THE LANCET
Digital Health

Supplementary appendix

This appendix formed part of the original submission and has been peer reviewed. We post it as supplied by the authors.

Supplement to: Byeon SK, Madugundu AK, Garapati K, et al. Development of a multiomics model for identification of predictive biomarkers for COVID-19 severity: a retrospective cohort study. *Lancet Digit Health* 2022; published online July 11. https://doi.org/10.1016/S2589-7500(22)00112-1.

Supplementary Material

Methods

Study design

Plasma samples were collected in two 10-ml EDTA and a 4.5-ml tube with sodium citrate and centrifuged at 2,000×g for 10 min and fractionated by automation (Janus robotic liquid handlers; PerkinElmer). Typically, 6-8 hours were elapsed between sample collection and freezing.

Cytokines

Regardless of SARS-CoV-2 status, all plasma samples were inactivated by adding TritonX-100 to plasma to final concentration of 1% TritonX-100, followed by incubation in the biosafety hood at room temperature for 2 hours. 1 µl of each sample was incubated with antibody probes at 4°C overnight. Matched pairs of antibodies attached to unique DNA oligonucleotides were used for each of the targeted plasma proteins. After binding, the extension mix was added and the products were extended and amplified using 17 cycles of PCR (Applied Biosystems 9700, Life Technologies). Next, 2.8 µl of each PCR product was added to the detection mix and loaded into the sample wells of a Fluidigm 96.96 Dynamic Array plate (Fluidigm Corporation) while kit primers were loaded into the primer wells. The Dynamic Array was primed in a Fluidigm HX IFC controller and then loaded into the Fluidigm Biomark imaging thermocycler for quantitative PCR. Quantification cycle values for each measurement were determined using Fluidigm's Real-Time PCR Analysis software and BiomarkDataCollection version 4.1.3. Data were normalized using the extension positive control and the negative control quantitation cycle values. Each detected protein produced a unique barcode resulting in a high-sensitivity, high-specificity readout with NGS. The Explore panel includes protein targets from Cardiometabolic, Inflammation, Neurology and Oncology panels. Pooled plasma samples were used to evaluate the average intra-assay and inter-assay coefficient of variance, which were calculated to be 13.5% and 23.8% across the 4 panels, respectively.

Lipidomics

Lipids were extracted from 75 µl of plasma using an automated BUME extraction after adding the identical amount of deuterated internal standards prior to extraction.[1] The extracts were dried under nitrogen and reconstituted in 0.25mL of dichloromethane:methanol (50:50) with 10mM ammonium acetate. The samples were analyzed via both positive and negative mode electrospray to maximize the coverage of lipids. The infusion-MS analysis using Sciex SelexION-5500 QTRAP was performed in multiple reaction monitoring mode (MRM) with a total of more than 1,100 MRMs, in which MS parameters were optimized for each species. Shimadzu LC was utilized for automated sample injection. Identification of lipids were based on retention index, precursor ion match (+/- 10 ppm) and MS/MS spectra. The lipid species were quantified by calculating their peak areas in reference to internal standards having the identical head groups, then multiplying by the concentration of internal standard. Total levels of lipid classes were calculated by adding the calculated concentrations of all species within each class together. Overall process variability was determined by calculating the median relative standard deviation (RSD) for all endogenous lipids present in the technical replicates and it was calculated to be 4.5%.

Metabolomics

From 225 µl of plasma, metabolites were extracted using the automated MicroLab STAR® system from Hamilton Company (Reno). For metabolite extraction, methanol was added to samples and shaken for 2 min (Glen Mills GenoGrinder 2000). Followed by centrifugation, the supernatant was collected, divided into 5 equal aliquots and dried. Each aliquot was eventually reconstituted in organic solvent appropriate for four methods of UPLC-MS/MS analyses and analyzed; one reversed phased UPLC-MS/MS in positive ion mode with elution gradient optimized for hydrophobic metabolites (methanol and acetonitrile containing 0.05% perfluoropentanoic acid and 0.1% formic acid, one reversed phased UPLC-MS/MS in positive ion mode with elution gradient optimized for hydrophilic metabolites (water and methanol with 0.05% perfluoropentanoic acid and 0.01% formic acid), one reversed phased UPLC-MS/MS in negative ion mode for basic metabolites (methanol and water with 6.5 mM ammonium bicarbonate at pH 8) and one for hydrophilic interaction chromatography UPLC-MS/MS with negative ion mode (water and acetonitrile with 10 mM ammonium formate at pH 10.8). One aliquot was saved as back up. For each method, a Waters ACQUITY UPLC was coupled to Q-Exactive Hybrid Quadrupole-Orbitrap mass spectrometers from Thermo Scientific. For reversed phase LC, Waters UPLC BEH C_{18} column (2.1 mm × 100 mm, 1.7 µm) was utilized as an analytical column

while Waters UPLC BEH Amide column (2.1 mm × 150 mm, 1.7 μm) was used in hydrophilic interaction chromatography. Data-dependent analysis was carried out for all methods with scan range of m/z 70-1000 and a mass resolution of 35,000. Identification of metabolites was made based on matching with a spectral library, which was built using purified standards. In addition to retention time index, m/z of precursor and product ions were taken into consideration with +/- 10 ppm mass tolerance. Metabolites were quantified by calculating their peak areas, followed by normalization using a quality control sample that was injected periodically to monitor MS performance. Instrument variability was determined by calculating the median RSD for the internal standards that were added to each sample prior to injection into the mass spectrometers, which was calculated to be 4.9%. Overall process variability was determined by calculating the median RSD for all endogenous metabolites across the technical replicates and it was calculated to be 8.6%.

Pathway and network analysis

The protein names from Olink data were mapped to the canonical gene symbols using DAVID Bioinformatics resource. The metabolite and lipid names were mapped to the Human Metabolome Database and Kyoto Encyclopedia of Gene and Genome compound accessions using the Metaboanalyst R package (version 5.0). The remapped gene symbols and compound accessions along with their corresponding $\log_2$ fold change and adjusted *P* value were loaded into Ingenuity Pathway Analysis (IPA; Qiagen). Candidates were shortlisted for pathway analysis using an absolute $\log_2$ fold-change$\geq$0.50 and an adjusted p$\leq$0.05. IPA was set to use the fold change direction when assessing the directionality of pathway change whenever possible. The software performed the core pathway analysis using the default settings. Pathways that have an adjusted p$\leq$0.05 were considered as statistically significant and utilized for further interpretation.

Machine learning analysis

Data pre-processing was implemented as follows. Candidate variables considered for inclusion in the model included both molecular markers (cytokines, lipids, and metabolites) and demographics (age, sex, race, ethnicity, and comorbidity). The dataset was split 80/20 into a training subset and a held-out test set observations, respectively, using standard stratified splitting methods provided by the Caret package in R.[2] The paired pre-COVID-19 and post-COVID-19 samples were not included in the stratified split and were instead placed in the test set directly. When multiple observations from a single individual were available, only the earliest observation (with at least one analyte in the target feature set) was retained. Any remaining missing values in molecular analyte data were assumed to be below the limit of detection and thus were imputed using the minimum value for that analyte across all samples. Demographic (age, sex, race, ethnicity) and comorbidity (Charlson Comorbidity Index) features were preprocessed as follows. Age was reported in years. Sex was binary encoded as either Male or Female. Race was predominately "White" for this cohort, so the race feature was binary encoded as either "White" or "Other". Similarly, ethnicity was predominately "Not Hispanic or Latino" for this cohort, so the ethnicity feature was binary encoded as either "Not Hispanic or Latino" or "Other". For both race and ethnicity, missing values encoded as their own class automatically by AutoGluon. Comorbidity values were median imputed. Class labels based on the WHO ordinal scale were binned into 4 levels as previously defined.

Exploratory analysis was performed within the 80% training set, leaving the 20% test set untouched. A 10-fold stratified cross-validation within our training set was used for model selection using the Caret package in R.[2] Classification methods were selected in agreement with established best practices to deal with the high dimensionality without overfitting.[3]

To this end, we compared LASSO to the ensemble method of AutoGluon's (Version 0.1.1b20210326) tabular prediction. AutoGluon outperformed LASSO in the cross-validation on training data, and thus was chosen as the classifier for final evaluation on the test set. AutoGluon performs model stacking of a large variety of machine learning algorithms (e.g., XGboost, random forests, deep neural networks, KNN, et cetera) to create an ensemble classifier. It avoids overfitting the data automatically through a variety of sophisticated machine learning methods, such as model ensembling/stacking, regularization applied to individual models in the stacked ensemble, and careful splitting and tracking of out-of-fold data points during training. AutoGluon was run with parameters "eval_metric='accuracy', presets='medium_quality_faster_train', and time_limit=6000", leaving all other parameters were left as default. We provide our AutoGluon code as a Supplementary Material. Inclusion of the demographic and comorbidity features did not statistically significantly change the performance of the "all analytes" feature set. Additionally, a model using only

demographic and comorbidity features had statistically significantly lower accuracy than the "all analytes" model in the held-out test set. This is consistent with results from other studies showing that demographics and comorbidity features did not improve model results.[4,5]

Biomarker evaluation was performed using feature importance scores of the final AutoGluon model. Feature importance scores were generated using the "feature_importance" function of AutoGluon-Tabular with the parameters "subsample_size = None", "num_shuffle_sets = 10", and "time_limit = 12000". Input variables with a positive feature importance score represent features that, when removed, cause the model to decrease in accuracy, while variables with a negative score cause the model performance to improve when they are removed from the model. Thus, the list of features with a positive feature importance score and a $p \leq 0.05$ were considered as candidate biomarkers for predicting COVID-19 severity.

