Cell Reports Medicine, Volume *3*

Supplemental information

Longitudinal characterization of circulating neutrophils uncovers phenotypes associated with severity in hospitalized COVID-19 patients

Thomas J. LaSalle, Anna L.K. Gonye, Samuel S. Freeman, Paulina Kaplonek, Irena Gushterova, Kyle R. Kays, Kasidet Manakongtreecheep, Jessica Tantivit, Maricarmen Rojas-Lopez, Brian C. Russo, Nihaarika Sharma, Molly F. Thomas, Kendall M. Lavin-Parsons, Brendan M. Lilly, Brenna N. Mckaig, Nicole C. Charland, Hargun K. Khanna, Carl L. Lodenstein, Justin D. Margolin, Emily M. Blaum, Paola B. Lirofonis, Or-Yam Revach, Arnav Mehta, Abraham Sonny, Roby P. Bhattacharyya, Blair Alden Parry, Marcia B. Goldberg, Galit Alter, Michael R. Filbin, Alexandra-Chloé Villani, Nir Hacohen, and Moshe Sade-Feldman

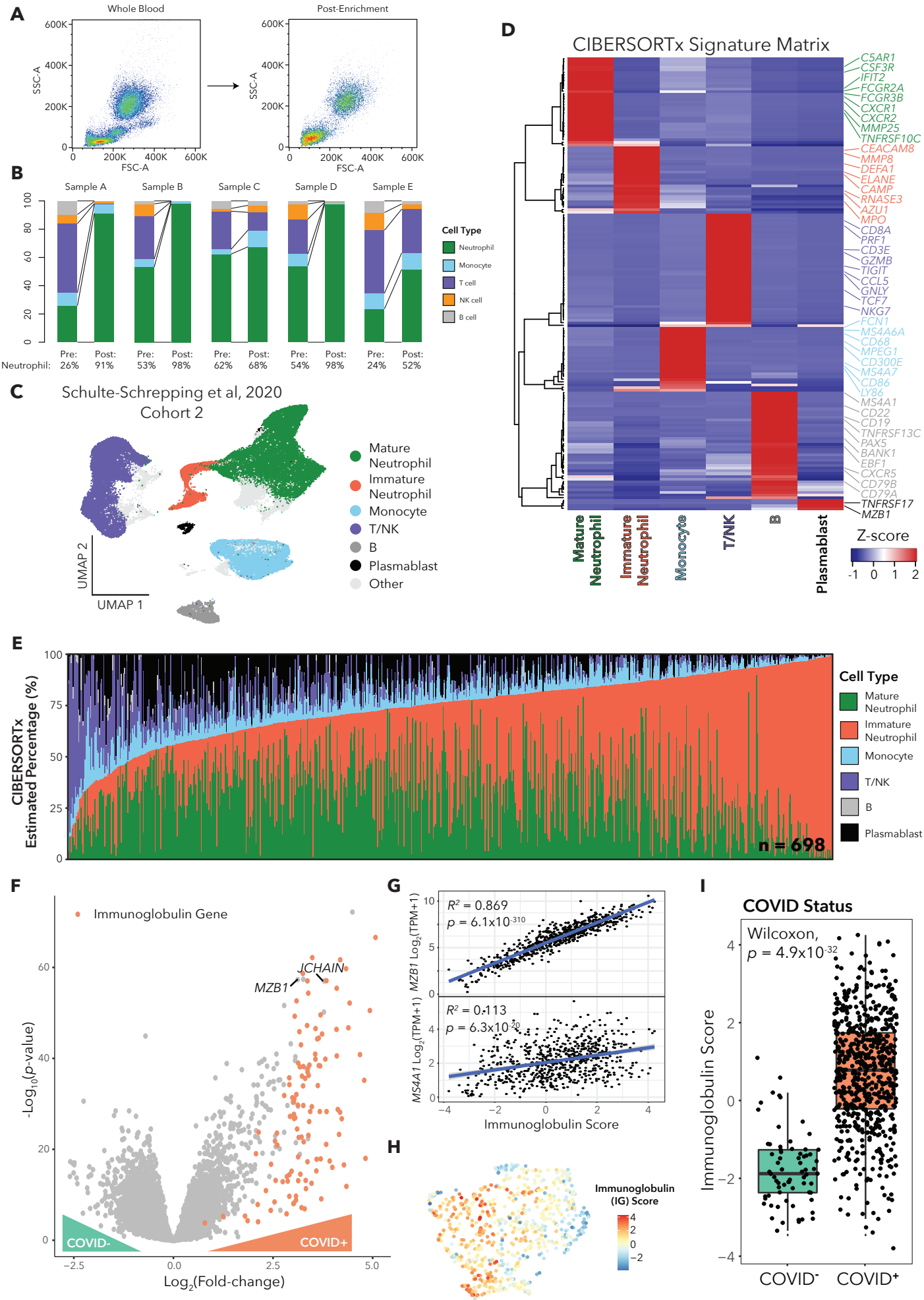
A
Whole Blood
Post-Enrichment
SSC-A
FSC-A
B
Sample A
Sample B
Sample C
Sample D
Sample E
Neutrophil:
Pre: 26%
Post: 91%
Pre: 53%
Post: 98%
Pre: 62%
Post: 68%
Pre: 54%
Post: 98%
Pre: 24%
Post: 52%
Cell Type
Neutrophil
Monocyte
T cell
NK cell
B cell
C
Schulte-Schrepping et al, 2020
Cohort 2
UMAP 1
UMAP 2
Mature Neutrophil
Immature Neutrophil
Monocyte
T/NK
B
Plasmablast
Other
D
CIBERSORTx Signature Matrix
C5AR1
CSF3R
IFIT2
FCGR2A
FCGR3B
CXCR1
CXCR2
MMP25
TNFRSF10C
CEACAM8
MMP8
DEFA1
ELANE
CAMP
RNASE3
AZU1
MPO
CD8A
PRF1
CD3E
GZMB
TIGIT
CCL5
GNLY
TCF7
NKG7
FCN1
MS4A6A
CD68
MPEG1
CD300E
MS4A7
CD86
LY86
MS4A1
CD22
CD19
TNFRSF13C
PAX5
BANK1
EBF1
CXCR5
CD79B
CD79A
TNFRSF17
MZB1
Mature Neutrophil
Immature Neutrophil
Monocyte
T/NK
B
Plasmablast
Z-score
E
CIBERSORTx Estimated Percentage (%)
n = 698
Cell Type
Mature Neutrophil
Immature Neutrophil
Monocyte
T/NK
B
Plasmablast
F
Immunoglobulin Gene
JCHAIN
MZB1
-Log10(p-value)
Log2(Fold-change)
COVID-
COVID+
G
R² = 0.869
p = 6.1x10-310
R² = 0.113
p = 6.3x10-20
MZB1 Log2(TPM+1)
MS4A1 Log2(TPM+1)
Immunoglobulin Score
H
Immunoglobulin (IG) Score
I
COVID Status
Wilcoxon,
p = 4.9x10-32
Immunoglobulin Score
COVID-
COVID+

Figure S1. Estimation of Sample Purity Using CIBERSORTx and Regressing Plasmablast Contamination, Related to Figure 1.

(A) Representative forward-scatter vs. side-scatter flow cytometry plot of healthy donor whole blood pre-enrichment (left) and post-enrichment (right) for neutrophils following partial red blood cell lysis and gating to remove doublets and some debris.

