Cell Genomics, Volume *3*

Supplemental information

Altered and allele-specific open chromatin landscape reveals epigenetic and genetic regulators of innate immunity in COVID-19

Bowen Zhang, Zhenhua Zhang, Valerie A.C.M. Koeken, Saumya Kumar, Michelle Aillaud, Hsin-Chieh Tsay, Zhaoli Liu, Anke R.M. Kraft, Chai Fen Soon, Ivan Odak, Berislav Bošnjak, Anna Vlot, Deutsche COVID-19 OMICS Initiative (DeCOI), Morris A. Swertz, Uwe Ohler, Robert Geffers, Thomas Illig, Jochen Huehn, Antoine-Emmanuel Saliba, Leif Erik Sander, Reinhold Förster, Cheng-Jian Xu, Markus Cornberg, Leon N. Schulte, and Yang Li

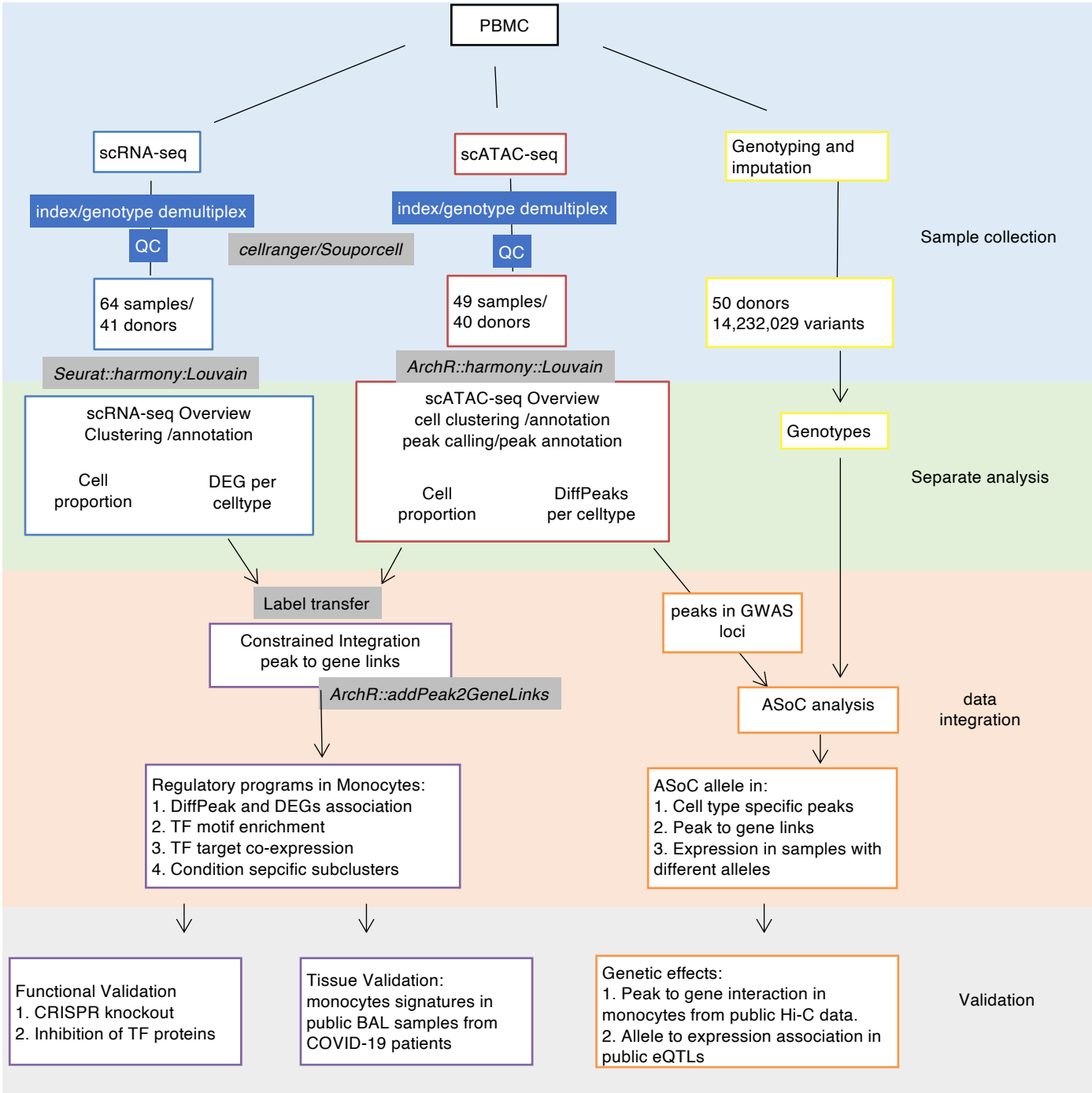

Figure S1. Overview of the analysis pipeline. Related to Figure 1 and STAR Methods.

