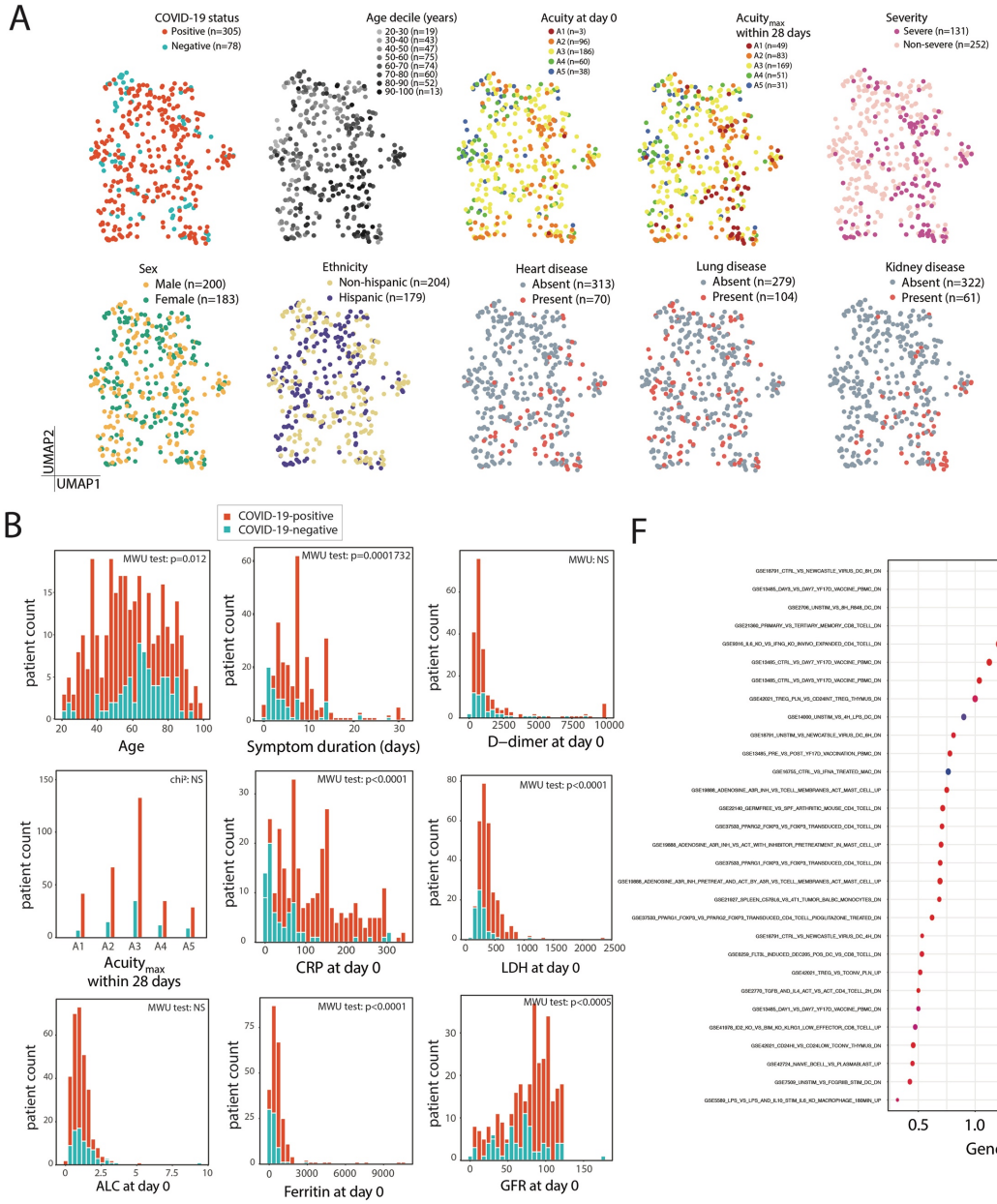
**Cell Reports Medicine, Volume *2***

## Supplemental information

# Longitudinal proteomic analysis of severe COVID-19 reveals survival-associated signatures, tissue-specific cell death, and cell-cell interactions

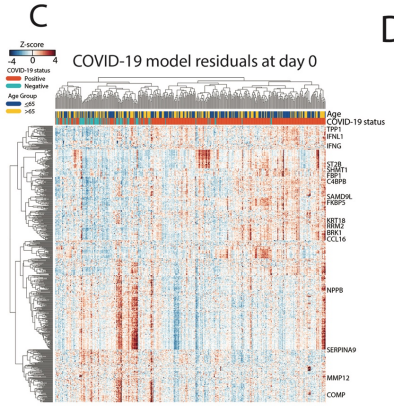
**Michael R. Filbin, Arnav Mehta, Alexis M. Schneider, Kyle R. Kays, Jamey R. Guess, Matteo Gentili, Bánk G. Fenyves, Nicole C. Charland, Anna L.K. Gonye, Irena Gushterova, Hargun K. Khanna, Thomas J. LaSalle, Kendall M. Lavin-Parsons, Brendan M. Lilley, Carl L. Lodenstein, Kasidet Manakongtreecheep, Justin D. Margolin, Brenna N. McKaig, Maricarmen Rojas-Lopez, Brian C. Russo, Nihaarika Sharma, Jessica Tantivit, Molly F. Thomas, Robert E. Gerszten, Graham S. Heimberg, Paul J. Hoover, David J. Lieb, Brian Lin, Debby Ngo, Karin Pelka, Miguel Reyes, Christopher S. Smillie, Avinash Waghray, Thomas E. Wood, Amanda S. Zajac, Lori L. Jennings, Ida Grundberg, Roby P. Bhattacharyya, Blair Alden Parry, Alexandra-Chloé Villani, Moshe Sade-Feldman, Nir Hacohen, and Marcia B. Goldberg**