(B) Bar plots displaying the composition of viable cells from healthy blood samples pre- and post-neutrophil enrichment with red blood cell lysis. Composition was determined by flow cytometry, broken down according to major lineage (neutrophil, monocyte, T cell, NK cell, B cell). Percentages below indicate the percentage of neutrophils in each sample. Average viable neutrophil percentage pre-enrichment: 44%. Average viable neutrophil percentage post-enrichment: 81%.

(C) UMAP (Uniform Manifold Approximation and Projection) plot of single-cell RNA-seq data of fresh whole blood from COVID-19-positive patients and controls from Schulte-Schrepping et al. Cohort 2. Cells are colored according to their major lineage (neutrophil, monocyte, T/NK, B, Plasmablast, Other), with neutrophils split between mature and immature.

(D) CIBERSORTx scaled expression signature matrix, generated from pseudobulked Schulte-Schrepping cell types using the “Create Signature Matrix” module, used to deconvolute the neutrophil-enriched bulk RNA-seq samples.

(E) Distribution of the CIBERSORTx estimated cell type percentages for each sample. Each column is a single sample, and columns are ordered by increasing Total Neutrophil content (Immature Neutrophil Fraction + Mature Neutrophil Fraction).

(F) Volcano plot showing differentially expressed genes between COVID-19-positive and COVID-19-negative samples on Day 0. Immunoglobulin genes (all of which are used in the score) are highlighted in red, and plasmablast marker genes *MZB1* and *JCHAIN* are annotated.

(G) Scatter plots and linear regression of immunoglobulin score versus $\log_2$(TPM+1) expression of (top) plasmablast marker gene *MZB1* and (bottom) B cell marker *MS4A1*. P values were obtained using lm() in R, which uses a t-test with the null hypothesis that the slope is zero.

(H) UMAP plot of all bulk RNA-seq samples color-coded by immunoglobulin score.

(I) Box plots comparing immunoglobulin score across COVID-19 status for all time points. Wilcoxon rank-sum test performed to determine significance.

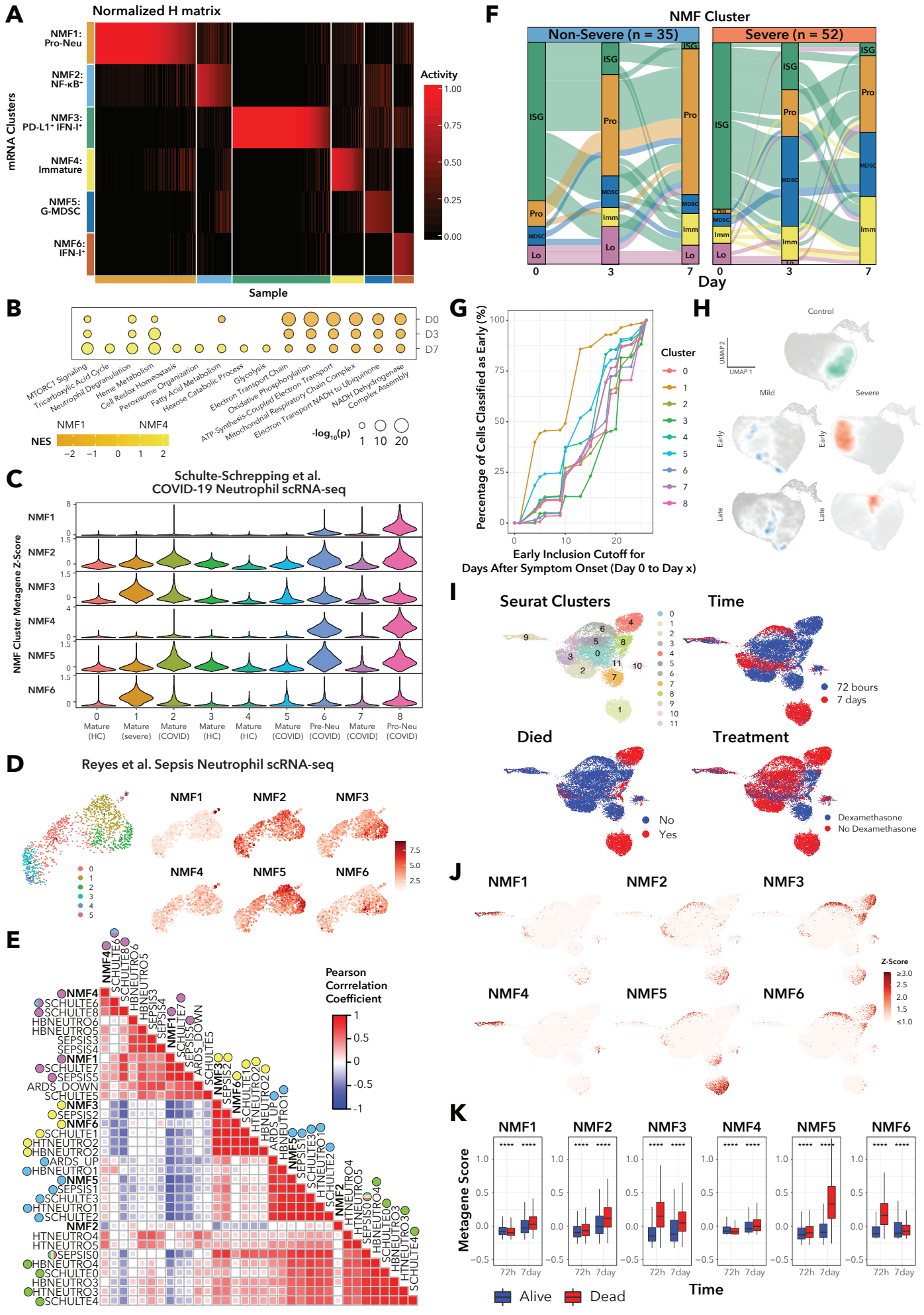
A
Normalized H matrix
mRNA Clusters
NMF1: Pro-Neu
NMF2: NF-κB+
NMF3: PD-L1+ IFN-I+
NMF4: Immature
NMF5: G-MDSC
NMF6: IFN-I+
Activity
1.00
0.75
0.50
0.25
0.00
Sample
B
D0
D3
D7
MTORC1 Signaling
Tricarboxylic Acid Cycle
Neutrophil Degranulation
Heme Metabolism
Cell Redox Homeostasis
Peroxisome Organization
Fatty Acid Metabolism
Hexose Catabolic Process
Glycolysis
Electron Transport Chain
Oxidative Phosphorylation
ATP-Synthesis-Coupled Electron Transport
Mitochondrial Respiratory Chain Complex
Electron Transport NADH to Ubiquinone
NADH Dehydrogenase
Complex Assembly
NMF1
NMF4
NES
−2 −1 0 1 2
-log10(p)
1 10 20
C
Schulte-Schrepping et al.
COVID-19 Neutrophil scRNA-seq
NMF Cluster Metagene Z-Score
NMF1
NMF2
NMF3
NMF4
NMF5
NMF6
0 Mature (HC)
1 Mature (severe)
2 Mature (COVID)
3 Mature (HC)
4 Mature (HC)
5 Mature (COVID)
6 Pre-Neu (COVID)
7 Mature (COVID)
8 Pro-Neu (COVID)
D
Reyes et al. Sepsis Neutrophil scRNA-seq
NMF1
NMF2
NMF3
NMF4
NMF5
NMF6
E
Pearson Corrrelation Coefficient
1
0.5
0
-0.5
-1
F
NMF Cluster
Non-Severe (n = 35)
Severe (n = 52)
ISG
Pro
MDSC
Imm
Lo
0
3
7
Day
G
Percentage of Cells Classified as Early (%)
Cluster
Early Inclusion Cutoff for
Days After Symptom Onset (Day 0 to Day x)
H
Control
Mild
Severe
Early
Late
UMAP 1
UMAP 2
I
Seurat Clusters
Time
72 bours
7 days
Died
No
Yes
Treatment
Dexamethasone
No Dexamethasone
J
NMF1
NMF2
NMF3
NMF4
NMF5
NMF6
Z-Score
K
NMF1
NMF2
NMF3
NMF4
NMF5
NMF6
Metagene Score
72h
7day
Time
Alive
Dead

Figure S2. Characterization of NMF Clustering Results and Comparisons with Single-Cell RNA-seq Data, Related to Figure 2.

