Single cell sequencing analysis uncovers genetics-influenced CD16+monocytes and memory CD8+T cells involved in severe COVID-19

Yunlong Ma[1,†], Fei Qiu[1,†], Chunyu Deng[1,†], Jingjing Li[2,†], Yukuan Huang[1,†], Zeyi Wu[1], Yijun Zhou[1], Yaru Zhang[1], Yichun Xiong[3], Yinghao Yao[3], Yigang Zhong[4], Jia Qu[1,] Jianzhong Su [1,3*]

[1] Institute of Biomedical Big Data, School of Ophthalmology & Optometry and Eye Hospital, School of Biomedical Engineering, Wenzhou Medical University, Wenzhou, 325027, China.

[2] State Key Laboratory for Diagnosis and Treatment of Infectious Diseases, the First Affiliated Hospital, Collaborative Innovation Center for Diagnosis and Treatment of Infectious Diseases, Zhejiang University School of Medicine, Hangzhou 310003, Zhejiang, China.

[3] Wenzhou Institute, University of Chinese Academy of Sciences, Wenzhou 325011, China.

[4] Department of Cardiology, Affiliated Hangzhou First People's Hospital, Zhejiang University school of Medicine, Hangzhou, China.

[†] These authors contributed equally to this work.

[*]**Correspondence should be addressed to**:

Jianzhong Su: sujz@wmu.edu.cn;

[1] Institute of Biomedical Big Data, School of Ophthalmology & Optometry and Eye Hospital, Wenzhou Medical University, Wenzhou, 325027, China.

[3] Wenzhou Institute, University of Chinese Academy of Sciences, Wenzhou 325011, China.

Abstract

Background: Understanding the host genetic architecture and viral immunity contributes to the development of effective vaccines and therapeutics for controlling the COVID-19 pandemic. Alterations of immune responses in peripheral blood mononuclear cells play a crucial role in the detrimental progression of COVID-19. However, the effects of host genetic factors on immune responses for severe COVID-19 remain largely unknown.

Methods: We constructed a powerful computational framework to characterize the host genetics-influenced immune cell subpopulations for severe COVID-19 by integrating GWAS summary statistics (N = 969,689 samples) with four independent scRNA-seq datasets (N = 606,534 cells).

Results: We found that 34 risk genes were significantly associated with severe COVID-19, and the number of highly-expressed genetics-risk genes increased with the severity of COVID-19. Three cell-subtypes that are CD16+monocytes, megakaryocytes, and memory CD8+T cells were significantly enriched by COVID-19-related genetic association signals. Notably, three causal risk genes of *CCR1*, *CXCR6*, and *ABO* were specifically expressed in these three cell types, respectively. *CCR1*$^{+}$CD16+monocytes and *ABO*$^{+}$ megakaryocytes with significant up-regulated genes including *S100A12*, *S100A8*, *S100A9*, and *IFITM1* confer higher risk to the cytokine storms among severe patients. *CXCR6*$^{+}$ memory CD8+ T cells exhibit a notable polyfunctionality of multiple immunologic features, including elevation of proliferation, migration, and chemotaxis. Moreover, we observed a prominent increase in cell-cell interactions of both *CCR1*$^{+}$ CD16+monocytes and *CXCR6*$^{+}$ memory CD8+T cells in severe patients compared to normal controls among both PBMCs and lung tissues, and elevated interactions with epithelial cells could contribute to enhance the resident to lung airway for against COVID-19 infection.

Conclusions: We uncover a major genetics-modulated immunological shift between mild and severe infection, including an increase in up-regulated genetic-risk genes, excessive secreted inflammatory cytokines, and functional immune cell subsets contributing high risk to severity,

which provides novel insights in parsing the host genetics-influenced immune cells for severe COVID-19.

Keywords: Single cell sequencing, GWAS, immune cells, inflammatory storm, COVID-19

Background

The coronavirus disease 2019 (COVID-19) outbreak, caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), has widely and severely jeopardized the health and economy systems of most countries worldwide. As of July 21th, 2021, there were more than 192.2 million confirmed patients with more than 4.12 million deaths in the whole world [1]. COVID-19 has distinct clinical manifestations ranging from asymptomatic to severe respiratory failure [2]. Mortalities of COVID-19 are largely derived from severe patients with interstitial pneumonia in both lungs and acute respiratory distress syndrome [3]. Many earlier studies [4-6] have shown that the number of severe COVID-19 patients who are elders and have comorbidities, such as diabetes and hypertension, has increased. In this connection, understanding the immunologic mechanism of severe COVID-19 and identifying novel vaccine targets to control the pandemic are of considerable interest.

Accumulating evidence have suggested that alterations of immune responses in peripheral blood mononuclear cells (PBMCs) and bronchoalveolar lavage fluid (BALF) play a crucial role in the detrimental progression of COVID-19 [7, 8]. There has been evidence that cytokine storm, usually found in severe COVID-19 patients, causes the adverse progression of COVID-19 [7]. Increased circulating levels of proinflammatory cytokine, including IL-10, IL-6 and TNF-α, have been reported to be associated with severe COVID-19 [7, 9]. Single-cell RNA sequencing (scRNA-seq) has been extensively utilized to reveal the immune responses of COVID-19 patients in both lung and peripheral blood [10-18]. Megakaryocytes and monocytes [11, 12], T cells exhaustion [14], lymphopenia [19], and increased levels of cytokines [20] may cause aberrant peripheral

immune activities in severe COVID-19 patients. Based on large-scale samples, previous studies identified that dysregulation of mTOR signaling pathway in dendritic cells [21] and aberrant myeloid cell subpopulations [16, 17] implicated in severe COVID-19. Su et al. [10] revealed an increase in inflammation and a sharp drop in blood nutrients between mild and moderate-to-severe COVID-19, and new subsets of immune cells emerged in moderate COVID-19 patients.

Genome-wide association study (GWAS) has emerged as a powerful approach to identify risk genes and genetic variants for complex diseases. By gathering population-based GWAS data worldwide, the COVID-19 Host Genetic Consortium has launched the "COVID-19 Host Genetics Initiative" project to facilitate COVID-19 host genetic research and identify genetic determinants of COVID-19 [22]. Subsequently, a growing number of GWASs have identified numerous significant genetic variants associated with COVID-19 susceptibility and severity [23-28]. Ellinghaus et al. [27] performed a meta-analysis of two independent GWAS datasets with 1,610 severe COVID-19 patients and 2,205 matched controls at seven hospitals in the Italian and Spanish epicenters, and identified two susceptibility loci of 3p21.31 and 9q34.2 to be significantly associated with severe COVID-19 at the genome-wide level. Based on a large-scale meta-analysis (N = 680,128), our group found that the *IFNAR2-IL10RB* gene cluster were significantly associated with COVID-19 susceptibility, and suggested that *IFNAR2* and *IL10RB* might have regulatory roles in the pulmonary immune response based on scRNA-seq data [25]. Consistently, Pairo-Gastineira et al. [24] conducted a GWAS study based on 2,244 critically ill COVID-19 patients and highlighted that several genes including *IFNAR2*, *DPP9*, and *OAS1* were significantly associated with severe COVID-19 at a genome-wide significance.

Two primary hypotheses were proposed for the involvement of immune genes in severe COVID-19 susceptibility. Whether the severe COVID-19-related risk genes associated with defective innate immune responses would induce persistent viral replication and resultant high viral loads, and whether an exaggerated genetically-mediated cytokine production contributes to the

hyper-inflammation and poor outcome among severe COVID-19. However, the effects of these genetic determinants on the peripheral immune cells for severe COVID-19 remain largely unknown. In view of a purely genetic study or single cell sequencing study cannot address this critical question, we here leveraged comprehensive computational methods to combine a large-scale GWAS summary dataset with scRNA-seq data for identifying host genetics-influenced immune cell subpopulations involved in the etiology of severe COVID-19.

Methods

Single cell RNA-seq data on severe COVID-19

In this study, we downloaded four independent scRNA-seq datasets on COVID-19 in PBMC and BALF from the ArrayExpress database (Dataset #1, the accession number is E-MTAB-9357 from Su et al. study [10]), and the Gene Expression Omnibus (GEO) database (Dataset #2, the accession number is GSE149689 from Lee et al. study [18], Dataset #3, the accession number is GSE150861 from Guo et al. study [11], and Dataset #4, the accession number is GSE158055 [29]). The first dataset contained 270 peripheral blood samples including 254 samples with different COVID-19 severity (i.e., mild N = 109, moderate N = 102, and severe N = 50) and 16 healthy controls for scRNA-seq analysis. There were eight patients in dataset #2 with COVID-19 of varying clinical severity, including asymptomatic, mild, and severe, and four healthy controls with PBMCs. The dataset #3 included five peripheral blood samples collected from two severe COVID-19 patients at three different time points during tocilizumab treatment, containing two different stages: severe stage and remission stage. Within the dataset #4, 12 BALF samples were collected from lung tissues, including three moderate and nine severe patients. For all the datasets, the sample collection process were reviewed and approved by Institutional Review Board at the institutions where samples were originally collected. The COVID-19 severity was evaluated by using the World Health Organization (WHO) ordinal scale (WOS), the National Early Warning Score (NEWS), or

the Diagnosis and Treatment of COVID-19 (Trail Version 6). Single-cell transcriptomes for these four datasets were gathered using the 10 × Genomics scRNA-seq platform.

Single cell RNA sequencing data processing

We performed normalization, clustering, and dimensionality reduction, differential expression gene (DEG) analysis, and visualization on these four independent scRNA-seq datasets with the Seurat R package [30]. The *SCTransform* function was used to scale and transform data, and linear regression model was applied to omit redundant variations caused by cellular complexity (i.e., cells expressed less than 200 genes or more than 2,500 genes were removed) or cellular quality (i.e., cells that had UMIs more than 10,000 and expressed reads of mitochondrial genes greater than 10% were removed). The *CellCycleSoring* function was applied to remove the effects of confounding factors. Principal component analysis (PCA) was carried out to extract principal components (PCs) that could explain most of datasets via using high variable genes. Top 20 PCs were utilized to conduct uniform manifold approximation and projection (UMAP) to embed the dataset into two dimensions. Subsequently, we constructed a shared nearest-neighbor graph (SNN) using the *FindNeighbors* function based on the top 20 PCs, and applied a graph-based modularity-optimization algorithm from the Louvain method [31] on this SNN for clustering the dataset with the cluster resolution set to 0.5. We used the *RunHarmony* function with PCA reduction method from harmony R package [32] to integrate samples to correct batch effects. The *FindConservedMarkers* function in Seurat was implemented to find differential expressed genes for determining cellular identity. Well-defined markers were used to annotate clusters, and uncharacterized clusters in the first round of clustering were extracted to run the second round of clustering (Supplemental Table S2). A total of 606,534 cells with 563,856 PBMC cells and 42,678 BALF cells were yielded from 300 samples based on the four independent scRNA-seq datasets (Supplemental Table S1 and Figure 1A). To allow comparison across samples and datasets, we

used a common dictionary of gene symbols to annotate genes and these unrecognized symbols were removed.

GWAS summary data on hospitalized COVID-19

The meta-GWAS summary data on severe COVID-19 round 4 (B2_ALL, Susceptibility [Hospitalized COVID-19 vs. Population]) were downloaded from the official website of the COVID-19 Host Genetic Consortium [22] (https://www.covid19hg.org/; analyzed file named: "COVID19_HGI_B2_ALL_leave_23andme_20201020.txt.gz"; released date of October 4 2020). There were 7,885 hospitalized COVID-19 patients and 961,804 control participants from 21 independent contributing studies. There was an overwhelming majority of participants in these contributing studies with European ancestry (93%). The meta-GWAS summary statistics contained P values, Wald statistic, inverse-variance meta-analyzed log Odds Ratio (OR) and related standard errors. The 1,000 Genomes Project European Phase 3 [33] was used as a panel for pruning. Results from 23&Me cohort GWAS summary statistics were excluded from our current analysis. Genetic variants without RefSNP number in the Human Genome reference builds 37 were filtered out, giving a total of 9,368,170 genetic variants satisfying the major allele frequency (MAF) over 0.0001 and the imputation score of greater tha 0.6. We used the *qqman* R package to figure both Manhattan plot and quantile-quantile (QQ) plot, and the web-based software of *LocusZoom* (http://locuszoom.sph.umich.edu/)[34] to visualize the regional association plots for significant risk loci.

Hierarchical clustering analysis

To examine the similarity of the transcriptome profiles between cell types across different COVID-19 severities (Supplemental Figure S4), we merged the counts of UMI for each cell type according to normal, mild, moderate, and severe COVID-19. In order to normalize gene expression,

we divided the counts of UMI for each gene by the counts of total UMI for all genes in each cell type and then multiplied by 100,000, as refer to the method in a previous study [18]. A median expression value of greater than 0.5 was used to calculate the relative change in each gene expression by dividing it by the median value for each gene, and the Pearson correlation coefficient (PPC) of the relative change in gene expression was used for current hierarchical clustering analysis.

Gene-based association analysis

To perform a gene-based genetic association analysis of the meta-GWAS summary statistics on severe COVID-19, we leveraged the updated SNP-wise Mean model of MAGMA [35]. In this model, MAGMA computes a test statistic:

$$T = \sum_i^N Z_i^2 = \boldsymbol{Z}^T \boldsymbol{Z}$$

where N is the number of SNPs mapped in a gene and $Z_i = \phi(p_i)$. Of note, ϕ is the cumulative normal distribution function and p_i is the marginal P value for a given SNP *i*. SNPs belonging to a specific gene were based on whether located in the gene body or within the +/- 20 kb upstream or downstream region of the gene. Furthermore, the model assumes $\boldsymbol{Z} \sim \text{MVN}(\boldsymbol{0}, \boldsymbol{S})$, where $\boldsymbol{S}$ is the LD matrix of the SNP genotypes. The LD matrix can be diagonalized and hence written as $\boldsymbol{S} = \boldsymbol{Q}\boldsymbol{A}\boldsymbol{Q}^T$, where $\boldsymbol{Q}$ is an orthogonal matrix and $\boldsymbol{A} = \text{diag}(\lambda_1, \lambda_2, ..., \lambda_N)$ with λ_j being the *j*th eigenvalue of $\boldsymbol{S}$. The 1,000 Genomes Project Phase 3 European Panel [33] was used for calculating the LD information among SNPs extracted from GWAS summary data on COVID-19. $\boldsymbol{D} \sim \text{MVN}(\boldsymbol{0}, \boldsymbol{I}_K)$ is a random variable, where $\boldsymbol{D} = \boldsymbol{A}^{-0.5}\boldsymbol{Q}^T\boldsymbol{Z}$. Then the sum of squared SNP Z-statistics as the following formula:

$$T = \boldsymbol{Z}^T\boldsymbol{Z} = (\boldsymbol{Q}\boldsymbol{A}^{0.5}\boldsymbol{D})^T\boldsymbol{Q}\boldsymbol{A}^{0.5}\boldsymbol{D} = \boldsymbol{D}^T\boldsymbol{A}\boldsymbol{D} = \sum_i^N \lambda_i D_i^2$$

with $D_i \sim N(0,1)$ and $D_i^2 \sim \chi_1^2$. Namely, T follows a mixture distribution of independent χ_1^2 random variables. A total of 19,138 genes were included in the current analysis. We used the Benjamini-Hochberg false discovery rate (FDR) method, in which a gene with a FDR ≤ 0.05 ($P \leq 6.8 \times 10^{-5}$) was interpreted as significant, to adjust for multiple testing.

Pathway enrichment analysis

We applied the *built-in* functions of MAGMA [35], using the results from GWAS summary statistics as its input, to examine genome-wide enriched biological pathways for severe COVID-19. We calculated competitive P values by examining the results that the combined effect of genes within a pathway is significantly greater than the combined effect of all other genes, and 10,000 permutations was used to adjust competitive P values. Additionally, we leveraged the over-representation algorithm of the WebGestalt (http://www.webgestalt.org) [36] along with the significant genes as an input list to conduct a pathway enrichment analysis using the KEGG pathway resource [37]. The number of genes in each pathway was set to between 5 and 2,000, and the Benjamini-Hochberg FDR was used for multiple correction. To cluster these identified KEGG pathways, we performed a multidimensional scaling (MDS) analysis based on the Jaccard distance method [38], and constructed a pathway-pathway interaction network for these significantly enriched pathways setting the Jaccard distance > 0.1.

