

1 **Implicating Gene and Cell Networks Responsible for**  
2 **Differential COVID-19 Host Responses via an Interactive**  
3 **Single Cell Web Portal**

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

8  
9 <sup>1</sup>Division of Biomedical Informatics, Cincinnati Children's Hospital Medical Center,  
10 Cincinnati, OH, 45229, USA

11 <sup>2</sup>Department of Biomedical Informatics, University of Cincinnati, Cincinnati, OH, 45229,  
12 USA

13 <sup>3</sup>Pediatric Rehabilitation, Dana-Dwek Children's Hospital, Tel Aviv Medical Center, Tel  
14 Aviv, 6423906, Israel

15 <sup>4</sup>Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, 6997801, Israel

16 <sup>5</sup>Department of Electrical Engineering and Computer Science, University of Cincinnati,  
17 Cincinnati, OH, 45221, USA

18 <sup>6</sup>Department of Neurology and Rehabilitation Medicine, University of Cincinnati College  
19 of Medicine, Cincinnati, OH, 45267, USA.

20 <sup>7</sup>Division of Allergy and Immunology, Department of Pediatrics, Cincinnati Children's  
21 Hospital Medical Center, University of Cincinnati, Cincinnati, OH, 45229, USA

22 <sup>8</sup>Department of Pediatrics, University of Cincinnati School of Medicine, Cincinnati, OH,  
23 45256, USA

24 <sup>9</sup>Lead contact

25 \*Correspondence: [bruce.aronow@cchmc.org](mailto:bruce.aronow@cchmc.org) (B.A.)

26

27

28

29

30

31

## 32 Summary

33 Numerous studies have provided single-cell transcriptome profiles of host responses to  
34 SARS-CoV-2 infection. Critically lacking however is a datamine that allows users to  
35 compare and explore cell profiles to gain insights and develop new hypotheses. To  
36 accomplish this, we harmonized datasets from COVID-19 and other control condition  
37 blood, bronchoalveolar lavage, and tissue samples, and derived a compendium of gene  
38 signature modules per cell type, subtype, clinical condition, and compartment. We  
39 demonstrate approaches to probe these via a new interactive web portal  
40 (<http://toppcell.cchmc.org/> COVID-19). As examples, we develop three hypotheses: (1)  
41 a multicellular signaling cascade among alternatively differentiated monocyte-derived  
42 macrophages whose tasks include T cell recruitment and activation; (2) novel platelet  
43 subtypes with drastically modulated expression of genes responsible for adhesion,  
44 coagulation and thrombosis; and (3) a multilineage cell activator network able to drive  
45 extrafollicular B maturation via an ensemble of genes strongly associated with risk for  
46 developing post-viral autoimmunity.

## 47 Keywords

48  
49 COVID-19, SARS-CoV-2, single-cell RNA-seq, host-pathogen cell atlas, interactive  
50 datamining, bronchoalveolar lavage, systems biology, antiviral host defense, platelets,  
51 inflammatory thrombosis, autoimmune disorder.

52

53

54

55

56

## 57 Introduction

58  
59 COVID-19 clinical outcomes are variable. The poorer outcomes due to this infection are  
60 highly associated with immunological and inflammatory responses to SARS-Cov-2  
61 infection (Shi et al., 2020; Tay et al., 2020) and many recent single cell expression  
62 profiling studies have characterized patterns of immunoinflammatory responses among  
63 individuals, mostly during acute infection phases. Different studies have revealed a  
64 spectrum of responses that range from lymphopenia (Cao, 2020; Wang et al., 2020),  
65 cytokine storms (Mehta et al., 2020; Pedersen and Ho, 2020), differential interferon  
66 responses (Blanco-Melo et al., 2020; Hadjadj et al., 2020) and emergency myelopoiesis  
67 (Schulte-Schrepping et al., 2020; Silvin et al., 2020). However, a variety of obstacles  
68 limit the ability of the research and medical communities to explore and compare these  
69 studies to pursue additional questions and gain additional insights that could improve  
70 our understanding of cell type specific responses to SARS-Cov-2 infection and their  
71 impact on clinical outcome.

72  
73 Whereas many studies have focused on the peripheral blood mononuclear cells  
74 (PBMC) (Arunachalam et al., 2020; Guo et al., 2020; Lee et al., 2020; Schulte-  
75 Schrepping et al., 2020; Wilk et al., 2020a) due to ease of procurement, other studies  
76 have profiled airway locations via bronchoalveolar lavage (BAL) (Grant et al., 2020; Liao  
77 et al., 2020), nasopharyngeal swabs, and bronchial brushes (Chua et al., 2020).  
78 Additional sampling sites that could also be infected or affected have also been  
79 approached in autopsy-derived materials from the central nervous system (Heming et  
80 al., 2021; Yang et al., 2020), and other sites (Delorey et al., 2021). Moreover, as major  
81 COVID-19 consortiums working on the collection and integration of each of their  
82 individual studies and interpreting important features of these individual datasets as  
83 downloadable datasets or browsable versions, such as single cell portal  
84 ([https://singlecell.broadinstitute.org/single\\_cell/covid19](https://singlecell.broadinstitute.org/single_cell/covid19)) and COVID-19 Cell Atlas  
85 (<https://www.covid19cellatlas.org/>), using these data beyond markers, cell types, and  
86 individual signatures is either not possible or not accomplishable across-datasets. Thus,  
87 a well-organized and systematic study of immune cells across tissues for in-depth

88 biological explorations is an unmet need for a deeper understanding of the underlying  
89 basis of the breadth of COVID-19 host defense and pathobiology.

90

91 Here we harmonized and analyzed eight high quality publicly available single-cell RNA-  
92 seq datasets from COVID-19 and immunologically-related studies that in total covered  
93 more than 480,000 cells isolated from peripheral blood, bronchial alveolar lavage and  
94 lung parenchyma samples, and assembled an integrated COVID-19 atlas  
95 (<https://toppcell.cchmc.org/>). We established a framework for deriving, characterizing,  
96 and establishing reference gene expression signatures from these harmonized datasets  
97 using modular and hierarchical approaches based on signatures per class, subclass,  
98 and signaling/activation and clinical status per each sample group. Leveraging these  
99 gene expression signature modules, we demonstrate datamining approaches that allow  
100 for the identification of a series of fundamental disease processes: (1) an intercellular  
101 monocytic activation cascade capable of mediating the emergence of  
102 hyperinflammatory monocyte-derived alveolar macrophages in severe COVID-19  
103 patients; (2) the generation of several alternatively differentiated platelet subtypes with  
104 dramatically different expression of sets of genes associated with critical platelet tasks  
105 capable of altering vascular and tissue responses to infectious agents; and (3) a  
106 multilineage and multi cell type cooperative signaling network with the potential to drive  
107 extrafollicular B maturation at a lesion site, but do so with high risk for the development  
108 of B cell-associated immunity. Additionally, immune hallmarks of COVID-19 patients  
109 were compared with other immune-mediated diseases using single-cell data from  
110 patients with influenza, sepsis, or multiple sclerosis. Consistent and varied  
111 compositional and gene patterns were identified across these implicating striking  
112 COVID-19 effects in some individuals.

## 113 Results

114

### 115 **Creating the First COVID-19 Signature Atlas Using ToppCell Portal**

116 To have a comprehensive coverage of cells, we collated single-cell data of COVID-19  
117 patients from eight public datasets, which in total contains 231,800 PBMCs, 101,800

118 BAL cells and 146,361 lung parenchyma cells from donors: 43 healthy; 22 mild; 42  
119 severe; and 2 convalescent patients (**Figure 1A, Table S1**).

120

121 In order to assemble an integrated atlas of human cell responses to COVID-19, we  
122 sought to harmonize metadata encompassing clinical information, sampling  
123 compartments, and cell and gene expression module designations. Doing so provides a  
124 rich framework for detecting perturbations of cell repertoire and differentiative state  
125 adaptations. We first integrated single cell RNA-seq data in Seurat (Stuart et al., 2019)  
126 and annotated cell types using canonical markers (**Table S2**). Further annotations of B  
127 cell and T cell subtypes were completed using the reference-based labeling tool  
128 Azimuth (Hao et al., 2020). Sub-clustering was applied for some cell types, such as  
129 neutrophils and platelets, to interrogate finer resolutions of disease-specific sub-  
130 populations (**Figure 1B**). Using the ToppCell toolkit (<https://toppcell.cchmc.org/>), we  
131 created over 3,000 hierarchical gene modules of the most significant differentially  
132 expressed genes (DEGs) for all cell classes and sub-clusters across compartments and  
133 disease severity (**Table S1**). These modules were then used to infer cell-cell  
134 interactions as well as upregulated pathways, which were further combined for  
135 functional comparative analysis in a specific cell manner in ToppCluster (Kaimal et al.,  
136 2010) (**Figure 1B**), such as sub-clusters of platelets. Integration of ToppCluster output  
137 of cells from multiple compartments and disease conditions built pathogenic maps,  
138 highlighted by the coagulation map of COVID-19 (**Figure S12**). In addition, perturbation  
139 of cell abundance was evaluated either in one cell population, or in multiple populations  
140 across diseases. Taken together, we investigated cell abundance changes, severity-  
141 associated signatures, mechanisms of COVID-19 specific symptoms and unique  
142 features of COVID-19 as an immune-mediated disease (**Figure 1B**).

143

#### 144 **Dynamic Changes and Balance of COVID-19 Immune Repository in Blood and** 145 **Lung**

146 After the aforementioned cell annotation procedure, we identified 28 and 24 distinct cell  
147 types in PBMC and BAL respectively (**Figure 2A, 2C; Table S2**). Shifts of Uniform  
148 Manifold Approximation and Projection (UMAP) of cell type distributions were observed

149 in both compartments of mild and severe patients (**Figure 2A, 2C, S1A and 3A**). In  
150 PBMC, conventional dendritic cells (cDC), plasmacytoid dendritic cells (pDC) and non-  
151 classical monocytes displayed a prominent reduction in severe patients (**Figure 2B,**  
152 **S1C**), consistent with prior reports (He et al., 2020; Laing et al., 2020; Wilk et al.,  
153 2020a). In contrast, severe patients demonstrated dramatic expansion of neutrophils,  
154 especially immature stages (**Figure S1C, S2**). Integration with evoked pathways in the  
155 following analysis implicated that neutrophil expansion was likely the consequence of  
156 emergency myelopoiesis (Wilk et al., 2020b). Additionally, a general down-regulation of  
157 T cell and NK cell was observed, consistent with lymphopenia reported in clinical  
158 practices (Pedersen and Ho, 2020; Terpos et al., 2020) (**Figure S1C, S2**). However, the  
159 trend of T cell subtypes varies across studies and individuals, apart from proliferative T  
160 cells which have a dramatic increase in mild and severe patients (**Figure S2**). Notably,  
161 plasmablasts substantially increased in COVID-19 patients, and especially so in severe  
162 patients, suggesting upregulated antibody production (De Biasi et al., 2020) (**Figure 2B**  
163 **and S1C**). Expansion of platelets is another significant change observed in severe  
164 patients, possibly leading to immunothrombosis in the lung, which could be closely  
165 associated with the severity of the disease (Middleton et al., 2020; Nicolai et al., 2020)  
166 (**Figure 2B and S1C**).

167  
168 In samples obtained from patients' lungs, we observed the depletion of FABP4<sup>high</sup>  
169 tissue-resident alveolar macrophages (TRAM) and dramatic expansion of FCN1<sup>high</sup>  
170 monocyte-derived alveolar macrophages (MoAM) in severe patients (**Figure 2C, 2D**  
171 **and S3D**). Mild patients exhibited a moderate reduction of tissue-resident  
172 macrophages, but no evidence of aggregation of monocyte-derived macrophages  
173 (**Figure 2C, 2D, S3A and S3D**). Dynamic changes of these two subtypes suggest  
174 increased tissue chemoattraction (Merad and Martin, 2020) and potential damage of  
175 patients' lungs (McGonagle et al., 2020). In addition, neutrophils were only identified in  
176 severe patients in the integrated BAL data (**Figure 2C and S3A**), which might be  
177 related with neutrophil extracellular traps (NETs) in the lung (Barnes et al., 2020).  
178 However, more samples are required to draw a solid conclusion. We also noted  
179 conventional dendritic cells decreased in the severe patients, which is consistent with

180 the trend of the counterpart in PBMC data. Opposite to the change in PBMC, an  
181 expansion of plasmacytoid dendritic cells is observed in both mild and severe patients  
182 (**Figure 2D**). Other cell types, including T cell and NK cell in the BAL, also have  
183 converse changes of their counterparts in PBMC, which could be attracted by lung  
184 macrophages or epithelial cells after infection or damages (Chua et al., 2020) (**Figure**  
185 **2D and Figure S3D**). These changes were consistently observed in lung parenchyma  
186 samples from severe COVID-19 patients (**Figure S4**). With cells well-annotated in the  
187 integrated COVID-19 atlas, we drew a global heatmap for cells in both blood and lung  
188 using ToppCell gene modules (top 50 DEG in each module) of all identified cell classes.  
189 While there was conservation of gene patterns involved in healthy donors and severe  
190 COVID-19 patients, there were substantial differences most notably in myeloid cells  
191 (**Figure 2E**). Such hierarchically ordered ToppCell gene modules were broadly used in  
192 visualization, large-scale comparisons and fine-resolution investigations in the following  
193 analyses.

194

### 195 **Myeloid Cell Atlas: Functionally Distinct Neutrophils at Different Levels of** 196 **Maturation and Derailed Macrophages in the Lung**

197 Dysregulated myeloid cells have been reported as an important marker of severe  
198 COVID-19 patients (Schulte-Schrepping et al., 2020; Silvin et al., 2020). In order to gain  
199 a deeper and comprehensive understanding of these cells, we applied the sub-  
200 clustering strategy on the integrated data of key cell types, such as neutrophils and  
201 macrophages, and then generated gene modules for comparative functional analysis  
202 and interactome inference. We successfully identified 5 neutrophil sub-clusters after the  
203 integration of PBMC and BAL data, including 3 FCGR3B<sup>+</sup> mature sub-clusters and 2  
204 FCGR3B<sup>-</sup> immature sub-clusters (**Figure 3A and S5B; Table S3**). They're mainly from  
205 severe patients and their gene modules were generated and subjected to comparative  
206 functional enrichment using ToppCell and ToppCluster (**Figure 3C, 3D and S5A**). We  
207 identified proliferative neutrophils (referred to as pro-neutrophils and Neu4) and  
208 MMP8<sup>high</sup> precursor immature neutrophils (referred to as pre-neutrophils and Neu2)  
209 (**Figure 3A and S5B**) consistent with prior studies (Schulte-Schrepping et al., 2020).  
210 While immune response genes and pathways were barely activated in the immature

211 neutrophils, they displayed upregulation of granule formation pathways and NETosis-  
212 associated proteins, including ELANE, DEFA4 and MPO, especially in Neu4 (Schulte-  
213 Schrepping et al., 2020; Wilk et al., 2020b) (**Figure 3C, S5B and S5C**). Upregulated  
214 myeloid leukocyte mediated immunity in Neu2 suggests involvement of this cell type in  
215 anti-viral function (**Figure S5D**). Yet, the absence of cytokine and interferon response  
216 pathways suggests the lack of mature immune responses (**Figure 3D**). Notably,  
217 compared to mature neutrophils (Neu0 and Neu1) in the blood, the extravasated  
218 hyperinflammatory sub-cluster (Neu3) from BAL of severe patients shows  
219 extraordinarily high expression of interferon-stimulated genes, as well as prominent  
220 upregulation of productions and responses to cytokines and interferons (**Figure 3C, 3D,**  
221 **S5B and S5D**).

222  
223 MoAM and TRAM were two main macrophage types in the BAL (**Figure 2C**); both are  
224 known to have distinct roles in immune responses in the lung (Liao et al., 2020). As  
225 described above, five sub-clusters among the expanded COVID-19 patient-specific  
226 MoAM (**Figure 3B; Table S3**) were found, where the loss of HLA class II genes and  
227 elevation of interferon-stimulated genes (ISGs) were consistently observed (**Figure 3F**  
228 **and S6A**). Relative to MoAM3,4, MoAM1,2,5 displayed an upregulation of interferon  
229 responses and cytokine production (**Figure 3D and S6B; Table S3**), indicating their  
230 pro-inflammatory characteristics. Notably, MoAM5 shows dramatic upregulation of IL-6  
231 secretion and cytokine receptor binding activities (**Figure S7A-D**). However, cells in this  
232 sub-cluster were mainly from one severe patient (**Figure S3C**). We still need more data  
233 to fully understand such dramatic upregulation of IL-6 secretion in some severe  
234 patients. Similar to MoAM, we also identified two distinct groups of TRAM in BAL  
235 (**Figure 3B and S6B**), including quiescent TRAM (TRAM1 and TRAM2) and activated  
236 TRAM (TRAM3). The quiescent group was mainly from healthy donors with enriched  
237 pathways of ATP metabolism (**Figure 3D**), while the activated group from mild and  
238 severe patients displays upregulation of ISGs and cytokine signaling pathways (**Figure**  
239 **S6B; Table S3**). However, the magnitude of activation and inflammatory responses in  
240 TRAM3 is smaller than MoAM1,2,5. Not surprisingly, stronger antigen processing and  
241 presentation activities were observed in TRAM3 relative to MoAM1,2,5 (**Figure 3D and**



242 **S6B; Table S3**). Collectively, we concluded that tissue-resident macrophages were  
243 greatly depleted in severe patients as the front-line innate immune responders in the  
244 lung. Pro-inflammatory monocyte-derived macrophages infiltrate into the lung, leading  
245 to the cytokine storm and damage of the lung. Large amounts of infiltration of MoAM  
246 were not observed in mild COVID-19 patients, probably due to the controlled infection,  
247 which could explain milder lung damages in those patients.

248  
249 To develop an understanding of the interaction network in the lung microenvironment of  
250 severe COVID-19 patients, we focused on signaling ligands, receptors and pathways  
251 using ToppCell and CellChat (**Figure 3E, S8A, S8B**). Notably, basal cells, MoAMs,  
252 neutrophils and T cells all contributed to the cytokine, chemokine and interleukin  
253 signaling networks. Strikingly, severe patient specific MoAM2 shows the broadest  
254 upregulation of signaling ligands, including CCL2, CCL3, CCL7, CCL8, CXCL9, CXCL9,  
255 CXCL10, CXCL11, IL6, IL15 and IL27, suggesting its role as a signaling network hub  
256 that is distinct from the other major signaling ligand-expressing cells of BAL such as  
257 epithelial and other myeloid cell types such as TRAM3 and proliferating myeloid cells  
258 (**Figure S8A**). Among the MoAM2 top signaling molecules, attractants CXCL8, CXCL9  
259 and CXCL10 are known to target CXCR3 on T cells, suggesting their role is to stimulate  
260 migration of T cells to the epithelial interface and into BAL fluid (**Figure 3E**) (Liao et al.,  
261 2020). In addition, many of MoAM2's ligands have the potential to cause autocrine  
262 signaling activation via IL6-IL6R, IL1RN-ILR2, CCL7-CCR1, CCL2-CCR1 and CCL4-  
263 CCR1, indicating its active roles in self-stimulation and development, which further  
264 amplify the attraction and migration of T cells and other immune cells. Notably, CCR1  
265 was also expressed in activated TRAM3, but with a lower level. Although IL6 expression  
266 level is relatively low compared to other ligands in BAL data, substantial expression of  
267 IL6R was observed in MoAMs. The CCL and CXCL signaling pathways of neutrophils  
268 are less strong than MoAMs (**Figure S8B**), but they displayed high expression levels of  
269 CXCR1 and CXCR2, which binds with a large number of the chemokines from MoAM  
270 and epithelial cells (**Figure 3E**). In addition, neutrophils exhibit an extraordinarily high  
271 level of IL1B, which could potentially in turn activate macrophages (**Figure S8A and**  
272 **S8B**). TRAM3 also displayed a unique pattern of signaling molecules, with a substantial

273 level of CCL23 which could potentially attract MoAM by the interaction with CCR1.  
274 Secretion of CXCL3 and CXCL5 in TRAM3 towards CXCR2 could be a potential  
275 chemoattraction pathway for neutrophils. In turn, neutrophils could activate TRAM3 by  
276 secreting IL1B, which binds with IL1RAP. Additionally, CD4+ T cells could also activate  
277 TRAM3 by IL10-IL10RB interaction (**Figure 3E, S8A and S8B**).

278

279 In addition to neutrophils and macrophages, the upregulation of ISGs was observed in  
280 classical monocytes of both mild and severe patients (cMono3, cMono4), while the  
281 reduction of the MHC class II cell surface receptor HLA-DR genes was only observed in  
282 severe patients (cMono4) (**Figure S9**). In cDCs, polarization of interleukin secretion was  
283 observed in mild-patient and severe-patient specific clusters (**Figure S10F**).