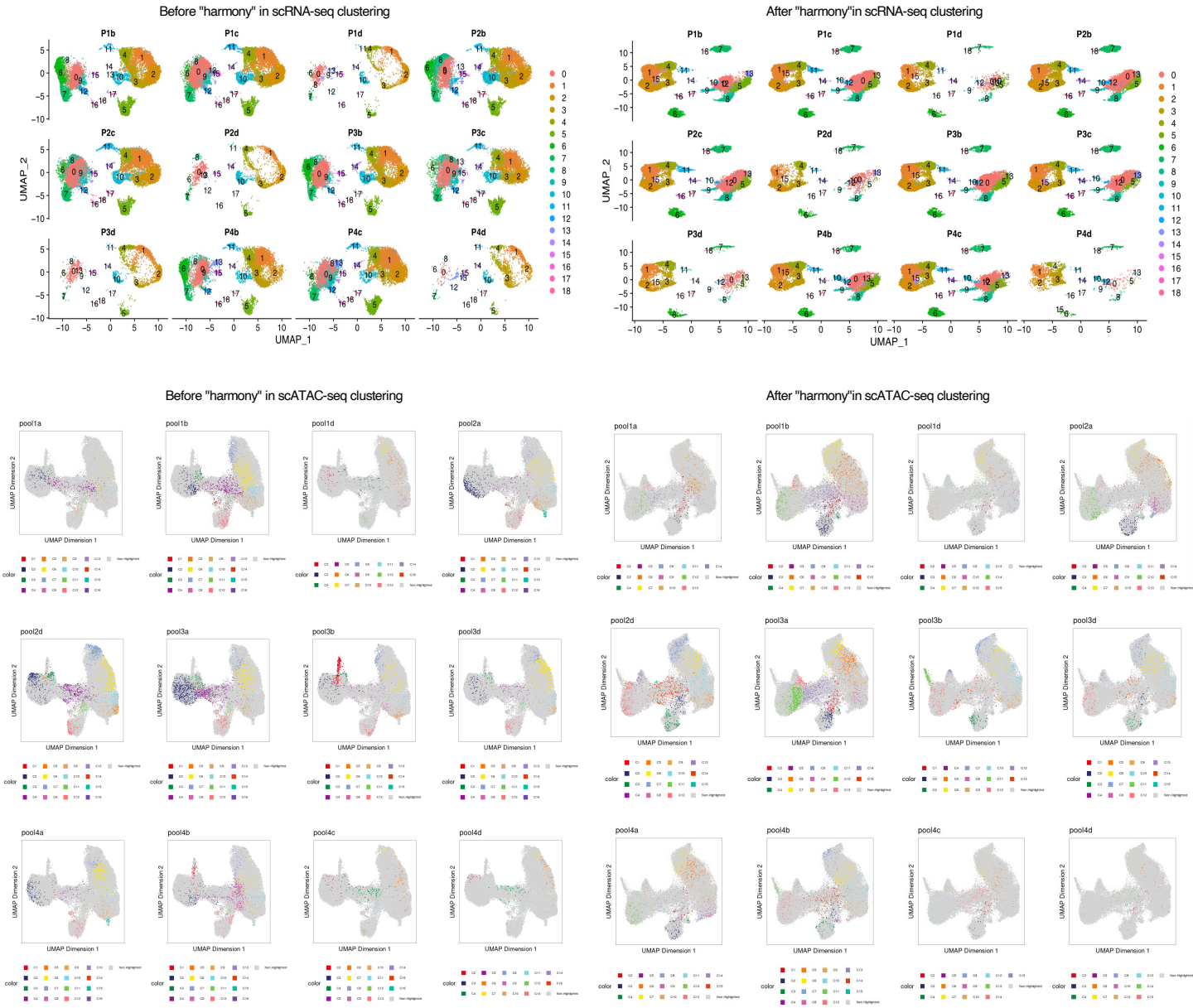

Figure S2. UMAP showing the batch effects before and after harmony integration from each scRNA-seq and scATAC-seq library. Related to STAR Methods.

