

Large-scale single-cell analysis reveals critical immune characteristics of COVID-19 patients

Xianwen Ren^{1,37}, Wen Wen^{2,3,4,37}, Xiaoying Fan^{5,11,37}, Wenhong Hou^{6,37}, Bin Su^{7,37}, Pengfei Cai^{8,37}, Jiesheng Li^{1,37}, Yang Liu^{9,37}, Fei Tang^{1,37}, Fan Zhang^{10,37}, Yu Yang^{1,37}, Jiangping He^{5,37}, Wenji Ma^{11,37}, Jingjing He^{12,37}, Pingping Wang^{13,37}, Qiqi Cao^{2,3,4}, Fangjin Chen³⁴, Yuqing Chen¹, Xuelian Cheng^{14,15}, Guohong Deng¹⁶, Xilong Deng²⁰, Wenyu Ding^{11,17}, Yingmei Feng⁷, Rui Gan¹⁰, Chuang Guo⁸, Weiqiang Guo³², Shuai He¹², Chen Jiang⁸, Juanran Liang¹⁹, Yi-min Li²¹, Jun Lin⁸, Yun Ling²³, Haofei Liu²⁴, Jianwei Liu⁵, Nianping Liu⁸, Yang Liu¹², Meng Luo¹³, Qiang Ma¹¹, Qibing Song²⁵, Wujianan Sun⁸, GaoXiang Wang²⁶, Feng Wang²⁷, Ying Wang²⁷, Xiaofeng Wen¹⁹, Qian Wu¹⁷, Gang Xu⁹, Xiaowei Xie^{14,15}, Xinxin Xiong¹², Xudong Xing²⁸, Hao Xu⁸, Chonghai Yin¹¹, Dongdong Yu²⁵, Kezhuo Yu¹, Jin Yuan⁹, Biao Zhang^{14,15}, Tong Zhang⁷, Jincun Zhao²¹, Peidong Zhao¹⁸, Jianfeng Zhou²⁶, Wei Zhou⁵, Sujuan Zhong¹⁷, Xiaosong Zhong³⁵, Shuye Zhang³⁶, Lin Zhu⁸, Ping Zhu^{14,15}, Bin Zou¹⁹, Jiahua Zou²⁹, Zengtao Zuo¹¹, Fan Bai¹, Xi Huang³⁰, Xiuwu Bian²⁴, Penghui Zhou^{12,*}, Qinghua Jiang^{13,*}, Zhiwei Huang^{10,*}, Jin-Xin Bei^{12,*}, Lai Wei^{19,*}, Xindong Liu^{24,*}, Tao Cheng^{14,15,*}, Xiangpan Li^{31,*}, Pingsen Zhao^{32,33,*}, Fu-Sheng Wang^{28,*}, Hongyang Wang^{2,3,4,*}, Bing Su^{27,*}, Zheng Zhang^{9,*}, Kun Qu^{8,*}, Xiaoqun Wang^{5,11*}, Jiekai Chen^{5,22,*}, Ronghua Jin^{7,*}, Zemin Zhang^{1,6,38,*}

1. Beijing Advanced Innovation Center for Genomics (ICG), BIOPIC and School of Life Sciences, Peking-Tsinghua Center for Life Sciences, Academy for Advanced Interdisciplinary Studies, Peking University, Beijing 100871, China
2. National Center for Liver Cancer Second Military Medical University, Shanghai, China
3. Eastern Hepatobiliary Surgery Hospital, Second Military Medical University, Shanghai, China
4. Ministry of Education (MOE) Key Laboratory on Signaling Regulation and Targeting Therapy of Liver Cancer, Second Military Medical University, Shanghai, China
5. Bioland Laboratory (Guangzhou Regenerative Medicine and Health GuangDong Laboratory), Guangzhou 510005, China
6. Institute of Cancer Research, Shenzhen Bay Laboratory, Shenzhen 518132, China
7. Beijing Youan Hospital, Capital Medical University, Beijing 100069, China
8. Department of oncology, The First Affiliated Hospital of USTC, Division of Molecular Medicine, Hefei National Laboratory for Physical Sciences at Microscale, Division of Life Sciences and Medicine, University of Science and Technology of China, Hefei, Anhui, 230021, China
9. Institute for Hepatology, National Clinical Research Center for Infectious Disease, Shenzhen Third People's Hospital, The Second Affiliated Hospital, School of Medicine, Southern University of Science and Technology, Shenzhen 518112, Guangdong Province, China
10. Center for Life Sciences, School of Life Science and Technology, Harbin Institute of Technology, 150080 Harbin, China
11. State Key Laboratory of Brain and Cognitive Science, CAS Center for Excellence in Brain Science and Intelligence Technology, Institute of Brain-Intelligence Technology (Shanghai), Institute of Biophysics, Chinese Academy of Sciences, Beijing, 100101, China
12. State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, Sun Yat-sen University Cancer Center, Guangzhou, 510060, China
13. Center for Bioinformatics, School of Life Science and Technology, Harbin Institute of Technology, China
14. State Key Laboratory of Experimental Hematology and National Clinical Research Center for Blood Diseases, Institute of Hematology and Blood Diseases Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College, Tianjin 300020, China
15. Center for Stem Cell Medicine and Department of Stem Cell & Regenerative Medicine, Chinese Academy of Medical Sciences and Peking Union Medical College, Tianjin, China
16. Department of Infectious Diseases, Southwest Hospital, Army Medical University, Chongqing, 400038, China
17. State Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University, Beijing, China
18. Analytical Biosciences Beijing Limited, Beijing, China
19. State Key Laboratory of Ophthalmology, Zhongshan Ophthalmic Center, Sun Yat-sen University, Guangzhou, 510060, China
20. Guangzhou Eighth People's Hospital, Guangzhou Medical University, Guangzhou 510060, China
21. State Key Laboratory of Respiratory Disease, National Clinical Research Center for Respiratory Disease, Guangzhou Institute of Respiratory Health, the First Affiliated Hospital of Guangzhou Medical University, Guangzhou, Guangdong, China 510120
22. Guangzhou Institutes of Biomedicine and Health, Chinese Academy of Sciences, Guangzhou 510530, China
23. Department of Infectious Disease, Shanghai Public Health Clinical Center, Shanghai, 201052, China
24. Institute of Pathology and Southwest Cancer Center, Southwest Hospital, Army Medical University, Chongqing 400038, P. R. China
25. Cancer Center, Renmin Hospital of Wuhan University, Wuhan, 430060, China
26. Department of Hematology, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430030, Hubei, China
27. Shanghai Institute of Immunology, Department of Microbiology and Immunology, Shanghai Jiao Tong University School of Medicine, Shanghai, 200025, China

- 58 28. Department of Infectious Diseases, Fifth Medical Center of Chinese PLA General Hospital, National Clinical Research Center for
59 Infectious Diseases, Beijing, China.
60 29. Cancer Center, Huanggang Hospital of Traditional Chinese Medicine, Huanggang, 438000, China
61 30. Center for Infection and Immunity, The Fifth Affiliated Hospital, Sun Yat-sen University, Zhuhai, Guangdong, 519000, China
62 31. Cancer Center, Renmin Hospital of Wuhan University, Wuhan, 430060, China
63 32. Department of Laboratory Medicine, Yuebei People's Hospital, Shantou University Medical College, Shaoguan, 512025, China.
64 33. Laboratory for Diagnosis of Clinical Microbiology and Infection, Medical Research Center, Shantou University Medical College,
65 Shaoguan, 512025, China
66 34. Center for Quantitative Biology, Peking University, Beijing, 100871, China
67 35. Beijing Shijitan Hospital, Capital Medical University, Beijing, China
68 36. Shanghai Public Health Clinical Center and Institute of Biomedical Sciences, Fudan University, Shanghai 201508, China
69
70 37. These authors contributed equally to this work
71 38. Lead contact

72
73 * Correspondence: zhouph@sysucc.org.cn (P.Z.), qhjiang@hit.edu.cn (Q.J.), huangzhiwei2009@gmail.com (Z.H.),
74 beijx@sysucc.org.cn (J.B.), weilai@gzzoc.com (L.W.), xindongliu@hotmail.com (X.L.), chengtao@ihcams.ac.cn (T.C.),
75 rm001227@whu.edu.cn (X.L.), zhaopingsen01@163.com (P.Z.), fswang302@163.com (F.W.), hywangk@vip.sina.com
76 (H.W.), bingusu@sjtu.edu.cn (B.S.), zhangzheng1975@aliyun.com (Z.Z.), qukun@ustc.edu.cn (K.Q.),
77 xiaoqunwang@ibp.ac.cn (X.W.), chen_jiekai@gibh.ac.cn (J.C.), ronghuajin_youan@126.com (R.J.), zemin@pku.edu.cn
78 (Z.Z.)
79
80
81

82 HIGHLIGHTS

- 83 ● Large-scale scRNA-seq analysis depicts the immune landscape of COVID-19
84 ● Lymphopenia and active T and B cell responses coexist and are shaped by age and
85 sex
86 ● SARS-CoV-2 infects diverse epithelial and immune cells, inducing distinct responses
87 ● Cytokine storms with systemic S100A8/A9 are associated with COVID-19 severity

88 SUMMARY

89 Dysfunctional immune response in the COVID-19 patients is a recurrent theme impacting
90 symptoms and mortality, yet the detailed understanding of pertinent immune cells is not
91 complete. We applied single-cell RNA sequencing to 284 samples from 205 COVID-19
92 patients and controls to create a comprehensive immune landscape. Lymphopenia and
93 active T and B cell responses were found to coexist and associated with age, sex and their
94 interactions with COVID-19. Diverse epithelial and immune cell types were observed to be
95 virus-positive and showed dramatic transcriptomic changes. Elevation of ANXA1 and
96 S100A9 in virus-positive squamous epithelial cells may enable the initiation of neutrophil
97 and macrophage responses via the ANXA1-FPR1 and S100A8/9-TLR4 axes. Systemic up-
98 regulation of S100A8/A9, mainly by megakaryocytes and monocytes in the peripheral blood,
99 may contribute to the cytokine storms frequently observed in severe patients. Our data
100 provide a rich resource for understanding the pathogenesis and designing effective
101 therapeutic strategies for COVID-19.
102

103 INTRODUCTION

104 The coronavirus disease 2019 (COVID-19) is an ongoing pandemic infectious disease,
105 caused with the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). Currently,
106 it has caused with around 29 million infections and close to 1 million deaths according to
107 the statistics of World Health Organization until September 15, 2020, with the fatality rate as
108 high as ~10% in specific regions. Although many COVID-19 patients experience
109 asymptomatic, mild or moderate symptoms, some patients progress to severe conditions
110 and even death. It is thus of paramount importance to understand the disease mechanisms

111 and the underlying factors associated with vulnerabilities, which are critical for controlling
112 the pandemic and alleviating the global crisis. It is also critical to systematically investigate
113 differences between clinical presentations (mild/moderate and severe), or between
114 treatment outcomes (disease progression and convalescence) of patients, as they can
115 provide important guidance to the development of effective therapeutics and vaccines.

116
117 Multiple studies have suggested the alterations of immune responses as one of the key
118 mechanisms for severe symptoms (Guo et al., 2020a; Schulte-Schrepping et al., 2020;
119 Silvin et al., 2020a; Wen et al., 2020; Zhang et al., 2020a; Zhang et al., 2020b). Patients
120 with severe COVID-19 might have a cytokine storm syndrome accompanying the hyper-
121 inflammatory response, which is a major cause of disease severity and death (Cao, 2020;
122 Del Valle et al., 2020; Huang et al., 2020; Liao et al., 2020; Mehta et al., 2020; Zhou et al.,
123 2020a). During the inflammatory responding process, pathogenic T cells and inflammatory
124 monocytes produced inflammatory cytokines (Zhou et al., 2020b) such as G-CSF (Costela-
125 Ruiz et al., 2020; Du et al., 2020), TNF- α (Jamilloux et al., 2020; Vabret et al., 2020), IL-6
126 and IL-1 (Abbasifard and Khorramdelazad, 2020; Costela-Ruiz et al., 2020; Del Valle et al.,
127 2020; Liao et al., 2020; Mehta et al., 2020; Yang et al., 2020), and drove downstream
128 hyper-inflammation. In contrast, some studies have argued against the presence of
129 cytokine storms (Kox et al., 2020; Wilk et al., 2020). Thus, the detailed immune responses
130 in COVID-19 patients with SARS-CoV-2 infection need to be more thoroughly investigated.

131
132 Single-cell RNA sequencing (scRNA-seq) is powerful at dissecting the immune responses
133 under various conditions at the finest resolution, and has been applied to COVID-19 studies
134 on limited scales (Chua et al., 2020; Guo et al., 2020a; He et al., 2020; Liao et al., 2020;
135 Wilk et al., 2020; Xie et al., 2020). While the current single cell studies of COVID-19 have
136 provided certain details of the cellular and molecular changes of patients after SARS-CoV-2
137 infection and even during convalescence (Mathew et al., 2020a), the small sample sizes of
138 such studies have raised concerns over the robustness and the generalization of such
139 findings. Here we applied scRNA-seq to a large cohort with 205 individuals, including
140 hospitalized COVID-19 patients with moderate or severe disease, and patients in the
141 convalescent stage, as well as healthy controls. With high-quality transcriptomics data of
142 ~1.5 million single cells, we reveal that SARS-CoV-2 could infect a wider range of cell types
143 than previous understanding, and induce distinct phenotypic changes in those infected cells.
144 Such heterogeneity of SARS-CoV-2 infection has important immunological implications as
145 such cells exhibit distinct interaction potentials with innate and adaptive immune cells. We
146 also observed critical changes in the peripheral blood discriminating mild/moderate from
147 severe COVID-19 patients in the disease progression or convalescence stages, and found
148 their association with patient sex and age. Further, our large cohort analysis provides a
149 unique opportunity to reveal the characteristics of cytokine storms in patients, and to further
150 illustrate the cell subpopulations that might contribute to the inflammatory responses and
151 the hyper-inflammatory genomic signatures under SARS-CoV-2 infection. Our findings may
152 have important implications to the research, treatment, control and prevention of COVID-19.

153 **RESULTS**

154 **Integrated analysis of the COVID-19 scRNA-seq data**

155 To systematically characterize the immune properties at single-cell resolution in the COVID-
156 19 patients, we formed a Single Cell Consortium for COVID-19 in China (SC4), which
157 consisted of researchers from 36 research institutes or hospitals from different regions of
158 China. Members of SC4 contributed COVID-19 related scRNA-seq data, mostly still

159 unpublished, for a total of 205 individuals, including 25 patients with mild/moderate
160 symptoms, 63 hospitalized patients with severe symptoms, and 92 recovered convalescent
161 persons, as well as 25 healthy controls (Figure 1A and Table S1). While most previous
162 studies did not discriminate whether convalescent individuals recovered from mild/moderate
163 or severe symptoms, we divided the convalescent group into two subgroups, 54 recovered
164 from mild/moderate symptoms and 38 recovered from severe symptoms, to investigate the
165 effects of disease severity on the immune status of recovered individuals. This cohort
166 covered a wide age range (from 6 to 92 years old), with the mild/moderate and severe
167 groups having significant age differences (Figure S1A), consistent with the epidemiological
168 observations that aged patients are prone to severe symptoms (Guo et al., 2020a; Hadjadj
169 et al., 2020; Silvin et al., 2020b; Wilk et al., 2020; Yu et al., 2020). Additionally, no
170 significant difference was noted in the sex composition between the mild/moderate and
171 severe groups (Figure S1B).

172
173 A total of 284 samples were collected for scRNA-seq, of which 249 were from peripheral
174 blood mononuclear cells (PBMCs) and 35 from the respiratory system, which was further
175 composed of 12 bronchoalveolar lavage fluid (BALF) samples, 22 sputum samples, and 1
176 sample for pleural fluid mononuclear cells. Some patients had multiple samples collected,
177 including seven patients with matching BALF and PBMC. Most samples were subjected to
178 scRNA-seq based on the 10X Genomics 5' sequencing platform to generate both the gene
179 expression and T cell receptor (TCR) or B cell receptor (BCR) data. The scRNA-seq raw
180 data were analyzed by a unified analysis pipeline, including the kallisto and bustools
181 programs (Bray et al., 2016; Melsted et al., 2019), to obtain the gene expression data of
182 individual cells and by the CellRanger program to obtain TCR and BCR sequences.

183
184 We applied a common set of stringent quality control criteria to ensure that the selected
185 data were from single and live cells and that their transcriptomic phenotypes were
186 comprehensively characterized. A total of 1,462,702 high-quality single cells were ultimately
187 obtained, with an average of 4,835 unique molecular identifiers (UMIs), representing 1,587
188 genes (Figures S1D and S1E). With the large-scale of data, we obtained 64 cell clusters,
189 covering diverse epithelial cells in the respiratory system, megakaryocytes, mast cells,
190 myeloid cells, and NK/T/B cells (Figure 1B). Such an information-rich resource (available at
191 <http://covid19.cancer-pku.cn/> for quick browsing) enabled accurate annotation and analysis
192 of these cell clusters at different resolutions (Figure 1C, Figure S1F-J and Table S2), which
193 allow the elucidation of potential molecular and cellular mechanisms underlying the
194 pathogenesis of SARS-CoV-2 infection and differences of human immune responses for
195 patients with distinct symptoms.

196
197 Notable differences could be observed in the immune compositions of healthy controls and
198 COVID-19 patients with mild/moderate or severe symptoms (Figure 1D) or between the
199 disease progression stages and convalescence (Figure 1E) based on the t-distributed
200 stochastic neighbor embedding (t-SNE) projection. The tissue preference of each cluster
201 was illustrated based on the ratio of observed to randomly expected cell numbers (Ro/e,
202 Figure 1F), partially reflecting the validity of cell clustering. Notably, various clusters of
203 proliferating CD8⁺ and CD4⁺ T, and plasma B cells were more enriched in BALF than
204 PBMCs, indicating activated adaptive immune responses in the lung (Figure 1F).

205
206 We first analyzed the compositional changes of the broad categories of immune cells for
207 PBMCs in different COVID-19 patient groups. Notably, the percentages of megakaryocytes
208 and monocytes in PBMCs were elevated, particularly in severe COVID-19 patients during

209 the disease progression stage (Figure 2A) (Guo et al., 2020a; Zhang et al., 2020b). While
210 NK cells did not show significant changes among the different patient groups (Figure 2A), B
211 cells were significantly increased in severe COVID-19 patients (Figure 2A)(Guo et al.,
212 2020a; Mathew et al., 2020b; Zhang et al., 2020b). By contrast, T cells and DCs were
213 decreased in severe COVID-19 patients (Figure 2A). These findings are consistent with
214 previous reports that lymphopenia is frequently observed in COVID-19 patients and that
215 impaired adaptive immunity may occur (Gao et al., 2020; Giamarellos-Bourboulis et al.,
216 2020; Kuri-Cervantes et al., 2020; Ni et al., 2020a; Yu et al., 2020).

217

218 **Elevated plasma B cells in COVID-19 patients**

219 As single-cell dissection can provide the finest resolution to investigate the compositional
220 changes among different COVID-19 patient groups, we then examined the heterogeneity of
221 sub-clusters within each major immune cell type. For B cells, XBP1+ plasma cells (B_c05-
222 MZB1-XBP1) showed the most significant compositional increases in PBMCs. For some
223 severe COVID-19 patients, the percentage of plasma cells could even reach 15% of CD45+
224 cells in PBMCs, but the levels in other COVID-19 patients and healthy controls were less
225 than 3% (Figure 2B). These cells highly expressed the genes encoding the constant
226 regions of IgA1, IgA2, IgG1 or IgG2 (Figure 2B), indicating their functions to secrete
227 antigen-specific antibodies to combat viral infection. This observation is consistent with the
228 recent report that the serum of severe COVID-19 patients had high titers of SARS-CoV-2-
229 specific antibodies (Tan et al., 2020b; Zhang et al., 2020c).

230

231 The increased plasma B cells in peripheral blood appeared to be derived from active
232 proliferation of plasmablasts and transitions from memory B cells based on the paired BCR
233 sequencing analyses. Both the extent of BCR clonal expansion and the diversity of the total
234 BCR repertoire of these cells were significantly increased in severe COVID-19 patients
235 (Figure 2C). Plasmablast cells (B_c06_MKI67), characterized by high expression of MKI67
236 and thus indicating a proliferative state, were elevated in the peripheral blood of severe
237 COVID-19 patients (Figure S2A) and shared the most clonotypes with plasma cells (Figure
238 2D). The memory B cell cluster expressing high levels of *CD27*, *CD80*, *AIM2*, *GRIP2*, and
239 *COCH* (B_c03-CD27-AIM2) was the second major source of plasma B cells in the
240 peripheral blood, which shared a large proportion of clonotypes with plasma cells and
241 plasmablasts (Figure 2D). Distinct from plasma cells and plasmablasts which were mainly
242 composed of IgAs and IgGs, B_c03-CD27-AIM2 had a higher proportion of IgMs (Figure
243 2E), indicating a precursor state.

244

245 We applied analysis of variance (ANOVA) to dissect the associations of compositional
246 changes of plasma B cells with disease severity, stage (progression or convalescence), age,
247 sex, or the interactions of these factors. We found that plasma B cells in blood were
248 specifically associated with the disease severity of COVID-19, and then disease stage, but
249 had no associations with age or sex observed (Figure 2F) (Takahashi et al., 2020). In fact,
250 for the mild/moderate disease, convalescent patients harbored higher levels of plasma B
251 cells than those in the disease progression stage. By contrast, the plasma B cell levels in
252 convalescent patients who recovered from severe disease were significantly lower than
253 those in the disease progression stage (Figure 2B). Interestingly, the precursors of plasma
254 B cells, *i.e.*, cells of B_c03-CD27-AIM2, appeared to be associated with sex differences
255 (Figure 2G). In females, the percentage of B_c03-CD27-AIM2 cells was significantly higher
256 than that of males (Figure 2G). Almost all B cell clusters were associated with disease
257 stages, implying the importance of humoral immune response changes between disease
258 progression and convalescence (Figure S2B and Table S3).

