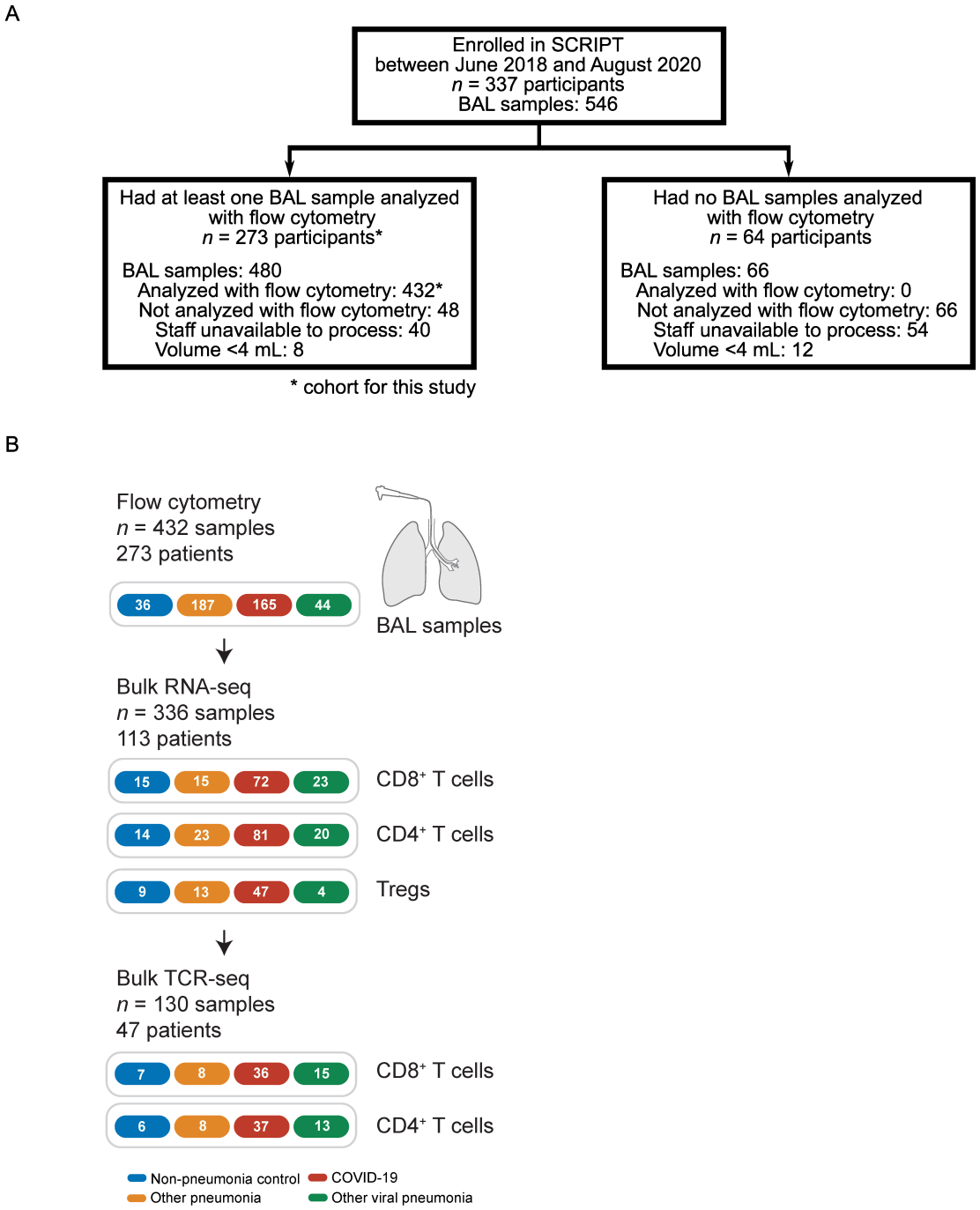
Supplemental Table and Figures

Supplemental Table 1

	Overall	Non-pneumonia control	Other pneumonia	COVID-19	Other viral pneumonia
Number	273	33	133	74	33
Age, median [IQR]	62.0 [42.0,71.0]	62.0 [42.0,70.0]	65.0 [51.0,72.0]	58.5 [44.5,66.8]	60.0 [55.0,69.0]
Female, n (%)	108 (39.6)	16 (48.5)	55 (41.4)	23 (31.1)	14 (42.4)
Ethnicity, n (%)					
Hispanic or Latino	60 (22.0)	5 (15.2)	13 (9.8)	37 (50.0)	5 (15.2)
Not Hispanic or Latino	204 (74.7)	27 (81.8)	118 (88.7)	33 (44.6)	26 (78.8)
Unknown or Not Reported	9 (3.3)	1 (3.0)	2 (1.5)	4 (5.4)	2 (6.1)
Race, n (%)					
Asian	10 (3.7)	1 (3.0)	5 (3.8)	3 (4.1)	1 (3.0)
Black or African American	59 (21.6)	7 (21.2)	28 (21.1)	18 (24.3)	6 (18.2)
Unknown or Not Reported	49 (17.9)	3 (9.1)	15 (11.3)	28 (37.8)	3 (9.1)
White	155 (56.8)	22 (66.7)	85 (63.9)	25 (33.8)	23 (69.7)
External transfer, n (%)	91 (33.3)	12 (36.4)	43 (32.3)	24 (32.4)	12 (36.4)
Admission BMI (kg/m2), median [IQR] *	29.2 [24.6,34.3]	26.4 [24.6,32.2]	27.0 [22.1,32.9]	31.9 [28.9,40.6]	28.3 [23.7,31.9]
SOFA score on ICU admission, median [IQR]	11.0 [8.0,13.0]	11.0 [8.0,13.0]	11.0 [8.0,14.0]	11.5 [9.0,13.0]	10.0 [6.0,14.0]
APS score on ICU admission, median [IQR]	90.0 [62.0,108.0]	80.0 [62.0,104.0]	87.0 [64.0,108.0]	91.0 [65.8,109.0]	83.0 [55.0,100.0]
Comorbidities, n (%)					
Myocardial infarction	20 (7.3)	5 (15.2)	12 (9.0)	2 (2.7)	1 (3.0)
Congestive heart failure	81 (29.7)	13 (39.4)	50 (37.6)	12 (16.2)	6 (18.2)
Peripheral vascular disease	63 (23.1)	8 (24.2)	39 (29.3)	9 (12.2)	7 (21.2)
Cerebrovascular disease	57 (20.9)	6 (18.2)	32 (24.1)	8 (10.8)	11 (33.3)
Dementia	13 (4.8)	3 (9.1)	4 (3.0)	2 (2.7)	4 (12.1)
Chronic pulmonary disease	96 (35.2)	13 (39.4)	53 (39.8)	17 (23.0)	13 (39.4)
Rheumatic disease	21 (7.7)	4 (12.1)	11 (8.3)	3 (4.1)	3 (9.1)
Peptic ulcer disease	27 (9.9)	5 (15.2)	16 (12.0)	1 (1.4)	5 (15.2)
Liver disease	80 (29.3)	12 (36.4)	44 (33.1)	11 (14.9)	13 (39.4)
Diabetes	104 (38.1)	10 (30.3)	41 (30.8)	35 (47.3)	18 (54.5)
Hemiplegia or paraplegia	19 (7.0)	3 (9.1)	12 (9.0)	2 (2.7)	2 (6.1)
Renal disease	77 (28.2)	14 (42.4)	41 (30.8)	11 (14.9)	11 (33.3)
Cancer	87 (31.9)	12 (36.4)	48 (36.1)	11 (14.9)	16 (48.5)
Immunocompromise	70 (25.6)	8 (24.2)	40 (30.1)	6 (8.1)	16 (48.5)
Biomarkers on day of first BAL procedure, median [IQR]					
C-reactive protein, mg/dL *	163.2 [92.0,259.5]	12.0 [11.0,55.0]	96.0 [22.5,176.0]	169.0 [116.0,275.5]	168.0 [168.0,168.0]
D-dimer, ng/mL *	1103.0 [492.5,3540.5]	1345.5 [405.8,2220.5]	2922.0 [886.0,5091.0]	896.5 [468.8,2892.6]	1954.2 [1170.5,2931.1]
Ferritin, ng/mL *	663.2 [300.9,1210.3]	92.4 [57.0,191.2]	631.8 [319.8,1183.7]	766.4 [330.1,1355.2]	–
Lactate, mmol/L *	1.5 [1.1,2.1]	1.7 [1.1,2.3]	1.6 [1.1,2.4]	1.4 [1.0,1.8]	1.3 [1.0,1.7]
Procalcitonin, ng/mL *	0.6 [0.2,2.8]	0.4 [0.1,1.6]	0.6 [0.3,2.9]	0.4 [0.1,2.9]	0.8 [0.3,3.8]
White blood cell count, per 1,000/μL *	11.6 [8.0,16.7]	12.6 [8.1,20.4]	12.6 [8.7,16.8]	10.1 [7.1,14.6]	10.7 [7.2,18.5]
Absolute neutrophil count, per 1,000/μL *	9.2 [5.7,13.7]	12.2 [6.0,18.8]	9.6 [6.3,15.3]	8.7 [5.5,11.2]	8.9 [5.2,13.0]
Absolute lymphocyte count, per 1,000/μL *	0.8 [0.5,1.4]	0.6 [0.4,0.8]	0.9 [0.5,1.7]	1.0 [0.6,1.3]	0.8 [0.2,1.6]
Medications					
Corticosteroids administered during admission, n (%)	162 (59.3)	20 (60.6)	80 (60.2)	39 (52.7)	23 (69.7)
Prednisone equivalents administered during admission, mg	70.0 [0.0,240.0]	80.0 [0.0,340.0]	92.0 [0.0,250.0]	15.0 [0.0,209.0]	96.0 [0.0,272.0]
Tocilizumab, n (%)	13 (4.8)	–	1 (0.8)	12 (16.2)	–
Sarilumab, n (%)	12 (4.4)	–	–	12 (16.2)	–
Remdesivir, n (%)	12 (4.4)	–	–	12 (16.2)	–
Remdesivir or placebo, n (%)	9 (3.3)	–	–	9 (12.2)	–
ICU course					
Ventilation duration (days), median [IQR]	10.0 [4.0,22.0]	3.0 [2.0,9.0]	9.0 [4.0,19.0]	20.0 [10.0,33.0]	7.0 [3.0,13.0]
ICU length of stay (days), median [IQR]	13.0 [6.0,23.0]	6.0 [4.0,11.0]	11.0 [5.0,21.0]	21.0 [14.0,35.0]	11.0 [9.0,20.0]
Tracheostomy, n (%)	70 (25.6)	3 (9.1)	29 (21.8)	32 (43.2)	6 (18.2)
Number of BALs, n (% of total)	432 (100)	36 (8.3)	187 (43.3)	165 (38.2)	44 (10.2)
Discharge disposition, n (%)					
Died	86 (31.5)	10 (30.3)	47 (35.3)	19 (25.7)	10 (30.3)
Home	76 (27.8)	12 (36.4)	27 (20.3)	30 (40.5)	7 (21.2)
LTACH	35 (12.8)	3 (9.1)	17 (12.8)	12 (16.2)	3 (9.1)
Rehab	50 (18.3)	4 (12.1)	27 (20.3)	9 (12.2)	10 (30.3)
SNF	18 (6.6)	2 (6.1)	12 (9.0)	4 (5.4)	–
Hospice	8 (2.9)	2 (6.1)	3 (2.3)	0 (0.0)	3 (9.1)