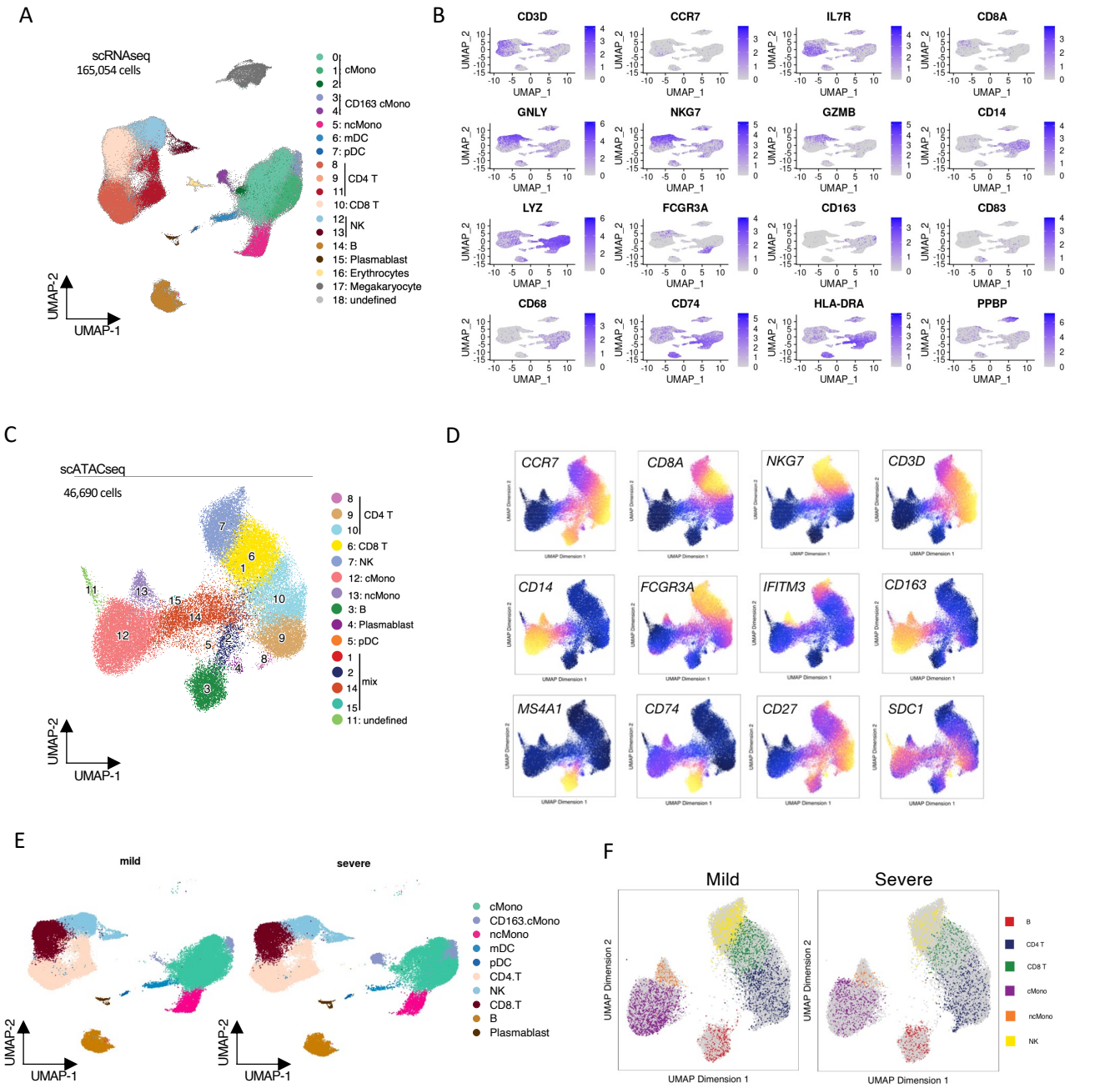

Figure S3. Marker genes and cell type annotation in scRNA-seq and scATAC-seq, related to Figure 1.
A. UMAP visualization showing the identified cell clusters from scRNA-seq, 165,054 cells from 64 samples (N = 37 hospitalized, N = 27 convalescent individuals). B. Expression of cell type markers in scRNA-seq clusters. C. UMAP visualization showing the identified cell clusters from scATAC-seq, 46,690 cells across 49 samples (N = 25 hospitalized, N = 24 convalescent individuals), 28 individuals shared between scRNA-seq and scATAC-seq profiling. D. Visualization of imputed marker gene activity scores in scATAC-seq clusters. E. UMAP showing the cell distribution between severe and mild patients in scRNA-seq. F. UMAP showing the cell distribution between severe and mild patients in scATAC-seq.