259

260 In summary, plasma B cells appeared to be significantly elevated in the peripheral blood
261 regarding either the composition, proliferation, or developmental transition from memory B
262 cells, and were more associated with disease severity. While their precursor cells were also
263 elevated, they were more prone to be influenced by sex differences, providing a plausible
264 explanation for the epidemiological observations on sex differences of COVID-19. Taken
265 together with the observation that plasma B cells were more enriched in BALF (Figure 1F),
266 these observations may suggest that humoral immune responses were actively initiated to
267 combat SARS-CoV-2 infection and contributed to disease severity.

268

269 **Elevated proliferative T cells in COVID-19 patients**

270 Similar to plasma B cells, proliferative CD8+ and CD4+ T cell clusters also showed an
271 enrichment in BALF (Figure 1F) and elevation in PBMCs of COVID-19 patients albeit with a
272 decrease of total T cells (Figures 2A and 3A) (Liao et al., 2020). A total of three proliferative
273 CD8+ T cell clusters identified in this study, including T_CD8_c10-MKI67-GZMK,
274 T_CD8_c11-MKI67-FOS, and T_CD8_c12-MKI67-TYROBP, were all increased in COVID-
275 19 patients but with different characteristics. T_CD8_c10-MKI67-GZMK, a proliferative
276 effector memory CD8+ T cell group characterized by high expression of *STMN1*, *HMGB2*,
277 *MKI67*, and *GZMK*, was increased in the convalescent stage of severe COVID-19 patients
278 (Figure 3B). Similarly, T_CD8_c11-MKI67-FOS also highly expressed *STMN1*, *HMGB2*,
279 and *MKI67*, but exhibited low levels of *GZMK* instead and high levels of *FOS*. This cluster
280 was increased in the disease progression stage of severe patients but not in convalescence
281 (Figure 3B). T_CD8_c12-MKI67-TYROBP was characterized by high expression of *STMN1*,
282 *HMGB2*, *MKI67*, and a NK cell marker gene *TYROBP*. This cluster was specifically
283 increased in mild/moderate patients during the disease progression stage but decreased in
284 the convalescence to a normal level as in healthy controls (Figure 3B). These observations
285 were consistent with the activation of T cell responses in the peripheral blood of COVID-19
286 patients as previously reported using flow cytometry or CyTOF techniques (Mathew et al.,
287 2020a; Sekine et al., 2020). However, the variations of proliferative CD8+ T cell clusters in
288 different severity and stages have not been noticed before, which may indicate the
289 complexity of T cell responses induced by SARS-CoV-2 infection in different patients.
290 Moreover, in contrast to plasma B cells that accounted for 6.39% of total B cells, each
291 proliferative CD8+ T cell cluster accounted for a much smaller proportion of the total CD8+
292 T cells (<1.63%) (Guo et al., 2020a; Liao et al., 2020; Mathew et al., 2020a).

293

294 Two proliferative CD4+ T cell clusters were also identified, with T_CD4_c13-MKI67-
295 CCL5^{low} characterized by high expression of *SELL* and low *CCL5* and T_CD4_c14-MKI67-
296 CCL5^{high} characterized by low *SELL* and high *CCL5*. The counts of T_CD4_c14-MKI67-
297 CCL5^{high} in PBMCs did not show significant differences among different COVID-19 patients.
298 By contrast, the T_CD4_c13-MKI67-CCL5^{low} counts were elevated in COVID-19 patients,
299 particularly in severe patients during the disease progression stage (Figure 3C). Similar to
300 plasma B cells, the diversity and clonality of this cluster were both increased in severe
301 patients with disease progression (Figure 3D), indicating an expanded TCR repertoire and
302 developmental transitions from other clusters. Unlike plasma B cells whose source cluster
303 B_c03-CD27-AIM2 was increased in peripheral blood (Figure S2C), the major source
304 cluster of proliferative CD4+ T cells T_c04_CD4-ANXA2 was decreased in COVID-19
305 patients, particularly in severe patients during the disease progression stage (Figures 3E
306 and 3F). This may partially explain the dichotomous and incomplete adaptive immunity
307 previously observed in COVID-19 patients (Catanzaro et al., 2020). ANOVA analyses
308 revealed that different from T_CD4_c13-MKI67-CCL5^{low} (Figure 3G), the percentage of

309 T_CD4_c04-ANXA2 was associated with disease severity, progression/convalescence,
310 and sex (Figure 3H). In particular, female patients generally had higher levels of
311 T_CD4_c04-ANXA2 than males (Figure 3H), indicating the sex differences of T cell
312 responses to SARS-CoV-2 infection (Takahashi et al., 2020).

313

314 In contrast to proliferative T cells that were elevated in PBMCs, other T cell clusters showed
315 decrease in COVID-19 patients albeit with varied magnitudes, consistent with the
316 lymphopenia that has been frequently observed in COVID-19 patients (Giamarellos-
317 Bourboulis et al., 2020; Kuri-Cervantes et al., 2020; Yu et al., 2020). The most significantly
318 decreased T cell clusters included $\gamma\delta$ T cells (T_c14_gdT-TRDV2), MAIT cells (T_CD8_c09-
319 SLC4A10), a CD8+ T cell cluster highly expressing *TYROBP*, *KLRF1*, *CD247* and *IL2RB*
320 (T_CD8_c08-IL2RB), and three CD4+ T cell clusters showing effector memory
321 characteristics (T_CD4_c09-GZMK-FOS^{low}, T_CD4_c11-GNLY, and T_CD4_c04-ANXA2).
322 ANOVA analysis suggested that these clusters were mainly associated with disease
323 severity rather than age or sex (Figures S2D and S2E and Table S3), implicating their
324 critical roles in the disease progression of COVID-19. In particular, decreases of $\gamma\delta$ T cells
325 and MAIT cells in the peripheral blood of COVID-19 patients (Figures S2F and Figure S2G)
326 have been supported by flow cytometry-based analyses, suggesting their potent
327 antimicrobial functions (Jouan et al., 2020).

328

329 While the decrease of $\gamma\delta$ T cells, MAIT cells, and effector memory T cells abovementioned
330 were primarily associated with disease severity, the decreases of naive and central memory
331 T cells were associated with the age but not sex difference of patients (Figure S2D and S2E
332 and Table S3). Such clusters included the naive CD8+ cluster T_CD8_c01-LEF1, the CD8+
333 central memory cluster T_CD8_c02-GPR183, the naive CD4+ cluster T_CD4_c01-LEF1,
334 and two CD69+ CD4+ clusters T_CD4_c06-NR4A2 and T_CD4_c05-FOS.

335

336 In summary, our scRNA-seq study recapitulated the lymphopenia in COVID-19 patients
337 frequently observed in previous studies (Dhama et al., 2020; Giamarellos-Bourboulis et al.,
338 2020; Kuri-Cervantes et al., 2020; Tan et al., 2020a; Yan et al., 2020; Yu et al., 2020). We
339 further confirmed the activation of both CD4+ and CD8+ T cell responses in PBMCs
340 recently found by flow cytometry-based immune profiling (Mathew et al., 2020a; Sekine et
341 al., 2020). With the high resolution provided by scRNA-seq, we revealed the existence of
342 distinct proliferative T cell clusters for both CD4+ and CD8+ T cells in COVID-19 patients,
343 and implicated their different roles in patients of different groups and stages. Our large
344 cohort also enabled us to dissect the impact of age and sex on the immune responses of
345 COVID-19 patients. We found that, rather than associated with T cell proliferation, age and
346 sex are more likely associated with the abundance of naive/central memory T cells and the
347 precursor cells of proliferative T cells, respectively, highlighting the complexity of human T
348 cell responses to SARS-CoV-2 infection.

349

350 **TCR/BCR usage patterns by COVID-19 patients**

351 Our scRNA-seq data also coupled with TCR/BCR repertoire sequencing and thus provided
352 a rich resource to investigate the TCR/BCR usage of COVID-19 patients, which is
353 instructive for the development of anti-SARS-CoV-2 therapeutics and vaccines. We first
354 examined whether identical TCRs or BCRs could be identified across COVID-19 patients.
355 We found that only a few TCRs or BCRs were shared between two patients, and no
356 identical TCRs or BCRs were shared beyond three patients. No TCRs or BCRs had
357 identical amino acid sequences in more than three patients for the complementarity
358 determining regions 3 (CDR3s) of β chains of TCRs or heavy chains of BCRs. We further

359 examined whether the amino acid sequences of CDR3s of published SARS-CoV-2-reacting
360 antibodies could be identified in the BCR repertoire of this cohort. We found that only one
361 non-clonal BCR had identical CDR3 in its heavy chain with a comprehensive compendium
362 containing 1,505 SARS-CoV-2-specific antibodies (Cao et al., 2020). Such scarcity of
363 common TCRs or BCRs was in contrast with previous studies on severe patients recovered
364 from enterovirus A71 infection and influenza vaccination (Chen et al., 2017; Jiang et al.,
365 2013), suggesting that SARS-CoV-2 infection might not impose dramatic selective pressure
366 on the somatic evolution of TCRs and BCRs.

367
368 Although no identical BCRs were found, we noticed that the BCR repertoire of COVID-19
369 patients had biased VDJ usage compared with that of healthy controls. We trained a
370 random forest classifier with the VDJ usage frequencies to discriminate COVID-19 patients
371 with mild/moderate or severe symptoms from healthy controls and found that the
372 classification accuracy measured by the values of area under curve (AUC) could reach as
373 high as 0.85. The most important VDJ combinations selected by the random forest
374 classifiers also had significant overlaps with those of experimentally verified SARS-CoV-2-
375 reacting antibodies (Figure S2H). Among the top 20 VDJ combinations important to
376 discriminate severe COVID-19 patients from healthy controls selected by random forests,
377 14 had identical VDJ usage with experimentally verified neutralizing antibodies. Of note, the
378 VDJ usage of the currently known SARS-CoV-2-neutralizing antibodies was biased towards
379 IGHV3 and IGHV1. In particular, more than 40 neutralizing antibodies used IGHV3–53.
380 Such observations are important to the development of effective diagnostics to trace human
381 infection history and the further refinement of the current neutralizing antibodies.

382
383 The diversity of TCR or BCR repertoires of various T and B clusters might also be
384 influenced by age, sex, COVID-19 severity, and disease stages. While age was mainly
385 associated with the abundance of only naive and central memory T cells in PBMC, ANOVA
386 analysis revealed that age might influence the decrease of TCR diversity in a wider range of
387 T cells, including naive, central memory, and diverse effector memory T cells (Figure S2I-J).
388 By contrast, sex differences were mainly associated with the BCR diversity of naive and
389 memory B cells (B_c01-TCL1A, B_c02-MS4A1-CD27, and B_c04-SOX5-TNFRSF1B) and
390 the TCR diversity of a subset of effector memory CD4+ T cells (T_CD4_c08-GZMK-FOS^{high})
391 (Figure S2I-K). After correcting the effects of age and sex, the decrease of diversity in MAIT
392 cells, naive B and CD8+ and CD4+ T cells, effector memory CD8+ T cells (T_CD8_c03-
393 GZMK and T_CD8_c04-COTL1), and a few CD69+ CD4+ T cell clusters (T_CD4_c03-
394 ITGA4, T_CD4_c04-ANXA2) remained independently associated with COVID-19 severity
395 (Figure S2I-K), highlighting the importance of these cells in COVID-19. Importantly, the
396 TCR diversity of one proliferative CD8+ T cell cluster, *i.e.*, T_CD8_c11-MKI67-FOS, was
397 associated with the triad interaction by disease severity, age, and sex (Figure S2J),
398 indicating the impacts of age and sex on disease severity. Similarly, the clonal expansion of
399 a central memory CD4+ T cell cluster highly expressing AQP3 (T_CD4_c02-AQP3) was
400 also associated with the triad interaction by disease severity, age, and sex (Table S3),
401 indicating that age and sex might impact the COVID-19 disease via multiple mechanisms.

402
403 Taken together, our data suggested that SARS-CoV-2 might not impose dramatic selective
404 pressure on the somatic evolution of TCRs or BCRs for COVID-19 patients, thus resulting
405 in few identical TCRs and BCRs across patients. However, preferential VDJ usage were
406 identified, which highly overlapped with the sequences of some known SARS-CoV-2-
407 neutralizing antibodies. The diversity of TCR and BCR repertoires of various T and B
408 clusters might be shaped by age, sex, disease severity, and stages together, although the

409 influences of these factors were heterogeneous on different cell types. In particular, triad
410 interactions among age, sex and disease severity were indicated for specific CD8+ and
411 CD4+ clusters, underscoring the complex T cell responses of COVID-19 patients and
412 providing important clues for future studies.

413

414 **SARS-CoV-2 detected in multiple epithelial and immune cell types with** 415 **interferon response phenotypes**

416 The enrichment of plasma B and proliferative T cells in BALF and the elevation of these
417 cells in PBMCs of COVID-19 patients highlighted the roles of these cells in combating
418 SARS-CoV-2 infection. To explore potential interactions between these cells and SARS-
419 CoV-2 infected cells, we examined the characteristics of cell types that harbored SARS-
420 CoV-2 sequences in our dataset. From the BALF samples of severe COVID-19 patients in
421 the disease progression stage, we identified viral RNAs of SARS-CoV-2 in three epithelial
422 cell types including ciliated, secretory, and squamous epithelial cells and a diverse set of
423 immune cells including neutrophils, macrophages, plasma B cells, T cells, and NK cells
424 (Figure 4A). The cell identities of these SARS-CoV-2-positive cells were well confirmed by
425 their corresponding molecular markers (Figure 4B), excluding the possibility of artefacts
426 caused by doublets during scRNA-seq. Because ACE2 and TMPRSS2 have been
427 recognized to play critical roles in mediating viral entry into the host cells for SARS-CoV-2,
428 we examined their expression levels in these cells (Figure 4C) (Netea et al., 2020). We
429 found that at least a subset of those epithelial cells expressed *ACE2* and *TMPRSS2*,
430 consistent with the notion that SARS-CoV-2 employs ACE2 and TMPRSS2 to invade these
431 cells. Interestingly, those immune cells, which did not express *ACE2* or *TMPRSS2*,
432 harbored even more viral RNA sequences than the epithelial cells (Figure 4D). The high
433 viral load reassured that the detection of SARS-CoV-2 RNAs in these immune cells was
434 unlikely caused by experimental contamination. Consistently, an independent scRNA-seq
435 study of COVID-19 patients also identified SARS-CoV-2 RNAs in neutrophils and
436 macrophages from the respiratory samples of COVID-19 patients (Bost et al., 2020).

437

438 Since interferon-stimulated genes (ISGs) are typically activated in virus-infected cells
439 (Schoggins and Rice, 2011), we next examined the expression of ISGs in these cells
440 (Figure 4E and Table S4). Because *IFIT1/2/3* and *IFITM1/2/3* are frequently observed to
441 increase after various viral infections (Zhang et al., 2016), the high expression of these
442 genes in these immune cells, particularly neutrophils and macrophages, may indicate ISG
443 activation in these cells. Compared with matched cell types in PBMC, almost all these *ISG*
444 genes exhibited elevated expression in these virus+ immune cells (Figures 4F and S3B and
445 Table S5). Compared with virus-negative immune cells of the same types in the BALF,
446 SARS-CoV-2+ epithelial cells, including ciliated, secretory, and squamous cells, as well as
447 those virus-positive neutrophils, exhibited higher levels of *ISG* expression (Figure 4F and
448 Table S5). Positive correlations between the viral loads estimated by the abundance of viral
449 RNAs and the *ISG* expression levels were observed for squamous epithelial cells but not
450 ciliated or secretory epithelial cells (Figures 4G and S3C). For immune cells, virus-positive
451 neutrophils exclusively demonstrated positive correlations between viral loads and *ISG*
452 levels (Figure 4G), but this phenomenon did not exist in other immune cell types. These
453 observations suggest that SARS-CoV-2 might be able to infect human cells beyond
454 traditionally assumed respiratory epithelial cells and could induce interferon responses.

455

456 In BALF from mild/moderate COVID-19 patients, fewer cells were obtained and no SARS-
457 CoV-2 RNAs were detected in cells from such samples, suggesting that the respiratory tract

458 of mild/moderate patients might be more intact and the viral titer was lower than those from
459 severe patients. Although type II alveolar (AT2) cells were reported vulnerable to SARS-
460 CoV-2 infection (Hou et al., 2020), our study revealed few AT2 cells in the BALF and no
461 detectable SARS-CoV-2 RNAs in AT2 cells, which is consistent with the previous finding
462 that lower respiratory tract cells had lower potential to be infected by SARS-CoV-2 than
463 those from nasal and upper respiratory tract (Hou et al., 2020; Sungnak et al., 2020).

464

465 **Distinct transcriptomic changes of ciliated, secretory, and squamous** 466 **epithelial cells after SARS-CoV-2 infection**

467 SARS-CoV-2 infection in different epithelial cells resulted in not only distinct interferon
468 responses, but also significant transcriptomic changes. For squamous epithelial cells,
469 SARS-COV-2+ cells exhibited elevated expression of a diverse set of genes such as *NT5E*,
470 *CLCA4*, and *SULT2B1* (Figure 5A). These genes were enriched in pathways such as
471 “response to virus”, “response to type I interferon” and “response to hypoxia”, consistent
472 with viral infection and the subsequent respiratory distress, reflecting the host immune
473 response via type I interferons (Figure 5B). By contrast, the numbers of genes with
474 significant changes after SARS-CoV-2 infection for ciliated and secretory epithelial cells
475 were much smaller than squamous cells, and few genes showed consistent changes in all
476 the three epithelial cell types (Figure 5C).

477

478 We next explored the impact of the above transcriptomic changes, especially on their
479 interaction potentials with immune cells. Annexin A1 (*ANXA1*), up-regulated in virus+
480 squamous epithelial cells (Figure 5D), is known to regulate the functions of neutrophils in
481 inflammation via its interactions with formyl peptide receptors (Sugimoto et al., 2016). This
482 prompted us to investigate the cellular interaction changes of epithelial cells with each other
483 and with immune cells after SARS-CoV-2 infection. Based on CSOmap that estimates cell-
484 cell interactions in three-dimensional space via ligand-receptor (LR) mediated cell self-
485 organization and competition (Ren et al., 2020), we estimated the cellular interaction
486 potentials in a computationally constructed pseudo-space and found that ciliated, secretory,
487 and squamous epithelial cells exhibited distinct interaction potentials after SARS-CoV-2
488 infection.

489

490 Ciliated epithelial cells exhibited lower interaction potentials with themselves and other cells
491 after SARS-CoV-2 infection, and thus would disperse in the outer compartment of the
492 pseudo-space (Figures 5E, S4A and S4B), consistent with the pathological phenomenon of
493 epithelial denudation of coronavirus infection in respiratory tract (Lee et al., 2003; Nicholls
494 et al., 2003). By contrast, squamous epithelial cells significantly enhanced their interacting
495 potentials with themselves after SARS-CoV-2 infection compared with those squamous
496 cells with no viral detection (Figure S4C). Such changes were consistent across COVID-19
497 patients (Figure 5F). Comparison across ciliated, secretory, and squamous epithelial cells
498 infected by SARS-CoV-2 also highlighted the dispersing tendency of ciliated cells and the
499 interacting potentials among squamous cells themselves (Figures 5G and 5H).

500

501 Such interaction distinctions not only existed among epithelial cells, but also impacted their
502 interactions with immune cells. Consistent with the dispersing nature of ciliated cells in the
503 outer compartment of the pseudo-space, no significant interactions were observed between
504 virus+ ciliated cells and immune cells. By contrast, virus+ secretory epithelial cells showed
505 significant interactions with neutrophils and macrophages in mild/moderate COVID-19
506 patients via the SCGB3A1-MARCO axis (Figures S4D and S4E), but such interactions were
507 subdued in severe COVID-19 patients due to the down-regulation of *MARCO* in neutrophils

508 and macrophages (Figure S4F). In severe patients, virus+ squamous cells showed
509 significant interactions with neutrophils and macrophages via the ANXA1-FPR1 and
510 S100A9/A8-TLR4 axes (Figure 5I). Neutrophils and macrophages exhibiting high interacting
511 potentials with virus+ squamous epithelial cells were also prone to be SARS-CoV-2 infected
512 (Figure 5J). As ANXA1-FPR1 and S100A9/A8-TLR4 interactions have been reported to
513 play critical roles in the recruitment of immune cells and inflammatory cascade under
514 various conditions including sepsis and tumor (Gavins et al., 2012; Laouedj et al., 2017;
515 Osei-Owusu et al., 2019; Vogl et al., 2007), they might also play important roles in the
516 pathogenesis of SARS-CoV-2 infection. In contrast to innate immune cells such as
517 neutrophils and macrophages, T and B cells did not show significant interactions with any of
518 the three types of virus+ epithelial cells (Figure S4G), implying a compromised adaptive
519 immune response. It is noteworthy that plasma B cells in BALF also tended to be SARS-
520 CoV-2-positive and displayed close interactions with virus+ neutrophils and squamous
521 epithelial cells via the S100A9/A8-TLR4 axes (Figure 5L).

522

523 We then investigated the cell types expressing *ANXA1*, *FPR1*, *S100A9*, *S100A8*, and *TLR4*
524 in both BALF and PBMC across COVID-19 patients to evaluate the possible inflammatory
525 cascade mediated by these LR pairs. It was evident that *ANXA1* was highly expressed in a
526 wide range of immune cells except B cells and naive T cells (Figures S5A and S5B) and its
527 receptor *FPR1* was highly expressed in neutrophils, macrophages, and monocytes (Figures
528 S5A and S5B). Interestingly, for most immune cell clusters in BALF, the expression levels
529 of *ANXA1* and *FPR1* were down-regulated in severe COVID-19 patients compared with
530 those of mild/moderate COVID-19 patients (Figure S5A). But in PBMCs, except for MAIT
531 cells (T_CD8_c09-SLC4A10) and $\gamma\delta$ T cells (T_gdT_c14-TRDV2), *ANXA1* and *FPR1* were
532 significantly up-regulated in many cell types in severe COVID-19 patients compared with
533 those of mild/moderate COVID-19 patients (Figure S5B). *S100A9* and *S100A8* were highly
534 expressed in neutrophils, macrophages, and monocytes in COVID-19 patients with
535 mild/moderate symptoms and had no expression in T, B, NK, or dendritic cells (Figures
536 S4H and S5B). However, for severe COVID-19 patients in the disease progression stage,
537 *S100A9* and *S100A8* were significantly up-regulated in almost all cell clusters for both
538 BALF and PBMCs (Figures S4H and S5B). In particular, T, B, NK, and dendritic cells had
539 no or minimal levels of *S100A9* and *S100A8* expression in mild/moderate COVID-19
540 patients (Figures S4H and S5B). By contrast, in severe COVID-19 patients, the levels of
541 *S100A9* and *S100A8* were significantly up-regulated in T, B, NK, and dendritic cells
542 (Figures S4H and S5B), indicating a systemic inflammatory response. *TLR4* did not exhibit
543 significant differences in PBMCs between severe and mild/moderate COVID-19 patients but
544 was significantly down-regulated in certain BALF monocyte and macrophage subsets
545 (Figure S5B).