Combining GWAS-based genetic signals with eQTL data

To uncover genetically-regulatory expression of genes associated with severe COVID-19, we conducted an integrative genomics analysis by using the S-PrediXcan [39] by combining meta-GWAS summary statistics with expression quantitative trait loci (eQTL) data for 49 tissues from the GTEx Project (version 8). S-PrediXcan mainly uses two linear regression models to analyze the association between predicted gene expression and severe COVID-19:

$$\boldsymbol{Y} = \boldsymbol{\alpha}_1 + \boldsymbol{X}_l \beta_l + \boldsymbol{\varepsilon}_1$$

$$\boldsymbol{Y} = \boldsymbol{\alpha}_2 + \boldsymbol{G}_g \gamma_g + \boldsymbol{\varepsilon}_2$$

where $\boldsymbol{\alpha}_1$ and $\boldsymbol{\alpha}_2$ are intercepts, $\boldsymbol{\varepsilon}_1$ and $\boldsymbol{\varepsilon}_2$ are independent error terms, $\boldsymbol{Y}$ is the n dimensional vector for n individuals, $\boldsymbol{X}_l$ is the allelic dosage for SNP l in n individuals, β_l is the effect size of SNP l, $\boldsymbol{G}_g = \sum_{i \in gene(g)} \omega_{ig} \boldsymbol{X}_i$ is the predicted expression calculated by ω_{lg} and $\boldsymbol{X}_l$, in which ω_{lg} is derived from the GTEx Project , and γ_g is the effect size of $\boldsymbol{G}_g$. The Z-score (Wald-statistic) of the association between predicted gene expression and severe COVID-19 can be transformed as:

$$Z_g = \frac{\hat{\gamma}_g}{\mathrm{se}(\hat{\gamma}_g)} \approx \sum_{i \in gene(g)} \omega_{ig} \frac{\hat{\sigma}_i}{\hat{\sigma}_g} \frac{\hat{\beta}_i}{\mathrm{se}(\hat{\beta}_i)}$$

where $\hat{\sigma}_g$ is the standard deviation of $\boldsymbol{G}_g$ and can be calculated from the 1,000 Genomes Project European Phase 3 Panel, $\hat{\beta}_l$ is the effect size from GWAS on COVID-19 and $\hat{\sigma}_l$ is the standard deviation of $\hat{\beta}_l$. S-PrediXcan was run for each of 49 tissues with 659,158 gene-tissue pairs.

Furthermore, to increase the power to discover significant genes whose expression has similar regulations across multi-tissues, we utilized the S-MultiXcan [40] to meta-analyze these results from above S-PrediXcan analysis. S-MultiXcan fits a linear regression model of severe COVID-19 on predicted expression from multiple tissue models jointly:

$$\boldsymbol{Y} = \sum_{j=1}^{p} \boldsymbol{T}_j g_j + \boldsymbol{e} = \boldsymbol{T}\boldsymbol{g} + \boldsymbol{e}$$

where $\tilde{\boldsymbol{T}}_j = \sum_{i \in gene(j)} \omega_i \boldsymbol{X}_i$ is the predicted expression of tissue j, and $\boldsymbol{T}_j$ is the standardization of $\tilde{\boldsymbol{T}}_j$ to $mean = 0$ and $standard\ deviation = 1$. g_j is the effect size for the predicted gene expression in tissue j, $\boldsymbol{e}$ is an error term with variance σ_e^2, and p is the number of included tissues. There were 22,326 genes across 49 GTEx tissues with integrated convergent evidence in S-MultiXcan, and a gene with a value of FDR ≤ 0.05 ($P \leq 3.8 \times 10^{-5}$) is considered to be significant.

In silico permutation analysis

To explore the concordance of results from both MAGMA analysis (Gene set #1: N = 944, P ≤ 0.05) and S-MultiXcan analysis (Gene set #2: N =1,274, P ≤ 0.05), we performed an *in silico* permutation analysis which consisted 100,000 times (N_{Total}) random selections [41, 42]. We first calculated the number of overlapped genes between Gene Set #1 and #2 ($N_{Observation}$ = 302), then employed the total number of genes in S-MultiXcan analysis as background genes ($N_{Background}$ = 22,326). By randomly selecting the same number of genes as Gene set #2 (N = 1,274) from the background genes, and after repeating it 100,000 times, we calculated the number of overlapped genes between Gene Set #1 and the sample we selected each time (N_{Random}). Finally, we calculated the empirically permuted P value using the following formula: $P = \frac{N_{Random} \geq N_{Observation}}{N_{Total}}$, and empirical P value ≤ 0.05 is considered to be significant.

Drug-gene interaction analysis

We conducted a drug-gene interaction analysis for identified genetics-risk genes by using protein-chemical interactions in the context of STRING-based PPI networks [43] and STITCH-based drug annotation information (v5.0, http://stitch.embl.de/) [44]. Only experimentally-validated gene-drug interactions with ranked confidence score were selected for constructing a drug-gene interaction network. To examine the potential therapeutic effects of highly-expressed genes in each immune cell, we conducted an enrichment analysis of 43 druggable categories based on the DGIdb database (https://www.dgidb.org/druggable_gene_categories). Additionally, we collected 1,263 human druggable proteins, which are therapeutic targets of clinical stage or approved drugs, from a previous study [26]. Among them, 704 proteins are targets for potential COVID-19-relevant drugs based on registers of clinical trials for COVID-19, approved

268 immunomodulatory/anticoagulant drugs, or have biological functions associated with SARS-CoV-
269 2 infection (Supplemental Table S11).
270
271 ***Integrated analysis of GWAS summary statistics and scRNA-seq data***
272 To identify genetically regulatory-related peripheral immune cells for severe COVID-19, we
273 implemented the RolyPoly algorithm [45] to incorporate GWAS summary statistics with scRNA-
274 seq data. Let $g(i)$ stands for the gene associated with SNP i, $S_j = \{i : g(i) = j\}$ be the SNP set with
275 multiple SNPs associated with the gene j, and $\boldsymbol{\beta}_{S_j}$ be a GWAS-based effect-size vector of S_j with
276 a *priori* assumption that $\boldsymbol{\beta}_{S_j} \sim \text{MVN}(\mathbf{0}, \sigma_j^2 \boldsymbol{I}_{|S_j|})$. Following the *priori*, RolyPoly gives a polygenic
277 linear model for $\boldsymbol{\beta}_{S_j}$:

278 $$\sigma_j^2 = \gamma_0 + \sum_{i=1}^{N} \gamma_i \alpha_{ji}$$

279 where γ_0 is an intercept term, $\alpha_{ji} (i = 1,2,...,N)$ are annotations such as cell-type-specific gene
280 expression, and γ_i are annotation coefficients for α_{ji}. To fit the observed and expected sum
281 squared SNP effect sizes related to each gene by using the method-of-moments estimators,
282 RolyPoly estimates γ_i by the following equation:

283 $$E(\sum_{i \in S_j} \hat{\beta}_i^2) = \sigma_j^2 \text{Tr}(\boldsymbol{R}_{S_j}^2) + |S_j| \sigma_e^2 n^{-1}$$

284 where $\boldsymbol{R}_{S_j}$ is the LD matrix of S_j and Tr represents the trace of a matrix. Finally, RolyPoly applies
285 the block bootstrap method with 1,000 iterations to estimate standard errors $\hat{\sigma}_{\gamma_i}$ for calculating a *t*-
286 statistic and corresponding P values. The PLINK (v1.90) [46] was used to calculate the LD between
287 SNPs within the 1 Mb window based on the 1,000 Genome Project European Phase 3 panel [33].
288 We restricted the analysis to SNPs in the autosomes, and any SNPs with MAF $\leq$ 5% were excluded.

The major histocompatibility complex region (Chr6: 25-35 Mbp) was also excluded due to the extensive LD in this region.

Defining cell state scores

We leveraged cell state scores to assess the immunological degree of each immune cell type expressed a pre-curated expression gene set [11, 14, 29]. The cell state scores (CTS) were calculated based on the average expression of genes from the pre-curated gene set in the respective cell with the following formula: CTS*k*(*m*) = average(RE(GS*k*, *m*)) - average(RE(GS*n*, m)), where GS*k* is a pre-defined gene set *k* in a given cell *m*, and GS*n* is a control gene set that was randomly chosen on the basis of aggregate expression levels bins, which obtain a comparable distribution of expression levels and over size to that of the pre-curated gene set. RE represents the relative expression of GS*k* or GS*n*. The *AddModuleScore* function in Seurat [30] was applied to calculate the CTS with default parameters. We used the inflammatory and cytokine genes (N = 324 genes, Supplemental Table S10), cytokine-cytokine receptor interactions (N = 294 genes), chemokine signaling pathway (N = 189 genes), T cell activation (GO: 0042110), response to interferon alpha (GO: 0035455), response to interferon beta (GO: 0035456), leukocyte migration (GO: 0050900), 5 well-defined proliferating markers (*MK167*, *TYMS*, *NKG7*, *IL7R*, and *CCR7*), 6 well-defined exhaustion markers (*LAG3*, *TIGIT*, *PDCD1*, *CTLA4*, *HAVCR2*, and *TOX*), and 12 cytotoxicity-associated genes (*PRF1*, *IFNG*, *GNLY*, *NKG7*, *GZMB*, *GZMA*, *GZMH*, *KLRK1*, *KLRB1*, *KLRD1*, *CTSW*, and *CST7*) to define inflammatory cytokine, chemokine, T cell activation, IFN-α/β response, migration, proliferation, exhaustion, and cytotoxicity score, respectively.

Cell-to-cell interaction analysis

To identify potential cellular interactions of *CCR1*$^{+}$ CD16+monocytes and *CXCR6*$^{+}$ memory CD8+T cells with other immune cells, we utilized the CellChat R package [47] for inferring the

predicted cell-to-cell communications based on two normalized scRNA-seq datasets (dataset #1 of PBMC and dataset #4 of BALF). CellChat algorithm could examine the significance of ligand-receptor interactions between two cell types depending on the expression of important factors, including stimulatory and inhibitory membrane-bound co-receptors, soluble agonists and antagonists. The communication probability of a signaling pathway was derived from the sum of probabilities of their ligand-receptor interactions. We only concentrated on the ligand-receptor interactions that significantly associated with severe COVID-19 compared with normal control.

Statistical analysis

The Wilcoxon sum-rank test was used to assess DEGs in mild, moderate, and severe COVID-19 group compared with normal control. The Mann-Kendall trend analysis was applied to evaluate the significance of cell state cells with elevated severities of COVID-19. Pathway- and disease-based enrichment analyses used the hypergeometric test to identify remarkable biological pathways and disease-terms. The Pearson correlation analysis was used to calculate the correlation coefficient of highly-expressed genes in *CCR1*$^+$ CD16+monocytes between moderate and severe patients. The paired Student's t test was used to calculate the significance of ligand/receptor interactions of *CCR1*$^+$ CD16+monocytes and *CXCR6*$^+$ memory CD8+T cells with other immune cells between normal control and severe COVID-19.

Results

The computational framework for integrating single-cell transcriptomes and GWAS on COVID-19

As shown in Figure 1, we devised a computational framework to parse the host genetics-modulated immune cell subpopulations implicated in severe COVID-19. It included three main parts: 1) integrative analysis that combined GWAS summary statistics with scRNA-seq data to

genetically map single-cell landscape for severe COVID-19 (Figure 1A); 2) identifying genetics-risk genes, pathways, and immune cell subpopulations that contributed to cytokine storms among severe patients (Figure 1B); and 3) uncovering the cellular interactions of genetics-modulated immune cell subsets, as well as their functions with cells in lung tissues (Figure 1C).

Identification of immune cell types associated with severe COVID-19

To parse the host genetics-influenced immune responses at single cellular level in PBMCs for severe COVID-19, we subjected three independent scRNA-seq datasets with 563,856 cells to UMAP based on highly variable genes using the Seurat (See Methods) [30]. There was identification of 13 distinct clusters unbiased by patients with different severities (Supplemental Figure S1). We leveraged well-known marker genes to assign these clusters to 13 distinct cell types, including mature B cells, megakaryocytes, naïve B cells, CD34+progenitors, dendritic cells, natural killer (NK) cells, CD14+monocytes, CD16+monocytes, memory CD4+T cells, naïve CD4+T cells, naïve CD8+T cells, memory CD8+T cells, and effector CD8+T cells (Supplemental Figures S2-S3).

While performing the hierarchical clustering analysis on the scRNA-seq profiles, we discovered that cell types were the primary determinants of their clustering, followed by disease severities, indicating both COVID-19 pathology and immune cell types might have crucial roles in altered patterns of immune transcriptome instead of technical artifacts (Supplemental Figure S4). As a vital feature for reflecting the alterations of immune responses, we examined the relative proportions of peripheral immune cells across different COVID-19 groups in comparison with normal group. The proportions of CD14+monocytes, megakaryocytes, and CD34+progenitors were significantly elevated in moderate and severe patients, whereas the proportions of CD16+monocytes, effector CD8+ T cells, memory CD8+T cells, memory CD4+T cells, naïve

CD4+T cells, and NK cells were significantly decreased with the increased severities (Supplemental Figure S5).

Identification of genetic risk loci associated with severe COVID-19

Through performing a meta-analysis of 21 independent GWAS studies from the COVID-19 Host Genetic Consortium, eight genomic loci were identified to be associated with hospitalized COVID-19 at a genome-wide significant level, including 1p22.2 (rs2166172, $P = 2.74 \times 10^{-8}$), 3p21.31 (rs35081325, $P = 3.32 \times 10^{-58}$, and rs33998492, $P = 3.59 \times 10^{-14}$), 6p21.33 (rs143334143, $P = 1.28 \times 10^{-10}$), 7p11.2 (rs622568, $P = 2.57 \times 10^{-8}$), 9q34.2 (rs505922, $P = 2.24 \times 10^{-9}$), 12q24.13 (rs2269899, $P = 3.24 \times 10^{-8}$), 19p13.3 (rs2109069, $P = 6.4 \times 10^{-13}$), and 21q22.11 (rs13050728, $P = 1.91 \times 10^{-11}$) (Figure 2A, Supplemental Table S3, Figure S6, and Materials S2). Among these eight loci, three loci, 1p22.2, 6p21.33 and 7p11.2, were newly identified. It should be noted that there were two independent genetic association signals (Index SNPs: rs35081325 and rs33998492) in the 3p21.31 locus for severe COVID-19 (Figure 2B and Supplemental Figure S7A-C). Using the Variant2Gene (V2G) algorithm [48], we prioritized *CXCR6* as a candidate causal gene for rs35081325 and causal gene *CCR1* for rs33998492 (Supplemental Method S1).

Furthermore, the index SNP of rs505922 ($P = 2.24 \times 10^{-9}$) in the 9q34.2 locus is highly LD with the reported SNP of rs657152 ($R^2 = 0.874$) [27] and rs8176719 ($R^2 = 0.876$) [25]. Based on the top-ranked V2G score for rs505922, we prioritized *ABO* as a potential causal gene contributing susceptibility to severe COVID-19. By performing a MAGMA gene-level association analysis, we observed that 25 genes including *CXCR6*, *CCR1*, *IFNAR2*, *IL10RB*, and *OAS1* were significantly associated with severe COVID-19 (FDR < 0.05, Supplemental Figure S8 and Table S4). GWAS-based pathway enrichment analysis revealed that 19 biological pathways, including cytokine-cytokine receptor interaction, influenza A, and TNF signaling, were significantly associated with hospitalized COVID-19 (Supplemental Figure S9 and Table S5).