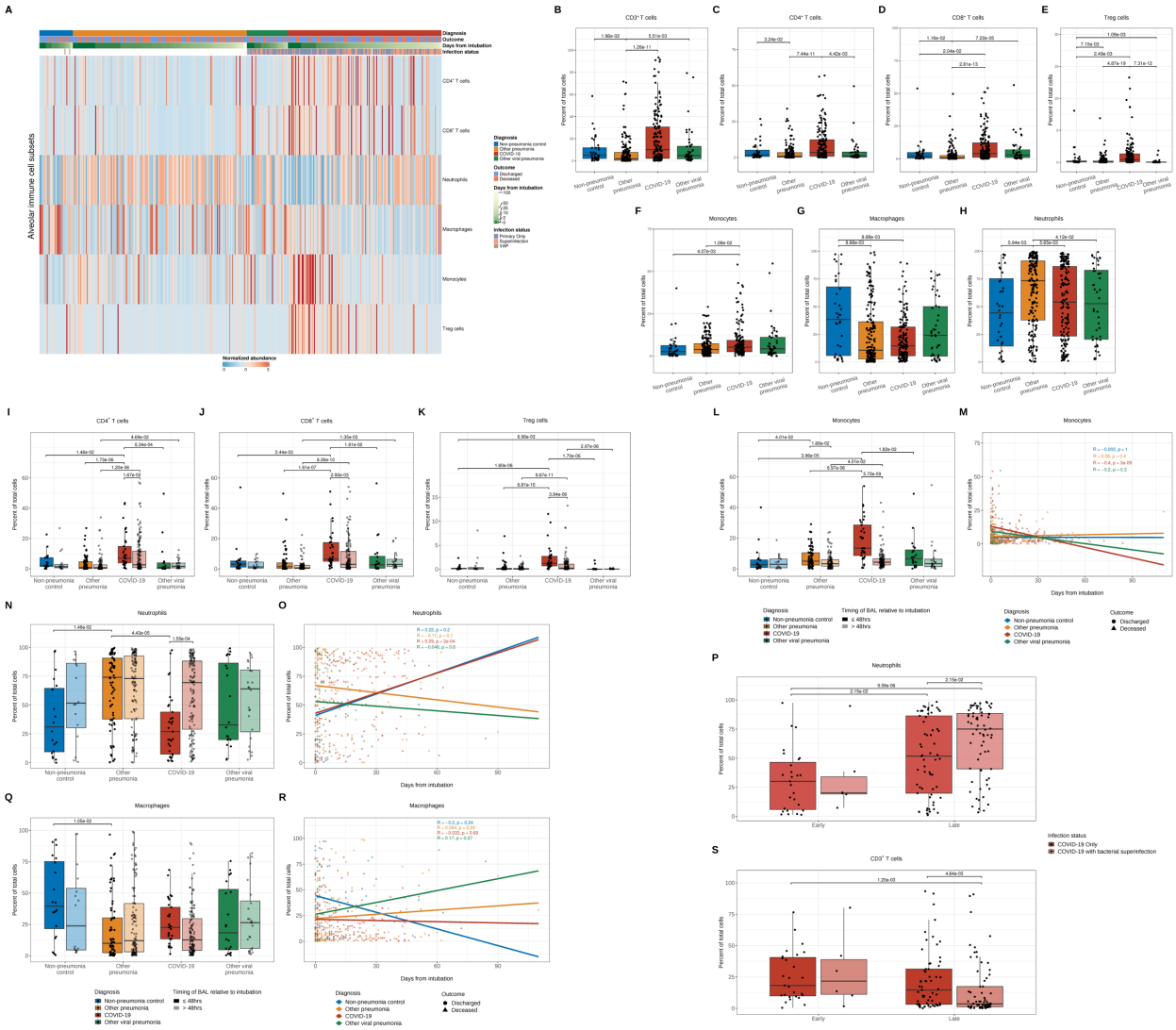
Supplemental Table 1. Description of the cohort. * Missing values, n: BMI, 1; C-reactive protein, 195; D-dimer, 174; ferritin, 208; lactate, 70; procalcitonin, 114; white blood cell count, 3; absolute neutrophil count, 96; absolute lymphocyte count 101. Empty cells represent data not available or not applicable. BMI, body mass index; SOFA, Sequential Organ Failure Assessment; APS, Acute Physiology Score from APACHE II; LTACH, long-term acute care hospital; SNF, skilled nursing facility; IQR, interquartile range.

Supplemental Figure 1



Supplemental Figure 1. (A) CONSORT diagram of patients included in this study. **(B)** Schematic depicting multi-step analysis of BAL fluid samples with flow cytometry, bulk RNA-sequencing, and bulk TCR-sequencing by diagnosis and T cell subset.

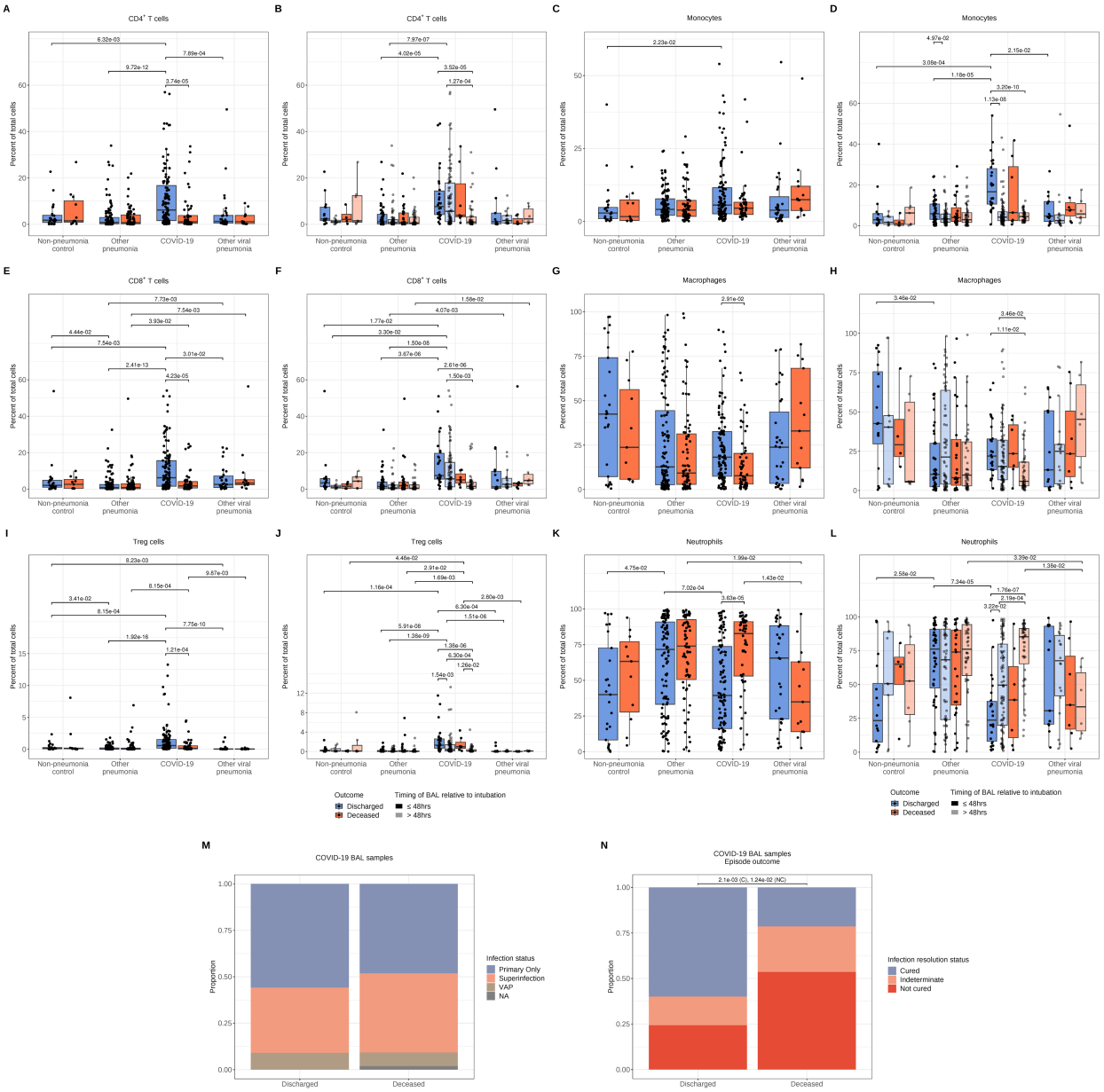
Supplemental Figure 2



Supplemental Figure 2. SARS-CoV-2 pneumonia is characterized by a lymphomonocytic alveolar infiltrate early following intubation. (A) Heatmap of flow cytometry analysis of alveolar immune cell subsets from BAL fluid samples ordered by duration of mechanical ventilation (blanks indicate chronically ventilated patients) and grouped by diagnosis, binary outcome (whether a given patient was discharged or died during hospitalization), and infection status (presence or absence of bacterial superinfection in patients with COVID-19 or other viral pneumonia). Blanks in these two groups refer to samples for which microbiological data were incomplete and infectious status could not be determined. The VAP (ventilator-associated pneumonia) flag

designates samples from non-pneumonia controls or patients with COVID-19 or other viral pneumonia who cleared the virus and then developed a bacterial pneumonia. **(B-H)** Percent of alveolar immune cell subsets detected in BAL fluid samples from flow cytometry analysis (q < 0.05, pairwise Wilcoxon rank-sum tests with FDR correction). **(I-L**, **N** and **Q)** Comparison of alveolar immune cell subset percentages between early (≤48 hours following intubation) and late (>48 hours following intubation) samples (q < 0.05, pairwise Wilcoxon rank-sum tests with FDR correction). **(M, O** and **R)** Correlation analysis between the percentage of alveolar immune cell subsets and duration of mechanical ventilation with Pearson correlation coefficient. **(P** and **S)** Comparison of neutrophil (P) and $CD3^+$ T cell (S) percentage grouped by the presence or absence of bacterial superinfection in early (≤48 hours following intubation) and late (>48 hours following intubation) COVID-19 samples (q < 0.05, pairwise Wilcoxon rank-sum tests with FDR correction).