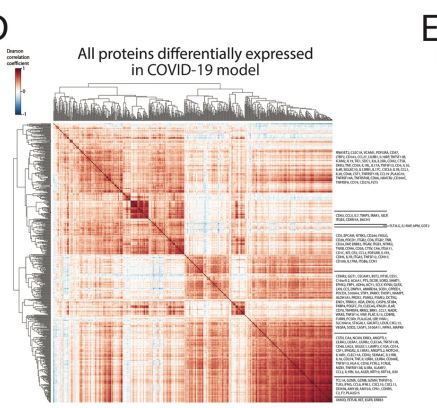
**Supplemental Figures**

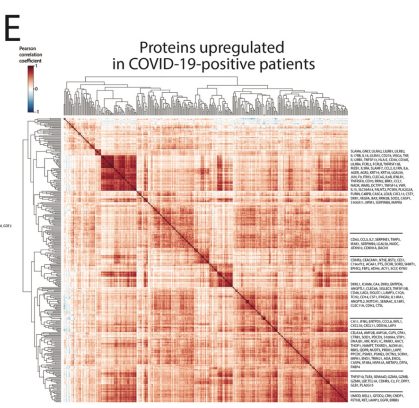

A
COVID-19 status
Positive (n=305)
Negative (n=78)
Age decile (years)
20-30 (n=19)
30-40 (n=43)
40-50 (n=47)
50-60 (n=75)
60-70 (n=74)
70-80 (n=60)
80-90 (n=52)
90-100 (n=13)
Acuity at day 0
A1 (n=3)
A2 (n=96)
A3 (n=186)
A4 (n=60)
A5 (n=38)
Acuity max within 28 days
A1 (n=49)
A2 (n=83)
A3 (n=169)
A4 (n=51)
A5 (n=31)
Severity
Severe (n=131)
Non-severe (n=252)
Sex
Male (n=200)
Female (n=183)
Ethnicity
Non-hispanic (n=204)
Hispanic (n=179)
Heart disease
Absent (n=313)
Present (n=70)
Lung disease
Absent (n=279)
Present (n=104)
Kidney disease
Absent (n=322)
Present (n=61)
UMAP2
UMAP1


B
COVID-19-positive
COVID-19-negative
MWU test: p=0.012
patient count
Age
MWU test: p=0.0001732
Symptom duration (days)
MWU: NS
D–dimer at day 0
chi²: NS
Acuity max within 28 days
A1
A2
A3
A4
A5
MWU test: p<0.0001
CRP at day 0
MWU test: p<0.0001
LDH at day 0
MWU test: NS
ALC at day 0
MWU test: p<0.0001
Ferritin at day 0
MWU test: p<0.0005
GFR at day 0

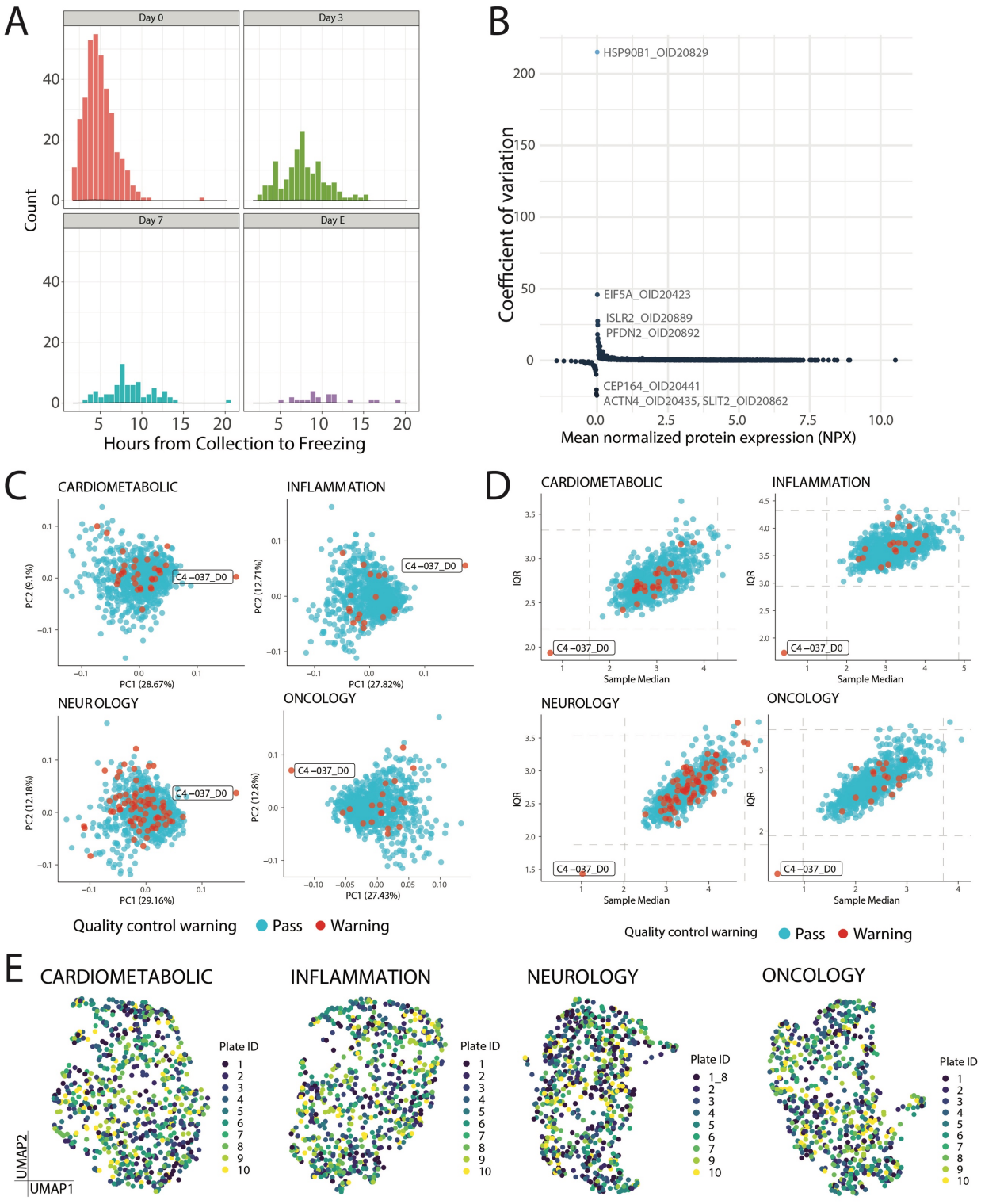

F
Count
20
40
60
p.adjust
0.00025
0.00050
0.00075
0.5
1.0
1.5
2.0
GeneRatio


C
Z-score
COVID-19 model residuals at day 0
COVID-19 status
Positive
Negative
Age Group
≤65
>65
Age
COVID-19 status
TPP1
IFNL1
IFNG
ST2B
SAMD9
FBP1
CXBP8
SAMD9L
FKBP5
KRT18
KRIN2
IRR1
CCL16
NPPB
SERPINA9
MMP12
COMP