(A) NMF Normalized H matrix of neutrophil-enriched bulk RNA-seq samples with CIBERSORTx Estimated Total Neutrophil Percentage above 50%. Clustering identified 6 subtypes. Activity corresponds to the probability that a sample is included in a given cluster. Samples are ordered according to activity value within a given cluster.

(B) Gene set enrichment analysis on genes differentially expressed between COVID-19-positive samples in clusters NMF1 versus NMF4 on Days 0, 3, and 7, specifically highlighting metabolic pathways. Bubble size corresponds to $-\log_{10}(p)$ and color corresponds to NES.

(C) Violin plots of the metagene z-score for each NMF cluster signature across the Schulte-Schrepping single-cell fresh whole blood neutrophil data.

(D) UMAPs of single-cell RNA-seq data of fresh whole-blood neutrophils from sepsis patients and healthy controls from Reyes *et al*. 2021. UMAPs are color-coded by Seurat cluster (left) and NMF cluster metagene scores (right) from each NMF cluster's marker genes.

(E) Pairwise Pearson correlation heatmap for the Z-scores of each gene set on all samples in the cohort. Color-coded dots indicate the network group membership from Figure 2C.

(F) Alluvial diagrams displaying the change in NMF cluster membership over time for patients who had all three blood draws which all passed quality control, split by non-severe (n = 35 patients) and severe (n = 52 patients). "ISG" indicates NMF3 and NMF6, "Pro" indicates NMF1, "MDSC" indicates NMF5, "Immature" indicates NMF4, and "Lo" indicates samples with CIBERSORTx Estimated Total Neutrophil Percentage less than 50%.

(G) Line chart showing the percentage of cells within a cluster that are categorized as early as a function of inclusion cutoff for days following symptom onset for the Schulte-Schrepping Cohort 2 fresh whole-blood neutrophil data. Data shown are for all COVID-19-positive samples. The early cutoff was defined as 0-10 days after disease onset by Schulte-Schrepping et al.

(H) Density plots of healthy controls, disease severity, and time point overlaid on the Schulte-Schrepping Cohort 2 neutrophil UMAP, classifying early time points as days 0-13 following symptom onset.

(I) UMAPs of single-cell RNA-seq data of circulating neutrophils in COVID-19 from Sinha et al. 2021. UMAPs are color-coded by Seurat clustering (top left), time point (top left), disease severity (bottom left), and dexamethasone treatment status (bottom right).

(J) UMAPs of circulating neutrophils from Sinha et al. color-coded by NMF cluster metagene scores.

(K) Box plots comparing NMF metagene scores between COVID-19 patients that survived severe disease and those that died in the Sinha et al. cohort. Quadruple asterisks indicate $p \leq 0.0001$. P values are for the Wilcoxon rank-sum test.

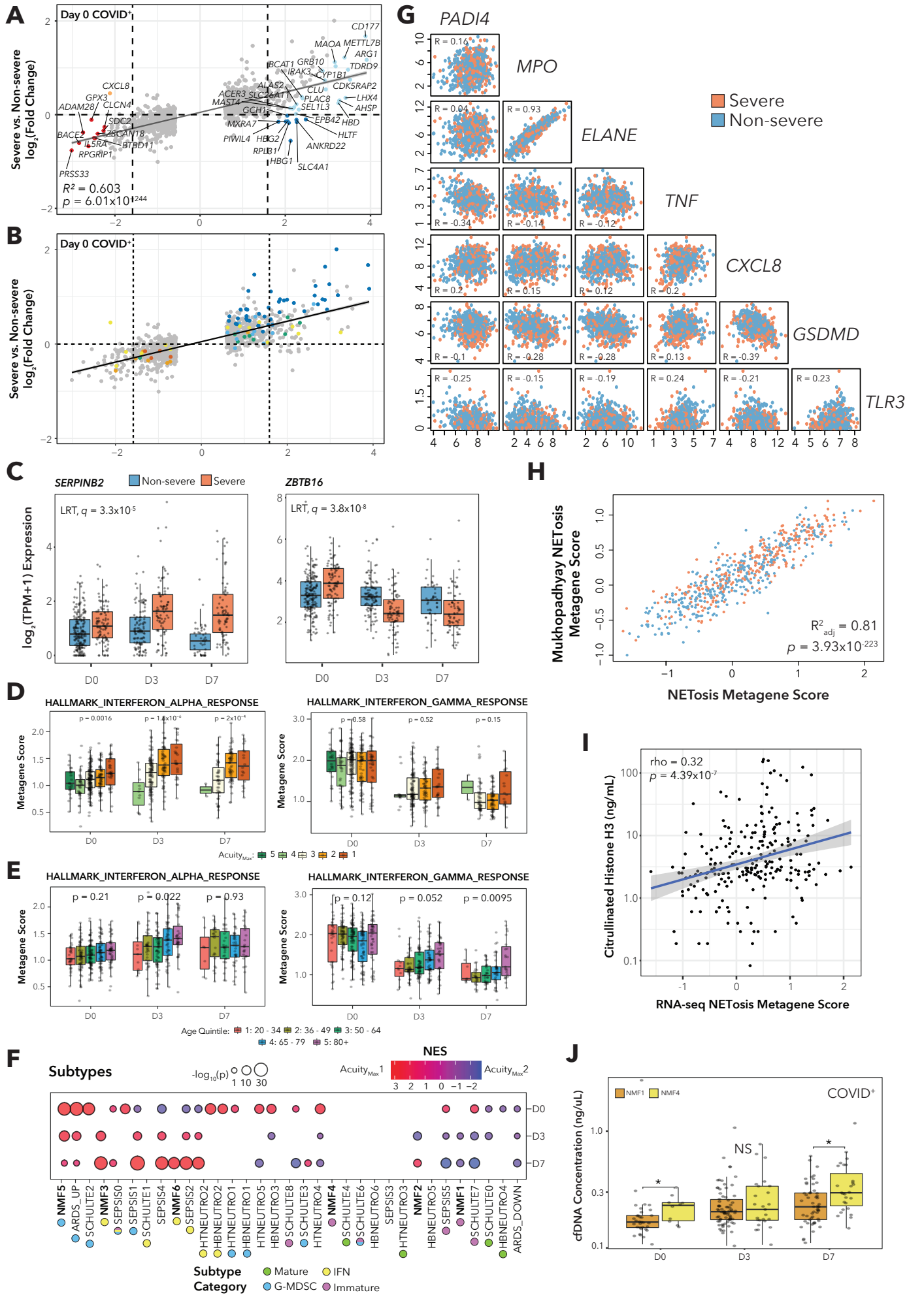
A
Day 0 COVID+
Severe vs. Non-severe log2(Fold Change)
R² = 0.603
p = 6.01x10-244
CD177
MAOA METTL7B
ARG1
TDRD9
IBCAT1 GRB10
IRAK3
CYP1B1
CDK5RAP2
ALAS2
CLU
PLAC8
LHX4
AHSP
ACER3 SLC25A37
MAST4
GCH1
SEL1L3
EPB42 HBD
MXRA7
PIWIL4 HBG2
RPL31
HBG1
HLTF
ANKRD22
SLC4A1
CXCL8
GPX3
ADAM28 CLCN4
SDC2
BACE2 ZSCAN18
IL5RA BTBD11
RPGRIP1
PRSS33
B
Day 0 COVID+
Severe vs. Non-severe log2(Fold Change)
G
PADI4
MPO
ELANE
TNF
CXCL8
GSDMD
TLR3
Severe
Non-severe
C
SERPINB2
Non-severe
Severe
ZBTB16
log2(TPM+1) Expression
LRT, q = 3.3x10-5
LRT, q = 3.8x10-8
D0
D3
D7
H
Mukhopadhyay NETosis Metagene Score
R²adj = 0.81
p = 3.93x10-223
NETosis Metagene Score
D
HALLMARK_INTERFERON_ALPHA_RESPONSE
HALLMARK_INTERFERON_GAMMA_RESPONSE
Metagene Score
p = 0.0016
p = 1.8x10-4
p = 2x10-4
p = 0.58
p = 0.52
p = 0.15
Acuity Max: 5 4 3 2 1
I
rho = 0.32
p = 4.39x10-7
Citrullinated Histone H3 (ng/mL)
RNA-seq NETosis Metagene Score
E
HALLMARK_INTERFERON_ALPHA_RESPONSE
HALLMARK_INTERFERON_GAMMA_RESPONSE
Metagene Score
p = 0.21
p = 0.022
p = 0.93
p = 0.12
p = 0.052
p = 0.0095
Age Quintile: 1: 20 - 34 2: 36 - 49 3: 50 - 64 4: 65 - 79 5: 80+
F
Subtypes
-log10(p) 1 10 30
NES
Acuity Max 1
Acuity Max 2
3 2 1 0 −1 −2
D0
D3
D7
NMF5 ARDS_UP SCHULTE2 NMF3 SEPSIS0 SEPSIS1 SCHULTE1 SEPSIS4 NMF6 SEPSIS2 HTNEUTRO2 HBNEUTRO2 HTNEUTRO1 HBNEUTRO1 HTNEUTRO5 HBNEUTRO3 SCHULTE8 SCHULTE3 HTNEUTRO4 NMF4 SCHULTE4 SCHULTE6 HBNEUTRO6 SEPSIS3 HTNEUTRO3 NMF2 HBNEUTRO5 SEPSIS5 NMF1 SCHULTE7 SCHULTE0 HBNEUTRO4 ARDS_DOWN
Subtype Category
Mature
IFN
G-MDSC
Immature
J
NMF1
NMF4
COVID+
cfDNA Concentration (ng/uL)
NS
*
*
D0
D3
D7

