Supplementary Information for:

COVID-19 severity is associated with immunopathology and multi-organ damage

Yan-Mei Chen[1,4], Yuanting Zheng[1, 4], Ying Yu[1, 4], Yunzhi Wang[1, 4], Qingxia Huang[1, 4], Feng Qian[1, 4], Lei Sun[2, 4], Zhi-Gang Song[1], Ziyin Chen[1], Jinwen Feng[1], Yanpeng An[1], Jingcheng Yang[1], Zhenqiang Su[1], Shanyue Sun[1], Fahui Dai[1], Qinsheng Chen[1], Qinwei Lu[1], Pengcheng Li[1], Yun Ling[1], Zhong Yang[1], Huiru Tang[1], Leming Shi[1], Li Jin[1], Edward C. Holmes[3], Chen Ding[1]*, Tong-Yu Zhu[1]*, Yong-Zhen Zhang[1]*

[1]Shanghai Public Health Clinical Center, State Key Laboratory of Genetic Engineering, School of Life Sciences and Human Phenome Institute, Fudan University, Shanghai, China.

[2]Institute of Developmental Biology and Molecular Medicine, Fudan University, Shanghai, China.

[3]Marie Bashir Institute for Infectious Diseases and Biosecurity, School of Life and Environmental Sciences and School of Medical Sciences, The University of Sydney, Sydney, New South Wales, Australia.

[4]These authors contributed equally: Yan-Mei Chen, Yuanting Zheng, Ying Yu, Yunzhi Wang, Qingxia Huang, Feng Qian, Lei Sun

*Correspondence to: Yong-Zhen Zhang, Email: zhangyongzhen@shphc.org.cn; Tong-Yu-Zhu, Email: zhutongyu@shphc.org.cn; Chen Ding, Email: chend@fudan.edu.cn.