546

547 In summary, our data indicated that SARS-CoV-2 infection in different types of epithelial
548 cells might trigger different transcriptomic changes and thus could modulate their
549 interactions with themselves and with immune cells. In particular, squamous epithelial cells
550 could up-regulate *ANXA1* and *S100A8/A9* after SARS-CoV-2 infection, enhancing their
551 interactions with neutrophils and macrophages via the axes of ANXA1-FPR1 and
552 S100A8/A9-TLR4. The systemic up-regulation of *ANXA1*, *FPR1*, and *S100A8/A9* in
553 immune cells from peripheral blood may indicate, at least partially, the molecular
554 mechanism of aberrant inflammation in severe COVID-19 patients. This hypothesis is
555 supported by a preliminary finding that small molecules targeting S100A8/A9 could inhibit
556 SARS-CoV-2-induced aberrant inflammation in mice (Guo et al., 2020b). Thus, S100A8/A9
557 should be further evaluated as therapeutic targets. Compared to innate immune cells,

558 adoptive immune cells including T and B cells did not show significant interactions with
559 SARS-CoV-2-positive epithelial cells in BALF by computational simulation, consistent with
560 previous findings (Chua et al., 2020; Wauters et al., 2020). These might suggest a
561 compromised adaptive immune response in severe patients. Furthermore, the
562 megakaryocytes in PBMCs, followed by monocytes, exhibited higher interaction potentials
563 with epithelial and immune cells in BALF than adaptive immune cells (Figure S4I),
564 suggesting the critical roles of these cells in the pathogenesis of COVID-19.

565

566 **Megakaryocytes and monocyte subsets as critical peripheral sources of** 567 **cytokine storms**

568 With our large scale scRNA-seq dataset, we next sought to investigate whether any crucial
569 cell subtypes in peripheral blood contribute to the bulk of inflammatory cytokine production.
570 We first defined a cytokine score and inflammatory score for each cell based on the
571 expressions of the collected cytokine genes and reported inflammatory response genes
572 (Liberzon et al., 2015) (Table S6), respectively, and used these two scores as indicators to
573 evaluate the levels of inflammatory cytokine storm for each cell. We found apparent
574 elevated expression of cytokine and inflammatory genes in patients, especially at the
575 severe progression stage (Figures 6A and S6A), indicating the existence of inflammatory
576 cytokine storm after SARS-CoV-2 infection. Seven cell subtypes, including three subtypes
577 of monocytes (Mono_c1-CD14-CCL3, Mono_c2-CD14-HLA-DPB1 and Mono_c3-CD14-
578 VCAN), three subtypes of T cells (T_CD4_c08-GZMK-FOS^{high}, T_CD8_c06-TNF and
579 T_CD8_c09-SLC4A10) and one subtype of megakaryocytes was detected with significantly
580 higher cytokine and inflammatory scores (Figure. S9B, Table S2, $P < 0.0001$), indicating
581 that these cells might be major sources of inflammatory storm. Interestingly,
582 megakaryocytes, which have not been reported in the inflammatory response in COVID-19
583 patients, may affect the functions of platelets at the disease stage, in consistent to a
584 previous study (Manne et al., 2020).

585 Each of the hyper-inflammatory subtypes highly expressed several cytokine genes that are
586 known to be involved in the inflammatory storm, such as *CCL3*, *IL1B*, *CXCL8*, *CCL4*, *CCL6*,
587 *IL32*, *LTB* and *TGFB1*, but with different patterns (Figure 6B), suggesting divergent
588 genomic signatures of these cells. We then investigated the proportion of each of the 7 cell
589 subtypes in every patient and found that these hyper-inflammatory cell subtypes were in
590 general slightly more frequent in patients at severe stage (Figure. S6C). When we clustered
591 these cell subtypes with each individual patient based on the proportions of the hyper-
592 inflammatory cell subtype in PBMCs, we found distinct enrichment of these cell subtypes in
593 different groups of patients (Figure 6C). Mono_c1-CD14-CCL3, known be associated with
594 tocilizumab-responding cytokine storm (Guo et al., 2020a), was highly enriched in a
595 subpopulation of severe onset patients likely to be accompanied by inflammatory storm
596 (Figures 6C and 6D). The proportion of Mono_c1-CD14-CCL3 subtype was also correlated
597 with the age of the corresponding patients (Figure. 6E). The hyper-inflammatory
598 megakaryocytes were enriched in another batch of severe onset patients, which could also
599 be under excessive inflammatory response (Figure. 6C and 6D).

600 By contrast, Mono_c2-CD14-HLA-DPB1 and Mono_c3-CD14-VCAN subtypes were widely
601 distributed in every disease stage, and the hyper-inflammatory T cells showed decreased
602 proportions in patients at the severe onset stage such as T_CD4_c08-GZMK-FOS^{high}
603 subtype (Figures 6C, 6D and S6B), although both of these two monocyte subtypes
604 exhibited increased proportions in elder convalescent patients (Figure 6E). Taken together,
605 these results suggest that Mono_c1-CD14-CCL3 and megakaryocytes were the major

606 sources triggering cytokine inflammatory storm, with both elevated cell ratios and
607 inflammatory scores in the severe onset patients. On the other hand, although the severity
608 of COVID-19 is correlated with lymphopenia, partially reflected by reduced T cells in PBMC
609 (Dhama et al., 2020; Tan et al., 2020a; Yan et al., 2020), certain T cell subtypes might
610 actually contribute to the inflammatory storm by enhanced expressing of proinflammatory
611 cytokines.

612 Next, we investigated the inflammatory signatures for each hyper-inflammatory cell subtype
613 and found unique pro-inflammatory cytokine gene expressions in each cell subtype (Figure
614 6F), suggesting diverse mechanisms by which these cell subtypes may contribute to the
615 cytokine storm. The hyper-inflammatory Mono_c1-CD14-CCL3 and megakaryocytes largely
616 expressed more cytokines, suggesting central roles of the two cell types in driving the
617 inflammatory storm. Specifically, Mono_c1-CD14-CCL3 highly expressed *CXCL8*, *TNF*,
618 *IL1RN*, *IL1B*, and *CCL3*, which we also detected with significantly higher levels in serum
619 from patients at the severe stage, especially those critically ill patients (Figures 6F and
620 S6D). Although the inflammatory megakaryocytes highly expressed the cell type identity
621 marker genes such as *PPBP* (Zhang et al., 1997), the expression level of these genes was
622 significantly decreased in patients compared to healthy controls, indicating a loss of
623 function of these cells after inflammatory activation (Figures 6F and 6G). Notably, the
624 T_CD8_c06-TNF subtype specifically and highly expressed *IFNG*, a pro-inflammatory
625 cytokine highly enriched in patients at the severe onset stage also confirmed by serum
626 cytokine detection (Figures 6F, G and S6D). Moreover, pro-inflammatory cytokines *CXCL8*
627 and *IFNG* showed significant age-dependent expressions in patients with disease
628 progression, while no significance was observed in healthy controls (Figure 6H). *PPBP*
629 showed no correlation with the age in either patients or healthy controls, suggesting that the
630 loss of function of megakaryocytes might not be age-dependent (Figure 6H). To assess the
631 dynamic changes of cytokines in COVID-19 patients with different periods, we compared
632 them with healthy controls for these seven hyper-inflammatory subtypes, and found that
633 *IFNG*, *IL6*, *CCL3*, *TNF*, *CXCL2*, *CXCL8*, *IL1RN*, etc, were highly expressed in cells of
634 severe patients with disease progression (Figure S6E).

635 We also observed eight cell subtypes with significantly higher cytokine scores even though
636 their inflammatory scores showed no difference to other cell clusters (Figure S6B, Table S7,
637 $p < 0.0001$). These cell subtypes exhibited uniform and relatively low expressions of
638 cytokine genes such as *IGF1*, *TXLNA*, *SCYL1*, *CCL5* and *IL16* (Figure 6F), likely not
639 involved in the cytokine storm. No significant differences were observed at the serum level
640 for these cytokines between the different groups of patients (Figure S6F). These genes
641 specific for hyper-inflammatory cells may serve as signatures for the inflammatory storm
642 and be helpful in deepening the understanding of COVID-19 pathogenesis.

643 **Interactions of hyper-inflammatory cell subtypes in lung and peripheral** 644 **blood**

645 The dysregulated cytokine responses associated with the inflammatory cytokine storm may
646 cause immunopathological injury to the lung, and large amount of infiltrating inflammatory
647 immune cells have been demonstrated in the pulmonary tissue of COVID-19 patient
648 (Bhaskar et al., 2020; Cao, 2020; Sun et al., 2020). We analyzed the expressions of
649 cytokines and inflammatory genes for each cell from the BALF samples, and compared the
650 inflammatory and cytokine scores among all the cell subtypes captured in BALF. No
651 enrichment of cytokine genes was observed from the epithelial cells, while subtypes of
652 macrophages and monocytes had the highest cytokine and inflammatory scores in the

653 severe onset samples (Figure 7A). Similar to our analysis on PBMCs, we identified five
654 hyper-inflammatory cell subtypes, including Macro_c2-CCL3L1, the three subtypes of
655 monocytes and the neutrophils (Figure 7B), suggesting that these cell subtypes might be
656 the major sources driving inflammatory storm in the lung tissue. Neither CD4+ nor CD8+ T
657 cells were detected with an elevated inflammatory score or the cytokine score in BALF
658 samples, which was different from those in PBMCs. Each hyper-inflammatory subtype
659 highly expressed specific cytokines; for example, Macro_c2-CCL3L1 specifically expressed
660 *CCL8*, *CXCL10/11*, and *IL6*. Mono_c1-CD14-CCL3, as one of the most notable
661 proinflammatory cell types in both peripheral blood and BALF, uniquely expressed high
662 expression levels of *IL1B*, *CCL20*, *CXCL2*, *CXCL3*, *CCL3*, *CCL4*, *HBEGF* and *TNF*. The
663 neutrophils also showed many uniquely expressed cytokines including *TNFSF13B*, *CXCL8*,
664 *FTH1*, *CXCL16* (Figure 7C).

665 To examine how hyper-inflammatory cells interacted with each other in driving the
666 inflammatory cytokine storm, we analyzed the ligand-receptor pairing patterns among
667 hyper-inflammatory cell subtypes in severe and moderate samples within PBMC and BALF
668 respectively (Figure S7). The interactions between PBMCs and BALF cells appeared to
669 show significant alterations (Figure 7D). Our data revealed elevated ligand-receptor
670 interactions of hyper-inflammatory cells in patients at severe compared to moderate stage.
671 Interestingly, cells in the peripheral blood of severe patients showed much lower
672 interactions with each other compared to those in BALF (Figure. S7A), except for the
673 megakaryocytes, which secreted *IL1B* and stimulated Mono_c1-CD14-CCL3 cells.
674 Mono_c1-CD14-CCL3 cells in BALF expressed *CCR5*, which could receive multiple
675 cytokine stimulations secreted by other cell types in both the lung tissue and the peripheral
676 blood. By contrast, the interactions of Macro_c2-CCL3L1 cells mainly relied on *CCR2* and
677 *IL1R2*. Collectively, these findings illustrated the molecular basis for the potential cell-cell
678 interactions at the pulmonary interface in an inflamed state, leading to a better
679 understanding of the mechanisms of SARS-CoV-2 infection.

680 DISCUSSION

681 Our SC4 alliance members generated scRNA-seq data for 284 clinical samples from 205
682 COVID-19 patients and healthy controls in China, and constructed an information-rich data
683 resource for dissecting the immune responses of COVID-19 patients at the single-cell
684 resolution. We observed a significant reduction of total T cells in the peripheral blood of
685 COVID-19 patients but no notable changes of NK cells, consistent with previous
686 observations (Liao et al., 2020). However, we did not observe a decrease of total B cells,
687 but instead noted elevation in some patients, particularly those with severe symptoms. This
688 contradicts previous studies based on flow cytometry (Giamarellos-Bourboulis et al., 2020),
689 which may reflect sampling fluctuation in small cohorts instead of technology bias, although
690 this has yet to be confirmed. Our findings indicate that T cell changes may be a major
691 cause of the lymphopenia in COVID-19 patients. Despite the overall reduction of total T
692 cells in the peripheral blood, proliferative CD4+ and CD8+ T cells were actually elevated in
693 peripheral blood and were enriched in lung samples, indicating activated cellular immune
694 responses to SARS-CoV-2 infection. Similarly, despite the conflicting reports on total B cell
695 levels, plasma B cells were consistently elevated in patient lung samples and blood,
696 supporting an activated humoral response (Gudbjartsson et al., 2020; Ni et al., 2020b). The
697 complex patterns of T and B cell subtype changes indicate that additional investigations are
698 needed to understand the detailed mechanism by which the cellular and humoral immune
699 responses are activated and compromised in COVID-19 patients.

700

701 Previous clinical and epidemiological studies have revealed obvious sex and age biases in
702 infection rates and disease severity of COVID-19 patients. Our data, covering a wide age
703 range and a sex-balanced COVID-19 cohort, proved to be powerful at dissecting the
704 associations of age and sex in the immune responses to SARS-CoV-2 infection. Our data
705 revealed an apparent involvement of age and sex in the diverse human immune responses
706 via multiple mechanisms, at least partially reflected at the immune cell sub-cluster level. In
707 general, plasma B and proliferative T cells were associated with disease severity, while
708 compositional differences of the precursor cells of these adaptive immune cell types were
709 more prone to be influenced by sex and age seemed to impact more on naive and central
710 memory cells. Of note, age and sex also seemed to impact the diversity of TCR/BCR
711 repertoires for a wide range of T and B cells, which may have clinical implications.

712
713 The single-cell resolution of our data also enabled us to examine the *in vivo* potential host
714 cells of SARS-CoV-2 and the transcriptomic changes caused by SARS-CoV-2 infection. We
715 observed the presence of SARS-CoV-2 RNAs in multiple epithelial cell types in the human
716 respiratory tract, including ciliated, secretory, and squamous cells. Although prominent type
717 I interferon responses could be identified in these cells, distinct transcriptomic changes
718 appeared to be caused by SARS-CoV-2 infection. Such distinctions were exhibited not only
719 in the correlations of interferon responses and viral load, but also in the genes of specific
720 immune relevance, including those encoding LR interactions which are pivotal to cell-cell
721 communications. Of hundreds of immune-relevant LR pairs, ANXA1-FPR1 and S100A8/A9-
722 TLR4 seemed to be critical in mediating the interactions of virus+ squamous epithelial cells
723 and neutrophils and macrophages. Although S100A8/A9 were not expressed in lymphoid
724 cells in mild/moderate COVID-19 patients, they were highly expressed in the T, B, and NK
725 cells of severe patients, likely contributing to the aberrant inflammation of these patients.
726 Coincidentally, small molecule inhibitors of S100A8/A9 could reduce the aberrant
727 inflammation and SARA-CoV-2 replication in mice (Guo et al., 2020b), supporting our
728 findings. Both S100A8/9 and FPR1 should be evaluated further as targets for modulating
729 the immune responses to SARS-CoV-2.

730
731 In addition to epithelial cells, RNAs of SARS-CoV-2 were also identified in various immune
732 cell types, including neutrophils, macrophages, plasma B cells, T and NK cells, often with
733 even higher levels than those in epithelial cells. The viral infection status of these cells
734 could also be supported by the prominent interferon responses in these cells. It is still not
735 clear how such immune cells would acquire viral sequences in the absence of either ACE2
736 or TMPRSS2, but it is evident that the pattern of SARS-CoV-2 infection is more complicated
737 than initial understanding. Such complexity needs to be thoroughly addressed before this
738 dreadful infectious disease can be effectively controlled.

739
740 The rich information of our data also allowed us to dissect the cellular origins of potential
741 cytokine storms. We found that megakaryocytes and a few monocyte subsets might be key
742 sources of a diverse set of cytokines highly elevated in COVID-19 patients with severe
743 disease progression. We suspect that in severe patients, infected epithelial cells would
744 secrete cytokines such as IL1RN into the peripheral, and monocytes expressing IL1R2
745 could be stimulated and in turn produce multiple proinflammatory cytokines such as CXCL8,
746 IL6, IL1B, and TNF (Figure 7E). Through IL1R2, these hyperactive monocytes could also
747 interact with dysfunctional megakaryocytes producing TGFB1, TNFSF4, PF4 and FTH1.
748 Meanwhile, the T cells in the blood go through lymphopenia, while the residual ones are
749 hyperactive in secreting many inflammatory cytokines such as IFNG and TNFSF8. Such
750 proinflammatory cytokines secreted by the cells in the blood could also infiltrate into the

751 lung tissue, and thus activating the tissue resident monocytes, macrophages and
752 neutrophils for further cytokine production. We acknowledge that this is only one of many
753 possible scenarios where an inflammatory storm could form, although our data revealed
754 key actors in the final cytokine screenplay.

755
756 In conclusion, we generated a large scRNA-seq dataset including ~1.5 million single cells
757 covering diverse disease severity and stages. Analyses based on this dataset revealed
758 multiple immune characteristics of SARS-CoV-2 infection with single-cell resolution. Such
759 data provide a critical resource and important insights in dissecting the pathogenesis of
760 COVID-19, and potentially help the development of effective therapeutics and vaccines
761 against SARS-CoV-2.
762

763 **ACKNOWLEDGMENTS**

764 We thank the Computing Platform of the Center for Life Science for their contributions. We also thank 10x
765 Genomics China team for their strong support in coordination and discussion of this study. We thank
766 Analytical BioSciences for their help in building the visualization website of this dataset. We thank Zhuo Zhou
767 of BIOPIC, Peking University for his helpful discussion on the interferon responses of SARS-CoV-2 infection.
768 We thank the USTC supercomputing center and the School of Life Science Bioinformatics Center for
769 providing supercomputing resources for this project. We thank the CAS interdisciplinary innovation team for
770 helpful discussion. We also thank W.Chen, C.Xiao, Y.Zheng, W.Su, Y. Zhang, C. Zhang, X.Wang, H.Ma, T.Jin,
771 X.Wang, H.We, B.Fu, L.Liu, J.Weng, X.Ma for their support to this project. This work was supported by the
772 Beijing Advanced Innovation Centre for Genomics at Peking University, the Ministry of Science and
773 Technology of the People's Republic of China (2020YFC0848700), the National Key R&D Program of China
774 (2017YFA0102900 to K.Q.), the National Natural Science Foundation of China grants (91940306, 81788101,
775 31970858, 31771428, 91640113 to K.Q., 31700796 to C.G., 81871479 to J.L., 81772165, 81974303,
776 31991171, 31530036 and 91742203), the China Primary Health Care Foundation-Youan Medical
777 Development Fund (BJYAYY-2020PY-01), the Beijing Municipal of Science and Technology Major Project
778 (Z2011000054200018), the Fundamental Research Funds for the Central Universities (YD2070002019 to
779 K.Q.).
780

781 **AUTHOR CONTRIBUTIONS**

782 Conceptualization, Z.Z., X.R., R.J., J.C., X.W, K.Q.,Z.Z., H.W., F.W., P.Z., X.L.,T.C., X.L., L.W., J.B., Z.H., Q.J.
783 and P.Z.; Resources, Z.Z., X.R., R.J., J.C., X.W, K.Q.,Z.Z., H.W., F.W., P.Z., X.L.,T.C., X.L., L.W., J.B., Z.H.,
784 Q.J. and P.Z.; Methodology, Z.Z., X.R., R.J., J.C., X.W, K.Q.,Z.Z., H.W., F.W., P.Z., X.L.,T.C., X.L., L.W., J.B.,
785 Z.H., Q.J., P.Z., W.W., X.F., W.H., B.S., P.C., J.L., Y.L., F.T., F.Z., Y.Y., J.H., W.M., X.X. and P.W.;
786 Investigation, Z.Z., X.R., R.J., J.C., X.W, K.Q.,Z.Z., H.W., F.W., P.Z., X.L.,T.C., X.L., L.W., J.B., Z.H., Q.J.,
787 P.Z., W.W., X.F., W.H., B.S., P.C., J.L., Y.L., F.T., F.Z., Y.Y., J.H., W.M., X.X. and P.W.; Validation, W.W. and
788 W.Y.; Writing Original Draft, Z.Z., X.R., R.J., J.C., X.W, K.Q.,Z.Z., H.W., F.W., P.Z., X.L.,T.C., X.L., L.W., J.B.,
789 Z.H., Q.J., P.Z., W.W., X.F., W.H., B.S., P.C., J.L., Y.L., F.T., F.Z., Y.Y., J.H., W.M., X.X. and P.W.
790

791 **DECLARATION OF INTERESTS**

792 Z.Z. is a founder of Analytical Bioscience and an advisor for InnoCare. All financial interests are unrelated to
793 this study. Other authors declare no competing interests.
794
795

