Expanded View Figures

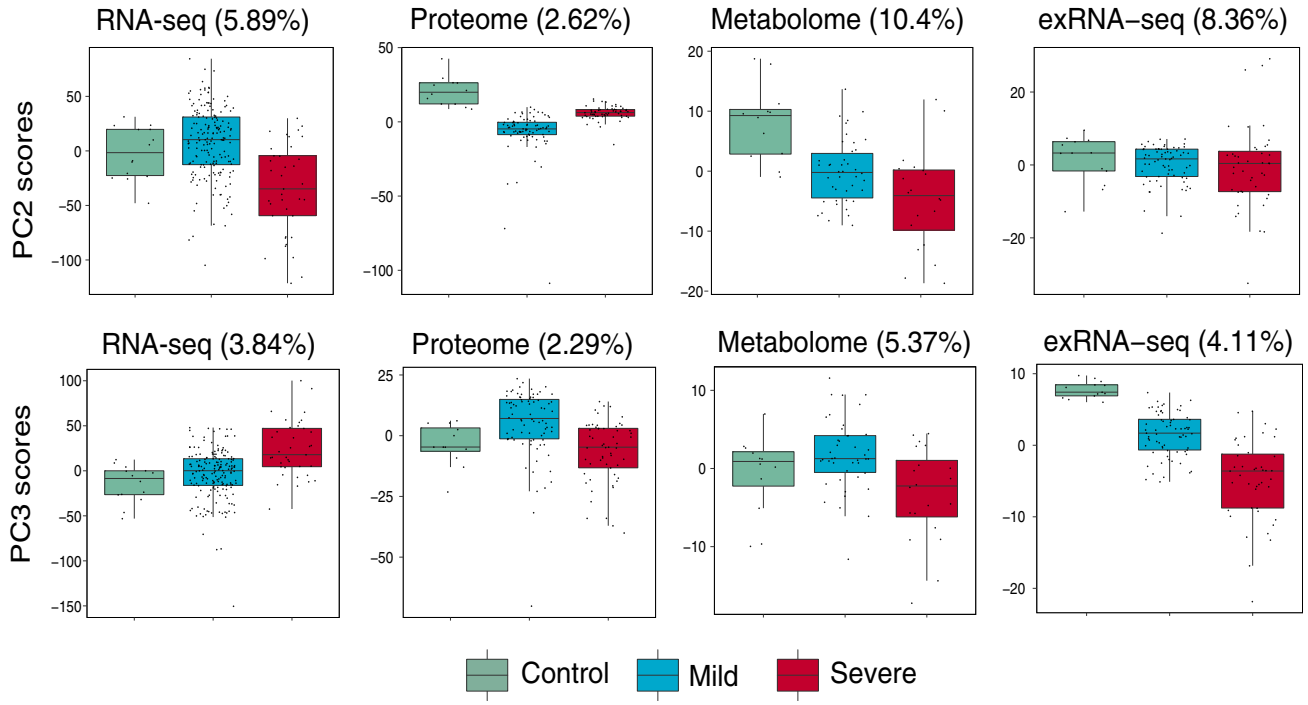

Figure EV1. Scores of principal components 2 and 3 (PC2 and PC3) of each sample from the RNA-seq, proteome, metabolome, and exRNA-seq principal component analyses.

For RNA-seq: control, $n = 14$; mild, $n = 179$; severe, $n = 37$; For proteome, control, $n = 12$; mild, $n = 80$; severe, $n = 57$; For metabolome, control, $n = 12$; mild, $n = 42$; severe, $n = 20$; For exRNA-seq: control, $n = 14$; mild, $n = 179$; severe, $n = 37$. The horizontal box lines in the boxplots represent the first quartile, the median, and the third quartile. Whiskers denote the range of points within the first quartile $-$ 1.5 $\times$ the interquartile range and the third quartile $+$ 1.5 $\times$ the interquartile range.