284 Collectively, dynamic changes of marker genes, transcriptional profiles, signaling  
285 molecules and biological activities reveal the heterogeneity of myeloid cell sub-clusters  
286 across disease severity (**Figure S11C**). Pro-inflammatory gene expression was found in  
287 all major myeloid cell types, including cMono4, DC1, DC9 in PBMC and Neu3, DC10,  
288 MoAM1, MoAM2, MoAM5 in BAL of COVID-19 patients. The reduction of MHC class II  
289 (HLA-II) genes is a common feature of classical monocytes and macrophages in  
290 COVID-19 patients and implies impaired capacity to activate T cell adaptive immunity.

291

## 292 **COVID-19 Coagulation and Immunothrombosis Map**

293 Individuals severely affected during acute phase COVID-19 infection, and in particular  
294 those with significantly elevated risk of death, frequently demonstrate striking  
295 dysregulation of coagulation and thrombosis characterized by hypercoagulability and  
296 microvascular thromboses (endothelial aggregations of platelets and fibrin) and highly  
297 elevated D-dimer levels. Yet, COVID-19 does not lead to wide scale consumption of  
298 fibrinogen and clotting factors (Iba et al., 2020a; Levi et al., 2020; Middleton et al., 2020;  
299 Nicolai et al., 2020; Rapkiewicz et al., 2020). At present, we lack a molecular or cellular  
300 explanation of the underlying basis of this pathobiology (Aid et al., 2020; Middleton et  
301 al., 2020). To evaluate candidate effectors of this pathobiology, we used a list of genes  
302 associated with abnormal thrombosis from mouse and human gene mutation  
303 phenotypes and identified parenchymal lung sample endothelial cells and platelets in

304 PBMC as cell types highly enriched with respect to genes responsible for the regulation  
305 of hemostasis (**Figure S12**). Because platelet counts were greatly elevated in severe  
306 versus mild individuals, we further examined platelet gene expression signatures and  
307 cell type differentiation and identified six distinct platelet sub-clusters shared across all  
308 datasets after data integration (**Figure 4A and 4B**). Severe-patient-specific PLT0 is an  
309 interesting sub-cluster with elevated integrin genes, including ITGA2B, ITGB1, ITGB3,  
310 ITGB5, as well as thrombosis-related genes, such as SELP, HPSE, ANO6 and PF4V1.  
311 Antibodies against the latter are associated with thrombosis including adverse reactions  
312 to recent COVID-19 vaccine ChAdOx1 nCoV-19 (Schultz et al., 2021). In addition,  
313 upregulated pathways of hemostasis, wound healing and blood coagulation were also  
314 observed in PLT0 (**Figure S13A; Table S4**). Importantly, PLT2 is an inflammatory sub-  
315 cluster with an upregulation of ISGs and interferon signaling pathways, while PLT4 is  
316 highlighted by upregulated post-transcriptional RNA splicing activities (**Figure S13A**  
317 **and S13C**).

318

319 Severity-associated gene patterns were also identified by selecting coagulation-  
320 associated genes modules (**Figure 4C; Table S4**), indicating distinct coagulation  
321 activities across platelets. Apart from pan-platelet genes, we found dramatic  
322 upregulation of genes involved in platelet activation, fibrinogen binding and blood  
323 coagulation in platelets of severe COVID-19, including procoagulant heparanase  
324 (HPSE) (Osterholm et al., 2013), Anoctamin-6 (ANO6) (Swieringa et al., 2018), and  
325 selectin P (SELP) (Sparkenbaugh and Pawlinski, 2013) (**Figure 4C and 4D**).  
326 Heparanase is an endoglycosidase that cleaves heparan sulfate constituents, a major  
327 component of anti-coagulation glycocalyx on the surface of vascular endothelium  
328 (Edovitsky et al., 2006; Iba et al., 2020b). Upregulated heparanase was related to  
329 upregulation of cell-matrix adhesion and coagulation (**Figure 4D**). Thrombotic vascular  
330 damages could be caused by the degradation function of heparinase enriched in  
331 platelets of severe patients. Elevation of ANO6 is known to trigger phospholipid  
332 scrambling in platelets, resulting in phosphatidylserine exposure which is essential for  
333 activation of the clotting system (Heemskerk et al., 2013). In addition, other upregulated  
334 genes involved in coagulation-associated activities were also observed, including

335 wound healing, fibrinolysis, platelet aggregation and activation (**Figure 4D**), which likely  
336 collectively contribute to the clotting issue of severe COVID-19 patients.

337

### 338 **Emergence of Developing Plasmablasts and B Cell Association with** 339 **Autoimmunity**

340 Autoimmune disorders in COVID-19 patients such as immune thrombocytopaenic  
341 purpura (ITP) is now recognized as a known disease complication (Ehrenfeld et al.,  
342 2020; Fujii et al., 2020; Gruber et al., 2020; Rodríguez et al., 2020; Zhou et al., 2020).  
343 However, little is known about the molecular and cellular mechanism behind it. To  
344 examine this further, we integrated B cells and plasmablasts from both PBMC and BAL  
345 and conducted systematic analysis (**Figure 5A, 5B, S14A and S14B**). Several COVID-  
346 19 specific sub-clusters were identified in B cells, such as ISG<sup>high</sup> activated B cells  
347 (cluster 7) (**Figure 5A**). Importantly, activated B cells showed dramatic upregulation of  
348 interferon signaling pathways and cytokine productions (**Figure S15A**), indicating its  
349 anti-virus characteristics. Notably, plasmablasts were mainly observed in severe  
350 COVID-19 patients, where a group of proliferative cells was identified and labeled as  
351 developing plasmablasts (**Figure 5B**). In contrast, non-dividing plasmablasts displayed  
352 upregulation of immunoglobulin genes (IGHA1, IGHA2, IGKC), B cell markers (CD79A)  
353 (Mason et al., 1995), interleukin receptors (IL2RG) and type II HLA complex (HLA-DOB)  
354 (**Figure 5C; Table S5**). In addition, non-dividing plasmablasts showed unique isotypes  
355 of immunoglobulin (Ig) in sub-regions of UMAP, whereas developing plasmablasts  
356 displayed obscure Ig types (**Figure S14E and S14F**). Antibody production activities  
357 were upregulated in non-dividing plasmablasts based on gene enrichment analysis  
358 (**Figure S15A; Table S5**). Collectively, we inferred that non-dividing plasmablasts had  
359 definite immunoglobulin isotypes and were actively involved in immune responses  
360 towards COVID infection, while developing plasmablasts were less mature but highly  
361 proliferative to replenish the repertoire of plasma cells.

362

363 Since there are few clues of gene associations of autoimmunity in COVID-19, we  
364 brought up a hypothesis-driven, prior knowledge-based approach to discover and  
365 prioritize genes for the specific phenotype (**Figure 5D**). First, gene modules of B cells

366 and other cells in severe patients were collected and subjected to ToppGene for  
367 enrichment analysis. Then we queried autoimmunity-associated terms in the enriched  
368 output and identified associated genes. After that, we retrieved interaction pairs using  
369 ToppCluster and CellChat database (Jin et al., 2020). In the end, we identified genes  
370 that are not only involved in autoimmunity, but have a mediator role in the immune  
371 signaling network. Using this approach, we observed several candidate pairs of genes,  
372 including TNFSF13B-TNSRSF13, IL10-IL10RA, IL21-IL21RA, IL6-IL6R, CXCL13-  
373 CXCR5, CXCL12-CXCR4, CCL21-CCR7, CCL19-CCR7 and CCL20-CCR6 in severe  
374 patients, which were enriched for autoimmune diseases, such as autoimmune thyroid  
375 diseases, lupus nephritis, autoimmune encephalomyelitis (Aust et al., 2004; Hirota et  
376 al., 2007; Kuwabara et al., 2009; Lee et al., 2010; Steinmetz et al., 2008). Candidate  
377 cytokine and chemokine ligand genes were expressed in various cell types in PBMC  
378 and BAL, including IL21 and CXCL13 from exhausted T cells of BAL, CXCL12 from  
379 mesenchymal cells, IL6 and CCL21 from endothelial cells, CCL19 from cDC and  
380 CCL20, TNFSF13B, and TNFSF13 from lung macrophages (**Figure 5E; Table S5**).  
381 These interaction pairs have been linked with auto-immunity (Klimatcheva et al., 2015;  
382 Wong et al., 2010). In addition, we analyzed single-cell studies (Arazi et al., 2019;  
383 Zhang et al., 2019) of rheumatoid arthritis and lupus nephritis patients and found that  
384 high expression levels of the candidate receptors in B cells and ligands in other cells  
385 were also observed, such as CXCL13 in helper T cells and CXCR5 in B cells in both  
386 studies (**Figure S15C, S15D**). However, more evidence is still required to infer the  
387 association between these interactions and autoimmunity in COVID-19 patients.  
388 Supported by the evidence above, we drew a network for potential mediator interactions  
389 of B cells and their associations with autoimmune disorders, where linkages with  
390 diseases, such as rheumatoid arthritis, systemic lupus erythematosus, were highlighted,  
391 as well as linkages with mouse phenotypes, such as abnormal immune tolerance and  
392 increased susceptibility to autoimmune disorder (**Figure 5F**). As a caveat, although  
393 using prior knowledge to prioritize gene and cell-associated functions and interactions  
394 may introduce biases, such approaches also have the potential to highlight key  
395 similarities and differences between different disease causes and clinical responses and  
396 improve our understanding of the molecular and cellular mechanisms at work.

397

## 398 **Functional Map and Immune Cell Interplay Landscape in COVID-19**

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

408

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

420

## 421 **Similarity and Heterogeneity Between COVID-19 and Other Immune-mediated** 422 **Diseases**

423 To further analyze COVID-19 specific immune signatures, we compared immune cells  
424 from COVID-19 patients with cells in other immune-mediated diseases, including severe  
425 influenza (Lee et al., 2020), sepsis (Reyes et al., 2020) and multiple sclerosis (Schafflick  
426 et al., 2020). 404,125 cells were included after the integration of PBMC single-cell  
427 datasets (**Figure 7A, S18; Table S7**). Dynamic changes of cell abundance were

428 compared in diseases versus healthy donors. Similar to COVID-19 patients, severe  
429 influenza patients also exhibited the reduction of non-classical monocytes, pDC, cDC  
430 and CD4+ TCM, but the effect of the former two types was smaller in magnitude (**Figure**  
431 **7B**). However, the reduction of non-classical monocytes is more significant in severe  
432 COVID-19 patients than severe influenza or mild COVID-19 patients (**Figure 7B**).  
433 Notably, NK cell reduction is associated with COVID-19 severity, whereas T cell  
434 depletion is a more dramatic perturbation in severe influenza. Within these  
435 comparisons, the expansion of plasmablasts is consistently observed, whereas the  
436 accumulation of platelets is unique to SARS-CoV-2 and in particular, to severe COVID-  
437 19 clinical status (**Figure 7B**).

438  
439 In addition to dynamic changes of cell ratios, we also investigated the regulation of  
440 immune mediator genes across various diseases (**Figure 7C; Table S7**). IL-6 is an  
441 important factor of cytokine storms in COVID-19 (Zhao, 2020). As shown in the  
442 heatmap, naive B cells are the main sources of IL-6 in COVID-19 patients while CD14+  
443 monocytes show the highest expression levels in severe influenza patients (**Figure 7C**).  
444 Specific ligands, including CXCL2, CXCL3, CCL20 were upregulated in both severe  
445 COVID-19 patients and severe influenza patients. CCR4 and IL2RA is uniquely high in  
446 CD4+ T cells of COVID-19 patients. Interestingly, most PBMC myeloid cell types  
447 displayed upregulated levels of interferon-stimulated genes in both COVID-19 and  
448 influenza, especially in COVID-19, where highest levels of ISGs in CD14+ Monocytes,  
449 cDC and pDC were observed.

450

451

## 452 Discussion

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

461  
462 Leveraging this approach, we identified three major candidate mechanisms capable of  
463 driving COVID-19 severity: (1) a cascade-like network of proinflammatory autocrine and  
464 paracrine ligand receptor interactions among subtypes of differentiating mononuclear,  
465 lymphoid, as well as other cell types; (2) the production of emergency platelets whose  
466 gene expression signatures implicate significantly elevated potential for adhesion,  
467 thrombosis, attenuated fibrinolysis, and potential to enhance the release of heparin-  
468 bound cytokines as well as further influence the activation of neutrophils causing further  
469 inflammatory cell recruitment and neutrophil netosis; and (3) the extrafollicular activation  
470 of naive and immature B cells via a multilineage network that includes monocytic  
471 subtypes and exhausted T cells of cytokines and interleukins with the potential to  
472 generate local antigen specific response to virus infected targets and collateral  
473 autoimmunity. More details will be discussed below.

474  
475 We identified dramatically expanded macrophages which were marked by the loss of  
476 HLA class II genes and upregulation of interferon-stimulated genes. It implicates a key  
477 role for these activated macrophages involved in signaling network and less so in  
478 activation of adaptive T cell immunity. Among them, MoAM2 displayed  
479 hyperinflammatory responses and extraordinary high levels of signaling molecules,  
480 which are involved in both autocrine (e.g. IL-6, CCL2, CCL4 and CCL8) and paracrine  
481 (e.g. CXCL2, CXCL9, CXCL10 and CXCL11) signaling pathways. The former pathway



482 contributed to the self-stimulation and development, which amplified the paracrine  
483 pathway for T cell and neutrophil chemoattraction. The latter two cell types in turn  
484 activated MoAMs with cytokines genes (CCL5, IL10 of T cells and IL1B of neutrophils,  
485 respectively). Based on the intercellular and multifactor complexity of the signaling  
486 cascade we have outlined, to effectively control a malignant inflammatory cascade, it  
487 may be essential to consider simultaneously targeting multiple nodes of this network of  
488 cytokines and interleukins. In addition, HLA-DR<sup>low</sup> monocytes, likely reflecting  
489 dysfunctional cells, were observed in severe infection. This, along with evidence of  
490 emergency myelopoiesis with immature circulating neutrophils into the circulation was  
491 detected in severe COVID-19. These neutrophils had transcriptional programs  
492 suggestive of dysfunction and immunosuppression not seen in patients with mild  
493 COVID-19. As such, we have presented evidence for the contribution of defective  
494 monocyte activation and dysregulated myelopoiesis to severe COVID.

495

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

512

513 We propose a signaling network which potentially shapes the differentiation of B cells  
514 towards the formation of autoantibodies. Proliferation and activation of inflammatory  
515 myeloid cells and the formation of exhausted CD4+ T helper around an area of direct or  
516 indirect viral tissue injury leads to the production of a set of interleukins and cytokines  
517 known to have both direct cell activating and maturing effects on naïve and immature B  
518 cells. Previous report had revealed the exaggerated extrafollicular B cell response,  
519 which is part of a mechanism that stimulates somatic mutation and maturation of B cells  
520 to produce plasma cells with specificity for antigens present in the vicinity of tissue  
521 damage sites (Farris and Guthridge, 2020). In the absence of macrophages or dendritic  
522 cells to restrict self vs non-self, the presence of IL-10, IL-21, CXCL13 CXCL10, IL-6 and  
523 others acting on receptors present in naïve and immature B cells leads to the selection  
524 and maturation of self-reactive maturation of B cells clones with formation of  
525 autoantibodies. Many of these COVID-19-activated genes (e.g. CXCL13, CCL19,  
526 CCL20, TNFRSF13) are known to be genetically associated with rheumatoid arthritis,  
527 lupus, and risk of developing autoimmune disease in humans and mouse models. The  
528 development of different patterns of autoimmunity may be the main hallmark of “Long  
529 Haul” Covid disease and could explain why some individuals develop different  
530 autoantibodies and suffer different forms of clinical consequences depending on which  
531 antigens drive the B-cell maturation. Thus, an additional prediction that could be made  
532 based on these findings and our network model is that among individuals treated with  
533 corticosteroids at the time these auto-immunogenic processes are activated, there  
534 should be a protective effect and lower likelihood of developing post acute sequela of  
535 Covid.

536

537 Consistent and varied compositional changes and gene patterns of immune cells were  
538 identified in COVID-19, influenza and sepsis. Expansion of plasmablasts, as well as the  
539 reduction of non-classical monocytes, are more significant changes in severe COVID-19  
540 patients, while the depletion of T cells is more dramatic in severe influenza patients. The  
541 accumulation is a unique immune hallmark of COVID-19 within the selected diseases,  
542 which contributes to the coagulation abnormalities and thrombosis, a key cause of

543 fatality in COVID-19 patients. Different signaling gene patterns were identified across  
544 immune-mediated diseases, with CCR4 only highly expressed in CD4+ T cells of  
545 COVID-19 patients, which might be related with extravasation of these cells (Spoerl et  
546 al., 2021). Upregulated interferon-stimulated genes of myeloid cells in PBMC revealed  
547 the inflammatory environment of COVID-19.

548

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

561

562 There are several limitations in our study. Different studies used various standards of  
563 COVID-19 severity definition. To generalize conclusions, we simplified disease  
564 conditions into several universal groups. Prospectively, a standardized definition of  
565 disease stages will assist to the accuracy of future studies. Additionally, the timing of  
566 sample collection was not considered as a variable in this study, rather disease stages  
567 were used to consolidate data across samples. We lack follow-up data of patients with  
568 sequela, which will be helpful for understanding the long-haul effects of the disease.

569

570

571

572

573

## 574 Acknowledgements

575 We thank Pablo Garcia-Nieto, Ambrose Carr and Jonah Cool and the Chan Zuckerberg  
576 Initiative for hosting the data on cellxgene. We acknowledge suggestions and help from  
577 Greta Beekhuis. Some figures were created using <https://biorender.com>. Funding for  
578 this study was provided by LungMap (U24 and HL148865), Digestive Health Center  
579 (P30, DK078392) and Harold C. Schott Foundation funding of the Harold C. Schott  
580 Endowed Chair, UC College of Medicine. We thank the support from Pediatric Cell Atlas  
581 and high performance computational cluster of CCHMC.

582

583

## 584 Author Contributions

585

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

590

## 591 Declaration of Interests

592 The authors declare no competing interests.

593

## 594 Resource availability

### 595 **Lead contact**

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

598

### 599 **Materials Availability**

600 This study did not generate new unique reagents.

601

### 602 **Data and Code Availability**

603 Public single-cell RNA-seq datasets of PBMC in COVID-19 patients are available on  
604 NCBI Gene Expression Omnibus and European Genome-phenome Archive, including  
605 [GSE150728](#), [GSE155673](#), [GSE150861](#), [GSE149689](#) and EGAS00001004571 (or  
606 [FastGenomics](#)). BAL single-cell RNA-seq datasets of COVID-19 patients are available  
607 on [GSE145926](#) and [GSE155249](#). Lung Parenchyma single-cell RNA-seq data are  
608 available on [GSE158127](#).  
609 Single-cell RNA-seq data of sepsis patients are available on the Single Cell Portal  
610 [SCP548](#) and [SCP550](#). Data of multiple sclerosis patients are available on [GSE128266](#).  
611 Data of severe influenza patients are available on [GSE149689](#).  
612 Gene modules of all datasets analyzed using ToppCell web portal are available on  
613 COVID-19 Atlas in [ToppCell](#), including gene modules from either a single dataset or an  
614 integrated dataset. Gene modules from the integration of specific cell types, such as B  
615 cells and neutrophils are also listed in ToppCell. More details are listed in Figure1A and  
616 Table S1. An interactive interface of integrated PBMC data and subclusters of immune  
617 cells will be public on [cellxgene](#).  
618 Codes of preprocessing, normalization, clustering and plotting of single-cell datasets will  
619 be available on [github](#).

620

## 621 Methods

622

### 623 **Single-cell RNA-seq data source**

624 To have a comprehensive understanding of immune cells in different repertoires, we  
625 collected 8 public COVID-19 single-cell RNA-seq datasets of multiple compartments,  
626 including peripheral blood mononuclear cells, bronchoalveolar lavage and lung biopsy,  
627 which in total covered over 43 healthy donors, 22 mild/moderate, 42 severe and 2  
628 convalescent COVID-19 patients. More details can be found in Figure 1A and Table S1.  
629 Lung biopsy samples were taken from the explanted lung or post-mortem lungs of  
630 COVID-19 patients(Bharat et al., 2020). Various criteria were used in these publications  
631 to describe COVID-19 severity. For example, we found asymptomatic, mild, moderate  
632 and floor COVID-19 patients under the definition of non-severe COVID-19 patients in  
633 our data sources. A recent paper used the WHO score of COVID-19 severity to

634 categorize disease conditions of patients(Wilk et al., 2020b), which is a more  
635 standardized and robust approach for the description of disease stages. However, in  
636 order to address the issue of missing information for disease stratification and to  
637 simplify the comparison, we grouped disease conditions into three groups, including  
638 healthy donors, mild COVID-19 patients and severe COVID-19 patients. Convalescent  
639 patients were excluded in some of our analysis for simplification. Sequencing data of  
640 healthy donors in Guo et al. was excluded since it was not from the same institute(Guo  
641 et al., 2020).

642 We also collected PBMC single-cell RNA-seq data from 29 sepsis patients(Reyes et al.,  
643 2020) and 4 multiple sclerosis(Schafflick et al., 2020) patients for comparative analysis  
644 of immune-mediated diseases (**Figure 1A; Table S1**). Data sources can be found in  
645 Data Availability.