796

797 Reference

- 798 Abbasifard, M., and Khorramdelazad, H. (2020). The bio-mission of interleukin-6 in the
799 pathogenesis of COVID-19: A brief look at potential therapeutic tactics. *Life Sci* 257,
800 118097.
- 801 Bhaskar, S., Sinha, A., Banach, M., Mittoo, S., Weissert, R., Kass, J.S., Rajagopal, S., Pai,
802 A.R., and Kutty, S. (2020). Cytokine Storm in COVID-19-Immunopathological Mechanisms,
803 Clinical Considerations, and Therapeutic Approaches: The REPROGRAM Consortium
804 Position Paper. *Front Immunol* 11, 1648.
- 805 Bost, P., Giladi, A., Liu, Y., Bendjelal, Y., Xu, G., David, E., Blecher-Gonen, R., Cohen, M.,
806 Medaglia, C., Li, H., *et al.* (2020). Host-Viral Infection Maps Reveal Signatures of Severe
807 COVID-19 Patients. *Cell* 181, 1475-1488 e1412.
- 808 Bray, N.L., Pimentel, H., Melsted, P., and Pachter, L. (2016). Near-optimal probabilistic
809 RNA-seq quantification. *Nat Biotechnol* 34, 525-527.
- 810 Cao, X. (2020). COVID-19: immunopathology and its implications for therapy. *Nat Rev*
811 *Immunol* 20, 269-270.
- 812 Cao, Y., Su, B., Guo, X., Sun, W., Deng, Y., Bao, L., Zhu, Q., Zhang, X., Zheng, Y., Geng,
813 C., *et al.* (2020). Potent Neutralizing Antibodies against SARS-CoV-2 Identified by High-
814 Throughput Single-Cell Sequencing of Convalescent Patients' B Cells. *Cell* 182, 73-84 e16.
- 815 Catanzaro, M., Fagiani, F., Racchi, M., Corsini, E., Govoni, S., and Lanni, C. (2020).
816 Immune response in COVID-19: addressing a pharmacological challenge by targeting
817 pathways triggered by SARS-CoV-2. *Signal Transduct Target Ther* 5, 84.
- 818 Chen, Z., Ren, X., Yang, J., Dong, J., Xue, Y., Sun, L., Zhu, Y., and Jin, Q. (2017). An
819 elaborate landscape of the human antibody repertoire against enterovirus 71 infection is
820 revealed by phage display screening and deep sequencing. *MAbs* 9, 342-349.
- 821 Chua, R.L., Lukassen, S., Trump, S., Hennig, B.P., Wendisch, D., Pott, F., Debnath, O.,
822 Thurmann, L., Kurth, F., Volker, M.T., *et al.* (2020). COVID-19 severity correlates with
823 airway epithelium-immune cell interactions identified by single-cell analysis. *Nat Biotechnol*
824 38, 970-979.
- 825 Costela-Ruiz, V.J., Illescas-Montes, R., Puerta-Puerta, J.M., Ruiz, C., and Melguizo-
826 Rodriguez, L. (2020). SARS-CoV-2 infection: The role of cytokines in COVID-19 disease.
827 *Cytokine Growth Factor Rev*.
- 828 Del Valle, D.M., Kim-Schulze, S., Huang, H.H., Beckmann, N.D., Nirenberg, S., Wang, B.,
829 Lavin, Y., Swartz, T.H., Madduri, D., Stock, A., *et al.* (2020). An inflammatory cytokine
830 signature predicts COVID-19 severity and survival. *Nat Med*.
- 831 Dhama, K., Khan, S., Tiwari, R., Sircar, S., Bhat, S., Malik, Y.S., Singh, K.P., Chaicumpa,
832 W., Bonilla-Aldana, D.K., and Rodriguez-Morales, A.J. (2020). Coronavirus Disease 2019-
833 COVID-19. *Clin Microbiol Rev* 33.
- 834 Du, F., Liu, B., and Zhang, S. (2020). COVID-19: the role of excessive cytokine release and
835 potential ACE2 down-regulation in promoting hypercoagulable state associated with severe
836 illness. *J Thromb Thrombolysis*.
- 837 Gao, L., Zhou, J., Yang, S., Chen, X., Yang, Y., Li, R., Pan, Z., Zhao, J., Li, Z., Huang, Q.,
838 *et al.* (2020). The dichotomous and incomplete adaptive immunity in COVID-19. *medRxiv*,
839 2020.2009.2005.20187435.
- 840 Gavins, F.N., Hughes, E.L., Buss, N.A., Holloway, P.M., Getting, S.J., and Buckingham, J.C.
841 (2012). Leukocyte recruitment in the brain in sepsis: involvement of the annexin 1-
842 FPR2/ALX anti-inflammatory system. *FASEB J* 26, 4977-4989.
- 843 Giamarellos-Bourboulis, E.J., Netea, M.G., Rovina, N., Akinosoglou, K., Antoniadou, A.,
844 Antonakos, N., Damoraki, G., Gkavogianni, T., Adami, M.E., Katsaounou, P., *et al.* (2020).

845 Complex Immune Dysregulation in COVID-19 Patients with Severe Respiratory Failure. *Cell*
846 *Host Microbe* 27, 992-1000 e1003.
847 Gudbjartsson, D.F., Norddahl, G.L., Melsted, P., Gunnarsdottir, K., Holm, H., Eythorsson,
848 E., Arnthorsson, A.O., Helgason, D., Bjarnadottir, K., Ingvarsson, R.F., *et al.* (2020).
849 Humoral Immune Response to SARS-CoV-2 in Iceland. *N Engl J Med*.
850 Guo, C., Li, B., Ma, H., Wang, X., Cai, P., Yu, Q., Zhu, L., Jin, L., Jiang, C., Fang, J., *et al.*
851 (2020a). Single-cell analysis of two severe COVID-19 patients reveals a monocyte-
852 associated and tocilizumab-responding cytokine storm. *Nat Commun* 11, 3924.
853 Guo, Q., Zhao, Y., Li, J., Liu, J., Guo, X., Zhang, Z., Cao, L., Luo, Y., Bao, L., Wang, X., *et*
854 *al.* (2020b). Small molecules inhibit SARS-COV-2 induced aberrant inflammation and viral
855 replication in mice by targeting S100A8/A9-TLR4 axis. *bioRxiv*, 2020.2009.2009.288704.
856 Hadjadj, J., Yatim, N., Barnabei, L., Corneau, A., Boussier, J., Smith, N., Pere, H., Charbit,
857 B., Bondet, V., Chenevier-Gobeaux, C., *et al.* (2020). Impaired type I interferon activity and
858 inflammatory responses in severe COVID-19 patients. *Science* 369, 718-724.
859 He, J., Cai, S., Feng, H., Cai, B., Lin, L., Mai, Y., Fan, Y., Zhu, A., Huang, H., Shi, J., *et al.*
860 (2020). Single-cell analysis reveals bronchoalveolar epithelial dysfunction in COVID-19
861 patients. *Protein Cell*.
862 Hou, Y.J., Okuda, K., Edwards, C.E., Martinez, D.R., Asakura, T., Dinnon, K.H., 3rd, Kato,
863 T., Lee, R.E., Yount, B.L., Mascenik, T.M., *et al.* (2020). SARS-CoV-2 Reverse Genetics
864 Reveals a Variable Infection Gradient in the Respiratory Tract. *Cell* 182, 429-446 e414.
865 Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., Zhang, L., Fan, G., Xu, J., Gu, X., *et*
866 *al.* (2020). Clinical features of patients infected with 2019 novel coronavirus in Wuhan,
867 China. *Lancet* 395, 497-506.
868 Jamilloux, Y., Henry, T., Belot, A., Viel, S., Fauter, M., El Jammal, T., Walzer, T., Francois,
869 B., and Seve, P. (2020). Should we stimulate or suppress immune responses in COVID-19?
870 Cytokine and anti-cytokine interventions. *Autoimmun Rev* 19, 102567.
871 Jiang, N., He, J., Weinstein, J.A., Penland, L., Sasaki, S., He, X.-S., Dekker, C.L., Zheng,
872 N.-Y., Huang, M., Sullivan, M., *et al.* (2013). Lineage Structure of the Human Antibody
873 Repertoire in Response to Influenza Vaccination. *Science Translational Medicine* 5,
874 171ra119.
875 Jouan, Y., Guillon, A., Gonzalez, L., Perez, Y., Boisseau, C., Ehrmann, S., Ferreira, M.,
876 Daix, T., Jeannet, R., François, B., *et al.* (2020). Phenotypical and functional alteration of
877 unconventional T cells in severe COVID-19 patients. *Journal of Experimental Medicine* 217.
878 Korsunsky, I., Millard, N., Fan, J., Slowikowski, K., Zhang, F., Wei, K., Baglaenko, Y.,
879 Brenner, M., Loh, P.R., and Raychaudhuri, S. (2019). Fast, sensitive and accurate
880 integration of single-cell data with Harmony. *Nat Methods* 16, 1289-1296.
881 Kox, M., Waalders, N.J.B., Kooistra, E.J., Gerretsen, J., and Pickkers, P. (2020). Cytokine
882 Levels in Critically Ill Patients With COVID-19 and Other Conditions. *JAMA*.
883 Kuri-Cervantes, L., Pampena, M.B., Meng, W., Rosenfeld, A.M., Ittner, C.A.G., Weisman,
884 A.R., Agyekum, R.S., Mathew, D., Baxter, A.E., Vella, L.A., *et al.* (2020). Comprehensive
885 mapping of immune perturbations associated with severe COVID-19. *Sci Immunol* 5.
886 Laouedj, M., Tardif, M.R., Gil, L., Raquil, M.A., Lachhab, A., Pelletier, M., Tessier, P.A., and
887 Barabe, F. (2017). S100A9 induces differentiation of acute myeloid leukemia cells through
888 TLR4. *Blood* 129, 1980-1990.
889 Lee, N., Hui, D., Wu, A., Chan, P., Cameron, P., Joynt, G.M., Ahuja, A., Yung, M.Y., Leung,
890 C.B., To, K.F., *et al.* (2003). A major outbreak of severe acute respiratory syndrome in
891 Hong Kong. *N Engl J Med* 348, 1986-1994.
892 Liao, M., Liu, Y., Yuan, J., Wen, Y., Xu, G., Zhao, J., Cheng, L., Li, J., Wang, X., Wang, F.,
893 *et al.* (2020). Single-cell landscape of bronchoalveolar immune cells in patients with
894 COVID-19. *Nat Med* 26, 842-844.

895 Liberzon, A., Birger, C., Thorvaldsdottir, H., Ghandi, M., Mesirov, J.P., and Tamayo, P.
896 (2015). The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell*
897 *Syst 1*, 417-425.

898 Lun, A.T., McCarthy, D.J., and Marioni, J.C. (2016). A step-by-step workflow for low-level
899 analysis of single-cell RNA-seq data with Bioconductor. *F1000Res 5*, 2122.

900 Manne, B.K., Denorme, F., Middleton, E.A., Portier, I., Rowley, J.W., Stubben, C.J., Petrey,
901 A.C., Tolley, N.D., Guo, L., Cody, M.J., *et al.* (2020). Platelet Gene Expression and
902 Function in COVID-19 Patients. *Blood*.

903 Mathew, D., Giles, J.R., Baxter, A.E., Greenplate, A.R., Wu, J.E., Alanio, C., Oldridge, D.A.,
904 Kuri-Cervantes, L., Pampena, M.B., D'Andrea, K., *et al.* (2020a). Deep immune profiling of
905 COVID-19 patients reveals patient heterogeneity and distinct immunotypes with
906 implications for therapeutic interventions. *bioRxiv*.

907 Mathew, D., Giles, J.R., Baxter, A.E., Oldridge, D.A., Greenplate, A.R., Wu, J.E., Alanio, C.,
908 Kuri-Cervantes, L., Pampena, M.B., D'Andrea, K., *et al.* (2020b). Deep immune profiling of
909 COVID-19 patients reveals distinct immunotypes with therapeutic implications. *Science 369*.

910 Mehta, P., McAuley, D.F., Brown, M., Sanchez, E., Tattersall, R.S., Manson, J.J., and Hlth
911 Across Speciality Collaboration, U.K. (2020). COVID-19: consider cytokine storm
912 syndromes and immunosuppression. *Lancet 395*, 1033-1034.

913 Melsted, P., Ntranos, V., and Pachter, L. (2019). The barcode, UMI, set format and
914 BUSTools. *Bioinformatics 35*, 4472-4473.

915 Netea, M.G., Giamarellos-Bourboulis, E.J., Dominguez-Andres, J., Curtis, N., van Crevel,
916 R., van de Veerdonk, F.L., and Bonten, M. (2020). Trained Immunity: a Tool for Reducing
917 Susceptibility to and the Severity of SARS-CoV-2 Infection. *Cell 181*, 969-977.

918 Ni, L., Cheng, M.-L., Zhao, H., Feng, Y., Liu, J., Ye, F., Ye, Q., Zhu, G., Li, X., Wang, P., *et*
919 *al.* (2020a). Impaired cellular immunity to SARS-CoV-2 in severe COVID-19 patients.
920 *medRxiv*, 2020.2008.2010.20171371.

921 Ni, L., Ye, F., Cheng, M.L., Feng, Y., Deng, Y.Q., Zhao, H., Wei, P., Ge, J., Gou, M., Li, X.,
922 *et al.* (2020b). Detection of SARS-CoV-2-Specific Humoral and Cellular Immunity in
923 COVID-19 Convalescent Individuals. *Immunity 52*, 971-977 e973.

924 Nicholls, J.M., Poon, L.L., Lee, K.C., Ng, W.F., Lai, S.T., Leung, C.Y., Chu, C.M., Hui, P.K.,
925 Mak, K.L., Lim, W., *et al.* (2003). Lung pathology of fatal severe acute respiratory syndrome.
926 *Lancet 361*, 1773-1778.

927 Osei-Owusu, P., Charlton, T.M., Kim, H.K., Missiakas, D., and Schneewind, O. (2019).
928 FPR1 is the plague receptor on host immune cells. *Nature 574*, 57-62.

929 Ren, X., Zhong, G., Zhang, Q., Zhang, L., Sun, Y., and Zhang, Z. (2020). Reconstruction of
930 cell spatial organization from single-cell RNA sequencing data based on ligand-receptor
931 mediated self-assembly. *Cell Res 30*, 763-778.

932 Schoggins, J.W., and Rice, C.M. (2011). Interferon-stimulated genes and their antiviral
933 effector functions. *Curr Opin Virol 1*, 519-525.

934 Schulte-Schrepping, J., Reusch, N., Paclik, D., Bassler, K., Schlickeiser, S., Zhang, B.,
935 Kramer, B., Krammer, T., Brumhard, S., Bonaguro, L., *et al.* (2020). Severe COVID-19 Is
936 Marked by a Dysregulated Myeloid Cell Compartment. *Cell*.

937 Sekine, T., Perez-Potti, A., Rivera-Ballesteros, O., Strålin, K., Gorin, J.-B., Olsson, A.,
938 Llewellyn-Lacey, S., Kamal, H., Bogdanovic, G., Muschiol, S., *et al.* (2020). Robust T cell
939 immunity in convalescent individuals with asymptomatic or mild COVID-19. *Cell*.

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

943 Silvin, A., Chapuis, N., Dunsmore, G., Goubet, A.G., Dubuisson, A., Derosa, L., Almiré, C.,
944 Henon, C., Kosmider, O., Droin, N., *et al.* (2020b). Elevated Calprotectin and Abnormal
945 Myeloid Cell Subsets Discriminate Severe from Mild COVID-19. *Cell*.
946 Sugimoto, M.A., Vago, J.P., Teixeira, M.M., and Sousa, L.P. (2016). Annexin A1 and the
947 Resolution of Inflammation: Modulation of Neutrophil Recruitment, Apoptosis, and
948 Clearance. *J Immunol Res* 2016, 8239258.
949 Sun, X., Wang, T., Cai, D., Hu, Z., Chen, J., Liao, H., Zhi, L., Wei, H., Zhang, Z., Qiu, Y., *et*
950 *al.* (2020). Cytokine storm intervention in the early stages of COVID-19 pneumonia.
951 *Cytokine Growth Factor Rev* 53, 38-42.
952 Sungnak, W., Huang, N., Becavin, C., Berg, M., Queen, R., Litvinukova, M., Talavera-
953 Lopez, C., Maatz, H., Reichart, D., Sampaziotis, F., *et al.* (2020). SARS-CoV-2 entry factors
954 are highly expressed in nasal epithelial cells together with innate immune genes. *Nat Med*
955 26, 681-687.
956 Takahashi, T., Ellingson, M.K., Wong, P., Israelow, B., Lucas, C., Klein, J., Silva, J., Mao,
957 T., Oh, J.E., Tokuyama, M., *et al.* (2020). Sex differences in immune responses that
958 underlie COVID-19 disease outcomes. *Nature*.
959 Tan, L., Wang, Q., Zhang, D., Ding, J., Huang, Q., Tang, Y.Q., Wang, Q., and Miao, H.
960 (2020a). Lymphopenia predicts disease severity of COVID-19: a descriptive and predictive
961 study. *Signal Transduct Target Ther* 5, 33.
962 Tan, W., Lu, Y., Zhang, J., Wang, J., Dan, Y., Tan, Z., He, X., Qian, C., Sun, Q., Hu, Q., *et*
963 *al.* (2020b). Viral Kinetics and Antibody Responses in Patients with COVID-19. *medRxiv*,
964 2020.2003.2024.20042382.
965 Traag, V.A., Waltman, L., and van Eck, N.J. (2019). From Louvain to Leiden: guaranteeing
966 well-connected communities. *Sci Rep* 9, 5233.
967 Vabret, N., Britton, G.J., Gruber, C., Hegde, S., Kim, J., Kuksin, M., Levantovsky, R., Malle,
968 L., Moreira, A., Park, M.D., *et al.* (2020). Immunology of COVID-19: Current State of the
969 Science. *Immunity* 52, 910-941.
970 Vogl, T., Tenbrock, K., Ludwig, S., Leukert, N., Ehrhardt, C., van Zoelen, M.A.D., Nacken,
971 W., Foell, D., van der Poll, T., Sorg, C., *et al.* (2007). Mrp8 and Mrp14 are endogenous
972 activators of Toll-like receptor 4, promoting lethal, endotoxin-induced shock. *Nature*
973 *Medicine* 13, 1042-1049.
974 Wauters, E., Van Mol, P., Garg, A.D., Jansen, S., Van Herck, Y., Vanderbeke, L., Bassez,
975 A., Boeckx, B., Malengier-Devlies, B., Timmerman, A., *et al.* (2020). Discriminating Mild
976 from Critical COVID-19 by Innate and Adaptive Immune Single-cell Profiling of
977 Bronchoalveolar Lavages. *bioRxiv*, 2020.2007.2009.196519.
978 Wen, W., Su, W., Tang, H., Le, W., Zhang, X., Zheng, Y., Liu, X., Xie, L., Li, J., Ye, J., *et al.*
979 (2020). Immune cell profiling of COVID-19 patients in the recovery stage by single-cell
980 sequencing. *Cell Discov* 6, 31.
981 Wilk, A.J., Rustagi, A., Zhao, N.Q., Roque, J., Martinez-Colon, G.J., McKechnie, J.L., Ivison,
982 G.T., Ranganath, T., Vergara, R., Hollis, T., *et al.* (2020). A single-cell atlas of the
983 peripheral immune response in patients with severe COVID-19. *Nat Med* 26, 1070-1076.
984 Wolf, F.A., Angerer, P., and Theis, F.J. (2018). SCANPY: large-scale single-cell gene
985 expression data analysis. *Genome Biol* 19, 15.
986 Wolock, S.L., Lopez, R., and Klein, A.M. (2019). Scrublet: Computational Identification of
987 Cell Doublets in Single-Cell Transcriptomic Data. *Cell Syst* 8, 281-291 e289.
988 Xie, X., Liu, M., Zhang, Y., Wang, B., Zhu, C., Wang, C., Li, Q., Huo, Y., Guo, J., Xu, C., *et*
989 *al.* (2020). Single-cell Transcriptomic Landscape of Human Blood Cells. *Natl Sci Rev*.
990 Yan, X., Li, F., Wang, X., Yan, J., Zhu, F., Tang, S., Deng, Y., Wang, H., Chen, R., Yu, Z.,
991 *et al.* (2020). Neutrophil to lymphocyte ratio as prognostic and predictive factor in patients
992 with coronavirus disease 2019: A retrospective cross-sectional study. *J Med Virol*.