Integrative analysis of GWAS on severe COVID-19 with GTEx eQTL data

To obtain combined signals from multiple tissues [49], we leveraged S-MultiXcan to meta-analyze the tissue-specific associations from 49 tissues in GTEx (see Methods), which showed that the genetically predicted expressions of 16 genes were significantly associated with severe COVID-19 (FDR < 0.05, Figure 2C and Supplemental Table S6). Of note, 14 of 16 genes (87.5%) were identified to be significant in MAGMA analysis (Supplemental Figure S10A-B). Through conducting S-PrediXcan analysis of blood and lung tissues that were linked with SARS-CoV-2 infection, we found eight genes whose genetically-regulated expression were significantly associated with severe COVID-19 (FDR < 0.05, Supplemental Table S7). Using *in silico* permutation analysis, we further observed that there existed a high consistence among results from MAGMA, S-PrediXcan, and S-MultiXcan analyses ($P < 1.0 \times 10^{-5}$, Supplemental Figure S11A-C). The aforementioned multiple genomic analyses identified 34 risk genes that showed supportive evidence of involvement in the etiology of COVID-19 (Supplemental Figure S12A-B).

Functional characterization of 34 risk genes for severe COVID-19

The result of a network-based enrichment analysis suggested that 22 of 34 risk genes were significantly enriched in a PPI subnetwork ($P = 2.85 \times 10^{-13}$, Figure 2D), which is consistent with the consensus that disease-related genes are more densely connected [50, 51]. To functionally characterize the drug targets of these genes, we conducted a drug-gene interaction analysis and identified 11 genes including *CCR1*, *IFNAR2*, *IL10RB*, and *OAS1* were targeted by at least one known drug (Figure 2D and Supplemental Figure S14), of which some genes including *CCR1*, *IFNAR2*, and *IL10RB* have been reported to be drug targets for treating severe COVID-19 patients [25, 26]. Furthermore, these 34 genes were significantly enriched in a functional module consisting of 10 biological pathways (Figure 2E, Supplemental Table S8 and Figure S13), among which two

top-ranked ones being cytokine-cytokine receptor interaction and chemokine signaling pathway (FDR < 0.05). Most of these enriched pathways have been reported to be implicated in COVID-19 [25, 52, 53].

Based on the expression profile of dataset #1, we conducted a hierarchical clustering analysis of these identified risk genes on COVID-19 severity, and found that these risk genes predisposed be highly-expressed in severe patients compared to normal group (Permuted P = 0.023, Figure 2F-G). Consistently, the number of significant enriched pathways were elevated with increased severities (Figure 2H). Genes in both cytokine-cytokine receptor interaction and chemokine signaling pathways showed significantly high expressions in the early phase of SARS-CoV-2 infection (Figure 2H), suggesting that these two pathways could play critical roles in the initiation of COVID-19.

Genetics-influenced peripheral immune cell types for severe COVID-19

To identify genome-wide genetics-influenced immune cells for severe COVID-19, we first leveraged a regression-based polygenic model to integrate GWAS summary data on severe COVID-19 with single-cell transcriptomic profiles (dataset #1) according to different COVID-19 severities (See methods). We found that CD16+monocytes were significantly associated with three phases of COVID-19, mature B cells showed remarkable associations with mild COVID-19, megakaryocytes were significantly associated with moderate and severe COVID-19, and memory CD8+T cells showed significant associations with severe COVID-19 (simulated P < 0.05, Figure 3A). Further, we used a generalized linear regression model to validate these severe COVID-19-associated cell types by conditioning on the 10% most specific genes for each type, and consistently found that CD16+monocytes and megakaryocytes showed notable associations with severe COVID-19 (P < 0.05, Supplemental Method S2). These results indicated that CD16+monocytes,

megakaryocytes, and memory CD8+T cells were more vulnerable to the influence of genetic components on severe-stage patients.

Based on the specificity algorithm used in MAGMA, we noticed that the top specific cell type of *CCR1* was CD16+monocytes, *CXCR6* was most specifically expressed in memory CD8+T cells, and *ABO* was specific to megakaryocytes (Supplemental Figure S15A), recalling that *CXCR6*, *CCR1* and *ABO* were prioritized to be candidate causal genes for severe COVID-19 based on the V2G score in above genetics-based analysis. Compared with other cell types, *CCR1* was primarily expressed in CD16+monocytes (24.77%), *CXCR6* was mainly expressed in memory CD8+T cells (40.29%), and the *ABO*-expressed cells were highly specific to megakaryocytes (54.63%) (Supplemental Figure S15B and Table S9). To gather additional empirical support, we analyzed the combined dataset of both datasets #2 and #3 as a validation and found *CCR1*, *CXCR6*, and *ABO* showed a consistent specificity in the three cell types (Supplemental Figure S16).

Given that the primary goal of current study was to characterize genetics-influenced peripheral immune cell types for severe COVID-19, the majority of our subsequent detailed analyses would be concentrated on three immune cell subpopulations: $CCR1^+$ CD16+monocytes, ABO^+ megakaryocytes, and $CXCR6^+$ memory CD8+T cells (Figure 3B).

$CCR1^+$ CD16+monocytes and ABO^+ megakaryocytes contributing higher risk to cytokine storm

The accumulating lines of evidence [29, 54] have suggested that subsets of monocytes and megakaryocytes might be the major resources of inflammatory storm. We sought to examine whether $CCR1^+$ CD16+monocytes and ABO^+ megakaryocytes play more important roles in cytokine storm among severe patients. As for $CCR1^+$ CD16+monocytes, we found that the inflammatory cytokine score was significantly higher than that of $CCR1^-$ CD16+monocytes ($P = 2.5\times10^{-7}$, Figure 4A). Consistently, the combined score of both cytokine-cytokine receptor interaction and chemokine signaling pathway was prominently higher in $CCR1^+$ CD16+monocytes

($P < 2.2 \times 10^{-16}$, Supplemental Figure 17A). Compared with *CCR1*$^{-}$ CD16+monocytes, there were 351 significantly highly-expressed genes in *CCR1*$^{+}$ CD16+monocytes, such as inflammatory and cytokine genes of *IL1B*, *IL27*, *CXCL10*, *CXCL8*, *CD14*, and *OSM* (FDR < 0.05, Figure 4B and Supplemental Table S11), which have been documented to be associated with the inflammatory response and chemotaxis of immune cells among COVID-19 patients [10, 15, 55, 56]. Functionally, 19 KEGG pathways were significantly overrepresented by the 351 highly-expressed genes (FDR < 0.05, Figure 4C and Supplemental Table S12), including cytokine-cytokine receptor interaction and chemokine signaling pathway, reminiscing that most of them were identified in above genetics-based pathway analysis. Additionally, these highly-expressed genes among *CCR1*$^{+}$ CD16+monocytes have a remarkably higher proportion of druggable genes and COVID-19-associated druggable genes ($P \leq 0.01$, Supplemental Figure S17 and Table S13).

The cell percentage of *CCR1*$^{+}$ CD16+monocytes showed a notable elevation among moderate and severe patients compared with normal controls ($P < 0.001$), with no significant difference between mild patients and normal controls ($P = 0.1$, Figure 4D). Furthermore, the inflammatory cytokine scores among *CCR1*$^{+}$ CD16+monocytes were significantly elevated with increased severities (Trend $P = 0.0013$, Figure 4E). In comparison with normal controls, mild, moderate, and severe patients displayed significantly up-regulated expressions (up-DEGs) with 14, 169, and 190 genes respectively (FDR < 0.05, Figure 4F and Supplemental Figure S17D). Notably, there existed a high correlation between up-DEGs of moderate and severe patients ($r = 0.937$, $P < 2.2 \times 10^{-16}$; Figure 4G), such as *S100A8*, *S100A9*, and *IFITM1* (Figure 4H-4J), indicating a similar expression pattern between moderate and severe patients. Accumulating release of massive amounts of calprotectin (*S100A8/S100A9*) in monocytes contributes to inflammatory response among severe COVID-19 patients [10, 16, 29].

Furthermore, these 190 up-DEGs were significantly enriched in disease-terms associated with viral infection and inflammation and 17 functional GO-terms (FDR < 0.05, Figure 4K,

Supplemental Figure S17E and Tables S14-S15), including interferon alpha/beta signaling and interferon gamma signaling. These interferon-related genes including *IRF3*, *IRF2*, *IFI6*, *IFITM1*, *ISG15*, and *ICAM1* may induce autoinflammatory and autoimmune conditions contributing to the innate immune cells against SARS-CoV-2 infection [57, 58]. Of note, a high proportion of 63.68% among 190 up-DEGs such as *CXCL8*, *IFITM1*, *S100A8*, and *S100A9* were annotated into 15 potential druggable gene categories (Supplemental Figure S17F-L and Table S16). These results indicated that interferon-related genes among *CCR1*$^+$ CD16+monocytes have instrumental effects in exacerbating inflammation among severe patients.

In addition, we found that *ABO*$^+$ megakaryocytes had a significantly higher inflammatory cytokine score than that in *ABO*$^-$ cells ($P < 0.001$, Supplemental Figure S18A-B). Compared with *ABO*$^-$ megakaryocytes, 424 genes were significantly highly-expressed in *ABO*$^+$ megakaryocytes (FDR < 0.05, Supplemental Figure S18C and Table S17). These 424 highly-expressed genes were significantly enriched in systemic lupus erythematosus, alcoholism, and platelet activation (FDR < 0.05, Supplemental Figure S18D and Table S18). Similar to *CCR1*$^+$ CD16+monocytes, the cell percentage of *ABO*$^+$ megakaryocytes was significantly elevated among moderate and severe patients ($P < 0.01$, Supplemental Figure S18E). Among *ABO*$^+$ megakaryocytes, 20 and 35 up-DEGs were notably associated with moderate and severe patients, respectively (FDR < 0.05, Supplemental Figure S18F-G). There was a highly overlapped rate of these up-DEGs between moderate and severe COVID-19 groups, including *ACP1*, *S100A8*, and *A100A9* (18/20 = 90%, Supplemental Figure S18F-N). These 35 up-DEGs were annotated to 12 druggable gene categories and significantly enriched in several disease terms (Supplemental Figure S18H and Tables S19-S20), such as shock and thrombocytopenia, which were reported to be associated with COVID-19 [59]. Overall, these results suggest that both *CCR1*$^+$ CD16+monocytes and *ABO*$^+$ megakaryocytes contribute higher risk to inflammatory storm among severe patients.

CXCR6^{+} *memory CD8+T cells convey risk to severe COVID-19*

Earlier studies [10, 60] have indicated that polyfunctional T cells play important roles in dominating the anti-viral infection immune response and can release a substantially higher amount of multiple distinct cytokines and chemokines in comparison to other T cells. It is plausible to infer that there exist subsets of memory CD8+T cells predisposing to be multi-functional for against SARS-CoV-2 infection. We calculated several immunological features to evaluate whether *CXCR6*$^{+}$ memory CD8+T cells have a higher polyfunctionality than *CXCR6*$^{-}$ memory CD8+T cells. Compared with *CXCR6*$^{-}$ memory CD8+T cells, we found that scores of cytokine, chemokine, IFN-α/β response, T cell activation, proliferation, and migration were significantly higher among *CXCR6*$^{+}$ memory CD8+T cells ($P < 0.05$, Figure 5A-D and Supplemental Figure S19A-C). There were 158 highly-expressed genes among *CXCR6*$^{+}$ memory CD8+T cells in comparison with *CXCR6*$^{-}$ cells (FDR < 0.05, Figure 5E). These highly-expressed genes were significantly enriched in two biological pathways of cytokine-cytokine receptor interaction and inflammatory bowel (FDR < 0.05, Supplemental Figure S19D and Table S21). The chemokine signaling pathway showed a suggestive enrichment ($P < 0.05$). These highly-expressed genes contained numerous pro-inflammatory cytokine and chemokine genes, such as *CCR1*, *CCR2*, *CCR5*, *CCR6*, *CCL3L1*, *IFNGR1*, *IL18R1*, *IL23R*, *MYC*, and *TNFSF14*, which may be associated with the activation of memory CD8+T cells.

Furthermore, the cell proportion of *CXCR6*$^{+}$ memory CD8+T cells was significantly higher among both mild and moderate COVID-19 than that among normal group ($P < 0.05$), whereas the cell proportion of *CXCR6*$^{+}$ memory CD8+T cells among severe COVID-19 was remarkably lower than that among normal group ($P = 0.012$, Figure 5F). Consistently, we found that the scores of chemokine, T cell activation, and migration were increased with the increasing patient severities among *CXCR6*$^{+}$ memory CD8+T cells (Trend $P < 0.05$, Figure 5G-I), and that lower cytotoxicity score and exhaustion score were observed among moderate-to-severe patients (Trend $P < 0.05$,

Supplemental Figure S19E-F). Additionally, we found 44, 42, and 53 up-DEGs that were notably associated with mild, moderate, and severe COVID-19, and there were six significant common genes across three phases of COVID-19, including *TCF7*, *GZMH*, *RAB5IF*, *CCND2*, *BIRC6*, and *NDUFAF3* (Figure 5J-K and Supplemental Figure S19G-N). The gene of *TCF7* was an essential factor in memory CD8+T cell differentiation [61], and *GZMH* was reported to mediate antiviral activity through direct cleavage of viral substrates [62]. These 108 up-DEGs were found to be significantly enriched in 22 functional GO-terms, including Fc-gamma receptor signaling pathway, regulation of leukocyte differentiation, and activation of immune response (Figure 5L-M and Supplemental Table S22). Overall, these results indicated that *CXCR6*$^+$ memory CD8+T cells have an enhanced propensity to be multi-functional and activated T cells involved in severe COVID-19.

Elevated cellular interactions may enhance the resident to lung airway for COVID-19

To gain refined insights into *CCR1*$^+$ CD16+monocytes and *CXCR6*$^+$ memory CD8+T cells, we examined the cellular interactions among cell populations in PBMCs and BALFs according to the COVID-19 disease status using the CellChat algorithm [47]. For *CCR1*$^+$ CD16+monocytes in PBMCs, we found a notable increase in cell-to-cell interactions with other immune cells among severe patients than that in normal controls ($P < 0.05$, Figure 6A and Supplemental Figure S20). There was no statistical difference in cellular communications of *CCR1*$^-$ CD16+monocytes with other cells between normal and COVID-19 patients ($P > 0.05$, Figure 6B). Compared with normal controls, *CCR1*$^+$ CD16+monocytes showed elevated interactions with megakaryocytes, memory CD8+T cells, NK, effector CD8+T cells, and CD14+monocytes among severe patients (Supplemental Figure S20). There were 14 ligand-receptor interactions observed to be remarkably dominated among severe patients (Figure 6C), including *ANXA1-FPR1*, *ITGB2-ICAM2/CD226*, *LGALS9-CD44*, *SELPLG-SELL/SELP*, *APP-CD74*, and *THBS1-CD36/CD47*.

With regard to *CXCR6*$^{+}$ memory CD8+T cells in PBMCs, the predicted cell-to-cell interactions showed a prominent elevation with increased severities of COVID-19 ($P < 0.05$, Figure 6D). Similar to *CCR1*$^{-}$ CD16+monocytes, we observed no obvious difference of cellular interactions between normal controls and COVID-19 patients among *CXCR6*$^{-}$ memory CD8+T cells ($P > 0.05$, Figure 6E). Compared with healthy individuals, *CXCR6*$^{+}$ memory CD8+T cells demonstrated higher cellular communications with CD14+monocytes, CD34+progenitors, dendritic cells, effector CD8+T cells, naïve CD8+T cells, memory CD4+T cell, naïve CD4+T cells, NK, and megakaryocytes among severe patients (Supplemental Figure S20). There were 20 elevated cellular interactions of *CXCR6*$^{+}$ memory CD8+T cells with other immune cells among severe patients, including *ADGRE5-CD55*, *ANXA1-FPR1*, *CCL3/CCL5-CCR1*, *CD99-CD99/PILRA*, *ICAM2-(ITGAL+ITGB2)*, and *ITGB2-ICAM2/CD226* (Figure 6F). These cell adhesion molecules (*ANXA1* and *ICMA2*), cytokine binding and receptor activity genes (*CD44*, *CD36*, *CD74*, *CXCR4*, and *THBS1*), and inflammatory genes (*FPR1* and *SELL*) have been reported to be associated with COVID-19 [16, 55, 63, 64].