D
All proteins differentially expressed in COVID-19 model

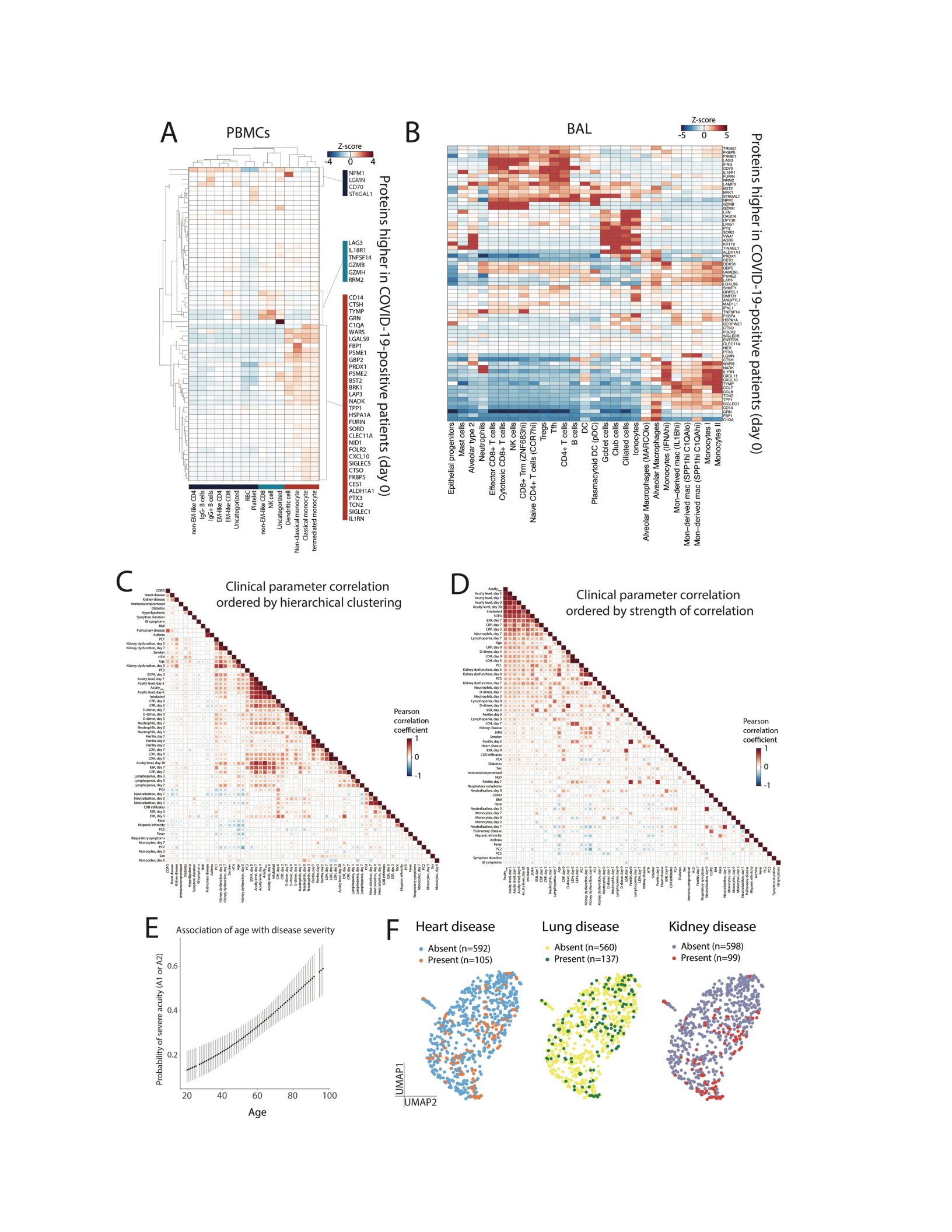

E
Proteins upregulated in COVID-19-positive patients

**Supplemental Figure 1**. Distributions of clinical characteristics and differentially-expressed plasma proteins between COVID-19-positive and COVID-19-negative patients in the study cohort. Related to Figure 1 and Tables S1, S2, and S3. (A) Unsupervised clustering (by UMAP), generated from plasma proteins of all patients on day 0, color-coded (left to right) by COVID-19 status, age decile, acuity level at day 0, $Acuity_{max}$ within 28 days, severity, gender, ethnicity, previously known heart disease, previously known lung disease, and previously known kidney disease. (B) Age, symptom duration prior to ED arrival, D-dimer, $Acuity_{max}$, CRP, LDH, absolute lymphocyte count (ALC), ferritin, and glomerular filtration rate (GFR). All measurements, except $Acuity_{max}$ and symptom duration are at day 0. Significance is by the Mann-Whitney U-test (MWU) and chi-squared test ($chi^2$); p-values as indicated. (C)-(F) Differentially-expressed proteins by COVID-19 status. (C) Identification of differentially-expressed proteins by COVID-19 status using linear model of Olink proteins with putative confounders as covariates (see **STAR methods**). Heatmap of residuals from this model for each protein found to be significantly differentially-expressed between COVID-19-positive and negative patients, using the model described in **Figure 1C-1E**. (D) Correlation heatmap of plasma protein levels of all proteins significantly differentially-expressed between COVID-19-positive and negative patients. (E) Correlation heatmap of plasma protein levels of all proteins significantly higher in COVID-19-positive patients than in COVID-19-negative patients. (F) Gene-set enrichment analysis for pathways enriched among plasma proteins differentially-expressed between COVID-19-positive and negative patients. Linear model fitting each Olink protein, with COVID-19 status as a main effect and putative confounders as covariates (see **STAR methods**). p-values were adjusted to control the false discovery rate (FDR) at <0.05, Benjamini-Hochberg method.