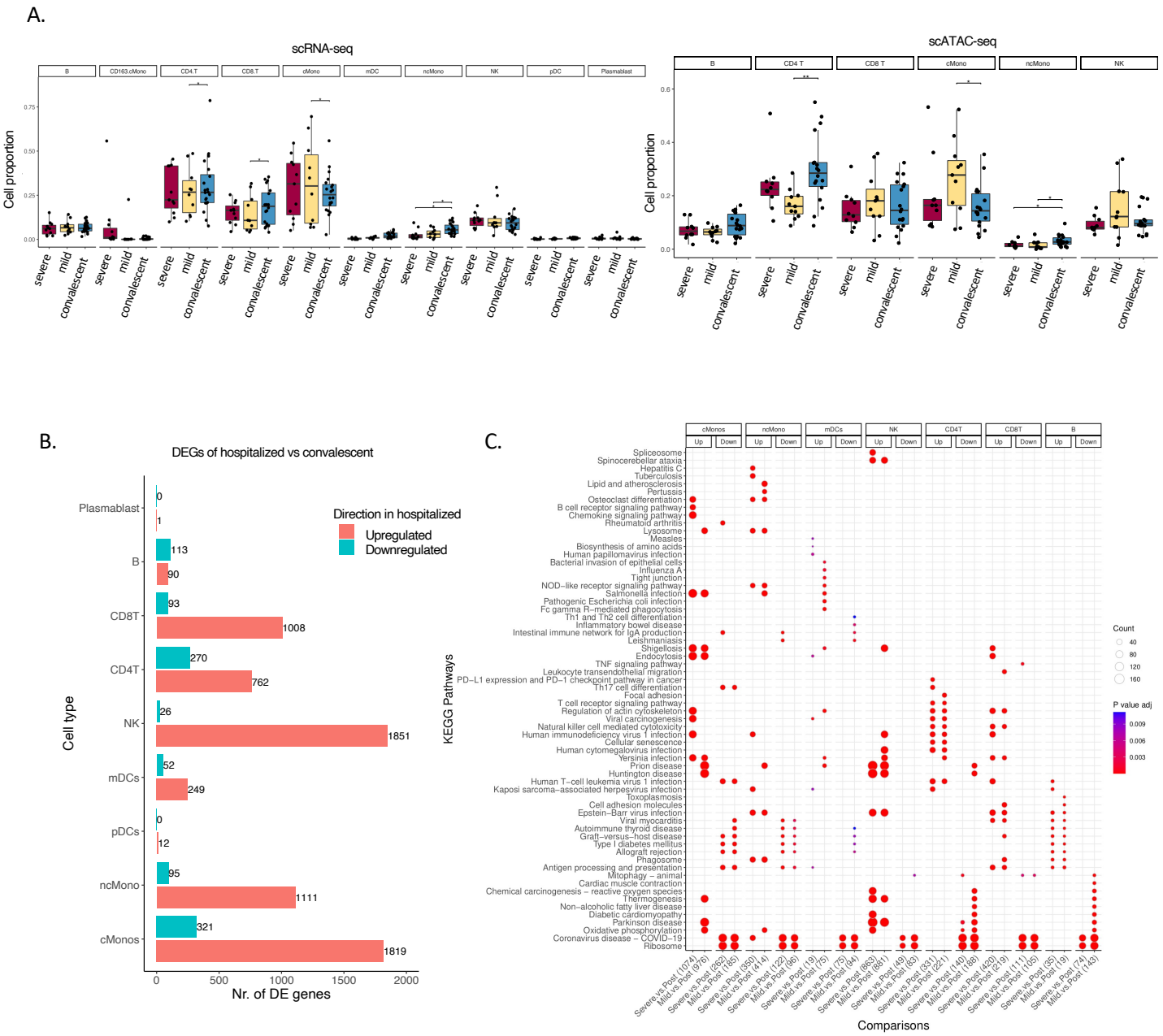

Figure S4. Severity-associated signatures in PBMC clusters, related to Figure 1.
A. Box plots showing the cell proportion of each cell type across severe, mild, and convalescent patients. B. Bar plots showing the number of significant differential expression genes (DEGs) identified in each cell type. DE analysis was performed to compare hospitalized vs convalescent individuals via Wilcoxon ranked-sum test. C. Pathway enrichments of the DEGs identified.