Among BALF cells, we also observed a remarkable increase in cellular interactions of *CCR1*$^{+}$ CD16+monocytes and *CXCR6*$^{+}$ memory CD8+T cells comparing to their corresponding negative cells ($P < 0.001$, Figure 6G-J and Supplemental Figure S21A). For example, enhanced ligand-receptor axes of *SELPLG-SELL*, *CCL5-CCR1*, *FN1-(ITGA4+ITGB1)*, *CD99-CD99*, and *APP-CD74* among *CCR1*$^{+}$ CD16+monocytes (Figure 6H), as well as *CXCL16-CXCR6*, *TNFSF14-TNFRSF14*, *ITGB2-CD226*, *CLEC2B/CLEC2C-KLRB1*, and *CCL3/CCL4-CCR5* among *CXCR6*$^{+}$ memory CD8+T cells (Figure 6J). Notably, there was a 60% increase in cellular interactions between *CCR1*$^{+}$ CD16+monocytes and epithelial cells compared with that of *CCR1*$^{-}$ CD16+monocytes (Supplemental Figure S21B). We also found a 33.33% increase in the interactions between *CXCR6*$^{+}$ memory CD8+T cells and epithelial cells compared with that of *CXCR6*$^{-}$ memory CD8+T cells (Supplemental Figure S21C), such as enhanced ligand-receptor

interactions including *TNF-TNFRSF1A*, *CXCL16-CXCR6*, and *CCL3-CCR5*. Previous studies [65, 66] have reported that the *CXCL16-CXCR6* axis modulates the localization of tissue-resident memory CD8+T cells to the lung airway. Overall, these results suggest that the increased cellular interactions with epithelial cells probably enhance the resident to the lung airway for against SARS-CoV-2 infection.

Discussion

By using large-scale genetics data, we identified eight genomic loci including three novel loci (e.g., 1p22.2, 6p21.33, and 7p11.2) that were significantly associated with severe COVID-19. Other five loci including 3p21.31, 9q34.2, 12q24.13, 19p13.3, and 21q22.11 have been reported to be involved in COVID-19 risk in previous studies [23-28]. Notably, we prioritized 34 risk genes, including potential causal genes of *CXCR6*, *CCR1*, and *ABO*, to be associated with severe COVID-19. The CXC motif chemokine receptor 6 (CXCR6), which is a G protein-coupled receptor with seven transmembrane domains, regulates the partitioning of resident memory T cells by recruiting lung tissue-resident memory CD8+T cells to airways [65]. *CCR1* gene encodes the CC motif chemokine receptor 1 (CCR1) belonging to a member of the beta chemokine receptor family. Several previous GWASs have reported genetic variants in *CCR1* are associated with COVID-19 susceptibility at a genome-wide significant level [25, 27]. For the *ABO* gene, it encodes protein relevant to the ABO blood group system. Both genetic and non-genetic studies [25, 27, 67] have showed the involvement of *ABO* gene in COVID-19 susceptibility, while the *ABO* gene encodes protein that is relevant to the ABO blood group system, and it was also notably associated with several thrombotic and coagulation-related traits including deep vein thrombosis and pulmonary heart disease, which have been reported to be risk factors and sequalae to severe COVID-19 [68, 69].

Understanding the immune responses of monocytes and memory T cells is fundamental to the rational design of innovative and effective strategies to develop better vaccines [70, 71], and contributes to reveal the pathogenesis of severe COVID-19 [29]. Our current analyses reveal that host genetic determinants have a prominent influence on the immune responses of CD16+monocytes, megakaryocytes, and memory CD8+T cells to severe COVID-19. Previous studies [11, 29, 54] showed that the influence caused by monocytes and megakaryocytes in inflammatory storms is noteworthy among severe COVID-19 patients. We found that *CCR1*$^{+}$ CD16+monocytes and *ABO*$^{+}$ megakaryocytes showed a significantly increased propensity to cause inflammatory storms among severe patients. The observations suggest highly-expressed interferon-related genes, including *S100A8*, *S100A9*, *S100A12*, *CD14*, *CXCL8*, *IGSF6*, *IRF3*, *IFI6*, *IFITM1*, and *IFITM3* among the two cell subsets contribute to exacerbate inflammation among severe patients. The inflammatory mediator of EN-RAGE encoded by *S100A12* was significantly correlated with COVID-19 [21], and *S100A8*, *S100A9*, *IRF3*, *IFI6*, *IFITM1*, and *IFITM3* have been reported to elicit autoinflammatory and autoimmune conditions in response to SARS-CoV-2 infection [10, 16, 29, 57, 58]. Double positive CD14+CD16+monocytes reported as tissue-infiltrative cells have a higher potency of antigen presentation and highly-expressed proinflammatory cytokines [72, 73]. Additionally, interferons are the mediators in several canonical host antiviral signaling to activate the expression of numerous required molecules of the early response to viral infection [74], and impaired type I interferon activity play important roles in severe COVID-19 [58]. Our findings described above suggest that *CCR1*$^{+}$ CD16+monocytes and *ABO*$^{+}$ megakaryocytes as a functional subset of myeloid cells convey higher risks to severe COVID-19.

Memory CD8+T cells could elicit improved immunological features that are critical in host protection from viral infectious [71]. After influenza virus infections, memory CD8+T cells reside in the lung for a couple of months and these resident memory T cells are necessary for effective immunity against secondary infection [75]. Among severe COVID-19 patients, we found that

635 *CXCR6*$^{+}$ memory CD8+T cells undertook several improved immunological features, including
636 higher scores of cytokine, chemokine, T cell activation, proliferation, and migration, which
637 suggests *CXCR6*$^{+}$ memory CD8+T cells potentially contribute to the protection of SARS-CoV-2
638 infection. Among these positive *CXCR6*$^{+}$ cells, numerous highly-expressed cytokine and
639 chemokine genes, including *CCR1*, *CCR2*, *IFNGR1*, and *MYC*, may work on activating memory T
640 cells. Earlier evidence indicated that MYC was rapidly but temporally induced during the early
641 stage of T cell activation [76]. The *CCR1* plays a pivotal role in the recruitment of effector immune
642 cells to the site of inflammation, and the pharmacologic inhibition of this gene may suppress
643 immune hyper-activation in severe COVID-19 [15]. Memory CD8+T cells obtained the capability
644 of transforming to effector cells by sensing inflammation from monocytes [71]. Thus, inflammatory
645 *CCR1*$^{+}$ CD16+monocytes among severe COVID-19 patients potentially accelerate the activation
646 of memory CD8+T cells.

647 Additionally, we observed a prominent decrease of the cell proportion of *CXCR6*$^{+}$ memory
648 CD8+T cells among severe patients. This decrease in peripheral blood among severe patients is
649 probably due to efflux to the site of viral infected lung tissue in answer to ongoing tissue damage.
650 Earlier studies [29, 77] have reported that functional CD8+T cell subsets manifest a notable
651 decrease in the peripheral blood of severe COVID-19 patients. Epithelium is the most vulnerable
652 tissue to be attacked by viral or microbial infection, thus the presence of resident memory CD8+T
653 cells are imperative for defending the debilitating infections for hosts [75]. In the current study, we
654 found an obvious increase in cellular interactions of *CXCR6*$^{+}$ memory CD8+T cells with
655 epitheliums. Enhanced ligand-receptor interactions including *TNF-TNFSFRSF1A*, *CXCL16-*
656 *CXCR6*, and *CCL3-CCR5* may contribute to the lung-residence of memory CD8+T cells. Previous
657 evidence demonstrated a major role for *CXCL16-CXCR6* interactions in regulating the resident of
658 virus-specific memory CD8+T cells [65, 66]. An earlier study showed a stronger interactions
659 between epithelial and immune cells among severe COVID-19 cases than that among moderate

cases [15]. We demonstrated that *CXCR6*$^{+}$ memory CD8+T cells mounted highly effective immune responses to against COVID-19, highlighting the remarkable biological plasticity in subsets of memory CD8+T cells differentiation.

The power of this study is limited by the lack of matched genetic data and scRNA-seq data in each sample for uncovering the genetic effects on immune cells for severe COVID-19. To reduce the influence of this limitation, we adopted a widely-used approach by integrating a large-scale GWAS summary statistics with enormous amount of single cell sequencing data, as referenced in previous studies [45, 78]. Based on our findings suggesting that host genetic components exert regulatory effects on immunological dysregulations for SRAS-CoV-2 infection, more studies are warranted for exploring the genetic modification of peripheral T cells to defend against lethal severe COVID-19.

Conclusions

In sum, we provide comprehensive insights that host genetic determinants are fundamental in influencing the peripheral immune responses to severe COVID-19. Both *CCR1*$^{+}$ CD16+monocytes and *ABO*$^{+}$ megakaryocytes contribute higher risk to the inflammatory storms among severe patients. *CXCR6*$^{+}$ memory CD8+T cells exhibit a notable polyfunctionality of several improved immunologic features implicated in the etiology of severe COVID-19. Further experiments to parse the molecular mechanism of these three cell subpopulations on severe COVID-19 patients are crucial for promoting personalized protective immunity.

Abbreviations

COVID-19: coronavirus disease 2019; SARS-CoV-2: severe acute respiratory syndrome coronavirus 2; GWAS: genome-wide association study; scRNA-seq: single cell RNA sequencing; PBMCs: peripheral blood mononuclear cells; BALF: bronchoalveolar lavage fluid; eQTL:

685 expression quantitative trait loci; GEO: the Gene Expression Omnibus database; WHO: the World
686 Health Organization; SNP: single nucleotide polymorphism; OR: odds ratio; MAF: minor allele
687 frequency; QQ: quantile-quantile; MAGMA: Multi-marker Analysis of GenoMic Annotation; LD:
688 linkage disequilibrium; FDR: false discovery rate; KEGG: the Kyoto Encyclopedia of Genes and
689 Genomes; PPC: the Pearson correlation coefficient; MDS: multidimensional scaling; OTG: the
690 Open Target Genetics; PPI: protein-protein interaction; up-DEG: significantly up-regulated
691 expression gene associated with COVID-19.
692
693 **Declarations**
694 **Acknowledgments**
695 We appreciate Prof. Yang Jian from Westlake University for providing helpful suggestions, and
696 our appreciation also goes to all the authors from the COVID-19 Host Genetic Consortium who
697 have deposited and shared GWAS summary data on public databases and goes to the authors who
698 publicly released the scRNA-seq datasets on PBMC and BALF with distinct COVID-19 severities.
699
700 **Authors' contributions**
701 J.S., and Y.M. conceived and designed the study. Y.M., F.Q., C.D., J.L., Y.K.H., Y.R.Z., Y.X.,
702 Y.G.Z., and Y.H.Y. contributed to management of data collection. Y.M., F.Q., C.D., Y.K.H., and
703 J.L. conducted bioinformatics analysis and data interpretation. Y.M., J.S., Z.W. and J.Q. wrote the
704 manuscripts. All authors reviewed and approved the final manuscript.
705
706 **Funding**
707 This study was funded by the National Natural Science Foundation of China (61871294 to J.S.),
708 the Scientific Research Foundation for Talents of Wenzhou Medical University (KYQD20201001
709 to Y.M.), and Science Foundation of Zhejiang Province (LR19C060001 to J.S).

Data and materials availability:

All the GWAS summary statistics used in this study can be accessed in the official websites (www.covid19hg.org/results). The GTEx eQTL data (version 8) were downloaded from Zenodo repository (https://zenodo.org/record/3518299#.Xv6Z6igzbgl). Three scRNA-seq datasets were downloaded from the GEO database (https://www.ncbi.nlm.nih.gov/gds/?term=GSE149689 and https://www.ncbi.nlm.nih.gov/gds/?term=GSE150861) and the ArrayExpress database (https://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-9357). All analysis code in the Methods is available in an online GitHub repository at https://github.com/mayunlong89/COVID19_scRNA.

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Competing interests:

The authors declare no competing interests.

Figure Legends

Figure 1. The workflow for this integrative genomic analysis. A) Combination of single cell RNA sequencing data and GWAS summary statistics on severe COVID-19 based on two independent methods. One method is regression-based polygenic model based on whole scRNA-seq profiles, and another is generalized linear regression model based on top 10% most specific

genes for each cell type. B) An increase in genetics-risk genes and cytokines for severe COVID-19. C) Cellular interaction analysis of genetics-influenced immune cell subsets with epithelial cells.

Figure 2. Risk genes and pathways associated with hospitalized COVID-19 from meta-GWAS summary data. A) Manhattan plot and quantile-quantile (QQ) plot of meta-GWAS analysis highlighting eight risk genetic loci for hospitalized COVID-19. The red horizontal line represents the genome-wide significance threshold of $P < 5 \times 10^{-8}$. The genomic inflation factor $\lambda = 1.02$. B) Nine index SNPs within eight genomic loci associated with hospitalized COVID-19. Left panel shows the P value of each index SNP, and right panel shows the odds ratio with 95% confidence interval. C) Circus plot showing the results of S-MultiXcan-based analysis. The inner ring demonstrates the 22 autosomal chromosomes (Chr1-22). In the outer ring, a circular symbol represents a specific gene and color marks the statistical significance of the gene for hospitalized COVID-19 (Red marks FDR < 0.05, orange indicates $6.96 \times 10^{-5} \leq P < 0.001$, light blue marks $0.001 \leq P \leq 0.05$, and dark blue indicates $P > 0.0$). D) PPI network of these 34 identified risk genes based on the STRING database (v11.0, https://string-db.org/). Orange ring represents druggable genes targeted by at least one known drug. E) Network module constructed by using the Jaccard distance showing the connectivity of 10 significant pathways enriched by 34 risk genes. F) Heatmap showing the results of hierarchical clustering analysis of 27 risk genes on COVID-19 severity. Seven risk genes did not expressed in the dataset #1, and the expression level of each gene was scaled. G) The proportion of highly-expressed genes among 27 risk genes in normal controls and in the three phases of COVID-19 (mild, moderate, and severe patients). Using 10,000 times of permutation analysis to calculate the significance of the observation (permuted P = 0.023). H) Plot showing an increase of the significantly enriched pathways in the network module with elevated COVID-19 severities. Orange color represents a significant enriched pathway (FDR ≤ 0.05) and gray color represents a non-significant enriched pathway (FDR > 0.05).

Figure 3. Integrative analysis identifies genetic associations between peripheral immune cells and severe COVID-19. A) Bar graph showing the results of the combination of scRNA-seq data and GWAS summary statistics on severe COVID-19 based on the RolyPoly among normal controls and patients with different severities (i.e., mild, moderate, and severe). The y-axis shows the 13 cell types, and x-axis shows mean negative log-transformation P value (-Log2(P)). Orange color indicates a cell type showing a significant association, and light blue represents there is no significant association. B) UMAP projections of peripheral immune cells colored by annotated cell types. The plot showing the region of CD16+monocytes, megakaryocytes, and memory CD8+T cells. Red dot represents positive gene expressions of *CCR1*$^{+}$, *ABO*$^{+}$, and *CXCR6*$^{+}$, and gray stands for negative cells.