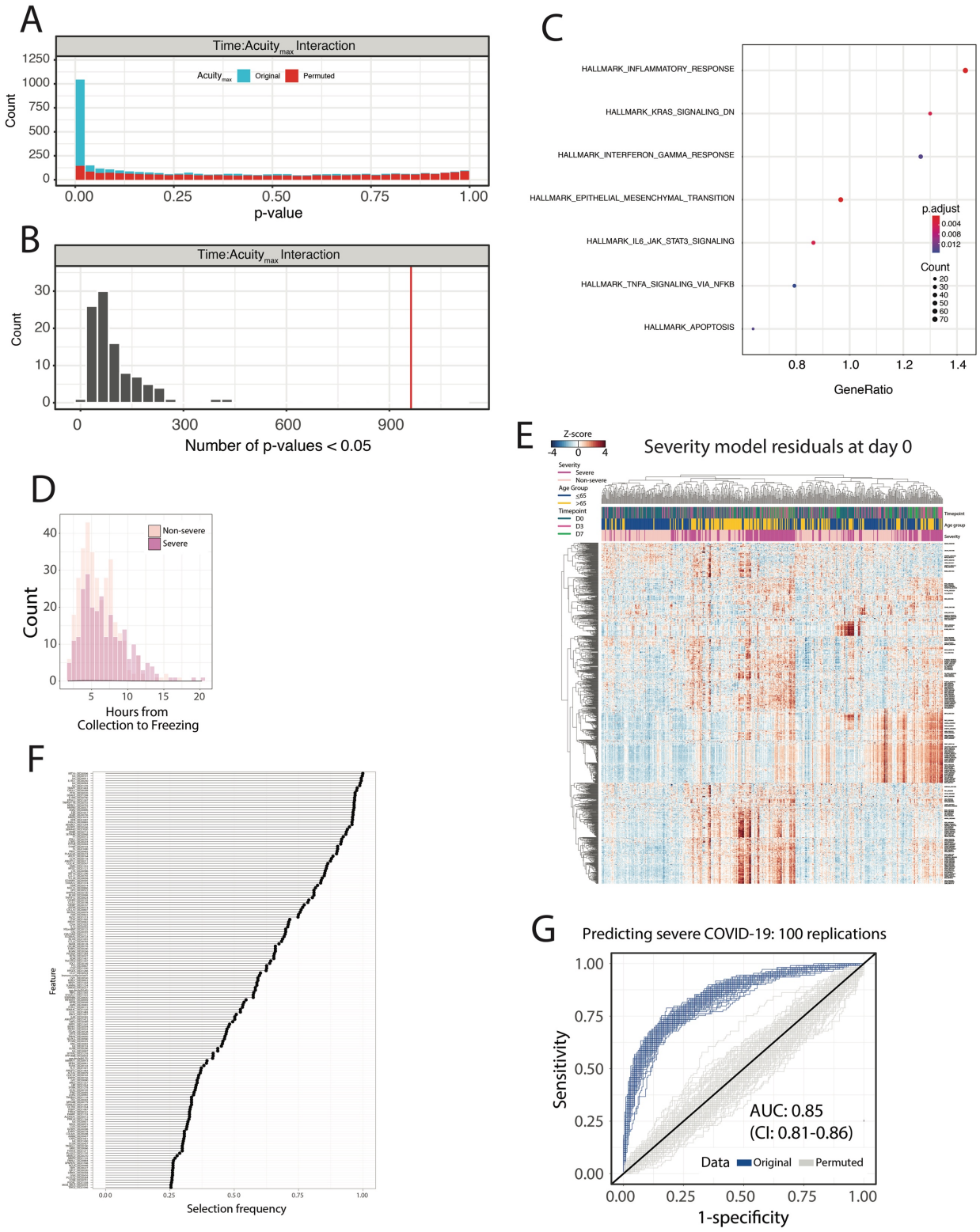



**Supplemental Figure 2**. Quality control for Olink plasma proteomics data. Related to Figures 1 and 2, and Table S2. (A) Histogram showing the time from collection to freezing for each sample, separated by time point of collection. E, event-driven samples (see **STAR methods**). (B) Histogram showing the

coefficient of variation versus mean normalized protein expression (NPX) for each of the 1356 assays performed by Olink. (C) Principal component plots using Olink assays in four predefined Olink panels (cardiometabolic, inflammation, neurology, and oncology), each with >300 protein assays. Shown are first two principal components (PC1 vs. PC2) for all patient samples. (D) Scatterplot showing interquartile range (IQR) versus sample median for each sample. In (C) and (D), sample C4-037_D0 was noted to be an outlier and was therefore excluded from downstream analyses. (E) Unsupervised clustering by UMAP of all samples by assays in each predefined panel as in (C-D), color-coded by plate label, confirming there were no plate-related biases.