SUPPLEMENTARY FIGURES

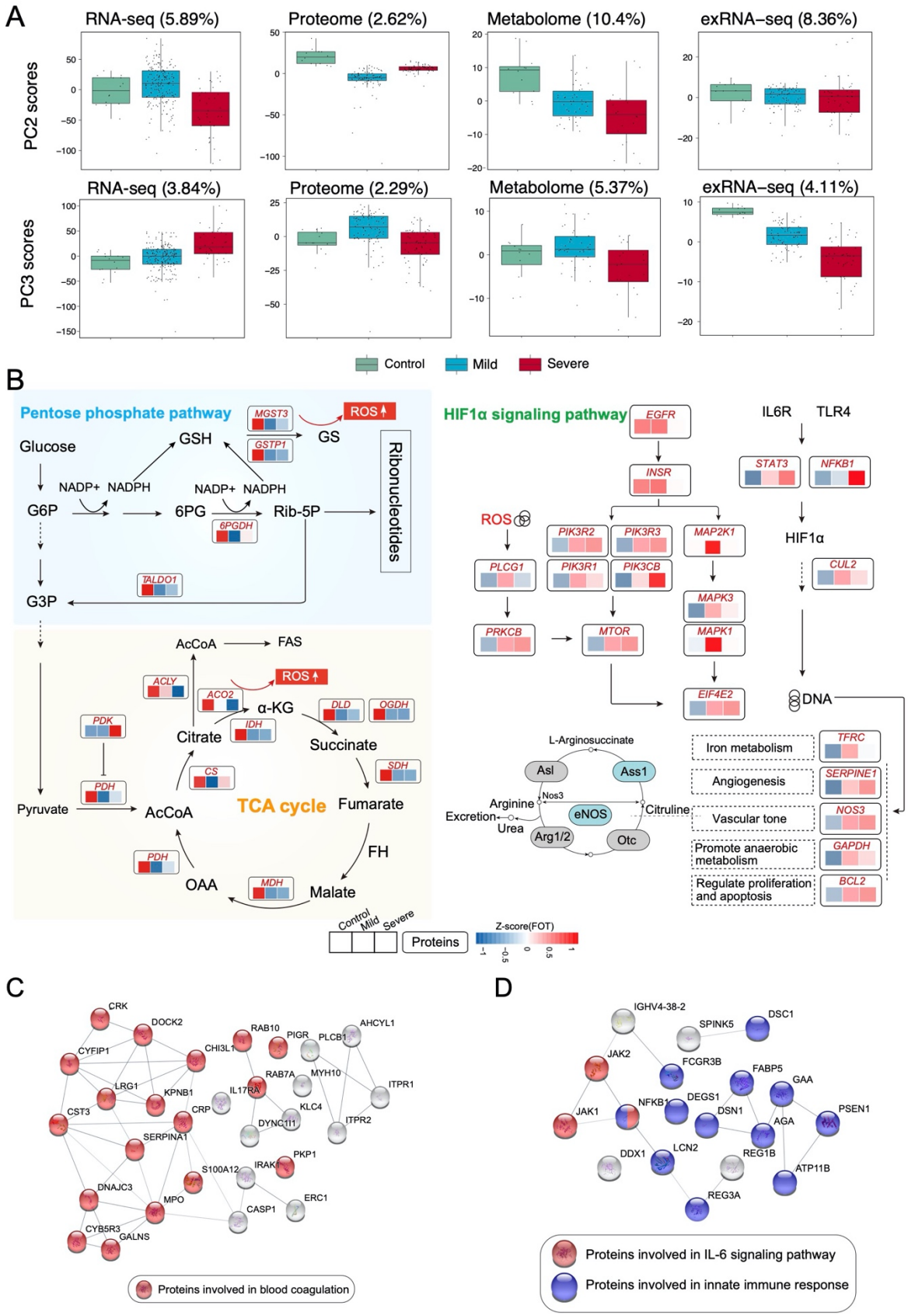


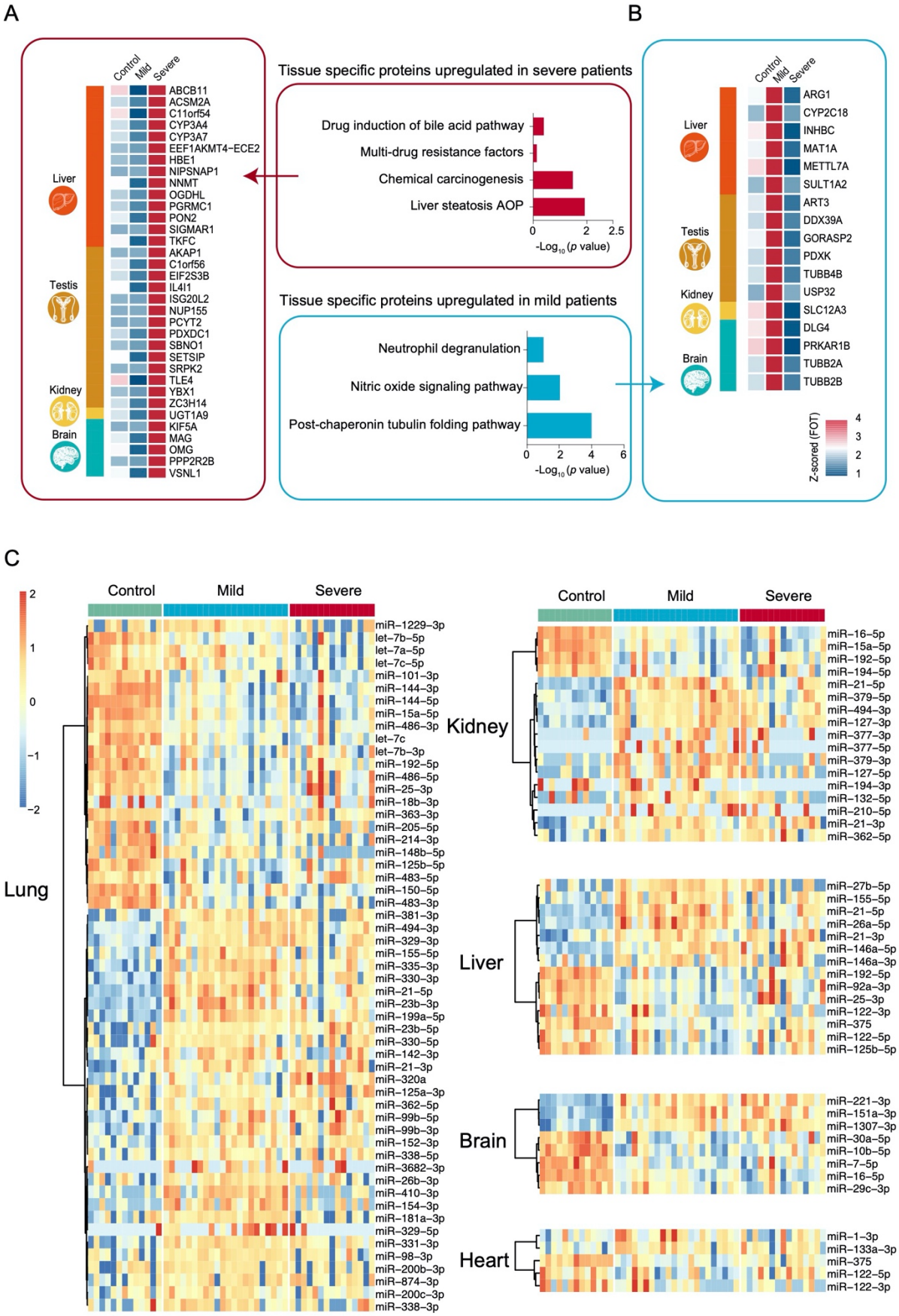
Fig. S1 Molecular variation associated with COVID-19 pathophysiology.

