Journal of proteome research

pubs.acs.org/jpr

Article

Integrative Modeling of Plasma Metabolic and Lipoprotein Biomarkers of SARS-CoV-2 Infection in Spanish and Australian COVID-19 Patient Cohorts

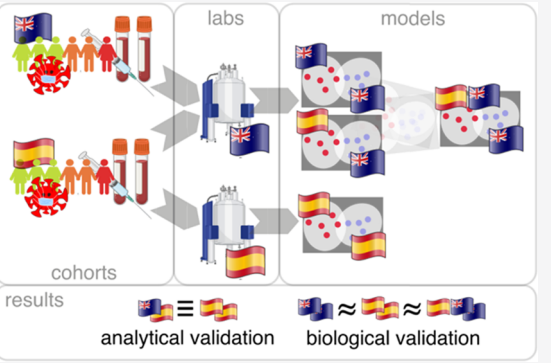
Reika Masuda,[○] Samantha Lodge,[○] Philipp Nitschke,[○] Manfred Spraul, Hartmut Schaefer, Sze-How Bong, Torben Kimhofer, Drew Hall, Ruey Leng Loo, Maider Bizkarguenaga, Chiara Bruzzone, Rubén Gil-Redondo, Nieves Embade, José M. Mato, Elaine Holmes, Julien Wist,* Oscar Millet,* and Jeremy K. Nicholson*

Cite This: *J. Proteome Res.* 2021, 20, 4139−4152

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Quantitative plasma lipoprotein and metabolite profiles were measured on an autonomous community of the Basque Country (Spain) cohort consisting of hospitalized COVID-19 patients (n = 72) and a matched control group (n = 75) and a Western Australian (WA) cohort consisting of (n = 17) SARS-CoV-2 positives and (n = 20) healthy controls using 600 MHz ^{1}H nuclear magnetic resonance (NMR) spectroscopy. Spanish samples were measured in two laboratories using one-dimensional (1D) solvent-suppressed and T_2-filtered methods with in vitro diagnostic quantification of lipoproteins and metabolites. SARS-CoV-2 positive patients and healthy controls from both populations were modeled and cross-projected to estimate the biological similarities and validate biomarkers. Using the top 15 most discriminatory variables enabled construction of a cross-predictive model with 100% sensitivity and specificity (within populations) and 100% sensitivity and 82% specificity (between populations). Minor differences were observed between the control metabolic variables in the two cohorts, but the lipoproteins were virtually indistinguishable. We observed highly significant infection-related reductions in high-density lipoprotein (HDL) subfraction 4 phospholipids, apolipoproteins A1 and A2,that have previously been associated with negative regulation of blood coagulation and fibrinolysis. The Spanish and Australian diagnostic SARS-CoV-2 biomarkers were mathematically and biologically equivalent, demonstrating that NMR-based technologies are suitable for the study of the comparative pathology of COVID-19 via plasma phenotyping.



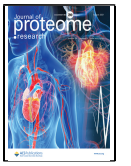
KEYWORDS: *COVID-19, SARS-CoV-2, NMR spectroscopy, plasma IVDr, metabolic phenotyping, diagnostic modeling, lipoproteins, phenoconversion, population cross-validation*

INTRODUCTION

SARS-CoV-2 infection gives rise to well-described, but highly variable, acute pulmonary disease and a complex set of multiorgan systemic pathologies and symptoms.[1−3] These pathologies are reflected in collective complex changes in the blood plasma biochemistry expressed in perturbations in multiple metabolic, amino acid, lipidic, and multiparametric lipoprotein signatures.[4,5] Nuclear magnetic resonance (NMR) spectroscopy of plasma/serum has long been known to be of value in characterizing metabolic diseases and is analytically extremely robust and reliable,[6,7] and detection of a wide range of chemical classes of low-molecular-weight (MW) metabolites can be achieved using a variety of NMR experiments.[8−10] NMR spectroscopic signatures of SARS-CoV-2 infection have uncovered metabolic perturbations reflecting the acute responses to infection.[4,5,11−16] "Post-Acute COVID-19 Syndrome" (PACS) patients also experience poor systemic recovery and exhibit multiple long-term symptoms and biochemical abnormalities.[17,18]