© 2020 The Authors

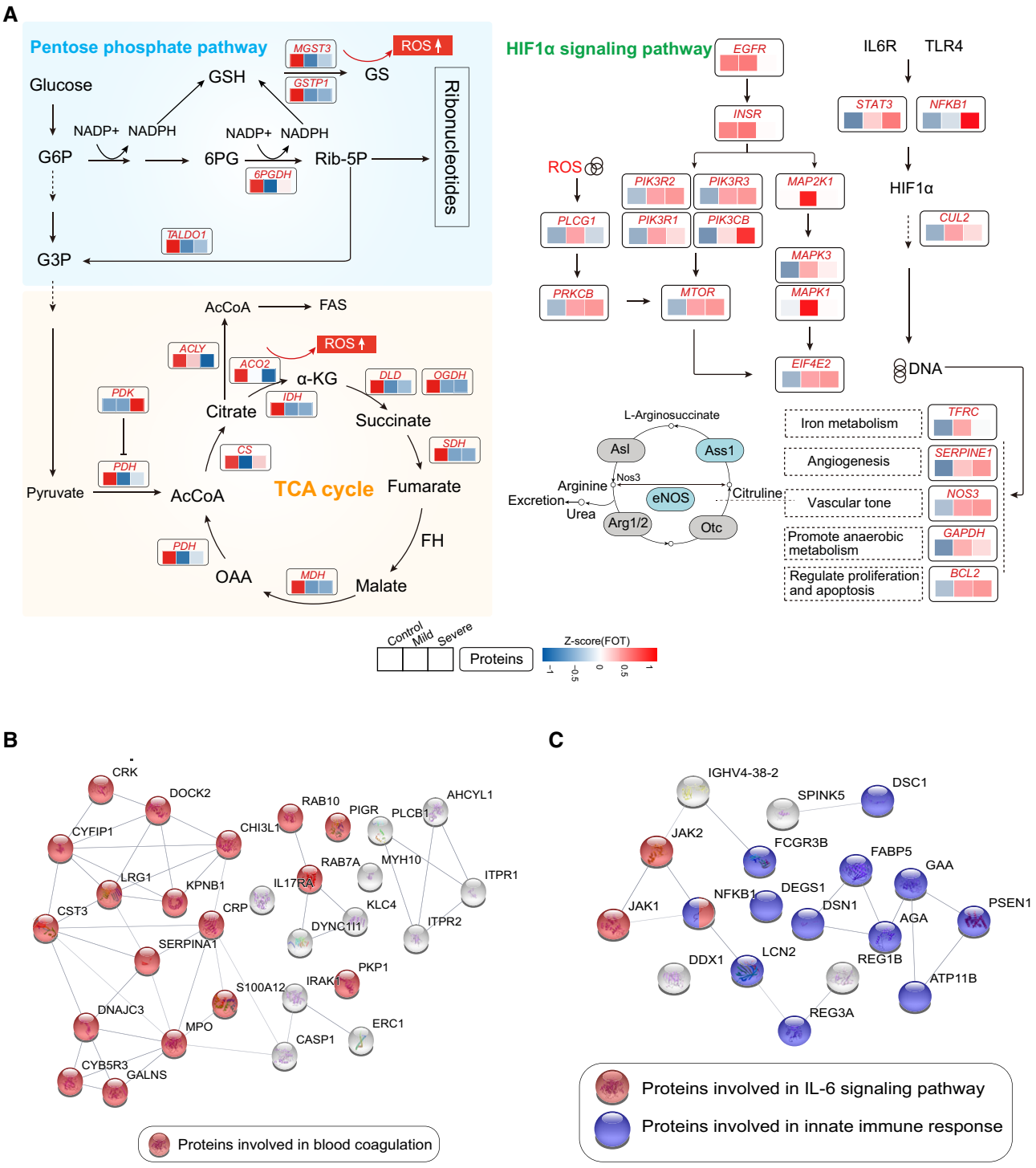

Figure EV2. Molecular variation associated with COVID-19 pathophysiology.

A Systematic summary of the proteins and signaling cascades significantly altered in healthy control patients (TCA, PEP) and in mild or severe patients (HIF-1α). Values for each protein in all samples analyzed (columns) are color-coded based on expression levels: low (blue) and high (red) z-scored FOT. Blue and gray circles present proteins that could and could not be detected in this study, respectively; eNOS, endothelial nitric oxide synthase; OTC, Ornithine carbamoyltransferase, mitochondrial; Asl, argininosuccinate lyase; Ass1, argininosuccinate synthase; Arg1/2, arginase-1/-2.

B Network indicating protein–protein interactions among module 1 enriched proteins.

C Network indicating protein–protein interactions among module 2 enriched proteins.