Figure 4. *CCR1*$^{+}$ CD16+momocytes contributes higher risk to cytokine storms among severe COVID-19 patients. A) Boxplot showing the difference in inflammatory cytokine score between *CCR1*$^{+}$ and *CCR1*$^{-}$ CD16+ monocytes. Two-side Wilcoxon sum-rank test was used. B) Volcano plot showing differentially expressed genes between *CCR1*$^{+}$ and *CCR1*$^{-}$ CD16+ monocytes. C) Significantly enriched pathways by 351 highly-expressed genes among *CCR1*$^{+}$ CD16+ monocytes. Color legend represents the log transformed FDR value (-Log10(FDR)). D) Bar graph showing the proportion of *CCR1*$^{+}$ CD16+ monocytes among normal, mild, moderate, and severe groups. E) Boxplot showing the inflammatory cytokine score of *CCR1*$^{+}$ CD16+ monocytes among normal, mild, moderate, and severe groups. The Mann-Kendall trend analysis was used. F) Bar graph showing the differentially up-DEGs among different COVID-19 patients compared with normal controls. Namely, mild COVID-19 vs. normal, moderate COVID-19 vs. normal, and severe COVID-19 vs. normal. Venn plot on top of bar showing the overlapped up-DEGs between moderate and severe patients. G) The correlation of up-DEGs between moderate and severe patients. Pearson

correlation analysis was used to calculate the correlation coefficient and P value. H)-J) Representative up-DEGs among *CCR1*$^{+}$ CD16+ monocytes showing significantly elevated expressions with increased COVID-19 severities. H) *S100A8*, I) *S100A9*, and J) *IFITM1*. K) Disease-terms enrichment analysis on 190 up-DEGs based on the GLAD4U database. The y-axis shows -Log10(FDR), and x-axis shows the enrichment ratio.

Figure 5. Multi-functionality of *CXCR6*$^{+}$ memory CD8+T cells for severe COVID-19. A)-D) Boxplots showing the difference in (A) cytokine score, (B) chemokine score, (C) IFN-α/β response score, and (D) T cell activation score between *CXCR6*$^{+}$ and *CXCR6*$^{-}$ memory CD8+T cells. Two-side Wilcoxon sum-rank test was used. E) Volcano plot showing differentially expressed genes between *CXCR6*$^{+}$ and *CXCR6*$^{-}$ memory CD8+T cells. F) Bar graph showing the proportion of *CXCR6*$^{+}$ memory CD8+T cells among normal, mild, moderate, and severe groups. G)-I) Boxplots showing the (G) chemokine score, (H) T cell activation score, and (I) migration score of *CXCR6*$^{+}$ memory CD8+T cells among normal, mild, moderate, and severe groups. The Mann-Kendall trend analysis was used. J) Venn plot showing the overlapped up-DEGs between pairwise comparisons: mild vs. normal, moderate vs. normal, and severe vs. normal. K) Representative gene of *GZMH* among *CXCR6*$^{+}$ memory CD8+T cells showing significantly elevated expressions with increased COVID-19 severities. L) Heatmap showing up-DEGs in *CXCR6*$^{+}$ memory CD8+T cells from pairwise comparisons: mild vs. normal, moderate vs. normal, severe vs. normal. The up-DEGs listed in the green panel were from mild vs. normal, yellow panel were from moderate vs. normal, and orange panel were from severe vs. normal. M) Scatter plot showing the enriched GO biological processes by 108 up-DEGs among *CXCR6*$^{+}$ memory CD8+T cells. The x-axis shows -Log10(FDR), and y-axis shows the enrichment ratio.

Figure 6. Cell-to-cell interactions of *CCR1*$^+$ CD16+momocytes and *CXCR6*$^+$ memory CD8+T cells with other cells in PBMC and BALF. A)-B) Boxplot showing the number of cellular interactions of (A) *CCR1*$^+$ CD16+ monocytes and (B) *CCR1*$^-$ CD16+ monocytes with other immune cells in PBMC between normal controls and patients with increased COVID-19 severities. C) Predicted cellular interactions of *CCR1*$^+$ CD16+ monocytes with other immune cells in PBMC, comparing severe COVID-19 vs. normal control. D)-E) Boxplot showing the number of cellular interactions of (D) *CXCR6*$^+$ memory CD8+T cells and (E) *CXCR6*$^-$ memory CD8+T cells with other immune cells in PBMC between normal controls and patients with increased COVID-19 severities. F) Predicted cellular interactions of *CXCR6*$^+$ memory CD8+T cells with other immune cells in PBMC, comparing severe COVID-19 vs. normal control. G) Boxplot showing an increase in cellular interactions with other cells in BALF for *CCR1*$^+$ CD16+ monocytes than *CCR1*$^-$ CD16+ monocytes. H) Predicted cellular interactions with other cells in BALF, comparing *CCR1*$^+$ CD16+ monocytes with *CCR1*$^-$ CD16+ monocytes. I) Boxplot showing an increase in cellular interactions with other cells in BALF for *CXCR6*$^+$ memory CD8+T cells than *CXCR6*$^-$ memory CD8+T cells. J) Predicted cellular interactions with other cells in BALF, comparing *CXCR6*$^+$ memory CD8+T cells with *CXCR6*$^-$ memory CD8+T cells. The circular size represents the significance of each ligand-receptor axis, and color represents the communication probability.

References

1. Dong E, Du H, Gardner L: **An interactive web-based dashboard to track COVID-19 in real time.** *Lancet Infect Dis* 2020, **20:**533-534.
2. Wu Z, McGoogan JM: **Characteristics of and Important Lessons From the Coronavirus Disease 2019 (COVID-19) Outbreak in China: Summary of a Report of 72 314 Cases From the Chinese Center for Disease Control and Prevention.** *JAMA* 2020.
3. Berlin DA, Gulick RM, Martinez FJ: **Severe Covid-19.** *N Engl J Med* 2020.
4. Richardson S, Hirsch JS, Narasimhan M, Crawford JM, McGinn T, Davidson KW, Barnaby DP, Becker LB, Chelico JD, Cohen SL, et al: **Presenting Characteristics, Comorbidities, and Outcomes Among 5700 Patients Hospitalized With COVID-19 in the New York City Area.** *JAMA* 2020, **323:**2052-2059.

5. Guan WJ, Ni ZY, Hu Y, Liang WH, Ou CQ, He JX, Liu L, Shan H, Lei CL, Hui DSC, et al: **Clinical Characteristics of Coronavirus Disease 2019 in China.** *N Engl J Med* 2020, **382:**1708-1720.
6. Xu L, Ma Y, Yuan J, Zhang Y, Wang H, Zhang G, Tu C, Lu X, Li J, Xiong Y, et al: **COVID-19 Quarantine Reveals That Behavioral Changes Have an Effect on Myopia Progression.** *Ophthalmology* 2021.
7. Pedersen SF, Ho YC: **SARS-CoV-2: a storm is raging.** *J Clin Invest* 2020, **130:**2202-2205.
8. Takahashi T, Ellingson MK, Wong P, Israelow B, Lucas C, Klein J, Silva J, Mao T, Oh JE, Tokuyama M, et al: **Sex differences in immune responses that underlie COVID-19 disease outcomes.** *Nature* 2020, **588:**315-320.
9. Chen G, Wu D, Guo W, Cao Y, Huang D, Wang H, Wang T, Zhang X, Chen H, Yu H, et al: **Clinical and immunological features of severe and moderate coronavirus disease 2019.** *J Clin Invest* 2020, **130:**2620-2629.
10. Su Y, Chen D, Yuan D, Lausted C, Choi J, Dai CL, Voillet V, Duvvuri VR, Scherler K, Troisch P, et al: **Multi-Omics Resolves a Sharp Disease-State Shift between Mild and Moderate COVID-19.** *Cell* 2020, **183:**1479-1495.e1420.
11. Guo C, Li B, Ma H, Wang X, Cai P, Yu Q, Zhu L, Jin L, Jiang C, Fang J, et al: **Single-cell analysis of two severe COVID-19 patients reveals a monocyte-associated and tocilizumab-responding cytokine storm.** *Nat Commun* 2020, **11:**3924.
12. Ren X, Wen W, Fan X, Hou W, Su B, Cai P, Li J, Liu Y, Tang F, Zhang F, et al: **COVID-19 immune features revealed by a large-scale single-cell transcriptome atlas.** *Cell* 2021.
13. Wen W, Su W, Tang H, Le W, Zhang X, Zheng Y, Liu X, Xie L, Li J, Ye J, et al: **Immune cell profiling of COVID-19 patients in the recovery stage by single-cell sequencing.** *Cell Discov* 2020, **6:**31.
14. Zhang JY, Wang XM, Xing X, Xu Z, Zhang C, Song JW, Fan X, Xia P, Fu JL, Wang SY, et al: **Single-cell landscape of immunological responses in patients with COVID-19.** *Nat Immunol* 2020, **21:**1107-1118.
15. Chua RL, Lukassen S, Trump S, Hennig BP, Wendisch D, Pott F, Debnath O, Thürmann L, Kurth F, Völker MT, et al: **COVID-19 severity correlates with airway epithelium-immune cell interactions identified by single-cell analysis.** *Nat Biotechnol* 2020, **38:**970-979.
16. Silvin A, Chapuis N, Dunsmore G, Goubet AG, Dubuisson A, Derosa L, Almire C, Hénon C, Kosmider O, Droin N, et al: **Elevated Calprotectin and Abnormal Myeloid Cell Subsets Discriminate Severe from Mild COVID-19.** *Cell* 2020, **182:**1401-1418.e1418.
17. Schulte-Schrepping J, Reusch N, Paclik D, Baßler K, Schlickeiser S, Zhang B, Krämer B, Krammer T, Brumhard S, Bonaguro L, et al: **Severe COVID-19 Is Marked by a Dysregulated Myeloid Cell Compartment.** *Cell* 2020, **182:**1419-1440 e1423.
18. Lee JS, Park S, Jeong HW, Ahn JY, Choi SJ, Lee H, Choi B, Nam SK, Sa M, Kwon JS, et al: **Immunophenotyping of COVID-19 and influenza highlights the role of type I interferons in development of severe COVID-19.** *Sci Immunol* 2020, **5**.
19. Cao X: **COVID-19: immunopathology and its implications for therapy.** *Nat Rev Immunol* 2020, **20:**269-270.
20. Del Valle DM, Kim-Schulze S, Huang HH, Beckmann ND, Nirenberg S, Wang B, Lavin Y, Swartz TH, Madduri D, Stock A, et al: **An inflammatory cytokine signature predicts COVID-19 severity and survival.** *Nat Med* 2020, **26:**1636-1643.
21. Arunachalam PS, Wimmers F, Mok CKP, Perera R, Scott M, Hagan T, Sigal N, Feng Y, Bristow L, Tak-Yin Tsang O, et al: **Systems biological assessment of immunity to mild versus severe COVID-19 infection in humans.** *Science* 2020, **369:**1210-1220.

22. **The COVID-19 Host Genetics Initiative, a global initiative to elucidate the role of host genetic factors in susceptibility and severity of the SARS-CoV-2 virus pandemic.** *Eur J Hum Genet* 2020, **28:**715-718.
23. Zhou S, Butler-Laporte G, Nakanishi T, Morrison DR, Afilalo J, Afilalo M, Laurent L, Pietzner M, Kerrison N, Zhao K, et al: **A Neanderthal OAS1 isoform protects individuals of European ancestry against COVID-19 susceptibility and severity.** *Nat Med* 2021, **27:**659-667.
24. Pairo-Castineira E, Clohisey S, Klaric L, Bretherick AD, Rawlik K, Pasko D, Walker S, Parkinson N, Fourman MH, Russell CD, et al: **Genetic mechanisms of critical illness in COVID-19.** *Nature* 2021, **591:**92-98.
25. Ma Y, Huang Y, Zhao S, Yao Y, Zhang Y, Qu J, Wu N, Su J: **Integrative Genomics Analysis Reveals a 21q22.11 Locus Contributing Risk to COVID-19.** *Hum Mol Genet* 2021.
26. Gaziano L, Giambartolomei C, Pereira AC, Gaulton A, Posner DC, Swanson SA, Ho YL, Iyengar SK, Kosik NM, Vujkovic M, et al: **Actionable druggable genome-wide Mendelian randomization identifies repurposing opportunities for COVID-19.** *Nat Med* 2021, **27:**668-676.
27. Ellinghaus D, Degenhardt F, Bujanda L, Buti M, Albillos A, Invernizzi P, Fernández J, Prati D, Baselli G, Asselta R, et al: **Genomewide Association Study of Severe Covid-19 with Respiratory Failure.** *N Engl J Med* 2020, **383:**1522-1534.
28. **Mapping the human genetic architecture of COVID-19.** *Nature* 2021.
29. Ren X, Wen W, Fan X, Hou W, Su B, Cai P, Li J, Liu Y, Tang F, Zhang F, et al: **COVID-19 immune features revealed by a large-scale single-cell transcriptome atlas.** *Cell* 2021, **184:**1895-1913.e1819.
30. Butler A, Hoffman P, Smibert P, Papalexi E, Satija R: **Integrating single-cell transcriptomic data across different conditions, technologies, and species.** *Nat Biotechnol* 2018, **36:**411-420.
31. Waltman L, van Eck NJ: **A smart local moving algorithm for large-scale modularity-based community detection.** *The European Physical Journal B* 2013, **86:**471.
32. Korsunsky I, Millard N, Fan J, Slowikowski K, Zhang F, Wei K, Baglaenko Y, Brenner M, Loh PR, Raychaudhuri S: **Fast, sensitive and accurate integration of single-cell data with Harmony.** *Nat Methods* 2019, **16:**1289-1296.
33. Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA, Abecasis GR: **A global reference for human genetic variation.** *Nature* 2015, **526:**68-74.
34. Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, Boehnke M, Abecasis GR, Willer CJ: **LocusZoom: regional visualization of genome-wide association scan results.** *Bioinformatics* 2010, **26:**2336-2337.
35. de Leeuw CA, Mooij JM, Heskes T, Posthuma D: **MAGMA: generalized gene-set analysis of GWAS data.** *PLoS Comput Biol* 2015, **11:**e1004219.
36. Wang J, Duncan D, Shi Z, Zhang B: **WEB-based GEne SeT AnaLysis Toolkit (WebGestalt): update 2013.** *Nucleic Acids Res* 2013, **41:**W77-83.
37. Kanehisa M, Goto S: **KEGG: kyoto encyclopedia of genes and genomes.** *Nucleic Acids Res* 2000, **28:**27-30.
38. Ma Y, Li J, Xu Y, Wang Y, Yao Y, Liu Q, Wang M, Zhao X, Fan R, Chen J, et al: **Identification of 34 genes conferring genetic and pharmacological risk for the comorbidity of schizophrenia and smoking behaviors.** *Aging (Albany NY)* 2020, **12:**2169-2225.
39. Barbeira AN, Dickinson SP, Bonazzola R, Zheng J, Wheeler HE, Torres JM, Torstenson ES, Shah KP, Garcia T, Edwards TL, et al: **Exploring the phenotypic consequences of**

tissue specific gene expression variation inferred from GWAS summary statistics. *Nat Commun* 2018, **9:**1825.