The clinical utility of metabolic biomarkers in the diagnosis, prognosis, and stratification of a disease is dependent upon the analytical and biological reproducibility of the metabolic signature across populations. The robustness of in vitro diagnostic NMR methods has been demonstrated in interlaboratory trials[19−21] comparing standards and individual sample profiles. In the case of major emergent diseases such as COVID-

Received: May 31, 2021
Published: July 12, 2021



https://doi.org/10.1021/acs.jproteome.1c00458
J. Proteome Res. 2021, 20, 4139−4152

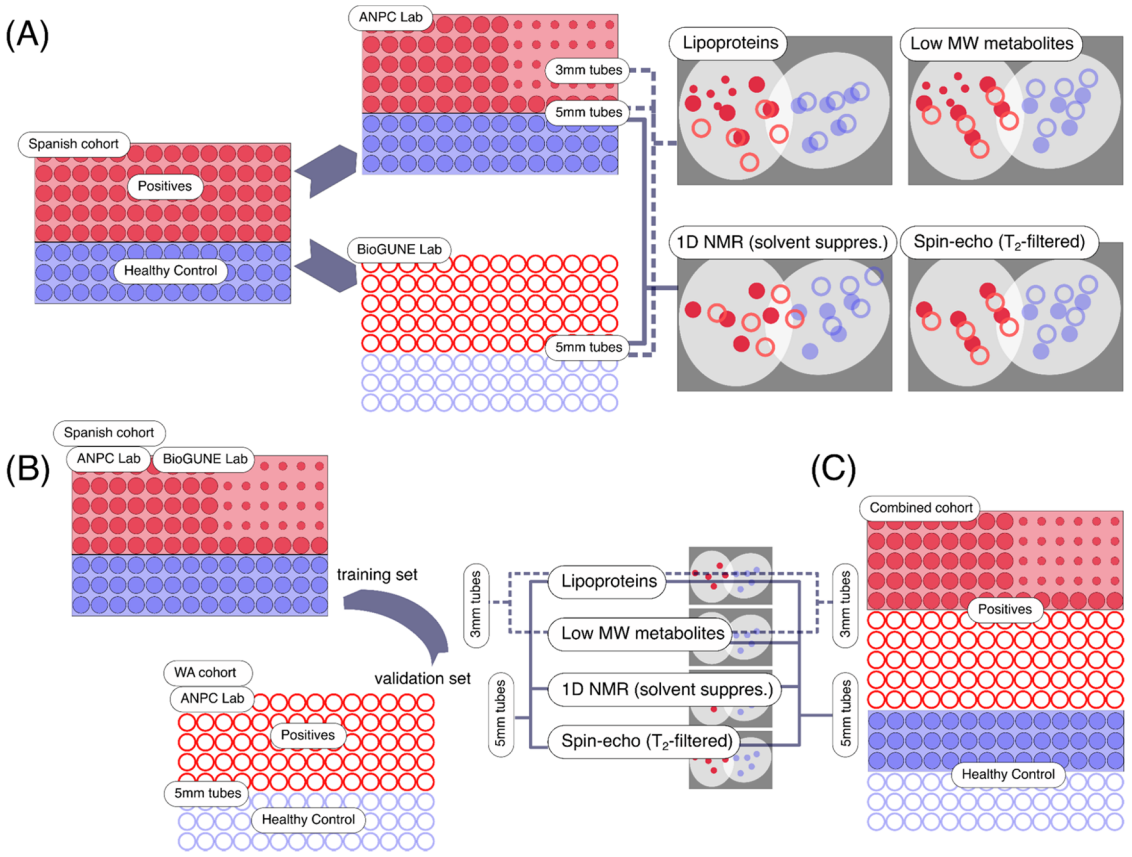


Figure 1. (A) Interlaboratory comparison of analytical methods: Biomarkers obtained by measuring the Spanish cohort in two different laboratories were compared. All 75 control and 72 SARS-CoV-2 positive samples were measured in 5 mm diameter tubes at the CIC bioGUNE laboratory. Measurements were repeated in 5 mm diameter tubes for 75 controls and 35 of the positive samples at the ANPC laboratory with the remaining 37 SARS-CoV-2 positive samples for the Spanish cohort being run in 3 mm diameter tubes (due to insufficient remaining volume). Quantification using the IVDr-extracted signals enabled data acquired using different experimental conditions to be merged, therefore data from the 3 mm tubes were used along with 5 mm tubes in interlaboratory comparisons of the targeted models (lipoproteins and low-MW metabolites). Only spectra measured in 5 mm tubes were used in the interlaboratory comparison of the full spectral profiles. (B) Cohort comparison: Full spectral data and IVDr-extracted data for lipoprotein and small-molecule parameters from two independent cohorts, from Spain and Western Australia (WA), were compared. The Spanish cohort was used to train and build a model, and the WA cohort was used as a validation data set (and vice versa). Again, data obtained from 3 mm tubes were only used in targeted models. (C) Fusion of cohorts: Statistical significance of SARS-CoV-2 infection "biomarkers" obtained by combining the two cohorts was compared with the results obtained for the individual models for the Spanish and Australian cohorts generated in step B compared with the one obtained by combining both cohorts.

19, it is important to understand the comparative biochemistry based on cross-validated methods, ideally performed with multiple technologies applied to the same samples to allow rapid and meaningful comparison of data from multiple cohorts. NMR spectroscopic characterization of SARS-CoV-2 infection has uncovered lipoprotein, acute phase reactive glycoprotein, and small-molecule abnormalities associated with the disease,[4,11,14,22,23] and multiple NMR spectroscopic experiments performed on the same samples have been applied to extract a range of diagnostic biochemical components and compartments that report on disease-induced phenoconversion[11,24] and biochemical recovery phenoreversion.[18]