993 Yang, Y., Shen, C., Li, J., Yuan, J., Wei, J., Huang, F., Wang, F., Li, G., Li, Y., Xing, L., *et al.*
 994 (2020). Plasma IP-10 and MCP-3 levels are highly associated with disease severity and
 995 predict the progression of COVID-19. *J Allergy Clin Immunol* *146*, 119-127 e114.
 996 Yu, G., Wang, L.G., Han, Y., and He, Q.Y. (2012). clusterProfiler: an R package for
 997 comparing biological themes among gene clusters. *OMICS* *16*, 284-287.
 998 Yu, K., He, J., Wu, Y., Xie, B., Liu, X., Wei, B., Zhou, H., Lin, B., Zuo, Z., Wen, W., *et al.*
 999 (2020). Dysregulated adaptive immune response contributes to severe COVID-19. *Cell Res*
 1000 *30*, 814-816.
 1001 Zhang, C., Gadue, P., Scott, E., Atchison, M., and Poncz, M. (1997). Activation of the
 1002 megakaryocyte-specific gene platelet basic protein (PBP) by the Ets family factor PU.1. *J*
 1003 *Biol Chem* *272*, 26236-26246.
 1004 Zhang, F., Gan, R., Zhen, Z., Hu, X., Li, X., Zhou, F., Liu, Y., Chen, C., Xie, S., Zhang, B.,
 1005 *et al.* (2020a). Adaptive immune responses to SARS-CoV-2 infection in severe versus mild
 1006 individuals. *Signal Transduct Target Ther* *5*, 156-156.
 1007 Zhang, J., Sze, D.M., Yung, B.Y., Tang, P., Chen, W.J., Chan, K.H., and Leung, P.H.
 1008 (2016). Distinct expression of interferon-induced protein with tetratricopeptide repeats (IFIT)
 1009 1/2/3 and other antiviral genes between subsets of dendritic cells induced by dengue virus
 1010 2 infection. *Immunology* *148*, 363-376.
 1011 Zhang, J.Y., Wang, X.M., Xing, X., Xu, Z., Zhang, C., Song, J.W., Fan, X., Xia, P., Fu, J.L.,
 1012 Wang, S.Y., *et al.* (2020b). Single-cell landscape of immunological responses in patients
 1013 with COVID-19. *Nat Immunol* *21*, 1107-1118.
 1014 Zhang, L., Pang, R., Xue, X., Bao, J., Ye, S., Dai, Y., Zheng, Y., Fu, Q., Hu, Z., and Yi, Y.
 1015 (2020c). Anti-SARS-CoV-2 virus antibody levels in convalescent plasma of six donors who
 1016 have recovered from COVID-19. *Aging (Albany NY)* *12*, 6536-6542.
 1017 Zhang, L., Yu, X., Zheng, L., Zhang, Y., Li, Y., Fang, Q., Gao, R., Kang, B., Zhang, Q.,
 1018 Huang, J.Y., *et al.* (2018). Lineage tracking reveals dynamic relationships of T cells in
 1019 colorectal cancer. *Nature* *564*, 268-272.
 1020 Zhou, C., Gao, C., Xie, Y., and Xu, M. (2020a). COVID-19 with spontaneous
 1021 pneumomediastinum. *Lancet Infect Dis* *20*, 510.
 1022 Zhou, Y., Fu, B., Zheng, X., Wang, D., Zhao, C., Qi, Y., Sun, R., Tian, Z., Xu, X., and Wei, H.
 1023 (2020b). Pathogenic T cells and inflammatory monocytes incite inflammatory storm in
 1024 severe COVID-19 patients. *Natl Sci Rev*, nwaa041.
 1025
 1026
 1027

1028 **STAR+METHODS**

1029 **KEY RESOURCES TABLE**

REAGENT RESOURCE	or	SOURCE	IDENTIFIER
Antibodies			
Biological Samples			

Critical Commercial Assays		
Fixation/Permeabilization Solution Kit	BD Biosciences	Cat #554714
SureSelectXT Target Enrichment System for Illumina Paired-End Multiplexed Sequencing Library Kit	Aglient	Cat #G9701
TruePrep DNA Library Prep Kit V2 for Illumina	Vazyme Biotech	Cat #TD503
Chromium Single Cell 3 0 Library and Bead kit	10x Genomics	Cat #PN-120237
Chromium Single Cell 30 Chip Kit v2	10x Genomics	Cat #PN-120236
Chromium i7 Multiplex Kit	10x Genomics	Cat #PN-120262
Hiseq 3000/4000 SBS kit	Illunima	Cat #FC-410-1003
Hiseq 3000/4000 PE cluster kit	Illunima	Cat #PE-410-1001
Deposited Data		
Data files for single-cell RNA sequencing (processed data)	This paper	The NCBI GEO database, and the access number is in the process
Data files for single-cell RNA sequencing (raw data)	This paper	The Genome Sequence Archive (GSA), and the access number is in the process
Oligonucleotides		
Software and Algorithms		

Harmony	Korsunsky et al., 2018	https://github.com/pardeike/HarmonyLun
Cellranger v2.3/v2.0.3	10x Genomics	https://support.10xgenomics.com/single-cell-gene-expression/software/pipelines/latest/ahta-is-cell-ranger
kb v0.24.4	Bray et al., 2016; Melsted et al., 2019	https://github.com/pachterlab/kb_python
kallisto v0.46.1	Bray et al., 2016	https://github.com/pachterlab/kallisto
bustools v0.39.3	Melsted et al., 2019	https://github.com/BUStools/bustools
STARTRAC	Zhang et al., 2018	https://github.com/Japrin/STARTRAC
Seurat 2.3.0/3.0	Butler et al., 2018	https://satijalab.org/seurat
scanpy 1.5.1	Wolf et al., 2018	https://scanpy.readthedocs.io/en/latest/
CSOmap	Ren, X. et al.	https://github.com/zhongguojie1998/CSOmap
Other		

1030

1031

1032 **LEAD CONTACT AND MATERIALS AVAILABILITY**

1033 Further information and requests for resources and reagents should be directed to and will
1034 be fulfilled by the Lead Contact, Zemin Zhang (zemin@pku.edu.cn).

1035

1036 **EXPERIMENTAL MODEL AND SUBJECT DETAILS**

1037

1038 **Ethics statement**

1039 This study was approved by the Ethics Committees of respective institutions, with written
1040 informed consents obtained from all participants before sample collection according to
1041 regular principles.

1042

1043 **Human subjects**

1044 A total of 183 patients with COVID-19 and 25 healthy individuals in this study were enrolled
1045 from 36 centers/ laboratories, with samples (n=284) collected. Samples of COVID-19 were
1046 further categorized into groups of moderate convalescence (n=83), moderate progression
1047 (n=33), severe convalescence (n=51) and severe progression (n=83) according to disease
1048 severity (moderate or severe) and stages (progression and convalescence) based on the
1049 Guidelines for Diagnosis and Treatment of Corona Virus Disease 2019 issued by the

1050 National Health Commission of China (7th edition). The sex ratio between female and male
1051 donor is 101:159. The age of the donors ranges from 6 to 92. Of all the 284 samples, 249
1052 samples were collected from PBMC, among which 77 samples have sorted B/T cells or
1053 both. 13 samples were collected from lung tissues, including 12 BALF samples and 1
1054 PFMC sample. We also collected 22 sputum samples from patients as well. Among all the
1055 samples, we have 7 paired lung BALF and PBMC samples. Single-cell transcriptome data
1056 for each sample was profiled using 10x Genomics scRNA-seq platform. Single-cell
1057 sequencing of TCRs (13 samples) and BCRs (53 samples) or both (11 samples) was also
1058 performed for part of the samples. Detailed clinical information and demographic
1059 characteristics of patient cohorts were shown in Table S1.
1060

1061 **METHOD DETAILS**

1062 **Sample collection**

1063 Blood samples that were not immediately processed for cell encapsulation were mixed with
1064 Whole Blood Cell Stabilizer (Cytodelics) and stored at -80°C freezer. The peripheral blood
1065 mononuclear cell (PBMCs) were isolated using standard density gradient centrifugation and
1066 then used for 10x single-cell RNA-seq. Bronchoalveolar lavage fluid (BALF) samples were
1067 collected from COVID-19 patients and processed with 2h according to WHO guidance.
1068 BALF was passed through 100- μm nylon cell strainer to obtain single cell suspensions with
1069 cooled RPMI 1640 complete medium. Cells in the BALF were freshly used for 10x single-
1070 cell RNA-seq. Sputum samples were collected from COVID-19 patients using an
1071 oropharyngeal swab or hypertonic saline induction. To reduce squamous cell contamination,
1072 subjects were asked to rinse their mouth with water and clear their throat. Samples were
1073 incubated in Dulbecco's Phosphate-Buffered Saline (DPBS) with agitation for 15 minutes
1074 and filtered through 40-micron strainers. Cells in the sputum were freshly used for 10x
1075 single-cell RNA-seq.
1076

1077 **Single cell RNA library preparation and sequencing**

1078 Cell suspensions were barcoded through the 10x Chromium Single Cell platform using
1079 Chromium Single Cell 5' Library, Chromium Single Cell 3' Library, Gel Bead and Multiplex
1080 Kit, and Chip Kit (10x Genomics). The loaded cell numbers range from 300-500,000 aiming
1081 for 300-14,000 single cells per reaction. Single-cell RNA libraries were prepared using the
1082 Chromium Single Cell 3' v2 Reagent (10x Genomics; PN-120237, PN-120236 and PN-
1083 120262), Chromium Single Cell 3' v3 Reagent (10x Genomics; PN-1000075, PN-1000073
1084 and PN-120262) the Chromium Single Cell 5' v2 Reagent (10x Genomics, 120237), and
1085 Chromium Single Cell V(D)J Reagent kits (10x Genomics, PN-1000006, PN-1000014, PN-
1086 1000020, PN-1000005) was used to prepare single-cell RNA libraries according to the
1087 manufacturer's instructions. Each sequencing library was generated with a unique sample
1088 index. The libraries were sequenced using either DIPSEQ, BGISEQ or Illumina platforms.
1089

1090 **Single-cell RNA-seq data processing**

1091 Single-cell sequencing data were aligned and quantified using kallisto/bustools (KB, v0.24.4)
1092 (Bray et al., 2016) against the GRCh38 human reference genome downloaded from 10x
1093 Genomics official website. Preliminary counts were then used for downstream analysis.
1094 Quality control was applied to cells based on three metrics step by step: the total UMI

1095 counts, number of detected genes and proportion of mitochondrial gene counts per cell.
1096 Specifically, cells with less than 1000 UMI counts and 500 detected genes were filtered, as
1097 well as cells with more than 10% mitochondrial gene counts. To remove potential doublets,
1098 for PBMC samples, cells with UMI counts above 25,000 and detected genes above 5,000
1099 are filtered out. For other tissues, cells with UMI counts above 70,000 and detected genes
1100 above 7,500 are filtered out. Additionally, we applied Scrublet (Wolock et al., 2019) to
1101 identify potential doublets. The doublet score for each single cell and the threshold based
1102 on the bimodal distribution was calculated using default parameters. The expected doublet
1103 rate was set to be 0.08, and cells predicted to be doublets or with doubletScore larger than
1104 0.25 were filtered. After quality control, a total of 1,598,708 cells were remained. We
1105 normalized the UMI counts with the deconvolution strategy implemented in the R package
1106 scran (Lun et al., 2016). Specifically, cell-specific size factors were computed by
1107 computeSumFactors function and further used to scale the counts for each cell. Then the
1108 logarithmic normalized counts were used for the downstream analysis.

1109 **Batch effect correction and cell subsets annotations**

1110 To integrate cells into a shared space from different datasets for unsupervised clustering,
1111 we used the harmony algorithm (Korsunsky et al., 2019) to do batch effect correction. To
1112 detect the most variable genes used for harmony algorithm, we performed variable gene
1113 selection separately for each sample. A consensus list of 1,500 variable genes was then
1114 formed by selecting the genes with the greatest recovery rates across samples, with ties
1115 broken by random sampling. All ribosomal, mitochondrial and immunoglobulin genes were
1116 then removed from the list. Next, we calculate a PCA matrix with 20 components using
1117 such informative genes and then feed this PCA matrix into HarmonyMatrix() function
1118 implemented in R package Harmony. We set sample and dataset as two technical
1119 covariates for correction with theta set as 2.5 and 1.5, respectively. The resulting batch-
1120 corrected matrix was used to build nearest neighbor graph using scanpy (Wolf et al., 2018).
1121 Such nearest neighbor graph was then used to find clusters by Louvain algorithm (Traag et
1122 al., 2019). The cluster-specific marker genes were identified using the rank_genes_groups
1123 function.

1124
1125 The first round of clustering (resolution = 0.3) identified six major cell types including T cells,
1126 NK cells, B cells, plasma B cells, myeloid cells and epithelial cells. To identify clusters
1127 within each major cell type, we performed a second round of clustering on T/NK, B/plasma
1128 B, myeloid and epithelial cells separately. The procedure of the second round of clustering
1129 is the same as first round, starting from low-rank harmony output (30 components) on the
1130 highly variable genes chosen as described above, with resolution ranging from 0.3 to 1.5.
1131 Each sub cluster was restrained to have at least 30 significantly highly expressed genes
1132 (FDR < 0.01, logFC > 0.25, t test) compared with other cells. Annotation of the resulting
1133 clusters to cell types was based on the known markers. Meanwhile, single cells expressing
1134 two sets of well-studied canonical markers of major cell types were labeled as doublets and
1135 excluded from the following analysis. Also, cells highly expressed *HBA*, *HBB* and *HBD*,
1136 which are the markers for erythrocytes, were also excluded. 136,006 cells were removed
1137 and a total of 1,462,702 cells were retained for downstream analysis.

1138 **Detection and processing of cells with viral RNA**

1139 To identify single cells with viral infection, we aligned raw scRNA-seq reads using
1140 kallisto/bustools(KB) against a customized reference genome, in which the SARS-CoV-2
1141 genome (Refseq-ID:NC_045512) was added as an additional chromosome to the human

1142 reference genome. Single cells with viral reads (UMI > 0) were retained. Cells with less
1143 than 200 genes expressed or more than 20% mitochondrial counts were excluded, as well
1144 as those labeled as doublet following aforementioned protocol.

1145
1146 The remaining cells were then used for dimension reduction and unsupervised clustering
1147 using Python package scanpy (Wolf et al., 2018) In brief, the top 500 genes with the highest
1148 variance were selected and the dimensionality of the data was reduced by principal
1149 component analysis (PCA) (30 components) first and then with t-SNE, followed by Louvain
1150 clustering (Traag et al., 2019) performed on the 30 principal components (resolution = 1).
1151 For t-SNE visualization, we directly fit the PCA matrix into the scanpy.api.tl.tsne function
1152 with perplexity of 30. To identify cell-type-specific gene markers, we selected genes that
1153 were differentially expressed across different cell types (FDR < 0.01, log fold change > 0.5)
1154 using the rank_genes_groups function. Clusters were annotated based on the expression
1155 of known marker genes.

1156 **TCR and BCR analysis**

1157 TCR/BCR sequences were assembled and quantified following Cell Ranger (v.3.0.2) vdj
1158 protocol against GRCh38 reference genome. Assembled contigs labeled as low-confidence,
1159 non-productive or with UMIs < 2 were discarded.

1160 To identify TCR clonotype for each T cell, only cells with at least one TCR α -chain (TRA)
1161 and one TCR β -chain (TRB) were remained. For a given T cell, if there are two or more α or
1162 β chains assembled, the highest expression level (UMI or reads) α or β chains was
1163 regarded as the dominated α or β chain in the cell. Each unique dominated α - β pair (CDR3
1164 nucleotide sequences and rearranged VDJ genes included) was defined as a clonotype. T
1165 cells with exactly the same clonotype constituted a T cell clone.

1166
1167 BCR clonotypes were identified similar to TCR. Only cells with at least one heavy chain
1168 (IGH) and one light chain (IGL or IGK) were kept. For a given B cell, if there are two or
1169 more IGH or IGL/IGK assembled, the highest expression level (UMI or reads) IGH or
1170 IGL/IGK was defined as the dominated IGH or IGL/IGK in the cell. Each unique dominated
1171 IGH-IGL/IGK pair (CDR3 nucleotide sequences and rearranged VDJ genes) was defined as
1172 a clonotype. B cells with exactly the same clonotype constituted a B cell clone.

1173
1174 220,968 T cells with TCR information and 282,464 B cells with BCR information were used
1175 to perform the STARTRAC analysis as we previously described (Zhang et al.,
1176 2018). STARTRAC-expa was used to quantified the potential clonal expansion level.
1177 TCR/BCR diversity was calculated as Shannon's entropy shown below:

$$H = - \sum_x p(x) * \log_2[p(x)]$$

1178 The $p(x)$ represents the frequency of a given TCR/BCR clone among all T/B cells with
1179 TCR/BCR identified.

1180

1181 **Comparing immune cell proportion**

1182 For samples from PBMC and BALF tissue, we calculated immune cell proportions for each
1183 major cell type and underlying cell subsets. In order to avoid bias caused by samples
1184 dominated by few cell types, we filtered samples containing FACS-sorted B/T cells and
1185 retained those samples with CD45+ cells > 1000. For each sample, cell type proportion was
1186 calculated by number of cells in certain cell type divided by total number of CD45+ cells. To

1187 identify changes in cell proportions between samples in different disease severity states,
1188 disease progression stages and sex, we performed Wilcoxon rank-sum test on the
1189 proportions of each major cell type and underlying cell subset across different phenotype
1190 groups. We performed correlation analysis to assess the association between cell type
1191 proportion and patient age. Only those cell types with statistically significant differences
1192 (FDR < 0.05) in proportions were shown.

1193 **ANOVA analysis**

1194 To further assess how different patients' phenotypes and their potential interactions
1195 influence cell type proportions, we performed multivariate ANOVA on cell type proportions
1196 and on diversity of BCR/TCR based on different patient phenotypes, including disease
1197 severity, disease progression stage, sex and age. All interactions between these variables
1198 were included in the models. To convert age into a categorical variable, we binned patient
1199 age into four groups: young (<18 years old), middle-age (18-50 years old), old-age (50-70
1200 years old) and the elderly (70+ years old). Interactions between variables were regarded as
1201 significantly associated with cell type proportions when FDR < 0.05.

1202 **Differential expression and GO term enrichment analysis**

1203 To investigate the impact of virus infection on epithelial cells, we identify differential
1204 expressed genes by performing two-sided unpaired Wilcoxon tests on all the expressed
1205 genes (expressed in at least 10% cells in either group of cells). P values were adjusted
1206 following Benjamini & Hochberg protocol. Top 100 highly expressed genes of each group
1207 were shown in the volcano plots. Based on these genes, enriched GO terms were then
1208 acquired for each group of cells using R package clusterProfiler (Yu et al., 2012).

1209 **Cell-cell interaction analysis by CSOmap**

1210 To illustrate the cell-cell interaction potential of cells with viral detection, we first created a
1211 set of datasets by joining 7 BALF samples with the virus+ dataset separately. Then, we
1212 used CSOmap (Ren et al., 2020) to construct a 3D pseudo space and calculate the
1213 significant interaction for each dataset. To investigate the interaction potentials of the cell
1214 types, we used two indexes, distances within cell type and normalized connection. Distance
1215 within each cell type is calculated based on the aforementioned 3D coordinates. The
1216 shorter the distance, the closer the cells are located in the 3D space, which indicates that
1217 they are more likely to interact with each other. To further investigate the interaction
1218 between different cell types, we made use of the CSOmap output connection matrix. For a
1219 cluster pair, normalized connection was calculated by dividing its corresponding connection
1220 value by the product of their respective cell numbers. Normalized connections were then
1221 multiplied by 10,000. Meanwhile, to highlight the key ligand-receptor pairs function in the
1222 interaction, we also examine the contribution output by CSOmap.
1223 In addition, normalized connections were also calculated on another set of cohorts where
1224 we combined virus+ dataset with samples with paired PBMC and BALF tissues, in order to
1225 investigate the interaction potential between cells from two tissues, PBMC and BALF.

1226 **Inflammatory and cytokine score related subtypes analysis.**

1227 Briefly, we firstly filtered out samples with less than 1000 cells. For PBMC, only subtypes
1228 with more than 1000 cells were included in the subsequent analysis. For BALF data
1229 analysis, we removed major cell types with less than 500 cells. To define inflammatory and
1230 cytokine score, we downloaded a gene set termed
1231 'HALLMARK_INFLAMMATORY_RESPONSE' from MSigDB (PMID: 26771021) and

1232 collected cytokine genes based on these references (see Table S1). Cytokine and
1233 inflammatory score were evaluated with the AddModuleScore function built in the Seurat
1234 (PMID: 31178118). To select the most promising hyper-inflammatory cell types, we
1235 performed Mann-Whitney rank test (single-tail) for each subtype's score versus all the other
1236 subtypes' score. Seven subtypes (Mono_c1-CD14-CCL3, Mono_c2-CD14-HLA-DPB1,
1237 Mono_c3-CD14-VCAN, T_CD4_c08-GZMK-FOS^{high}, T_CD8_c06-TNF, T_CD8_c09-
1238 SLC4A10 and Mega) in PBMC were defined as hyper-inflammatory cell types with
1239 significantly statistical parameters ($P < 0.0001$) in both cytokine and inflammatory score. In
1240 addition, we defined 8 subtypes (T_CD8_c08-IL2RB, T_CD4_c11-GNLY, NK_c01-FCGR3A,
1241 T_CD8_c05-ZNF683, T_CD8_c04-COTL1, T_CD8_c07-TYROBP, T_CD8_c03-GZMK and
1242 T_gdT_c14-TRDV2) with significantly statistical parameters ($P < 0.0001$) only in cytokine
1243 score. For subtypes in BALF, we defined 5 subtypes (Macro_c2-CCL3L1, Mono_c1-CD14-
1244 CCL3, Mono_c2-CD14-HLA-DPB1, Mono_c3-CD14-VCAN, Neu) as hyper-inflammatory
1245 types with the same standard threshold as PBMC ($P < 0.0001$).

1246 **Cell ratio and cytokine marker analysis of hyper-inflammatory and** 1247 **cytokine subtypes (integrated with Statistics section)**

1248 To explore whether there are state-specific of COVID-19 patients enriched subtypes, we
1249 performed hierarchical clustering with setting standard scale (0-1) for 7 hyper-inflammatory
1250 and 8 cytokine subtypes respectively. Then, we used the Wilcoxon rank-sum test to
1251 calculate the significance of cell proportion of each subtype in states (moderate
1252 convalescent, moderate progression, severe convalescent, severe progression) compared
1253 with healthy control. We also applied the ordinary least square method to calculate the
1254 correlation between age and cell proportion in different states of COVID-19 patients. For the
1255 significance of cytokine expression level with state and age, we performed Wilcoxon rank-
1256 sum test and ordinary least square to assess the P values.

1257 **Cell-cell communication analysis between PBMC and BALF by iTALK**

1258 To identify and visualize the possible cell-cell interactions in terms of cytokine storm
1259 between the highly inflammation-correlated cell types evaluated by the inflammation score
1260 within each tissue and the crosstalk between lung and circulating blood, we employed an R
1261 package iTALK introduced by Wang et al. (Wang et al., 2019, bioRxiv,
1262 <https://www.biorxiv.org/content/10.1101/507871v1>). Cytokine/chemokine category ($n = 320$)
1263 in the ligand-receptor database was selectively used for our purpose. Wilcoxon rank sum
1264 test was used to identify the differentially expressed genes (DEGs) between severe onset
1265 and moderate onset patient groups for each cell type. DEGs were then matched and paired
1266 against the ligand-receptor database to construct a putative cell-cell communication
1267 network. An interaction score defined as the product of the log fold change of ligand and
1268 receptor was used to rank these interactions. In addition, the expression level of both ligand
1269 and receptor were also considered. We defined severe gained interaction if a ligand gene
1270 was upregulated in severe onset group and its paired gene upregulated or remains no
1271 change. We defined severe lost interaction if a ligand(receptor) gene was downregulated in
1272 severe onset group regardless of the expression level of its paired gene.

1273 **Cytokine analysis of serum by using multiplex bead-based immunoassay**

1274 Human cytokines in the serum were measured by Bio-plex Pro TM Human Cytokine
1275 Screening 48 plex Bio-PlexTM 200 System (# 12007283, Bio-Rad, US) and Bio-PlexTM
1276 200 System (Bio-Rad). Bio-Plex ProTM assays are essentially immunoassays formatted on
1277 magnetic beads and are built upon three core elements of xMAP technology, fluorescently

1278 dyed microspheres (also called beads), a dedicated flow cytometer with two lasers to
1279 measure the different molecules bound to the surface of the beads, and a high-speed digital
1280 signal processor that efficiently manages the fluorescence data.