(A) 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.

(B) 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.

(C) Network indicating protein-protein interactions among module 1 enriched proteins.

(D) Network indicating protein-protein interactions among module 2 enriched proteins.


A
B
C
Control
Mild
Severe
Liver
Testis
Kidney
Brain
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
Tissue specific proteins upregulated in severe patients
Drug induction of bile acid pathway
Multi-drug resistance factors
Chemical carcinogenesis
Liver steatosis AOP
-Log10 (p value)
Tissue specific proteins upregulated in mild patients
Neutrophil degranulation
Nitric oxide signaling pathway
Post-chaperonin tubulin folding pathway
-Log10 (p value)
ARG1
CYP2C18
INHBC
MAT1A
METTL7A
SULT1A2
ART3
DDX39A
GORASP2
PDXK
TUBB4B
USP32
SLC12A3
DLG4
PRKAR1B
TUBB2A
TUBB2B
Z-scored (FOT)
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
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

35

Fig. S2 Comparative analysis of tissue injury in mild and severe COVID-19 patients.

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

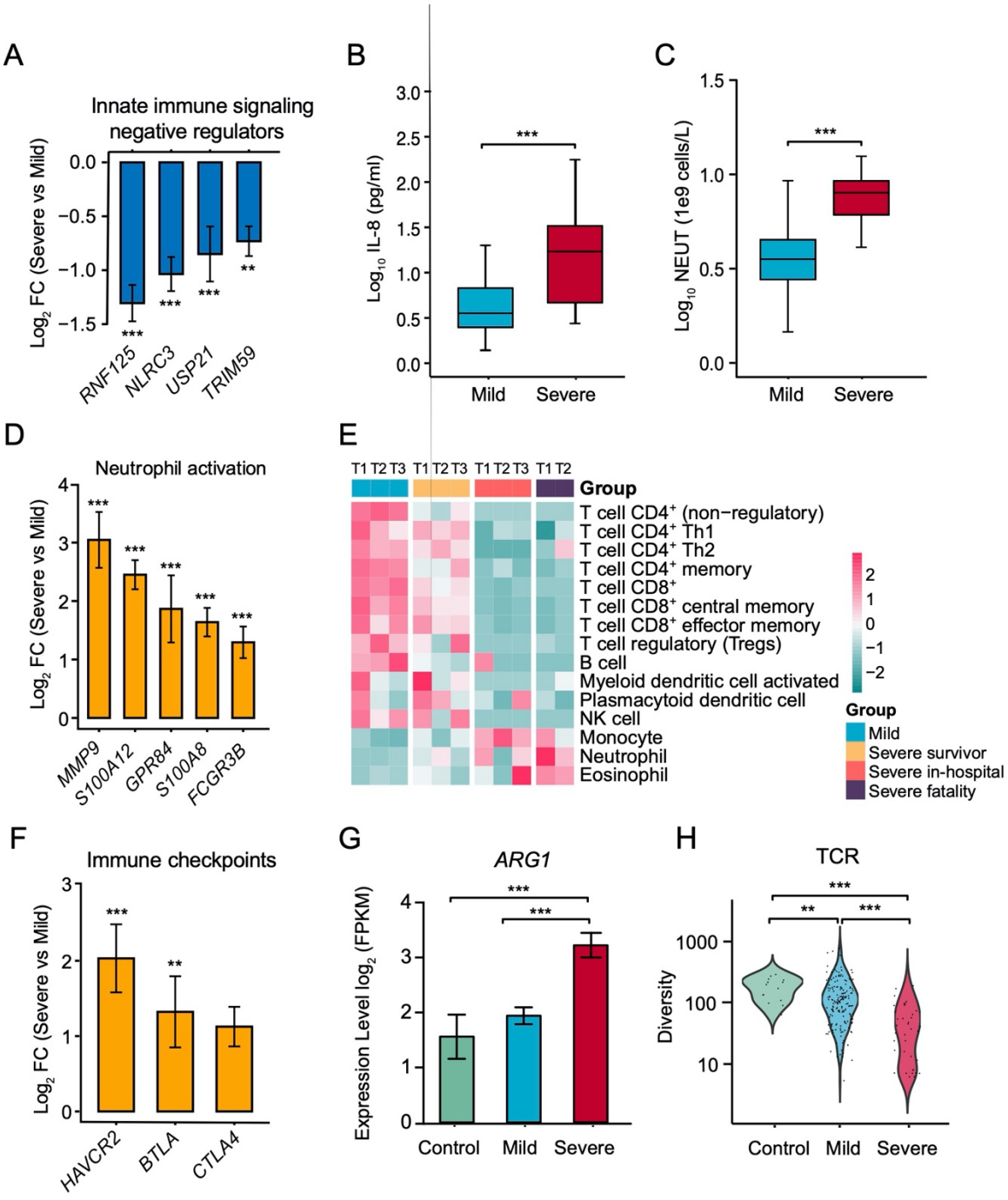


Fig. S3 Immune characteristics in COVID-19 patients.

(A) Normalized gene expression of innate immune signaling negative regulators in mild versus severe COVID-19 patient comparisons.

(B) Level of plasma cytokines IL-8 between mild and severe patient groups.

(C) Absolute neutrophil count (NEUT) between mild and severe patient groups.

(D) Normalized expression of neutrophil activation genes in mild versus severe COVID-19 patient comparisons.

(E) Cell type enrichment analysis of the RNA sequencing data using the xCell tool among COVID-19 patient subgroups from longitudinal samples.

(F) Expression levels of immune checkpoints, normalized to *CD3G* mild versus severe patient comparisons.

(G) Expression levels of *ARG1* between control and COVID-19 patient subgroups.

(H) Comparison of TCR diversity between healthy controls and COVID-19 patient subgroups. FC, fold change. Data are represented as means ± SEM. ** $p < 0.01$; ***$p < 0.001$ (t-test).