Proteomics and glycoproteomics

We had 24 individual COVID-19 patients enrolled for whom we also had samples available in the biorepository that were collected before these individuals contracted COVID-19. After samples were aliquoted for the multi-omics experiments with proteomics (proximity extension assay), lipidomics and metabolomics, there was adequate amount of plasma available for mass spectrometry-based total proteomics and glycoproteomics experiments for 21 patients of 24 patients with relative quantitation using tandem mass tags (TMT). As the limit of multiplexing and relative quantitation with tandem mass tags is 16 samples per experiment, we divided these 21 patients (and their pre-COVID-19 and COVID-19 samples numbering 42) into three separate batches of 8, 8 and 5 patients. Detailed information for these patients is provided in Supplementary Table 4. Plasma samples were diluted 1:50 and protein concentration was determined by BCA assay. For 21 individuals, plasma samples collected before and after COVID-19, equal amount of protein was trypsin digested. Dried plasma samples were reconstituted in 100 µl of 8 M urea in 50 mM triethylammonium bicarbonate (TEAB), pH 8.5 and reduction was carried out with dithiothreitol at final concentration of 10 mM (Sigma) at 37°C. Reduced samples were cooled to room temperature (RT) and 40 mM iodoacetamide (Sigma) was added. Samples were incubated for 15 minutes in dark at room temperature. TEAB was used to dilute samples 1:10 and sequencing-grade trypsin was added to a final amount ratio of 1:20 (trypsin:total protein). The digested peptides were labeled with tandem mass tag (16-plex TMT) reagents as per the manufacturer's instructions (Thermo Fisher). Briefly, peptides were resuspended in 50 mM TEAB (pH 8.0) and mixed with TMT reagents which was dissolved in anhydrous acetonitrile. After incubating for 1 hour at RT, the reaction was quenched by 1M tris buffer, pH 8.5. After labeling check, labelled samples were pooled into one and either size-exclusion chromatography or basic pH reversed-phase fractionation was performed. An aliquot of dried peptides was resuspended in 100 µl of 0.1% formic acid and injected into Superdex peptide 10/300 column (GE Healthcare) also equilibrated with 0.1% formic acid. The early fractions were collected starting at 10 minutes after injection (total run time of 130 minutes) and concatenated to 12 fractions, which were analyzed by LC-MS/MS. Another aliquot of total peptides was cleaned up by C_{18} TopTips (Glygen) and fractionated by bRPLC on a reversed phase C_{18} column (4.6 mm ×250 mm column) using an Ultimate 3000 UHPLC System. 5 mM ammonium formate, pH 9 and 5 mM ammonium formate in 90% acetonitrile, pH 9 were used as solvent A and B. Ninety-six fractions were collected for a total of 120 min and were concatenated into 12 fractions and analyzed by LC-MS/MS.

A modification of previously published LC-MS/MS parameters were used.[6] Specifically, 12 fractions from SEC were selected based on the UV profile (214 nm, 220 nm) and separately 12 concatenated fractions of bRPLC were analyzed by Orbitrap Eclipse mass spectrometer (Thermo Fisher Scientific). Liquid chromatography for separation of peptides was performed on an EASY-Spray column (75 µm × 50 cm, PepMap RSCL C_{18}, Thermo Fisher Scientific) packed with 2 µm C_{18} particles. The column was maintained at 50 °C. Solvent A and B were 0.1% formic acid in water and 0.1% formic acid in acetonitrile, respectively. Injected peptides were trapped on a trap column (100 µm × 2 cm, Acclaim PepMap100 Nano-Trap, Thermo Fisher Scientific) at a flow rate of 20 µl/min. Every run was 130 minutes with flow rate being 300 nl/min. The gradient used for separation was equilibration at 3% solvent B from 0 to 4 min, 3% to 10% sol B from 4 to 10 min, 10% to 35% sol B from 10.1 to 125 min, 35% to 80% sol B from 125 to 145 minutes. followed by equilibration for next run at 5% sol B for 5 min. Ionization of eluting peptides was performed using an EASY-Spray source at an electric potential of 2 kV. All experiments were done in data dependent acquisition mode with top 15 ions isolated at a window of 0.7 or 1.2 m/z and default charge state of +2. Only precursors with charge states ranging from +2 to +7 were considered for MS/MS events. Stepped collision energy was applied to fragment precursors at normalized collision energies of 15, 25, 40. MS precursor mass range was set to 375 to 2000 m/z and 100 to 2000 for MS/MS. Automatic gain control for MS and MS/MS were 1×10^6 and 1×10^5 and injection

time to reach AGC were 50 ms and 200 or 250 ms respectively. Exclude isotopes feature was set to "ON" and 60 s dynamic exclusion was applied. Data acquisition was performed with option of Lock mass (441.1200025 m/z) for all data. Database searching was performed using publicly available software pGlyco Version 2.2.0.[7,8] Glycan database containing 8,092 entries, available with the software was used and Uniprot Human Reviewed protein sequences (20,432 entries) were used as proteins sequence fasta file. Cleavage specificity was set to fully tryptic with 2 missed cleavages and precursor and fragment tolerance were set to 5 and 20 ppm. Cysteine carbamidomethylation was set as fixed modification and oxidation of methionine as variable modification. The results were filtered to retain only entries which had 1% FDR at glycopeptide level. Reporter ion quantification was performed in Proteome Discoverer 2.5 using "reporter ion quantifier" node and Ids were matched with quantitation on a scan-to-scan basis (MS/MS). Glycopeptide PSMs were combined to reflect unique glycopeptides per search and reporter ion intensities were summed up. Spectra were manually verified for glycan oxonium ions and quality. All sialic acid glycopeptides' spectra were verified for presence of sialic acid-specific oxonium ions; 274.09, 292.1 and 657.23. Core fucosylated glycopeptides spectra were checked for at least one peptide+HexNAc+Fuc ion. Proteomics dataset was searched using Sequest in Proteome Discoverer 2.4.

scRNA-seq analysis

Publically available scRNA-seq data was analyzed from three COVID-19 studies using the in-house platform at nference that currently hosts 127 single-cell studies profiling 2.7 million cells.[9] Three COVID-19 studies were identified which profiled the expression of cells in patients with varying symptomatic profiles ranging from mild/moderate to severe. [10-12] The study by Bost et al. profiled 50,615 cells from bronchoalveolar lavage fluid of three mild, six severe COVID-19 patients and 8 healthy controls. [12] The study by Wilk et al. isolated cells from peripheral blood samples derived from only ICU admitted COVID-19 patients. [11] For our downstream analyses, we defined severe COVID-19 patients as those that required ventilator support and the rest as moderate/mild symptomatic cases. This resulted in 4 patients/samples with severe COVID-19 symptoms and 4 patients/samples with moderate symptoms and 6 healthy controls with a total of 44,721 cells. Finally, the study by Chua et al. profiled 160,528 cells from nasopharyngeal and bronchial samples tissue isolated from 19 COVID-19 patients and 5 control patients.[10] The severity of COVID-19 in this study was assessed as per the WHO guidelines. We specifically compared the expression profiles of cells isolated from 8 patients that presented mild/moderate symptoms and 11 patients with severe COVID-19 symptoms. All the differential expression analyses of scRNA-seq datasets were performed using non-parametric Wilcoxon rank sum test implemented in FindMarkers function of SEURAT package (version 4.0.4) in R programming environment (4.1.0).[13] For comparing scRNA-seq data with the plasma proteomics cohort, we computed the direction of gene expression between severity groups in each study cohort.

Statistics

Statistical significance of individual molecules was carried out for the severe and critical patients against the outpatients using Student's t-test with two tail distributions assuming equal variance. Resulting p-values were corrected for multiple testing using the Benjamini-Hochberg method to calculate adjusted p-values. For proteomics and glycoproteomics analysis of pre-COVID-19 and COVID-19 matched patient plasma samples, one-way ANOVA was performed to calculate significance between the three severity subgroups and individual fold-changes were calculated for all patients and a $p \leq 0.05$ was considered significant. Metaboanalyst 5.0 was employed to perform principal component analysis after scaling the data using autoscaling option.[14]

Statistical significance of model performances between the four multi-omics models (All Lipids, All metabolites, Olink protein panel, and All analytes) and the two baseline models (one with only interleukin-6 and the other using the cytokine storm panel of cytokines) shown in Supplementary Figure 1D were computed using Fisher's exact test. Confidence intervals in Supplementary Figure 1D were computed using a two-sided exact binomial test at a confidence level of 0.95.

Results

Principal component analysis of -omics datasets

Principal component analysis was carried out using data from proteomics alone (Supplementary Figure 2A), lipidomics alone (Supplementary Figure 2B) or the integrated multi-omics dataset (Supplementary Figure 2C). Although we observed a separation of the uninfected individuals (control) from hospitalized patients, no such separation could be discerned between the control and outpatient groups from this analysis. Further, among the hospitalized patients, patients who were classified as severe could not be separated from the critically ill group.