1281

1282 Sample preparation

1283 Whole blood from COVID-19 patients and healthy controls were drawn into collection tubes
1284 containing anticoagulant. Centrifugation the tubes at 1,000 x g for 15 min at 4°C and
1285 transfer the serum to a clean polypropylene tube, followed by another centrifugation at
1286 10,000 x g for 10 min at 4°C to completely remove platelets and precipitates. Dilute
1287 samples fourfold (1:4) by adding 1 volume of sample to 3 volumes of Bio-Plex sample
1288 diluent. Fifty microliter of each sample were used to assay.

1289

1290 Preparation of coupled beads

1291 Coupled beads were diluted to a 1x concentration according to the instruction. Required
1292 volume of Bio-Plex assay buffer was added to a 15 ml polypropylene tube. Vortex the stock
1293 coupled beads at medium speed for 30 sec. Carefully open the cap and pipet any liquid
1294 trapped in the cap back into the tube. Dilute coupled beads to 1x by pipetting the required
1295 volume into the 15 ml tube and vortex. Protect the beads from light with aluminum foil and
1296 equilibrate to room temperature prior to use.

1297

1298 Assay running

1299 Add coupled beads, standards and samples to each well of the assay plate. Cover plate
1300 with a new sheet of sealing tape and protect from light with aluminum foil. Incubate on
1301 shaker at 850 ± 50 rpm at room temperature (RT). While the samples were incubating,
1302 calculate and prepare the volume of detection antibodies and detection antibody diluent
1303 needed. After washing the plate three times with 100 μ l wash buffer, transfer 25 μ l detection
1304 antibodies to each well using a multichannel pipet. Cover plate with a new sheet of sealing
1305 tape and protect from light with aluminum foil. Incubate on shaker at 850 ± 50 rpm for 30
1306 min at room temperature.

1307

1308 Read plate and calculation

1309 Bio-Plex Manager™ software was used for data acquisition and analysis.

1310

1311 DATA AND CODE AVAILABILITY

1312 The raw sequencing and processed gene expression data in this paper have been
1313 deposited into GSA (Genome Sequence Archive in BIG Data Center, Beijing Institute of
1314 Genomics, Chinese Academy of Sciences) and the NCBI GEO database, respectively.
1315 Visualization of this dataset can be found at <http://covid19.cancer-pku.cn>.

1316

1317 Figure legends

1318 Figure 1. Single-cell atlas of multiple tissue types from healthy individuals and 1319 COVID-19 patients

1320 (A) A flowchart depicting the overall design of the study.

1321 (B) t-Stochastic Neighborhood Embedding (t-SNE) representations of integrated single-cell
1322 transcriptomes of 1,462,702 cells derived from our healthy controls and COVID-19 patients.

1323 Cells are colour-coded by 64 cell subsets from 6 major cell types.

1324 (C) Violin plots of selected marker genes (rows) for major cell subpopulations (columns)
1325 ordered by cell lineage relationships. NK, natural killer cells; Mono, monocyte; Macro,
1326 macrophage; DC, dendritic cells; Neu, neutrophil; Mega, megakaryocyte; Epi, epithelial
1327 cells.
1328 (D-E) t-SNE representations of integrated single-cell transcriptomes of 1,462,702 cells
1329 coloured by disease symptoms (D) and disease progression stages (E).
1330 (F) Tissue preference of each cluster estimated by Ro/e. Ro/e denotes the ratio of observed
1331 to expected cell number.
1332 See also Figure S1.
1333

1334 **Figure 2. Dynamic changes of B cell composition across disease conditions**

1335 (A) Differences in immune cell composition across disease conditions for PBMC. Conditions
1336 are shown in different colors. Each boxplot represents one cell cluster. All differences with
1337 adjusted P -value < 0.05 are indicated; two-sided unpaired Wilcoxon was used for analysis.
1338 (B) Changes of XBP1+ plasma cells proportion across disease conditions. Composition of
1339 XBP1+ plasma cell BCRH cgene. All differences with P -value < 0.05 are indicated; two-
1340 sided unpaired Wilcoxon was used for analysis.
1341 (C) Differences of XBP1+ plasma cells clonal expansion and BCR diversity across disease
1342 conditions. BCR clonal expansion level is calculated by STARTRAC-expa. Shannon's
1343 entropy reveals the diversity of BCR repertoire. All differences with P -value < 0.05 are
1344 indicated; two-sided unpaired Wilcoxon was used for analysis.
1345 (D) Transition between XBP1+ plasma cells and other B cell sub clusters (left). Clonotypes
1346 of clones contain XBP1+ plasma cells (right); only shows clones with more than 5 cells.
1347 (E) Composition of B_c03-CD27-AIM2 memory cells BCRH cgene.
1348 (F) ANOVA of XBP1+ plasma cells proportion.
1349 (G) ANOVA of B_c03-CD27-AIM2 memory cells proportion (left) and differences of B_c03-
1350 CD27-AIM2 memory cells proportion between male and female (right). Two-sided unpaired
1351 Wilcoxon test.
1352 See also Figure S2.
1353

1354 **Figure 3. Differences in T cell composition across disease conditions**

1355 (A) Changes of proliferating CD4 and CD8 T cells across disease conditions for PBMC.
1356 Conditions are shown in different colors. All differences with adjusted P -value < 0.05 are
1357 indicated; two-sided unpaired Wilcoxon was used for analysis.
1358 (B) Differences of three CD8 proliferating T cell sub clusters proportion across disease
1359 conditions. All differences with P -value < 0.05 are indicated; two-sided unpaired Wilcoxon
1360 was used for analysis.
1361 (C) Differences of two CD4 proliferating T cell sub clusters proportion across disease
1362 conditions. All differences with P -value < 0.05 are indicated; two-sided unpaired Wilcoxon
1363 was used for analysis.
1364 (D) Differences of T_CD4_c13-MKI67-CCL5^{low} proliferating cells clonal expansion and TCR
1365 diversity across disease conditions. TCR clonal expansion level is calculated by
1366 STARTRAC-expa. Shannon's entropy reveals the diversity of BCR repertoire. All
1367 differences with P -value < 0.05 are indicated; two-sided unpaired Wilcoxon was used for
1368 analysis.
1369 (E) Transition between T_CD4_c13-MKI67-CCL5^{low} proliferating cells and other CD4 cell
1370 sub clusters (left) and clonotypes of clones contain T_CD4_c13-MKI67-CCL5^{low}
1371 proliferating cells (right); only shows clones with more than 5 cells.

1372 (F) Differences of T_c04_CD4-ANXA2 T cell proportion across disease conditions. All
1373 differences with P -value < 0.05 are indicated; two-sided unpaired Wilcoxon was used for
1374 analysis.
1375 (G) ANOVA of T_CD4_c13-MKI67-CCL5^{low} proliferating cells proportion.
1376 (H) ANOVA of T_c04_CD4-ANXA2 T cell proportion (left) and differences of T_c04_CD4-
1377 ANXA2 T cell proportion between male and female (right). Two-sided unpaired Wilcoxon
1378 test.
1379 See also Figure S2.
1380

1381 **Figure 4 Landscape Of Cell Types Detected SARS-Cov-2 Sequences and Their** 1382 **Antiviral Response.**

1383 (A) Uniform Manifold Approximation and Projection (UMAP) of all cells with SARS-CoV-2
1384 genome UMI > 0 after quality control containing 3085 cells in total.
1385 (B) Characteristic markers we chose to identify each cell type. The purple box indicates
1386 immune cell types (top), and the red one indicates epithelial cell types (bottom).
1387 (C) UMAP showing expression level of known SARS-CoV-2 infected receptor ACE2 (left)
1388 and TMPRSS2 (right). Each dot denotes a single cell and colored by its expression level of
1389 the gene.
1390 (D) UMAP showing the viral load of each cell. The darker colors in the bar indicate a higher
1391 viral load in cells.
1392 (E) UMAP showing the activation of Interferon-stimulated genes (ISGs) in cells with viral
1393 detection.
1394 (F) Violin plots showing differential expression of ISGs between cells with viral detection
1395 (virus+) and cells without (virus-) in PBMC-derived neutrophils (left panel), BALF-derived
1396 neutrophils (middle panel) and squamous cells (right panel). The y axis represents the
1397 expression level of each gene. logCP10K, log-transformed counts per 10,000. Two-sided
1398 unpaired Wilcoxon test.
1399 (G) Scatter plots showing the correlation between viral load and ISGs in neutrophil (left
1400 panel) and squamous cells (right panel). The line in scatter plots represent the result of
1401 linear regression. Each point in the graph represents one single cell, colored by cell types.
1402 The x axis shows virus load in each cell while the y axis represents the expression level of
1403 one of the ISG genes. Correlation coefficient (R) and probability (p) are acquired using
1404 Pearson's correlation.

1405 See also Figures S3 and Tables S4.

1406 **Figure 5. The Impact of Virus Infection on Expression and Cell-cell interaction** 1407 **of Epithelial Subtypes**

1408 (A) Volcano plot showing differentially expressed genes between squamous cells with or
1409 without viral detection. Adjusted P -value < 0.05 , Two-sided unpaired Wilcoxon test. *ANXA1*
1410 is denoted in dark blue.
1411 (B) Enriched GO terms of genes highly expressed in virus+ squamous cells shown in (A).
1412 (C) Venn plot showing the intersection of genes upregulated in epithelial cells with viral
1413 detection. Each compartment is colored by the number of genes.
1414 (D) Violin plot showing the expression of *ANXA1* in squamous cells with or without viral
1415 detection. Two-sided unpaired Wilcoxon test.
1416 (E) Boxplot showing the pseudo space distance within ciliated cells. Each dot represents an
1417 individual patient. Two-sided paired Wilcoxon test.

1418 (F) Boxplot showing the pseudo space distance within squamous cells. Each dot represents
1419 an individual patient. Two-sided paired Wilcoxon test.
1420 (G) Violin plot showing the pseudo space distance within each type of epithelial cells in one
1421 example. Two-sided unpaired Wilcoxon test.
1422 (H) Boxplot showing the median of pseudo space distance within each type of epithelial
1423 cells of all the patients with BALF data. Each dot represents an individual patient. Two-
1424 sided unpaired Wilcoxon test.
1425 (L) Boxplot showing the normalized connection between squamous cells and virus-detected
1426 plasma B cells of all the patients with BALF data. Each dot represents an individual patient.
1427 Two-sided unpaired Wilcoxon test.
1428 (I) Pie chart showing the ligand-receptor contribution proportion between virus+ squamous
1429 and Macro_c6-VCAN in one example. Ligand-receptor pairs with contribution less than 0.05
1430 were merged as 'Other LR's'.
1431 (J) Boxplot showing the normalized connection between squamous cells and virus-
1432 macrophage (left), virus+ macrophage (middle) and virus+ neutrophils (left). Each dot
1433 represents an individual patient. Two-sided unpaired Wilcoxon test.
1434 See also Figures S4 and S5.
1435

1436 **Figure 6. Mono_c1-CD14-CCL3 and megakaryocytes in peripheral blood**
1437 **appear as dominant source for inflammatory cytokine storm.**

1438 (A) t-SNE plots of PBMC cells colored by major cell types (top left panel), inflammatory cell
1439 types (top right panel), cytokine score (middle panel) and inflammatory score (bottom
1440 panel).
1441 (B) Violin plots of selected cytokine genes for seven hyper-inflammatory cell subtypes.
1442 (C) Heatmap of an unsupervised clustering of cell proportion of seven hyper-inflammatory
1443 cell subtypes in all samples analyzed.
1444 (D) Box plots of the cell proportion of Mono_c1-CD14-CCL3, Mega and T_CD4_c08-GZMK-
1445 FOS^{high} clusters from healthy controls (n=20), moderate convalescent (n=48), moderate
1446 onset (n=18), severe convalescent (n=35) and severe onset (n=38) patients. Two-sided
1447 Wilcoxon rank-sum test.
1448 (E) Ordinary least squares model of age to cell proportion of Mono_c1-CD14-CCL3,
1449 Mono_c2-CD14-HLA-DPB1 and Mono_c3-CD14-VCAN clusters from healthy controls
1450 (n=20), convalescent (n=83) and onset (n=56) patients. P value was assessed with F-
1451 statistic for ordinary least squares model.
1452 (F) Heatmap of cytokines genes' expression among seven hyper-inflammatory cell
1453 subtypes. Seven hyper-inflammatory cell subtypes are colored in red and others are
1454 colored in grey.
1455 (G) Box plots of the cytokines' expression of Mono_c1-CD14-CCL3, Mega and
1456 T_CD8_c06-TNF clusters from healthy controls (n=20), moderate convalescent (n=48),
1457 moderate onset (n=18), severe convalescent (n=35) and severe onset (n=38) patients.
1458 Two-sided Wilcoxon rank-sum test.
1459 (H) Ordinary least squares model of age to cytokines' expression of Mono_c1-CD14-CCL3,
1460 Mega and T_CD8_c06-TNF clusters from healthy controls (n=20), convalescent (n=48+35)
1461 and onset (n=18+38) patients. P value was assessed with F-statistic for ordinary least
1462 squares model.
1463 In (D) and (G), the box represents the second, third quartiles and median, whiskers each
1464 extend 1.5 times the interquartile range; dots represent outliers. In panel (B), (C) and (F),
1465 Mono_c1, Mono_c2, Mono_c3, T_CD4_c08, T_CD8_c09, T_CD8_c06 and Mega
1466 correspond to Mono_c1-CD14-CCL3, Mono_c2-CD14-HLA-DPB1, Mono_c3-CD14-VCAN,

1467 T_CD4_c08-GZMK-FOS^{high}, T_CD8_c09-SLC4A10, T_CD8_c06-TNF and Mega,
1468 respectively. In panel (F), T_CD4_c11, T_CD8_c03, T_CD8_c04, T_CD8_c05, T_CD8_c07,
1469 T_gdT_c14, T_CD8_c08, NK_c01 correspond to clusters of T_CD4_c11-GNLY,
1470 T_CD8_c03-GZMK, T_CD8_c04-COTL1, T_CD8_c05-ZNF683, T_CD8_c07-TYROBP,
1471 T_gdT_c14-TRDV2, T_CD8_c08-IL2RB and NK_c01-FCGR3A, respectively. DC, dendritic
1472 cells. Mega, megakaryocytes. Mono, monocytes.
1473

1474 **Figure 7. The interactions of hyper-inflammatory cell subtypes in lung and** 1475 **peripheral blood.**

1476 (A) t-SNE plots of BALF cells colored by major cell types (top panel), cytokine score (middle
1477 panel) and inflammatory score (bottom panel).

1478 (B) Boxplots of the inflammatory score (top panel) and cytokine score (bottom panel) of cell
1479 subtypes. Significance was evaluated with Wilcoxon rank-sum test. **** P < 0.0001.

1480 (C) Heatmap of an unsupervised clustering of cytokine genes' expression among five
1481 hyper-inflammatory cell subtypes.

1482 (D) Circos plot showing the prioritized interactions mediated by ligand-receptor pairs
1483 between inflammation-related cell types from BALF and PBMC, respectively. The outer ring
1484 displays color coded cell types and the inner ring represents the involved ligand-receptor
1485 interacting pairs. The line width and arrow width are proportional to the log fold change
1486 between severe onset and moderate onset patient groups in ligand and receptor,
1487 respectively. Colors and types of lines are used to indicate different types of interactions as
1488 shown in the legend. The bar plot at bottom indicates the interaction score for each
1489 interaction which serves to measure the interaction strength.

1490 (E) Summary illustration depicting the potential cytokine/receptor interactions of hyper-
1491 inflammatory cell subtypes involved in the cytokine storm.

1492 DC, dendritic cells. Epi, epithelial cells. Macro, macrophage cells. Mono, monocytes. Neu,
1493 neutrophils.

1494

1495

1496 **Supplementary Figures**

1497 **Figure S1. Basic information of the dataset quality and cell subsets in major** 1498 **cell lineages, Related to Figure 1**

1499 (A) Sorted age span of donors color-coded by disease symptoms.

1500 (B) Distribution of sex in donors with different disease symptoms. Chi-square test.

1501 (C-E) Distribution of unique molecular identifier (UMI) count per cell (C), gene count per cell
1502 (D), and percentage of mitochondrial transcripts per cell (E) detected for cells in various
1503 tissue types. PBMC, peripheral blood mononuclear cells; BALF, bronchoalveolar lavage
1504 fluid; PFMC/Sputum, pleural effusion/sputum.

1505 (F-J) Violin plots of selected marker genes (rows) for cell subsets (columns) within each cell
1506 lineage, including 6 B/plasma B cell clusters (F), 23 Myeloid cell clusters (G), 3 NK cell
1507 clusters (H), 4 Epithelial cell clusters (I) and 28 T cell clusters (J).

1508

1509 **Figure S2. ANOVA of cell composition and clonal expansion, Related to**
1510 **Figures 2 and 3**

1511 (A) Differences of B_c06-MKI67 cells proportion across disease conditions. All differences
1512 with P -value < 0.05 are indicated; two-sided unpaired Wilcoxon test.

1513 (B) ANOVA of B cells proportion

1514 (C) Differences of B_c03-CD27-AIM2 cells proportion across disease conditions. All
1515 differences with P -value < 0.05 are indicated; two-sided unpaired Wilcoxon test.

1516 (D) ANOVA of CD4 T cells proportion

1517 (E) ANOVA of CD8 T cells proportion

1518 (F) Differences of T_c14_gdT-TRDV2 cells proportion across disease conditions. All
1519 differences with P -value < 0.05 are indicated; two-sided unpaired Wilcoxon test.

1520 (G) Differences of T_c09_CD8-SLC4A10 cells proportion across disease conditions. All
1521 differences with P -value < 0.05 are indicated; two-sided unpaired Wilcoxon test.

1522 (H) V gene usage of SARS-Cov-2 neutralized antibodies.

1523 (I-K) ANOVA of the diversity of TCR/BCR repertoire, estimated by Shannon's entropy.

1524 **Figure S3. Differential Expression Analysis for ISG Genes between other Virus+**
1525 **and Virus- cells in PBMC or BALF, Related to Figures 4.**

1526 (A) The ISG genes in virus+ differentially expressed comparison against virus- in residential
1527 cell types including immune cells and epithelial cells. Two sided unpaired Wilcoxon test.

1528 (B) Violin plots showing the expression of ISGs in BALF. Two sided unpaired Wilcoxon test.

1529 (C) Scatter plots showing the correlation between viral load and expression level of viral
1530 load. Pearson's correlation.

1531 **Figure S4. Detailed Investigation of interacting potentials of epithelial cells**
1532 **with viral detection, Related to Figure 5**

1533 (A) 2D pseudo space calculated by CSOmap, showing the location of ciliated cells. Each
1534 dot denotes a single cell, colored by cell type.

1535 (B) Violin plot showing the distance calculated from space shown in (A) within each ciliated
1536 cell group. Two-sided unpaired Wilcoxon test.

1537 (C) Violin plot showing the distance within each squamous cell group. Two-sided unpaired
1538 Wilcoxon test.

1539 (D) Bar plot showing the mean of normalized connections of the interaction between virus+
1540 secretory and Macro_c1-C1QC in patients categorized by two states. Error bar, s.e.m.

1541 (E) Pie chart showing the ligand-receptor contribution proportion between virus+ secretory
1542 and Macro_c6-VCAN in one example. Ligand-receptor pairs with contribution less than 0.05
1543 were merged as 'Other LRs'.

1544 (F) Dotplot showing the mean expression level of MARCO in BALF samples. Pct,
1545 percentage of expressed cells.

1546 (G) Boxplot of normalized connection between major cell types and ciliated (top), secretory
1547 (middle) and squamous (bottom) cells with viral detection. Kruskal-Wallis Rank Sum Test.

1548 (H) Dot plots showing the expression of S100A9(left) and S100A8 (right) in PBMC samples.
1549 Each dot is colored by the means of the expression and sized by the scaled means (Z
1550 scores). Blue boxes highlight expressions in severe onset patients.

1551 (I) Boxplot of normalized connection between PBMC-derived cell types and BALF. Each dot
1552 represents a sample. Kruskal-Wallis Rank Sum Test.

1553

1554 **Figure S5. The Expression of Genes in PBMC and BALF samples, Related to**
1555 **Figure 5**

1556 (A) Dot plots showing the expression of *ANXA1* (top), *FPR1* (middle) and *TLR4* (bottom) in
1557 PBMC samples.

1558 (B) Dot plots showing the expression of *ANXA1* (first panel), *FPR1* (second panel) *S100A9*
1559 (third panel), *S100A8* (fourth panel) and *TLR4* (bottom panel) in BALF samples.

1560 Each dot is colored by the means of the expression and sized by the scaled means (*Z*
1561 scores).

1562

1563 **Figure S6. Identification of hyper-inflammatory subtypes associated with**
1564 **cytokine storm in PBMCs**

1565 (A) *t*-SNE plots of PBMC cells colored by cytokine score (top panel) and inflammatory score
1566 (bottom panel).

1567 (B) The proportion of subtypes from healthy controls (color, *n*=20), severe onset (color,
1568 *n*=38) and average of all samples (*n*=159) (top panel); the inflammatory score (middle panel)
1569 and cytokine score (bottom panel) of subtypes from healthy controls (*n*=20), moderate
1570 convalescent (*n*=48), moderate onset (*n*=18), severe convalescent (*n*=35) and severe onset
1571 (*n*=38) patients. Significance was evaluated with Mann-Whitney rank test for each subtype
1572 versus all the other subtypes. **** *P* < 0.0001.

1573 (C) Barplots of subtypes' (seven hyper-inflammatory cell types, eight cytokine cell types and
1574 others) frequencies across each individual samples from healthy controls (*n*=20), moderate
1575 convalescent (*n*=48), moderate onset (*n*=18), severe convalescent (*n*=35) and severe onset
1576 (*n*=38) patients.

1577 (D) Bar graphs showing cytokine concentration at the serum levels of CCL3, IFNG, IL1RN
1578 and TNF from healthy controls (*n*=5), convalescent (*n*=7), non-severe (*n*=4), severe (*n*=4),
1579 death case (*n*=7) patients. Shown are *P* values by student *t*-test.