In addition to analytical variation across laboratories, metabolic variations between populations can contribute to observed interstudy differences in the metabolic signatures of a disease.[25,26] Genes and the environment interact to create metabolic phenotypes and disease risks in individuals and whole populations,[26,27] which can help stratify patients into disease subtypes, and also affect responses of individuals to disease therapies.[7] In the case of SARS-CoV-2 infection, we have shown that metabolic signatures of pathological conditions produce greater contributions to metabolic phenotypes over the background physiological variation and can be predictively modeled to extract population risk biomarkers.[28] However, there are few real-world patient diagnostic studies, where both cross-validation of the analytical methods and population differences are independently evaluated in different laboratories and then comodeled to improve biomarker selection. Here, we present a combined analytical and biological interlaboratory cross-validation study of two independent patient cohorts from Spain and Australia. We demonstrate the robustness of the NMR-based diagnostic approach and biomarker integrity for the biochemical sequelae of SARS-CoV-2 infection and extract integrated cross-population biomarkers for acute phase SARS-CoV-2 infections.

MATERIALS AND METHODS

Participant Enrollment and Sample Collection

Autonomous Community of the Basque Country (Spain). The cohort consisted of (i) patients who tested positive for SARS-CoV-2 infection from upper and/or lower respiratory tract swabs by the reverse transcription-polymerase

https://doi.org/10.1021/acs.jproteome.1c00458
J. Proteome Res. 2021, 20, 4139−4152

chain reaction (RT-PCR) (n = 72) and (ii) 75 healthy control participants (Figure 1A and Table S1). All serum samples were collected by the Basque Biobank for research (BIOEF). Healthy serum samples were collected before the COVID-19 pandemic from the active population while the COVID-19 samples were collected at the Cruces University Hospital (Barakaldo, Spain) from patients who presented compatible symptoms, confirmed by a RT-PCR assay on nasal swab samples. All participants provided informed consent to clinical investigations, according to the Declaration of Helsinki, and all data were anonymized to protect their confidentiality. The sample-handling protocol was evaluated and approved by the Comité de Ética de Investigación con medicamentos de Euskadi (CEIm-E, PI+CES-BIOEF 2020-04 and PI219130). Shipment of human samples to ANPC had the approval of the Ministry of Health of the Spanish Government. Samples were stored at −80 °C.