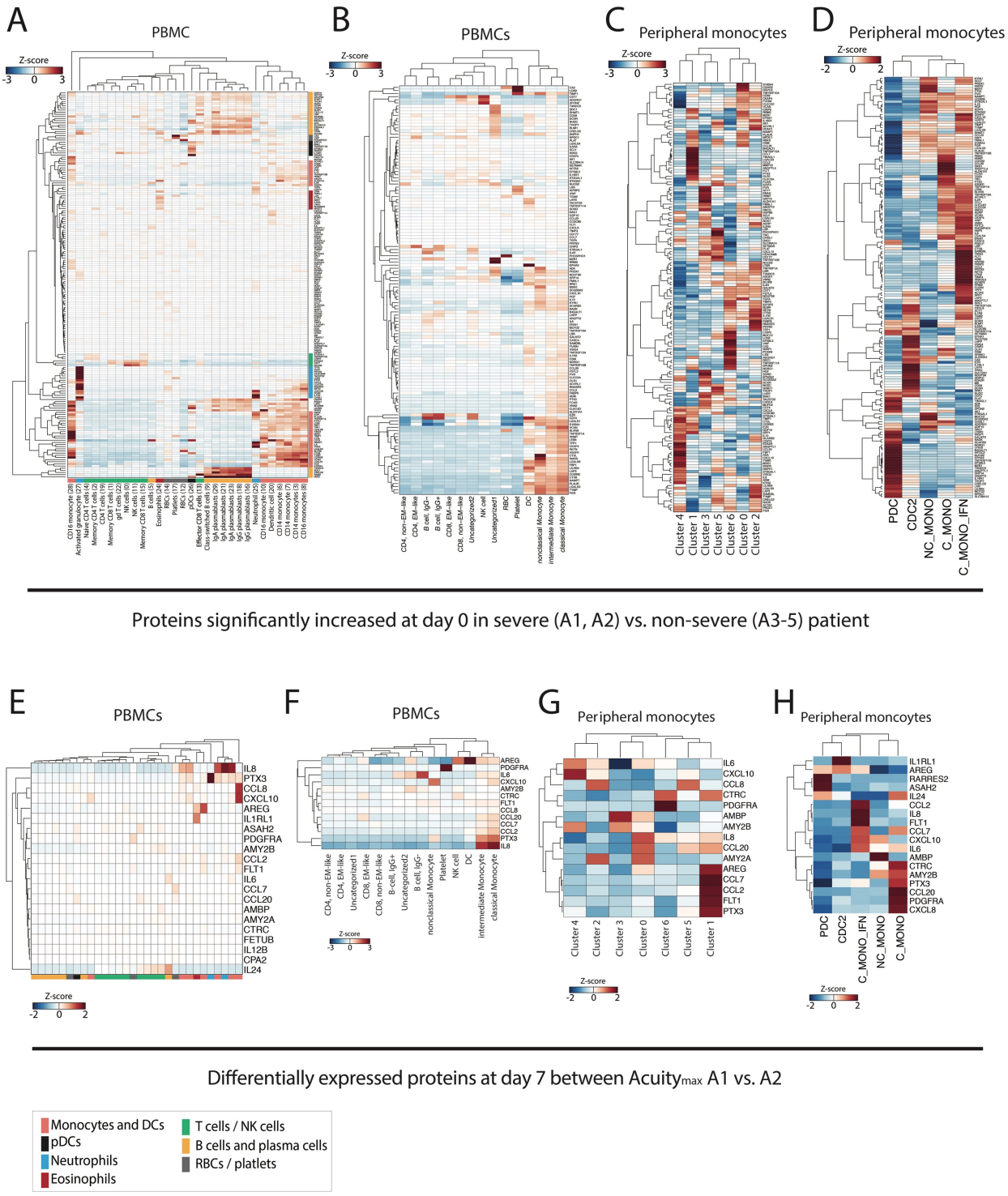

A
PBMCs
Z-score
-4
0
4
NPM1
LGMN
CD70
ST6GAL1
LAG3
IL18R1
TNFSF14
GZMB
GZMH
RRM2
CD14
CTSH
TYMP
GRN
C1QA
WARS
LGALS9
FBP1
PSME1
GBP2
PRDX1
PSME2
BST2
BRK1
LAP3
NADK
TPP1
HSPA1A
FURIN
SORD
CLEC11A
NID1
FOLR2
CXCL10
SIGLEC5
CTSO
FKBP5
CES1
ALDH1A1
PTX3
TCN2
SIGLEC1
IL1RN
Proteins higher in COVID-19-positive patients (day 0)
non-EM-like CD4
IgG- B cells
IgG+ B cells
EM-like CD4
EM-like CD8
Uncategorized
RBC
Platelet
non-EM-like CD8
NK cell
Uncategorized
Dendritic cell
Non-classical monocyte
Classical monocyte
termediated monocyte
B
BAL
Z-score
-5
0
5
Proteins higher in COVID-19-positive patients (day 0)
Epithelial progenitors
Mast cells
Alveolar type 2
Neutrophils
Effector CD8+ T cells
Cytotoxic CD8+ T cells
NK cells
CD8+ Trm (ZNF683hi)
Naive CD4+ T cells (CCR7hi)
Tregs
Tfh
CD4+ T cells
B cells
DC
Plasmacytoid DC (pDC)
Goblet cells
Club cells
Ciliated cells
Ionocytes
Alveolar Macrophages (MARCOlo)
Alveolar Macrophages
Monocytes (IFNAhi)
Mon–derived mac (IL1Bhi)
Mon–derived mac (SPP1hi C1QAlo)
Mon–derived mac (SPP1hi C1QAhi)
Monocytes I
Monocytes II
C
Clinical parameter correlation
ordered by hierarchical clustering
Pearson
correlation
coefficient
1
0
-1
D
Clinical parameter correlation
ordered by strength of correlation
Pearson
correlation
coefficient
1
0
-1
E
Association of age with disease severity
Probability of severe acuity (A1 or A2)
0.6
0.4
0.2
20
40
60
80
100
Age
F
Heart disease
Absent (n=592)
Present (n=105)
Lung disease
Absent (n=560)
Present (n=137)
Kidney disease
Absent (n=598)
Present (n=99)
UMAP1
UMAP2

**Supplemental Figure 3**. Inference of cell of origin and clinical correlates derived by mapping gene expression of differentially-expressed plasma proteins (elevated in COVID-19-positive versus COVID-19-negative patients) onto scRNAseq datasets. Related to Figures 1 and 2, and Table S2. (A)-(B) Heatmaps of mean expression of COVID-19-related proteins (y-axis) in immune cell subtypes (x-axis) from peripheral blood cell scRNAseq dataset in Lee *et al*.[1] (A) and BAL COVID-19 scRNAseq dataset in Bost *et al*.[2] (B). non-EM-like, non-effector memory-like; EM-like, effector memory-like; Trm, resident memory CD8+ T cells; DC, dendritic cells; Mon-derived mac, monocyte-derived macrophages. (C)-(D) Correlation heatmap of all clinically associated variables collected in this study ordered by hierarchical clustering (C) or correlation to $Acuity_{max}$ (D). For the purpose of displaying positive correlation with increasing disease severity, inverse numbering was assigned to A1-A5 for these calculations, and kidney function was measured as the inverse of the glomerular filtration rate (GFR). PC1 to PC5 are the loadings of the first five principal components for the PCA plot of all samples using all proteins. SOFA, sequential organ failure assessment score; monocytes, monocyte count; neutrophils, absolute neutrophil count. (E) Logistic regression model predicting probability of severe acuity ($Acuity_{max}$ of A1 or A2) using patient age. (F) Unsupervised clustering by UMAP using plasma proteins of all COVID-19-positive patients at days 0, 3 and 7, color-coded (left to right) by previously known heart disease, previously known lung disease, and previously known kidney disease.