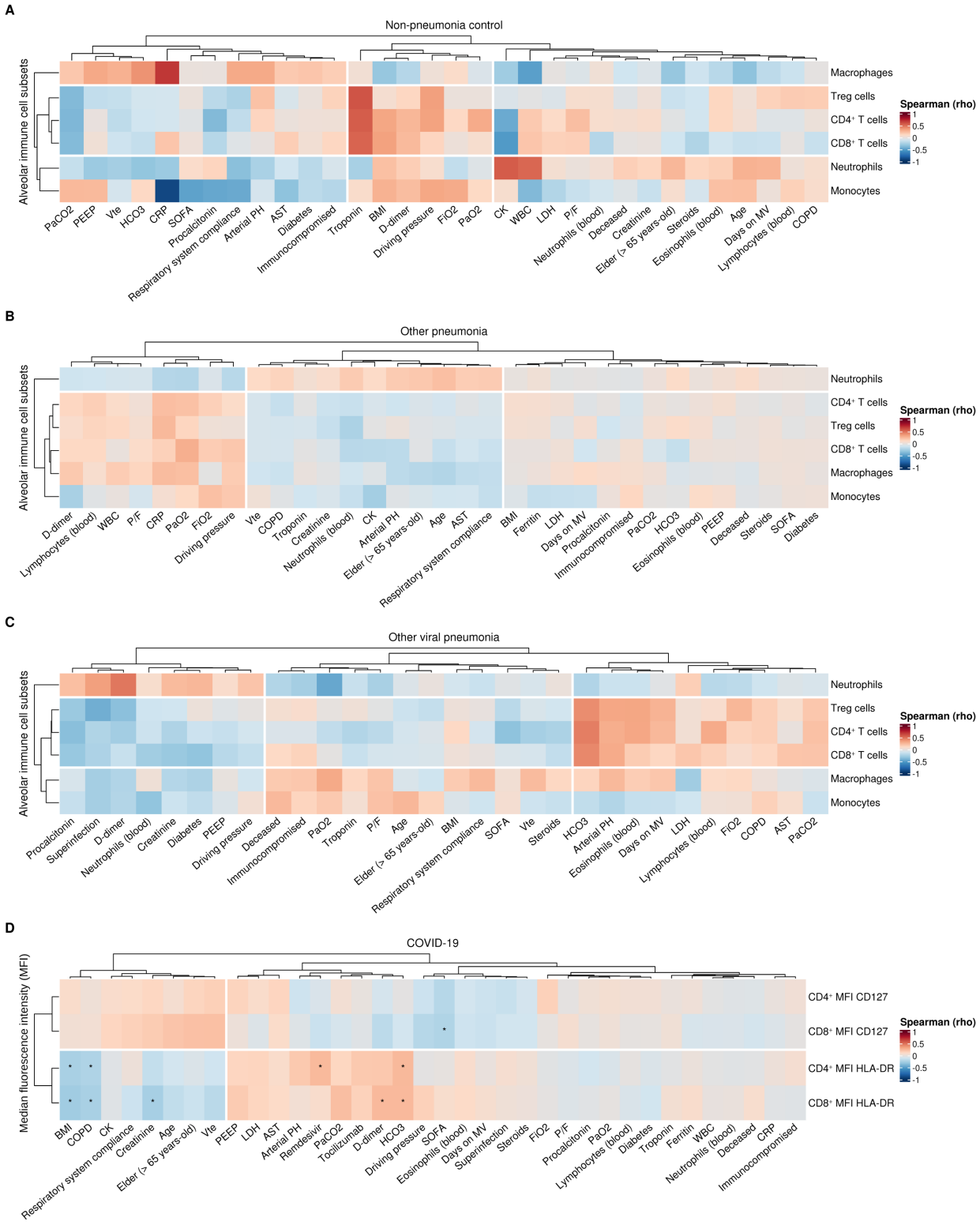
Supplemental Figure 3



Supplemental Figure 3. Persistent alveolar T cell enrichment throughout the course of severe SARS-CoV-2 pneumonia is associated with discharge from hospital. (A-L) Comparison of alveolar immune cell subset percentages by binary outcome (C, E, G, I, and K) and between timing of BAL sampling (B, D, F, H, J and L) (q <0.05, pairwise Wilcoxon rank-sum tests with FDR correction). **(M)** Proportion of BAL fluid samples from patients with COVID-19, comparing presence or absence of bacterial superinfection with binary outcome (not significant

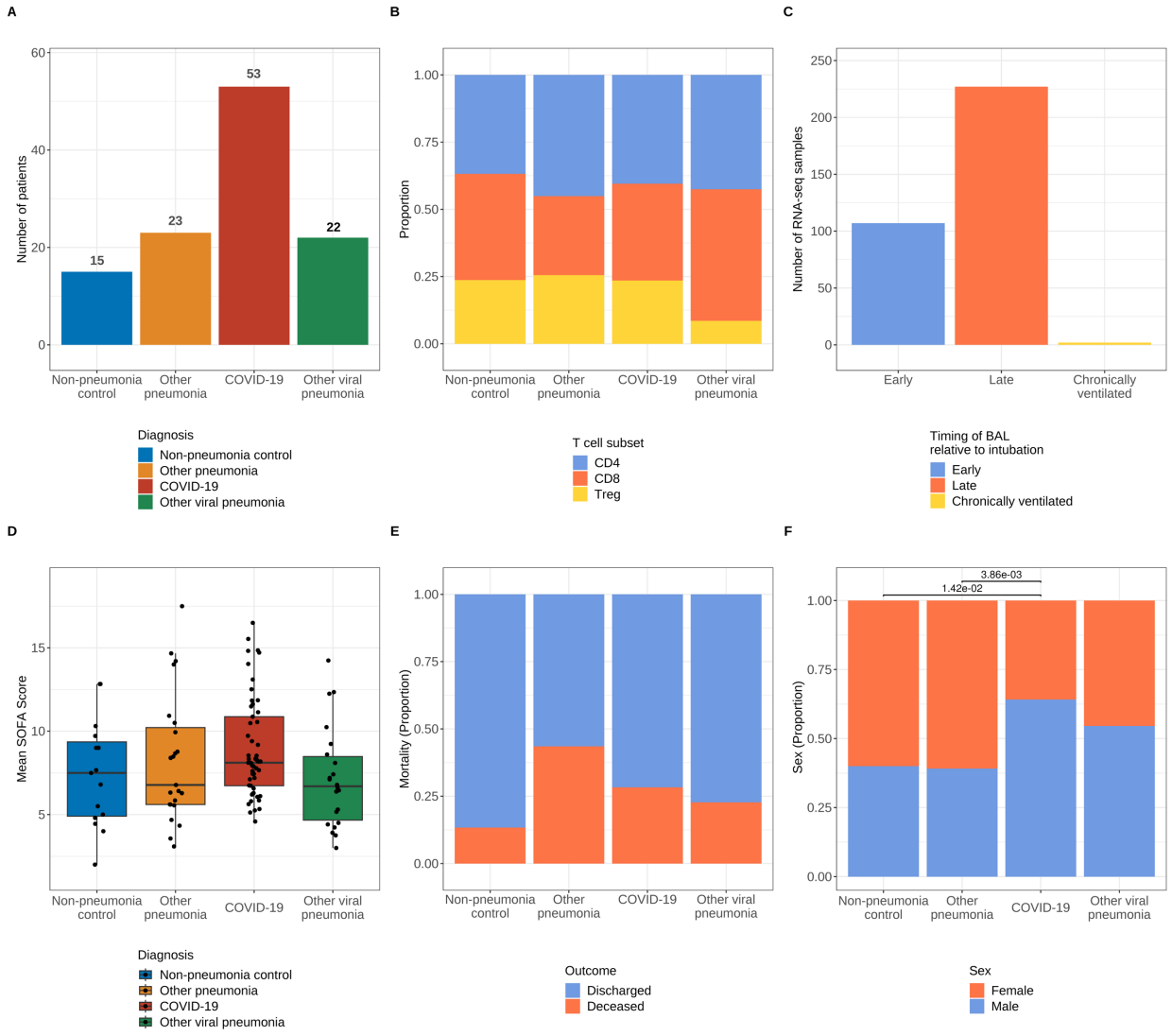
by Fisher exact test). **(N)** Proportion of BAL fluid samples from patients with COVID-19, comparing pneumonia episode outcome status with binary outcome (q < 0.05, Fisher exact test with FDR correction). C (Cured) and NC (Not cured).