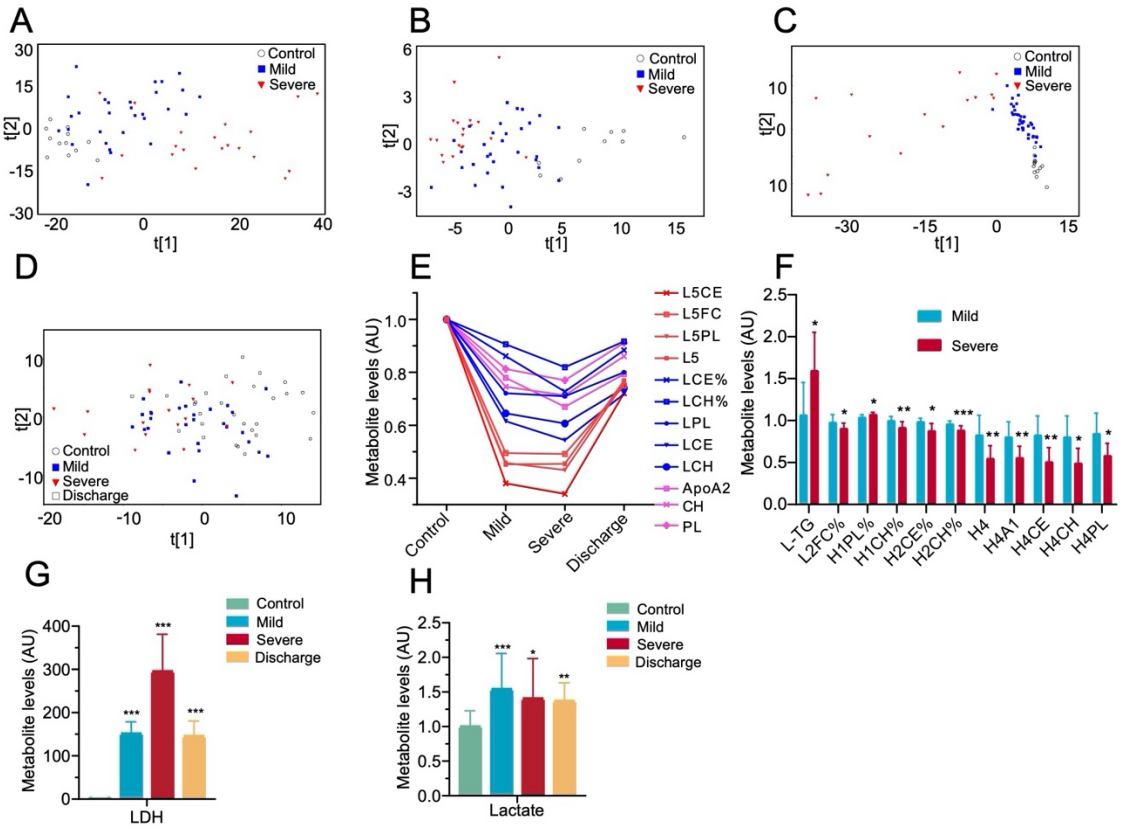


Fig. S4 COVID-19-caused changes in plasma metabolomics and clinical biochemistry associated with disease severity.

(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 (G), lactate dehydrogenase (H) in blood plasma. AU: The metabolite concentration of each sample is normalized using the

77 average of the control group; LDH: lactate dehydrogenase; H1: HDL1; H2: HDL2; H3: HDL3;
78 H4: HDL4; L: LDL; L2: LDL2; TG: triglycerides; FC: free cholesterol; CE: cholesteryl esters;
79 CH: total cholesterol (i.e., FC + CE); PL: total phospholipids; L5CE, L5FC, L5PL: cholesteryl
80 esters (CE), free cholesterol (FC) and total phospholipids (PL) in LDL5 (L5); ApoA2: total
81 ApoA2 (in both nascent and mature HDL); CH: total cholesterol (both FC and CE); H4A1,
82 H4A2, H4CH, H4FC, H4CE, H4PL: ApoA1, ApoA2, CH, FC, CE and PL in HDL4; LDPL,
83 LDL-CH, LDL-CE: PL, CH, CE in LDL; L-CE%, L-FC%: percentages of CE and FC in all
84 lipids of LDL. ***$p < 0.001$.
85