Analysis for enriched pathways and cellular processes

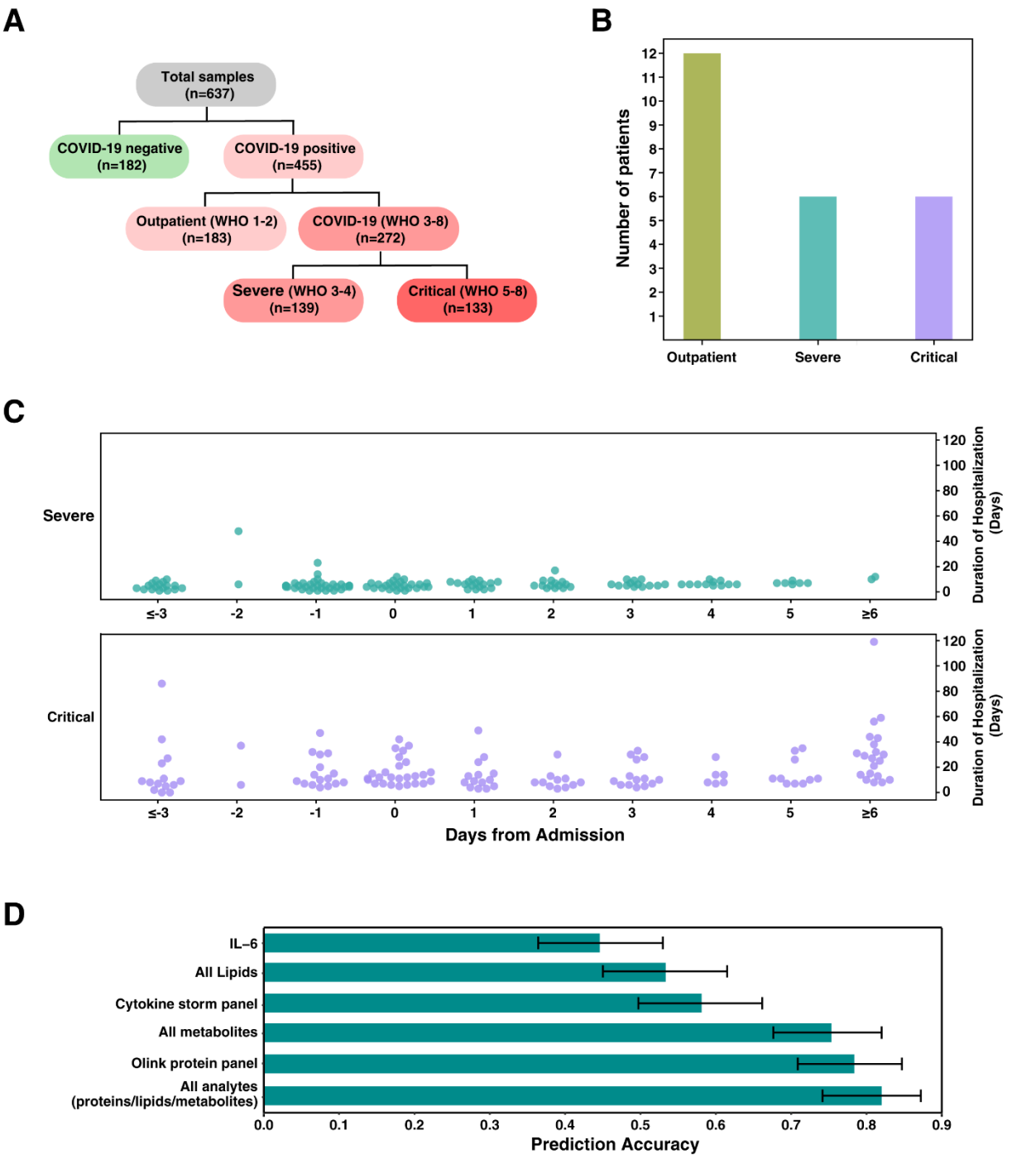
Pathway and network analysis to uncover distinguishing pathways between critical and severe cases showed an enrichment of processes related to protein synthesis and degradation (Supplementary Figure 3). EGFR-mediated signaling pathway molecules were enriched, along with those associated with protein post-translational modifications. Protein degradation via cathepsins and proteases was also elevated, which along with elevated levels of THOP1, TPP1 and metabolites like hydroxyproline, points to increased turnover of extracellular matrix proteins and remodeling. A similar analysis to distinguish severe patients from outpatients enriched pathways involved in cellular organization, function and death (Supplementary Figure 4A). On the other hand, critical patients were separated from outpatients by an enrichment of pathways related to cellular response to therapeutics, signaling and death (Supplementary Figure 3B).

A multi-omic biomarker panel to predict COVID-19 severity

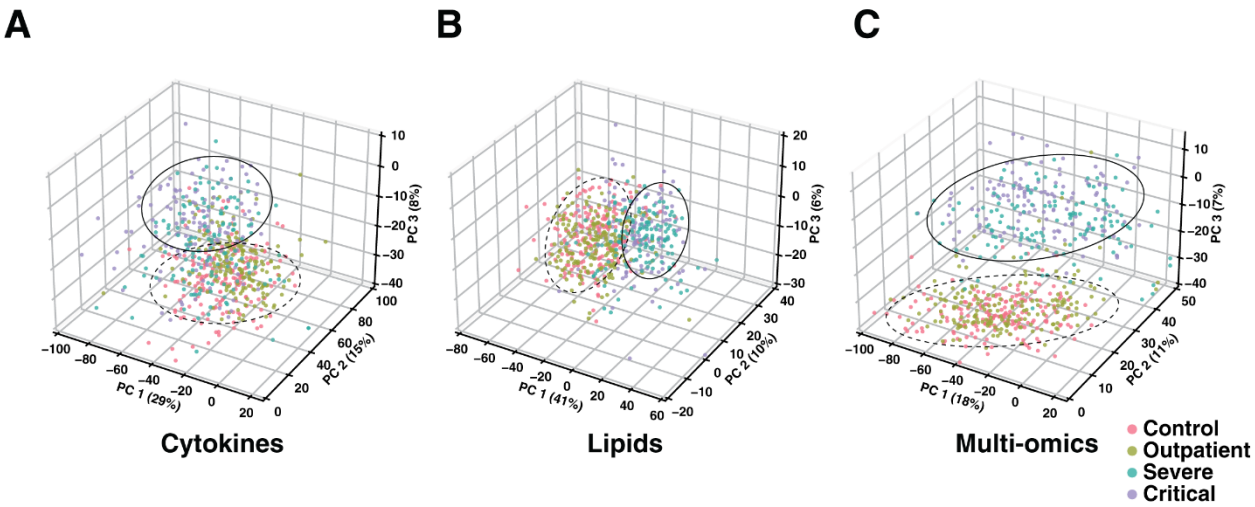
Individualized predictions can be generated using the model developed in this study. This process would require the generation of a dataset based on the cytokine, lipid, and metabolite panels as well as demographic data for the individual. This dataset would then be preprocessed as described in the methods section using the same codebase developed here. Then the predictive model trained on our dataset could be used to generate predictions based on the new data. Results are reported as a probability the new individual is a member of each of the four classes this predictive model was trained on. The class with the highest probability is the predicted class. While this process describes how the model may be used directly for predictions using new data, the model was not intended to be used in a clinical setting without further validation and refinement.

Cellular mRNA expression of predictive cytokine markers

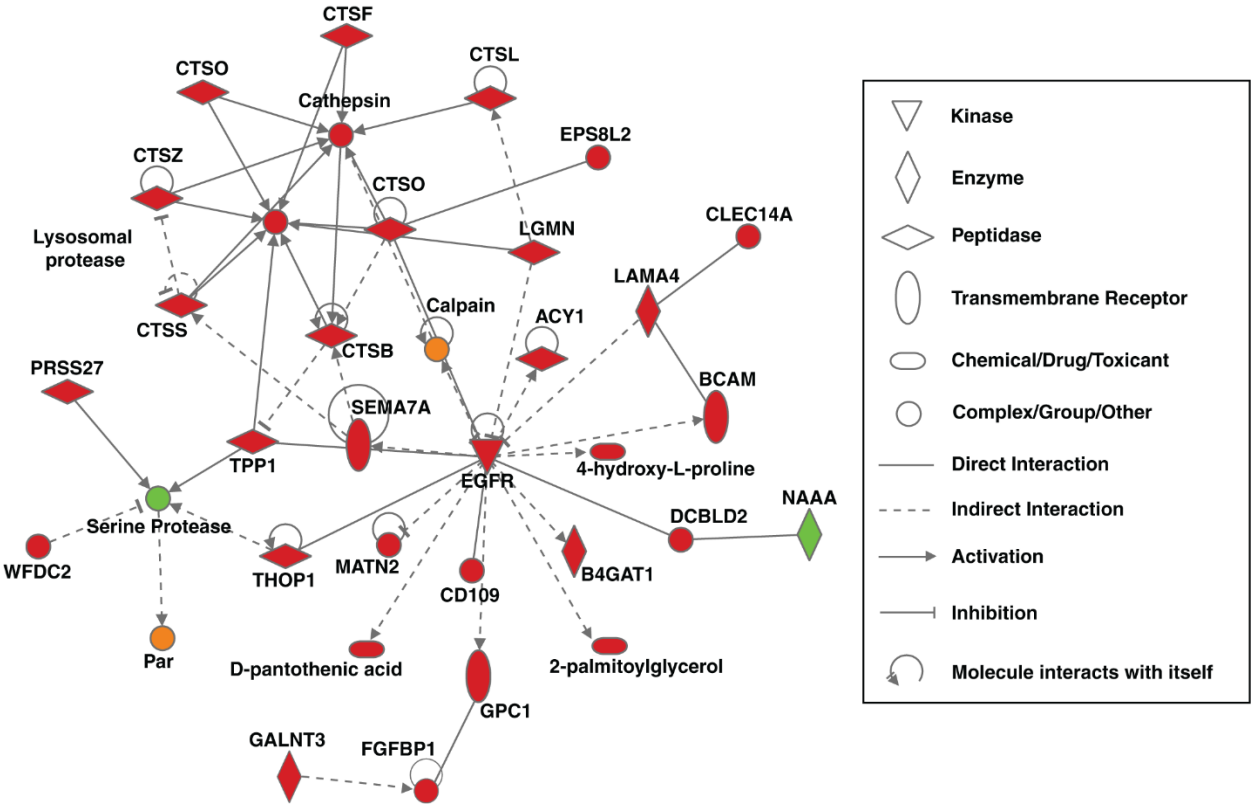
We observed a similar upregulation based on scRNA-seq data in the severe group for 17 proteins that were identified from our cytokine analysis.[10-12] These 17 upregulated genes were *TNF, PFDN2, PSME2, CXCL11, BTN3A2, CAPG, TGFA, FLT1, LTA4H, CASP1, DDX58, OLR1, CCL2, IL1B, LGALS9, MMP7* and *HMOX1* (Figure 3 and Supplementary Figures 4 and 5). This upregulation at the mRNA level was defined as significant increase in expression in at least one cell type and in at least one of these three studies in the critical patient groups as compared to the severe group. We observed some genes with similar signals across multiple sc-RNA-seq studies and cell types. For example, *CCL2* was consistently overexpressed in 8 cell types in two studies.[10,12] C-C motif chemokine 2 (CCL2) is a chemotactic factor for monocytes and basophils and a ligand for CCR2 receptor, which suggests increased activity of monocytes and basophils in severely ill patients. *LTA4H* was consistently overexpressed in two single cell studies.[11,12] Leukotriene A-4 hydrolase (LTA4H) is a bifunctional enzyme with catalytic activity for biosynthesis of the proinflammatory mediator leukotriene B4 and its upregulation in the severe group suggests dominance of the leukotriene pathway. Macrophage-capping protein (CAPG) is known to play an important role in macrophage function.[15] Loss of *CAPG* in bone marrow macrophages profoundly inhibits macrophage colony stimulating factor–stimulated ruffling. Correspondingly, *CAPG* was overexpressed in epithelial cells, *CD14, CD16*, monocytes and neutrophils, aligning with the functional requirement for the performance across all three studies. *BTN3A2* was significantly elevated in five cell types (gamma delta T cells, DC, neutrophil, CD4 and CD8 family).[11] Interleukin-1 beta (IL1B) functions to induce T-cells, neutrophils influx and activation and was elevated in two of the studies.[10,12]



