Supplementary Materials for

Plasma metabolome and cytokine profile reveal glycylylproline modulating antibody-fading in convalescent COVID-19 patients

Zhu Yang\textsuperscript{a,b,c,1}, Di Wu\textsuperscript{d,e,1}, Shanxin Lu\textsuperscript{a,b,1}, Yang Qiu\textsuperscript{d,e,1}, Zhengyi Hua\textsuperscript{a,b}, Fancheng Tan\textsuperscript{a,b}, Cixiong Zhang\textsuperscript{a}, Lei Zhang\textsuperscript{a}, Ding-Yu Zhang\textsuperscript{e,f}, Xi Zhou\textsuperscript{d,e,f,2}, Zongwei Cai\textsuperscript{c,2}, You Shang\textsuperscript{e,f,g,2}, Shu-Hai Lin\textsuperscript{a,b,2}

\textsuperscript{a} State Key Laboratory for Cellular Stress Biology, School of Life Sciences, Faculty of Medicine and Life Sciences, Xiamen University, Fujian 361102, China; 
\textsuperscript{b} National Institute for Data Science in Health and Medicine, Xiamen University, Fujian 361102, China; 
\textsuperscript{c} State Key Laboratory of Environmental and Biological Analysis, Department of Chemistry, Hong Kong Baptist University, Hong Kong SAR, China; 
\textsuperscript{d} State Key Laboratory of Virology, Wuhan Institute of Virology, Center for Biosafety Mega-Science, CAS, Wuhan 430071, China; 
\textsuperscript{e} Joint Laboratory of Infectious Diseases and Health, Wuhan Institute of Virology \& Wuhan Jinyintan Hospital, Wuhan Jinyintan Hospital, Wuhan 430023, China; 
\textsuperscript{f} Center for Translational Medicine, Jinyintan Hospital, Wuhan 430023, China; 
\textsuperscript{g} Department of Critical Care Medicine, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430030, China; 

\textsuperscript{1} Z.Y., D.W., S.L., and Y.Q. contributed equally to this work.

\textsuperscript{2} Corresponding authors, Email:: shuhai@xmu.edu.cn (S.H.L.); you_shanghust@163.com (Y.S.); zwcai@hkbu.edu.hk (Z.C.); zhouxi@wh.iov.cn (X.Z.).
Supplementary Methods

Ethics and Human Subjects

All work performed in this study was approved by the Wuhan Jinyintan Hospital Ethics Committee and written informed consents were obtained from patients (No. KY-2020-83). Diagnosis of SARS-CoV-2 infection was based on reverse transcription-polymerase chain reaction (RT-PCR) following the criteria from the New Coronavirus Pneumonia Prevention and Control Program (6th edition) published by the National Health Commission of China. Healthy subjects were recruited from healthcare workers and laboratory workers at Wuhan Jinyintan Hospital and Wuhan Institute of Virology, Chinese Academy of Sciences. None of them had previously experienced SARS-CoV-2 infection.

Chemicals and Reagents

The anti-SARS-CoV-2 RBD IgG ELISA kit was purchased from Wuhan Anbo Bio Co., Ltd. (China). The Q300 absolutely quantitative metabolomics kit was purchased from Metabo-Profile Co., Ltd. (Shanghai, China). All of the standards of targeted metabolites were the products of Sigma-Aldrich (St. Louis, MO, USA), Steraloids Inc. (Newport, RI, USA), and TRC Chemicals (Toronto, ON, Canada). Formic acid (Optima grade) was purchased from Sigma-Aldrich (St. Louis, MO, USA). Acetonitrile (Optima LC-MS), isopropanol (Optima LC-MS), and methanol (Optima LC-MS) were obtained from Thermo-Fisher Scientific (FairLawn, NJ, USA). Ultrapure water was produced by a Mill-Q Reference system equipped with an LC-MS Pak filter (Millipore, Billerica, MA, USA). Gly-pro (Cat# HY-W016887) and Sitagliptin (Cat# HY-13749) were purchased from MedChemExpress LLC (Shanghai, China), and Cbz-pro (Cat# C0713) was obtained from TCI Shanghai (China).