Figure S3. Genes and Pathways that Vary with Time According to Severity, and Patterns of NETosis Scores in RNA and Plasma, Related to Figures 2, 3, and 4.

(A) Scatter plots and linear regression comparing $\log_2$(fold-change) of COVID-19-positive severe versus non-severe on Day 0 to $\log_2$(fold-change) of ARDS versus Healthy Volunteers at its only time point. For each gene, the difference between the ARDS and COVID-19 fold-changes was calculated. Genes with a difference in fold-change greater than two standard deviations are color-coded according to the legend.

(B) Scatter plot from (A), color-coded according to whether the ARDS microarray differentially expressed gene is a COVID-19 NMF cluster marker gene.

(C) Box plots of $\log_2$(TPM+1) expression over time of *SERPINB2* and *ZBTB16*, two genes which show significant interactions between Day and $\text{Severity}_{\text{Max}}$ according to the DESeq2 Likelihood Ratio Test.

(D) Box plots displaying (top) the HALLMARK_INTERFERON_ALPHA_RESPONSE metagene score and (bottom) the HALLMARK_INTERFERON_GAMMA_RESPONSE metagene score, separated by Day and $\text{Acuity}_{\text{Max}}$. Indicated p values are for the Kruskal-Wallis test within each Day.

(E) Box plots displaying (top) the HALLMARK_INTERFERON_ALPHA_RESPONSE metagene score and (bottom) the HALLMARK_INTERFERON_GAMMA_RESPONSE metagene score, separated by Day and Age quintile. Indicated p values are for the Kruskal-Wallis test within each Day.

(F) Gene set enrichment analysis for the differentially expressed genes between COVID-19-positive $\text{Acuity}_{\text{Max}}1$ (death) and $\text{Acuity}_{\text{Max}}2$ (intubation with survival) patients on Days 0, 3, and 7. Gene sets correspond to the neutrophil states in Figure 2F. Bubble size is scaled to $-\log_{10}$(p-value) and color corresponds to normalized enrichment score (NES).

(G) Scatter plots comparing $\log_2$(TPM+1) expression of genes contributing to the NETosis metagene score (*PADI4*, *MPO*, *ELANE*, *TNF*, *CXCL8*, *GSDMD*, *TLR3*). Points are color-coded according to $\text{Severity}_{\text{Max}}$. Pearson correlation coefficients are indicated.

(H) Scatter plot comparing the NETosis metagene score to a previously defined NETosis gene set from Mukhopadhyay et al. (*CR1*, *ITGAM*, *CFH*, *CFB*, *C5*, *C5AR1*, *C3*, *CFP*, *MPO*, *ELANE*, *CTSG*, *HMGB1*, *AGER*, *TLR2*, *TLR4*, *H4C1*, *TF*, *TFPI*, *F2*, *FGB*, *PLG*, *VWF*, *PF4*, *CCL5*, *DNASE1*, *ITGB2*, *CD33*, *CEACAM8*). R^2 and p value determined using lm() in R.

(I) Scatter plots comparing the RNA-seq NETosis Metagene Score versus citrullinated histone H3 in patient plasma as measured by ELISA. Linear regression was calculated using the lm() function in R, and rho refers to the Spearman's rank correlation coefficient.

(J) Box plots comparing cell-free DNA concentration in plasma from COVID-19-positive patients across time, separated by NMF1 versus NMF4. Single asterisk denotes Wilcoxon rank-sum test $p < 0.05$.