Supplemental Figure 4



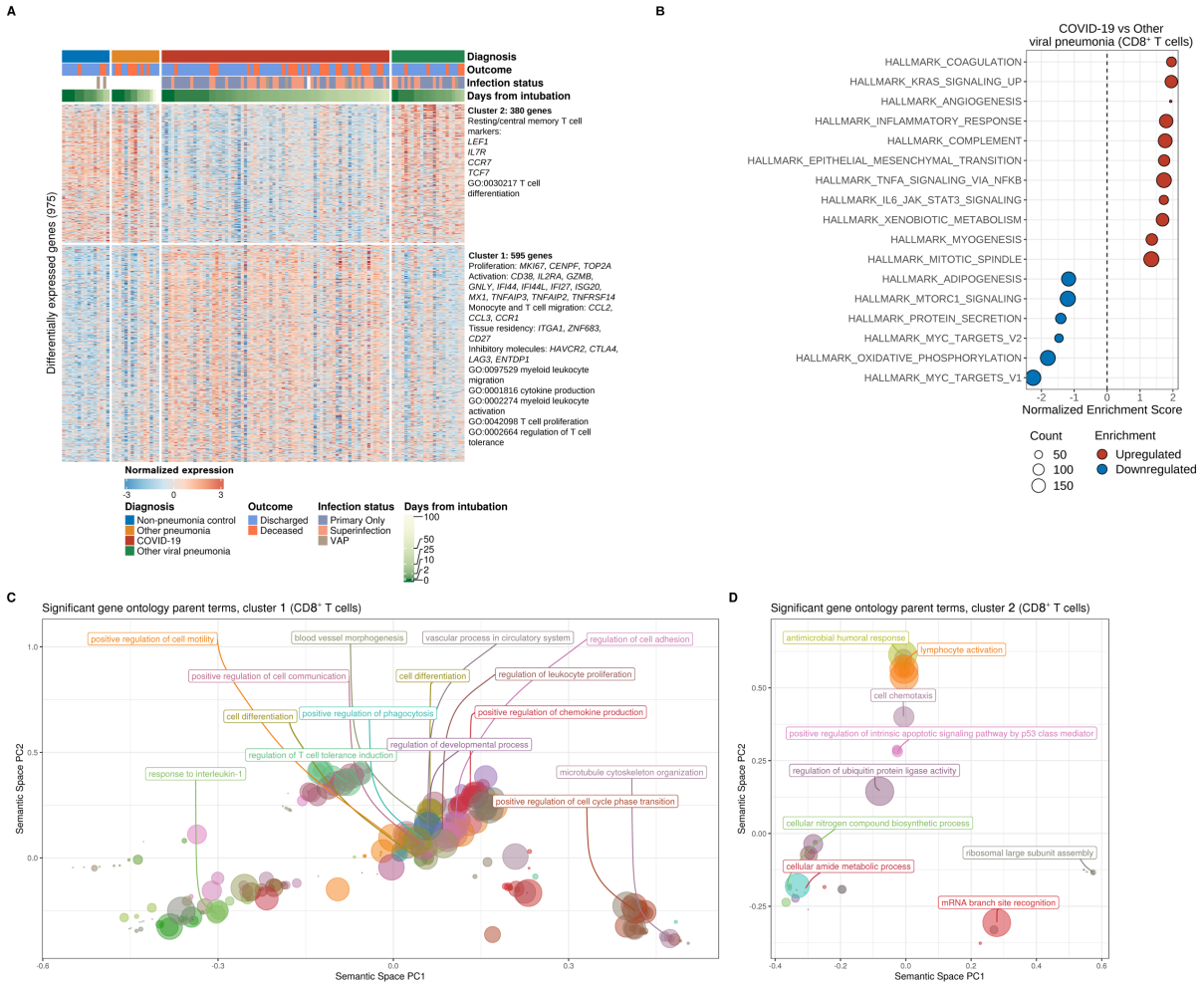
Supplemental Figure 4. Correlation of alveolar immune cell subset abundance with clinical outcomes differs between patients with distinct etiologies of severe pneumonia. (A-C). Correlation analysis between the percentage of alveolar immune cell subsets and clinical, physiologic, and laboratory variables as a function of diagnostic group. No significant values after calculating Spearman rank correlation coefficient with FDR correction. **(D)** Correlation between T cell subset surface expression of CD127 and HLA-DR in the alveolar space with clinical, laboratory, and physiological variables in COVID-19 samples. Spearman rank correlation coefficient with FDR correction ($q < 0.05$ [*]). Abbreviations: PaCO2 (partial arterial carbon dioxide pressure), HCO3 (bicarbonate), Days on MV (days on mechanical ventilation), SOFA (Sequential Organ Failure Assessment), WBC (peripheral white blood cells), CK (creatinine kinase), Vte (minute ventilation), LDH (lactate dehydrogenase), FiO2 (fraction of inspired oxygen), CRP (C-reactive protein), PEEP (positive end-expiratory pressure), BMI (body mass index), AST (aspartate aminotransferase), PaO2 (partial arterial oxygen pressure), P/F (ratio of partial arterial oxygen pressure to fraction of inspired oxygen), COPD (chronic obstructive pulmonary disease), MFI (median fluorescence intensity).

Supplemental Figure 5



Supplemental Figure 5. Composition and demographics of bulk RNA-sequencing samples. **(A)** Number of participants grouped by diagnosis. **(B)** Proportion of samples grouped by T cell subset and diagnosis. **(C)** Number of samples categorized by timing of BAL. **(D)** Distribution of Sequential Organ Failure Assessment (SOFA) scores by patient. Nonsignificant after pairwise Wilcoxon rank-sum tests with FDR correction). **(E)** Mortality by patient. Nonsignificant after pairwise χ^2 tests for homogeneity of proportions with FDR correction. **(F)** Sex by patient (pairwise χ^2 tests for homogeneity of proportions with FDR correction).

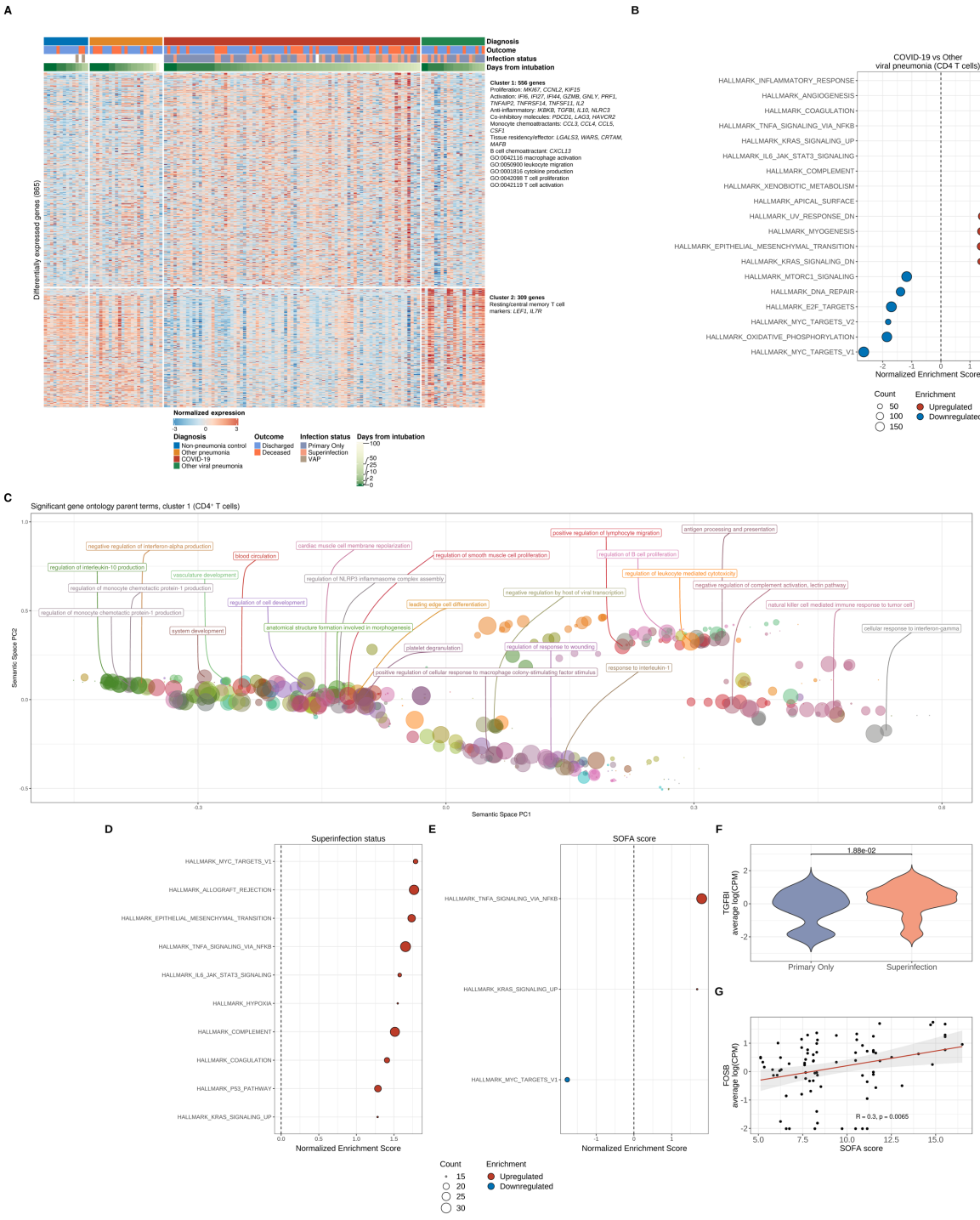
Supplemental Figure 6



Supplemental Figure 6. SARS-CoV-2 pneumonia is characterized by a transcriptional program enriched for processes associated with monocyte and T cell activation, migration, and angiogenesis in alveolar CD8⁺ T cells. (A) *K*-means clustering of 975 differentially expressed genes (q < 0.05, likelihood-ratio test with FDR correction) across pneumonia diagnoses. Columns represent unique samples grouped by diagnosis and are ordered by duration of mechanical ventilation. Column headers are color-coded by diagnosis, binary outcome (whether a given patient was discharged or died during hospitalization), duration of mechanical ventilation (blanks indicate chronically ventilated patients), and infection status (presence or absence of bacterial superinfection in patients with COVID-19 or other viral pneumonia). The VAP

(ventilator-associated pneumonia) flag designates samples from non-pneumonia controls or patients with COVID-19 or other viral pneumonia who cleared the virus and then developed a bacterial pneumonia. Representative genes and significant gene ontology (GO) biological processes are shown for each cluster. **(B)** Gene set enrichment analysis (GSEA) of Hallmark gene sets for the pairwise comparison between COVID-19 samples and other viral pneumonia samples. Count denotes pathway size after removing genes not detected in the expression dataset. Enrichment denotes significant (q < 0.25 with FDR correction) upregulated (red) and downregulated (blue) pathways by normalized enrichment score. **(C-D)** Gene ontology (GO) parent term annotation after grouping significant terms (following classical or over-representation enrichment analysis, q < 0.05 with multiple testing correction using the Benjamini-Hochberg method) by semantic similarity. Reduced GO terms are depicted in a scatter plot where distance between points represent the similarity between terms and axes are the first two components after applying a principal coordinates analysis to the dissimilarity matrix. Points are color-coded by unique terms and size denotes the number of genes within each GO term.