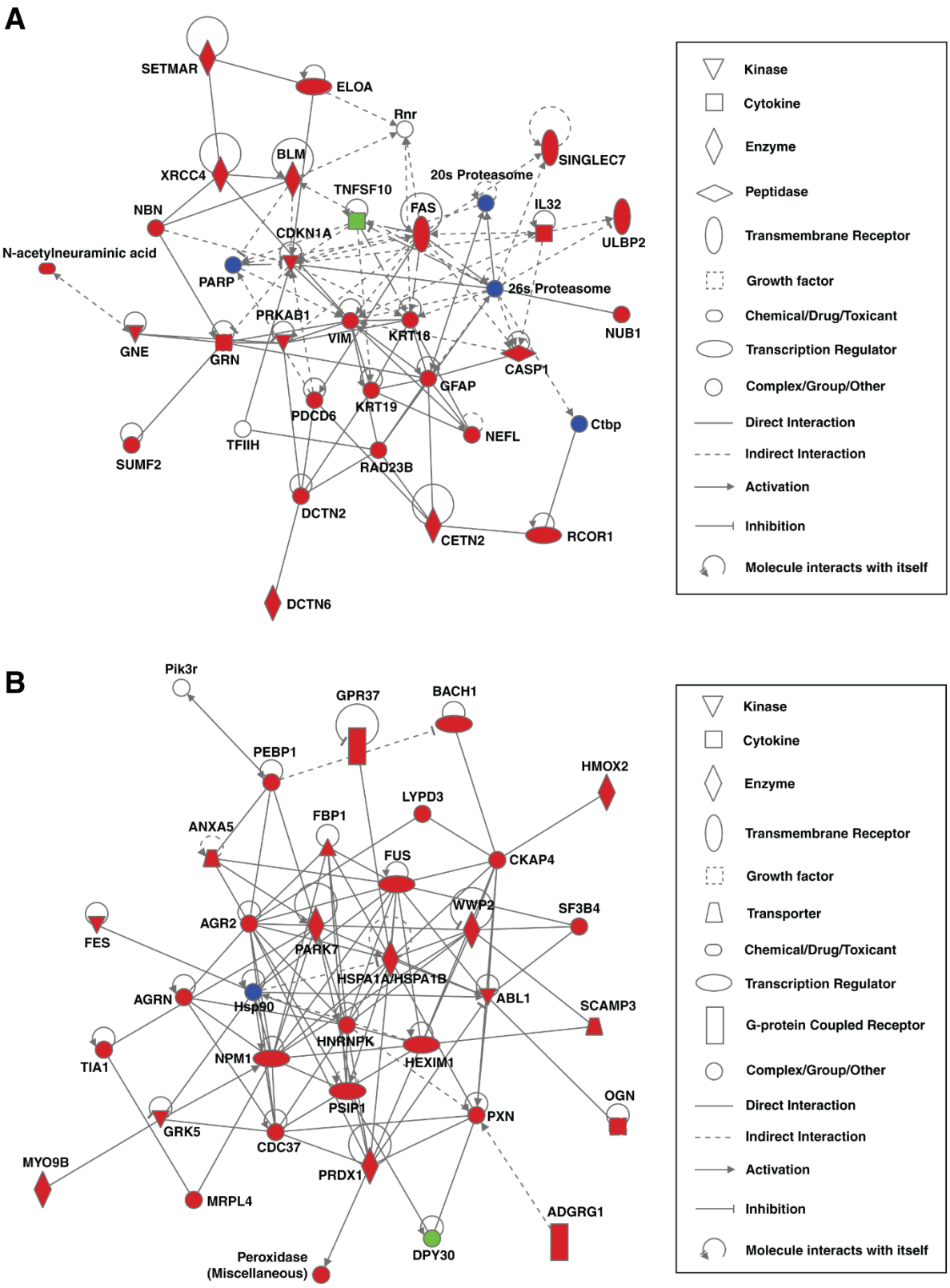
Supplementary Figure 1. Multi-Omic analysis of COVID-19 patients. (A) Distribution of controls and patients across various grades of severity. The patients are binned according to WHO ordinal scale for clinical improvement. (B) Distribution of patients with COVID-19 where and matched pre-COVID-19 samples were available and binned as per the WHO ordinal scales of clinical outcomes. (C) Distribution of sample collection time point and the duration of hospitalization for the admitted COVID-19 patients. (D) Bar plot showing the prediction accuracies in the held-out test set of classification models constructed using various selected molecular feature sets and features across -omics platform (Whiskers in bar plot indicate 95% confidence interval of mean).



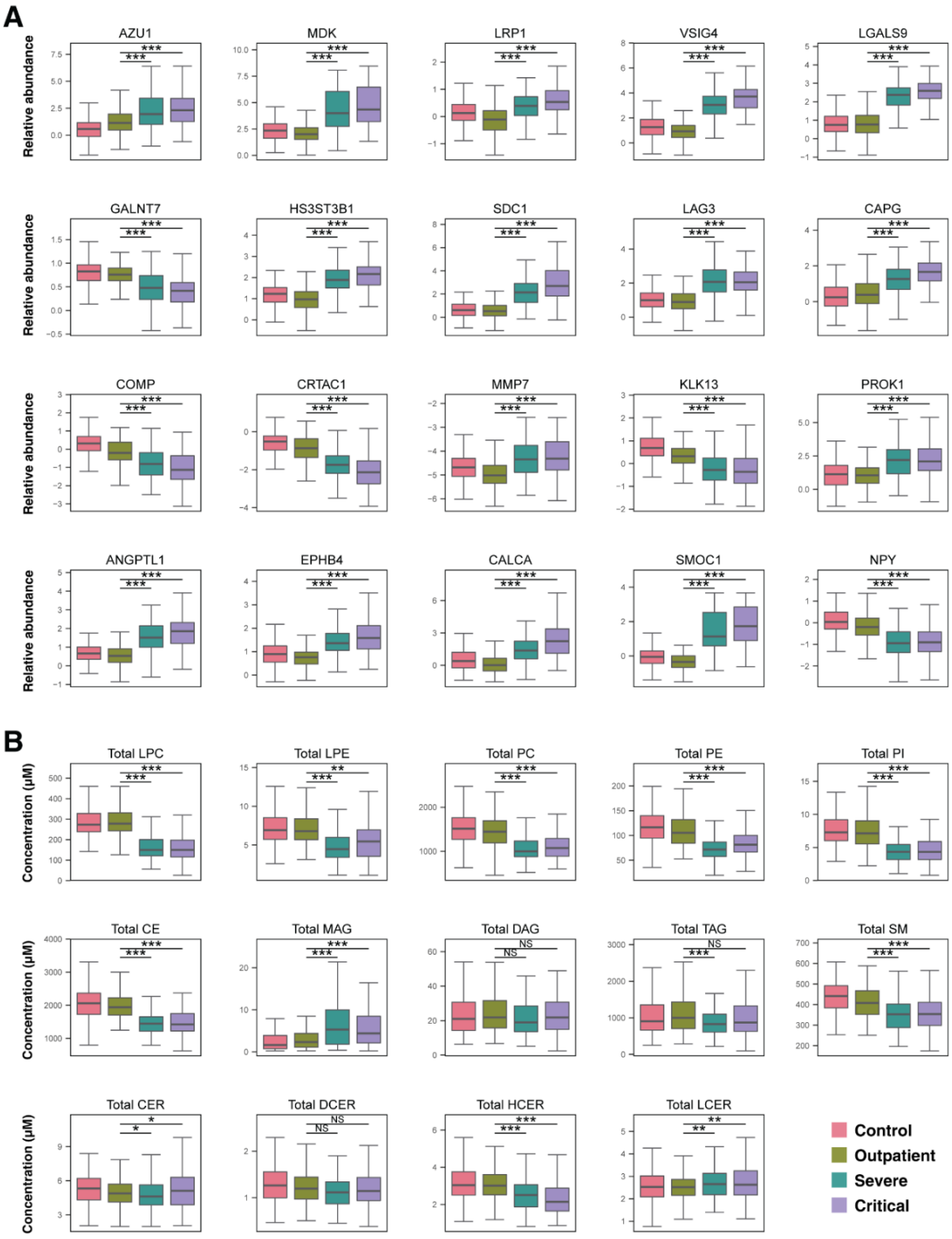
Supplementary Figure 2. Principal Component Analysis shows clustering of severity subgroups across patients and controls using (A) 1,463 unique proteins, (B) 902 lipids, and (C) 3,383 multi-omic analytes (1,463 proteins, 902 lipids and 1,018 metabolites).