646

#### 647 **Data preprocessing and normalization**

648 For datasets with raw UMI counts, we first removed cells with less than 300 detected  
649 genes or less than 600 UMI counts. Then cells with more than 15% counts of  
650 mitochondrial genes were filtered out. Genes expressed in less than 5 cells were  
651 removed. After quality control, we finally harvested 483,765 high-quality cells from 8  
652 studies (**Table S1**). We normalized the total UMI counts per gene to 1 million (CPM)  
653 and applied  $\log_2(\text{CPM}+1)$  transformation for heatmap visualization and downstream  
654 differential gene expression analysis. Steps above were done in Scanpy(Wolf et al.,  
655 2018).

656 For some datasets that only provide processed and normalized *h5ad* or *rd*s files, we  
657 checked their preprocessing procedures in the original publications and confirmed that  
658 stringent quality control procedures were used. Most of them used the default  
659 normalization approach in the Seurat or Scanpy pipeline. We transferred them to  
660  $\log_2(\text{CPM}+1)$  to make data consistently normalized. We also prepared corresponding  
661 raw count files for data integration.

662

#### 663 **Integration of PBMC datasets and BAL datasets using Reciprocal PCA in Seurat**

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

681

### 682 **Cell Annotations using canonical markers after unsupervised clustering**

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

695

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

705

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

711

## 712 **Cell Annotations using Azimuth**

713 To better annotate T cells in our study, we applied Azimuth  
714 (<https://satijalab.org/azimuth/>), a tool for reference-based single-cell analysis developed  
715 in Seurat version 4.0(Hao et al., 2020). High-quality PBMC single-cell data in Azimuth  
716 was used as the reference for label projection. After removing annotations with low  
717 prediction scores or low mapping scores, we got a collection of well-annotated T cell  
718 subtypes, including CD4+ Cytotoxic T cell, CD4+ Naive T cell, CD4+ Central Memory T  
719 cell, CD8+ Naive T cell, CD8+ Effector Memory cell, gamma-delta T cell, double-  
720 negative T cell. CD4+ Effector Memory T cell and CD8+ Central Memory T cell were  
721 found by Azimuth but removed later because of low scores. Apart from annotations of T  
722 cell subtypes, we also found CD56-bright NK cell, intermediate B cell and Memory B cell  
723 using Azimuth.

724

## 725 **Sub-clustering for specific cell types**



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

734

### 735 **Generation of ToppCell gene modules**

736 ToppCell (<https://toppcell.cchmc.org/>) was designed to parallelly analyze transcriptional  
737 profiles of single-cell datasets by organizing differential expressed gene modules in a  
738 customized hierarchical order. In our study, we hierarchically annotated cells with  
739 multiple layers, including compartments, disease conditions, lineages, cell classes and  
740 sub-clusters. All the cells were grouped into specific hierarchical categories. For  
741 example, “PBMC\_severe COVID-19\_myeloid cells\_classical-monocytes\_cMono1”  
742 represents cells belonging to cMono1 (a sub-cluster of classical monocytes) in PBMC of  
743 severe COVID-19 patients. With hierarchically ordered cell annotations, we calculated  
744 their DEGs in a hierarchical way as well. We defined customized ranges for  
745 comparisons and applied t-test based on normalized expression values. More details  
746 can be seen on ToppCell website. Usually, the top 200 most differentially genes in each  
747 comparison were picked up as the gene modules for the selected cell group, which are  
748 the starting point of downstream analysis, including gene enrichment in ToppGene and  
749 interaction inference in ToppCluster. All gene modules in our study were curated in  
750 COVID-19 Atlas (<https://toppcell.cchmc.org/biosystems/go/index3/COVID-19 Atlas>) and  
751 ImmuneMap (<https://toppcell.cchmc.org/biosystems/go/index3/ImmuneMap>) on the  
752 ToppCell website.

753

### 754 **Gene Enrichment Analysis using ToppGene**

755 Abundant gene modules were generated with ToppCell. After that, we used ToppGene  
756 (<https://toppgene.cchmc.org/>) for gene enrichment analysis. Genes in each gene

757 module were sent to ToppGene platform as input for enrichment in different domains.  
758 GO-Molecular Function, GO-Biological Process and GO-Cellular Component and  
759 Mouse Phenotype were usually used for enrichment. P values of enrichment results  
760 were adjusted using the Benjamini-Hochberg procedure.

761

### 762 **Generation of Functional Association Heatmap using ToppCluster**

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

773

### 774 **Cell Interaction Inference in immunothrombosis activities and cytokine signaling** 775 **pathways**

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

782

783 ToppCell was used to infer interactions in immunothrombosis. We first selected genes  
784 related to coagulation or immunothrombosis pathways from subtypes of endothelial  
785 cells, platelets, neutrophils, classical monocytes and monocyte-derived macrophages  
786 by filtering the output of ToppCluster (**Figure S12A**). Then we used CellChatDB as the  
787 knowledge base to find the subset of genes participating in cell-cell interaction, including

788 genes involved in signaling via secretion, cell-cell contact and extracellular matrix  
789 interaction. These genes in each cluster were sent to ToppCluster to infer the  
790 interaction network using protein-protein interactions (PPI) between those genes.

791

### 792 **Statistics of Cell Proportion Changes in Different Disease Stages**

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

797 To investigate the dynamic changes of cell proportions across various immune-  
798 mediated diseases, we followed the approach in recent literature (Lee et al., 2020)  
799 (**Figure 7B**). For each disease condition, we computed the relative ratio of each cell  
800 type in individual disease samples divided by individual healthy samples. Log<sub>2</sub>  
801 transformed values were shown in the box plot. Then we calculated relative ratios of  
802 each cell type between all sample pairs of healthy donors as a control. To compute the  
803 significance, we used a two-sided Kolmogorov-Smirnov (KS) test using relative ratios in  
804 diseases and those values in healthy donors.

805

### 806 **Generation of Volcano Plots**

807 We first calculated differential expressed genes using *tl.rank\_genes\_groups* in Scanpy.  
808 Adjusted p values and log fold changes in the output were used as the input of volcano  
809 plots. R package *EnhancedVolcano* (Blighe et al., 2018) was used to draw figures.

810

### 811 **Construction of COVID-19 Functional Enrichment Map**

812 In order to characterize functional properties of cell types and subtypes observed in  
813 BAL, PBMC, and lung parenchymal samples from control, mild, and severe COVID-19  
814 patient samples, we used the library of gene expression signatures (“Gene Module  
815 Report” from [ToppCell](#)) as an input to the ToppCluster enrichment analyzer web server  
816 (Kaimal et al 2010). Using categories of *Gene Ontology*, *Human Phenotype*, *Mouse*  
817 *Phenotype*, *Pathway* and *Protein Interaction*, a matrix was constructed using minus log  
818 P enrichment values for each celltype gene list and then all cells and enriched features

819 could be clustered and ordered based on their shared or distinct properties that could  
820 then be associated with lineage, cell subclass, tissue compartment, and disease state.

821

## 822 Supplementary Information

823 Table S1. Metadata of data sources and generated resources in this study, relative to  
824 Figure 1.

825

826 Table S2. Severity-associated cell distribution in multiple compartments of COVID-19  
827 patients, relative to Figure 2.

828

829 Table S3. Gene modules and DEGs of neutrophil and macrophage sub-clusters, relative  
830 to Figure 3.

831

832 Table S4. Gene modules and pro-thrombosis genes in platelets of COVID-19 patients,  
833 relative to Figure 4.

834

835 Table S5. Developing plasmablast signatures and autoimmunity-associated signatures  
836 of COVID-19 B cells, relative to Figure 5.

837

838 Table S6. Enrichment scores of COVID-19 Functional Map, relative to Figure 6.

839

840 Table S7. Dynamic changes of abundances and signatures across immune-mediated  
841 diseases, relative to Figure 7.

842

843

## 844 Reference

845 Aid, M., Busman-Sahay, K., Vidal, S.J., Maliga, Z., Bondoc, S., Starke, C., Terry, M.,  
846 Jacobson, C.A., Wrijil, L., Ducat, S., et al. (2020). Vascular Disease and Thrombosis in  
847 SARS-CoV-2-Infected Rhesus Macaques. *Cell* 183, 1354–1366.e13.

848 Arazi, A., Rao, D.A., Berthier, C.C., Davidson, A., Liu, Y., Hoover, P.J., Chicoine, A.,  
849 Eisenhaure, T.M., Jonsson, A.H., Li, S., et al. (2019). The immune cell landscape in

- 850 kidneys of patients with lupus nephritis. *Nat. Immunol.* *20*, 902–914.
- 851 Arunachalam, P.S., Wimmers, F., Mok, C.K.P., Perera, R.A.P.M., Scott, M., Hagan, T.,  
852 Sigal, N., Feng, Y., Bristow, L., Tak-Yin Tsang, O., et al. (2020). Systems biological  
853 assessment of immunity to mild versus severe COVID-19 infection in humans. *Science*  
854 *369*, 1210–1220.
- 855 Aust, G., Sittig, D., Becherer, L., Anderegg, U., Schütz, A., Lamesch, P., and  
856 Schmücking, E. (2004). The role of CXCR5 and its ligand CXCL13 in the  
857 compartmentalization of lymphocytes in thyroids affected by autoimmune thyroid  
858 diseases. *Eur. J. Endocrinol.* *150*, 225–234.
- 859 Barnes, B.J., Adrover, J.M., Baxter-Stoltzfus, A., Borczuk, A., Cools-Lartigue, J.,  
860 Crawford, J.M., Daßler-Plenker, J., Guerci, P., Huynh, C., Knight, J.S., et al. (2020).  
861 Targeting potential drivers of COVID-19: Neutrophil extracellular traps. *J. Exp. Med.*  
862 *217*.
- 863 Bharat, A., Querrey, M., Markov, N.S., Kim, S., Kurihara, C., Garza-Castillon, R.,  
864 Manerikar, A., Shilatifard, A., Tomic, R., Politanska, Y., et al. (2020). Lung  
865 transplantation for patients with severe COVID-19. *Sci. Transl. Med.* *12*.
- 866 Blanco-Melo, D., Nilsson-Payant, B.E., Liu, W.-C., Uhl, S., Hoagland, D., Møller, R.,  
867 Jordan, T.X., Oishi, K., Panis, M., Sachs, D., et al. (2020). Imbalanced Host Response  
868 to SARS-CoV-2 Drives Development of COVID-19. *Cell* *181*, 1036–1045.e9.
- 869 Blighe, K., Rana, S., and Lewis, M. (2018). EnhancedVolcano: Publication-Ready  
870 Volcano Plots With Enhanced Colouring and Labeling. (2019). R Package Version 1.
- 871 Cao, X. (2020). COVID-19: immunopathology and its implications for therapy. *Nat. Rev.*  
872 *Immunol.* *20*, 269–270.
- 873 Chua, R.L., Lukassen, S., Trump, S., Hennig, B.P., Wendisch, D., Pott, F., Debnath, O.,  
874 Thürmann, L., Kurth, F., Völker, M.T., et al. (2020). COVID-19 severity correlates with  
875 airway epithelium–immune cell interactions identified by single-cell analysis. *Nat.*  
876 *Biotechnol.* *38*, 970–979.
- 877 De Biasi, S., Lo Tartaro, D., Meschiari, M., Gibellini, L., Bellinazzi, C., Borella, R.,  
878 Fidanza, L., Mattioli, M., Paolini, A., Gozzi, L., et al. (2020). Expansion of plasmablasts  
879 and loss of memory B cells in peripheral blood from COVID-19 patients with pneumonia.  
880 *Eur. J. Immunol.* *50*, 1283–1294.
- 881 Delorey, T.M., Ziegler, C.G.K., Heimberg, G., Normand, R., Yang, Y., Segerstolpe, A.,  
882 Abbondanza, D., Fleming, S.J., Subramanian, A., Montoro, D.T., et al. (2021). A single-  
883 cell and spatial atlas of autopsy tissues reveals pathology and cellular targets of SARS-  
884 CoV-2. *bioRxiv*.
- 885 Edovitsky, E., Lerner, I., Zcharia, E., Peretz, T., Vlodaysky, I., and Elkin, M. (2006). Role  
886 of endothelial heparanase in delayed-type hypersensitivity. *Blood* *107*, 3609–3616.

- 887 Ehrenfeld, M., Tincani, A., Andreoli, L., Cattalini, M., Greenbaum, A., Kanduc, D.,  
888 Alijotas-Reig, J., Zinserling, V., Semenova, N., Amital, H., et al. (2020). Covid-19 and  
889 autoimmunity. *Autoimmun. Rev.* *19*, 102597.
- 890 Farris, A.D., and Guthridge, J.M. (2020). Overlapping B cell pathways in severe COVID-  
891 19 and lupus. *Nat. Immunol.* *21*, 1478–1480.
- 892 Fujii, H., Tsuji, T., Yuba, T., Tanaka, S., Suga, Y., Matsuyama, A., Omura, A., Shiotsu,  
893 S., Takumi, C., Ono, S., et al. (2020). High levels of anti-SSA/Ro antibodies in COVID-  
894 19 patients with severe respiratory failure: a case-based review. *Clin. Rheumatol.* *39*,  
895 3171–3175.
- 896 Goodall, K.J., Poon, I.K.H., Phipps, S., and Hulett, M.D. (2014). Soluble heparan sulfate  
897 fragments generated by heparanase trigger the release of pro-inflammatory cytokines  
898 through TLR-4. *PLoS One* *9*, e109596.
- 899 Grant, R.A., Morales-Nebreda, L., and Markov, N.S. (2020). Alveolitis in severe SARS-  
900 CoV-2 pneumonia is driven by self-sustaining circuits between infected alveolar  
901 macrophages and T cells. *bioRxiv*.
- 902 Gruber, C.N., Patel, R.S., Trachtman, R., Lepow, L., Amanat, F., Krammer, F., Wilson,  
903 K.M., Onel, K., Geanon, D., Tuballes, K., et al. (2020). Mapping Systemic Inflammation  
904 and Antibody Responses in Multisystem Inflammatory Syndrome in Children (MIS-C).  
905 *Cell* *183*, 982–995.e14.
- 906 Guo, C., Li, B., Ma, H., Wang, X., Cai, P., Yu, Q., Zhu, L., Jin, L., Jiang, C., Fang, J., et  
907 al. (2020). Single-cell analysis of two severe COVID-19 patients reveals a monocyte-  
908 associated and tocilizumab-responding cytokine storm. *Nat. Commun.* *11*, 3924.
- 909 Hadjadj, J., Yatim, N., Barnabei, L., Corneau, A., Boussier, J., Smith, N., Péré, H.,  
910 Charbit, B., Bondet, V., Chenevier-Gobeaux, C., et al. (2020). Impaired type I interferon  
911 activity and inflammatory responses in severe COVID-19 patients. *Science* *369*, 718–  
912 724.
- 913 Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W.M., Zheng, S., Butler, A., Lee, M.J.,  
914 Wilk, A.J., Darby, C., Zagar, M., et al. (2020). Integrated analysis of multimodal single-  
915 cell data.
- 916 He, R., Lu, Z., Zhang, L., Fan, T., Xiong, R., Shen, X., Feng, H., Meng, H., Lin, W.,  
917 Jiang, W., et al. (2020). The clinical course and its correlated immune status in COVID-  
918 19 pneumonia. *J. Clin. Virol.* *127*, 104361.
- 919 Heemskerk, J.W.M., Mattheij, N.J.A., and J M E (2013). Platelet-based coagulation:  
920 different populations, different functions. *Journal of Thrombosis and Haemostasis* *11*, 2–  
921 16.
- 922 Heming, M., Li, X., Räuber, S., Mausberg, A.K., Börsch, A.-L., Hartlehnert, M., Singhal,  
923 A., Lu, I.-N., Fleischer, M., Szepanowski, F., et al. (2021). Neurological Manifestations  
924 of COVID-19 Feature T Cell Exhaustion and Dedifferentiated Monocytes in

- 925 Cerebrospinal Fluid. *Immunity* 54, 164–175.e6.
- 926 Hirota, K., Yoshitomi, H., Hashimoto, M., Maeda, S., Teradaira, S., Sugimoto, N.,  
927 Yamaguchi, T., Nomura, T., Ito, H., Nakamura, T., et al. (2007). Preferential recruitment  
928 of CCR6-expressing Th17 cells to inflamed joints via CCL20 in rheumatoid arthritis and  
929 its animal model. *J. Exp. Med.* 204, 2803–2812.
- 930 Iba, T., Levy, J.H., Connors, J.M., Warkentin, T.E., Thachil, J., and Levi, M. (2020a).  
931 The unique characteristics of COVID-19 coagulopathy. *Crit. Care* 24, 360.
- 932 Iba, T., Connors, J.M., and Levy, J.H. (2020b). The coagulopathy, endotheliopathy, and  
933 vasculitis of COVID-19. *Inflamm. Res.* 69, 1181–1189.
- 934 Jayatilleke, K.M., and Hulett, M.D. (2020). Heparanase and the hallmarks of cancer. *J.*  
935 *Transl. Med.* 18, 453.
- 936 Jin, S., Guerrero-Juarez, C.F., Zhang, L., Chang, I., Myung, P., Plikus, M.V., and Nie,  
937 Q. (2020). Inference and analysis of cell-cell communication using CellChat.
- 938 Kaimal, V., Bardes, E.E., Tabar, S.C., Jegga, A.G., and Aronow, B.J. (2010).  
939 ToppCluster: a multiple gene list feature analyzer for comparative enrichment clustering  
940 and network-based dissection of biological systems. *Nucleic Acids Res.* 38, W96–  
941 W102.
- 942 Klimatcheva, E., Pandina, T., Reilly, C., Torno, S., Bussler, H., Scrivens, M., Jonason,  
943 A., Mallow, C., Doherty, M., Paris, M., et al. (2015). CXCL13 antibody for the treatment  
944 of autoimmune disorders. *BMC Immunol.* 16, 6.
- 945 Kuwabara, T., Ishikawa, F., Yasuda, T., Aritomi, K., Nakano, H., Tanaka, Y., Okada, Y.,  
946 Lipp, M., and Kakiuchi, T. (2009). CCR 7 Ligands Are Required for Development of  
947 Experimental Autoimmune Encephalomyelitis through Generating IL-23-Dependent  
948 Th17 Cells. *The Journal of Immunology* 183, 2513–2521.
- 949 Laing, A.G., Lorenc, A., Del Molino Del Barrio, I., Das, A., Fish, M., Monin, L., Muñoz-  
950 Ruiz, M., McKenzie, D.R., Hayday, T.S., Francos-Quijorna, I., et al. (2020). Author  
951 Correction: A dynamic COVID-19 immune signature includes associations with poor  
952 prognosis. *Nat. Med.* 26, 1951.
- 953 Lee, H.-T., Shiao, Y.-M., Wu, T.-H., Chen, W.-S., Hsu, Y.-H., Tsai, S.-F., and Tsai, C.-Y.  
954 (2010). Serum BLC/CXCL13 concentrations and renal expression of CXCL13/CXCR5 in  
955 patients with systemic lupus erythematosus and lupus nephritis. *J. Rheumatol.* 37, 45–  
956 52.
- 957 Lee, J.S., Park, S., Jeong, H.W., Ahn, J.Y., Choi, S.J., Lee, H., Choi, B., Nam, S.K., Sa,  
958 M., Kwon, J.-S., et al. (2020). Immunophenotyping of COVID-19 and influenza  
959 highlights the role of type I interferons in development of severe COVID-19. *Sci*  
960 *Immunol* 5.
- 961 Levi, M., Thachil, J., Iba, T., and Levy, J.H. (2020). Coagulation abnormalities and