© 2020 The Authors

Figure EV3. Comparative analysis of tissue injury in mild and severe COVID-19-infected patients.

A, B The heatmap indicates expression patterns of tissue-enhanced biomarkers among the healthy control, mild, and severe patient groups. (A), tissue-enhanced proteins upregulated in severe patients; (B), tissue-enhanced proteins upregulated in mild patients. Values for each protein in all samples analyzed (columns) are color-coded based on expression levels: low (blue) and high (red) z-scored FOT. The bar plots indicate GO processes and pathways enriched by tissue-specific proteins upregulated in mild patients (A) and severe patients (B) (Fisher's exact test, $P < 0.05$).

C exRNA data related to tissue injury were collected from publications. The majority of tissue injury-related exRNAs across all tissues analyzed showed differential expression. Differentially expressed exRNAs were identified using a t-test < 0.05 and fold change> 2 or < 0.5.

© 2020 The Authors

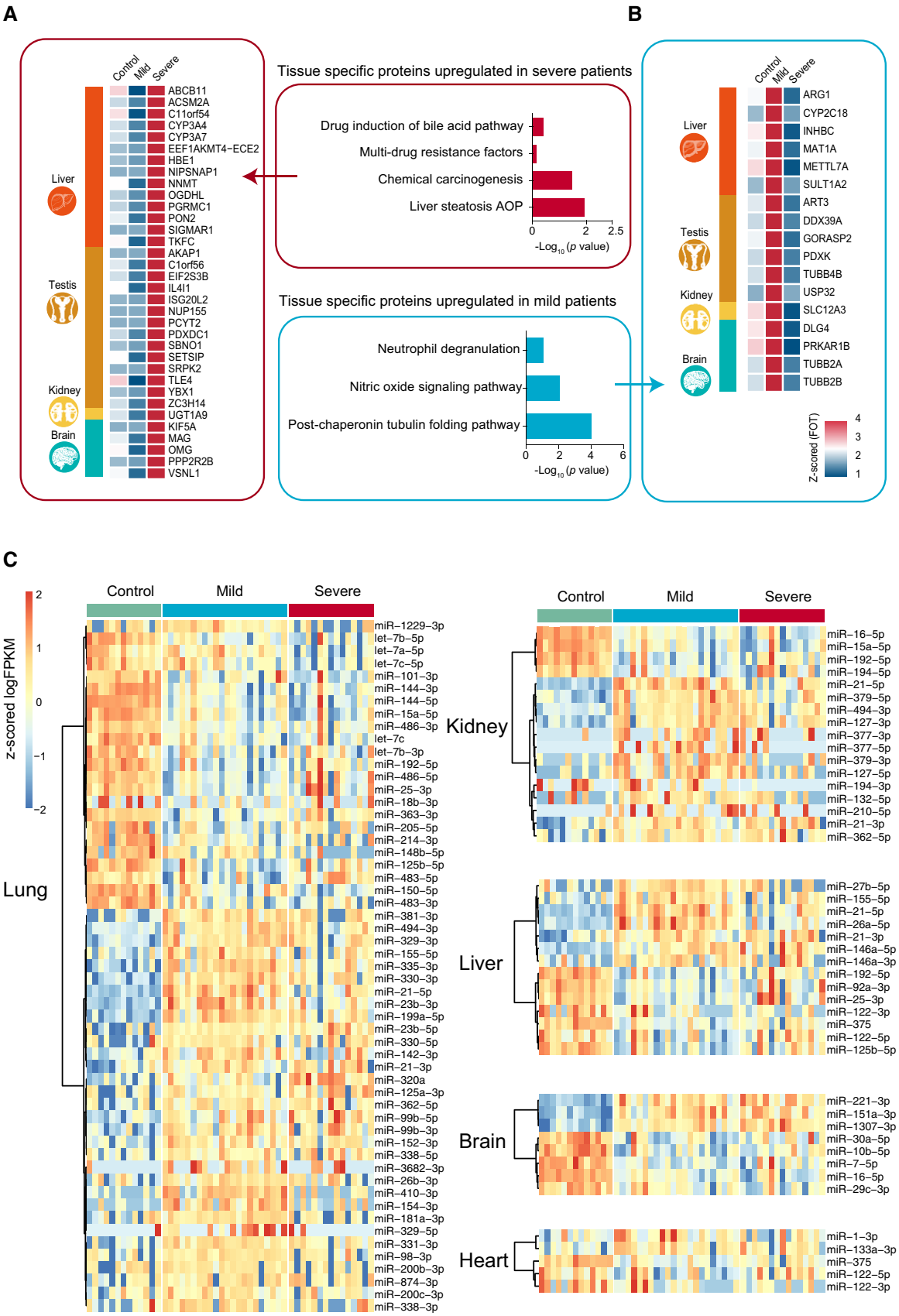
A
Control
Mild
Severe
ABCB11
ACSM2A
C11orf54
CYP3A4
CYP3A7
EEF1AKMT4-ECE2
HBE1
NIPSNAP1
NNMT
OGDHL
PGRMC1
PON2
SIGMAR1
TKFC
AKAP1
C1orf56
EIF2S3B
IL4I1
ISG20L2
NUP155
PCYT2
PDXDC1
SBNO1
SETSIP
SRPK2
TLE4
YBX1
ZC3H14
UGT1A9
KIF5A
MAG
OMG
PPP2R2B
VSNL1
Liver
Testis
Kidney
Brain
Tissue specific proteins upregulated in severe patients
Drug induction of bile acid pathway
Multi-drug resistance factors
Chemical carcinogenesis
Liver steatosis AOP
0
2
2.5
-Log10 (p value)