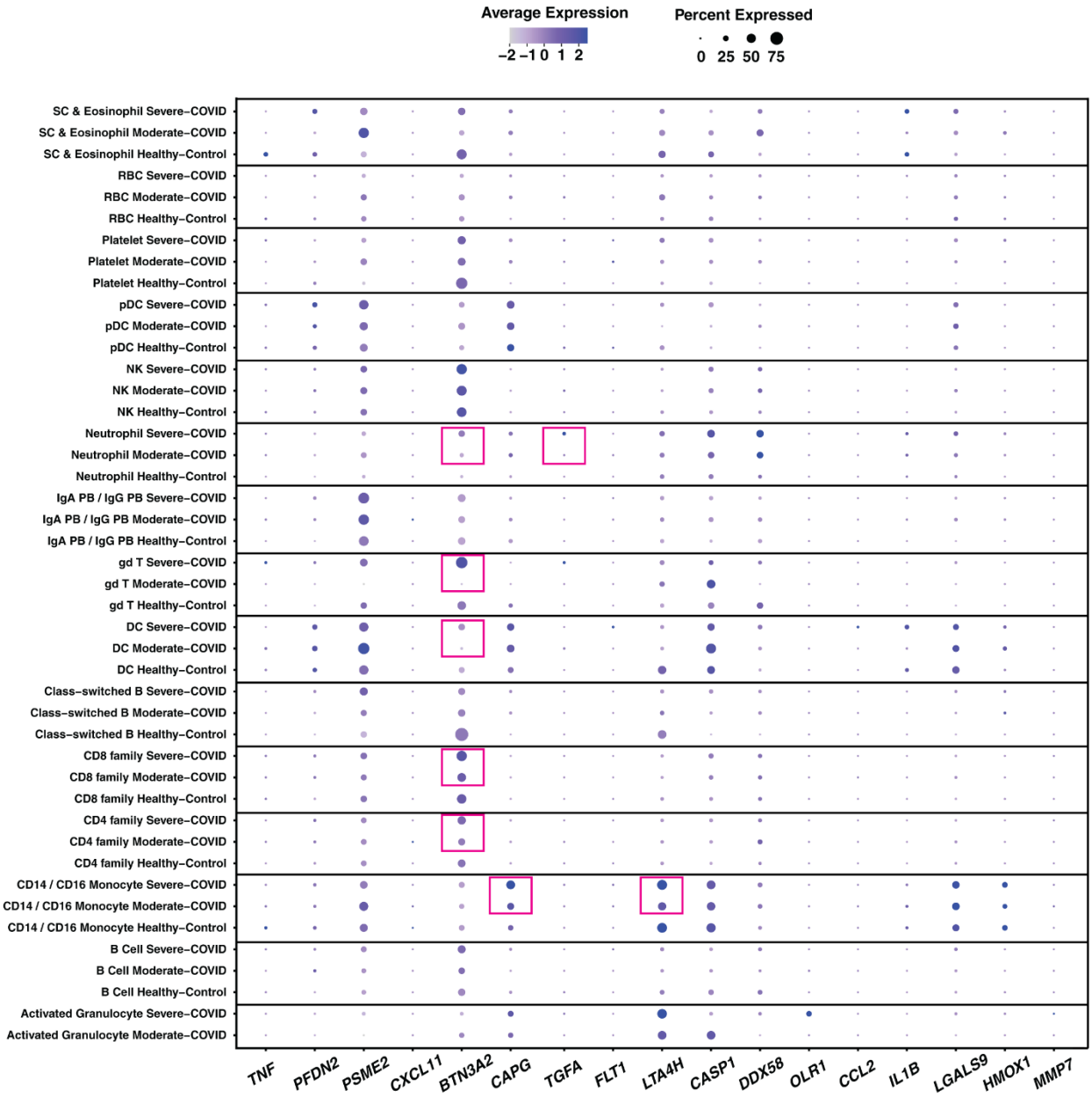
Supplementary Figure 3. Top scoring network analysis of molecules showing changes in critical vs. severe cases of COVID-19. Red and green nodes indicate upregulated and downregulated molecules, respectively. Orange nodes indicate molecules with predicted activation.



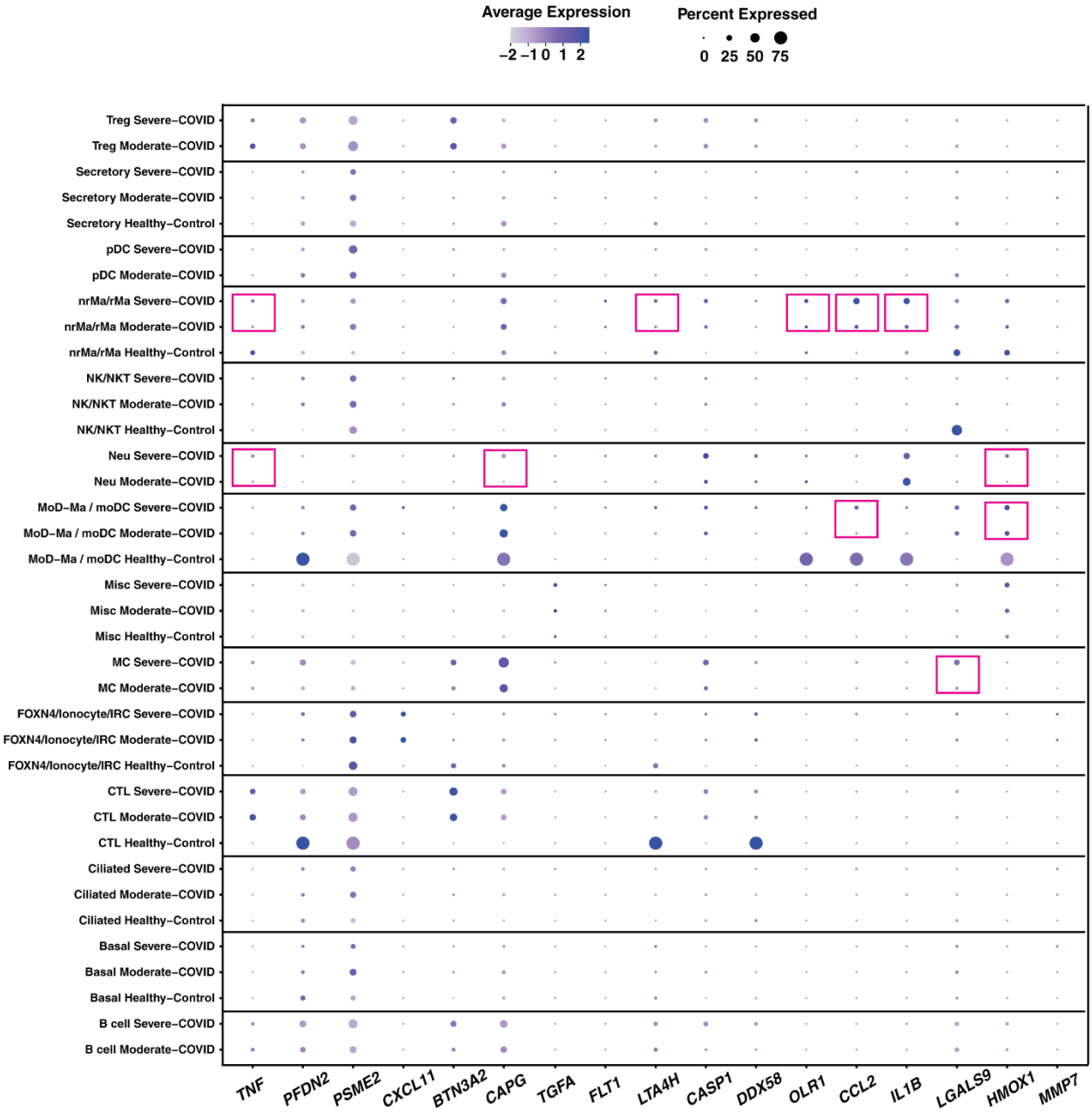
Supplementary Figure 4. Top scoring networks of (A) severe vs. outpatient COVID-19 cases and (B) critical vs. outpatient COVID-19 cases using Ingenuity Pathway Analysis. Red and green nodes indicate upregulated and downregulated molecules, respectively. Blue nodes indicate molecules with predicted inhibition. White node represents molecules not detected in the study.