Supplemental Figure 7

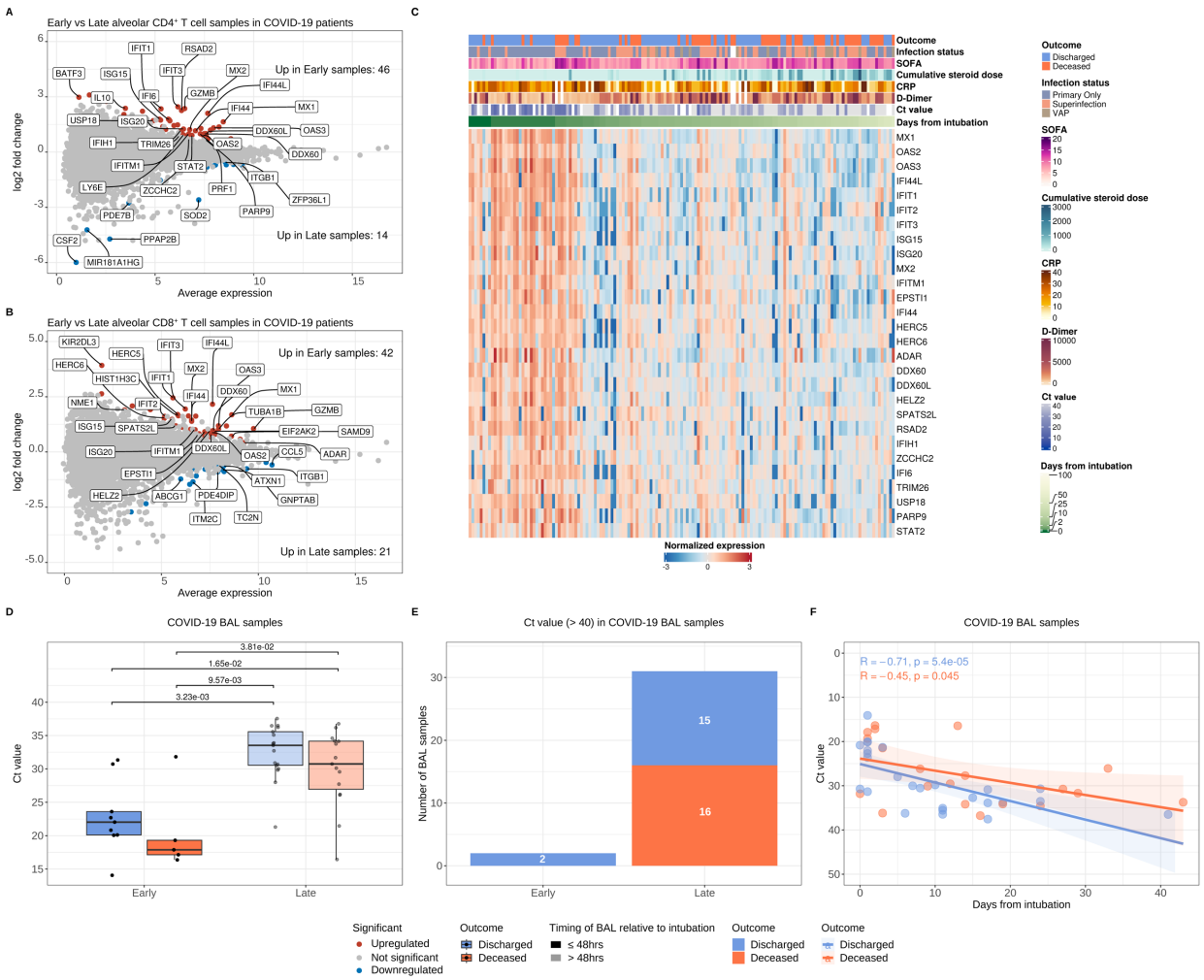


Supplemental Figure 7. SARS-CoV-2 pneumonia is characterized by a transcriptional program enriched for processes associated with monocyte, B cell and T cell activation, migration, and angiogenesis in alveolar CD4⁺ T cells. (A) *K*-means clustering of 865

differentially expressed genes (q < 0.05, likelihood-ratio test with FDR correction) across pneumonia diagnoses. Columns represent unique samples grouped by diagnosis and are ordered by duration of mechanical ventilation. Column headers are color-coded by diagnosis, binary outcome (whether a given patient was discharged or died during hospitalization), duration of mechanical ventilation (blanks indicate chronically ventilated patients), and infection status (presence or absence of bacterial superinfection in patients with COVID-19 or other viral pneumonia). The VAP (ventilator-associated pneumonia) flag designates samples from non-pneumonia controls or patients with COVID-19 or other viral pneumonia who cleared the virus and then developed a bacterial pneumonia. Representative genes and significant gene ontology (GO) biological processes are shown for each cluster. **(B)** Gene set enrichment analysis (GSEA) of Hallmark gene sets for the pairwise comparison between COVID-19 samples and other viral pneumonia samples. Count denotes pathway size after removing genes not detected in the expression dataset. Enrichment denotes significant (q < 0.25 with FDR correction) upregulated (red) and downregulated (blue) pathways by normalized enrichment score. **(C)** Gene ontology (GO) parent term annotation after grouping significant terms (following classical or over-representation enrichment analysis, q < 0.05 with multiple testing correction using the Benjamini-Hochberg method) by semantic similarity. Reduced GO terms are depicted in a scatter plot where distance between points represent the similarity between terms and axes are the first two components after applying a principal coordinates analysis to the dissimilarity matrix. Points are color-coded by unique terms and size denotes the number of genes within each GO term. **(D-E)** GSEA of COVID-19 samples after performing correlation analysis of differentially expressed genes in $CD4^+$ T cells and clinical variables of interest with Spearman rank correlation coefficient computation. Count denotes pathway size after removing genes not detected in the expression dataset. Enrichment denotes significant (q < 0.25 with FDR correction) upregulated (red) and downregulated (blue) pathways by normalized enrichment score. **(F-G)** Leading edge analysis

reveals selected core genes driving pathway enrichment signal in clinical variables, which are annotated for the superinfection and SOFA score variables.

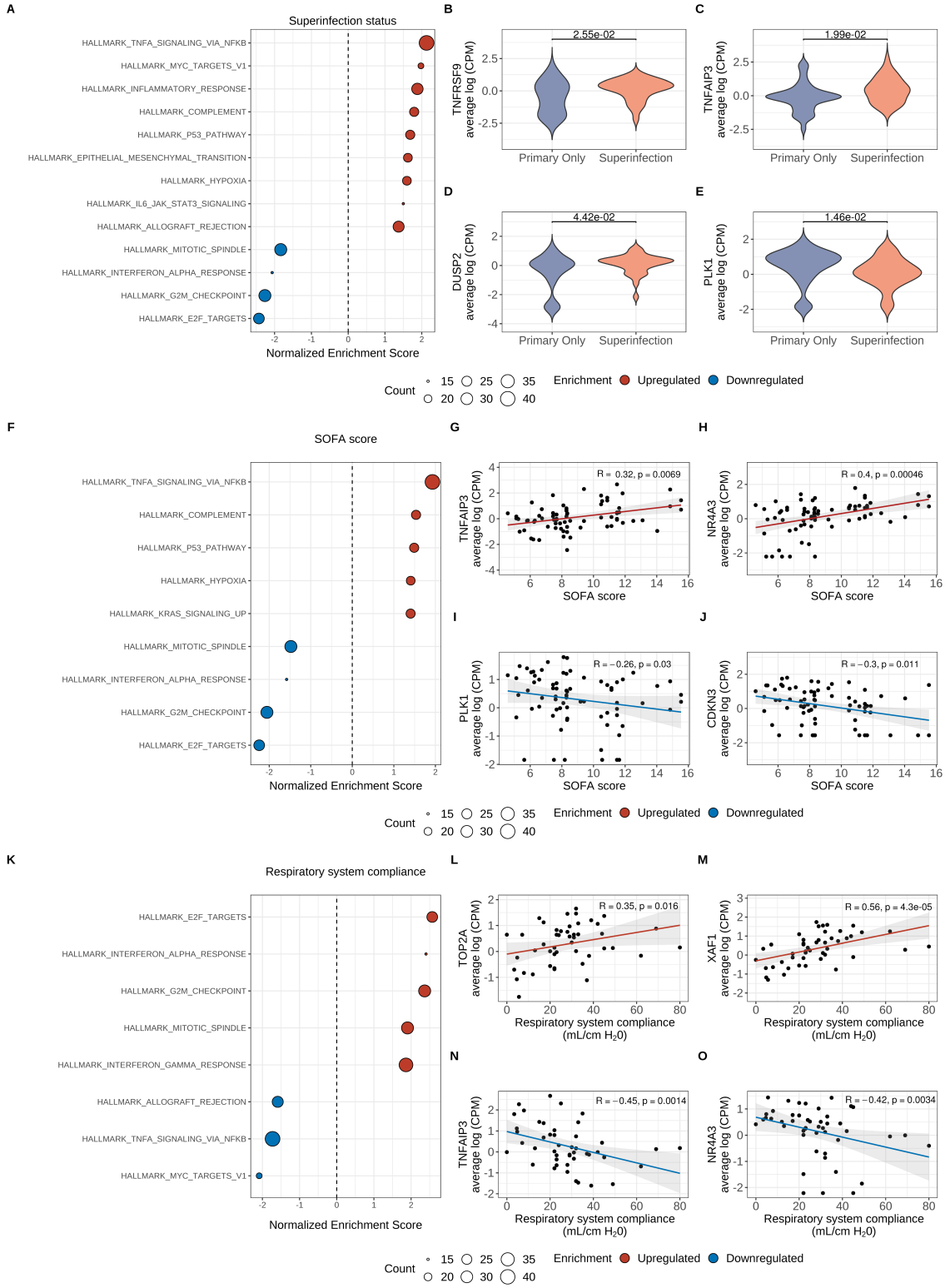
Supplemental Figure 8



Supplemental Figure 8. The early T cell response during severe SARS-CoV-2 pneumonia is dominated by an interferon signaling transcriptional program. (A-B) MA plot of differentially expressed genes in CD4$^+$ T cells (81 samples from 46 patients with COVID-19 pneumonia) (A) and CD8$^+$ T cells (72 samples from 46 patients with COVID-19 pneumonia) (B), comparing early (≤48 hours following intubation) versus late (>48 hours following intubation)