- 962 thrombosis in patients with COVID-19. *The Lancet Haematology* 7, e438–e440.
- 963 Liao, M., Liu, Y., Yuan, J., Wen, Y., Xu, G., Zhao, J., Cheng, L., Li, J., Wang, X., Wang,  
964 F., et al. (2020). Single-cell landscape of bronchoalveolar immune cells in patients with  
965 COVID-19. *Nat. Med.* 26, 842–844.
- 966 Mason, D.Y., Cordell, J.L., Brown, M.H., Borst, J., Jones, M., Pulford, K., Jaffe, E.,  
967 Ralfkiaer, E., Dallenbach, F., and Stein, H. (1995). CD79a: a novel marker for B-cell  
968 neoplasms in routinely processed tissue samples. *Blood* 86, 1453–1459.
- 969 McGonagle, D., Sharif, K., O'Regan, A., and Bridgewood, C. (2020). The Role of  
970 Cytokines including Interleukin-6 in COVID-19 induced Pneumonia and Macrophage  
971 Activation Syndrome-Like Disease. *Autoimmun. Rev.* 19, 102537.
- 972 Mehta, P., McAuley, D.F., Brown, M., Sanchez, E., Tattersall, R.S., Manson, J.J., and  
973 HLH Across Speciality Collaboration, UK (2020). COVID-19: consider cytokine storm  
974 syndromes and immunosuppression. *Lancet* 395, 1033–1034.
- 975 Merad, M., and Martin, J.C. (2020). Author Correction: Pathological inflammation in  
976 patients with COVID-19: a key role for monocytes and macrophages. *Nat. Rev.*  
977 *Immunol.* 20, 448.
- 978 Middleton, E.A., He, X.-Y., Denorme, F., Campbell, R.A., Ng, D., Salvatore, S.P.,  
979 Mostyka, M., Baxter-Stoltzfus, A., Borczuk, A.C., Loda, M., et al. (2020). Neutrophil  
980 extracellular traps contribute to immunothrombosis in COVID-19 acute respiratory  
981 distress syndrome. *Blood* 136, 1169–1179.
- 982 Nicolai, L., Leunig, A., Brambs, S., Kaiser, R., Weinberger, T., Weigand, M.,  
983 Muenchhoff, M., Hellmuth, J.C., Ledderose, S., Schulz, H., et al. (2020).  
984 Immunothrombotic Dysregulation in COVID-19 Pneumonia Is Associated With  
985 Respiratory Failure and Coagulopathy. *Circulation* 142, 1176–1189.
- 986 Osterholm, C., Folkersen, L., Lengquist, M., Pontén, F., Renné, T., Li, J., and Hedin, U.  
987 (2013). Increased expression of heparanase in symptomatic carotid atherosclerosis.  
988 *Atherosclerosis* 226, 67–73.
- 989 Pedersen, S.F., and Ho, Y.-C. (2020). SARS-CoV-2: a storm is raging. *J. Clin. Invest.*  
990 130, 2202–2205.
- 991 Rapkiewicz, A.V., Mai, X., Carsons, S.E., Pittaluga, S., Kleiner, D.E., Berger, J.S.,  
992 Thomas, S., Adler, N.M., Charytan, D.M., Gasmi, B., et al. (2020). Megakaryocytes and  
993 platelet-fibrin thrombi characterize multi-organ thrombosis at autopsy in COVID-19: A  
994 case series. *EClinicalMedicine* 24, 100434.
- 995 Reyes, M., Filbin, M.R., Bhattacharyya, R.P., Billman, K., Eisenhaure, T., Hung, D.T.,  
996 Levy, B.D., Baron, R.M., Blainey, P.C., Goldberg, M.B., et al. (2020). An immune-cell  
997 signature of bacterial sepsis. *Nat. Med.* 26, 333–340.
- 998 Rodríguez, Y., Novelli, L., Rojas, M., De Santis, M., Acosta-Ampudia, Y., Monsalve,

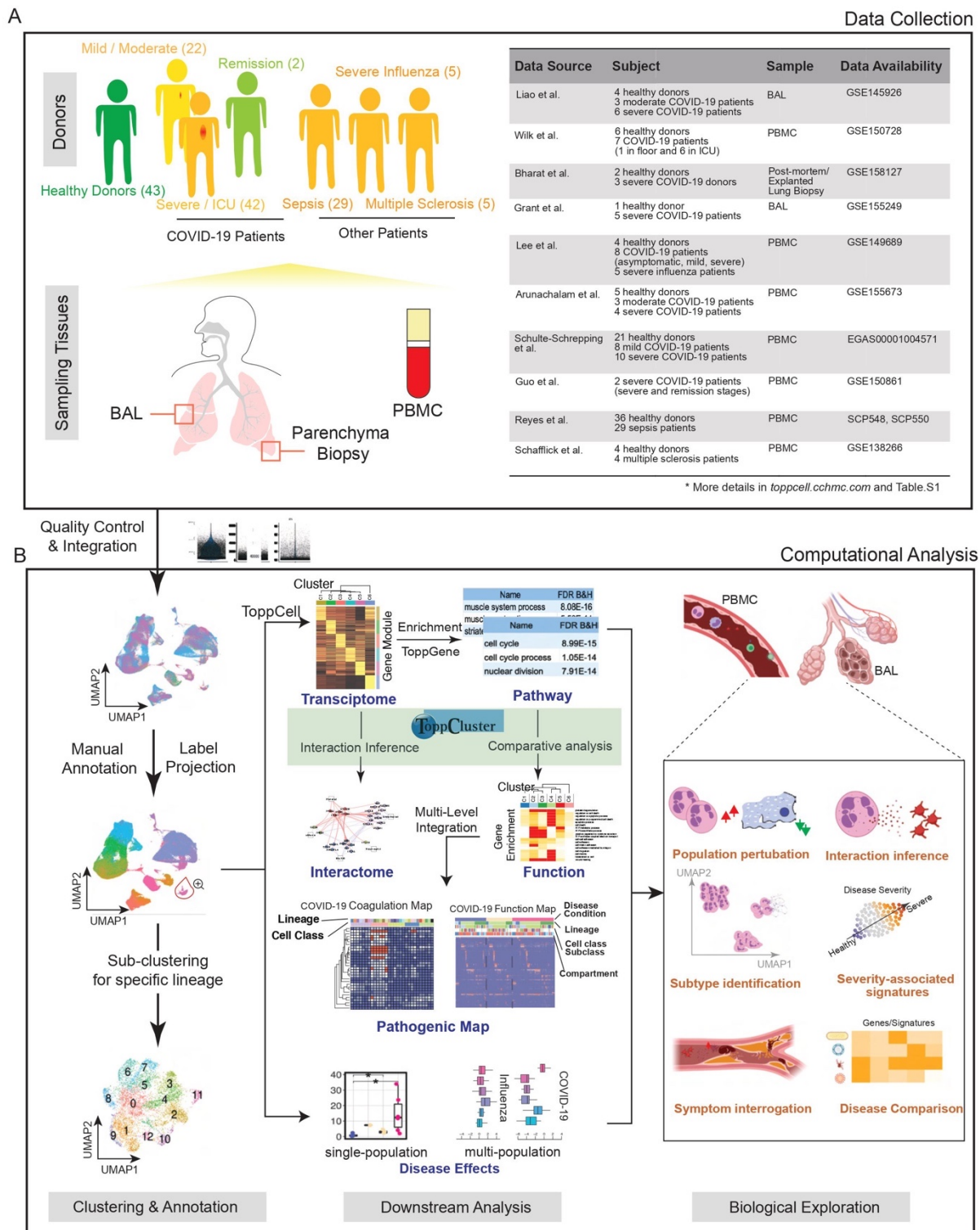


- 999 D.M., Ramírez-Santana, C., Costanzo, A., Ridgway, W.M., Ansari, A.A., et al. (2020).  
1000 Autoinflammatory and autoimmune conditions at the crossroad of COVID-19. *J.*  
1001 *Autoimmun.* *114*, 102506.
- 1002 Schafflick, D., Xu, C.A., Hartlehnert, M., Cole, M., Schulte-Mecklenbeck, A., Lautwein,  
1003 T., Wolbert, J., Heming, M., Meuth, S.G., Kuhlmann, T., et al. (2020). Integrated single  
1004 cell analysis of blood and cerebrospinal fluid leukocytes in multiple sclerosis. *Nature*  
1005 *Communications* *11*.
- 1006 Schulte-Schrepping, J., Reusch, N., Paclik, D., Baßler, K., Schlickeiser, S., Zhang, B.,  
1007 Krämer, B., Krammer, T., Brumhard, S., Bonaguro, L., et al. (2020). Severe COVID-19  
1008 Is Marked by a Dysregulated Myeloid Cell Compartment. *Cell* *182*, 1419–1440.e23.
- 1009 Schultz, N.H., Sørvoll, I.H., Michelsen, A.E., Munthe, L.A., Lund-Johansen, F., Ahlen,  
1010 M.T., Wiedmann, M., Aamodt, A.-H., Skattør, T.H., Tjønnfjord, G.E., et al. (2021).  
1011 Thrombosis and Thrombocytopenia after ChAdOx1 nCoV-19 Vaccination. *N. Engl. J.*  
1012 *Med.*
- 1013 Shi, Y., Wang, Y., Shao, C., Huang, J., Gan, J., Huang, X., Bucci, E., Piacentini, M.,  
1014 Ippolito, G., and Melino, G. (2020). COVID-19 infection: the perspectives on immune  
1015 responses. *Cell Death Differ.* *27*, 1451–1454.
- 1016 Silvin, A., Chapuis, N., Dunsmore, G., Goubet, A.-G., Dubuisson, A., Derosa, L., Almire,  
1017 C., Hénon, C., Kosmider, O., Droin, N., et al. (2020). Elevated Calprotectin and  
1018 Abnormal Myeloid Cell Subsets Discriminate Severe from Mild COVID-19. *Cell* *182*,  
1019 1401–1418.e18.
- 1020 Sparkenbaugh, E., and Pawlinski, R. (2013). Interplay between coagulation and  
1021 vascular inflammation in sickle cell disease. *Br. J. Haematol.* *162*, 3–14.
- 1022 Spoerl, S., Kremer, A.N., Aigner, M., Eisenhauer, N., Koch, P., Meretuk, L., Löffler, P.,  
1023 Tenbusch, M., Maier, C., Überla, K., et al. (2021). Upregulation of CCR4 in activated  
1024 CD8+ T cells indicates enhanced lung homing in patients with severe acute SARS-CoV-  
1025 2 infection. *Eur. J. Immunol.*
- 1026 Steinmetz, O.M., Velden, J., Kneissler, U., Marx, M., Klein, A., Helmchen, U., Stahl,  
1027 R.A.K., and Panzer, U. (2008). Analysis and classification of B-cell infiltrates in lupus  
1028 and ANCA-associated nephritis. *Kidney Int.* *74*, 448–457.
- 1029 Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., 3rd,  
1030 Hao, Y., Stoeckius, M., Smibert, P., and Satija, R. (2019). Comprehensive Integration of  
1031 Single-Cell Data. *Cell* *177*, 1888–1902.e21.
- 1032 Swieringa, F., Spronk, H.M.H., Heemskerk, J.W.M., and van der Meijden, P.E.J. (2018).  
1033 Integrating platelet and coagulation activation in fibrin clot formation. *Res Pract Thromb*  
1034 *Haemost* *2*, 450–460.
- 1035 Tay, M.Z., Poh, C.M., Rénia, L., MacAry, P.A., and Ng, L.F.P. (2020). The trinity of  
1036 COVID-19: immunity, inflammation and intervention. *Nat. Rev. Immunol.* *20*, 363–374.

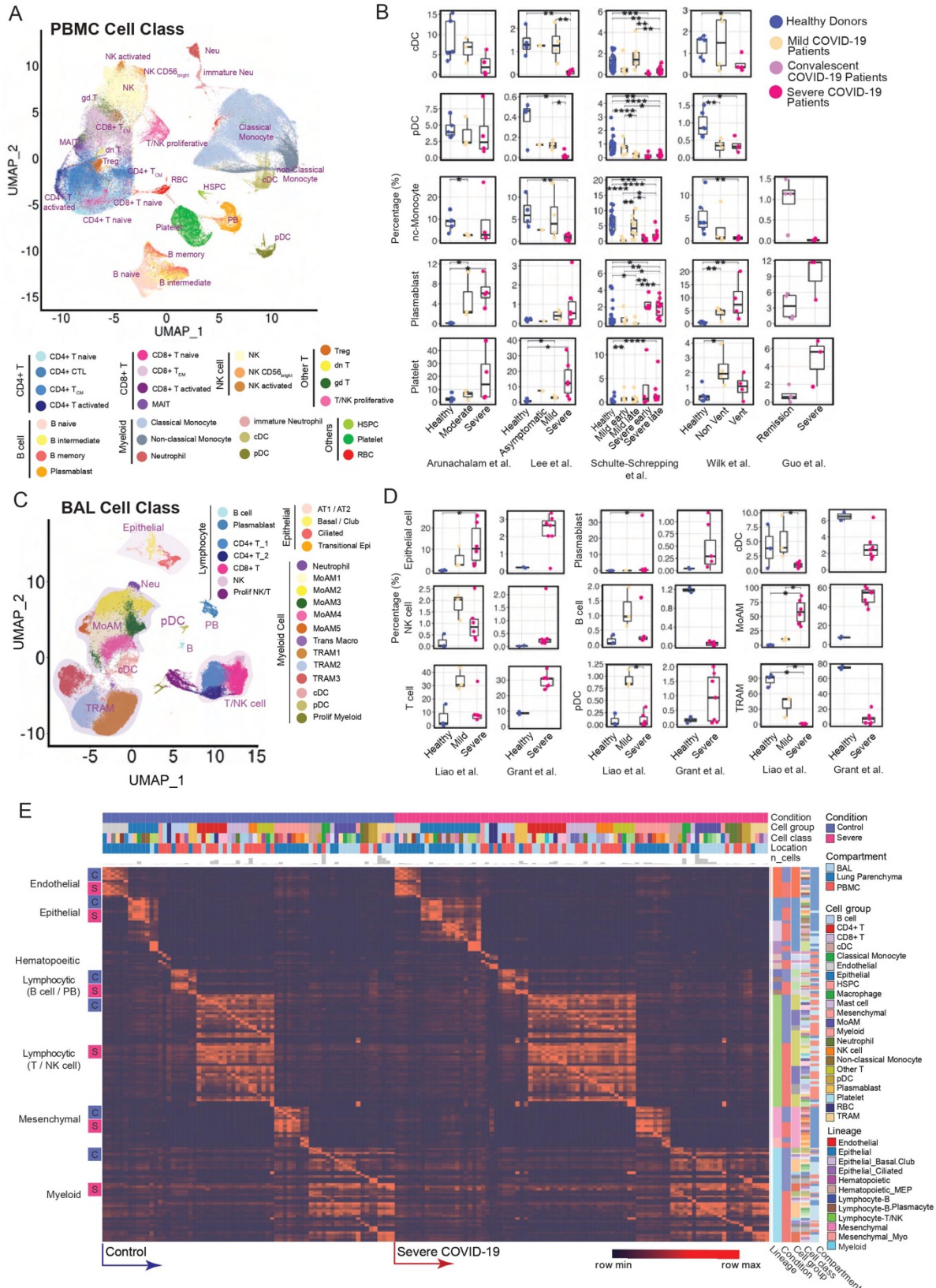
- 1037 Terpos, E., Ntanasis-Stathopoulos, I., Elalamy, I., Kastritis, E., Sergentanis, T.N.,  
1038 Politou, M., Psaltopoulou, T., Gerotziapas, G., and Dimopoulos, M.A. (2020).  
1039 Hematological findings and complications of COVID-19. *Am. J. Hematol.* *95*, 834–847.
- 1040 Thålin, C., Hisada, Y., Lundström, S., Mackman, N., and Wallén, H. (2019). Neutrophil  
1041 Extracellular Traps: Villains and Targets in Arterial, Venous, and Cancer-Associated  
1042 Thrombosis. *Arterioscler. Thromb. Vasc. Biol.* *39*, 1724–1738.
- 1043 Wang, F., Nie, J., Wang, H., Zhao, Q., Xiong, Y., Deng, L., Song, S., Ma, Z., Mo, P.,  
1044 and Zhang, Y. (2020). Characteristics of Peripheral Lymphocyte Subset Alteration in  
1045 COVID-19 Pneumonia. *J. Infect. Dis.* *221*, 1762–1769.
- 1046 Wilk, A.J., Rustagi, A., Zhao, N.Q., Roque, J., Martínez-Colón, G.J., McKechnie, J.L.,  
1047 Ivison, G.T., Ranganath, T., Vergara, R., Hollis, T., et al. (2020a). A single-cell atlas of  
1048 the peripheral immune response in patients with severe COVID-19. *Nature Medicine* *26*,  
1049 1070–1076.
- 1050 Wilk, A.J., Lee, M.J., Wei, B., Parks, B., Pi, R., Martínez-Colón, G.J., Ranganath, T.,  
1051 Zhao, N.Q., Taylor, S., Becker, W., et al. (2020b). Multi-omic profiling reveals  
1052 widespread dysregulation of innate immunity and hematopoiesis in COVID-19.
- 1053 Wolf, F.A., Angerer, P., and Theis, F.J. (2018). SCANPY: large-scale single-cell gene  
1054 expression data analysis. *Genome Biol.* *19*, 15.
- 1055 Wong, C.K., Wong, P.T.Y., Tam, L.S., Li, E.K., Chen, D.P., and Lam, C.W.K. (2010).  
1056 Elevated production of B cell chemokine CXCL13 is correlated with systemic lupus  
1057 erythematosus disease activity. *J. Clin. Immunol.* *30*, 45–52.
- 1058 Yang, A.C., Kern, F., Losada, P.M., Maat, C.A., and Schmartz, G. (2020). Broad  
1059 transcriptional dysregulation of brain and choroid plexus cell types with COVID-19.  
1060 bioRxiv.
- 1061 Zhang, F., Wei, K., Slowikowski, K., Fonseka, C.Y., Rao, D.A., Kelly, S., Goodman,  
1062 S.M., Tabechian, D., Hughes, L.B., Salomon-Escoto, K., et al. (2019). Defining  
1063 inflammatory cell states in rheumatoid arthritis joint synovial tissues by integrating  
1064 single-cell transcriptomics and mass cytometry. *Nat. Immunol.* *20*, 928–942.
- 1065 Zhao, M. (2020). Cytokine storm and immunomodulatory therapy in COVID-19: Role of  
1066 chloroquine and anti-IL-6 monoclonal antibodies. *Int. J. Antimicrob. Agents* *55*, 105982.
- 1067 Zhou, Y., Han, T., Chen, J., Hou, C., Hua, L., He, S., Guo, Y., Zhang, S., Wang, Y.,  
1068 Yuan, J., et al. (2020). Clinical and autoimmune characteristics of severe and critical  
1069 cases of COVID-19. *Clin. Transl. Sci.* *13*, 1077–1086.
- 1070 Zuo, Y., Zuo, M., Yalavarthi, S., Gockman, K., Madison, J.A., Shi, H., Woodard, W.,  
1071 Lezak, S.P., Lugogo, N.L., Knight, J.S., et al. (2021). Neutrophil extracellular traps and  
1072 thrombosis in COVID-19. *J. Thromb. Thrombolysis* *51*, 446–453.