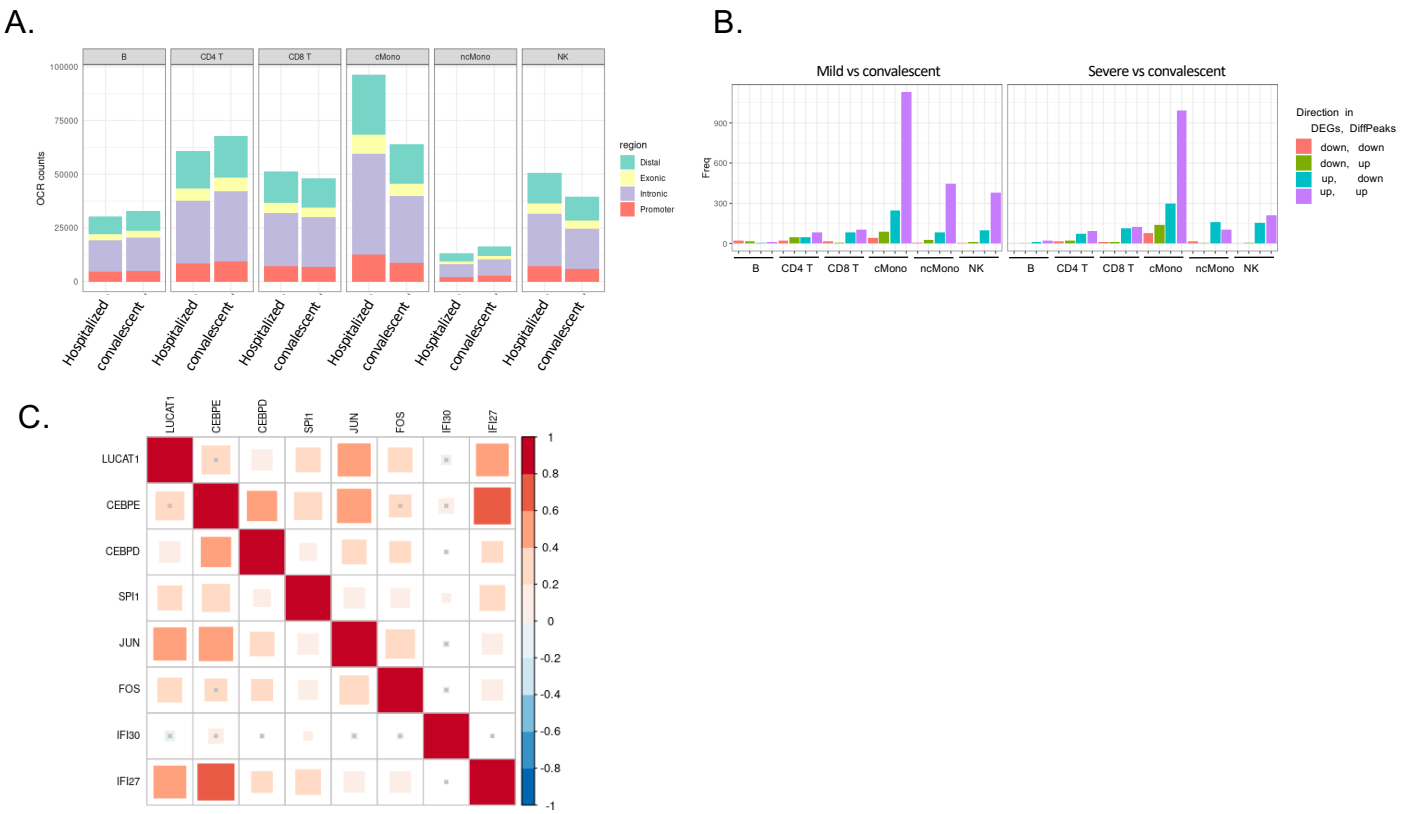

Figure S5. Chromatin accessibility signatures of PBMC clusters in COVID-19 patients, related to Figure 2.
A. Bar plots showing the number of regulatory elements in the open chromatin region (OCR) identified in each cell type. These were further classified into 'distal', 'extronic', 'intronic' and 'promoter' groups based on genome annotation. B. Barplots showing the overlap of significant differential expression genes (DEGs) and nominal differential accessible peaks (DAP) with shared or opposite regulation direction in hospitalized COVID-19 patients compared with convalescent patients. C. Heatmap showing the expression correlation of LUCAT1, IFI30 and transcriptional factor (TF) genes in classical monocytes of convalescent patients.

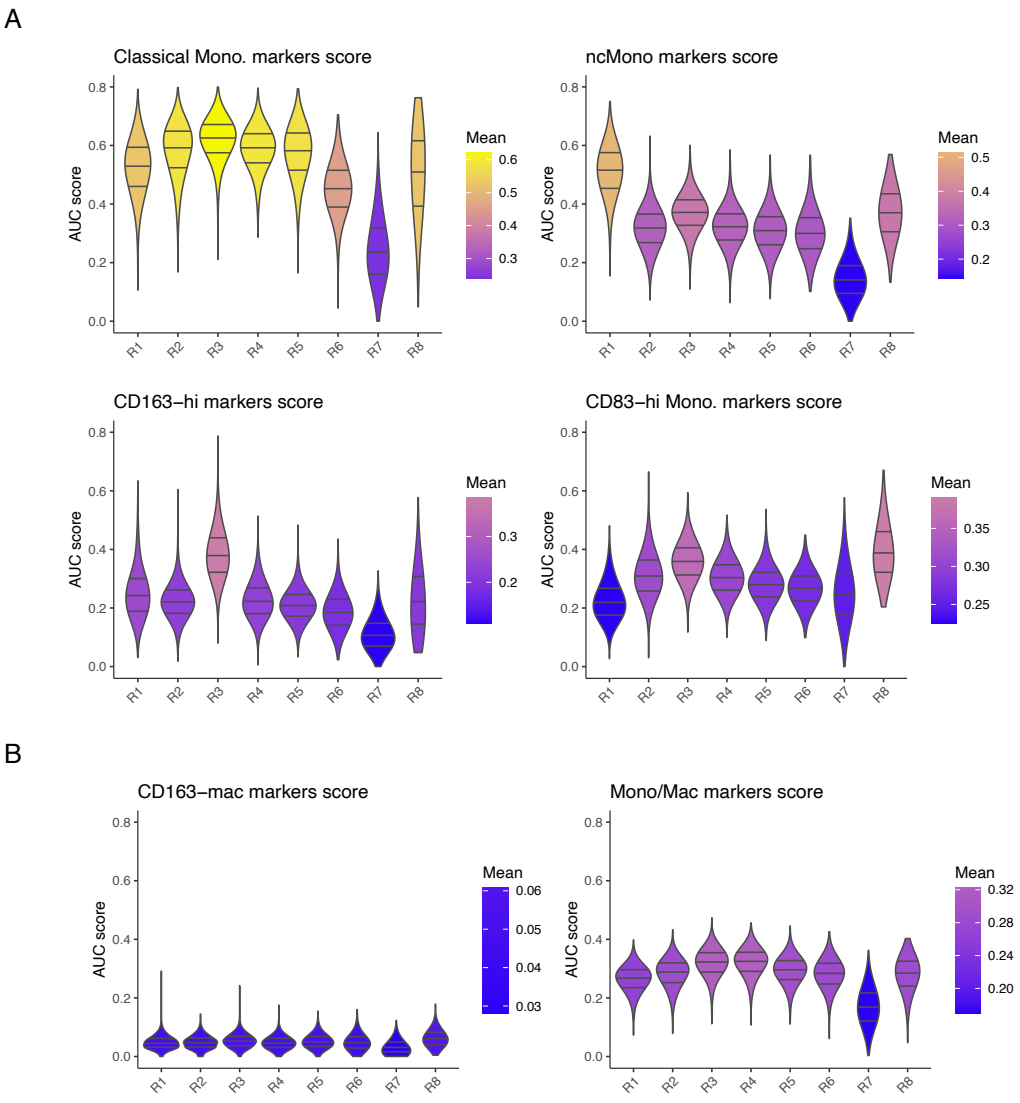

Figure S6. Violin plots showing the AUCell-based gene signature scores for each sub-clusters, related to Figure 3.

A, AUC scores from monocytes populations in PBMC (Schulte-Schrepping[4]); B, AUC scores from monocytes/macrophages (Mono/Mϕ) and profibrotic pulmonary macrophages (CD163/LGMN-Mϕ) in bronchoalveolar lavage (BAL) fluid (Wendisch[26]).

