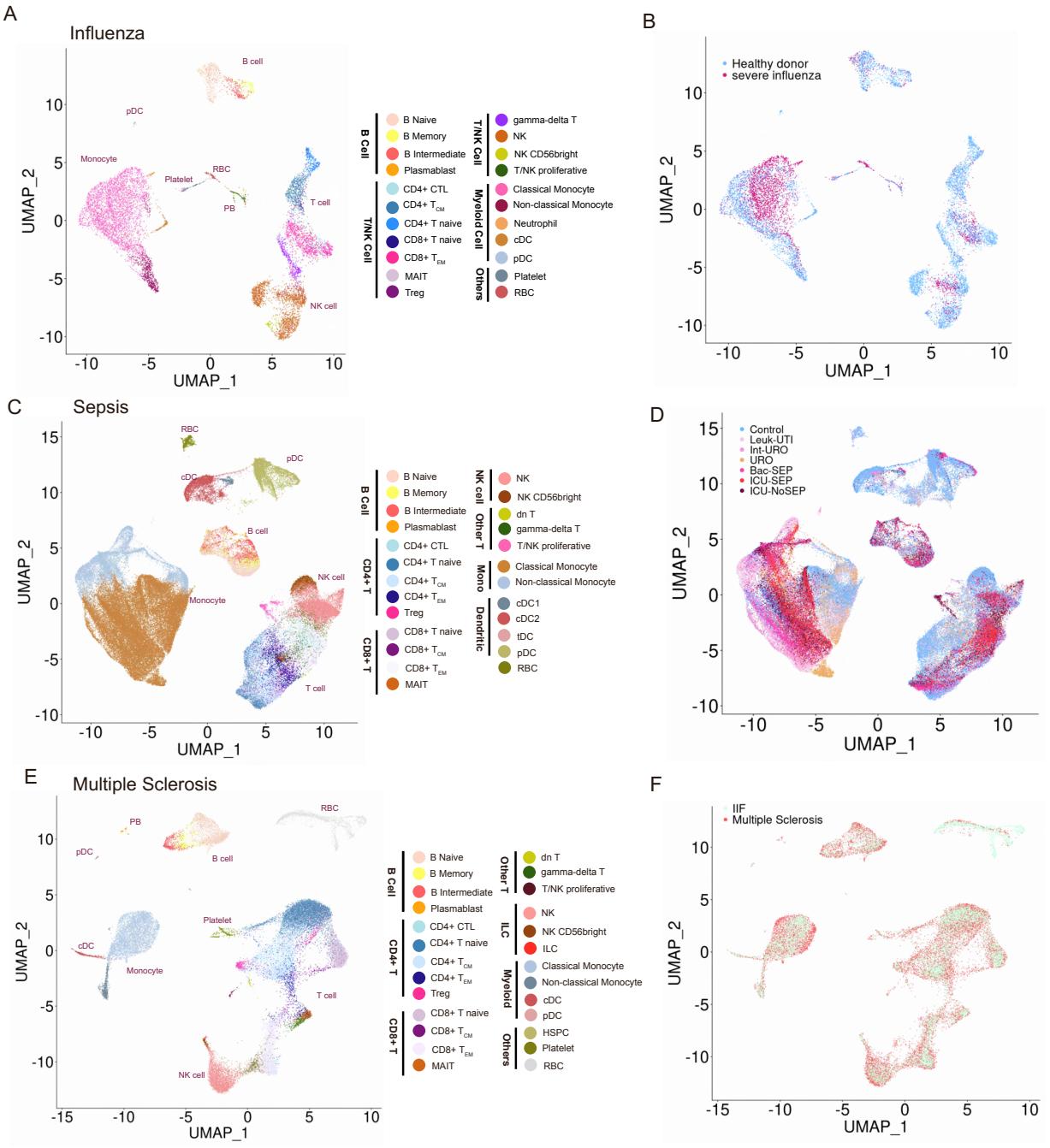

Figure S18. Various cell types in immune-mediated diseases, relative to Figure 7. (A, C, E) Distributions of cell types identified in influenza (A), sepsis (C) and multiple sclerosis (E) patients were shown on UMAPs. (B, D, F) Distributions of disease conditions in influenza (B), sepsis (D) and multiple sclerosis (F) patients were shown on UMAPs.