Tissue specific proteins upregulated in mild patients
Neutrophil degranulation
Nitric oxide signaling pathway
Post-chaperonin tubulin folding pathway
0
2
4
6
-Log10 (p value)
B
Control
Mild
Severe
ARG1
CYP2C18
INHBC
MAT1A
METTL7A
SULT1A2
ART3
DDX39A
GORASP2
PDXK
TUBB4B
USP32
SLC12A3
DLG4
PRKAR1B
TUBB2A
TUBB2B
Liver
Testis
Kidney
Brain
Z-scored (FOT)
4
3
2
1
C
Control
Mild
Severe
z-scored logFPKM
2
1
0
−1
−2
Lung
miR−1229−3p
let−7b−5p
let−7a−5p
let−7c−5p
miR−101−3p
miR−144−3p
miR−144−5p
miR−15a−5p
miR−486−3p
let−7c
let−7b−3p
miR−192−5p
miR−486−5p
miR−25−3p
miR−18b−3p
miR−363−3p
miR−205−5p
miR−214−3p
miR−148b−5p
miR−125b−5p
miR−483−5p
miR−150−5p
miR−483−3p
miR−381−3p
miR−494−3p
miR−329−3p
miR−155−5p
miR−335−3p
miR−330−3p
miR−21−5p
miR−23b−3p
miR−199a−5p
miR−23b−5p
miR−330−5p
miR−142−3p
miR−21−3p
miR−320a
miR−125a−3p
miR−362−5p
miR−99b−5p
miR−99b−3p
miR−152−3p
miR−338−5p
miR−3682−3p
miR−26b−3p
miR−410−3p
miR−154−3p
miR−181a−3p
miR−329−5p
miR−331−3p
miR−98−3p
miR−200b−3p
miR−874−3p
miR−200c−3p
miR−338−3p
Control
Mild
Severe
Kidney
miR−16−5p
miR−15a−5p
miR−192−5p
miR−194−5p
miR−21−5p
miR−379−5p
miR−494−3p
miR−127−3p
miR−377−3p
miR−377−5p
miR−379−3p
miR−127−5p
miR−194−3p
miR−132−5p
miR−210−5p
miR−21−3p
miR−362−5p
Liver
miR−27b−5p
miR−155−5p
miR−21−5p
miR−26a−5p
miR−21−3p
miR−146a−5p
miR−146a−3p
miR−192−5p
miR−92a−3p
miR−25−3p
miR−122−3p
miR−375
miR−122−5p
miR−125b−5p
Brain
miR−221−3p
miR−151a−3p
miR−1307−3p
miR−30a−5p
miR−10b−5p
miR−7−5p
miR−16−5p
miR−29c−3p
Heart
miR−1−3p
miR−133a−3p
miR−375
miR−122−5p
miR−122−3p

Figure EV3.