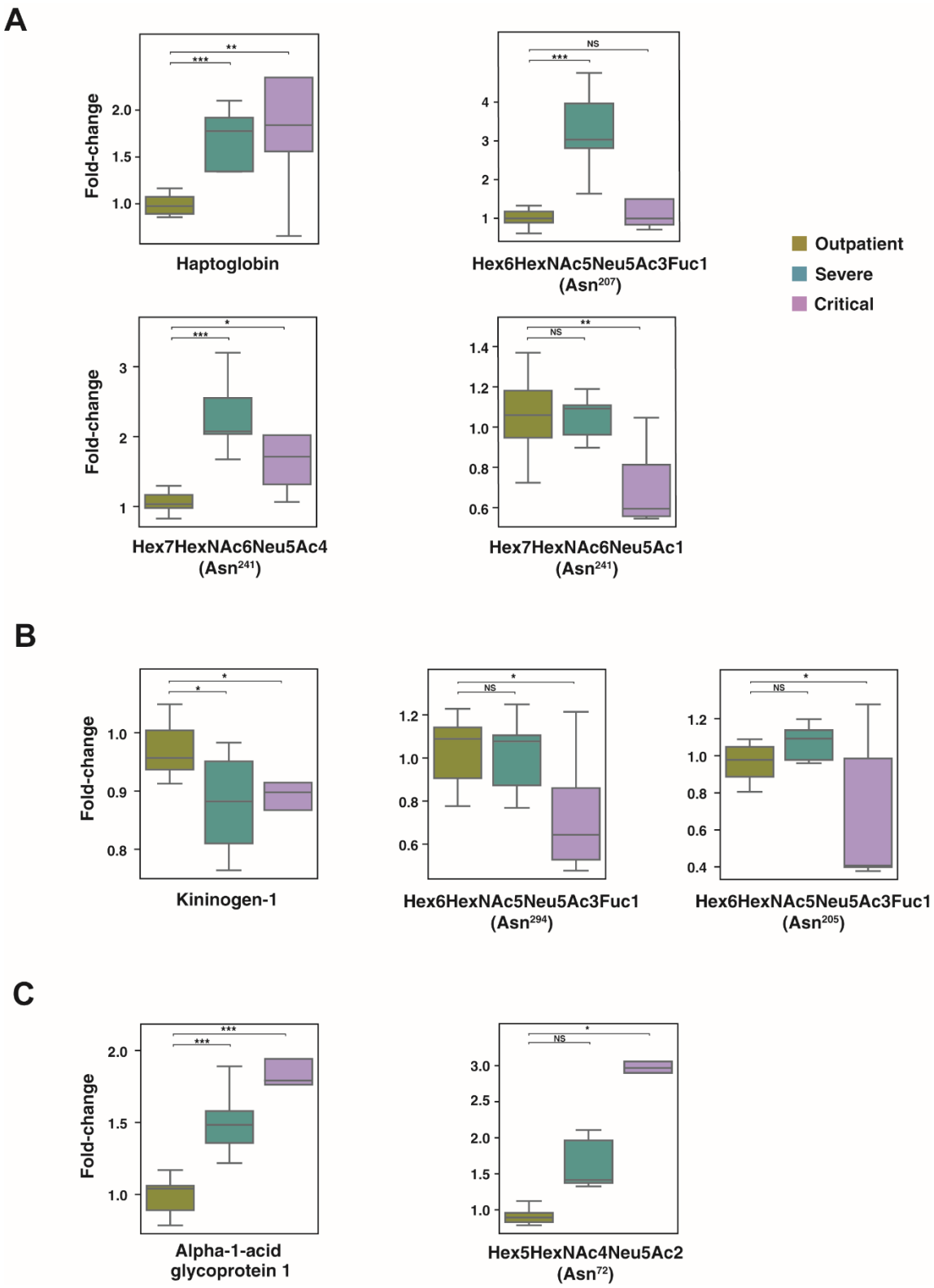
Supplementary Figure 5: Box plots of (A) cytokines and other proteins, and (B) 14 classes of lipids. NS - No significance; *Adjusted $p \leq 0.05$; **Adjusted $p \leq 0.01$; ***Adjusted $p \leq 0.001$. Abbreviations: LPC – lysophosphatidylcholine; LPE – lysophosphatidylethanolamine; PC – phosphatidylcholine; PE – phosphatidylethanolamine; PI – phosphatidylinositol; CE – cholesteryl ester; MAG – monoglyceride; DAG – diglyceride; TAG – triglyceride; SM – sphingomyelin; CER – ceramide; DCER – dihydroceramide; HCER – hexosylceramide; LCER - lactosylceramide



Supplementary Figure 6: Genes upregulated significantly in severe COVID-19 compared to moderate/severe COVID-19 in Wilk's study[11] and the current study. The pink boxes highlighted in the figure denotes those genes which are upregulated in severe as compared to moderate/mild cohort in both scRNA-seq and plasma proteomics analysis. Abbreviations: SC - stem cells; RBC - red blood cells; pDCs - plasmacytoid dendritic cells; NK - Natural Killer; gD T cells - Gamma delta T cells; DCs - conventional dendritic cells; Treg - regulatory T cell; NK/NKT - Natural killer/Natural killer T cells; Neu - Neutrophils; MoDC - Monocyte-derived dendritic cells; CTL - cytotoxic T lymphocytes.



Supplementary Figure 7: Genes upregulated significantly in severe COVID-19 compared to moderate/severe COVID-19 in Chua's study[10] and the current study. The pink boxes highlighted in the figure denote genes which are upregulated in severe as compared to moderate/mild cohort in both scRNA-seq and plasma proteomics analysis. Abbreviations: SC - stem cells; RBC - red blood cells; pDCs - plasmacytoid dendritic cells; NK - Natural Killer; gD T cells - Gamma delta T cells; DCs - conventional dendritic cells; Treg - regulatory T cell; NK/NKT - Natural killer/Natural killer T cells; Neu - Neutrophils; MoDC - Monocyte-derived dendritic cells; CTL - cytotoxic T lymphocytes.



Supplementary Figure 8: Box plots showing variation at protein level and their N-linked glycosylation sites across different patient groups for proteins: Haptoglobin (A), Kininogen-1 (B) and Alpha-1-acid glycoprotein (C). NS - No significance; **Adjusted $p \leq 0.05$; **Adjusted $p \leq 0.01$; ***Adjusted $p \leq 0.001$. Abbreviations: Hex – Hexose; HexNAc – N-acetylhexosamine; Neu5Ac – N-acetylneuraminic acid; Fuc – Fucose.

Supplementary Table 1: Complete list of 102 predictors of COVID-19 severity discovered in the study and their corresponding importance score and p-value

Molecules	Importance score	p-value	Omics	Molecules	Importance score	p-value	Omics
linoleamide (18:2n6)	0.079	<0.0001	M	eicosapentaenoate (EPA; 20:5n3)	0.004	0.0026	M
LTA4H	0.05	<0.0001	P	LILRA5	0.004	0.026	P
oleamide	0.044	<0.0001	M	PPP3R1	0.004	0.0026	P
KRT19	0.036	<0.0001	P	androstenediol (3alpha, 17alpha) monosulfate	0.004	0.0026	M
CCL7	0.027	<0.0001	P	arachidonate (20:4n6)	0.004	0.0026	M
heme	0.023	<0.0001	M	delta-CEHC	0.004	0.0026	M
VSIG4	0.021	<0.0001	P	cortolone glucuronide	0.004	0.0026	M
5-methyluridine (ribothymidine)	0.017	0.0001	M	carnitine of C10H14O2	0.004	0.0026	M
IL15	0.016	<0.0001	P	CLEC6A	0.004	0.0026	P
LGALS4	0.013	<0.0001	P	theobromine	0.004	0.0026	M
lactate	0.012	0.0014	M	cysteine	0.004	0.0026	M
PTGDS	0.0093	<0.0001	P	2-hydroxydecanoate	0.0033	0.0075	M
CXCL9	0.0087	<0.0001	P	LPE(18:1)	0.0033	0.0075	L
ITGAV	0.008	0.0005	P	TAG55:7-FA20:3	0.0033	0.0075	L
leucylglycine	0.008	0.0001	M	ITGB1	0.0033	0.0075	P
OBP2B	0.008	0.0001	P	PE(O-16:0/22:6)	0.0033	0.0075	L
LRP1	0.0073	<0.0001	P	DUOX2	0.0033	0.0075	P
2,6-dihydroxybenzoic acid	0.0073	<0.0001	M	RTBDN	0.0033	0.0075	P
PC(16:0/20:1)	0.0067	0.0005	L	IDUA	0.0033	0.0075	P
eicosanedioate (C20-DC)	0.006	0.021	M	N-methylpipecolate	0.0033	0.0075	M
docosadienoate (22:2n6)	0.006	0.0019	M	4-hydroxyphenylpyruvate	0.0033	0.0075	M
bilirubin (E,Z or Z,E)	0.006	0.0019	M	HBQ1	0.0033	0.0075	P
KRT18	0.006	0.015	P	PI(18:0/20:3)	0.0033	0.048	L
gamma-glutamylisoleucine	0.006	<0.0001	M	phenylacetylcarnitine	0.0033	0.0075	M
DPEP1	0.006	<0.0001	P	AMY2B	0.0033	0.0075	P
IGFBP6	0.006	<0.0001	P	CDH6	0.0033	0.0075	P
4-aminophenol sulfate (2)	0.006	<0.0001	M	ursodeoxycholate	0.0033	0.026	M
PE(P-18:1/18:1)	0.0053	0.0054	L	PLA2G15	0.0027	0.018	P
CTRC	0.0053	0.0001	P	GOLM2	0.0027	0.018	P
AOC3	0.0053	0.0001	P	pentadecanoate (15:0)	0.0027	0.018	M
gamma-glutamylthreonine	0.0053	0.0001	M	RARRES2	0.0027	0.018	P
ITGB6	0.0053	0.0001	P	ANGPT2	0.0027	0.018	P
glucose	0.0053	0.0001	M	CXCL11	0.0027	0.018	P
KLK10	0.0053	0.0001	P	AMY2A	0.0027	0.018	P
PE(O-16:0/18:1)	0.0047	0.012	L	CA1	0.0027	0.018	P
acetoacetate	0.0047	0.0048	M	NRP1	0.002	0.041	P
CTSB	0.0047	0.0007	P	IRAG2	0.002	0.041	P
DAG(18:1/18:2)	0.0047	0.0007	L	NME3	0.002	0.041	P
ENPP5	0.0047	0.0007	P	arabinose	0.002	0.041	M
CEACAM5	0.0047	0.0007	P	DAG(16:1/20:4)	0.002	0.041	L
ICOSLG	0.0047	0.0007	P	3-methyl catechol sulfate	0.002	0.041	M
4-hydroxychlorothalonil	0.0047	0.0007	M	TBC1D5	0.002	0.041	P
CEP85	0.0047	0.0007	P	TAG56:6-FA18:3	0.002	0.041	L
SERPINB5	0.0047	0.0007	P	DCER(18:1)	0.002	0.041	L
CD1C	0.0047	0.0007	P	sebacate (C10-DC)	0.002	0.041	M
heneicosapentaenoate (21:5n3)	0.0047	0.0007	M	CRACR2A	0.002	0.041	P
CGA	0.0047	0.0007	P	HAGH	0.002	0.041	P
NTF4	0.0047	0.0007	P	cys-gly, oxidized	0.002	0.041	M
RAB6A	0.0047	0.0007	P	trizma acetate	0.002	0.041	M
Ep-CAM	0.004	0.0026	P	PE(16:0/20:4)	0.002	0.041	L
DPP7	0.004	0.0026	P	cysteinylglycine disulfide	0.002	0.041	M