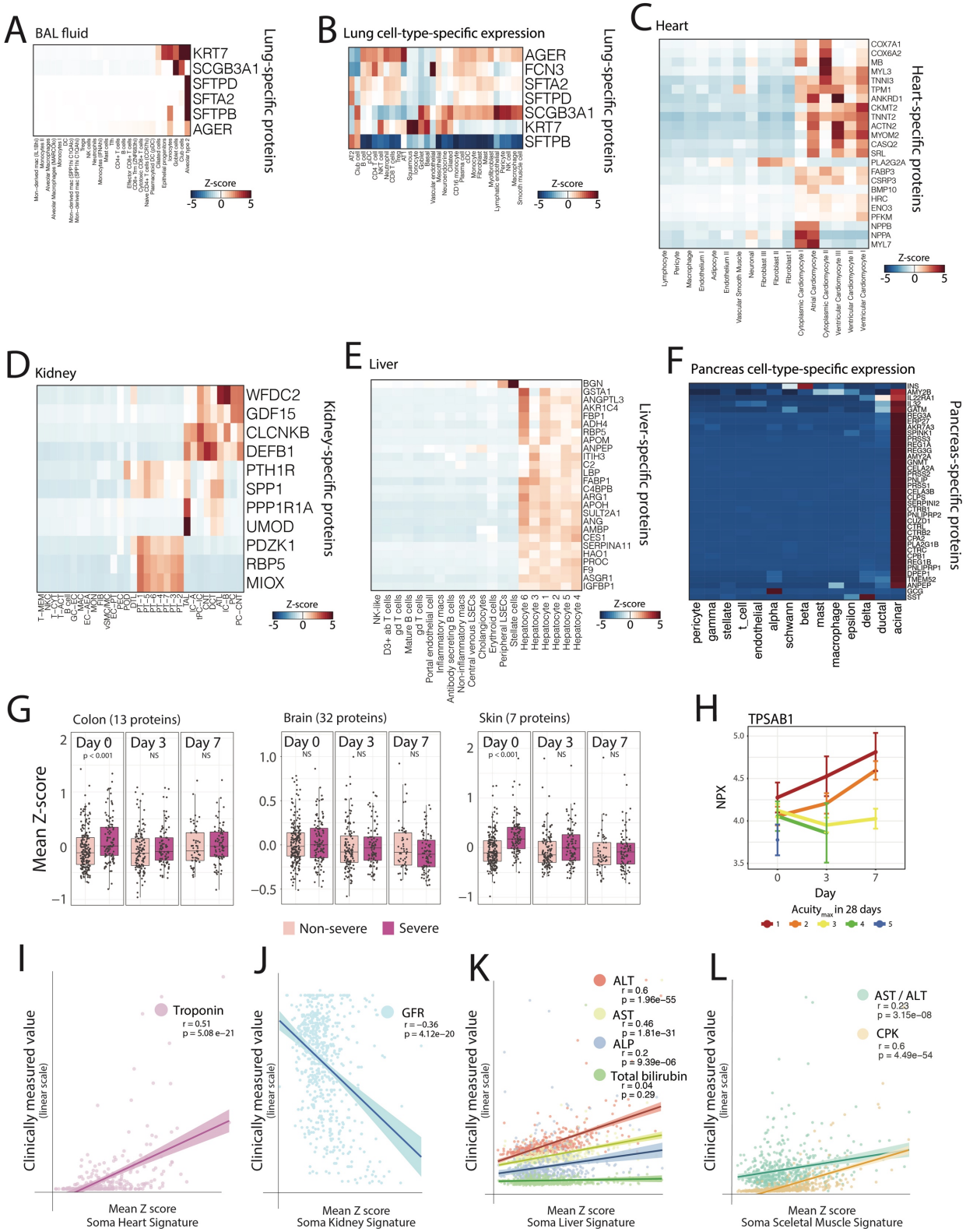

**Supplemental Figure 4**. Permutation analysis for linear mixed model on Olink data, with Acuity$_{max}$ and time as main effects. Related to Figure 2, and Tables S2 and S3. (A) Linear mixed model fitting each

Olink protein, with $Acuity_{max}$, timepoint, and the interaction of the two terms as main effects and putative confounders as covariates (see **STAR methods**). Significance of the three model terms was determined with an F-test using Satterthwaite degrees of freedom and type III sum of squares. p-values for the three model terms of interest were adjusted to control the FDR at < 0.05, Benjamini-Hochberg method. Group differences were calculated for each protein passing the 0.05 FDR threshold with p-values adjusted using the Tukey method. Shown are the distribution of p-values obtained from this linear mixed model and the distribution of p-values obtained from permutations of random $Acuity_{max}$ assignments to all patients. This result shows that the obtained distribution of p-values does not happen by chance. (B) Distribution of the number of p-values < 0.05 obtained from 100 permutations of random $Acuity_{max}$ assignments for the linear mixed model described in (A). The red line indicates the number of significant assays obtained from our true model with correct patient $Acuity_{max}$ assignments. (C)-(G) Protein associations with disease severity. (C) Gene-set enrichment analysis for pathways enriched among plasma proteins differentially-expressed between severe and non-severe COVID-19-positive patients. (D) Time from sample collection to freezing for all samples, color-coded by COVID-19 severity, indicating no clear difference between samples from severe and samples from non-severe patients. (E) Linear model fitting each Olink protein, with putative confounders as covariates (see **STAR methods**). Heatmap of residuals from this model for each protein that was found to be significant for interaction term in the model described in **Fig. 2C**. (F) Frequency of protein feature selection over the 100 iterations of the predictor in (G). (G) Receiver operating characteristic (ROC) curve showing predictive performance of an elastic net logistic regression classifier of disease severity using Olink plasma proteins for each patient at day 0. Shown are curves for each of the 100 repeats of 5-fold cross validation. Measures of performance are reported as medians and 95% confidence intervals. Shown in the figure are the ROC curves for the original data with the true severity labels and the curves for our data with random (permuted) severity assignments.