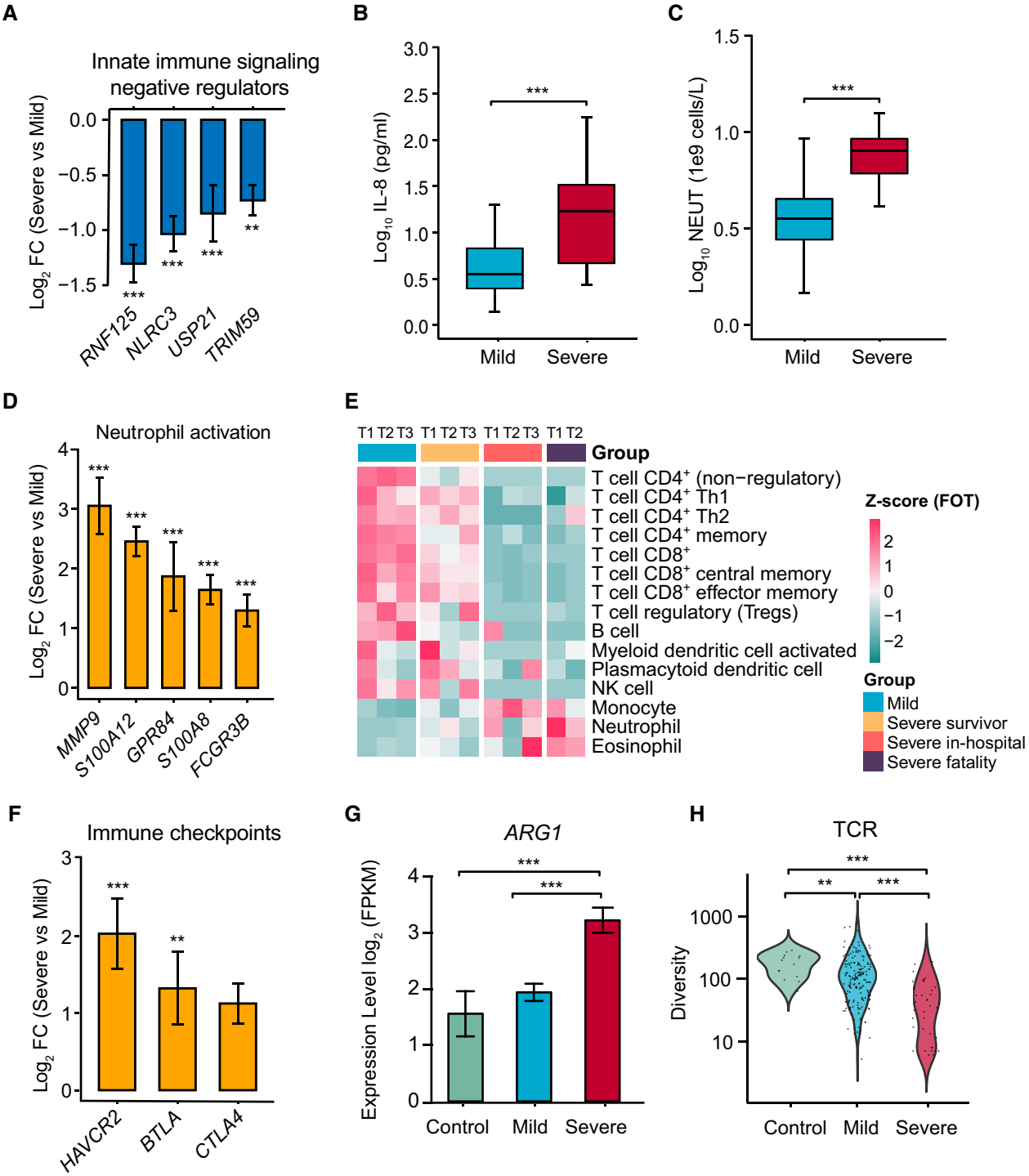

Figure EV4.

© 2020 The Authors

Figure EV4. Immune characteristics in COVID-19-infected patients.

A Normalized gene expression of innate immune signaling negative regulators in mild versus severe COVID-19-infected patient comparisons. Mild, $n = 179$; severe, $n = 37$. Differences between groups were estimated using Student's *t*-test. $^{**}P < 0.01$; $^{***}P < 0.001$.

B Level of plasma cytokines IL-8 between mild and severe patient groups. Mild, $n = 35$; severe, $n = 16$. Differences between groups were estimated using Student's *t*-test. $^{***}P < 0.001$. The horizontal box lines in the boxplots represent the first quartile, the median, and the third quartile. Whiskers denote the range of points within the first quartile − 1.5 × the interquartile range and the third quartile + 1.5 × the interquartile range.