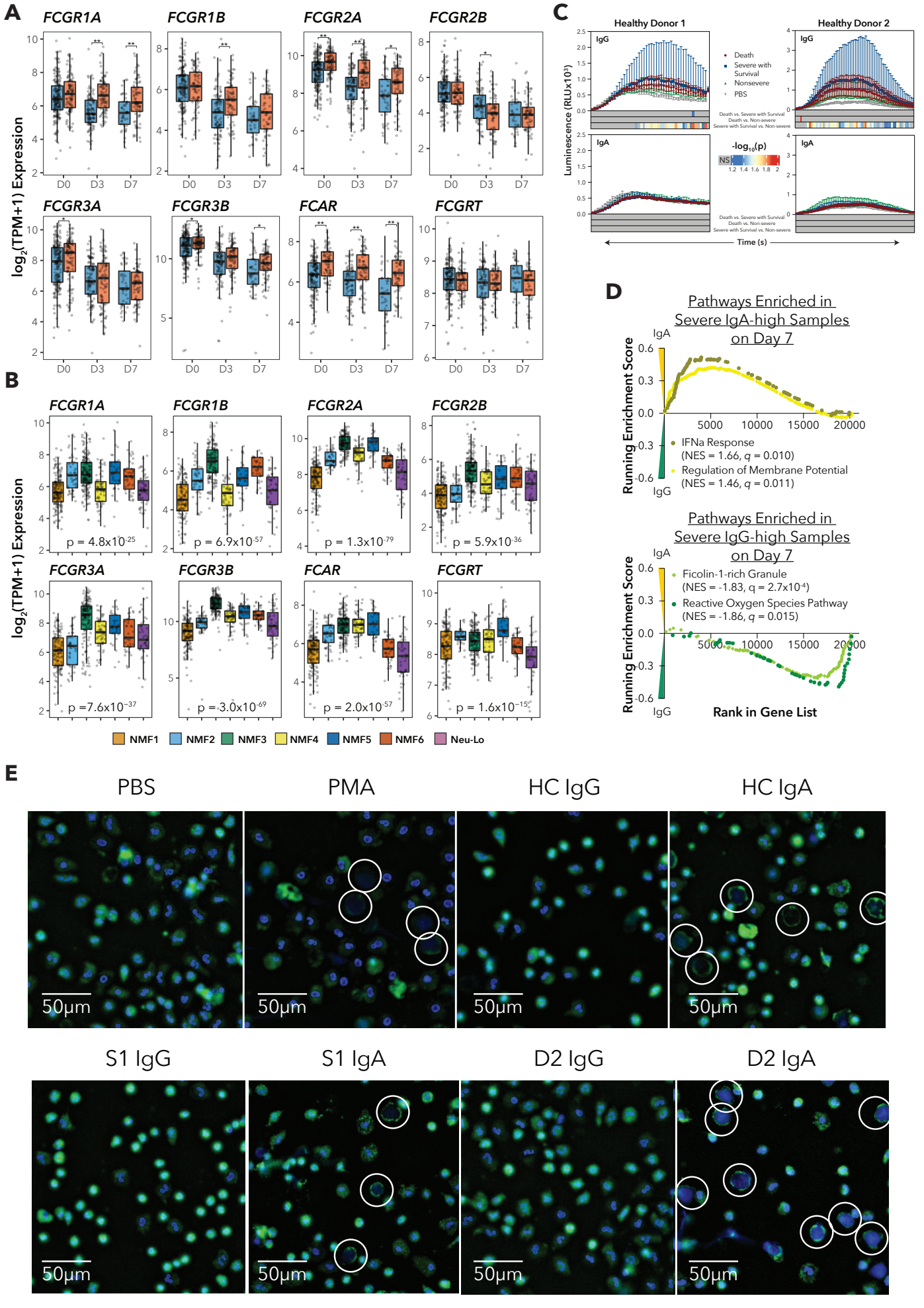
A
FCGR1A
FCGR1B
FCGR2A
FCGR2B
FCGR3A
FCGR3B
FCAR
FCGRT
log2(TPM+1) Expression
D0
D3
D7
B
p = 4.8x10-25
p = 6.9x10-57
p = 1.3x10-79
p = 5.9x10-36
p =7.6x10-37
p =3.0x10-69
p = 2.0x10-57
p = 1.6x10-15
NMF1
NMF2
NMF3
NMF4
NMF5
NMF6
Neu-Lo
C
Healthy Donor 1
Healthy Donor 2
IgG
IgA
Luminescence (RLUx10³)
Death
Severe with Survival
Nonsevere
PBS
-log10(p)
NS
Time (s)
D
Pathways Enriched in Severe IgA-high Samples on Day 7
Running Enrichment Score
IFNa Response (NES = 1.66, q = 0.010)
Regulation of Membrane Potential (NES = 1.46, q = 0.011)
Pathways Enriched in Severe IgG-high Samples on Day 7
Ficolin-1-rich Granule (NES = -1.83, q = 2.7x10-4)
Reactive Oxygen Species Pathway (NES = -1.86, q = 0.015)
Rank in Gene List
E
PBS
PMA
HC IgG
HC IgA
S1 IgG
S1 IgA
D2 IgG
D2 IgA
50μm

Figure S4. Fc Receptor Expression, and Differential Effects of IgG versus IgA Antibodies on Neutrophil Effector Functions, Related to Figure 5.

(A)-(B) Box plots of $\log_2$(TPM+1) expression of Fc receptors (*FCGR1A, FCGR1B, FCGR2A, FCGR2B, FCGR3A, FCGR3B, FCAR, FCGRT*) across (A) Day and $\text{Severity}_{\text{Max}}$ and (B) NMF cluster. P values indicate the Kruskal-Wallis test: *: $p < 0.05$, **: $p < 0.001$.

(C) Point-range plots showing the luminescence of the reactive oxygen species reagent, luminol, over time when neutrophils from two healthy donors are exposed to IgG:S or IgA:S immune complexes using purified IgG and IgA antibodies from serum of patients who died (n = 12), patients with severe disease who survived (n = 12), and patients with non-severe disease (n = 12), or PBS. Point ranges are plotted as median +/- interquartile range. Color bar beneath each plot displays the log-transformed P values for the Wilcoxon rank-sum test between (top) Death vs. Severe with survival, (middle) Death vs. Non-severe, and (bottom) Severe with survival vs. Non-severe values at each time point, with gray values indicating no significant difference.

(D) GSEA enrichment plots for pathways enriched between samples with higher IgA:IgG or higher IgG:IgA ratios for COVID-19-positive samples from severe patients on Day 7. Pathways enriched in IgA-high samples are HALLMARK_INTERFERON_ALPHA_RESPONSE and GO_REGULATION_OF_MEMBRANE_POTENTIAL, and pathways enriched in IgG-high samples are HALLMARK_REACTIVE_OXYGEN_SPECIES_PATHWAY and GO_FICOLIN_1_RICH_GRANULE.

(E) Representative fluorescence microscopy images of neutrophils treated with the following conditions: PBS, 100nM PMA + L-glu, one healthy control IgG + L-glu and IgA + L-glu, one severe with survival IgG + L-glu and IgA + L-glu, and one death IgG + L-glu and IgA + L-glu. Cells were stained for DNA (DAPI) and neutrophil elastase. Images were captured at 20x magnification. Blue; DAPI. Green; Neutrophil elastase. Circled cells indicate similar neutrophil cell death morphologies between PMA- and IgA-treated cells.