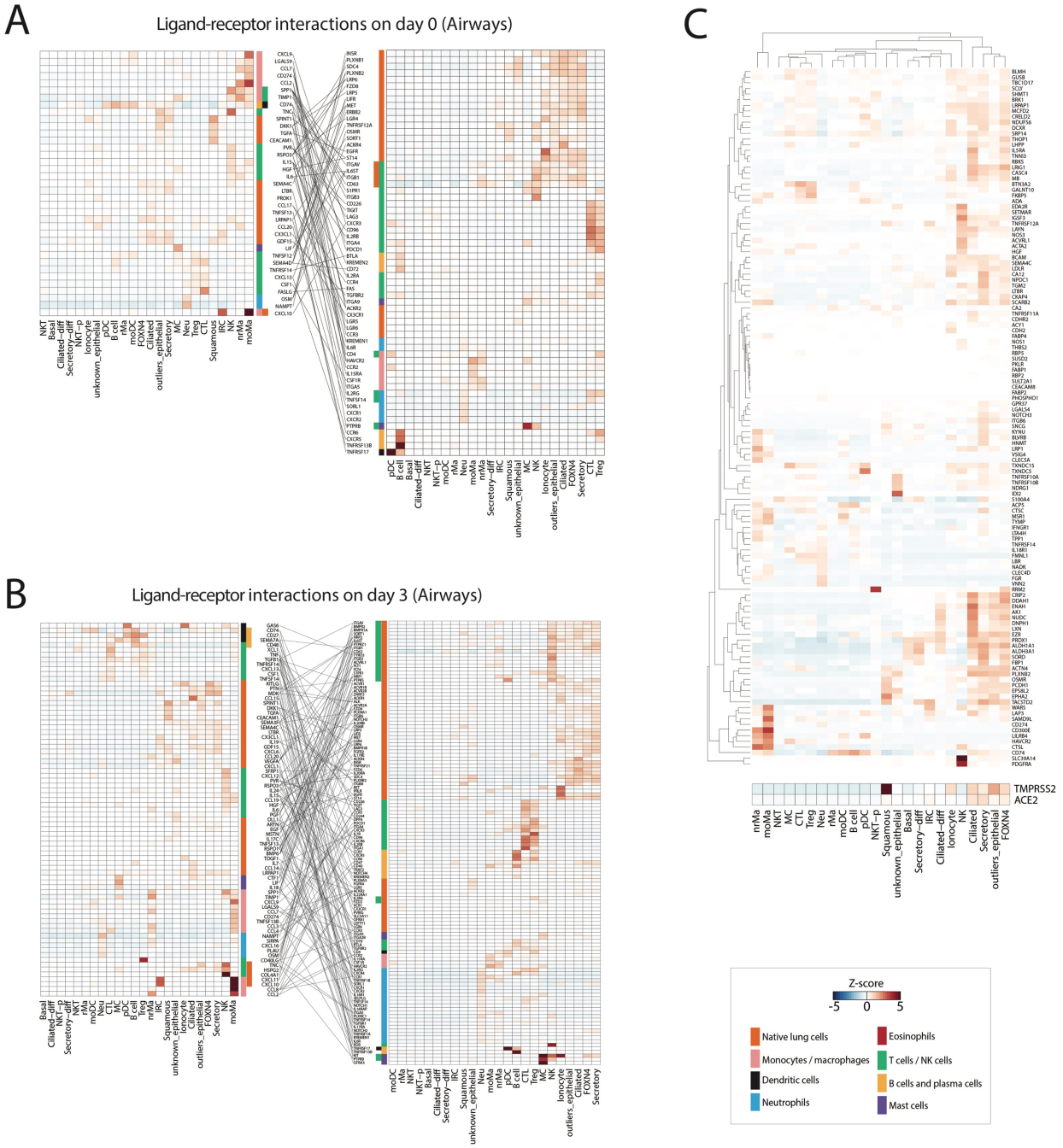



**Supplemental Figure 5**. Expression of severity-associated plasma proteins in PBMCs from COVID-19 patients. Related to Figures 1 and 5, and Tables S2 and S4. (A) Gene expression of Olink plasma proteins expressed more highly in severe COVID-19 patients shown within PBMCs using scRNAseq data.[3] (B)-

(D) Expression of Olink plasma proteins expressed more highly in severe COVID-19 patients within PBMCs show within another independent dataset[1] (B), or circulating monocytes, from two independent datasets[4, 5] (C)-(D). (E)-(H) Expression of differentially-expressed Olink proteins at day 7 between patients whose maximum classification was A1 (death) versus A2 (ARDS but survived) within day 28, for two independent PBMC datasets[1, 3] (E)-(F) and two independent peripheral monocyte datasets[4, 5] (G)-(H). non-EM-like, non-effector memory-like; EM-like, effector memory-like; DC, dendritic cells; pDC, plasmacytoid DCs; CDC2, C_MONO, NC_MONO, dendritic and monocyte clusters as defined in [4, 5].


A
BAL fluid
KRT7
SCGB3A1
SFTPD
SFTA2
SFTPB
AGER
Lung-specific proteins
Z-score
-5 0 5
B
Lung cell-type-specific expression
AGER
FCN3
SFTA2
SFTPD
SCGB3A1
KRT7
SFTPB
Lung-specific proteins
Z-score
-5 0 5
C
Heart
COX7A1
COX6A2
MB
MYL3
TNNI3
TPM1
ANKRD1
CKMT2
TNNT2
ACTN2
MYOM2
CASQ2
SRL
PLA2G2A
FABP3
CSRP3
BMP10
HRC
ENO3
PFKM
NPPB
NPPA
MYL7
Heart-specific proteins
Z-score
-5 0 5
D
Kidney
WFDC2
GDF15
CLCNKB
DEFB1
PTH1R
SPP1
PPP1R1A
UMOD
PDZK1
RBP5
MIOX
Kidney-specific proteins
Z-score
-5 0 5
E
Liver
BGN
GSTA1
ANGPTL3
AKR1C4
FBP1
ADH4
RBP5
APOM
ANPEP
ITIH3
C2
LBP
FABP1
C4BPB
ARG1
APOH
SULT2A1
ANG
AMBP
CES1
SERPINA11
HAO1
PROC
F9
ASGR1
IGFBP1
Liver-specific proteins
Z-score
-5 0 5
F
Pancreas cell-type-specific expression
INS
AMY2B
IL22RA1
IL22
GATM
REG3A
PNP2
AKR7A3
SPINK1
PRSS3
REG1B
REG3G
AMY2A
GNMT
CEL.HSA
PRSS2
PNLIP
PRSS1
CELA3B
CPB
SERPINI2
CTRB1
PNLIPRP2
CUZD1
CTRL
CTRB2
CPA2
PLA2G1B
CTRC
CPB1
REG1B
PNLIPRP1
QPEP1
TMEM52
ANPEP
GCG
SST
pericyte
gamma
stellate
t_cell
endothelial
alpha
schwann
beta
mast
macrophage
epsilon
delta
ductal
acinar
Pancreas-specific proteins
G
Colon (13 proteins)
Brain (32 proteins)
Skin (7 proteins)
Mean Z-score
Day 0
Day 3
Day 7
p < 0.001
NS
Non-severe
Severe
H
TPSAB1
NPX
Day
Acuity_max in 28 days
1 2 3 4 5
I
Troponin
r = 0.51
p = 5.08 e−21
Clinically measured value
(linear scale)
Mean Z score
Soma Heart Signature
J
GFR
r = −0.36
p = 4.12e−20
Clinically measured value
(linear scale)
Mean Z score
Soma Kidney Signature
K
ALT
r = 0.6
p = 1.96e−55
AST
r = 0.46
p = 1.81e−31
ALP
r = 0.2
p = 9.39e−06
Total bilirubin
r = 0.04
p = 0.29
Clinically measured value
(linear scale)
Mean Z score
Soma Liver Signature
L
AST / ALT
r = 0.23
p = 3.15e−08
CPK
r = 0.6
p = 4.49e−54
Clinically measured value
(linear scale)
Mean Z score
Soma Sceletal Muscle Signature