Sample Collection and Grouping

All the convalescent patients (both CA and CO groups) were discharged from the hospital after two consecutive negative results from throat swab tests for SARS-CoV-2. All patients were followed up for months after discharging and their plasma samples were collected at 40 ~ 70 days after the first onset of the disease symptoms onset. All blood
samples were collected with potassium-EDTA blood collection tubes after overnight fasting. We classified the convalescent patients into the CA and CO groups according to the results of the ELISA test of anti-SARS-CoV-2 IgG following the manufacturer’s instructions. Briefly, 100 µL of diluted plasma samples (1:100 to 1:800 dilution) was added to pre-coated plates, and the plates were then incubated at 37°C for 1 h. After washing, 100 µL of horseradish peroxidase (HRP)-conjugated RBD protein of SARS-CoV-2 was added into each well, followed by 30 more min of incubation at 37°C. After washing, the OD value at 450 nm (A450) was determined. The cutoff for negative was calculated by summing 0.090 and the average A450 of negative-control. A sample was determined negative when its A450 was below this cutoff value. The remained plasma samples were then stored at −80°C.

**Detection of Cytokines in Plasma**

The cytokines in the plasma samples of the convalescent patients (both CA and CO groups) and the healthy subjects were detected with the Human Inflammation Array Q3 Kit (RayBiotech Life, Peachtree Corners, GA, USA) following the manufacturer’s instruction. Briefly, the Cytokine Standard Mix was first reconstituted by adding 500 µl of Sample Diluent. Then the dissolved standard was serially 3-fold diluted with Sample Diluent to prepare eight cytokine standard solutions of various concentrations. Glass slides were blocked with 100 µL Sample Diluent in each well followed by a 30-minute incubation at room temperature. Then 100 µL standard cytokines or samples were loaded to the slide and incubated at room temperature for 2 hours. Reconstitution of the detection antibody was prepared by adding 1.4 mL of Sample Diluent to Biotinylated Antibody Cocktail. Then 80 µL of the detection antibody cocktail was added to each well of the washed glass, which was incubated at room temperature for another 2 hours. Cy3 equivalent dye-conjugated streptavidin was dissolved in 1.4 mL of Sample Diluent. After slide washing, 80 µL of Cy3 equivalent dye-conjugated streptavidin solution was added to each well and incubated at room temperature for 1 hour in darkness. After washing, the signals were detected with a laser scanner system Axon GenePix and quantified according to the readouts of cytokine standards.
Absolutely Quantitative Metabolomics of Plasma Samples

Metabolomics of plasma samples was performed and quantified according to the previously described method with minor modifications [1]. In brief, the standard solutions (5.0 mg/mL) were made by dissolving the accurately weighed chemicals in the appropriate solutions including water, methanol, sodium hydroxide solution, or hydrochloric acid solution. Then the stock calibration solutions were mixed from the appropriate amounts of individual stock solutions following the instruction of the manufacturer. After thawing at 4°C, 20-μL aliquots of the samples were added to a 96-well plate. Also were added to the plate the calibration solutions of eight various concentrations, quality control (QC) samples (equally mixed samples), as well as solvent blank. Then 120 μL of the standard solution was added to each well. The microporous plate was covered with aluminum foil, placed on a constant-temperature mixer, and vibrated at 10°C, 650 rpm for 20 min. After centrifugation at 4000g for 20 min, 30μL supernatant from each well was transferred to a new 96-well plate. The derivative reagents of the Q300 Kit were added to all wells and the plate was covered and incubated at 1200 rpm at 30°C to carry out derivatization for 60 min. After derivatization, 330 μL of precooled 50% methanol solution was added to each well and mixed with the samples at 1200 rpm at 10°C for 5 minutes. Then the plate was centrifuged at 4000g, 4°C for 30 minutes. Finally, the supernatant was further transferred to a new 96-well plate and put into the automatic sampler of UPLC-MS analysis.