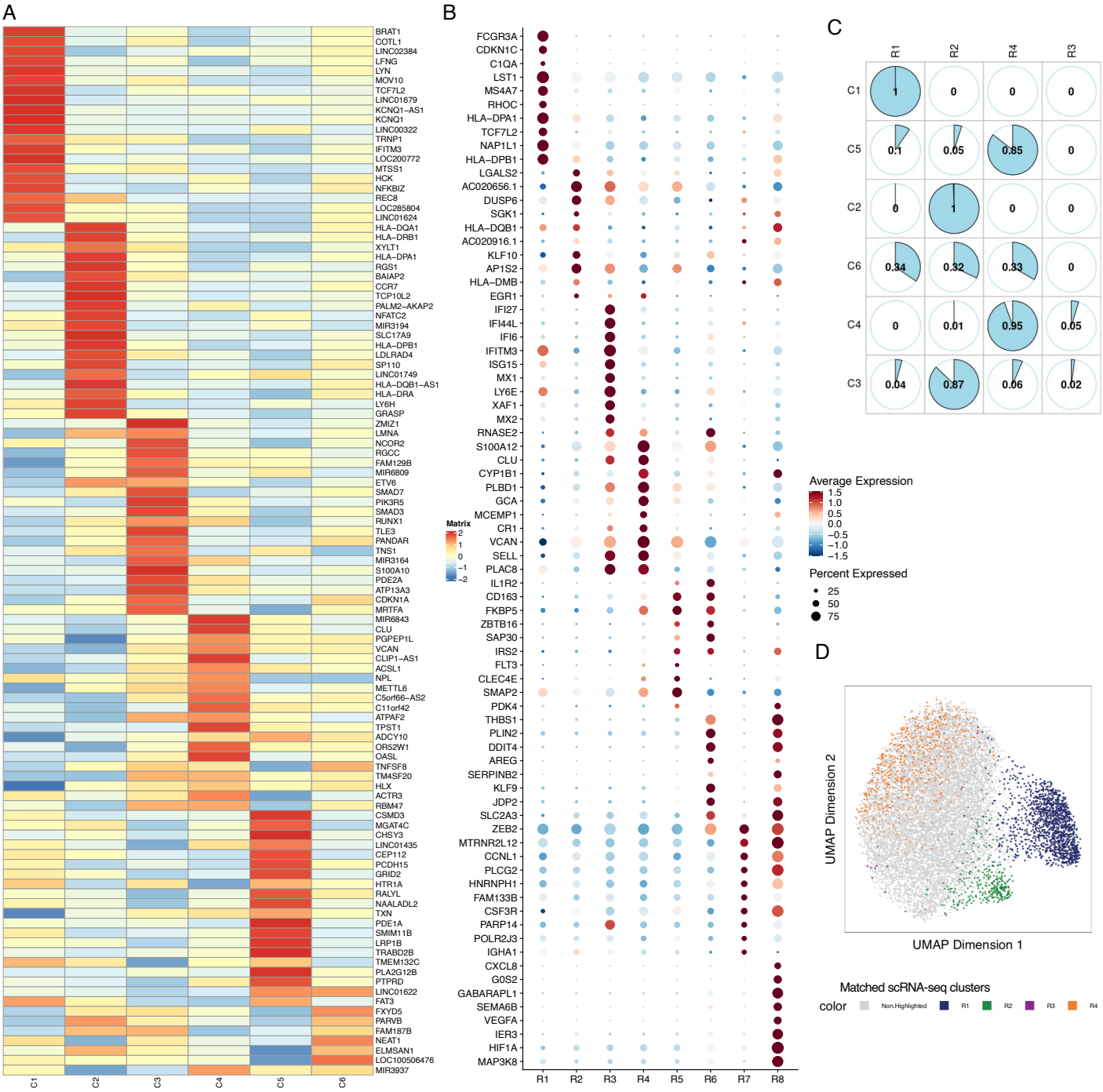

Figure S7. Cross-platform linkage across monocyte sub-clusters identified in scATAC-seq and scRNA-seq, related to Figure 3.

A. Heatmap showing the imputed activity scores of top-20 marker genes identified in monocytes sub-clusters of scATAC-seq. B. Dot plot showing the expression of top-10 marker genes identified in monocytes sub-clusters of scRNA-seq. C. Pie chart showing the percentage of matched cells in each scATAC-seq sub-clusters aligning to each scRNA-seq sub-clusters. D. UMAP showing the distribution of matched identities of cells in scATAC-seq datasets. Cells with a predicted linkage score < 0.6 were regarded as not matched cells and are colored as gray.