1073

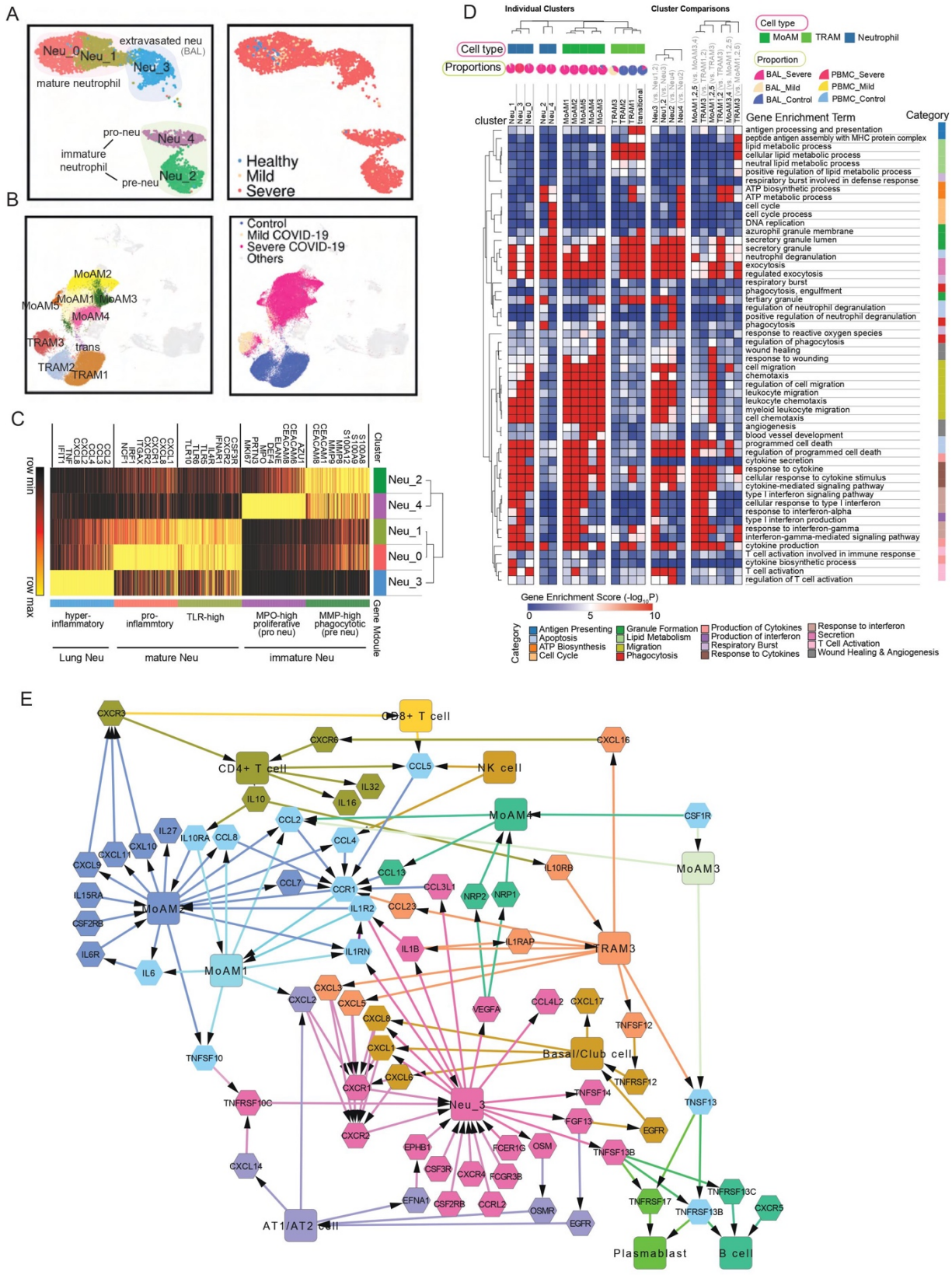
## Figures



**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>).



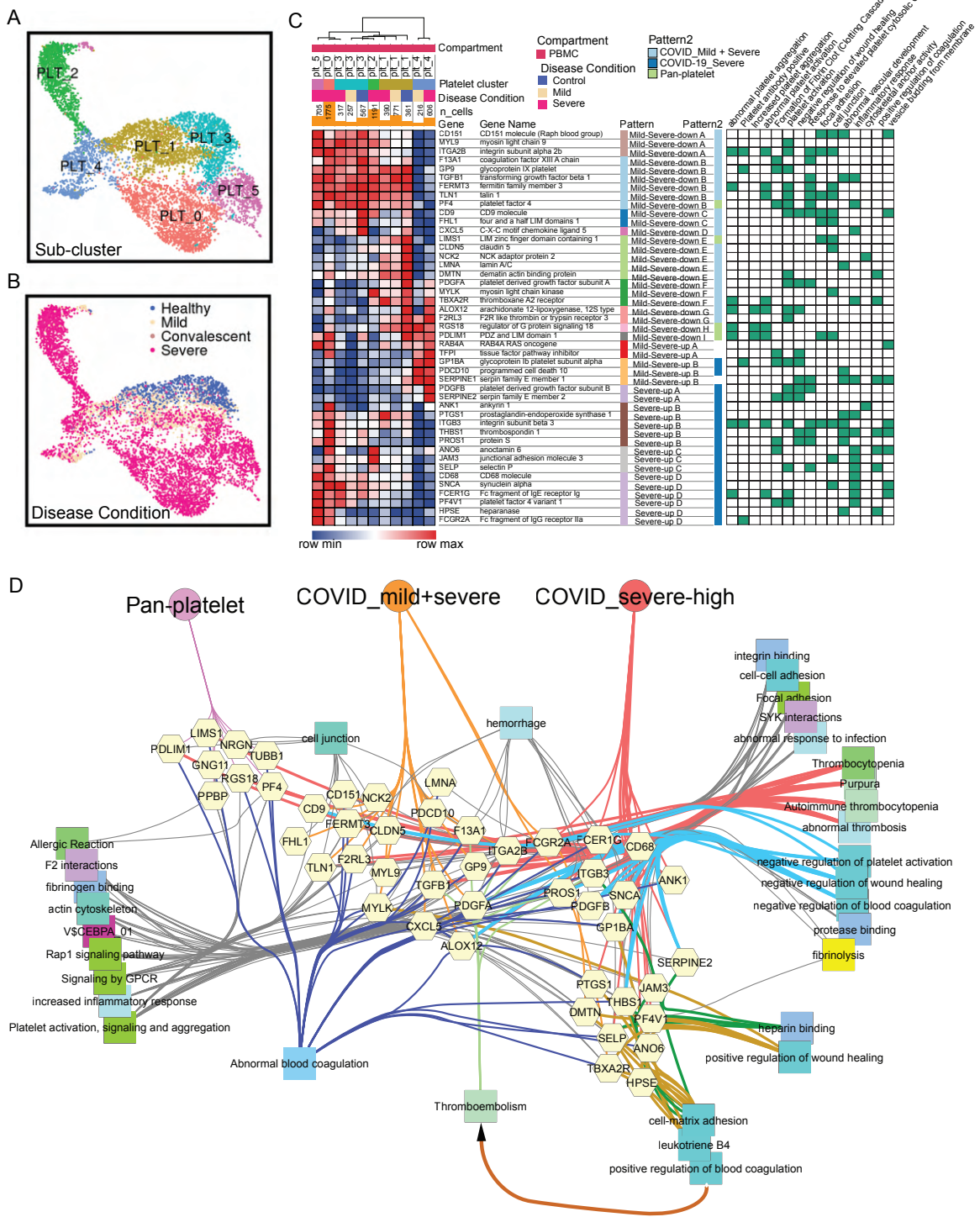
**Fig. 2. Modularized representation of cell type specific gene signatures and dynamic changes of cell abundance.** (A) Uniform Manifold Approximation and Projection (UMAP) of 28 distinct cell types identified in the integrated peripheral blood mononuclear cell (PBMC) data. (B) Comparative analysis of 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. \*:  $p \leq 0.05$ ; \*\*:  $p \leq 0.01$ ; \*\*\*:  $p \leq 0.001$ ; \*\*\*\*:  $p \leq 0.0001$ . (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.



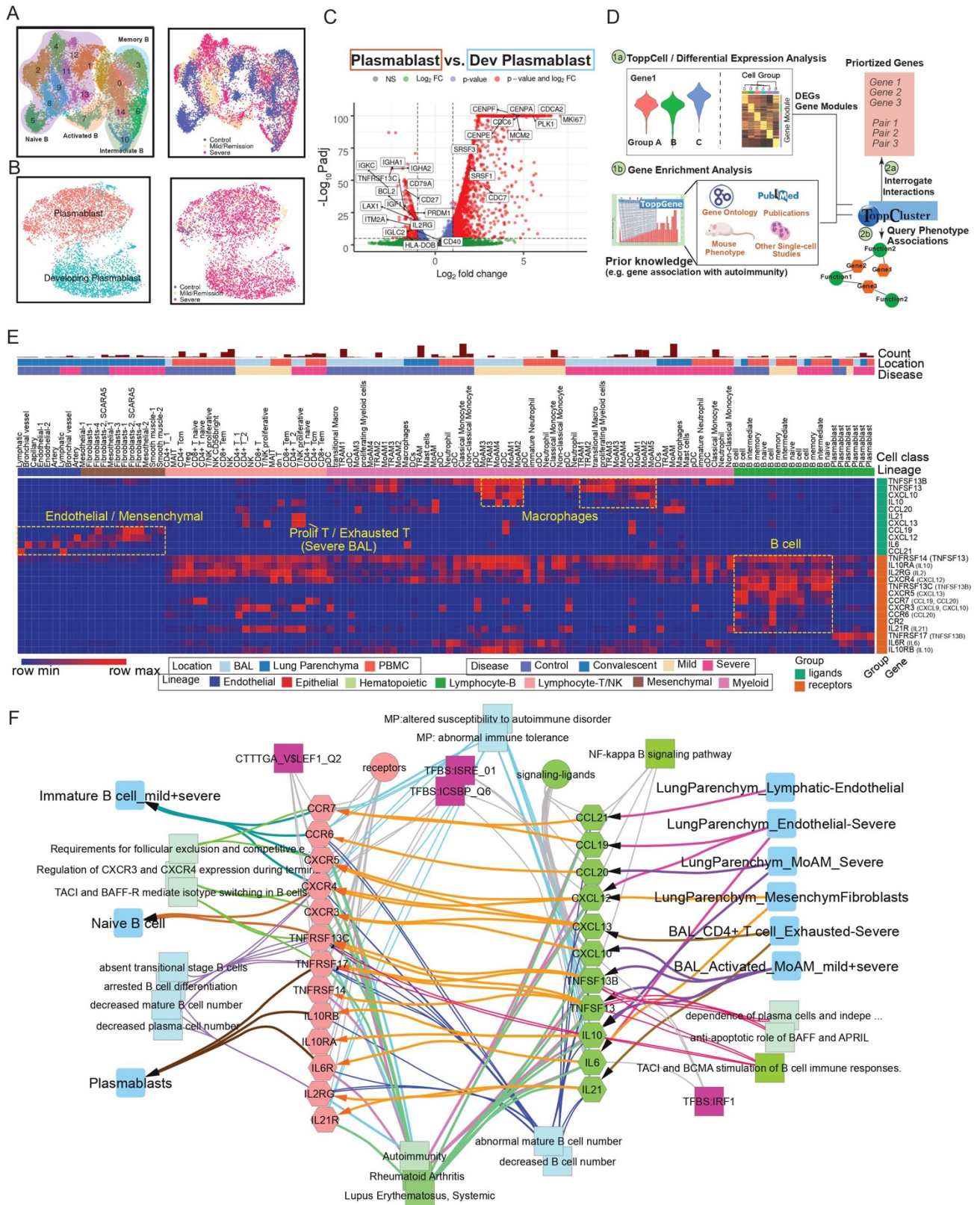
**Fig. 3. Functional analysis of compartment-specific immature and subtype-differentiated neutrophils and monocytic macrophages in COVID-19 patients. (A)**

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  $-\log_{10}(\text{adjusted p-value})$ , 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.



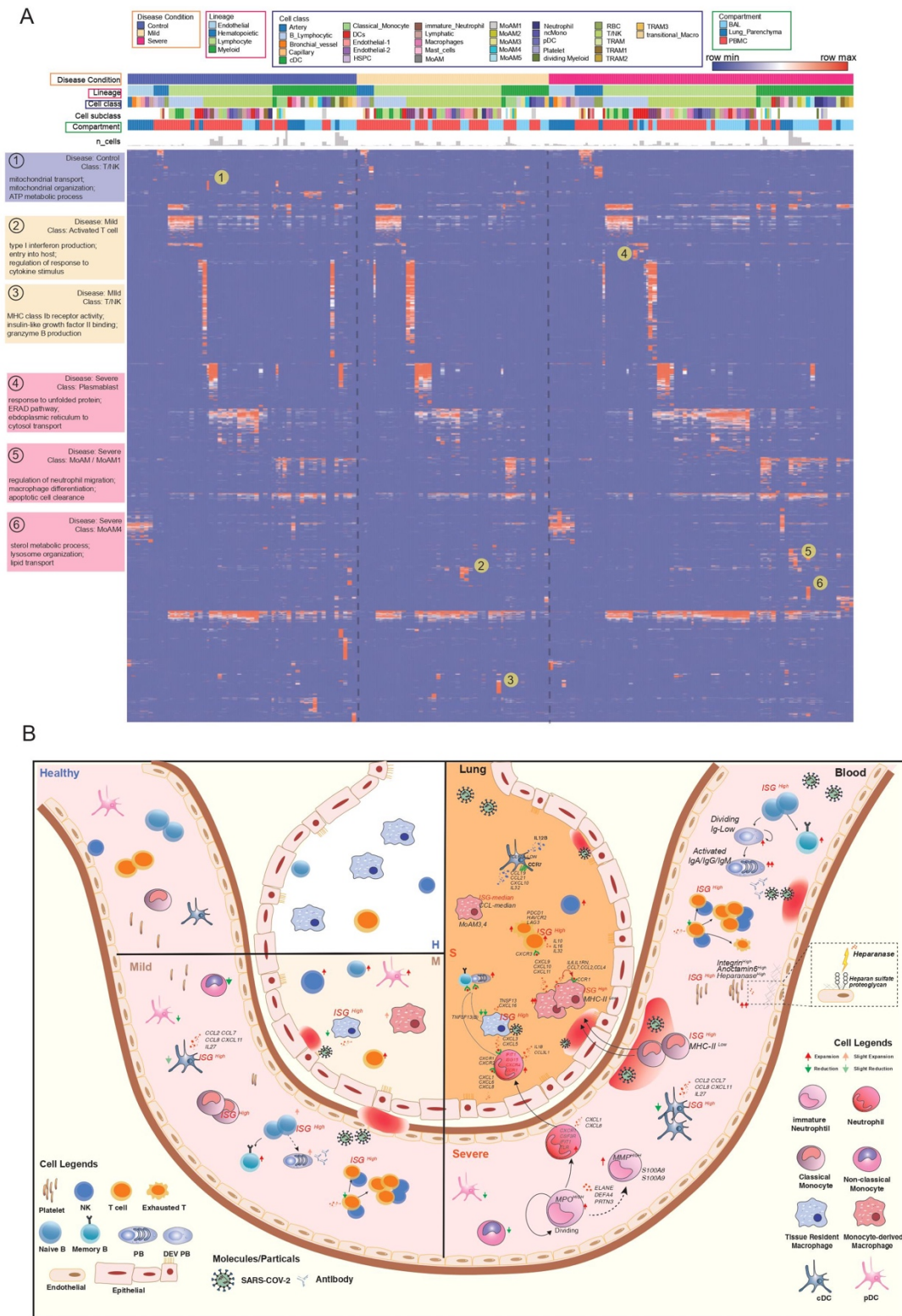


**Fig. 4. COVID-19 driven reprogramming of platelets leads to drastically altered expression of genes associated with platelet adhesion, activation, coagulation and thrombosis.** (A-B) Uniform Manifold Approximation and Projections (UMAPs) show distributions of sub-clusters (A) and COVID-19 conditions (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.



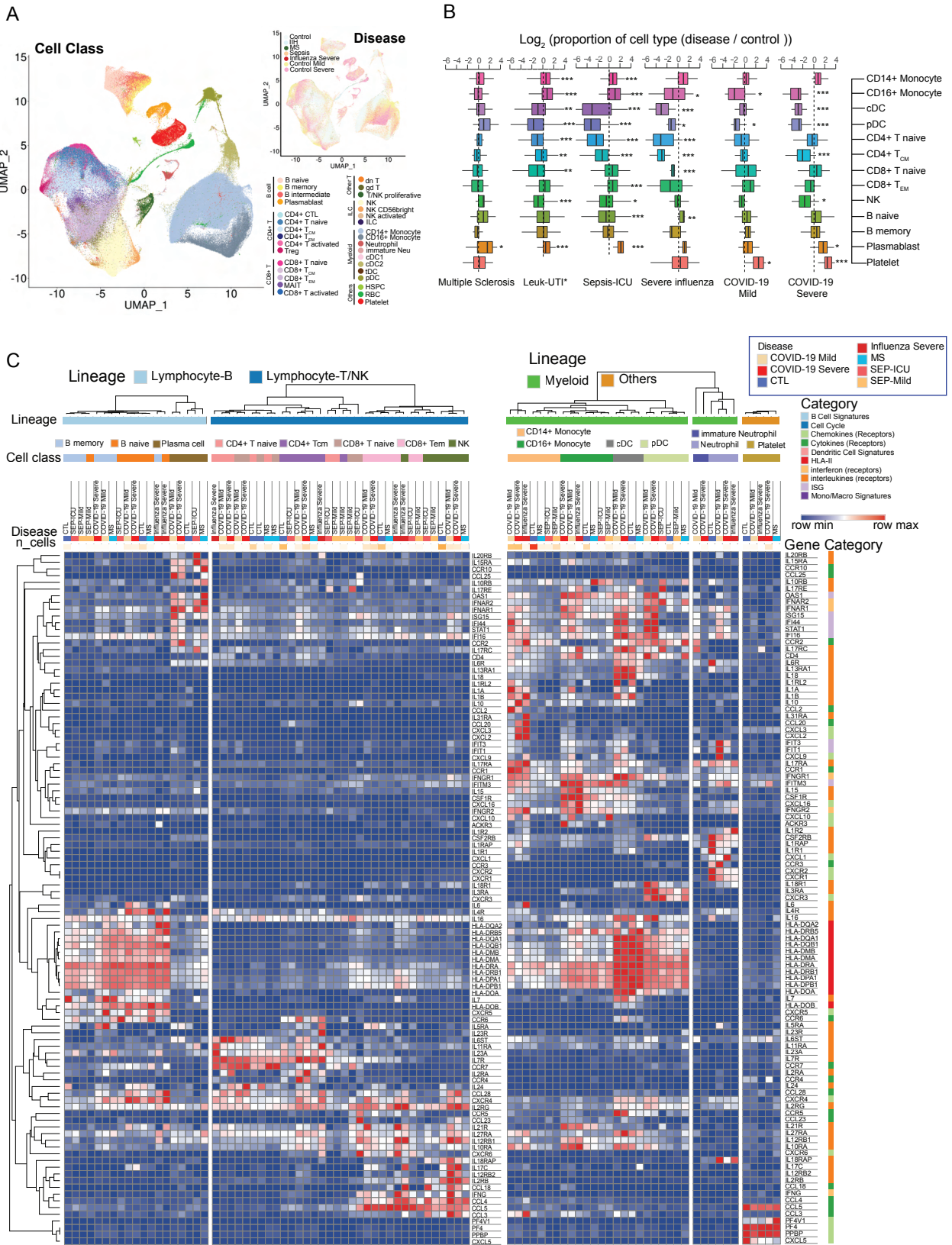
**Fig. 5. Implicating a multi-lineage cell network capable of driving extrafollicular B cell maturation and the emergence of humoral autoimmunity in COVID-19**

**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.



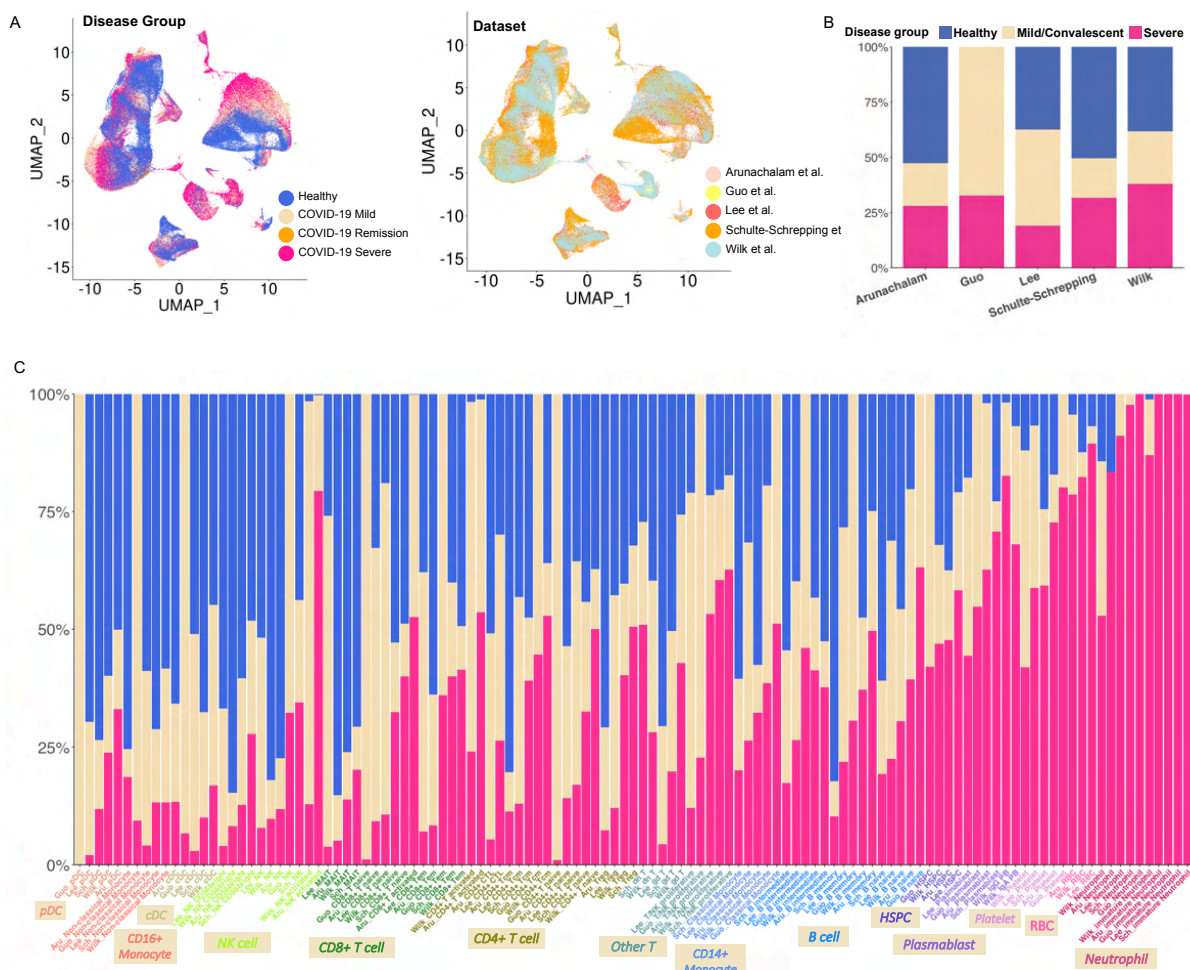
**Fig. 6. Comparative analysis of cell type specific gene signatures associated with lineage, class, subclass, compartment, and disease state in the COVID-19 atlas.**

(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.