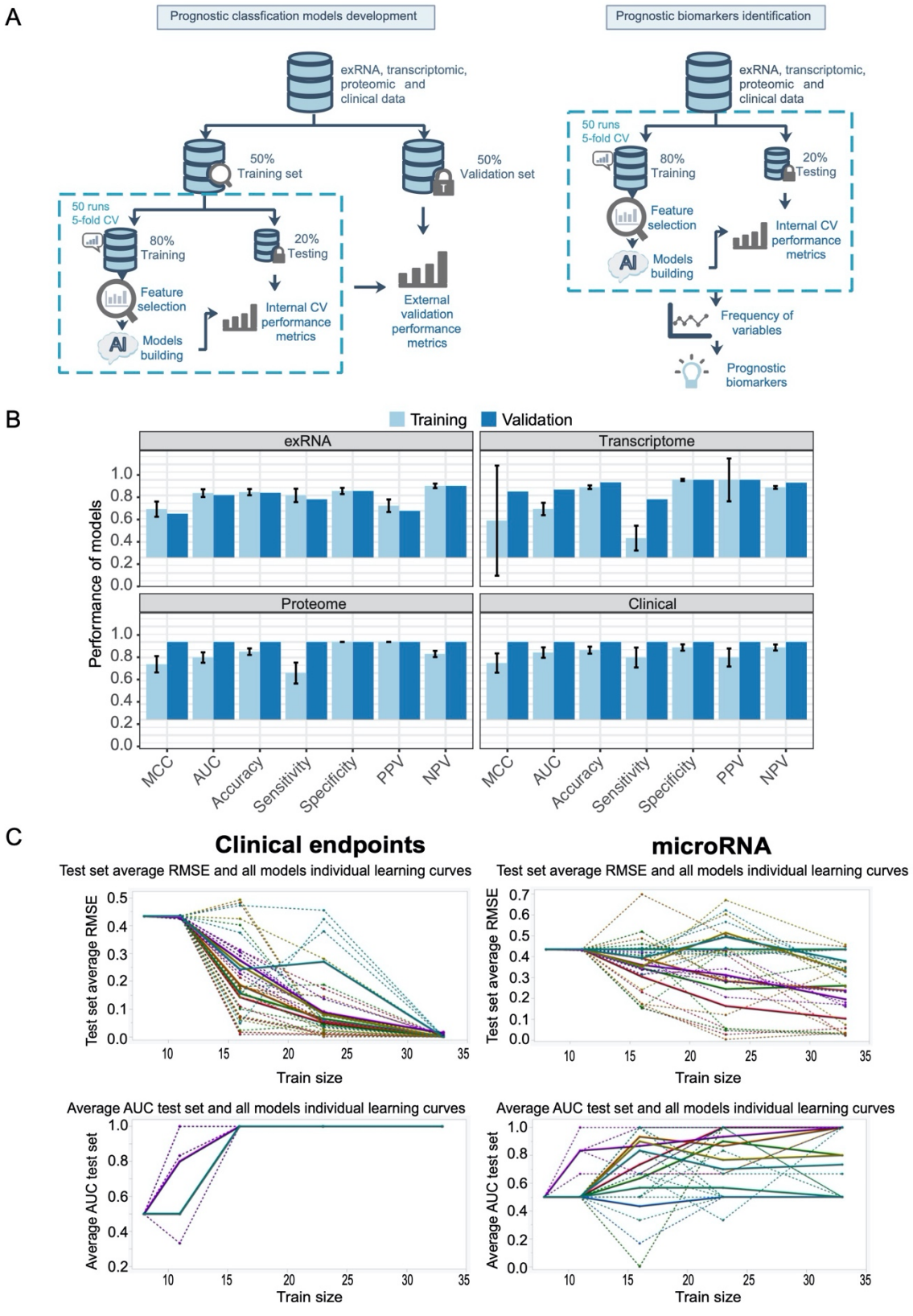


Fig. S5 Training and validation set performance in exRNA-, transcriptome-, proteome-, and clinical- models.

(A) Workflow of prediction model construction.

(B) Performance of AI models in the training and validation set based on exRNA, transcriptome, proteome, and the corresponding clinical covariate data sets. The model performance of 5-fold cross-validation was assessed using Matthews correlation coefficient (MCC), AUC, accuracy, sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV).

(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. The LCMC suggested that with up to 15 samples, 8 partition tree models reached AUC as 1 for clinical endpoints, but more than 23 samples were needed for one model and more than 30 samples need for three models to reach AUC as 1 for microRNA.