*Abbreviations: P - protein; M - metabolite; L – lipid

Supplementary Table 2. Measure of precisions of the prediction accuracies in the held-out test set of classification models

Features	Groups	True positive	False negative	False positive	Precision	Precision - Lower 95% CI	Precision - Upper 95% CI	Recall	Recall - Lower 95% CI	Recall - Upper 95% CI
All cytokines and circulatory proteins	Control	52	3	13	0.80	0.68	0.89	0.95	0.85	0.99
	Outpatient	19	14	3	0.86	0.65	0.97	0.58	0.39	0.75
	Severe	24	0	16	0.60	0.43	0.75	1.00	0.86	1.00
	Critical	21	15	0	1.00	0.84	1.00	0.58	0.41	0.74
All lipids	Control	35	20	18	0.66	0.52	0.78	0.64	0.50	0.76
	Outpatient	14	20	21	0.40	0.24	0.58	0.41	0.25	0.59
	Severe	17	8	25	0.40	0.26	0.57	0.68	0.46	0.85
	Critical	14	22	6	0.70	0.46	0.88	0.39	0.23	0.57
All metabolites	Control	54	1	5	0.92	0.81	0.97	0.98	0.90	1.00
	Outpatient	29	5	2	0.94	0.79	0.99	0.85	0.69	0.95
	Severe	16	9	23	0.41	0.26	0.58	0.64	0.43	0.82
	Critical	14	22	7	0.67	0.43	0.85	0.39	0.23	0.57
All omics	Control	53	2	2	0.96	0.87	1.00	0.96	0.87	1.00
	Outpatient	30	4	2	0.94	0.79	0.99	0.88	0.73	0.97
	Severe	19	6	17	0.53	0.35	0.70	0.76	0.55	0.91
	Critical	21	15	6	0.78	0.58	0.91	0.58	0.41	0.74
Interleukin-6	Control	24	31	22	0.52	0.37	0.67	0.44	0.30	0.58
	Outpatient	22	11	35	0.39	0.26	0.52	0.67	0.48	0.82
	Severe	13	11	23	0.36	0.21	0.54	0.54	0.33	0.74
	Critical	7	29	2	0.78	0.40	0.97	0.19	0.08	0.36
Discussed cytokines	Control	45	10	15	0.75	0.62	0.85	0.82	0.69	0.91
	Outpatient	18	15	10	0.64	0.44	0.81	0.55	0.36	0.72
	Severe	21	3	24	0.47	0.32	0.62	0.88	0.68	0.97
	Critical	13	23	2	0.87	0.60	0.98	0.36	0.21	0.54
Cytokine storm panel	Control	41	14	25	0.62	0.49	0.74	0.75	0.61	0.85
	Outpatient	18	15	17	0.51	0.34	0.69	0.55	0.36	0.72
	Severe	11	13	14	0.44	0.24	0.65	0.46	0.26	0.67
	Critical	16	20	6	0.73	0.50	0.89	0.44	0.28	0.62

Supplementary Table 3: Top scoring networks and molecules involved in SARS-CoV-2.

Comparison	Molecules in Network	Score	Focus Molecules	Top Diseases and Functions
Patients with critical COVID-19 vs. severe COVID-19	2-palmitoylglycerol,ACY1,B4GAT1,BCAM,calpain,Cathepsin,CD109,CLEC14A,CTSB, CTSD,CTSF,CTSL,CTSO,CTSS,CTSZ,D-pantothenic acid,DCBLD2,EGFR,EPS8L2, FGFBP1,GALNT3,GPC1,LAMA4,LGMN,Lysosomal Protease,MATN2,NAAA,Par, PRSS27,SEMA7A,Serine Protease,THOP1,TPP1,trans-4-hydroxy-L-proline,WFDC2	51	30	[Post-Translational Modification, Protein Degradation, Protein Synthesis]
	ADAM22,ADAM23,ADAMTS13,ADRB,AREG,CG,CLEC11A,Ecm,EGFR ligand,EGLN, FABP4,Fcer1,FSH,gondoic acid,Gpcr,GUSB,HAVCR1,IGFBPL1,IL1R2,IL6R,LY75, MEP1B,Metalloprotease,MMP12,MMP9,MYOC,PAPPA,PRSS8,PTPase,SCARF2,SDC1,SP INT2,STC1,Vegf,WARS1	36	24	[Cellular Movement, Hematological System Development and Function, Immune Cell Trafficking]
	17-hydroxydocosahexaenoic acid,bilirubin,CALCA,CCL15,CCL16,CCL21,CCL23, chemokine,CLEC10A,Collagen type V,ENPP2,ENTPD5,HAVCR2,IDO,Ifn,IL17C, IL17D,IL2,Il8r,immune complex,LIFR,LTBR,MIR124,NCR1,NFkB (family),OSCAR, Pro-inflammatory Cytokine,SDC4,SIGLEC1,SPINT1,Tlr,Tnf receptor,TNFRSF9, Villin,VSIG4	34	23	[Cellular Movement, Hematological System Development and Function, Immune Cell Trafficking]
	Alpha catenin,ANXA10,AXL,BAG3,CCL19,CCL20,CCL3,CDH2,CDH5,CST5, CXCL10,CXCL16,CXCL9,Erm,Fc gamma receptor,FCRL2,G protein beta gamma,Hsp90,Ifn gamma,IgG2b,IL13,KRT19,LILRB4,Mek,MFGE8,NEFL,NOMO1 (includes others),OSM, Ppp2c,SRC (family),SYK/ZAP,TAM receptor kinase,TNFRSF12A,TNFSF13B,TYRO3	34	23	[Cell-mediated Immune Response, Cellular Movement, Hematological System Development and Function]
	3 beta HSD,Abl1/2,ACAN,BCAN,CNTN1,CNTN2,CSPG,Cytokeratin,ESAM,FETUB, FSTL3,FXR ligand-FXR-Retinoic acid-RXRα,gelatinase,HNMT,IGFBP2,IGFBP7, Kallikrein,KIR,L1CAM,LTBP2,Mir200,MSMB,NCAN,NPTXR,plasminogen activator,PLAT,SMOC2,SMOOTH MUSCLE ACTIN,sodium channel,SPOCK1, Tenascin,TGFB1,TNR,WFIKKN1,WFIKKN2	32	22	[Carbohydrate Metabolism, Cell-To-Cell Signaling and Interaction, Cellular Assembly and Organization]
Patients with critical COVID-19 vs. outpatients of COVID-19	ABL1,ADGRG1,AGR2,AGRN,ANXA5,BACH1,CDC37,CKAP4,DPY30,FBP1,FES,FUS,G PR37,GRK5,HEXIM1,HMOX2,HNRNPK,Hsp90,HSPA1A/HSPA1B,LYPD3,MRPL46,MY O9B,NPM1,OGN,PARK7,PEBP1,peroxidase (miscellaneous),Pik3r,PRDX1,PSIP1,PXN, SCAMP3,SF3B4,TIA1,WWP2	44	32	[Cell Death and Survival, Cell-To-Cell Signaling and Interaction, Cellular Response to Therapeutics]
	AMFR,APEX1,BAG3,BIRC2,BLM,CETN2,DCTN2,DCTN6,DCTPP1,DPP10,DTX3,E3 cofactor-E3 ring-target,E3 RING,E3 ring-target,ELOA,FABP5,FAS,FEN1,GPKOW, KRT18, LYAR,MAD1L1,MNDA,PDCD6,PMVK,PRKAB1,RAD23B,RNASET2,Rnr,SETMAR,SRP 14,SSB,SUGT1,TPT1,XRCC4	42	31	[Cell Death and Survival, Cellular Compromise, DNA Replication, Recombination, and Repair]
	ADM,ALT,CASP8,Caspase 3/7,CD109,CD34,CD46,CD69,CD70,CLEC7A,CLSPN,FADD, FASLG,GOT,GZMA,GZMB,GZMH,IKKA/B,IL12A,IL2RA,KDR,LGALS1,LGALS3,LTA, MERTK,MMP13,PARP1,PLXNA4,SELE,SELP,SPINK1,TNFRSF1A,TNFRSF1B,trypsin,Z BTB16	40	30	[Cell Death and Survival, Connective Tissue Disorders, Inflammatory Response]
	aldo,apyrase,ATP5PO,ATP6AP2,ATP6V1D,ATP6V1F,CANT1,CCL2,CLEC14A,CLTA,DA PP1,Enolase,Fcer1,glutamine,GOPC,GUSB,H -transporting two-sector ATPase,IGF2R, LAMP3,leukotrieneB4,lipoxygenase,LTA4H,MITD1,NDRG1,NPDC1,PDP1,PRTFDC1,RA B6A,SIGLEC5,SLAMF7,SNX9,STAMBP,TBC1D23,TJAP1,Vacuolar H ATPase	35	28	[Cellular Compromise, Cellular Movement, Inflammatory Response]