40. Barbeira AN, Pividori M, Zheng J, Wheeler HE, Nicolae DL, Im HK: **Integrating predicted transcriptome from multiple tissues improves association detection.** *PLoS Genet* 2019, **15:**e1007889.
41. Ma X, Wang P, Xu G, Yu F, Ma Y: **Integrative genomics analysis of various omics data and networks identify risk genes and variants vulnerable to childhood-onset asthma.** *BMC Med Genomics* 2020, **13:**123.
42. Xu M, Li J, Xiao Z, Lou J, Pan X, Ma Y: **Integrative genomics analysis identifies promising SNPs and genes implicated in tuberculosis risk based on multiple omics datasets.** *Aging (Albany NY)* 2020, **12:**19173-19220.
43. von Mering C, Huynen M, Jaeggi D, Schmidt S, Bork P, Snel B: **STRING: a database of predicted functional associations between proteins.** *Nucleic Acids Res* 2003, **31:**258-261.
44. Szklarczyk D, Santos A, von Mering C, Jensen LJ, Bork P, Kuhn M: **STITCH 5: augmenting protein-chemical interaction networks with tissue and affinity data.** *Nucleic Acids Res* 2016, **44:**D380-384.
45. Calderon D, Bhaskar A, Knowles DA, Golan D, Raj T, Fu AQ, Pritchard JK: **Inferring Relevant Cell Types for Complex Traits by Using Single-Cell Gene Expression.** *Am J Hum Genet* 2017, **101:**686-699.
46. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ: **Second-generation PLINK: rising to the challenge of larger and richer datasets.** *Gigascience* 2015, **4:**7.
47. Jin S, Guerrero-Juarez CF, Zhang L, Chang I, Ramos R, Kuan CH, Myung P, Plikus MV, Nie Q: **Inference and analysis of cell-cell communication using CellChat.** *Nat Commun* 2021, **12:**1088.
48. Ghoussaini M, Mountjoy E, Carmona M, Peat G, Schmidt EM, Hercules A, Fumis L, Miranda A, Carvalho-Silva D, Buniello A, et al: **Open Targets Genetics: systematic identification of trait-associated genes using large-scale genetics and functional genomics.** *Nucleic Acids Res* 2021, **49:**D1311-d1320.
49. Battle A, Brown CD, Engelhardt BE, Montgomery SB: **Genetic effects on gene expression across human tissues.** *Nature* 2017, **550:**204-213.
50. Wang Q, Chen R, Cheng F, Wei Q, Ji Y, Yang H, Zhong X, Tao R, Wen Z, Sutcliffe JS, et al: **A Bayesian framework that integrates multi-omics data and gene networks predicts risk genes from schizophrenia GWAS data.** *Nat Neurosci* 2019, **22:**691-699.
51. Ma Y, Li MD: **Establishment of a Strong Link Between Smoking and Cancer Pathogenesis through DNA Methylation Analysis.** *Sci Rep* 2017, **7:**1811.
52. Auwul MR, Rahman MR, Gov E, Shahjaman M, Moni MA: **Bioinformatics and machine learning approach identifies potential drug targets and pathways in COVID-19.** *Brief Bioinform* 2021.
53. More SA, Patil AS, Sakle NS, Mokale SN: **Network analysis and molecular mapping for SARS-CoV-2 to reveal drug targets and repurposing of clinically developed drugs.** *Virology* 2021, **555:**10-18.
54. Manne BK, Denorme F, Middleton EA, Portier I, Rowley JW, Stubben C, Petrey AC, Tolley ND, Guo L, Cody M, et al: **Platelet gene expression and function in patients with COVID-19.** *Blood* 2020, **136:**1317-1329.
55. Shaath H, Vishnubalaji R, Elkord E, Alajez NM: **Single-Cell Transcriptome Analysis Highlights a Role for Neutrophils and Inflammatory Macrophages in the Pathogenesis of Severe COVID-19.** *Cells* 2020, **9**.
56. Rydyznski Moderbacher C, Ramirez SI, Dan JM, Grifoni A, Hastie KM, Weiskopf D, Belanger S, Abbott RK, Kim C, Choi J, et al: **Antigen-Specific Adaptive Immunity to**

SARS-CoV-2 in Acute COVID-19 and Associations with Age and Disease Severity. *Cell* 2020, **183:**996-1012.e1019.

57. King KR, Aguirre AD, Ye YX, Sun Y, Roh JD, Ng RP, Jr., Kohler RH, Arlauckas SP, Iwamoto Y, Savol A, et al: **IRF3 and type I interferons fuel a fatal response to myocardial infarction.** *Nat Med* 2017, **23:**1481-1487.
58. Hadjadj J, Yatim N, Barnabei L, Corneau A, Boussier J, Smith N, P é é H, Charbit B, Bondet V, Chenevier-Gobeaux C, et al: **Impaired type I interferon activity and inflammatory responses in severe COVID-19 patients.** *Science* 2020, **369:**718-724.
59. Lippi G, Plebani M, Henry BM: **Thrombocytopenia is associated with severe coronavirus disease 2019 (COVID-19) infections: A meta-analysis.** *Clin Chim Acta* 2020, **506:**145-148.
60. Ma C, Cheung AF, Chodon T, Koya RC, Wu Z, Ng C, Avramis E, Cochran AJ, Witte ON, Baltimore D, et al: **Multifunctional T-cell analyses to study response and progression in adoptive cell transfer immunotherapy.** *Cancer Discov* 2013, **3:**418-429.
61. Akondy RS, Fitch M, Edupuganti S, Yang S, Kissick HT, Li KW, Youngblood BA, Abdelsamed HA, McGuire DJ, Cohen KW, et al: **Origin and differentiation of human memory CD8 T cells after vaccination.** *Nature* 2017, **552:**362-367.
62. Andrade F, Fellows E, Jenne DE, Rosen A, Young CS: **Granzyme H destroys the function of critical adenoviral proteins required for viral DNA replication and granzyme B inhibition.** *Embo j* 2007, **26:**2148-2157.
63. Li Y, Hou G, Zhou H, Wang Y, Tun HM, Zhu A, Zhao J, Xiao F, Lin S, Liu D, et al: **Multi-platform omics analysis reveals molecular signature for COVID-19 pathogenesis, prognosis and drug target discovery.** *Signal Transduct Target Ther* 2021, **6:**155.
64. Bruchez A, Sha K, Johnson J, Chen L, Stefani C, McConnell H, Gaucherand L, Prins R, Matreyek KA, Hume AJ, et al: **MHC class II transactivator CIITA induces cell resistance to Ebola virus and SARS-like coronaviruses.** *Science* 2020, **370:**241-247.
65. Wein AN, McMaster SR, Takamura S, Dunbar PR, Cartwright EK, Hayward SL, McManus DT, Shimaoka T, Ueha S, Tsukui T, et al: **CXCR6 regulates localization of tissue-resident memory CD8 T cells to the airways.** *J Exp Med* 2019, **216:**2748-2762.
66. Takamura S, Kato S, Motozono C, Shimaoka T, Ueha S, Matsuo K, Miyauchi K, Masumoto T, Katsushima A, Nakayama T, et al: **Interstitial-resident memory CD8(+) T cells sustain frontline epithelial memory in the lung.** *J Exp Med* 2019, **216:**2736-2747.
67. Zhao J, Yang Y, Huang H, Li D, Gu D, Lu X, Zhang Z, Liu L, Liu T, Liu Y, et al: **Relationship between the ABO Blood Group and the COVID-19 Susceptibility.** *Clin Infect Dis* 2020.
68. Klok FA, Kruip M, van der Meer NJM, Arbous MS, Gommers D, Kant KM, Kaptein FHJ, van Paassen J, Stals MAM, Huisman MV, Endeman H: **Incidence of thrombotic complications in critically ill ICU patients with COVID-19.** *Thromb Res* 2020, **191:**145-147.
69. Grillet F, Behr J, Calame P, Aubry S, Delabrousse E: **Acute Pulmonary Embolism Associated with COVID-19 Pneumonia Detected with Pulmonary CT Angiography.** *Radiology* 2020, **296:**E186-e188.
70. Poran A, Harjanto D, Malloy M, Arieta CM, Rothenberg DA, Lenkala D, van Buuren MM, Addona TA, Rooney MS, Srinivasan L, Gaynor RB: **Sequence-based prediction of SARS-CoV-2 vaccine targets using a mass spectrometry-based bioinformatics predictor identifies immunogenic T cell epitopes.** *Genome Med* 2020, **12:**70.
71. Soudja SM, Ruiz AL, Marie JC, Lauvau G: **Inflammatory monocytes activate memory CD8(+) T and innate NK lymphocytes independent of cognate antigen during microbial pathogen invasion.** *Immunity* 2012, **37:**549-562.

72. Ziegler-Heitbrock L: **The CD14+ CD16+ blood monocytes: their role in infection and inflammation.** *J Leukoc Biol* 2007, **81:**584-592.
73. Kawanaka N, Yamamura M, Aita T, Morita Y, Okamoto A, Kawashima M, Iwahashi M, Ueno A, Ohmoto Y, Makino H: **CD14+,CD16+ blood monocytes and joint inflammation in rheumatoid arthritis.** *Arthritis Rheum* 2002, **46:**2578-2586.
74. Hambleton S, Goodbourn S, Young DF, Dickinson P, Mohamad SM, Valappil M, McGovern N, Cant AJ, Hackett SJ, Ghazal P, et al: **STAT2 deficiency and susceptibility to viral illness in humans.** *Proc Natl Acad Sci U S A* 2013, **110:**3053-3058.
75. Samji T, Khanna KM: **Understanding memory CD8(+) T cells.** *Immunol Lett* 2017, **185:**32-39.
76. Nie Z, Hu G, Wei G, Cui K, Yamane A, Resch W, Wang R, Green DR, Tessarollo L, Casellas R, et al: **c-Myc is a universal amplifier of expressed genes in lymphocytes and embryonic stem cells.** *Cell* 2012, **151:**68-79.
77. Jouan Y, Guillon A, Gonzalez L, Perez Y, Boisseau C, Ehrmann S, Ferreira M, Daix T, Jeannet R, Fran çois B, et al: **Phenotypical and functional alteration of unconventional T cells in severe COVID-19 patients.** *J Exp Med* 2020, **217**.
78. Finucane HK, Reshef YA, Anttila V, Slowikowski K, Gusev A, Byrnes A, Gazal S, Loh PR, Lareau C, Shoresh N, et al: **Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types.** *Nat Genet* 2018, **50:**621-629.

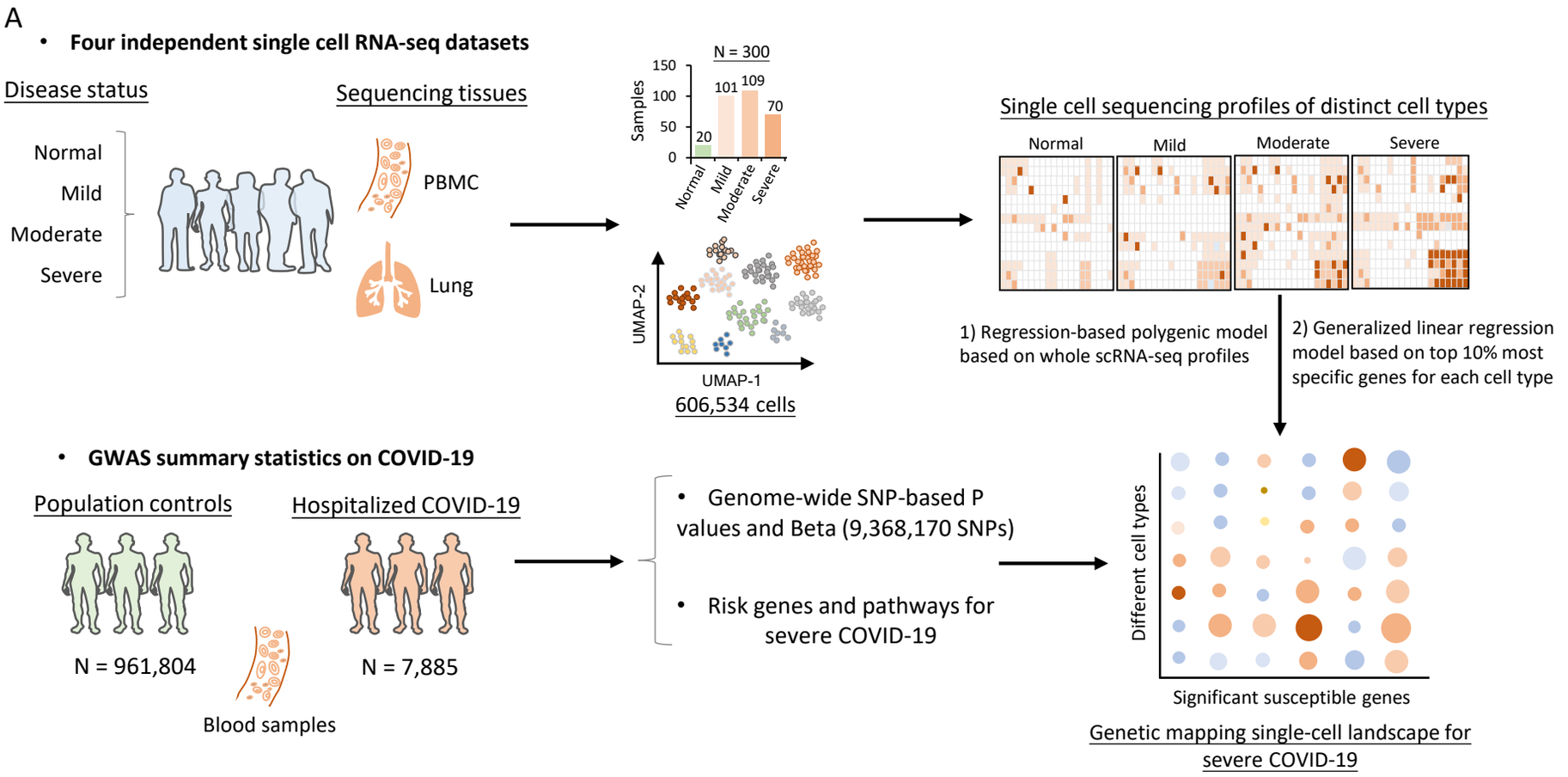

A
Four independent single cell RNA-seq datasets
Disease status
Normal
Mild
Moderate
Severe
Sequencing tissues
PBMC
Lung
N = 300
Samples
150
100
50
0
20
101
109
70
Normal
Mild
Moderate
Severe
UMAP-2
UMAP-1
606,534 cells
Single cell sequencing profiles of distinct cell types
Normal
Mild
Moderate
Severe
1) Regression-based polygenic model based on whole scRNA-seq profiles
2) Generalized linear regression model based on top 10% most specific genes for each cell type
GWAS summary statistics on COVID-19
Population controls
N = 961,804
Hospitalized COVID-19
N = 7,885
Blood samples
Genome-wide SNP-based P values and Beta (9,368,170 SNPs)
Risk genes and pathways for severe COVID-19
Different cell types
Significant susceptible genes
Genetic mapping single-cell landscape for severe COVID-19
B
An increase in genetics-risk genes and cytokines for severe COVID-19
Cytokine-cytokine receptor interaction
Chemokine signaling pathway
Genetics-related risk pathways
S100A8, S100A9, S100A12, IFITM1
Inflammatory and cytokine genes
Immune responses
Genetics risk genes
COVID-19 severity
C
Genetics-influenced immune cell subpopulations for severe COVID-19
CCR1+ CD16+monocytes
ABO+ megakaryocytes
CXCR6+ memory CD8+T cells
Interactions
Epithelial cells
Elevated cell-to-cell interactions enhance the resident to lung airway