The Shim-pack UFLC SHIMADZU CBM A ultrahigh-performance liquid chromatography (UHPLC) system (SHIMADZU, Japan) coupled with QTRAP 6500+ triple quadrupole mass spectrometer (Sciex, Washington, USA) was used to analyze the metabolomics. ACQUITY UPLC BEH C18 1.7 μm VanGuard pre-column (2.1mm × 5 mm) and ACQUITY UPLC BEH C18 1.7 μm analytical column (2.1 × 100 mm) (Waters Corporation, Milford, MA, USA) were applied to this system. The mobile phases A and B were 0.1% formic acid solution in water and acetonitrile-IPA mixture (70:30, v/v), respectively. A 5-μL injection of each sample was maintained at 40°C and the flow rate was 0.40 mL/min. The mobile phase gradient programme was: 0-1 min (5% B), 1-11 min (5-78% B), 11-13.5 min (78-95% B), 13.5-14 min (95-100% B), 14-16 min (100% B), 16-16.1 min (100-5% B), and 16.1-18 min (5% B). The mass spectrometer was operated in both positive and negative modes, with capillary voltages of 1.5 and 2.0 kV, respectively.
The source and desolvation temperatures were 150°C and 550 °C, respectively. The flow rate of desolvation gas was 1000 L/hour.

The raw MS data were processed using a website-based platform named Targeted Metabolome Batch Quantification (TMBQ, Metabo-profile Co., Ltd., Shanghai, China). Peaks A calibration curve was plotted for each standard based on the results from calibration solutions of various concentrations. The metabolite concentration in each sample was calculated from the equation \( y = ax + b \) fitted to the calibration curve. The correctness of all calibration curves and integrated peak areas corresponding to each compound was checked manually.

**Mouse Immunization and Serum Antibody Measurement.**

Single doses (50 μg) SARS-CoV-2 RBD protein domain was intraperitoneally (i.p.) injected to 48 female BALB/c mice (6 weeks old) at week 0. Then the mice were randomly assigned into four groups (12 mice each). Two groups (Gly-Pro and Gly-Pro&sitagliptin) of mice were injected with Gly-Pro (3.7μmol/kg/day) daily and one of them (Gly-Pro&sitagliptin) was also gavaged with sitagliptin (10 mg/kg/day). The third group of mice was injected with N-Benzylxycarbony-L-Proline (Cbz-pro) at a dose of 60mg/kg/day. PBS of the same volume was injected into the last group, serving as the control. All mice were weighed weekly and blood samples were collected from individual mice per week.

The total IgG antibodies against the RBD domain of SARS–CoV-2 were measured as described. In brief, a 96-well microtiter plate was first coated with HEK 293F-expressed RBD protein (25 ng/well) before use. Serial 2-fold dilutions of 50 μL of mouse antisera (starting at a 1:100 dilution) were added to the blocked plate, which was then incubated at 37 °C for 2 hours. HRP-conjugated anti-mouse IgG was then added to each well and incubated for 2 more hours at 37 °C. Absorbance at 450 nm (OD\(_{450}\)) was measured after color development and the area under the curve (AUC) of reciprocal dilution was integrated for each mouse sample.

Two weeks post-immunization, the lymph nodes and spleens were collected from all four groups of mice. Lymph node (for GC B and Tfh cell analysis) and spleen (for plasma cell analysis) cells were washed with PBS and subjected to analysis by FACS (Becton
Dickinson Pont de Claix, France) using the following antibodies: Alexa Flour 700 anti-mouse CD45, Brilliant Violet 605 anti-mouse CD185 CXCR5, Brilliant Violet 421 anti-mouse CD279 PD-1, PerCP/Cy5.5 anti-mouse/human CD45R/B220, Brilliant Violet 510 anti-human CD95, APC anti-mouse/human GL7, APC anti-mouse/rat/human CD27, PE anti-mouse CD11c, FITC anti-mouse I-A/I-E, PE anti-mouse CD138 (Biolegend). The data were collected with 1×10^5 cells per sample. Appropriate conjugated isotype-matched mAbs were used as controls.