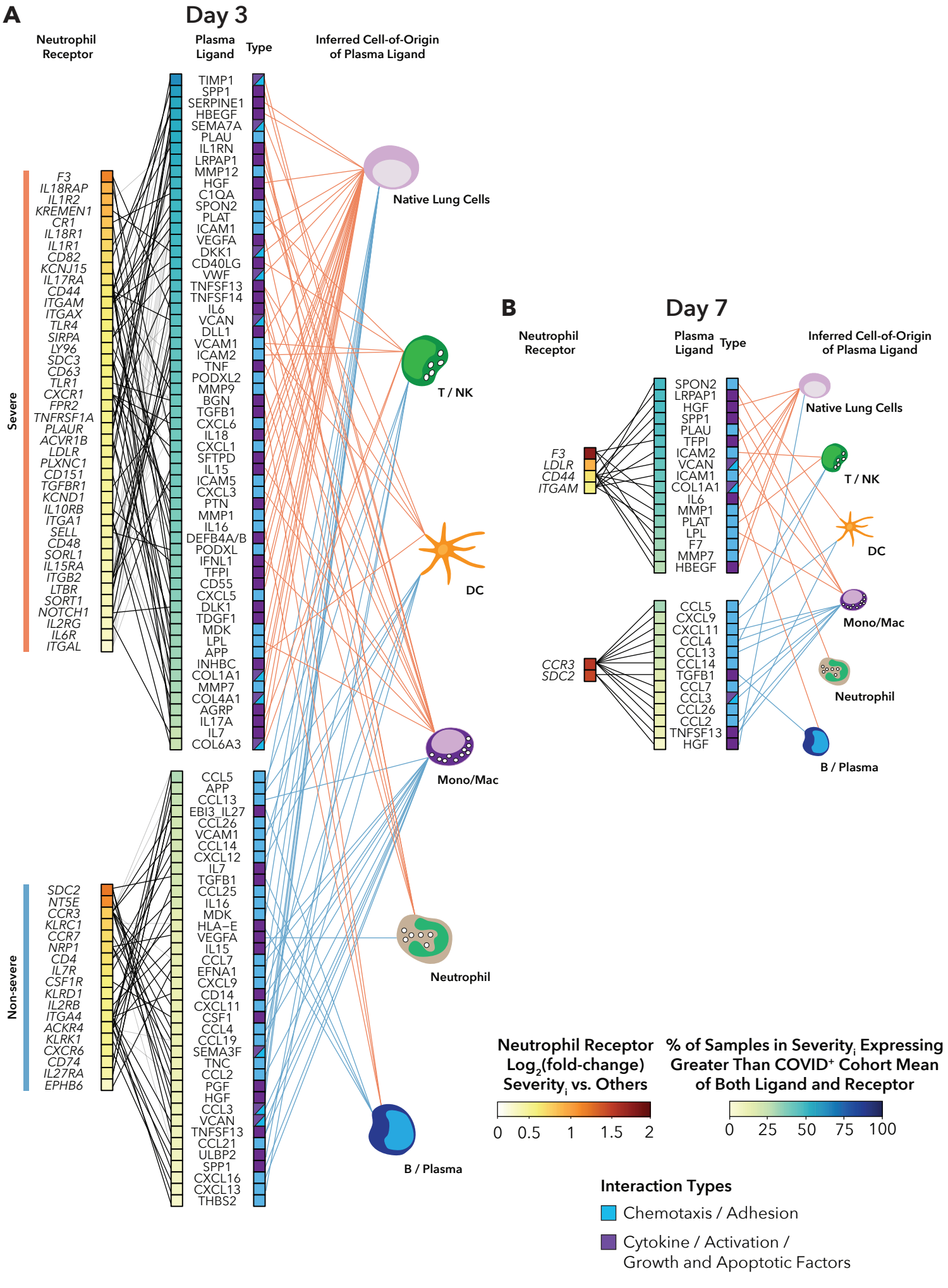
A
Day 3
Neutrophil Receptor
Plasma Ligand
Type
Inferred Cell-of-Origin of Plasma Ligand
Severe
F3
IL18RAP
IL1R2
KREMEN1
CR1
IL18R1
IL1R1
CD82
KCNJ15
IL17RA
CD44
ITGAM
ITGAX
TLR4
SIRPA
LY96
SDC3
CD63
TLR1
CXCR1
FPR2
TNFRSF1A
PLAUR
ACVR1B
LDLR
PLXNC1
CD151
TGFBR1
KCND1
IL10RB
ITGA1
SELL
CD48
SORL1
IL15RA
ITGB2
LTBR
SORT1
NOTCH1
IL2RG
IL6R
ITGAL
TIMP1
SPP1
SERPINE1
HBEGF
SEMA7A
PLAU
IL1RN
LRPAP1
MMP12
HGF
C1QA
SPON2
PLAT
ICAM1
VEGFA
DKK1
CD40LG
VWF
TNFSF13
TNFSF14
IL6
VCAN
DLL1
VCAM1
ICAM2
TNF
PODXL2
MMP9
BGN
TGFB1
CXCL6
IL18
CXCL1
SFTPD
IL15
ICAM5
CXCL3
PTN
MMP1
IL16
DEFB4A/B
PODXL
IFNL1
TFPI
CD55
CXCL5
DLK1
TDGF1
MDK
LPL
APP
INHBC
COL1A1
MMP7
COL4A1
AGRP
IL17A
IL7
COL6A3
Native Lung Cells
T / NK
DC
Mono/Mac
Neutrophil
B / Plasma
Non-severe
SDC2
NT5E
CCR3
KLRC1
CCR7
NRP1
CD4
IL7R
CSF1R
KLRD1
IL2RB
ITGA4
ACKR4
KLRK1
CXCR6
CD74
IL27RA
EPHB6
CCL5
APP
CCL13
EBI3_IL27
CCL26
VCAM1
CCL14
CXCL12
IL7
TGFB1
CCL25
IL16
MDK
HLA−E
VEGFA
IL15
CCL7
EFNA1
CXCL9
CD14
CXCL11
CSF1
CCL4
CCL19
SEMA3F
TNC
CCL2
PGF
HGF
CCL3
VCAN
TNFSF13
CCL21
ULBP2
SPP1
CXCL16
CXCL13
THBS2
B
Day 7
Neutrophil Receptor
Plasma Ligand
Type
Inferred Cell-of-Origin of Plasma Ligand
F3
LDLR
CD44
ITGAM
SPON2
LRPAP1
HGF
SPP1
PLAU
TFPI
ICAM2
VCAN
ICAM1
COL1A1
IL6
MMP1
PLAT
LPL
F7
MMP7
HBEGF
CCR3
SDC2
CCL5
CXCL9
CXCL11
CCL4
CCL13
CCL14
TGFB1
CCL7
CCL3
CCL26
CCL2
TNFSF13
HGF
Native Lung Cells
T / NK
DC
Mono/Mac
Neutrophil
B / Plasma
Neutrophil Receptor Log2(fold-change) Severity_i vs. Others
0 0.5 1 1.5 2
% of Samples in Severity_i Expressing Greater Than COVID+ Cohort Mean of Both Ligand and Receptor
0 25 50 75 100
Interaction Types
Chemotaxis / Adhesion
Cytokine / Activation / Growth and Apoptotic Factors

Figure S5. Ligand-receptor Interaction Analysis for Severe versus Non-severe Patients, Related to Figure 7.

Ligand-receptor analysis for differentially expressed ligands in plasma and receptors on neutrophils between COVID-19-positive severe and non-severe samples on (A) Day 3 and (B) Day 7. Ligands and receptors are color-coded by severity. Receptors are color-scaled according to the $\log_2$(fold-change) between severity groups. Ligands are color-scaled according to the percentage of samples within the severity group for which the ligand and receptor are both expressed above the overall mean expression.