Figure 1

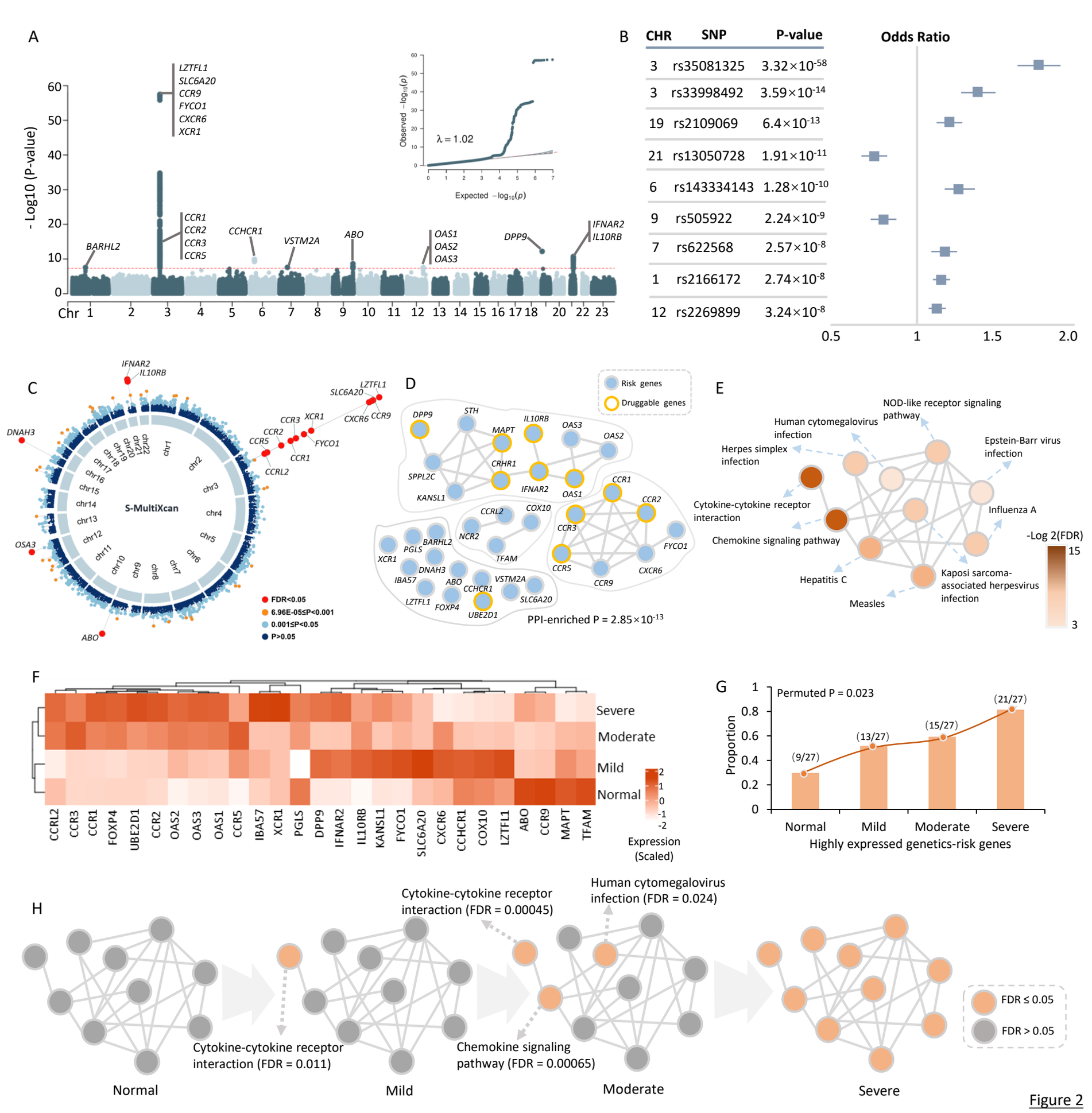
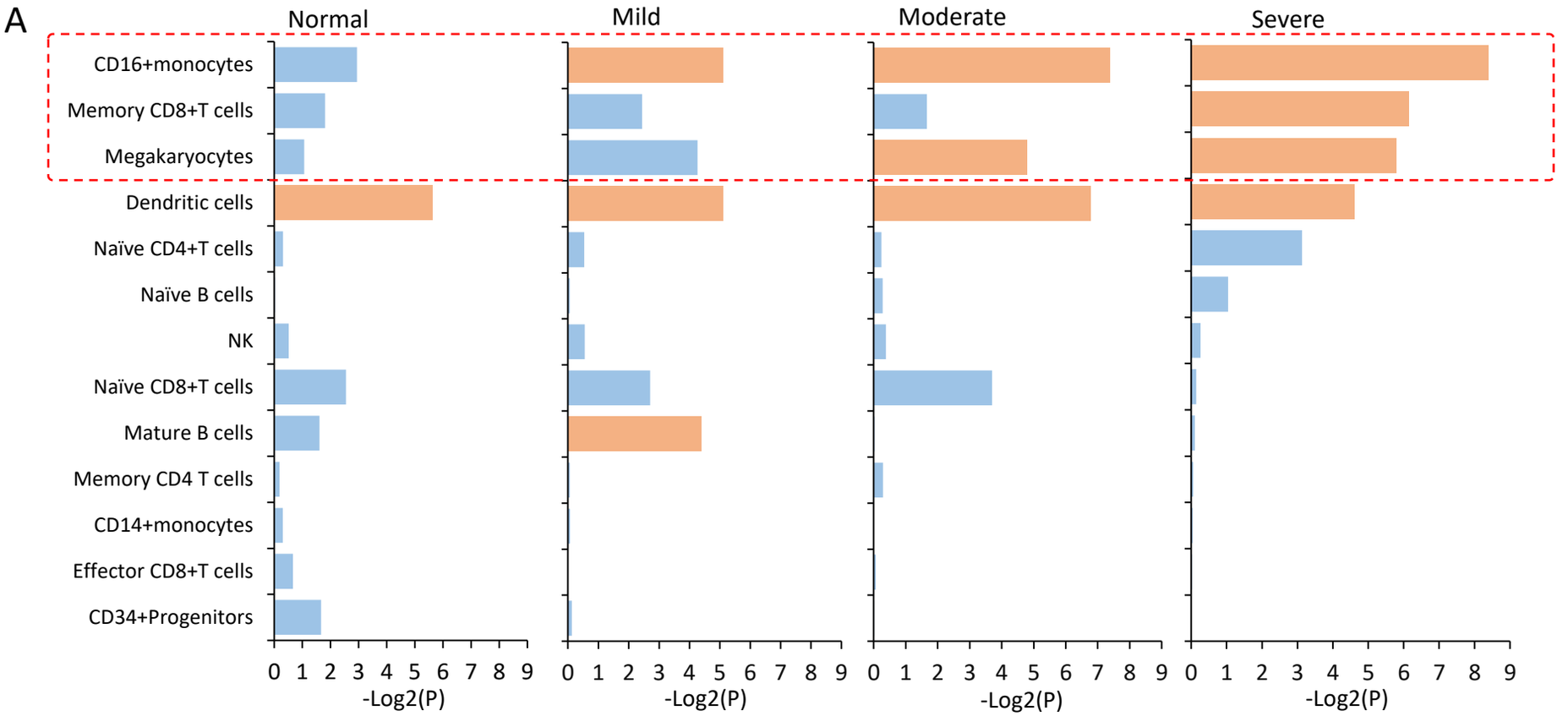

A
- Log10 (P-value)
Chr
LZTFL1
SLC6A20
CCR9
FYCO1
CXCR6
XCR1
CCR1
CCR2
CCR3
CCR5
BARHL2
CCHCR1
VSTM2A
ABO
OAS1
OAS2
OAS3
DPP9
IFNAR2
IL10RB
λ = 1.02
Observed −log10(p)
Expected −log10(p)
B
CHR SNP P-value
3 rs35081325 3.32×10-58
3 rs33998492 3.59×10-14
19 rs2109069 6.4×10-13
21 rs13050728 1.91×10-11
6 rs143334143 1.28×10-10
9 rs505922 2.24×10-9
7 rs622568 2.57×10-8
1 rs2166172 2.74×10-8
12 rs2269899 3.24×10-8
Odds Ratio
C
S-MultiXcan
FDR<0.05
6.96E-05≤P<0.001
0.001≤P<0.05
P>0.05
D
Risk genes
Druggable genes
PPI-enriched P = 2.85×10-13
E
NOD-like receptor signaling pathway
Human cytomegalovirus infection
Herpes simplex infection
Epstein-Barr virus infection
Cytokine-cytokine receptor interaction
Influenza A
Chemokine signaling pathway
Hepatitis C
Kaposi sarcoma-associated herpesvirus infection
Measles
-Log 2(FDR)
F
Severe
Moderate
Mild
Normal
Expression (Scaled)
G
Permuted P = 0.023
Proportion
(9/27)
(13/27)
(15/27)
(21/27)
Normal
Mild
Moderate
Severe
Highly expressed genetics-risk genes
H
Cytokine-cytokine receptor interaction (FDR = 0.00045)
Human cytomegalovirus infection (FDR = 0.024)
Cytokine-cytokine receptor interaction (FDR = 0.011)
Chemokine signaling pathway (FDR = 0.00065)
Normal
Mild
Moderate
Severe
FDR ≤ 0.05
FDR > 0.05

Figure 2


A
Normal
Mild
Moderate
Severe
CD16+monocytes
Memory CD8+T cells
Megakaryocytes
Dendritic cells
Naïve CD4+T cells
Naïve B cells
NK
Naïve CD8+T cells
Mature B cells
Memory CD4 T cells
CD14+monocytes
Effector CD8+T cells
CD34+Progenitors
0 1 2 3 4 5 6 7 8 9
-Log2(P)

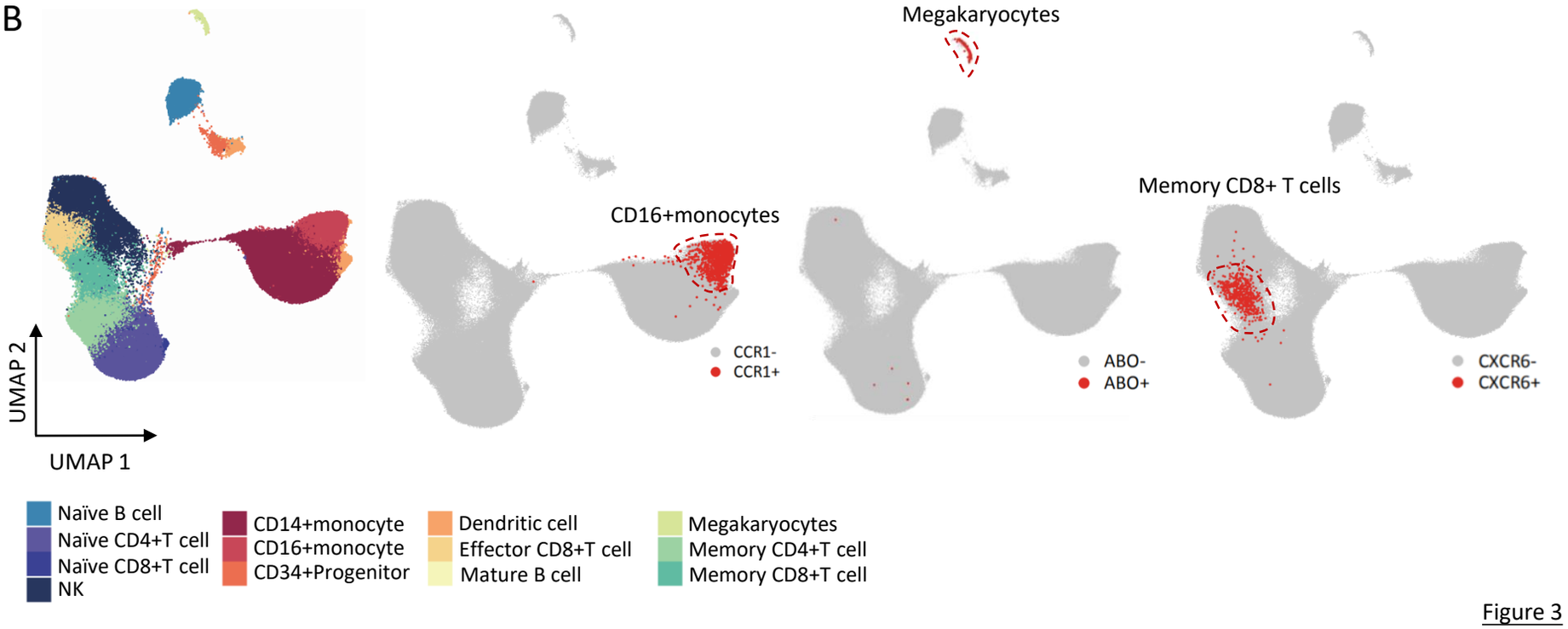

B
Megakaryocytes
CD16+monocytes
Memory CD8+ T cells
CCR1-
CCR1+
ABO-
ABO+
CXCR6-
CXCR6+
UMAP 2
UMAP 1
Naïve B cell
Naïve CD4+T cell
Naïve CD8+T cell
NK
CD14+monocyte
CD16+monocyte
CD34+Progenitor
Dendritic cell
Effector CD8+T cell
Mature B cell
Megakaryocytes
Memory CD4+T cell
Memory CD8+T cell

Figure 3

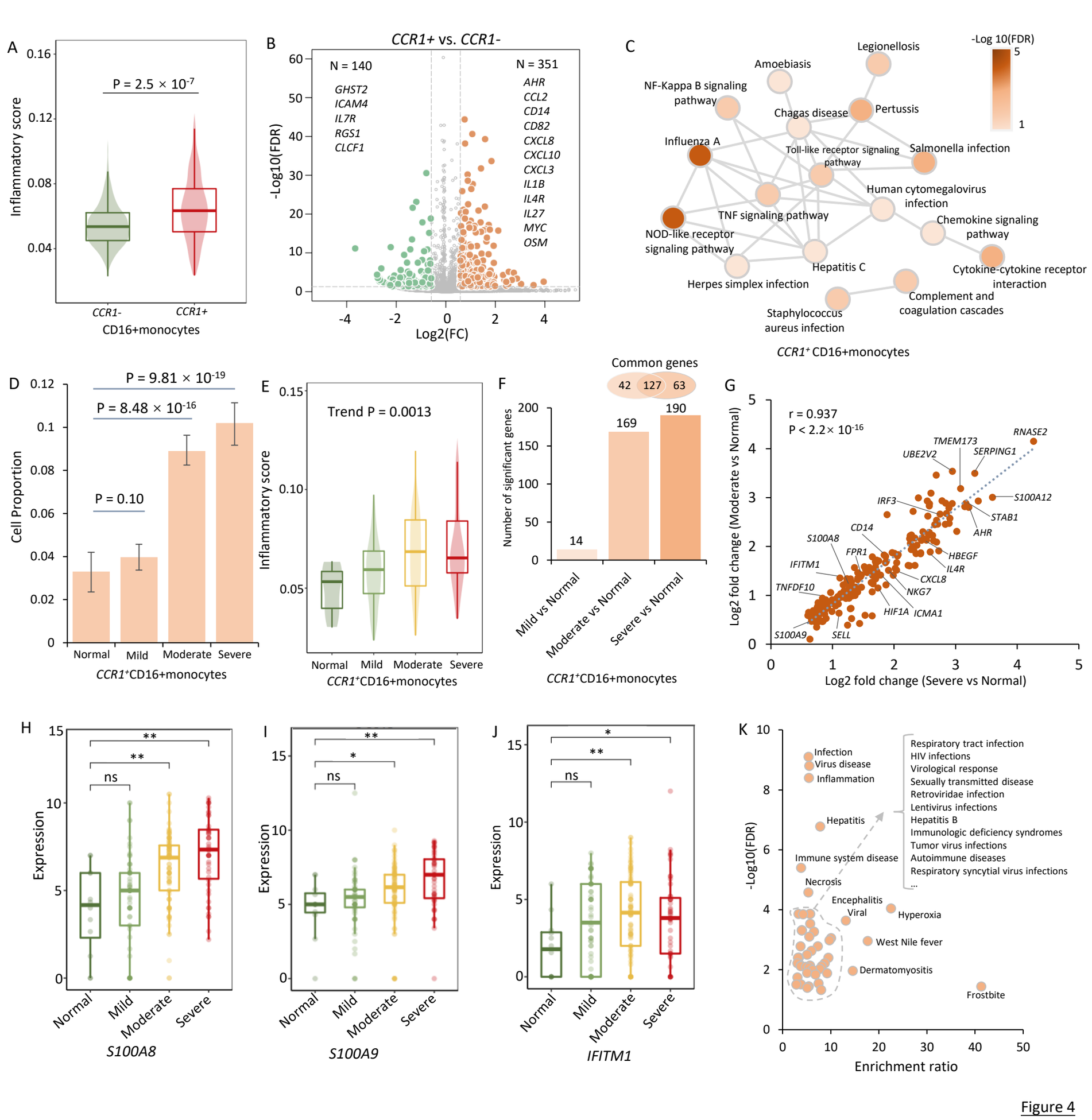

A
P = 2.5 × 10-7
Inflammatory score
CCR1-
CCR1+
CD16+monocytes
B
CCR1+ vs. CCR1-
N = 140
GHST2
ICAM4
IL7R
RGS1
CLCF1
N = 351
AHR
CCL2
CD14
CD82
CXCL8
CXCL10
CXCL3
IL1B
IL4R
IL27
MYC
OSM
-Log10(FDR)
Log2(FC)
C
-Log 10(FDR)
Legionellosis
Amoebiasis
NF-Kappa B signaling pathway
Chagas disease
Pertussis
Influenza A
Toll-like receptor signaling pathway
Salmonella infection
Human cytomegalovirus infection
TNF signaling pathway
Chemokine signaling pathway
NOD-like receptor signaling pathway
Hepatitis C
Cytokine-cytokine receptor interaction
Herpes simplex infection
Complement and coagulation cascades
Staphylococcus aureus infection
CCR1+ CD16+monocytes
D
P = 9.81 × 10-19
P = 8.48 × 10-16
P = 0.10
Cell Proportion
Normal
Mild
Moderate
Severe
CCR1+CD16+monocytes
E
Trend P = 0.0013
Inflammatory score
Normal
Mild
Moderate
Severe
CCR1+CD16+monocytes
F
Common genes
42
127
63
Number of significant genes
14
169
190
Mild vs Normal
Moderate vs Normal
Severe vs Normal
CCR1+CD16+monocytes
G
r = 0.937
P < 2.2× 10-16
RNASE2
TMEM173
UBE2V2
SERPING1
S100A12
IRF3
STAB1
AHR
CD14
S100A8
FPR1
HBEGF
IL4R
IFITM1
CXCL8
TNFDF10
NKG7
HIF1A
ICMA1
S100A9
SELL
Log2 fold change (Moderate vs Normal)
Log2 fold change (Severe vs Normal)
H
ns
**
**
Expression
Normal
Mild
Moderate
Severe
S100A8
I
ns
*
**
Expression
Normal
Mild
Moderate
Severe
S100A9
J
ns
**
*
Expression
Normal
Mild
Moderate
Severe
IFITM1
K
Infection
Virus disease
Inflammation
Hepatitis
Immune system disease
Necrosis
Encephalitis
Viral
Hyperoxia
West Nile fever
Dermatomyositis
Frostbite
Respiratory tract infection
HIV infections
Virological response
Sexually transmitted disease
Retroviridae infection
Lentivirus infections
Hepatitis B
Immunologic deficiency syndromes
Tumor virus infections
Autoimmune diseases
Respiratory syncytial virus infections
...
-Log10(FDR)
Enrichment ratio