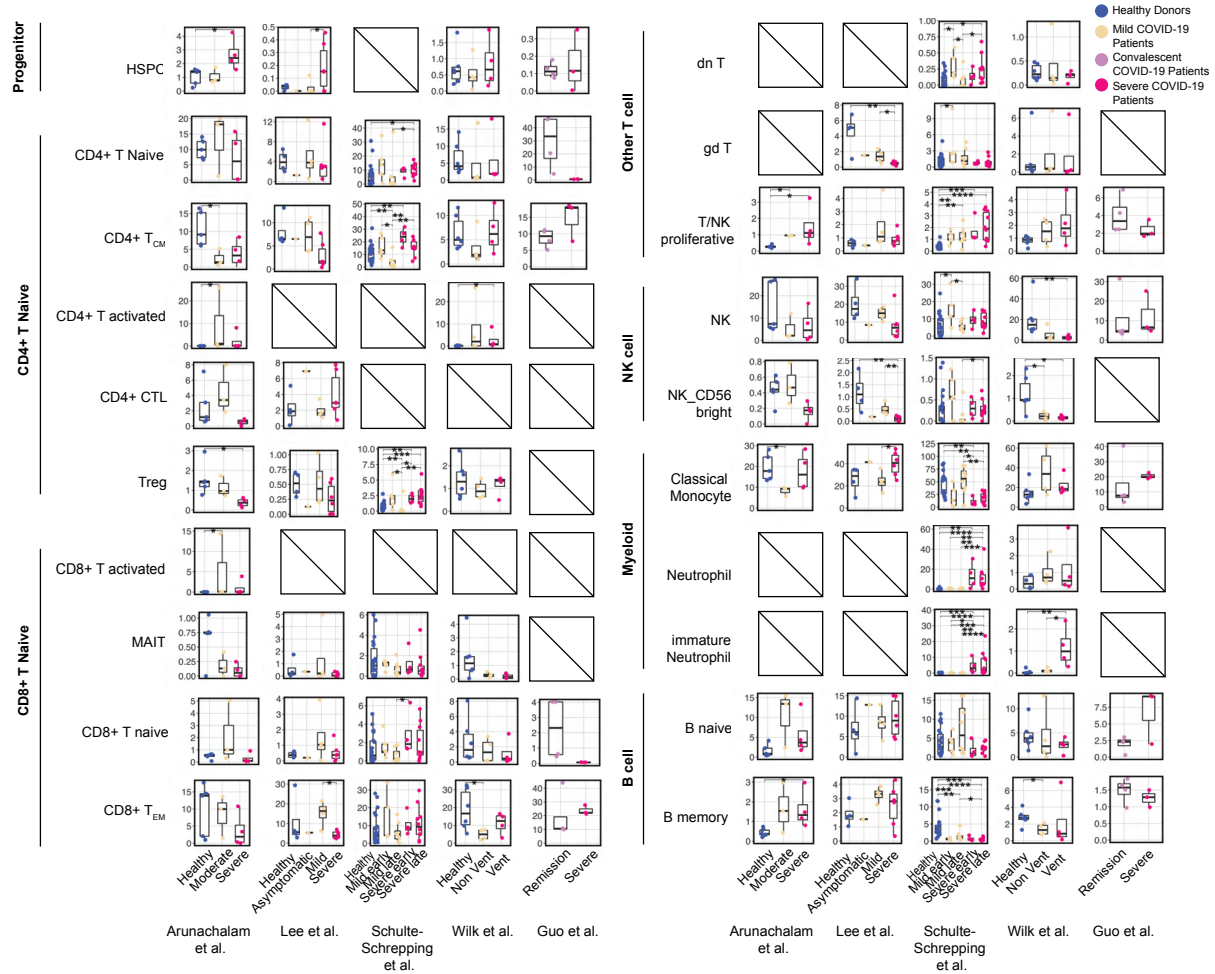


**Fig. 7. Comparative analysis of differentially-expressed immunoregulatory genes between COVID-19 and other immune-mediated diseases.** (A) Uniform Manifold Approximation and Projection (UMAP) shows the distributions of cell types (Left) and diseases (Top right) after the integration of datasets in 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 controls. Log<sub>2</sub>(ratio) was calculated to show the levels of changes. \*, p<0.05, \*\*, p<0.01, \*\*\*, p<0.001. Statistical models can be found in the Methods. Leuk-UTI: sepsis patients that enrolled into UTI with leukocytosis (blood WBC ≥ 12,000 per mm<sup>3</sup>) but no organ dysfunction. (C) Normalized expression values of key genes involved in immune signaling and responses are shown for cell types across multiple diseases. Lowly expressed genes (maximal average expression level across all cell types in the heatmap is less than 0.5 after Log<sub>2</sub>CPM normalization) were removed.

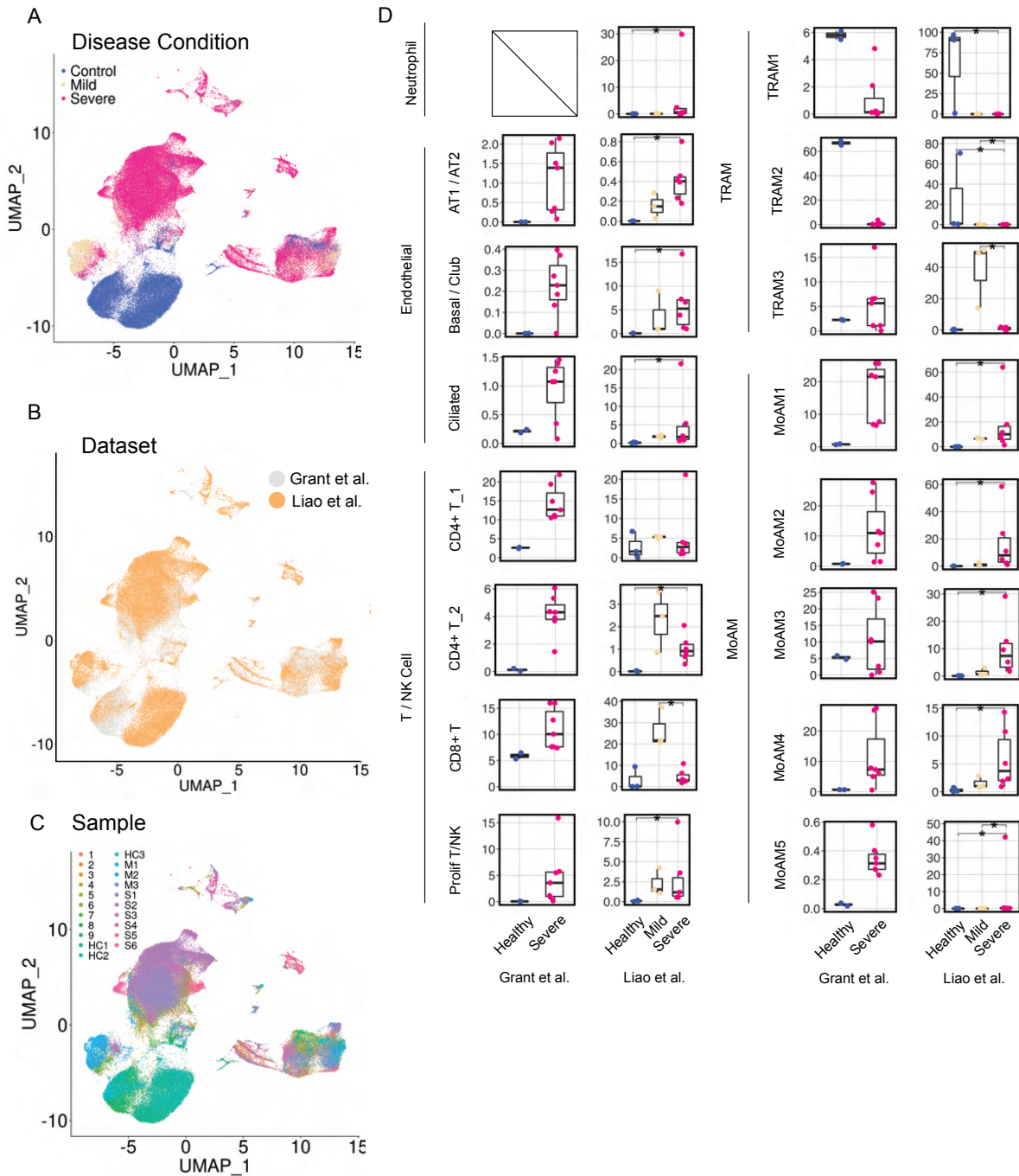




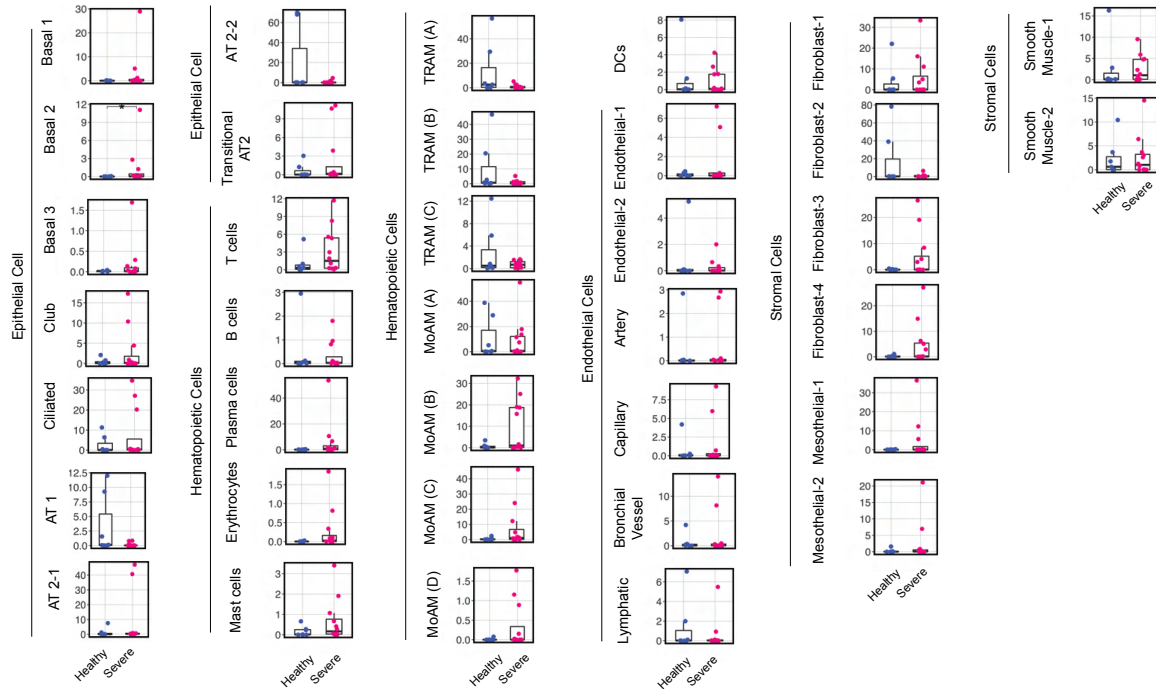
**Figure S1. Cell distribution and abundance in the integrated COVID-19 PBMC data, relative to Figure 2. (A)** Distributions of COVID-19 conditions (Left) and data sources (Right) for the integrated PBMC data are shown on the same UMAP of Figure 2A. **(B)** Bar plot depicts distributions of disease conditions 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.



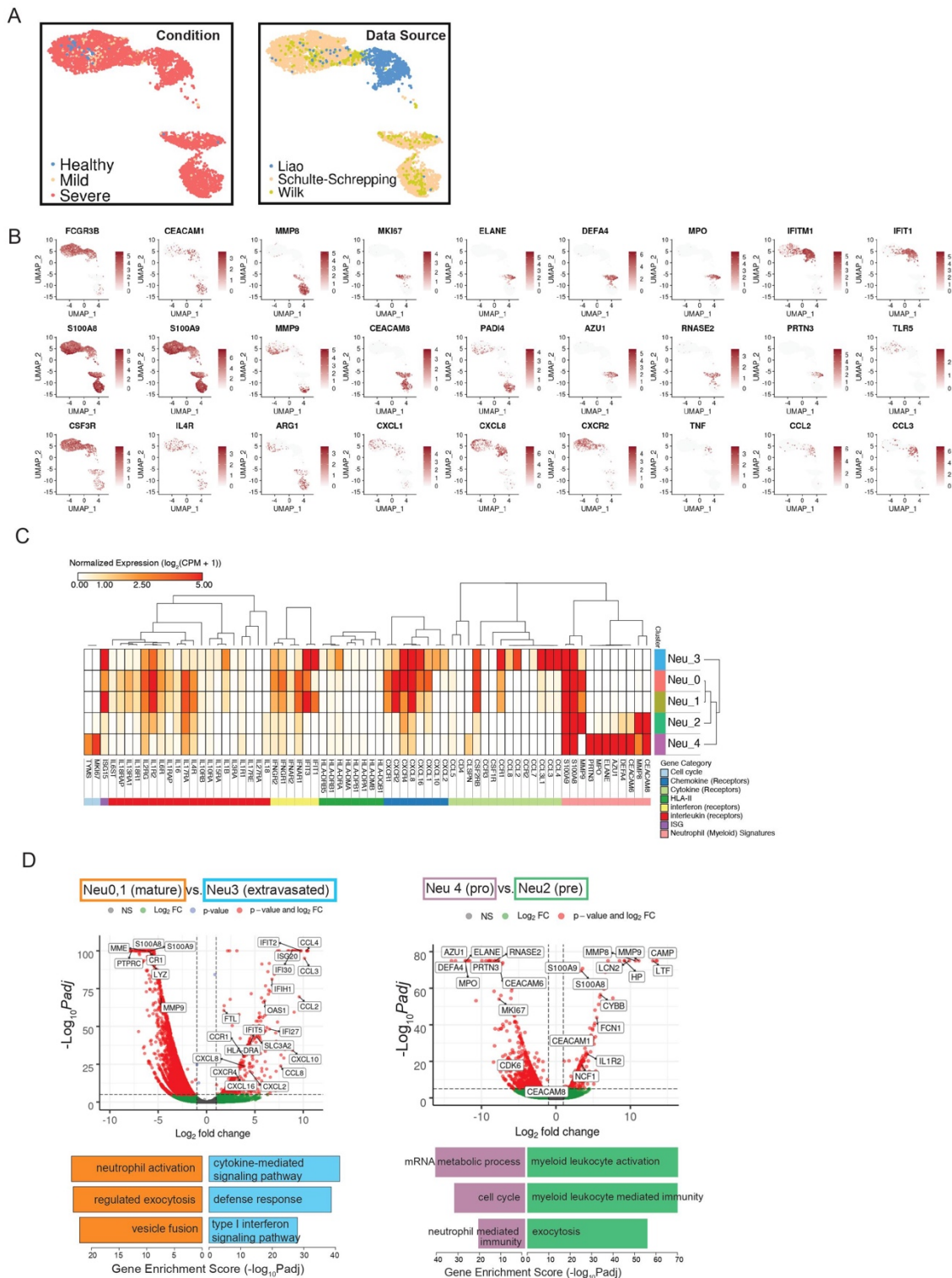
**Figure S2. Dynamic changes of cell type abundances in five COVID-19 PBMC datasets, relative to Figure 2.** Relative abundances and differences of major cell types in each single cell dataset are shown and compared to controls per each disease condition, per each single-cell dataset. Box plots of all 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.

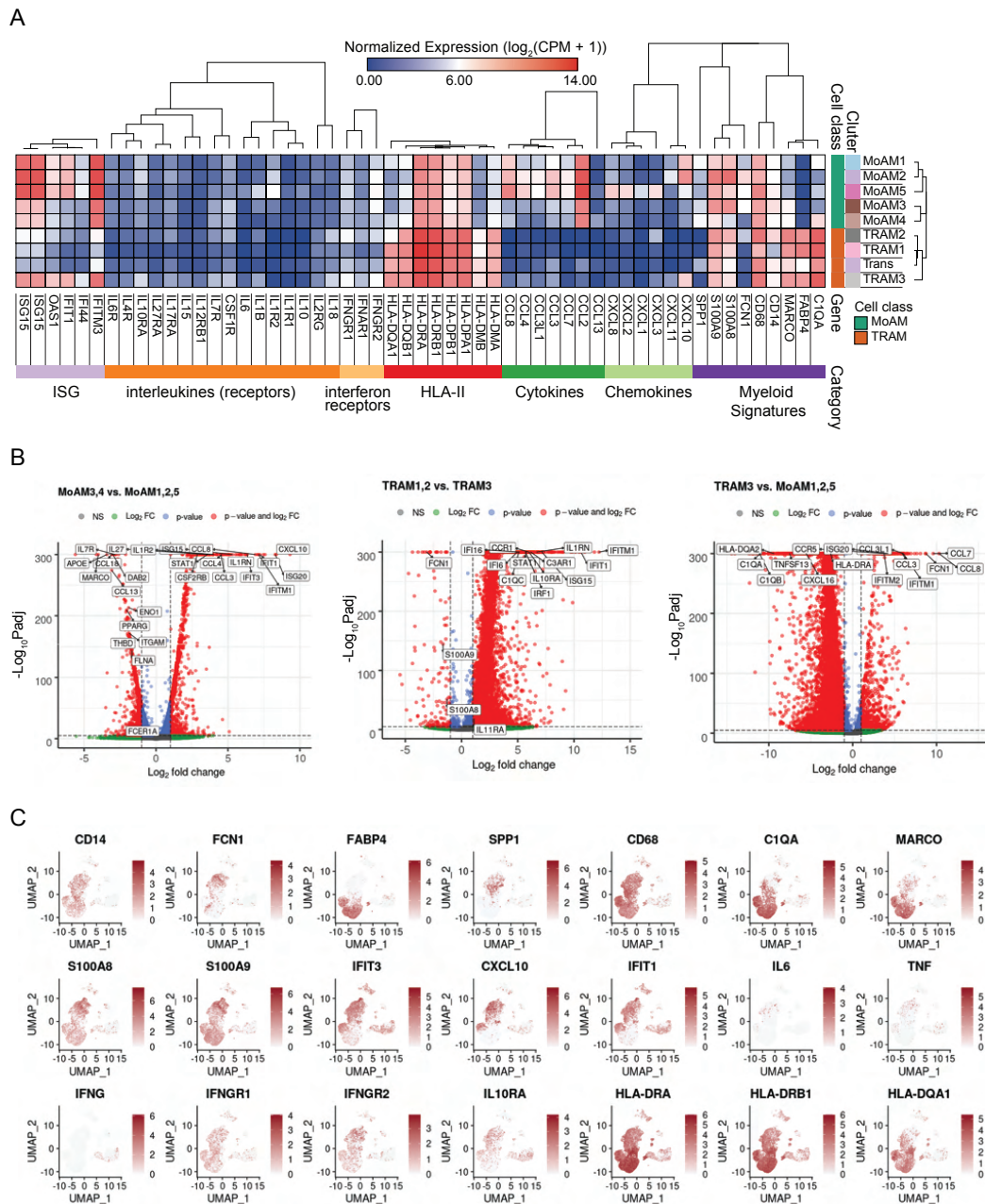


**Figure S4. Cell type abundance changes in COVID-19 lung parenchyma dataset, relative to Figure 2.** Box plots depict percentages of cell types in control samples and severe COVID-19 samples. We used cell type clusters identified in the original publication but modified cell naming of macrophage subtypes 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**