C Absolute neutrophil count (NEUT) between mild and severe patient groups. Mild, $n = 50$; severe, $n = 16$. Differences between groups were estimated using Student's *t*-test. $^{***}P < 0.001$. The horizontal box lines in the boxplots represent the first quartile, the median, and the third quartile. Whiskers denote the range of points within the first quartile − 1.5 × the interquartile range and the third quartile + 1.5 × the interquartile range.

D Normalized expression of neutrophil activation genes in mild versus severe COVID-19-infected patient comparisons. Mild, $n = 179$; severe, $n = 37$. Data were represented as means ± SEM. Differences between groups were estimated using Student's *t*-test. $^{***}P < 0.001$.

E Cell type enrichment analysis of the RNA sequencing data using the X-Cell tool among COVID-19-infected patient subgroups from longitudinal samples. The colors of the heatmap represent the Z-scaled values of fractions of each cell type from different stages and subgroups. The red color stands for a higher proportion whereas the blue color denotes a lower cell population.

F Expression levels of immune checkpoints, normalized to *CD3G* mild versus severe patient comparisons. Mild, $n = 179$; severe, $n = 37$. Data were represented as means ± SEM. Differences between groups were estimated using Student's *t*-test. $^{**}P < 0.01$; $^{***}P < 0.001$.

G Expression levels of *ARG1* between control and COVID-19-infected patient subgroups. Control, $n = 14$; mild, $n = 179$; severe, $n = 37$. Data were represented as means ± SEM. Differences between groups were estimated using Student's *t*-test. $^{***}P < 0.001$.

H Comparison of TCR diversity between healthy controls and COVID-19-infected patient subgroups. Control, $n = 14$; mild, $n = 179$; severe, $n = 37$. Differences between groups were estimated using Student's *t*-test. $^{**}P < 0.01$; $^{***}P < 0.001$.