1580 (E) The differential expression distribution of cytokines for severe onset (*n*=38), moderate
1581 onset (*n*=18) and convalescent (*n*=48+35) versus healthy control (*n*=20). The triangle
1582 represents severe onset versus healthy controls. Circle stands for moderate onset versus
1583 healthy controls. The square stands for convalescent versus healthy controls. All rings in
1584 the plot from the inside to the outside represent the range of *P* value, which are *P* > 0.05,
1585 0.01 < *P* ≤ 0.05, 0.001 < *P* ≤ 0.01, 0.0001 < *P* ≤ 0.001 and *P* < 0.0001 respectively. Red
1586 indicates positive and blue indicates negative. Size for the triangle, circle and square
1587 means log₂ (fold change). Two-sided unpaired *t* test. HC, healthy control.

1588 (F) Bar graphs showing cytokine concentration at the serum levels of CCL5 and IL16 from
1589 healthy controls (*n*=5), convalescent (*n*=7), non-severe (*n*=4), severe (*n*=4), death case
1590 (*n*=7) patients. Kruskal-Wallis H-test between non-severe, severe and death case.

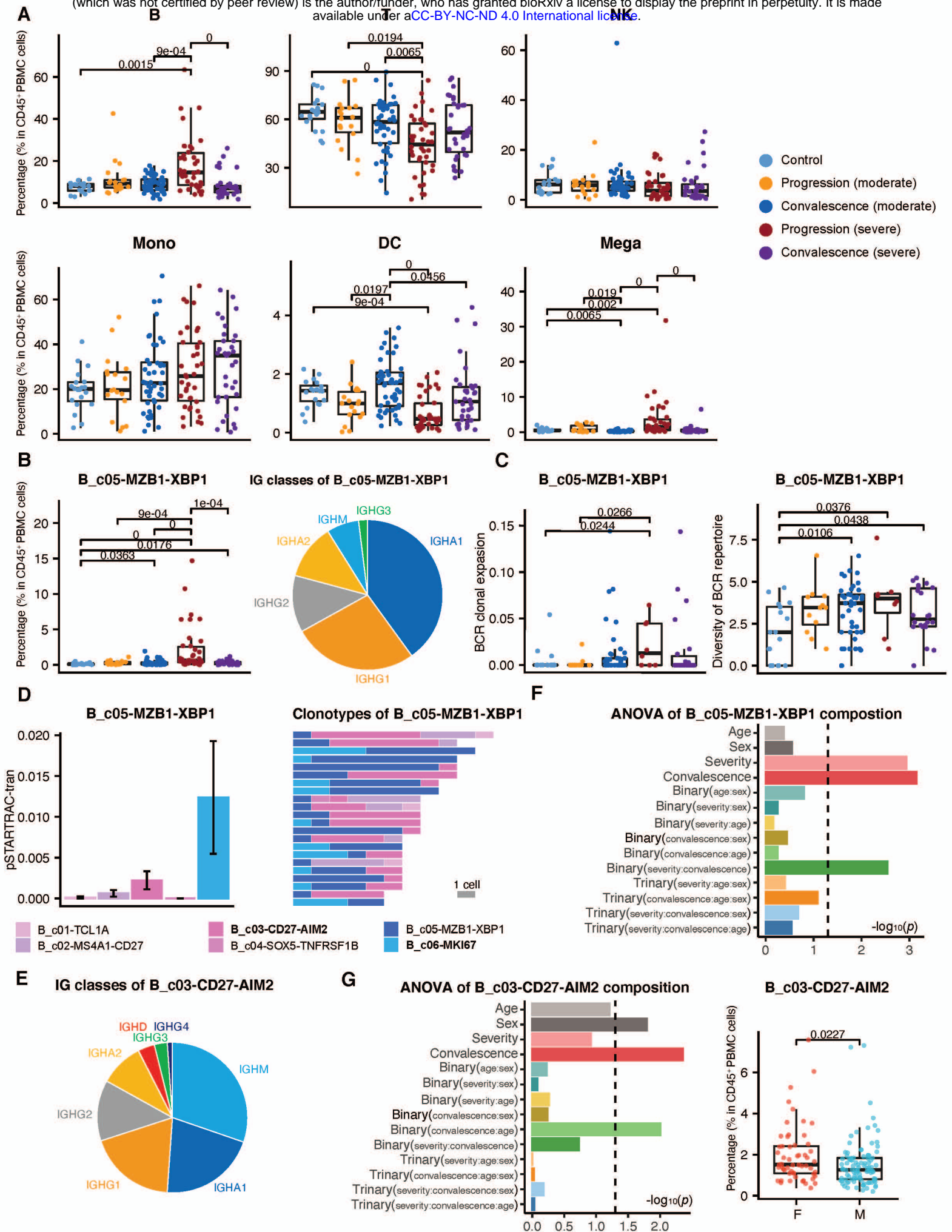
1591 In panel (D) and panel (F), all points are shown and bars represent mean with the 95%
1592 confidence intervals. DC, dendritic cells. Mega, megakaryocytes. Mono, monocytes.

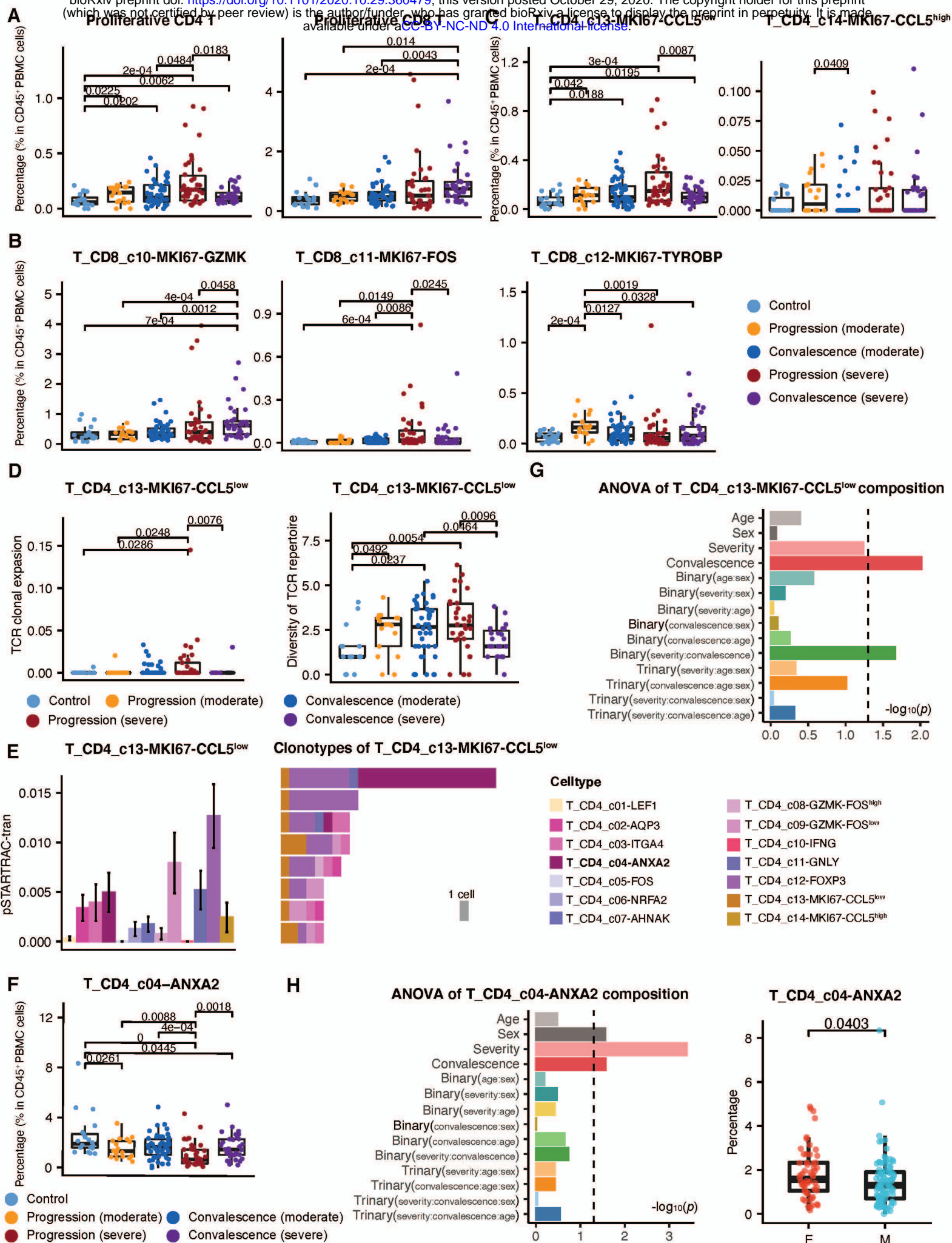
1593

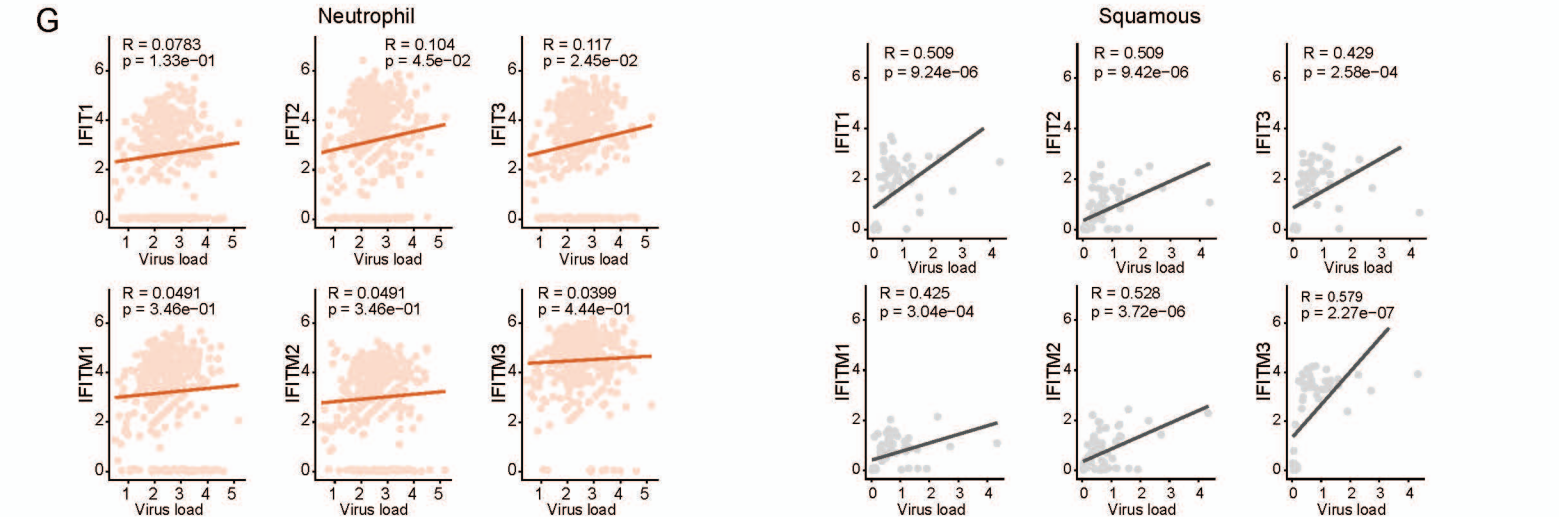
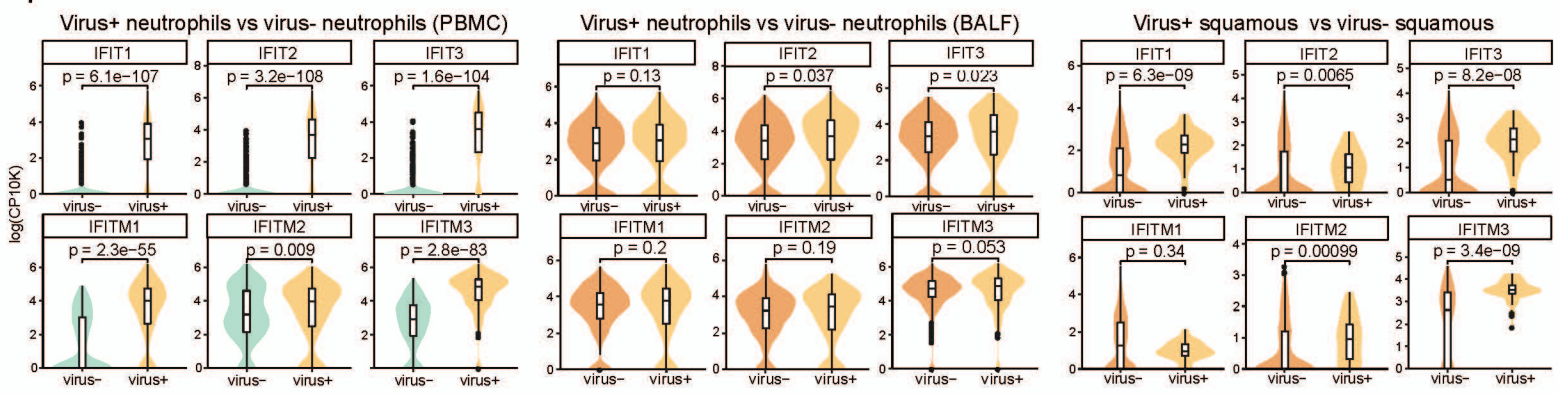
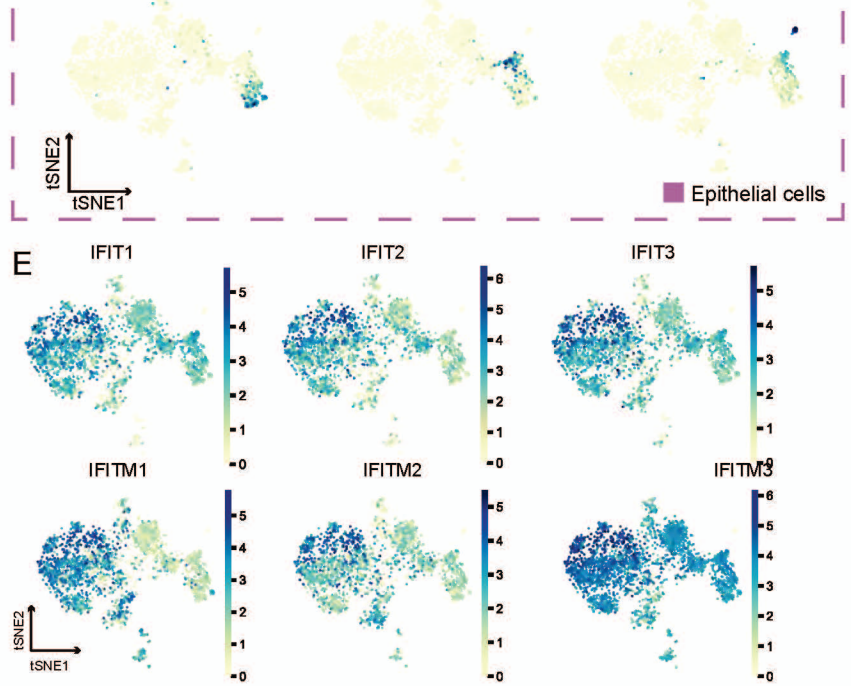
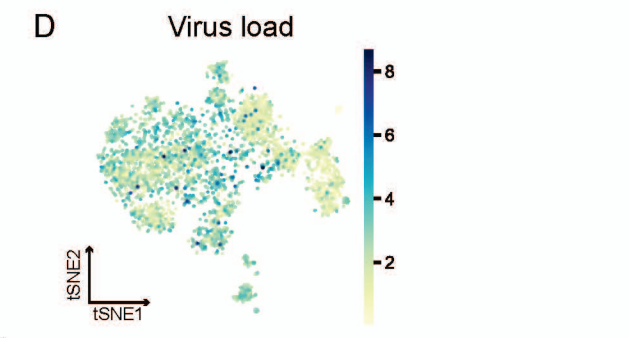
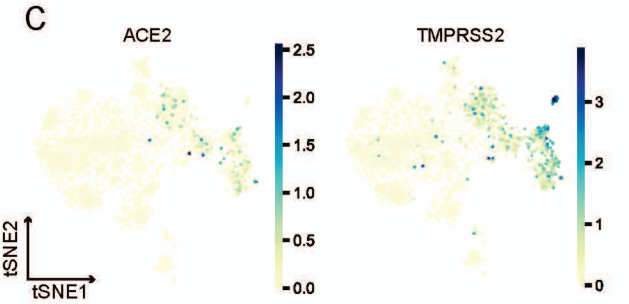
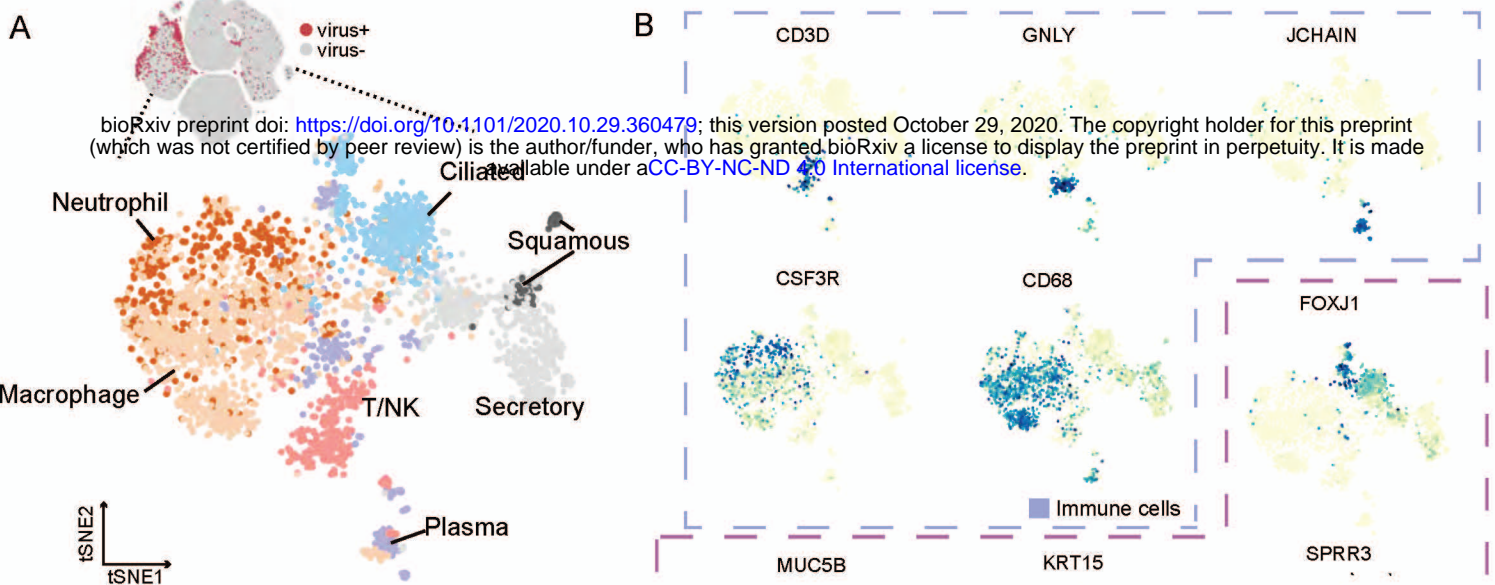
1594 **Figure S7. Intercellular interaction alterations among cell types between**
1595 **severe and moderate onset sample groups.**

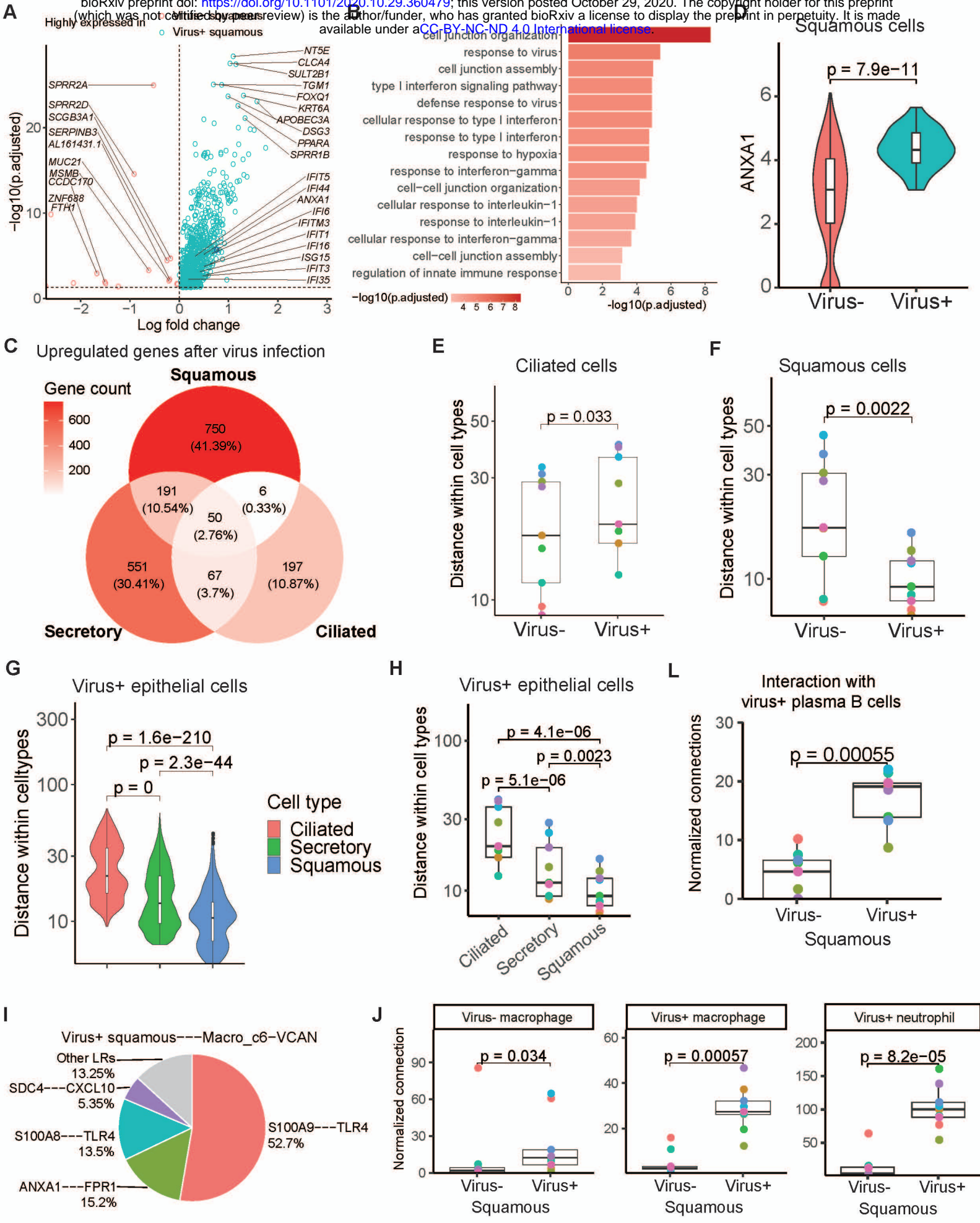
1596 (A). Circos plot showing the prioritized interactions mediated by ligand-receptor pairs
1597 between inflammation-related cell subtypes for each tissue, namely, PBMC (left panel) and
1598 BALF (right panel). The outer ring displays color coded cell types and the inner ring
1599 represents the involved ligand-receptor interacting pairs. The line width and arrow width are
1600 proportional to the log fold change between severe onset and moderate onset patient