**Supplemental Figure 6.** Organ-specific cellular death signatures (from SomaScan data). (A)-(B) Expression of derived lung specific plasma proteins (secreted, membrane and intracellular) within BAL fluid from COVID-19 patients.[2] Related to Figures 5 and 6. (A) and an integrated analysis of several lung scRNAseq datasets (B). (C) Heatmap showing cell-type expression of severity-associated intracellular proteins mapped to cell types in scRNAseq data of normal cardiac tissue.[6] Kidney, pancreatic and liver-specific cellular death signatures. (D) Heatmap showing cell-type expression of severity-associated intracellular proteins mapped to normal kidney tissue.[7] (E) Heatmap showing cell-type expression of severity-associated intracellular proteins mapped to normal liver tissue.[8] (F) Heatmap showing cell-type expression of severity-associated intracellular proteins mapped to normal pancreatic tissue expression matrix obtained from the Itai Yanai lab.[9] (G) Expression of tissue-specific plasma protein signatures in severe versus non-severe patients at each timepoint in select tissues. (H) Expression of the mast cell marker tryptase (TPSAB1 assay by Olink) over time, by $Acuity_{max}$ level. (I)-(L) Organ damage plasma protein signature correlations with indicated clinical laboratory measurements at day 0. GFR, glomerular filtration rate; ALT, alanine aminotransferase; AST, aspartate aminotransferase; ALP, alkaline phosphatase; CPK, creatinine phosphokinase.




**Supplemental Figure 7.** Cellular communication and intracellular severity-associated proteins in the airway. Related to Figures 5 and 6, and Table S2. (A)-(B) Ligand-receptor relationships of severity-associated plasma ligands in an independent dataset of nasopharyngeal and bronchial samples from COVID-19 patients[10, 11] for ligands significant on day 0 (A) or day 3 (B). (C) Heatmap showing mean gene expression per cell type of severity-associated intracellular plasma proteins at D0 derived from

Olink data that map to scRNAseq of the nasopharyngeal and bronchial airway,[10, 11] with TMPRSS2 and ACE2 expression indicated. NKT, natural killer (NK) T cells; Ciliated-diff, differentiating secretory cells; Secretory-diff, differentiating secretory cells; NKT-p, proliferating NK T cells; pDC, plasmacytoid dendritic cells; rMa, resident macrophages; moDC, monocyte-derived dendritic cells; MC, mast cells; Neu, neutrophils; Treg, regulatory T cells; CTL, cytotoxic T cells; IRC, IFNg responsive cells; nrMa, non-resident macrophages; moMa, monocyte-derived macrophages; rMa, resident macrophages.

**Supplemental References**


1. Lee JS, Park S, Jeong HW, Ahn JY, Choi SJ, Lee H, Choi B, Nam SK, Sa M, Kwon JS, et al. (2020). Immunophenotyping of COVID-19 and influenza highlights the role of type I interferons in development of severe COVID-19. Sci Immunol 5. 2020/07/12.

2. Bost P, Giladi A, Liu Y, Bendjelal Y, Xu G, David E, Blecher-Gonen R, Cohen M, Medaglia C, Li H, et al. (2020). Host-Viral Infection Maps Reveal Signatures of Severe COVID-19 Patients. Cell 181, 1475-1488 e1412. 2020/06/02.

3. Wilk AJ, Rustagi A, Zhao NQ, Roque J, Martinez-Colon GJ, McKechnie JL, Ivison GT, Ranganath T, Vergara R, Hollis T, et al. (2020). A single-cell atlas of the peripheral immune response in patients with severe COVID-19. Nat Med 26, 1070-1076. 2020/06/10.

4. Arunachalam PS, Wimmers F, Mok CKP, Perera R, Scott M, Hagan T, Sigal N, Feng Y, Bristow L, Tak-Yin Tsang O, et al. (2020). Systems biological assessment of immunity to mild versus severe COVID-19 infection in humans. Science 369, 1210-1220. 2020/08/14.

5. Schulte-Schrepping J, Reusch N, Paclik D, Bassler K, Schlickeiser S, Zhang B, Kramer B, Krammer T, Brumhard S, Bonaguro L, et al. (2020). Severe COVID-19 Is Marked by a Dysregulated Myeloid Cell Compartment. Cell 182, 1419-1440 e1423. 2020/08/19.

6. Tucker NR, Chaffin M, Fleming SJ, Hall AW, Parsons VA, Bedi KC, Jr., Akkad AD, Herndon CN, Arduini A, Papangeli I, et al. (2020). Transcriptional and Cellular Diversity of the Human Heart. Circulation 142, 466-482. 2020/05/15.

7. Menon R, Otto EA, Hoover P, Eddy S, Mariani L, Godfrey B, Berthier CC, Eichinger F, Subramanian L, Harder J, et al. (2020). Single cell transcriptomics identifies focal segmental glomerulosclerosis remission endothelial biomarker. JCI Insight 5. 2020/02/29.

8. MacParland SA, Liu JC, Ma XZ, Innes BT, Bartczak AM, Gage BK, Manuel J, Khuu N, Echeverri J, Linares I, et al. (2018). Single cell RNA sequencing of human liver reveals distinct intrahepatic macrophage populations. Nat Commun 9, 4383. 2018/10/24.

9. Baron M, Veres A, Wolock SL, Faust AL, Gaujoux R, Vetere A, Ryu JH, Wagner BK, Shen-Orr SS, Klein AM, et al. (2016). A Single-Cell Transcriptomic Map of the Human and Mouse Pancreas Reveals Inter- and Intra-cell Population Structure. Cell Syst 3, 346-360 e344. 2016/10/28.

10. Chua RL, Lukassen S, Trump S, Hennig BP, Wendisch D, Pott F, Debnath O, Thurmann L, Kurth F, Volker MT, et al. (2020). COVID-19 severity correlates with airway epithelium-immune cell interactions identified by single-cell analysis. Nat Biotechnol 38, 970-979. 2020/06/28.

11. Blanco-Melo D, Nilsson-Payant BE, Liu WC, Uhl S, Hoagland D, Moller R, Jordan TX, Oishi K, Panis M, Sachs D, et al. (2020). Imbalanced Host Response to SARS-CoV-2 Drives Development of COVID-19. Cell 181, 1036-1045 e1039. 2020/05/18.