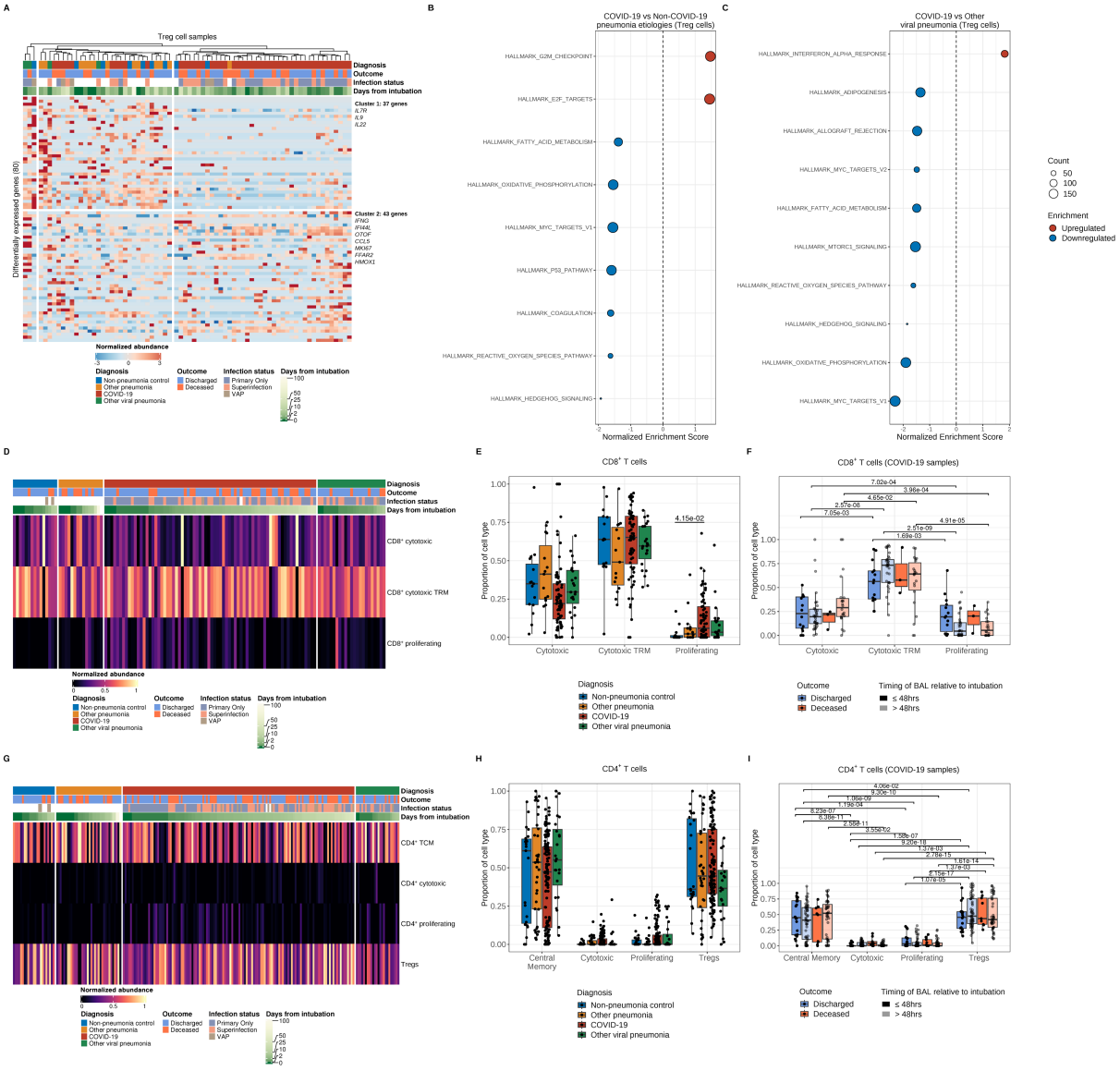
COVID-19 samples. Significantly upregulated genes in early samples are shown in red, and significantly upregulated genes in late samples are shown in blue ($q < 0.05$, likelihood-ratio tests with FDR correction). Genes shown in gray are not significantly differentially expressed. Representative significant genes are annotated. **(C)** Heatmap of longitudinal analysis of interferon-stimulated genes in combined $CD4^+$ and $CD8^+$ T cells of patients with severe SARS-CoV-2 pneumonia. Columns represent unique T cell samples and are color-coded by binary outcome, infection status, severity of illness (SOFA score), cumulative steroid dose (mg of hydrocortisone), C reactive protein (CRP), D-dimer, viral load (Ct value), and ordered by duration of mechanical ventilation. Blanks indicate missing values. **(D)** Comparison of SARS-CoV-2 viral load (Ct value) by binary outcome and BAL sampling time in COVID-19 samples that underwent RNA-sequencing ($q < 0.05$, pairwise Wilcoxon rank-sum tests with FDR correction). **(E)** Number of COVID-19 BAL samples grouped by binary outcome and sampling time with a Ct value above limit of detection (>40). **(F)** Correlation analysis of COVID-19 Ct values grouped by binary outcome and duration of mechanical ventilation with Spearman rank correlation coefficient.

Supplemental Figure 9



Supplemental Figure 9. Distinct activation molecular signatures in alveolar CD8$^+$ T cells predict clinical outcomes in patients with severe SARS-CoV-2 pneumonia. (A-O) Gene set enrichment analysis (GSEA) of COVID-19 samples after performing correlation analysis of differentially expressed genes in CD8$^+$ T cells and clinical variables of interest with Spearman rank correlation coefficient computation. Count denotes pathway size after removing genes not detected in the expression dataset. Enrichment denotes significant ($q < 0.25$ with FDR correction) upregulated (red) and downregulated (blue) pathways by normalized enrichment score. Leading edge analysis reveals selected core genes driving pathway enrichment signal in clinical variables, which are annotated for the superinfection (B-E), severity of illness (SOFA score) (G-J), and respiratory system compliance (L-O) variables.

Supplemental Figure 10

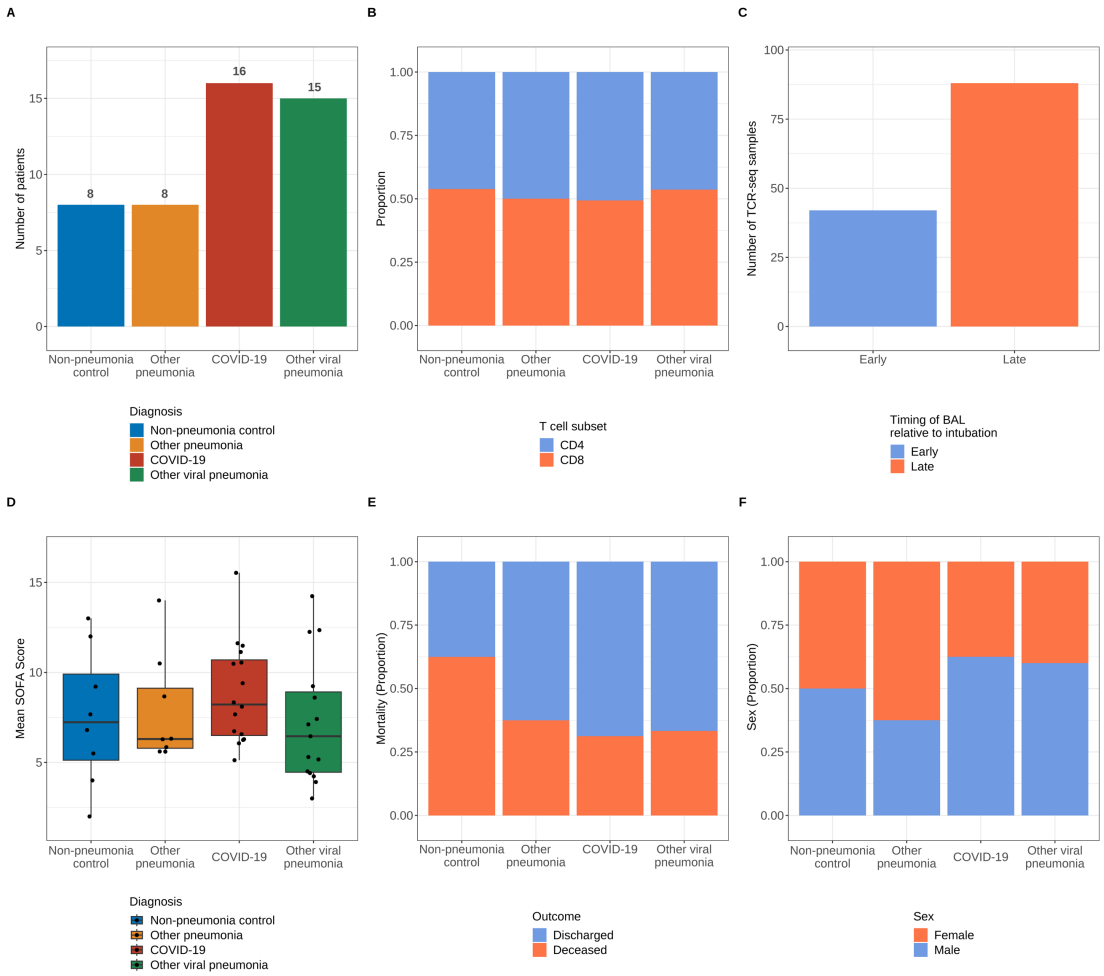


Supplemental Figure 10. Cell-type deconvolution of bulk RNA-sequencing reveals a dominant activated memory phenotype of alveolar T cells during severe pneumonia. (A) *K*-means clustering of 80 differentially expressed genes (q < 0.05, likelihood-ratio test with FDR correction) in Treg cells across pneumonia diagnoses. Columns represent unique samples and column headers are color-coded by diagnosis, binary outcome (whether a given patient was discharged or died during hospitalization), duration of mechanical ventilation (blanks indicate