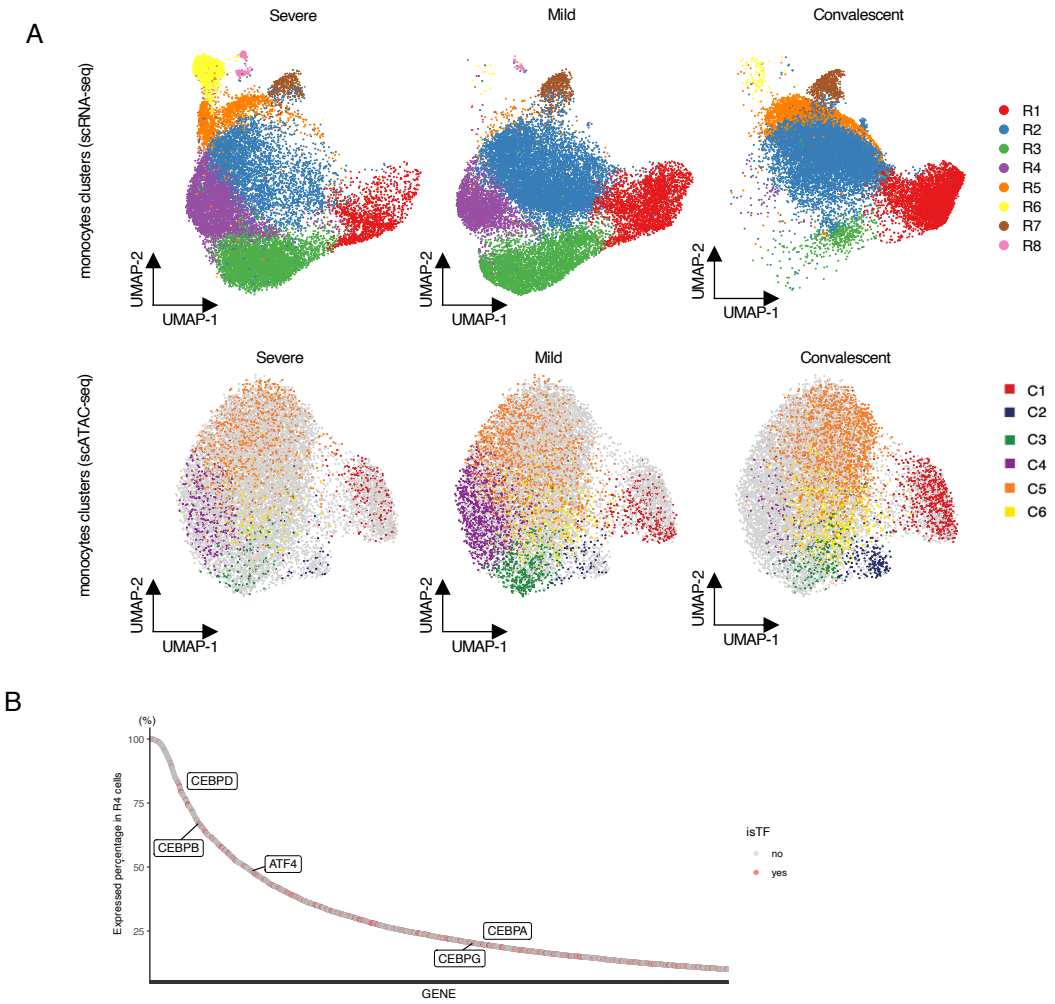

Figure S8. Cell distribution and gene ranks in monocyte subclusters, related to Figure 3.
A. UMAPs showing the distribution of monocytes subclusters across severe, mild and convalescent patients from the scRNA-seq (top row) and scATAC-seq (bottom row), respectively. B. Gene expressed percentage of cells from hospitalized COVID-19 patients in R4 clusters. TF genes are marked in RED.

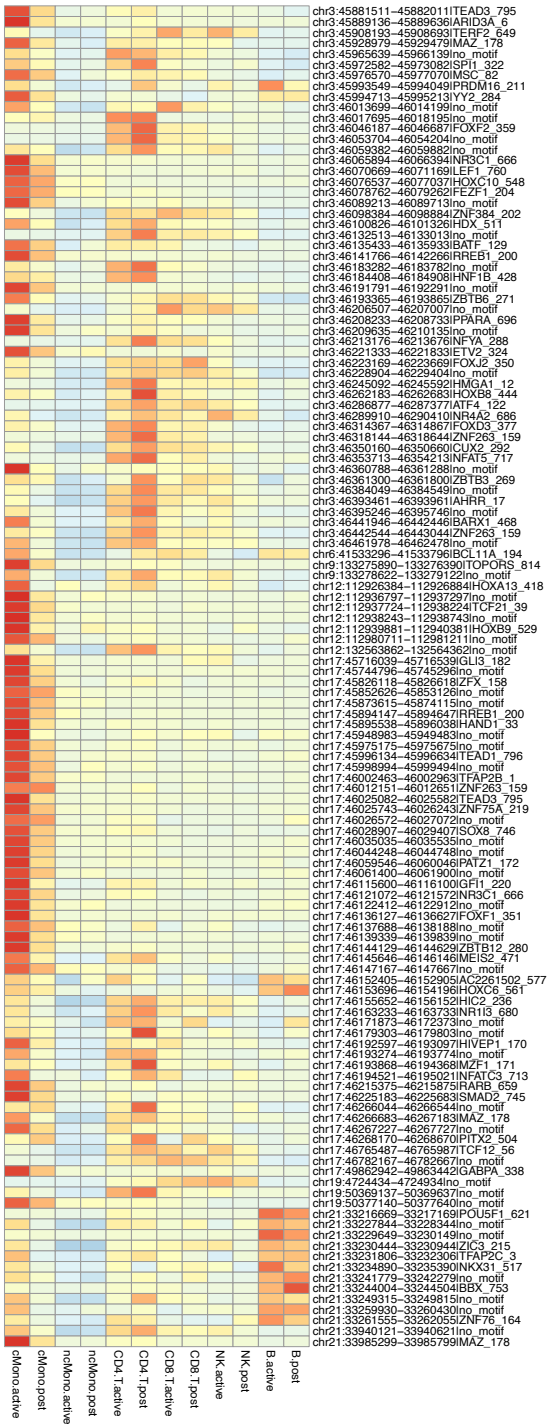

Figure S9. Full heatmap showing chromatin accessibility of peaks detected with hospitalized COVID-19 risk variants, related to Figure 4B. TF motifs that bind to peaks are indicated on the row names.