© 2020 The Authors

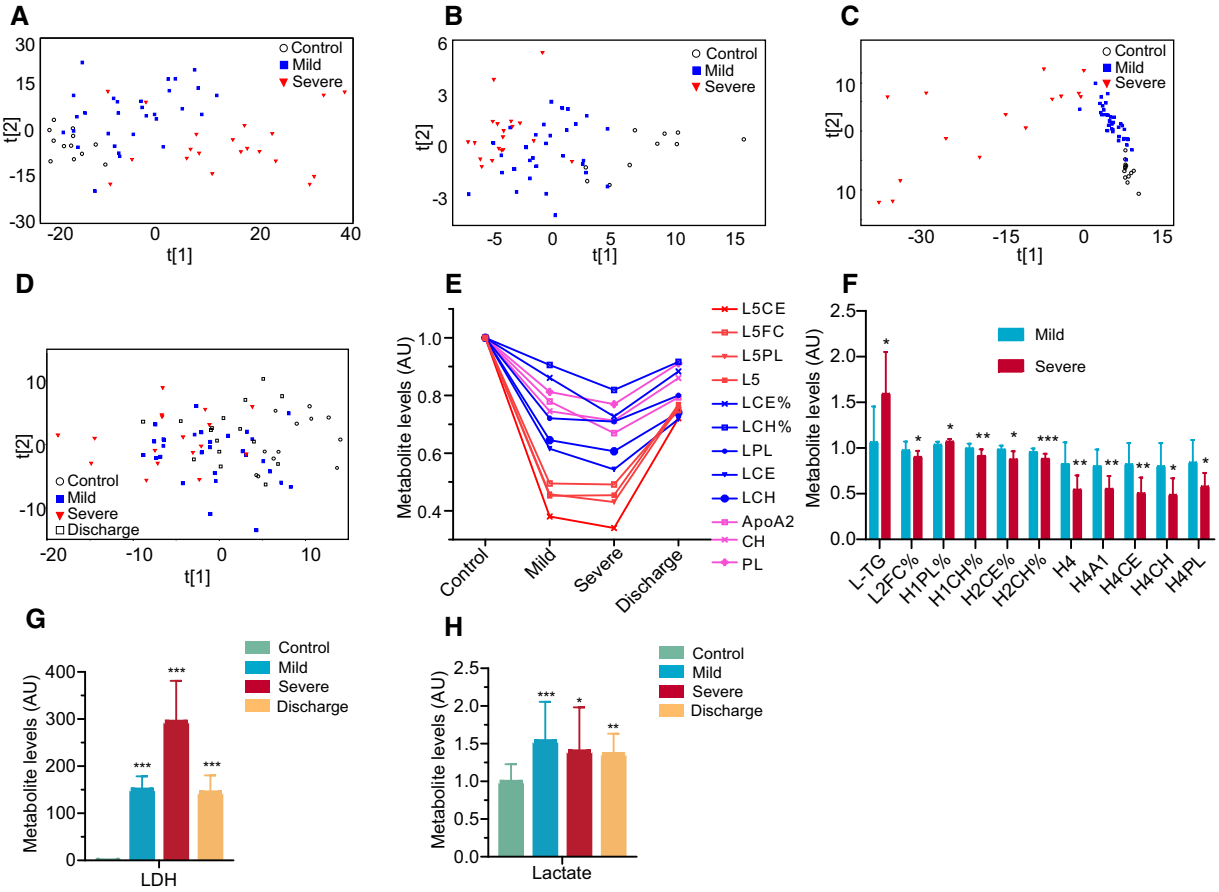

Figure EV5. COVID-19-caused changes in plasma metabolomics and clinical biochemistry associated with disease severity.

A–C COVID-19 severity was associated with plasma metabolomic phenotypes defined by all MS-detectable lipidomic compounds (A), hydrophilic metabolites (B), and all NMR-detectable metabolite signals (C).

D PCA scores revealed a variation in plasma metabolomic trajectory among healthy controls, patients with mild and severe COVID-19, and upon discharge. The discharge group comprised all patients (mild and severe) that were recovered and discharged.

E–H COVID-19 severities are associated with changes in levels of the compositional components of lipoprotein subclasses (E and F), lactate (H), lactate dehydrogenase (G) in blood plasma. Data in bar plots are represented as means ± SD. Control, $n = 12$; mild, $n = 29$; severe, $n = 17$; discharge, $n = 16$; Differences between groups were estimated using Student's *t*-test. $^{*}P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$. HDL: high-density lipoprotein; IDL: intermediate-density lipoprotein; LDL: low-density lipoprotein; VLDL: very low-density lipoprotein; TG: triglycerides; FC: free cholesterol; CE: cholesteryl esters; CH: total cholesterol (i.e., FC + CE); PL: total phospholipids; L5CE, L5FC, L5PL:CE, FC and PL in LDL5; L5:LDL5; LCE, LCH: CE and CH in LDL; LCE%, LCH%: percentages of LCE and LCH in total lipids of LDL; LPL: PL in LDL; L-TG: TG in LDL; L2FC%: percentages of L2FC (FC in LDL2) in total lipids of LDL2; H1PL%: percentages of H1PL (PL in HDL1) in total lipids of HDL1; H1CH%: percentages of H1CH (CH in HDL1) in total lipids of HDL1; H2CE%: percentages of H2CE (CE in HDL2) in total lipids of HDL2; H2CH%: percentages of H2CH (CH in HDL2) in total lipids of HDL2; H4A1, H4CE, H4CH, H4PL: ApoA1, CE, CH and PL in HDL4; AU: The metabolite concentration of each sample was normalized using the average of the control group; LDH: lactate dehydrogenase.

Figure EV6. Training and validation set performance in exRNA-, transcriptome-, proteome-, and clinical- models.

A Workflow of prediction model construction.

B Performance of AI models in the training and validation set based on exRNA ($n = 37$), transcriptome ($n = 63$), proteome ($n = 31$), and the corresponding clinical covariate data sets ($n = 37$). Data are represented as means ± SD.

C Learning curve model comparison (LCMC) revealing sample size effects on the accuracy and variability of the predictive models using cross-validation. Each individual root means square error (RMSE) learning curve and the average for each of 8 models is shown. Each color represents a model. A solid line represents the average for each model, whereas a dashed line represents one random iteration for each model.