chronically ventilated patients), and infection status (presence or absence of bacterial superinfection in patients with COVID-19 or other viral pneumonia). The VAP (ventilator-associated pneumonia) flag designates samples from non-pneumonia controls or patients with COVID-19 or other viral pneumonia who cleared the virus and then developed a bacterial pneumonia. Samples were clustered using Ward's minimum variance clustering method. Representative genes are shown for each cluster. **(B-C)** Gene set enrichment analysis (GSEA) of Hallmark gene sets for the pairwise comparison between COVID-19 samples and combined non-COVID-19 samples (non-pneumonia control, other pneumonia, and other viral pneumonia) (B) or the pairwise comparison between COVID-19 samples and other viral pneumonia samples (C). Count denotes pathway size after removing genes not detected in the expression dataset. Enrichment denotes significant (q < 0.25 with FDR correction) upregulated (red) and downregulated (blue) pathways by normalized enrichment score. **(D)** Heatmap demonstrating the proportion of alveolar $CD8^+$ T cell subsets from deconvolution analysis. Columns represent unique samples grouped by diagnosis and are ordered by duration of mechanical ventilation. Column headers are color-coded by diagnosis, binary outcome, duration of mechanical ventilation (blanks indicate chronically ventilated patients), and infection status (presence or absence of bacterial superinfection in patients with COVID-19 or other viral pneumonia). The VAP (ventilator-associated pneumonia) flag designates samples from non-pneumonia controls or patients with COVID-19 or other viral pneumonia who cleared the virus and then developed a bacterial pneumonia. **(E)** Proportion of alveolar $CD8^+$ T cell subsets across different pneumonia etiologies. (q < 0.05, pairwise Wilcoxon rank-sum tests with FDR correction). **F)** Proportion of alveolar $CD8^+$ T cell subsets by outcome and timing of BAL fluid sampling in COVID-19 patients (q < 0.05, pairwise Wilcoxon rank-sum tests with FDR correction). **(G)** Heatmap demonstrating the proportion of alveolar $CD4^+$ T cell subsets from deconvolution analysis. Columns represent unique samples grouped by diagnosis and are ordered by duration of mechanical ventilation. Column headers are color-coded by diagnosis, binary outcome, duration of mechanical ventilation (blanks

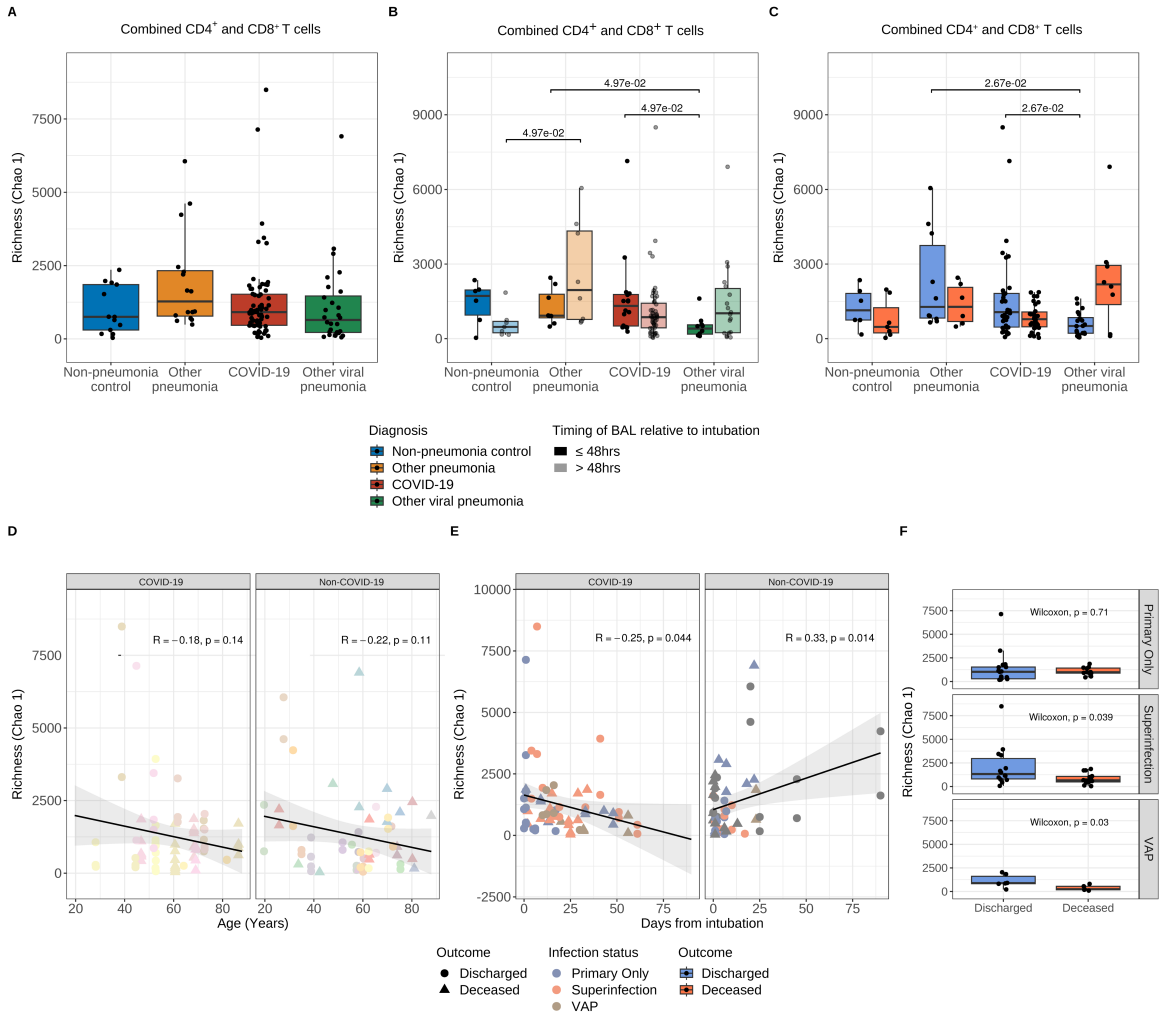
indicate chronically ventilated patients), and infection status (presence or absence of bacterial superinfection in patients with COVID-19 or other viral pneumonia). The VAP (ventilator-associated pneumonia) flag designates samples from non-pneumonia controls or patients with COVID-19 or other viral pneumonia who cleared the virus and then developed a bacterial pneumonia. **(H)** Proportion of alveolar $CD4^+$ T and Treg cell subsets across different pneumonia etiologies ($q < 0.05$, pairwise Wilcoxon rank-sum tests with FDR correction). **(I)** Proportion of alveolar $CD4^+$ T cell subsets by binary outcome and timing of BAL fluid sampling in COVID-19 patients ($q < 0.05$, pairwise Wilcoxon rank-sum tests with FDR correction).

Supplemental Figure 11



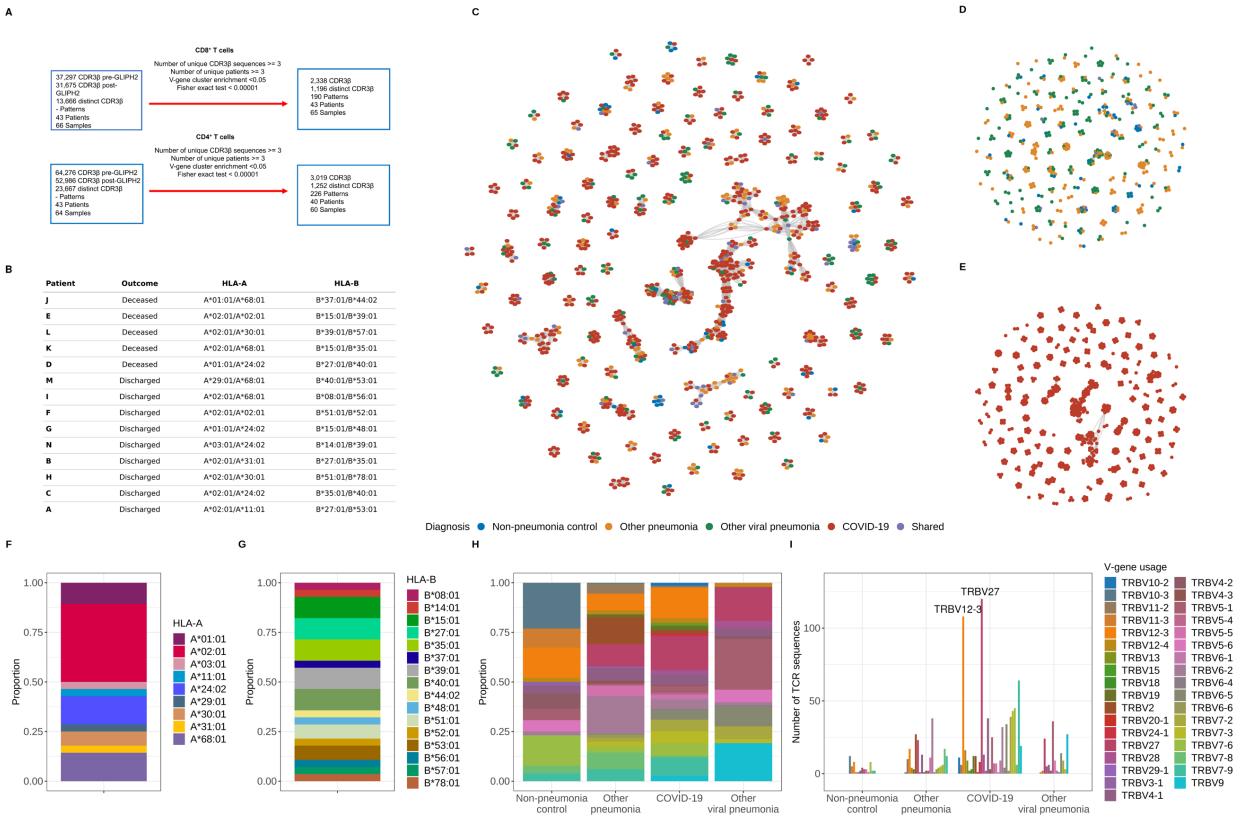
Supplemental Figure 11. Composition and demographics of bulk TCR-sequencing samples. (A) Number of participants grouped by diagnosis. **(B)** Proportion of samples grouped by T cell subset and diagnosis. **(C)** Number of samples categorized by timing of BAL. **(D)** Distribution of SOFA scores by patient. Nonsignificant after pairwise Wilcoxon rank-sum tests with FDR correction. **(E)** Mortality by patient. Nonsignificant after pairwise χ^2 tests for homogeneity of proportions with FDR correction. **(F)** Sex by patient. Nonsignificant after pairwise χ^2 tests for homogeneity of proportions with FDR correction.