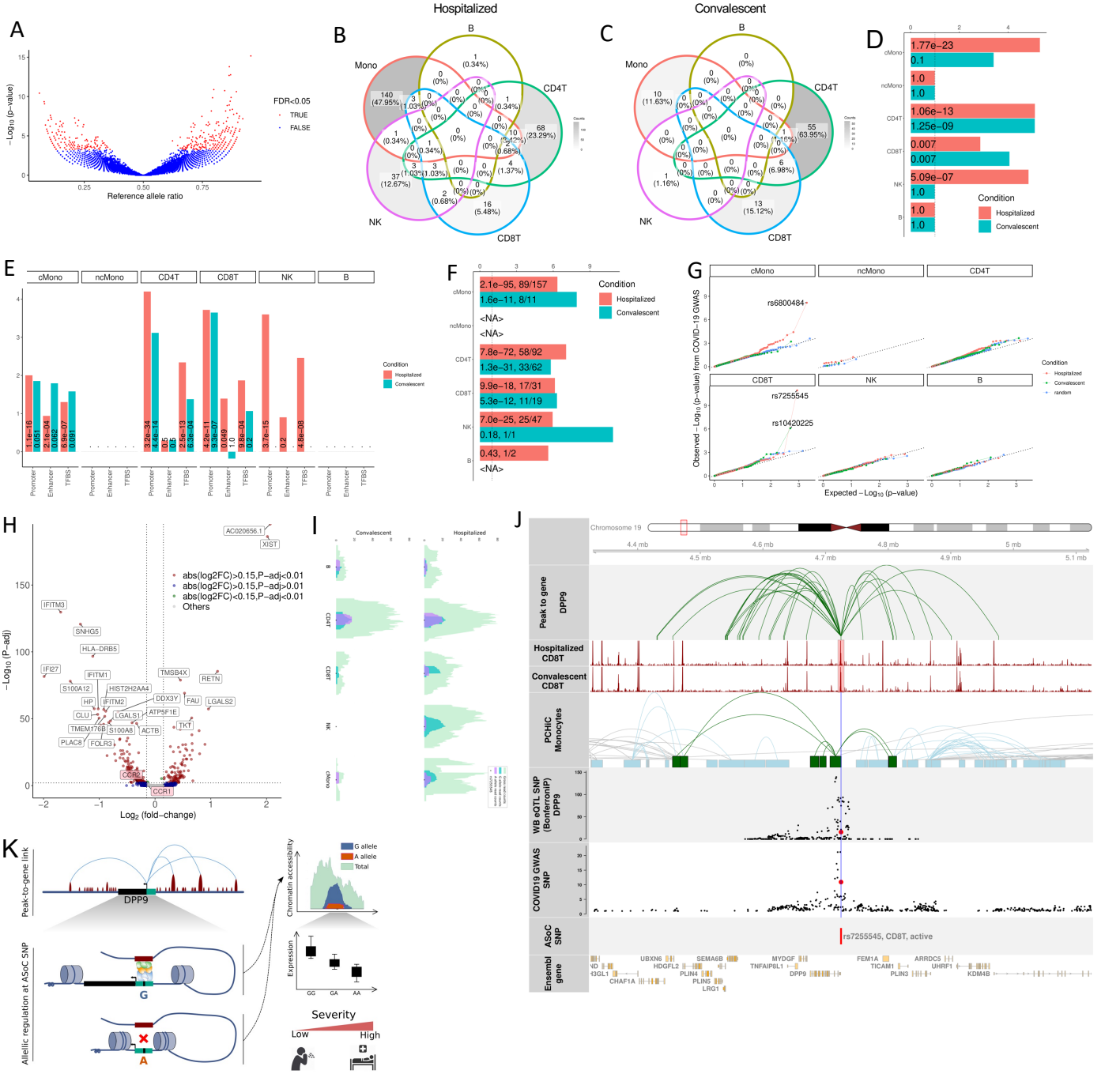

Figure S10. Supplementary information of ASoC analysis, Related to Figure 5.
A. Ratio of reference allele of each heterozygous SNPs used in the ASoC analysis. B. Number of identified ASoC SNPs in each cell type from hospitalized COVID-19 patients. C. Number of identified ASoC SNPs in each cell type from convalescent COVID-19 participants. D. Enrichment of ASoC SNPs that were located in open chromatin regions identified in our scATAC-seq analysis. E. Enrichment of ASoC SNPs that were located in promoter, enhancer and TFBS per cell type per condition. F. Enrichment of ASoC SNPs that were assigned to genes by eQTL analysis from eQTLGen consortium. G. QQ plot showing COVID-19 GWAS P-values (“Hospitalized covid vs. population”, release 6) of ASoC SNPs per cell type per condition. H. Volcano plot showing the differentially expressed genes between individuals carrying risk alleles (CC) and non-risk allele (TT). I. Allele-specific read counts at rs7255545 per cell type per condition. J. Data integration illustrated a potential regulatory program showing the effect of rs7255545 in COVID-19. K. Schematic plot showing the potential epigenetic and genetic regulating program at DPP9 locus under COVID-19 scenario.