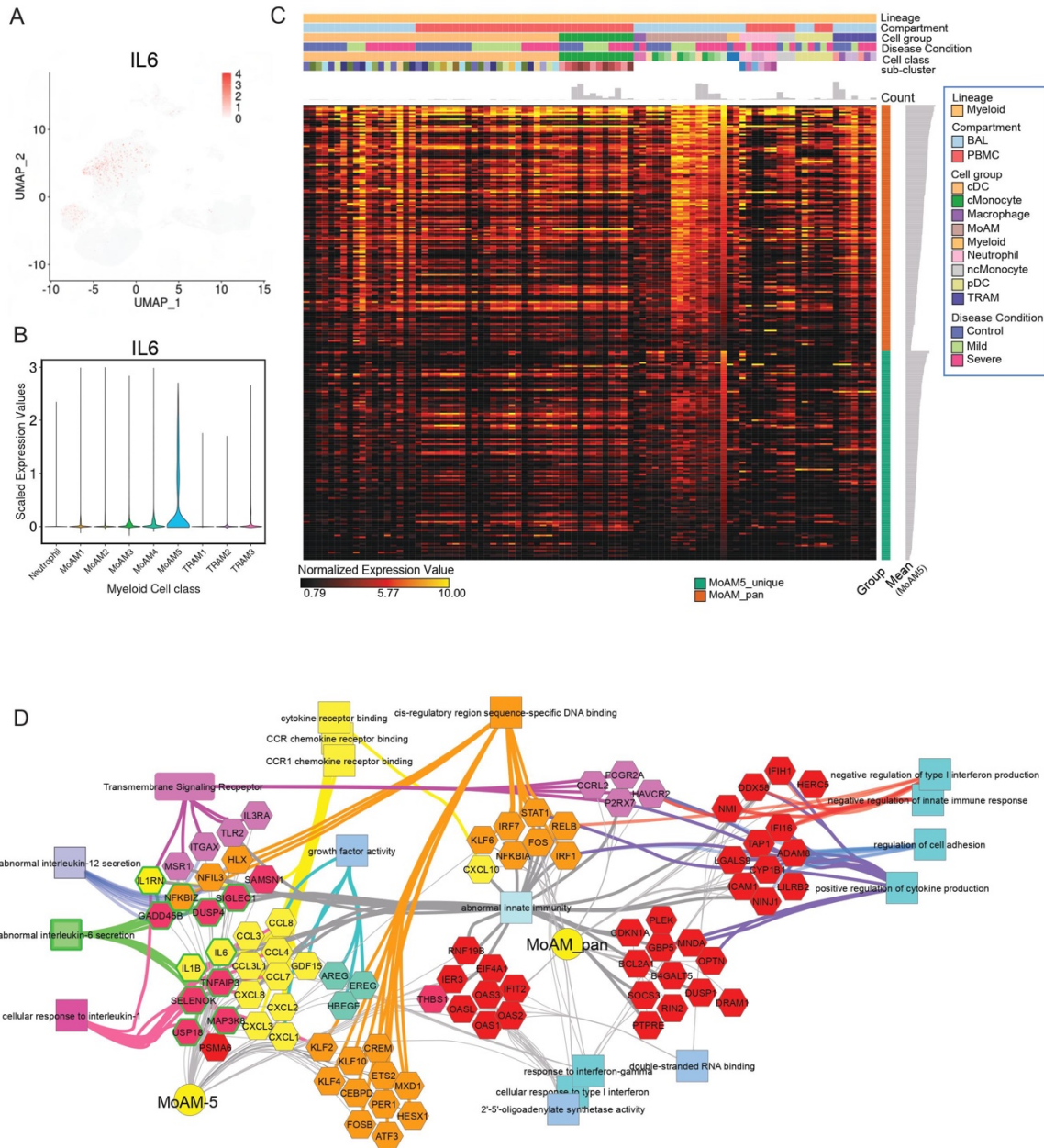
(Right) for the integrated neutrophil data on the same UMAP of Figure 3A. **(B)** UMAPs of neutrophil sub-cluster-associated genes from Figure 3C. Normalized expression values for each gene were used. **(C)** Normalized expression values of neutrophil-associated genes and other important immune signatures 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  $\text{Log}_2\text{CPM}$  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  $\log_2\text{CPM}$  normalization) were removed from the gene pool of MHC-II, cytokines, chemokines, ISGs, interleukins, interferons and their

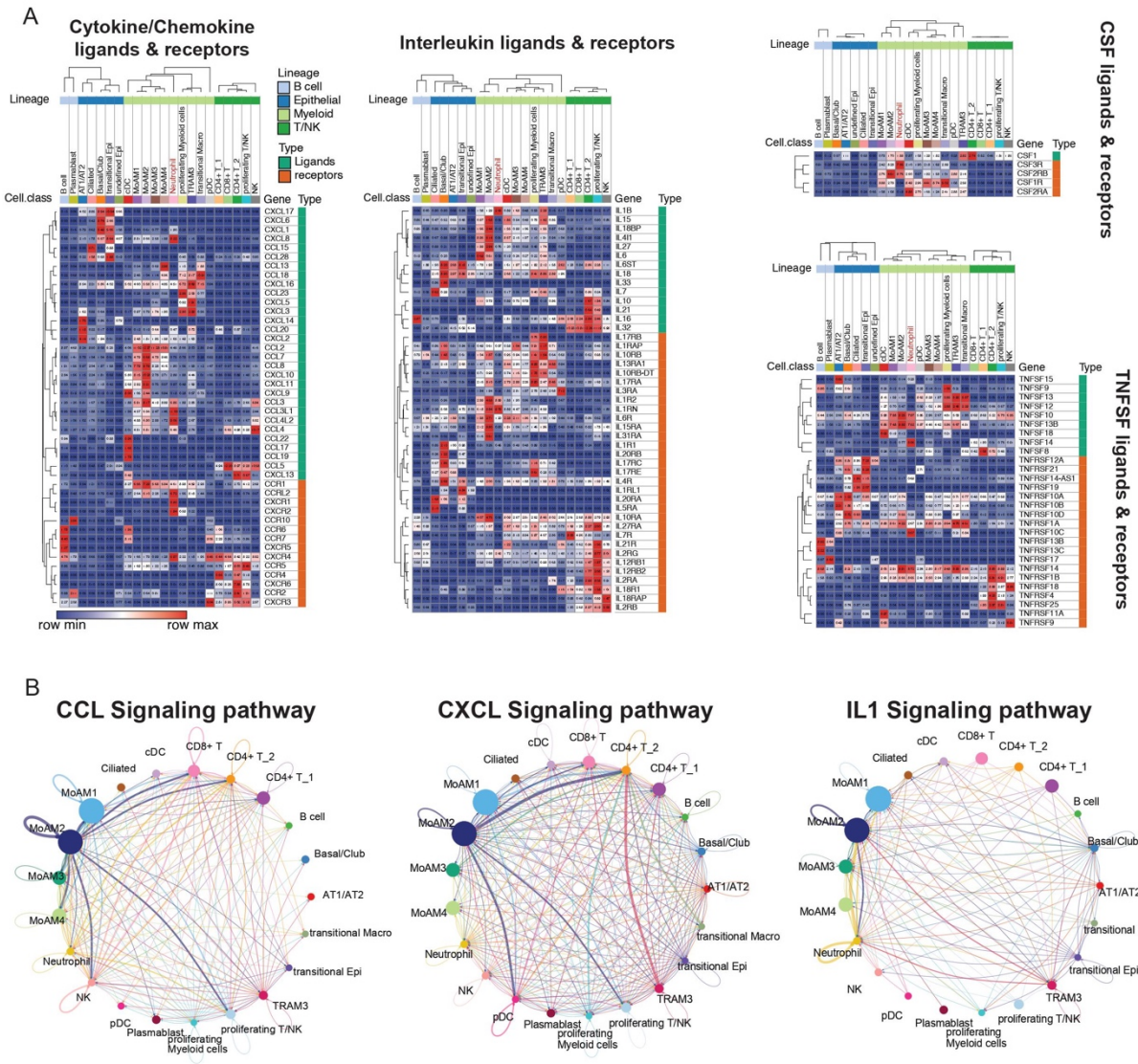
receptors. **(B)** Volcano plots were drawn for DEGs of MoAM3,4 versus MoAM1,2,5 (Left) and TRAM 1,2 versus TRAM3 (Middle) and TRAM3 versus MoAM1,2,5 (Right). Statistical methods are the same with Figure 5C. **(C)** Normalized expression values were shown on the same UMAP of Figure 3B for important genes including macrophage signatures, ISGs, interferons, receptors and MHC-II.





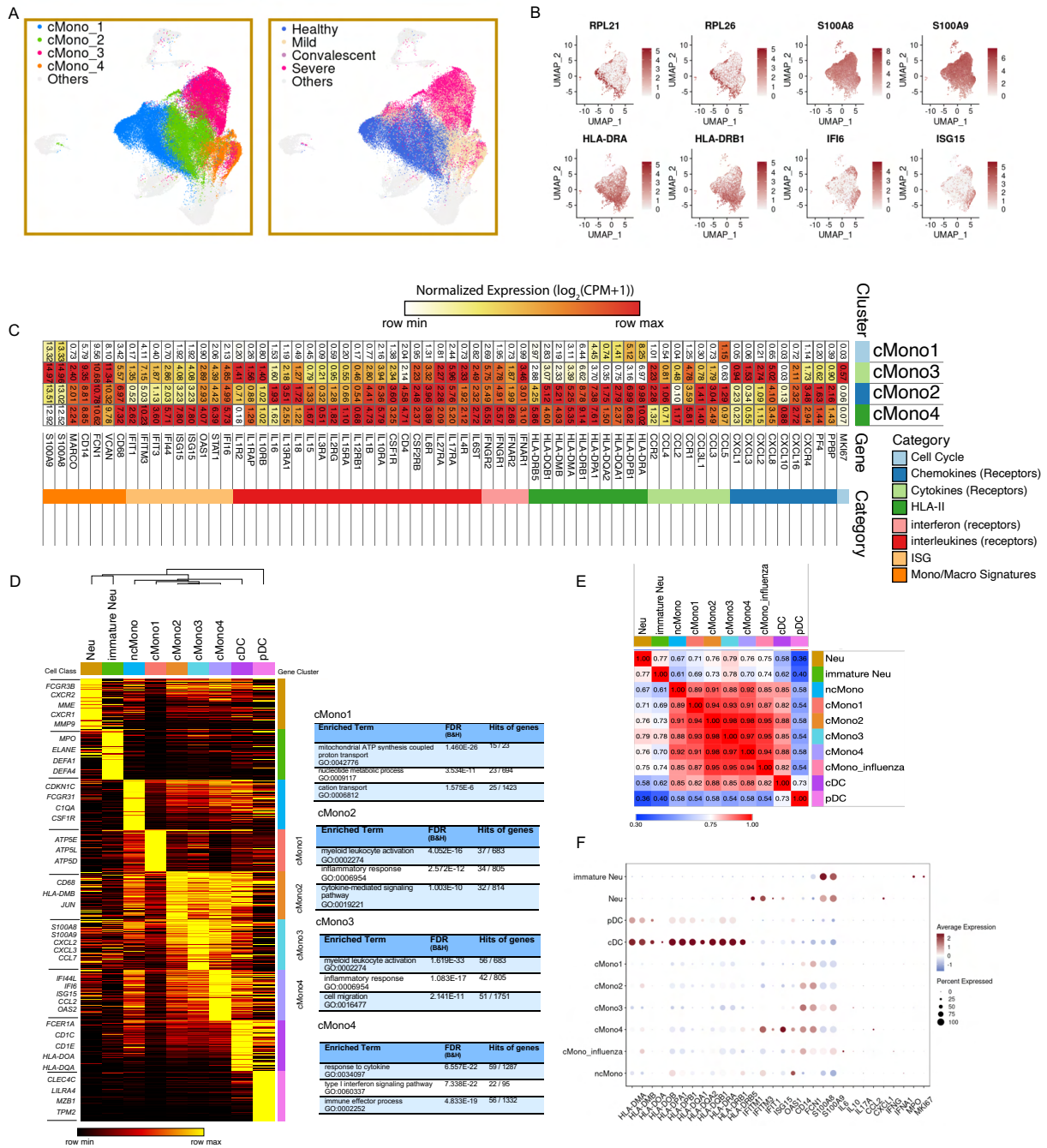
**Figure S7. A uniquely-activated monocyte-derived cell type (MoAM5) exhibits a broad signature of cytokines, chemokines, and interleukins including IL6, relative to Figure 3. (A)** Normalized expression values of IL6 on the same reference UMAP of integrated BAL data as Figure 3B. **(B)** Scale expression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



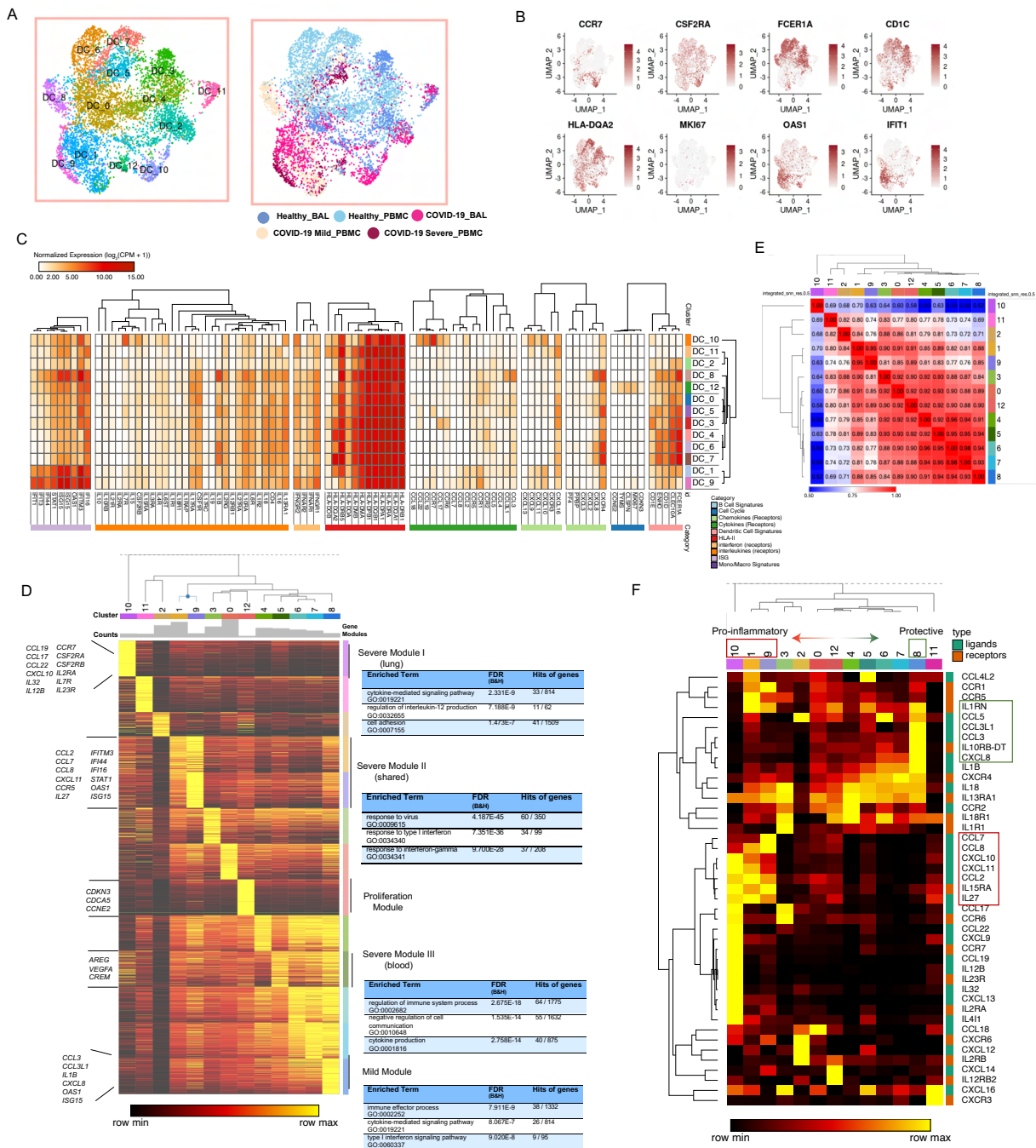
**Figure S8. Cell type and cell subtype-specific divisions of cytokine, chemokine, and interleukin signaling pathways in BAL of severe COVID-19 patients, relative to Figure 3. (A)** Heatmap of expression patterns of ligands and receptors in cytokine, chemokine, interleukin, CSF and TNFSF signaling pathways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



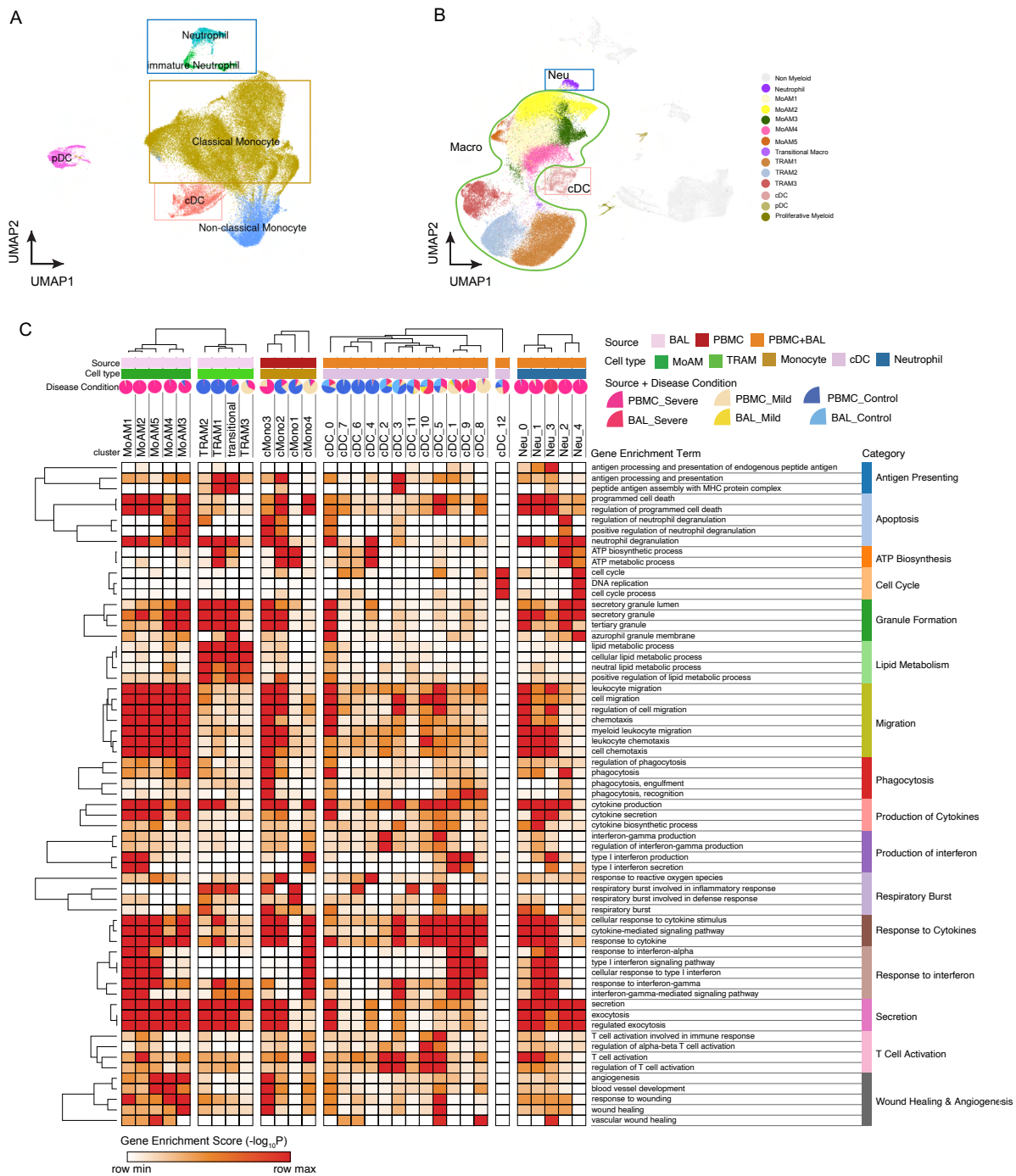


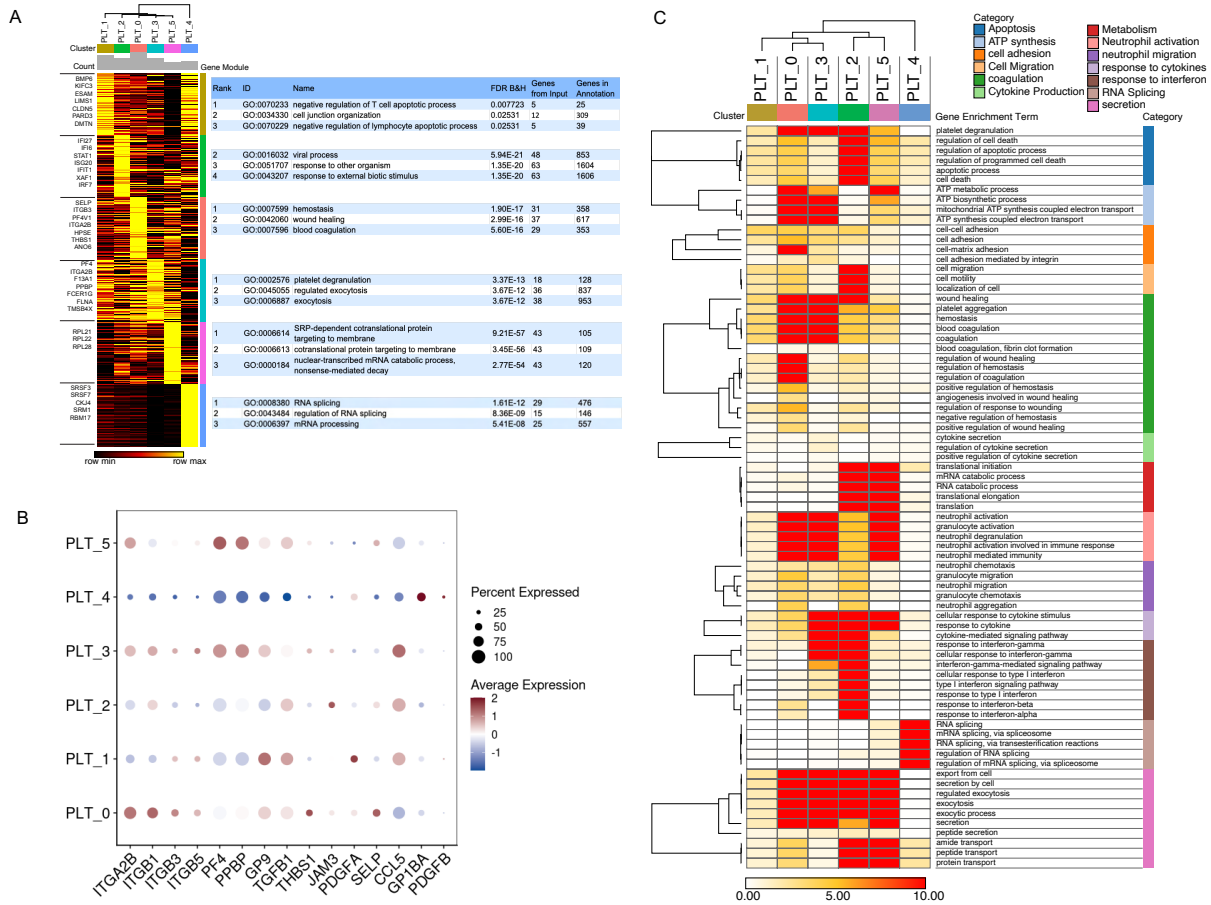
Figure S11. Landscape of myeloid cells in the integrated PBMC and BAL data, relative to Figure 3.