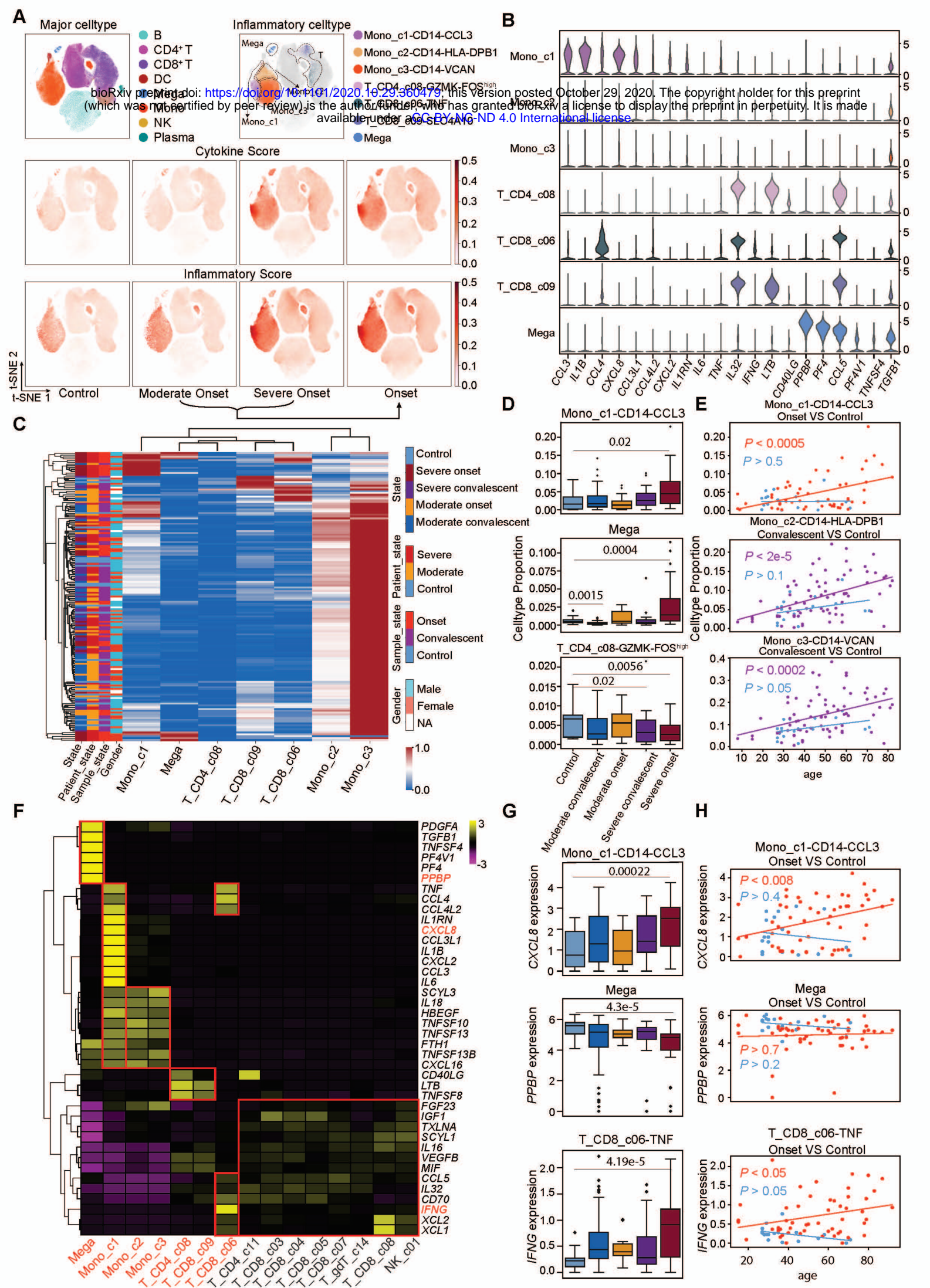
1601 groups in ligand and receptor, respectively. Colors and types of lines are used to indicate
1602 different types of interactions as shown in the legend. The barplot at bottom indicates the
1603 interaction score for each ligand-receptor interaction which serves to measure the
1604 interaction strength.
1605 DC, dendritic cells. Epi, epithelial cells. Macro, macrophage cells. Mono, monocytes. Neu,
1606 neutrophils. Mega, megakaryocytes.
1607
1608

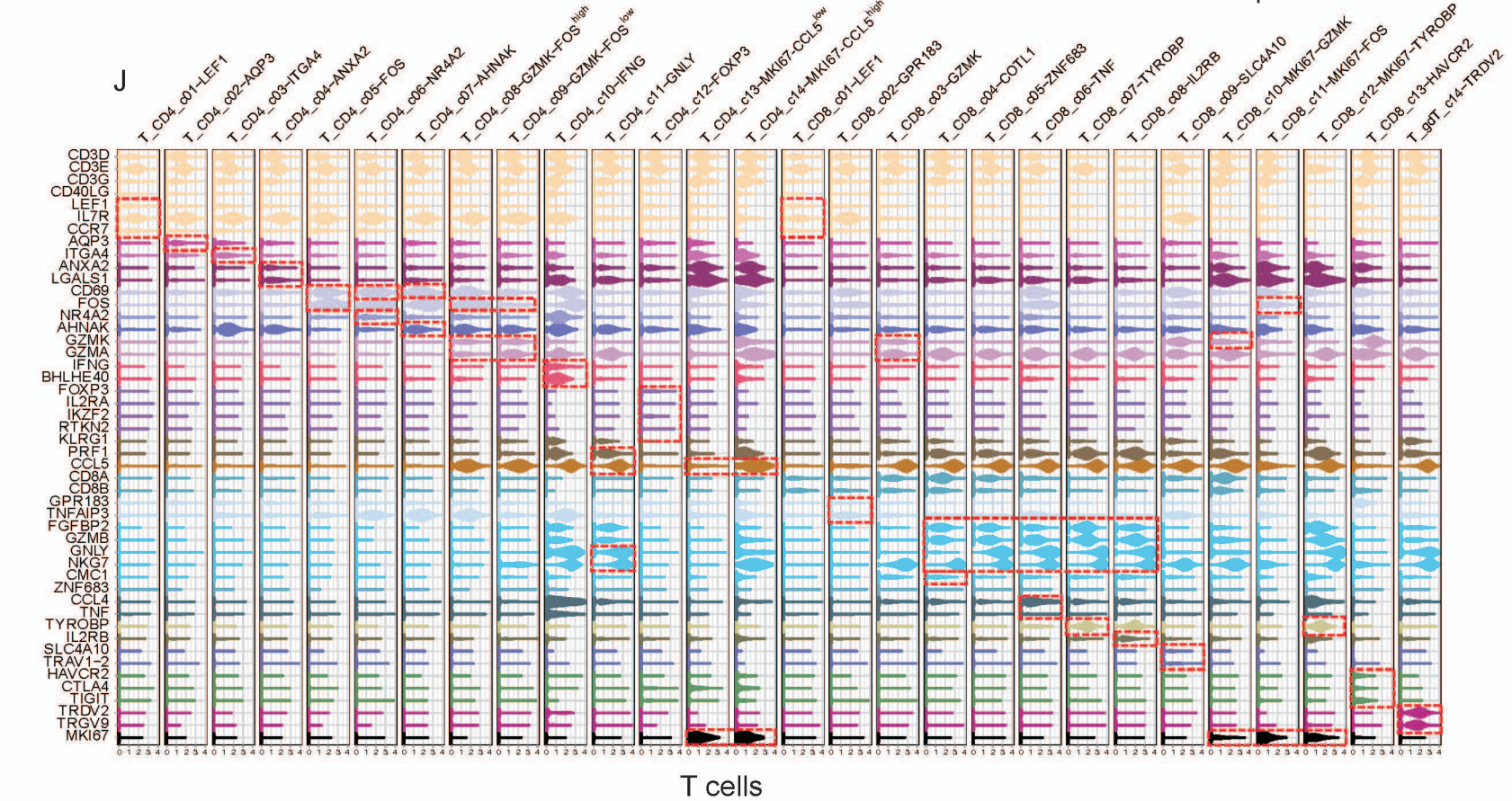
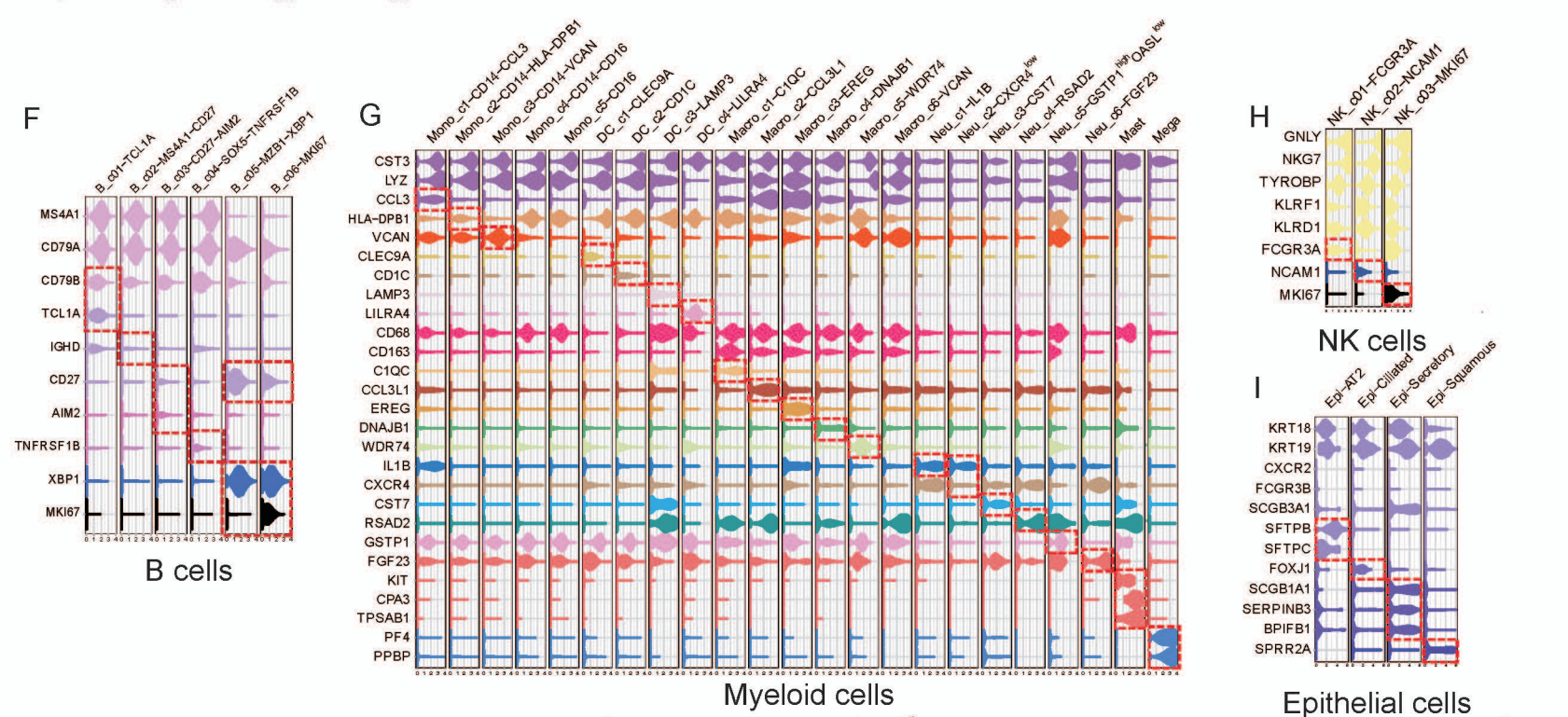
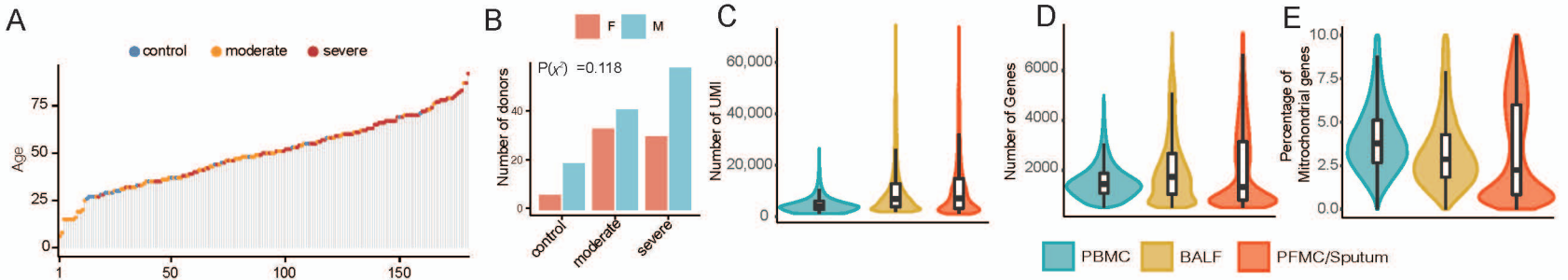


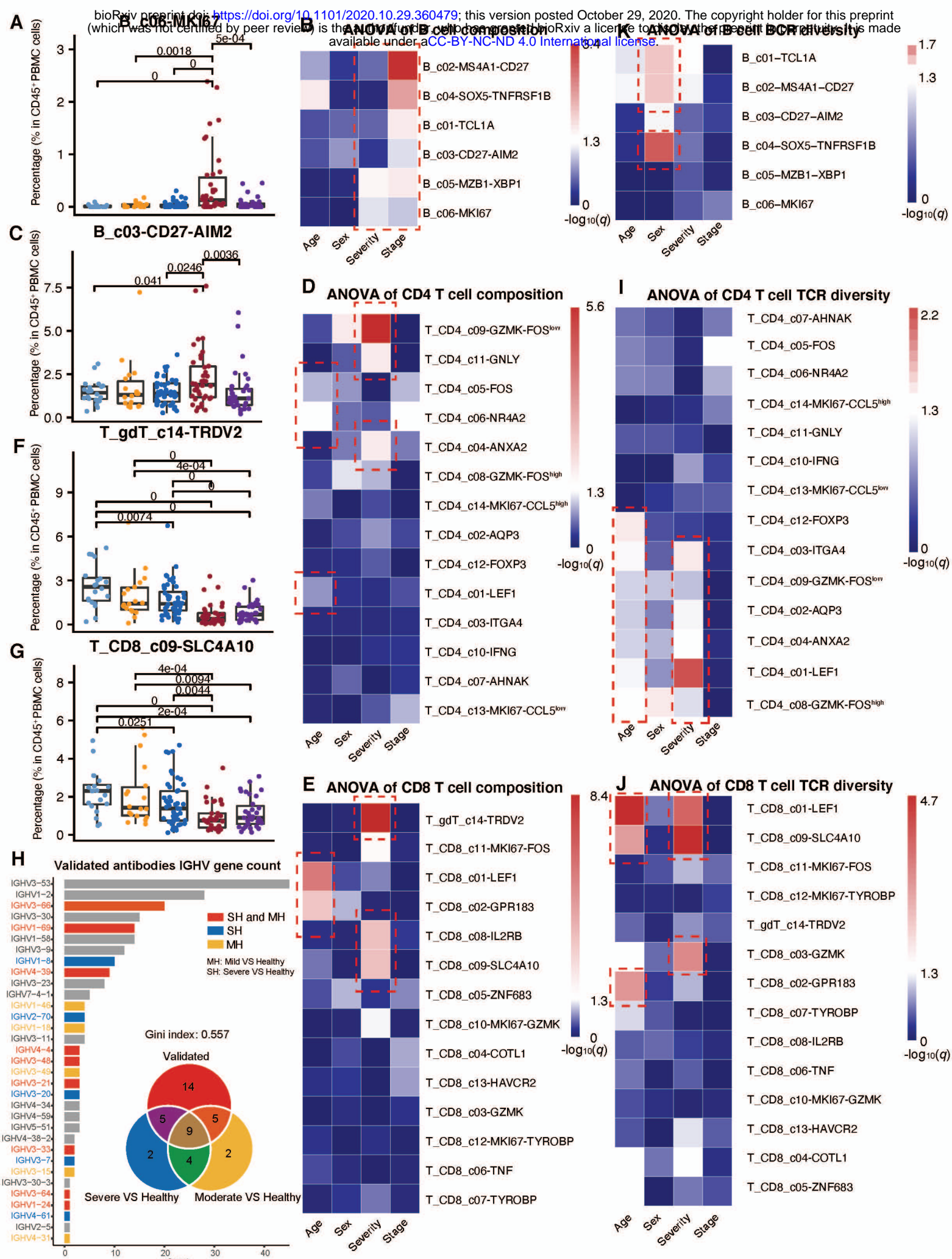






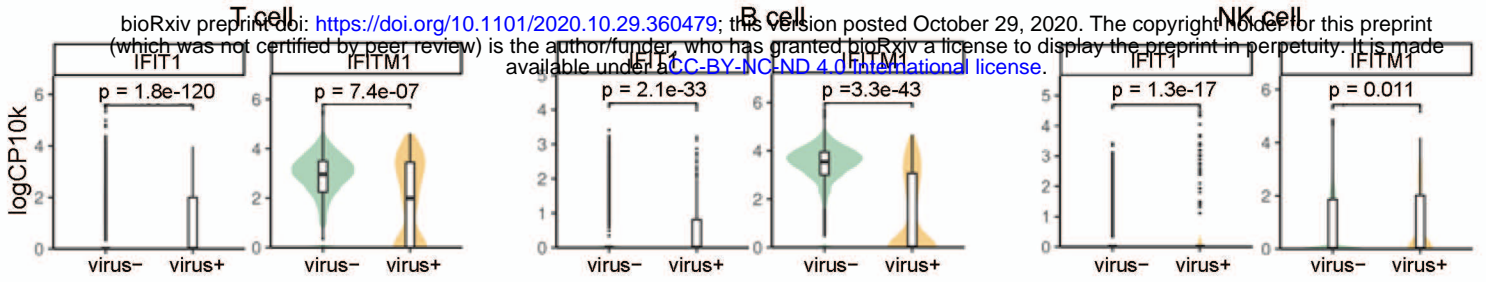






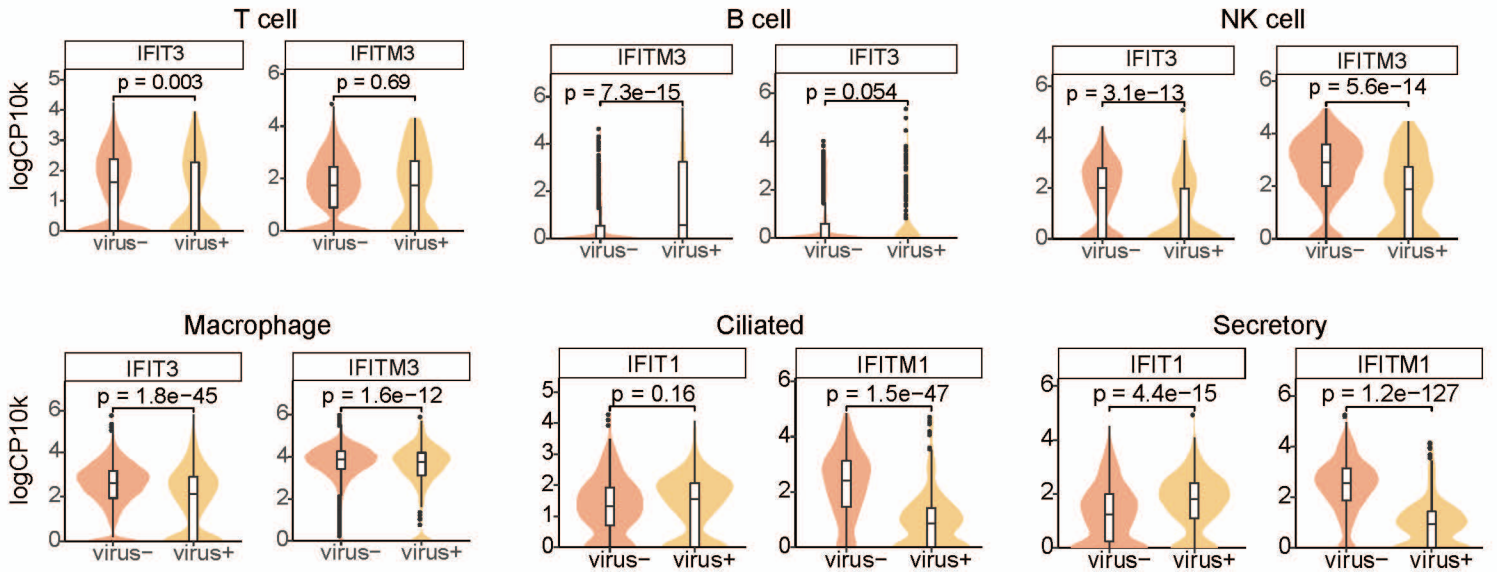
A

PBMC



B

BALF



C