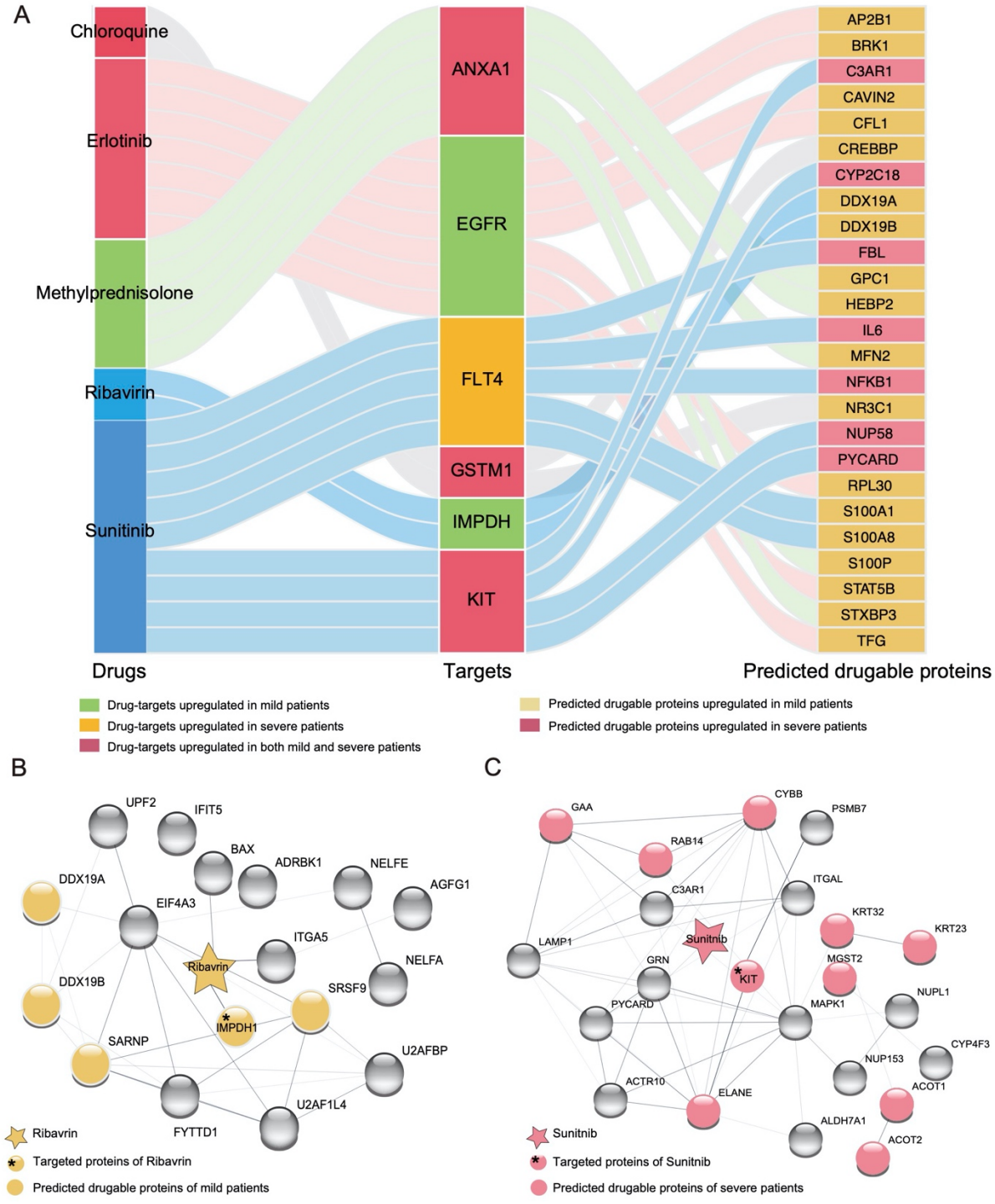


Fig. S6 A drug-protein network for predicting novel drug candidates.

(A) The snaky plot indicates the correlation network of drug, drug targets and predicted-drug targets.

(B) Network indicating the interactions between Ribavrin, IMPDH1 (target of Ribavrin) and predicted druggable proteins in mild COVID-19 patients.

(C) Network indicating the interactions between Sunitinib, KIT (target of Sunitinib) and predicted druggable proteins in severe COVID-19 patients.