Supplemental Figure 12



Supplemental Figure 12. SARS-CoV-2 pneumonia complicated by secondary bacterial pneumonia is associated with lower TCR repertoire diversity. (A-C) Alpha diversity estimation with Chao 1 in combined alveolar $CD4^+$ and $CD8^+$ T cells in patients grouped by diagnosis (A), timing of BAL fluid collection relative to intubation (B), and binary outcome (C) (q < 0.05, pairwise Wilcoxon rank-sum tests with FDR correction). **(D-E)** Correlation analysis between combined alveolar $CD4^+$ and $CD8^+$ T cell richness (Chao 1) and age (D) and duration of mechanical ventilation (E) using Pearson correlation. Data points are color-coded by unique patients (D) or infection status (E) and shaped according to binary outcome. **(F)** Alpha diversity estimation with Chao1 in combined alveolar $CD8^+$ and $CD4^+$ T cells in patients grouped by binary outcome and infection status (presence or absence of bacterial superinfection in patients with COVID-19 or other viral pneumonia). The VAP (ventilator-associated pneumonia) flag designates samples from non-pneumonia controls or patients with COVID-19 or other viral pneumonia who cleared the virus and then developed a bacterial pneumonia. Wilcoxon rank sum tests p-values are shown.

Supplemental Figure 13



Supplemental Figure 13. Alveolar CD8$^+$ T cell receptor specificity analysis in SARS-CoV-2 pneumonia. (A) GLIPH2 specificity filtering criteria and numbers. **(B)** Inferred HLA-A and HLA-B alleles in patients with COVID-19. **(C-E)** Network analysis of TCR sequences in patients with severe pneumonia and respiratory failure from all diagnosis categories (C), non-COVID-19 groups (D), or COVID-19 (E). Nodes represent unique TCR (CDR3β) sequences and are color-coded by diagnosis. Shared TCR sequences by at least two different pneumonia categories are colored in purple. Edges constitute patterns or specificity groups identified through the GLIPH2 algorithm. **(F-G)** Proportion of HLA-A (F) and HLA-B (G) molecules identified in patients with COVID-19. *n* of patients = 14 **(H-I)** TCRβ (V) gene usage analysis of identified sequences following processing using the GLIPH2 pipeline and cross-matching with the MIRA class I dataset. Proportion of genes by pneumonia category (*n* of patients = non-pneumonia control [4], other pneumonia [7], COVID-

19 [14], and other viral pneumonia [8]) (H) and absolute number of TCRβ sequences for each V region by pneumonia category (I) are shown. Dominant genes are annotated in (I).

Supplemental Figure 14

Supplemental Figure 14. Alveolar CD4$^+$ T cell responses during severe SARS-CoV-2 pneumonia. (A) Proportion of alveolar CD4$^+$ T cell responses by SARS-CoV-2 protein. TCR sequences identified in all samples from patients with COVID-19 were cross-referenced with the MIRA II dataset to identify reactivity against specific SARS-CoV-2 antigens. *n* of patients = 12 and *n* of samples = 22 **(B)** Inferred HLA-DP, HLA-DQ, and HLA-DR alleles in patients with COVID-19. **(C-E)** Network analysis of TCR sequences in patients with severe pneumonia and respiratory failure from all diagnosis categories (C), non-COVID-19 groups (D), or COVID-19 (E). Nodes represent unique TCR (CDR3β) sequences and are color-coded by diagnosis. Shared TCR sequences by at least two different pneumonia categories are colored in purple. Edges constitute

patterns or specificity groups identified through the GLIPH2 algorithm. **(F)** Network analysis of shared TCR sequences recognizing SARS-CoV-2 epitopes. Nodes represent unique patients in the COVID-19 group (labeled here using the lettering scheme from Figure 4G), edges constitute shared TCR sequences by at least two patients mapped to a MIRA class II dataset epitope pool, and width of edges (magnitude) denote total number of shared TCR sequences. Edges are color-coded by SARS-CoV-2 antigens. **(G)** Immunoprevalence of SARS-CoV-2 epitopes in patients with COVID-19 was calculated by counting the number of events when a given epitope was shared by at least two patients. Total counts from all eight identified epitopes are represented as percentage (%). **(H)** Overall number of TCR sequences mapped to a given SARS-CoV-2 epitope in patients with COVID-19 was calculated by counting all events of TCRs recognizing an epitope. Total counts from all eight identified epitopes are represented as percentage (%). * denotes other epitopes are present within MIRA class I dataset peptide pool. **(I-J)** TCRβ (V) gene usage analysis of identified sequences following GLIPH2 pipeline and cross-matching with MIRA class II dataset. Proportion of genes by pneumonia category. *n* of patients = non-pneumonia control (1), other pneumonia (4), COVID-19 (12), and other viral pneumonia (2) (I) and absolute number of TCRβ sequences for each V region by pneumonia category (J) are shown. Dominant genes are annotated in (J). **(K)** Scatter plot of SARS-CoV-2 epitope prevalence in patients with COVID-19 (*n* = 12) and without COVID-19 (unexposed, *n* = non-pneumonia control [1], other pneumonia [4], and other viral pneumonia [2]). Dots are color coded by SARS-CoV-2 antigen. Dot size corresponds to the number of detected TCR sequences recognizing a given antigen. **(L)** SARS-CoV-2 epitope prevalence in samples that underwent $CD4^+$ TCR sequencing grouped by COVID-19 status. Wilcoxon rank sum test *p*-values are shown.

Supplemental Data Files

Supplemental File 1. Differentially expressed genes in $CD8^+$ T cell bulk RNA-sequencing samples from K-means clustering analysis (cluster 1).

Supplemental File 2. Differentially expressed genes in $CD8^+$ T cell bulk RNA-sequencing samples from K-means clustering analysis (cluster 2).

Supplemental File 3. Differentially expressed genes in $CD4^+$ T cell bulk RNA-sequencing samples from K-means clustering analysis (cluster 1).

Supplemental File 4. Differentially expressed genes in $CD4^+$ T cell bulk RNA-sequencing samples from K-means clustering analysis (cluster 2).

Supplemental File 5. Significant gene ontology processes identified in $CD8^+$ T cell bulk RNA-sequencing samples from K-means clustering analysis (cluster 1).

Supplemental File 6. Significant gene ontology processes identified in $CD8^+$ T cell bulk RNA-sequencing samples from K-means clustering analysis (cluster 2).

Supplemental File 7. Significant gene ontology parent terms and terms identified in $CD8^+$ T cell bulk RNA-sequencing samples from K-means clustering analysis (cluster 1).

Supplemental File 8. Significant gene ontology parent terms and terms identified in $CD8^+$ T cell bulk RNA-sequencing samples from K-means clustering analysis (cluster 2).

Supplemental File 9. Significant gene ontology processes identified in $CD4^+$ T cell bulk RNA-sequencing samples from K-means clustering analysis (cluster 1).

Supplemental File 10. Significant gene ontology parent terms and terms identified in $CD4^+$ T cell bulk RNA-sequencing samples from K-means clustering analysis (cluster 1).

Supplemental File 11. GLIPH2 analysis establishing specificity groups from $CD4^+$ T cell bulk TCR-sequencing samples.

Supplemental File 12. GLIPH2 analysis establishing specificity groups from $CD8^+$ T cell bulk TCR-sequencing samples.

Supplemental File 13. Cross-reference of $CD4^+$ T cell repertoire sequences with Multiplex Identification of Antigen-Specific T-Cell Receptors Assay (MIRA) dataset.

Supplemental File 14. Cross-reference of $CD8^+$ T cell repertoire sequences with Multiplex Identification of Antigen-Specific T-Cell Receptors Assay (MIRA) dataset.

Supplemental File 15. (A) Multiplex Identification of Antigen-Specific T-Cell Receptors Assay (MIRA) peptide deconvolution of Class I targets. **(B)** Hierarchical distribution of Class I immunodominant epitopes by patient from the bulk TCR-sequencing subset. **(C)** Pairwise similarity and average conservation score estimation between SARS-CoV-2 and HCoV epitopes.

Supplemental File 16. Anonymized metadata by patient and samples for bulk RNA-sequencing and TCR-sequencing analysis.

Supplemental File 17. Raw counts for $CD4^+$ T cell bulk RNA-sequencing samples.

Supplemental File 18. Raw counts for $CD8^+$ T cell bulk RNA-sequencing samples.

Supplemental File 19. Raw counts for Treg cell bulk RNA-sequencing samples.

Supplemental File 20. MiXCR-processed raw sequencing files with TCR repertoire data for $CD4^+$ T cell bulk TCR-sequencing samples.

Supplemental File 21. MiXCR-processed raw sequencing files with TCR repertoire data for $CD8^+$ T cell bulk TCR-sequencing samples.

Supplemental File 22. Differential expression analysis in $CD8^+$ T cells of pairwise comparison between COVID-19 samples and combined non-COVID-19 samples from Figure 2B.

Supplemental File 23. Differential expression analysis in $CD4^+$ T cells of pairwise comparison between COVID-19 samples and combined non-COVID-19 samples from Figure 3B.

Supplemental File 24. Differential expression analysis in $CD8^+$ T cells of pairwise comparison between COVID-19 samples and other viral pneumonia samples from Supplemental Figure 6B

Supplemental File 25. Differential expression analysis in $CD4^+$ T cells of pairwise comparison between COVID-19 samples and other viral pneumonia samples from Supplemental Figure 7B

Supplemental File 26. Differential expression analysis in Treg cells of pairwise comparison between COVID-19 samples and combined non-COVID-19 samples from Supplemental Figure 10B.

Supplemental File 27. Differential expression analysis in Treg cells of pairwise comparison between COVID-19 samples and other viral pneumonia samples from Supplemental Figure 10C

Supplemental File 28. Differential expression analysis in $CD4^+$ T cells of pairwise comparison between early (≤48 hours after intubation) and late (>48 hours after intubation) COVID-19 samples from Supplemental Figure 8A.

Supplemental File 29. Differential expression analysis in $CD8^+$ T cells of pairwise comparison between early (≤48 hours after intubation) and late (>48 hours after intubation) COVID-19 samples from Supplemental Figure 8B.

Supplemental File 30. The NU SCRIPT Study Investigators.