© 2020 The Authors

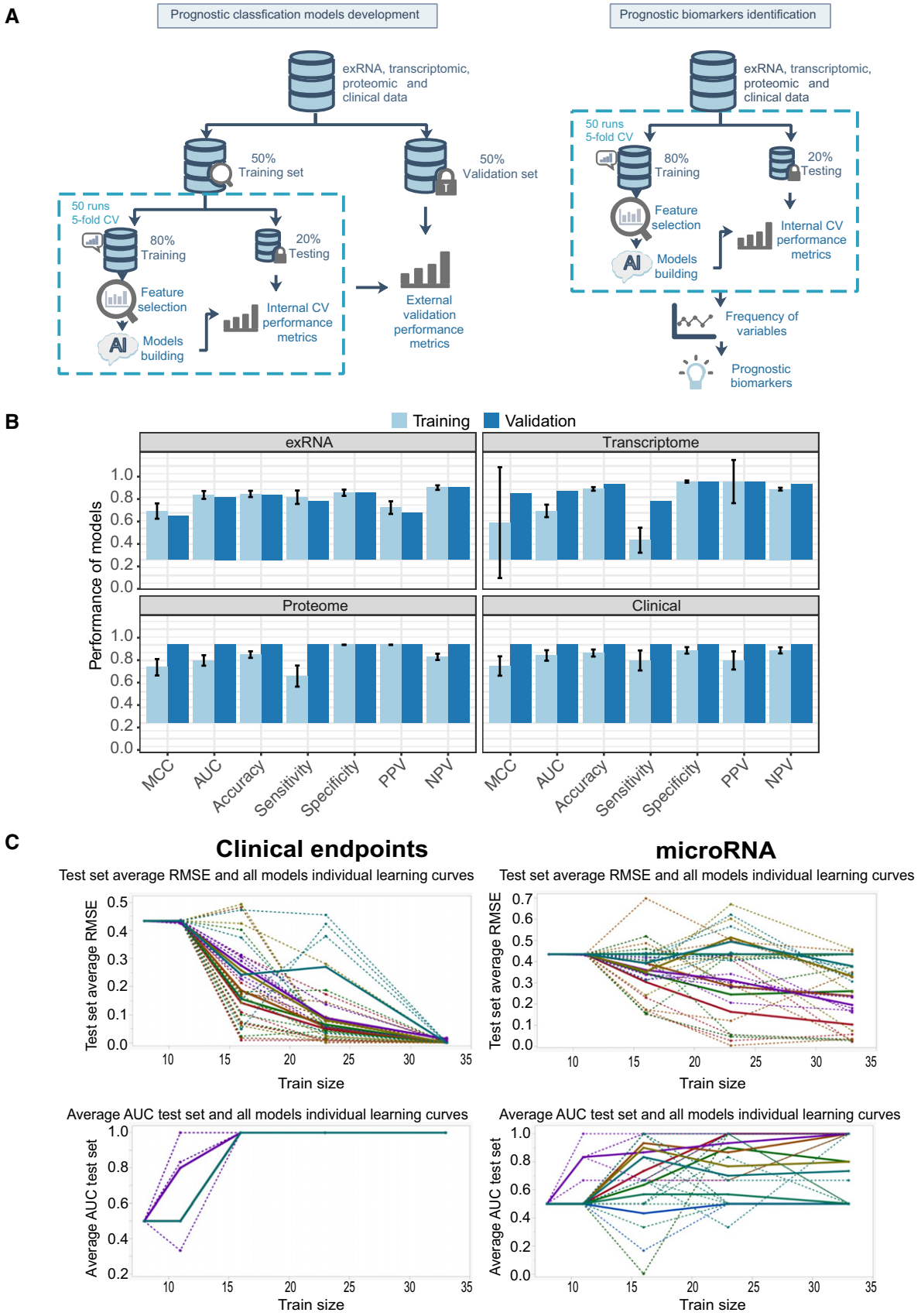
A
Prognostic classfication models development
Prognostic biomarkers identification
exRNA, transcriptomic, proteomic and clinical data
50% Training set
50% Validation set
50 runs 5-fold CV
80% Training
20% Testing
Feature selection
Models building
Internal CV performance metrics
External validation performance metrics
Frequency of variables
Prognostic biomarkers
B
Training
Validation
exRNA
Transcriptome
Proteome
Clinical
Performance of models
MCC
AUC
Accuracy
Sensitivity
Specificity
PPV
NPV
C
Clinical endpoints
microRNA
Test set average RMSE and all models individual learning curves
Test set average RMSE
Train size
Average AUC test set and all models individual learning curves
Average AUC test set

Figure EV6.