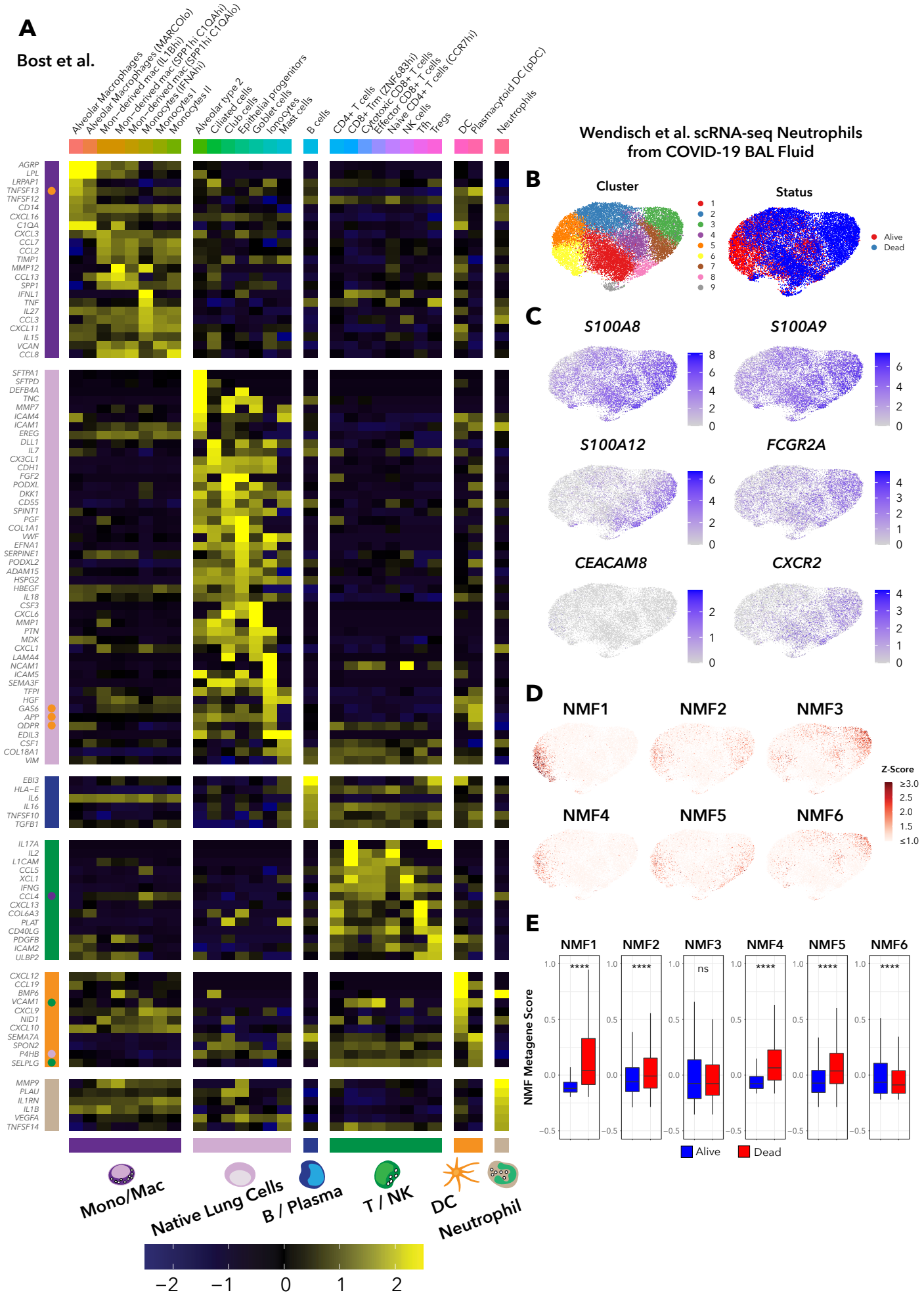
A
Bost et al.
Alveolar Macrophages
Alveolar Macrophages (MARCOlo)
Alveolar Macrophages (IL1Bhi)
Mon−derived mac (IL1Bhi)
Mon−derived mac (SPP1hi C1QAhi)
Mon−derived mac (SPP1hi C1QAlo)
Monocytes (IFNAhi)
Monocytes I
Monocytes II
Alveolar type 2
Ciliated cells
Club cells
Epithelial progenitors
Goblet cells
Ionocytes
Mast cells
B cells
CD4+ T cells
CD8+ Trm (ZNF683hi)
Cytotoxic CD8+ T cells
Effector CD8+ T cells
Naive CD4+ T cells (CCR7hi)
NK cells
Tfh
Tregs
DC
Plasmacytoid DC (pDC)
Neutrophils
AGRP
LPL
LRPAP1
TNFSF13
TNFSF12
CD14
CXCL16
C1QA
CXCL3
CCL7
CCL2
TIMP1
MMP12
CCL13
SPP1
IFNL1
TNF
IL27
CCL3
CXCL11
IL15
VCAN
CCL8
SFTPA1
SFTPD
DEFB4A
TNC
MMP7
ICAM4
ICAM1
EREG
DLL1
IL7
CX3CL1
CDH1
FGF2
PODXL
DKK1
CD55
SPINT1
PGF
COL1A1
VWF
EFNA1
SERPINE1
PODXL2
ADAM15
HSPG2
HBEGF
IL18
CSF3
CXCL6
MMP1
PTN
MDK
CXCL1
LAMA4
NCAM1
ICAM5
SEMA3F
TFPI
HGF
GAS6
APP
QDPR
EDIL3
CSF1
COL18A1
VIM
EBI3
HLA−E
IL6
IL16
TNFSF10
TGFB1
IL17A
IL2
L1CAM
CCL5
XCL1
IFNG
CCL4
CXCL13
COL6A3
PLAT
CD40LG
PDGFB
ICAM2
ULBP2
CXCL12
CCL19
BMP6
VCAM1
CXCL9
NID1
CXCL10
SEMA7A
SPON2
P4HB
SELPLG
MMP9
PLAU
IL1RN
IL1B
VEGFA
TNFSF14
Mono/Mac
Native Lung Cells
B / Plasma
T / NK
DC
Neutrophil
−2
−1
0
1
2
Wendisch et al. scRNA-seq Neutrophils from COVID-19 BAL Fluid
B
Cluster
1
2
3
4
5
6
7
8
9
Status
Alive
Dead
C
S100A8
S100A9
S100A12
FCGR2A
CEACAM8
CXCR2
D
NMF1
NMF2
NMF3
NMF4
NMF5
NMF6
Z-Score
≥3.0
2.5
2.0
1.5
≤1.0
E
NMF1
NMF2
NMF3
NMF4
NMF5
NMF6
NMF Metagene Score
****
****
ns
****
****
****
Alive
Dead

Figure S6. Inferring Cell-of-origin for Plasma Ligands Utilizing Single-cell RNA-seq Data and Neutrophil Subtypes in BAL Fluid, Related to Figure 7.

(A) Heatmap displaying single-cell RNA-sequencing (scRNA-seq) average scaled expression values per cell type for the genes encoding the protein ligands found to be differentially expressed in plasma between NMF clusters or severity groups for the ligand-receptor analysis in Figure 7 and Figure S5. scRNA-seq data is from bronchoalveolar lavage fluid (Bost et al. 2020). Column breaks indicate major cell lineages (Mono/Mac, Native lung cells, B/Plasma, T/NK, DC, Neutrophil). Row breaks indicate which genes have the highest average expression in a given major cell lineage. Color-coded dots indicate that the highest-to-second-highest difference in average scaled expression was less than 0.1, and thus the ligand was assigned to both lineages.

(B) UMAPs of single-cell RNA-seq data of neutrophils from COVID-19 BAL fluid from Wendisch et al. 2021. UMAPs are color-coded by Seurat clustering (left) and disease severity (right).

(C) UMAPs of BAL fluid neutrophils color-coded according to scaled gene expression of *S100A8*, *S100A9*, *S100A12*, *FCGR2A*, *CEACAM8*, and *CXCR2*.

(D) UMAPs of BAL fluid neutrophils color-coded by NMF cluster metagene scores.

(E) Box plots comparing NMF metagene scores between patients with WHO severity 7 (severe disease with survival) and severity 8 (patients who died). ns; not significant. P values are for the Wilcoxon rank-sum test. Quadruple asterisks indicate $p \leq 0.0001$.