Patients with critical COVID-19 vs. outpatients of COVID-19	2-palmitoylglycerol,ACY1,adenine-riboflavin dinucleotide,Ap2,Ap2 alpha,B4GALT1, B4GAT1,Cg Beta,CNPY4,D-pantothenic acid,EGFR,Endophilin,F3-F7,gamma-glutamylalanine,HARS1,homocitrulline,Keratin II, 6,LRIG1,LRP11,MICOS10-NBL1/NBL1,MUC13,MUC16,Mucin,N-acetylneuraminic acid,PHOSPHO1,PON2,PROK1, S-glutathionyl-L-cysteine,SHMT1,SPINK5,TFF3,trans-4-hydroxy-L-proline, transglutaminase,TSPAN1,uracil	33	27	[Immunological Disease, Inflammatory Disease, Inflammatory Response]
	14-3-3,CD177,CG,CLMP,Complement,CXCL1,CXCL16,CXCL3,CXCL5, DEFB4A/DEFB4B,ESM1,GNRH,IL1,IL12B,IL18,IL1RN,IL32,IL33,ITGB2,JINK1/2,LCN2, LGALS9,MAP2K1/2,OSM,PRKAR1A,PRTN3,SERPINB1,SORCS2,THBD,Tnf (family), TNFRSF11B,TNFSF13,TNFSF14,TYMP,VCAN	33	27	[Cellular Movement, Hematological System Development and Function, Immune Cell Trafficking]
	ADA,Adaptor protein 1,AGER,AIF1,beta-lactose,CCL11,CCL17,CCL18,CCL19,CCL3, CCL4,CCL7,Cd1,CD274,CXCL10,CXCL11,CXCL6,CXCL9,FIBRINOGEN (family), HMOX1,Ifn gamma,IFNGR1,IL12 (family),IL4R,IL7,LDL,LILRB4,LPL,MAPK9, MMP12,MMP9,p85 (pik3r),TH1 Cytokine,TNFRSF12A,XCL1	33	27	[Cellular Movement, Hematological System Development and Function, Immune Cell Trafficking]
	5-hydroxytryptamine,ACP5,AMPK,ANGPT2,BCL2L11,CD302,DRG2,EGLN1,ENO2, FGF23,FKBP5,FLT3,FOXO1,FOXO3,HDL,INPPL1,Lh,MAP1LC3,MAPT,MAX,MILR1, MLN,MPO,NAMPT,NGF,p70 S6k,PEPCK,PI3K (family),PLAUR,PPP3R1,PRDX5,SOD2, TSH,VCAM1,VEGFA	33	27	[Connective Tissue Disorders, Free Radical Scavenging, Inflammatory Disease]
Patients with severe COVID-19 vs. outpatients of COVID-19	20s proteasome,26s Proteasome,BLM,CASP1,CDKN1A,CETN2,Ctbp,DCTN2,DCTN6, ELOA,FAS,GFAP,GNE,GRN,IL32,KRT18,KRT19,N-acetylneuraminic acid,NBN,NEFL, NUB1,PARP,PDCD6,PRKAB1,RAD23B,RCOR1,Rnr,SETMAR,SIGLEC7,SUMF2,TFIIH, TNFSF10,ULBP2,VIM,XRCC4	42	29	[Cell Death and Survival, Cellular Assembly and Organization, Cellular Function and Maintenance]
	AGR2,AXL,CDC37,CGA,CKAP4,CRELD2,DPY30,DRG2,EGLN,ERP44,EZR,FKBP5,FUS ,GNRH,GOPC,GRK5,GTPase,HARS1,Hsp90,HSPA1A/HSPA1B,KIFBP,Lh,MYO9B,NOM O1 (includes others),NUDC,Pkc(s),PLXNB3,PREB,PSIP1,SEPTIN9,SESTD1,SPOCK1, SUGT1,THOP1,TNFRSF10B	42	29	[Digestive System Development and Function, Gastrointestinal Disease, Organ Morphology]
	Adaptor protein 1,AIF1,CCL11,CCL17,CCL20,CCL3,CCL7,CD74,CEBPB,CHIT1,CXCL10, CXCL16,CXCL3,CXCL5,CXCL9,FGF23,FOSB,FOXO1,FOXO3,Gsk3,Ifn gamma,IL15, IL1RL1,IL1RN,IL7,IRAK4,LILRB4,MAP1LC3,MAPK9,METAP1D,Mhc2 Alpha,PI3K (family),PRTN3,SERPINB1,XCL1	42	29	[Cellular Movement, Hematological System Development and Function, Inflammatory Response]
	AIFM1,BAX,BID,CD274,CD38,CD69,CLMP,CTSD,DGKZ,FADD,FASLG,GBP4,GC-GCR dimer,IFN alpha/beta,IFNGR1,Ikb,IKK (complex),IL12RB1,IL13RA1,IL18R1,IL4R, IL5RA,IRAK1,LCN2,MOG,PECAM1,SPP1,STAT5a/b,Tgf beta,TNFRSF11A,TNFRSF11B, TNFRSF1A,TNFRSF1B,Ubiquitin,ZBTB16	40	28	[Cell Death and Survival, Cellular Growth and Proliferation, Lymphoid Tissue Structure and Development]
	AKR1B1,Aldose Reductase,BMP6,CC2D1A,CLTA,CORO1A,Creb,CST5,DECR1,Ferritin, FLI1,FMNL1,IL11,LAP3,LYAR,MATN2,MED18,mediator,Mir122a,b,MRPL46,MVK,Oste ocalcin,PAK4,PLIN3,PRKRA,RAB6A,RAS,ROCK,Rsk,Smad,TACC3,TARBP2,THPO,TIA 1,TJAP1	33	25	[Cellular Function and Maintenance, Cellular Growth and Proliferation, Hematological System Development and Function]

Supplementary Table 4: Details of patients with matched pre-COVID-19 samples studied by mass spectrometry-based total proteomics and glycoproteomics

Patient ID	WHO ordinal scale	Gender	Age	Ethnicity	Race	Sample collection date: Pre-COVID-19	Sample collection date: COVID-19	TMT Batch
1	1	F	66	Not Hispanic or Latino	White	2013-05-22	2020-09-25	1
2	1	F	62	Not Hispanic or Latino	White	2012-11-21	2020-09-22	1
3	1	M	73	NA	White	2015-08-02	2020-08-13	1
4	5	M	96	Not Hispanic or Latino	White	2014-07-29	2020-07-08	1
5	5	F	83	Not Hispanic or Latino	White	2014-07-30	2020-07-28	1
6	4	M	90	Not Hispanic or Latino	White	2014-04-14	2020-07-28	1
7	5	M	80	Not Hispanic or Latino	White	2014-03-04	2020-07-07	1
8	6	F	69	Not Hispanic or Latino	White	2013-05-14	2020-05-20	1
9	4	M	73	Not Hispanic or Latino	White	2015-08-19	2020-09-29	2
10	4	F	82	Hispanic or Latino	White	2015-03-05	2020-08-18	2
11	4	M	76	Not Hispanic or Latino	White	2014-05-27	2020-06-29	2
12	1	M	59	Not Hispanic or Latino	White	2013-05-06	2020-09-28	2
13	1	F	50	Not Hispanic or Latino	White	2015-05-04	2020-10-05	2
14	1	M	67	Not Hispanic or Latino	White	2014-10-27	2020-10-20	2
15	1	F	61	Not Hispanic or Latino	White	2015-04-05	2020-09-03	2
16	1	F	48	Hispanic or Latino	White	2015-06-22	2020-08-27	2
17	1	F	58	Not Hispanic or Latino	White	2014-01-31	2020-09-04	3
18	1	M	78	Not Hispanic or Latino	White	2014-05-12	2020-08-25	3
19	5	M	77	Not Hispanic or Latino	White	2014-10-06	2020-10-20	3
20	3	F	82	Not Hispanic or Latino	White	2014-02-17	2020-10-28	3
21	1	F	66	Not Hispanic or Latino	White	2015-01-09	2020-10-29	3

References

1. Lofgren L, Stahlman M, Forsberg GB, Saarinen S, Nilsson R, Hansson GI. The BUME method: a novel automated chloroform-free 96-well total lipid extraction method for blood plasma. J Lipid Res 2012; 53(8): 1690-700.
2. Kuhn M. Building Predictive Models in R Using the caret Package. J Stat Softw 2008; 28(5): 1-26.
3. Reel PS, Reel S, Pearson E, Trucco E, Jefferson E. Using machine learning approaches for multi-omics data analysis: A review. Biotechnology Advances 2021; 49: 107739.
4. Williams SA, Kivimaki M, Langenberg C, et al. Plasma protein patterns as comprehensive indicators of health. Nat Med 2019; 25(12): 1851-7.
5. Overmyer KA, Shishkova E, Miller IJ, et al. Large-Scale Multi-omic Analysis of COVID-19 Severity. Cell Syst 2021; 12(1): 23-40 e7.
6. Mun DG, Renuse S, Saraswat M, et al. PASS-DIA: A Data-Independent Acquisition Approach for Discovery Studies. Anal Chem 2020; 92(21): 14466-75.
7. Liu MQ, Zeng WF, Fang P, et al. pGlyco 2.0 enables precision N-glycoproteomics with comprehensive quality control and one-step mass spectrometry for intact glycopeptide identification. Nat Commun 2017; 8(1): 438.
8. Zeng WF, Liu MQ, Zhang Y, et al. pGlyco: a pipeline for the identification of intact N-glycopeptides by using HCD- and CID-MS/MS and MS3. Sci Rep 2016; 6: 25102.
9. Venkatakrishnan AJ, Puranik A, Anand A, et al. Knowledge synthesis of 100 million biomedical documents augments the deep expression profiling of coronavirus receptors. Elife 2020; 9.
10. Chua RL, Lukassen S, Trump S, et al. COVID-19 severity correlates with airway epithelium-immune cell interactions identified by single-cell analysis. Nat Biotechnol 2020; 38(8): 970-9.
11. Wilk AJ, Rustagi A, Zhao NQ, et al. A single-cell atlas of the peripheral immune response in patients with severe COVID-19. Nat Med 2020; 26(7): 1070-6.
12. Bost P, Giladi A, Liu Y, et al. Host-Viral Infection Maps Reveal Signatures of Severe COVID-19 Patients. Cell 2020; 181(7): 1475-88 e12.
13. Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat Biotechnol 2018; 36(5): 411-20.
14. Pang Z, Chong J, Zhou G, et al. MetaboAnalyst 5.0: narrowing the gap between raw spectra and functional insights. Nucleic Acids Res 2021; 49(W1): W388-W96.
15. Witke W, Li W, Kwiatkowski DJ, Southwick FS. Comparisons of CapG and gelsolin-null macrophages: demonstration of a unique role for CapG in receptor-mediated ruffling, phagocytosis, and vesicle rocketing. J Cell Biol 2001; 154(4): 775-84.