Figure 4

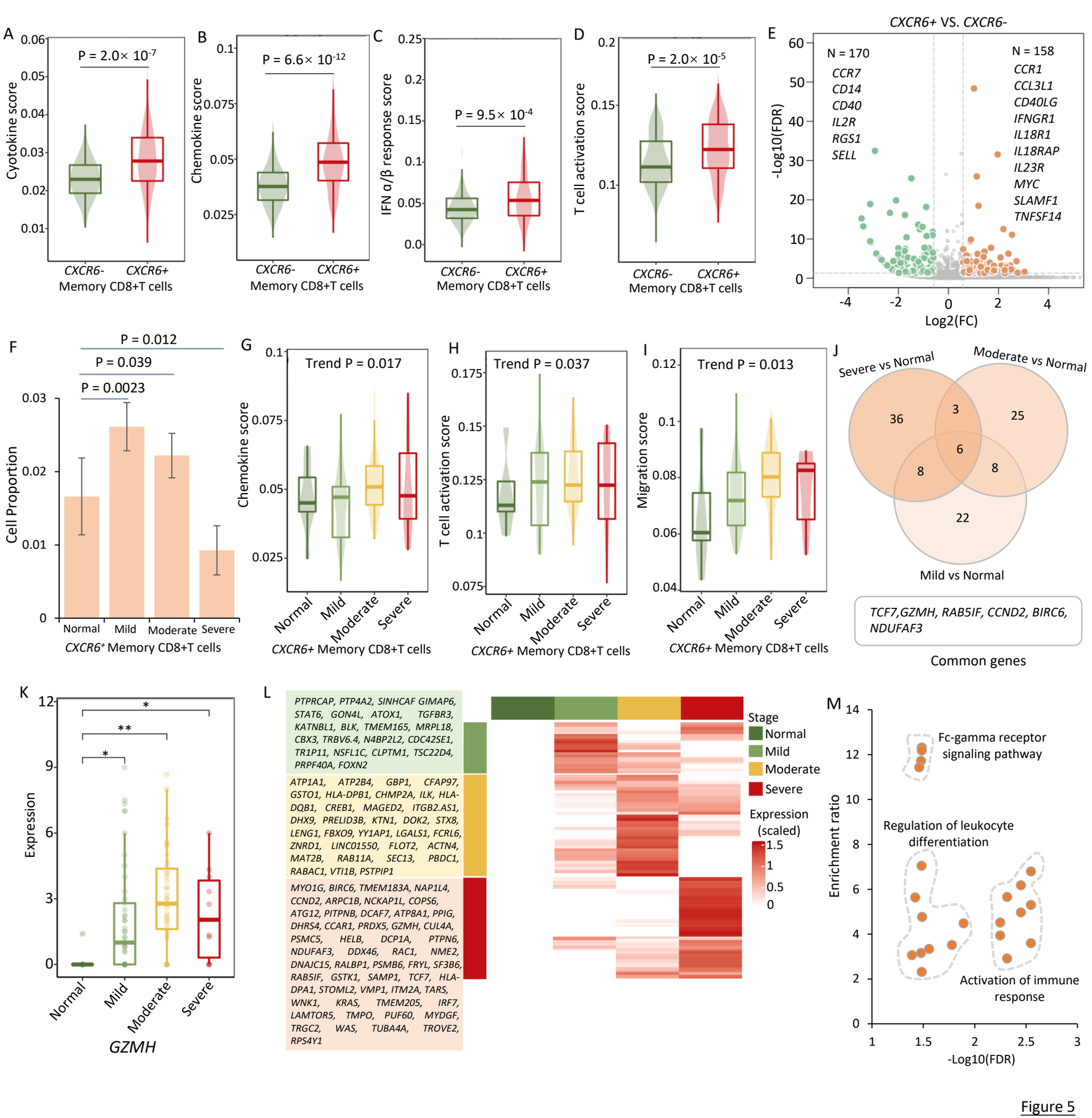

A
Cytokine score
P = 2.0× 10-7
CXCR6-
CXCR6+
Memory CD8+T cells
B
Chemokine score
P = 6.6× 10-12
C
IFN α/β response score
P = 9.5× 10-4
D
T cell activation score
P = 2.0× 10-5
E
CXCR6+ VS. CXCR6-
N = 170
CCR7
CD14
CD40
IL2R
RGS1
SELL
N = 158
CCR1
CCL3L1
CD40LG
IFNGR1
IL18R1
IL18RAP
IL23R
MYC
SLAMF1
TNFSF14
-Log10(FDR)
Log2(FC)
F
P = 0.012
P = 0.039
P = 0.0023
Cell Proportion
Normal
Mild
Moderate
Severe
CXCR6+ Memory CD8+T cells
G
Trend P = 0.017
Chemokine score
CXCR6+ Memory CD8+T cells
H
Trend P = 0.037
T cell activation score
CXCR6+ Memory CD8+T cells
I
Trend P = 0.013
Migration score
CXCR6+ Memory CD8+T cells
J
Severe vs Normal
Moderate vs Normal
Mild vs Normal
36
3
25
6
8
8
22
TCF7,GZMH, RAB5IF, CCND2, BIRC6, NDUFAF3
Common genes
K
Expression
GZMH
L
PTPRCAP, PTP4A2, SINHCAF GIMAP6, STAT6, GON4L, ATOX1, TGFBR3, KATNBL1, BLK, TMEM165, MRPL18, CBX3, TRBV6.4, N4BP2L2, CDC42SE1, TR1P11, NSFL1C, CLPTM1, TSC22D4, PRPF40A, FOXN2
ATP1A1, ATP2B4, GBP1, CFAP97, GSTO1, HLA-DPB1, CHMP2A, ILK, HLA-DQB1, CREB1, MAGED2, ITGB2.AS1, DHX9, PRELID3B, KTN1, DOK2, STX8, LENG1, FBXO9, YY1AP1, LGALS1, FCRL6, ZNRD1, LINC01550, FLOT2, ACTN4, MAT2B, RAB11A, SEC13, PBDC1, RABAC1, VTI1B, PSTPIP1
MYO1G, BIRC6, TMEM183A, NAP1L4, CCND2, ARPC1B, NCKAP1L, COPS6, ATG12, PITPNB, DCAF7, ATP8A1, PPIG, DHRS4, CCAR1, PRDX5, GZMH, CUL4A, PSMC5, HELB, DCP1A, PTPN6, NDUFAF3, DDX46, RAC1, NME2, DNAJC15, RALBP1, PSMB6, FRYL, SF3B6, RAB5IF, GSTK1, SAMP1, TCF7, HLA-DPA1, STOML2, VMP1, ITM2A, TARS, WNK1, KRAS, TMEM205, IRF7, LAMTOR5, TMPO, PUF60, MYDGF, TRGC2, WAS, TUBA4A, TROVE2, RPS4Y1
Stage
Normal
Mild
Moderate
Severe
Expression (scaled)
M
Fc-gamma receptor signaling pathway
Regulation of leukocyte differentiation
Activation of immune response
Enrichment ratio
-Log10(FDR)

Figure 5

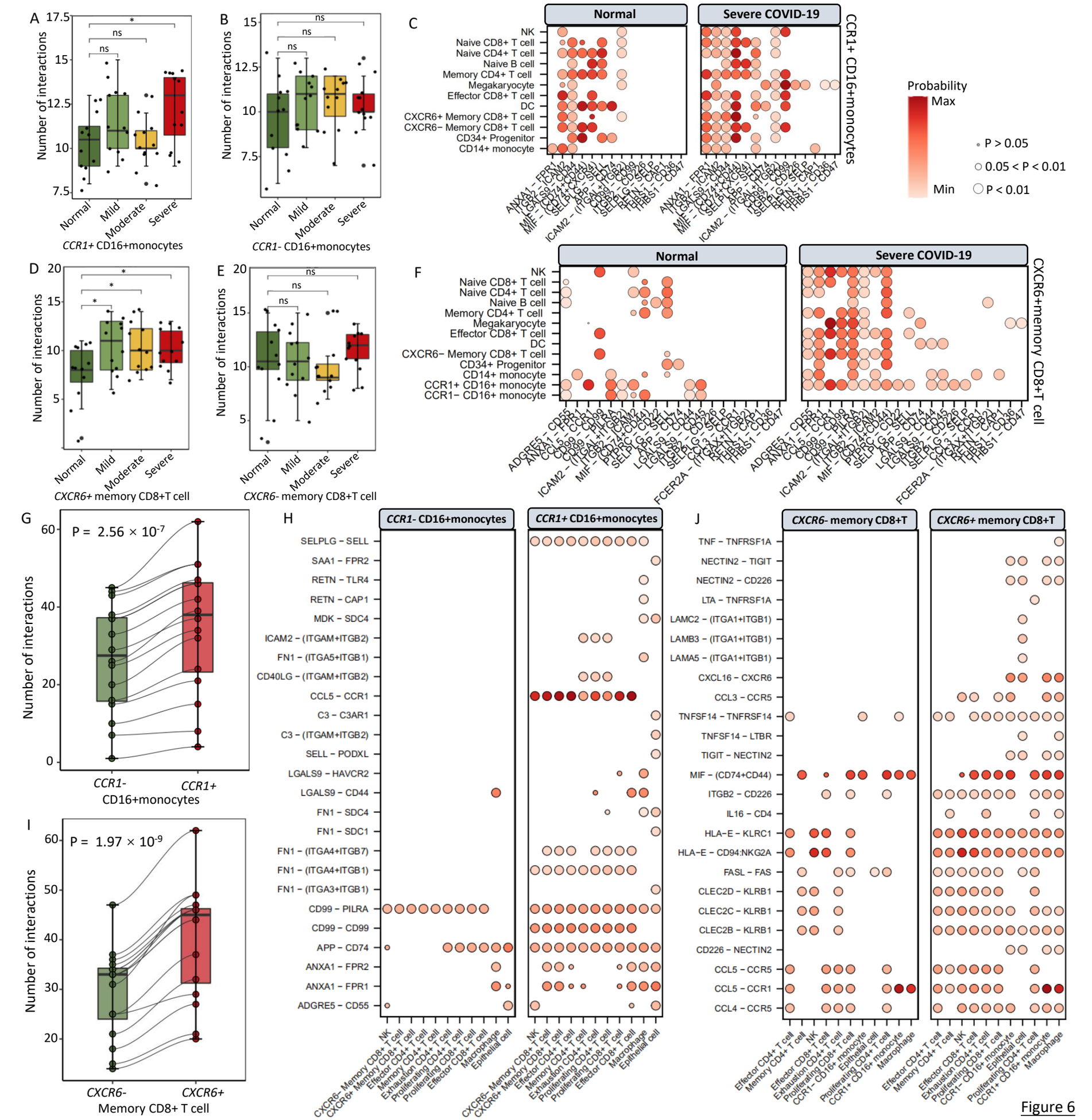

A
B
C
D
E
F
G
H
I
J
Number of interactions
Normal
Mild
Moderate
Severe
ns
*
CCR1+ CD16+monocytes
CCR1- CD16+monocytes
CXCR6+ memory CD8+T cell
CXCR6- memory CD8+T cell
Severe COVID-19
Probability
Max
Min
P > 0.05
0.05 < P < 0.01
P < 0.01
NK
Naive CD8+ T cell
Naive CD4+ T cell
Naive B cell
Memory CD4+ T cell
Megakaryocyte
Effector CD8+ T cell
DC
CXCR6+ Memory CD8+ T cell
CXCR6− Memory CD8+ T cell
CD34+ Progenitor
CD14+ monocyte
CCR1+ CD16+ monocyte
CCR1− CD16+ monocyte
P = 2.56 × 10-7
P = 1.97 × 10-9
CCR1-
CCR1+
CD16+monocytes
CXCR6-
CXCR6+
Memory CD8+ T cell
CCR1- CD16+monocytes
CCR1+ CD16+monocytes
CXCR6- memory CD8+T
CXCR6+ memory CD8+T
SELPLG – SELL
SAA1 – FPR2
RETN – TLR4
RETN – CAP1
MDK – SDC4
ICAM2 – (ITGAM+ITGB2)
FN1 – (ITGA5+ITGB1)
CD40LG – (ITGAM+ITGB2)
CCL5 – CCR1
C3 – C3AR1
C3 – (ITGAM+ITGB2)
SELL – PODXL
LGALS9 – HAVCR2
LGALS9 – CD44
FN1 – SDC4
FN1 – SDC1
FN1 – (ITGA4+ITGB7)
FN1 – (ITGA4+ITGB1)
FN1 – (ITGA3+ITGB1)
CD99 – PILRA
CD99 – CD99
APP – CD74
ANXA1 – FPR2
ANXA1 – FPR1
ADGRE5 – CD55
TNF – TNFRSF1A
NECTIN2 – TIGIT
NECTIN2 – CD226
LTA – TNFRSF1A
LAMC2 – (ITGA1+ITGB1)
LAMB3 – (ITGA1+ITGB1)
LAMA5 – (ITGA1+ITGB1)
CXCL16 – CXCR6
CCL3 – CCR5
TNFSF14 – TNFRSF14
TNFSF14 – LTBR
TIGIT – NECTIN2
MIF – (CD74+CD44)
ITGB2 – CD226
IL16 – CD4
HLA-E – KLRC1
HLA-E – CD94:NKG2A
FASL – FAS
CLEC2D – KLRB1
CLEC2C – KLRB1
CLEC2B – KLRB1
CD226 – NECTIN2
CCL5 – CCR5
CCL5 – CCR1
CCL4 – CCR5
Effector CD4+ T cell
Memory CD4+ T cell
Effector CD8+ T cell
Exhaustion CD4+ T cell
Proliferating CD4+ T cell
Proliferating CD8+ T cell
CCR1- CD16+ monocyte
CCR1+ CD16+ monocyte
Epithelial cell
Macrophage

Figure 6