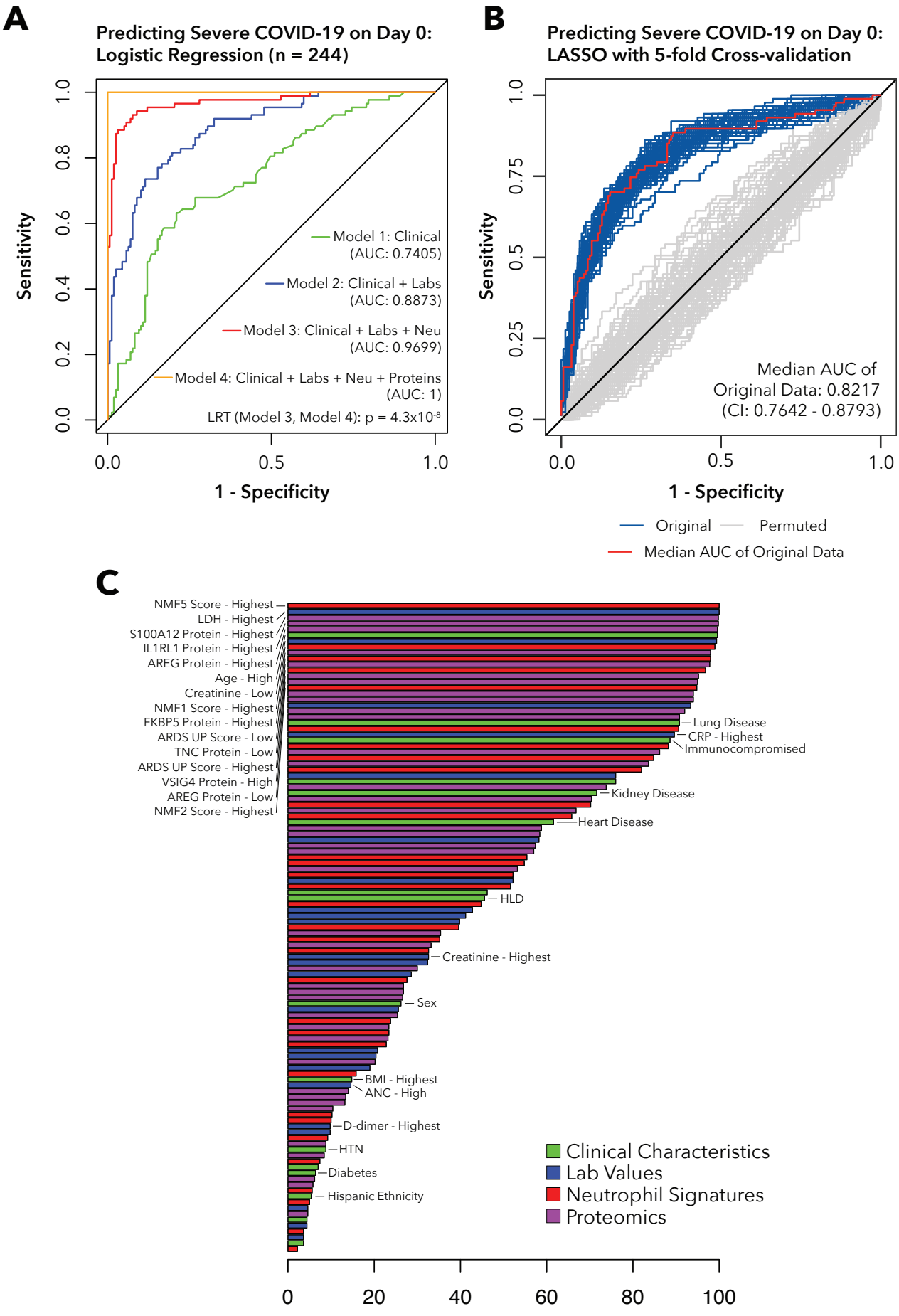
A
Predicting Severe COVID-19 on Day 0:
Logistic Regression (n = 244)
Sensitivity
1 - Specificity
0.0
0.2
0.4
0.6
0.8
1.0
0.5
Model 1: Clinical
(AUC: 0.7405)
Model 2: Clinical + Labs
(AUC: 0.8873)
Model 3: Clinical + Labs + Neu
(AUC: 0.9699)
Model 4: Clinical + Labs + Neu + Proteins
(AUC: 1)
LRT (Model 3, Model 4): p = 4.3x10-8
B
Predicting Severe COVID-19 on Day 0:
LASSO with 5-fold Cross-validation
Sensitivity
1 - Specificity
0.25
0.75
Median AUC of
Original Data: 0.8217
(CI: 0.7642 - 0.8793)
Original
Permuted
Median AUC of Original Data
C
NMF5 Score - Highest
LDH - Highest
S100A12 Protein - Highest
IL1RL1 Protein - Highest
AREG Protein - Highest
Age - High
Creatinine - Low
NMF1 Score - Highest
FKBP5 Protein - Highest
ARDS UP Score - Low
TNC Protein - Low
ARDS UP Score - Highest
VSIG4 Protein - High
AREG Protein - Low
NMF2 Score - Highest
Lung Disease
CRP - Highest
Immunocompromised
Kidney Disease
Heart Disease
HLD
Creatinine - Highest
Sex
BMI - Highest
ANC - High
D-dimer - Highest
HTN
Diabetes
Hispanic Ethnicity
Clinical Characteristics
Lab Values
Neutrophil Signatures
Proteomics
0
20
40
60
80
100

Figure S7. Predicting Severe COVID-19 on Day 0 Utilizing Clinical Data, Neutrophil Transcriptomics, and Neutrophil-expressed Plasma Proteins, Related to Figures 6 and 7.

(A) Receiver operating characteristic (ROC) curve for predictive performance of logistic regression models predicting COVID-19 disease severity on Day 0. Ten samples were dropped from the models in Figure 3B due to missing proteomics data for a total of 244 samples. Model 1 includes only clinical characteristics: age, gender, ethnicity, heart disease, diabetes, hypertension, hyperlipidemia, lung disease, kidney disease, immunocompromised status, BMI (AUC: 0.7405). Model 2 adds the following clinical laboratory values: ANC, ALC, Creatinine, CRP, D-dimer, LDH (AUC: 0.8873). Model 3 incorporates the following neutrophil gene signature scores, broken into quintiles: NMF1, NMF2, NMF3, NMF4, NMF5, NMF6, ARDS Up - Juss, ARDS Down - Juss (AUC: 0.9699). Model 4 adds the following plasma proteins, broken into expression quintiles: TNC, TNFRSF10C, S100A12, HGF, F9, AREG, MMP8, IL1RL1, FKBP5, VSIG4. Significance of improvement of model determined with the likelihood ratio test.

(B) ROC curve of predictive performance of a least absolute shrinkage and selection operator (LASSO) model of COVID-19 disease severity on Day 0 using the parameters from Model 4. Prediction was performed with repeated 5-fold cross-validation with 100 repeats for both the original data and permutated labels of severity. Shown in red is the ROC curve for the cross-validation repeat with the median AUC across all repeats.

(C) Bar plot displaying the selection frequency for each factor in the LASSO regression model. Bars are color coded by variable type, corresponding to the four models shown in (B). Lab values and gene signatures variables are broken into quintiles with levels: 1 = lowest, 2 = low, 3 = mid, 4 = high, 5 = highest.