**Data Processing and Sample Classification with Machine-Learning Approaches**

The concentrations of the quantified inflammatory factors and metabolites were processed using R. To estimate the difference of serum inflammatory factors and metabolites between groups, the least-squares (LS) means were estimated with a linear model (LM) adjusted for age, performed by the EMMEANS package (v1.6.3) followed by Benjamini-Hochberg (BH) correction for multiple hypothesis test. The Inflammatory factors and metabolites with false discovery rate (FDR, a.k.a. q) < 0.05 between every two groups were considered significant. Three-way ANOVA with the factors groups, age, and sex was performed using the RSTATIX package (v0.7.0) to estimate the effects of age, sex and their interactions with the grouping of subjects. The features showing significantly (BH-corrected q < 0.05) affected by age, sex, or the interactions involving them were excluded from the following feature selection. Features (i.e., the quantified inflammatory factors and metabolites) were selected with the least absolute shrinkage and selection operator (LASSO) regression, conducted with the GLMNET package (v4.1.2) in. Multinomial logistic regression with 5-fold validation was performed 10,000 times due to the potential instability in small data set training. Regularization penalty λ, the hyperparameter of LASSO, was determined based on the averaged misclassification error (MCE) and its standard error (SE). The λ resulting in the largest MCE smaller than or equal to minimum MCE + 1SE was used to select features using the full dataset. Different machine learning algorithms were applied using these selected features. Random forest was done with the R package randomForest (v4.7.1), while neural network, support vector machine, ensemble of subspace discriminant, linear discriminant analysis, and k-nearest
neighbors was conducted with machine learning Apps of Matlab (version R2018b, MathWorks, Natick, Massachusetts, USA).

Principal component analysis (PCA) was performed using the ROPLS package (v1.24.0) in R. Pearson’s correlation coefficients were calculated from the scaled concentration (i.e., z-score) of inflammatory factors and metabolites with the STATS package (v3.6.2) in R. FDR was estimated by the Benjamini- Hochberg (BH) correction for multiple hypothesis test.
Supplementary information

Figure S1. The cytokine alterations between three groups.
Figure S1. The cytokine alterations between three groups (Cont’d).
Figure S1. The cytokine alterations between three groups (Cont'd).
Figure S2. The Z-score of quantified metabolites from all three groups.
Figure S3. The metabolites of significant difference between at least two groups.
Figure S3. The metabolites of significant difference between at least two groups (Cont’d).
Figure S3. The metabolites of significant difference between at least two groups (Cont’d).
Figure S3. The metabolites of significant difference between at least two groups (Cont’d).
Figure S4. The metabolites of significant difference between at least two groups in males or in females.
Figure S4. The metabolites of significant difference between at least two groups in males or in females (Cont’d).
Figure S5. The metabolites with significantly different levels between the male and female subjects of the CO group. a and b, the metabolites with signification higher levels in the female (a) and male (b) patients of the CO group but showing no difference in the H group.
Figure S6. The cytokine and metabolite pair showing correlation in the CO group but not the CA group. a and f, scatter plot of the inflammatory factor-metabolite pairs IFNg vs. GMP (a), IFNg vs. b-Neu5ac (b), IFNg vs. R5P (c), GM-CSF vs fumarate (d), GM-CSF vs b-Neu5ac (e), and MIP-1a vs phenylpyruvate (f).
Figure S7. LASSO-based feature selection. (a) Mis-classification error (MCE) of 10000 times of 5-fold cross-validation. Mean ± SE (standard error) is plotted. The dashed lines annotated the values of the hyperparameter $\lambda$ that gave the minimum mean MCE (min) and that gave the most regularized model such that MCE within one SE of the minimum (1 SE). (b) the $\ell_2$ norm of each feature's coefficient vector with the change of $\lambda$. 
Figure S8. The machine learning models used to classify the CA, CO and H groups. a-f, the confusion matrices of machine learning models using neural network (a), support vector machine (b), subspace discriminant (c), linear discriminant analysis (d), random forest (e), and k-nearest neighbors (f).
Figure S9. The coefficients of different features result from the binomial logistical regression applied to classify the CA and CO groups. a-c, the results for all six features of individual AUC > 0.75 (a), the three features of the highest AUCs (b), and only two features Gly-pro and MCSF (c).

Figure S10. The bodyweights of RBD-immunized mice. The error bars show the standard deviation of 12 female BALB/c mice in each group.
Figure S11. Left: Representative flow cytometry gating strategy for identification of GC B (CD45+B220+CD95+GL-7+), Tfh (CD45+CD4+CD185+ PD-1+), and plasma (B220+CD27+CD138+); Right: Proposed working model modulated by glycylylproline (Gly-pro).
References:

[1] Xie G, et al. (2021) A Metabolite Array Technology for Precision Medicine. *Anal Chem.* 93(14):5709-5717.