**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)**

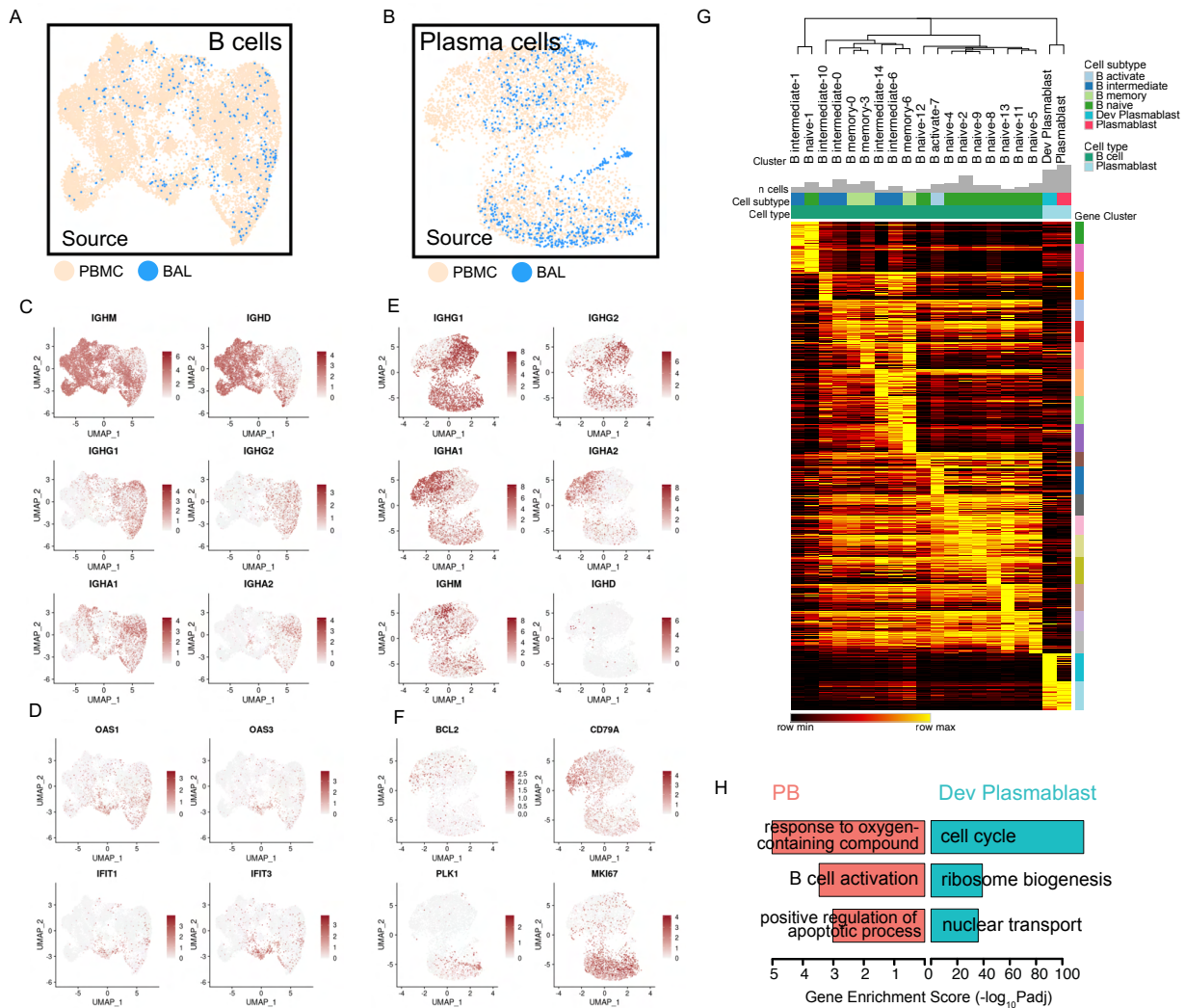
data. Cell types which were further clustered are highlighted in different colors. **(C)** The heatmap shows associations between subclusters of myeloid cells and myeloid-cell-associated pathways, such as antigen presenting, T cell activation, phagocytosis etc. Gene enrichment scores, defined as  $-\log_{10}(\text{adjusted p value})$ , were calculated as the strength of associations. Pie charts showed the proportions of COVID-19 conditions in each sub-cluster.



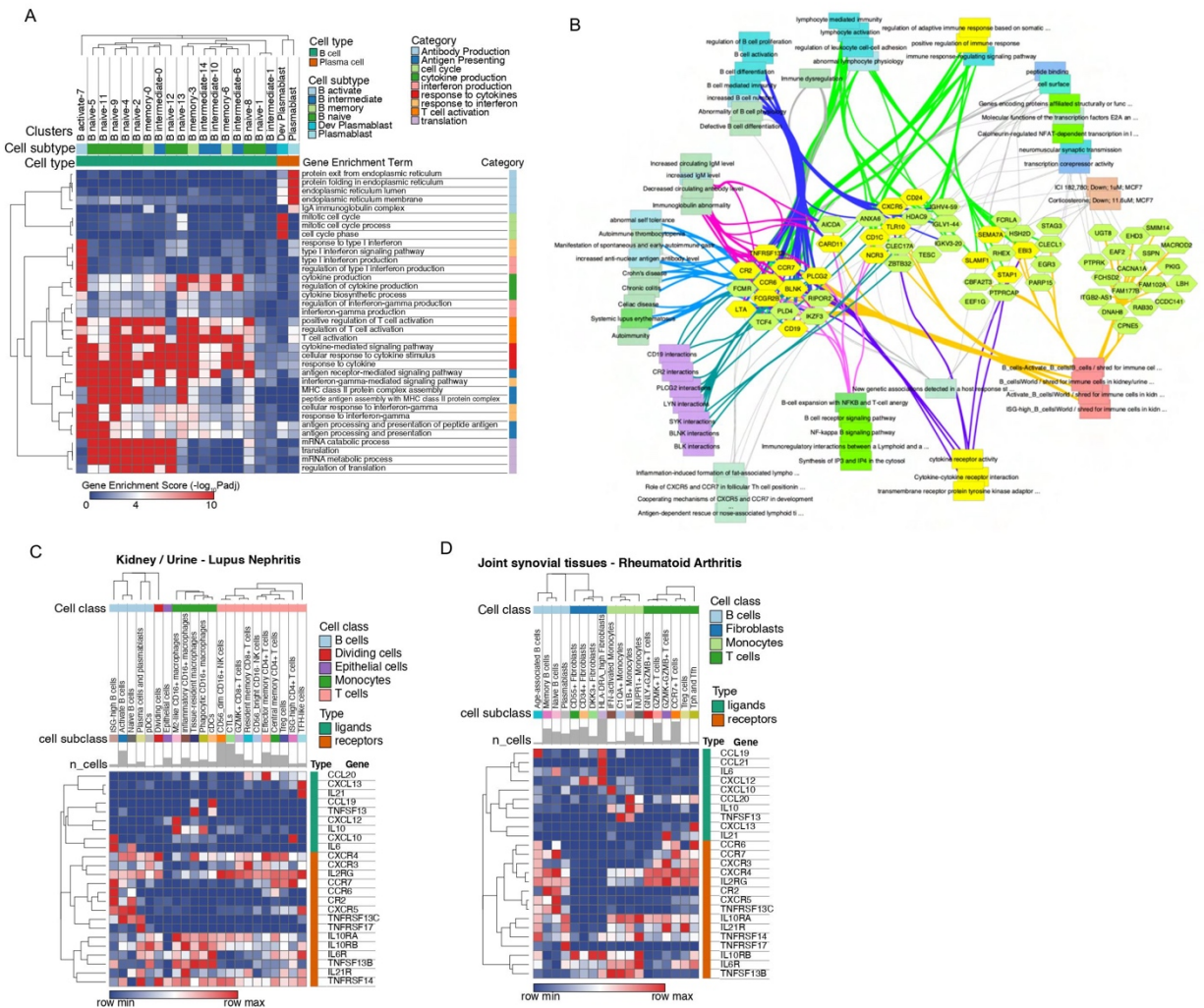
potential gene-gene interactions in immunothrombosis of COVID-19 patients. CellChat and ToppCell/ToppGene protein-protein ligand receptor and cell adhesion interaction databases were used to find interaction pairs among upregulated genes. **(C)** A new network derived from **(B)** shows integrin-associated interactions between platelets and other cells.



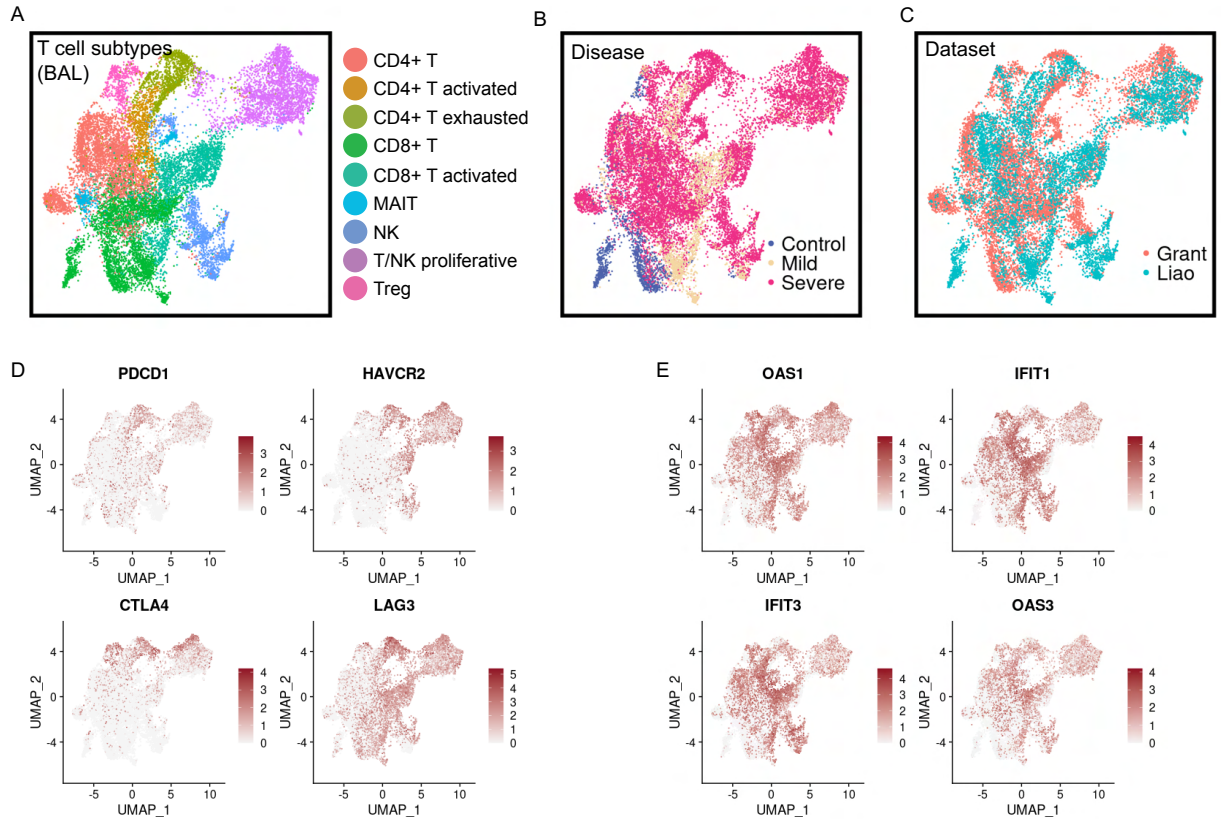
**Figure S13. Emergence of platelet subtypes implicating functionally significant alternative roles in hemostasis, coagulation, wound response, and neutrophil recruitment and activation, relative to Figure 4. (A)** The heatmap shows TopCell gene modules of 6 platelet sub-clusters in COVID-19 PBMC. Each 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 TopGene 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  $-\log_{10}(\text{adjusted p value})$ , 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.

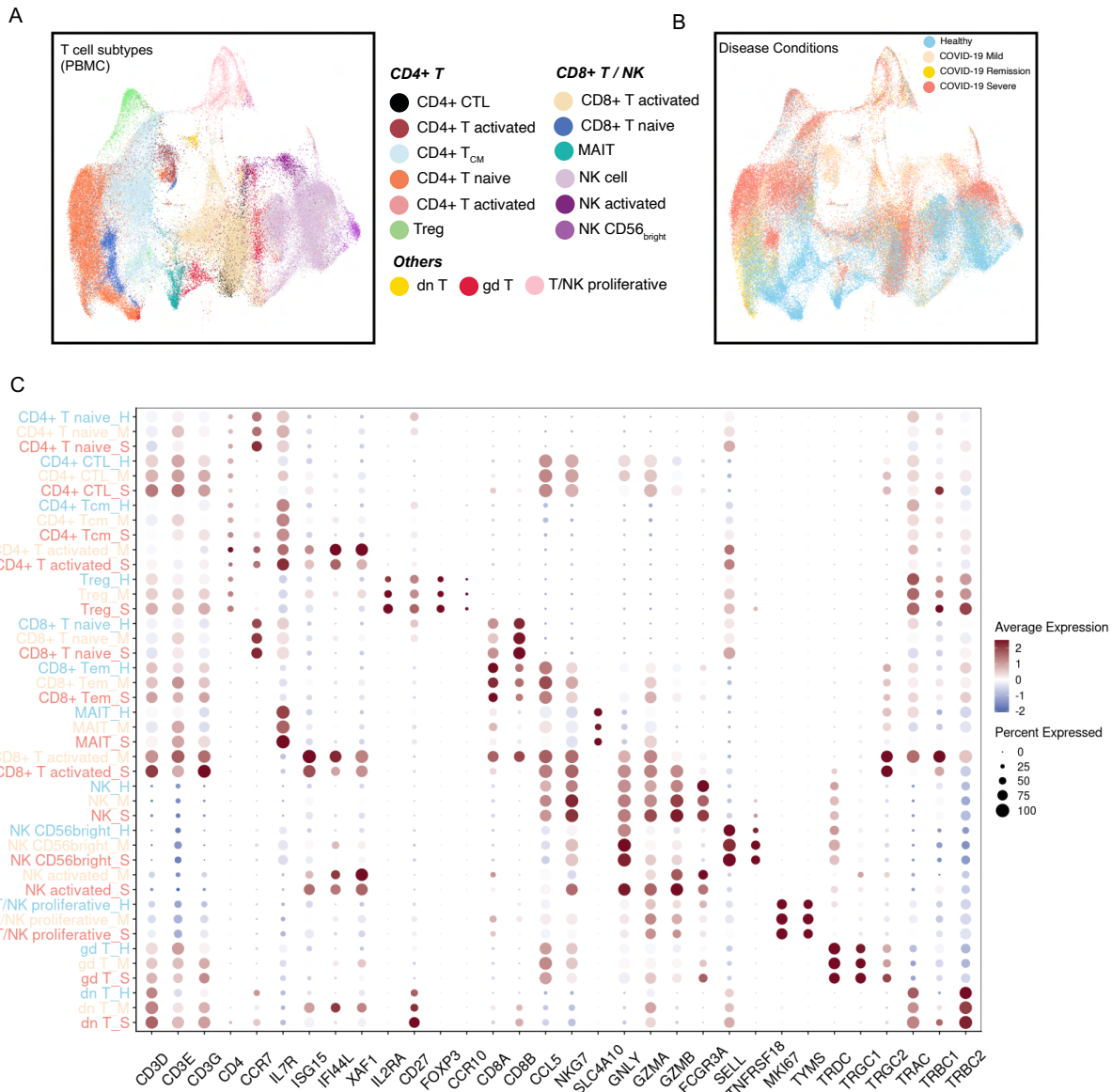


**Figure S15. Gene Enrichment analysis of B cell subtypes and autoimmune-associated signatures, relative to Figure 5. (A)** Heatmap shows gene enrichment scores of B-cell-associated pathways for each B cell sub-cluster and plasmablast subtype. **(B)** Pathway and function association network of upregulated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

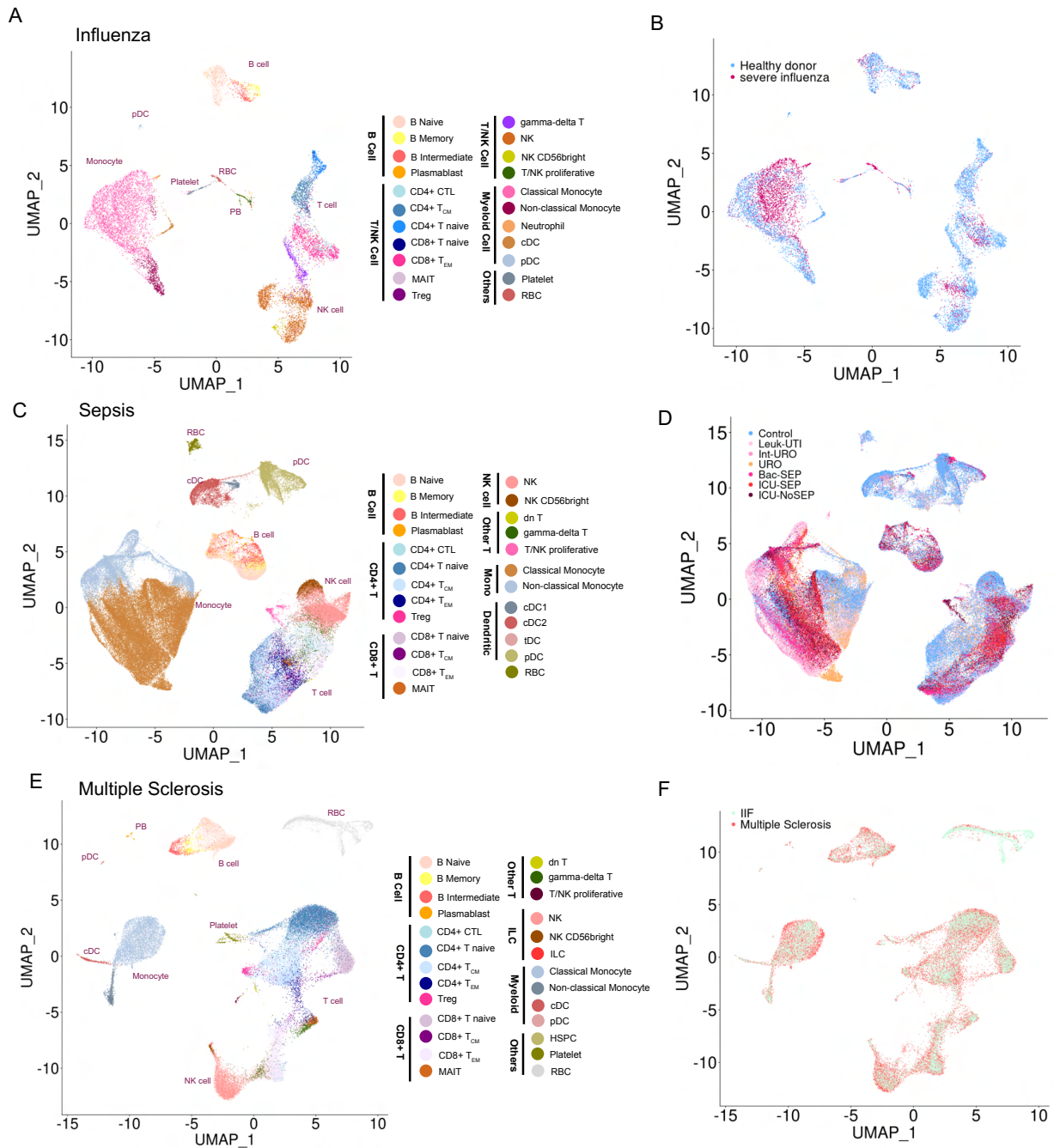


**Figure S16. Distinct subtypes of T cells and NK cells in COVID-19 BAL data. (A-C)** UMAPs of subtypes (A), COVID-19 conditions (B) and data sources (C) of T cells and NK cells in the integrated BAL data. (D-E) UMAPs of normalized expression values of exhausted T cell markers (D) and ISGs (E).





**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.