Western Australian Test Cohort. Blood plasma samples were collected from a cohort of adult individuals in a study initiated at Fiona Stanley Hospital in the Western Australia South Metropolitan Health Service catchment as part of the International Severe Acute Respiratory and Emerging Infections Consortium (ISARIC)/World Health Organisation (WHO) pandemic trail framework (SMHS Research Governance Office PRN:3976 and Murdoch University Ethics No. 2020/052). Healthy control participants from Western Australia were enrolled as volunteers. Study details were provided, and written consent was obtained prior to data collection in accordance with the ethical governance (Murdoch University Ethics No. 2020/053). Patients who were presented with COVID-19 disease symptoms and subsequently tested positive for SARS-CoV-2 infection from upper and/or lower respiratory tract swabs by RT-PCR (n = 17 participants) were recruited from the Fiona Stanley and Royal Perth Hospitals (Western Australian cohort) and additional healthy control participants (n = 20) were recruited (Tables S2 and S3). Plasma samples were stored at −80 °C. Sample processing was performed according to Bruker IVDr protocols for the small-molecule and lipoprotein data[19] and according to our recommended procedures for IVDr metabolic analysis of COVID-19 plasma samples.[28] We have previously published an analysis of this patient group with respect to metabolic biomarker discovery as determined by NMR spectroscopy and mass spectrometry.[4] In the present study, we use these data as a test set to project onto the Spanish cohort training set models measured and generated independently in the ANPC and the CIC bioGUNE laboratories in Spain (Figure 1B).

^{1}H NMR Sample Preparation at CIC bioGUNE. Samples were stored at −80 °C until the day of analysis, when they were defrosted at room temperature for 30 min. NMR samples were prepared in a SamplePro Tube (Bruker Biospin) robot system for liquid handling with integrated temperature control. Every sample was automatically prepared as a mixture of phosphate buffer (75 mM Na_2HPO_4, 2 mM NaN_3, 4.6 mM sodium trimethylsilyl propionate-[2,2,3,3-$^2H_4$] (TSP) in H_2O/D_2O 4:1, pH 7.4 ± 0.1) and serum at a 1:1 ratio for a final volume of 600 μL into 5 mm SampleJet NMR tubes. Samples were then manually shaken for several seconds and stored at 5 °C inside the SampleJet automatic sample changer until measurement (<24 h).

^{1}H NMR Sample Preparation at the ANPC. Plasma samples were thawed at 20 °C for 30 min and then centrifuged for 10 min at 13 000g at 4 °C. For the Western Australian test cohort, all plasma samples were prepared in 5 mm outer diameter SampleJet NMR tubes, following the recommended procedures for in vitro analytical and diagnostic procedures[29] using 300 μL of plasma mixed with 300 μL of phosphate buffer (75 mM Na_2HPO_4, 2 mM NaN_3, 4.6 mM sodium trimethylsilyl propionate-[2,2,3,3-$^2H_4$] (TSP) in H_2O/D_2O 4:1, pH 7.4 ± 0.1).

For the Spanish sample cohort, all healthy control samples were prepared in 5 mm outer diameter SampleJet NMR tubes. For the SARS-CoV-2 positive samples, 35 were prepared in 5 mm outer diameter SampleJet NMR tubes as described for the Western Australian cohort while the remaining SARS-CoV-2 positive samples (n = 37), with a limited sample volume, were prepared in 3 mm outer diameter SampleJet NMR tubes (Figure 1A). For the 3 mm outer diameter SampleJet NMR tubes, 90 μL of plasma was mixed with 90 μL of phosphate buffer (75 mM Na_2HPO_4, 2 mM NaN_3, 4.6 mM sodium trimethylsilyl propionate-[2,2,3,3-$^2H_4$] (TSP) in H_2O/D_2O 4:1, pH 7.4 ± 0.1) and 180 μL transferred into an NMR tube.

^{1}H NMR Spectroscopy Data Acquisition and Processing Parameters. NMR spectroscopic analyses were performed with two 600 MHz Bruker Avance III HD spectrometers, each equipped with a 5 mm BBI probe and fitted with the Bruker SampleJet robot cooling system set to 5 °C. A full quantitative calibration was completed prior to the analysis using a protocol described elsewhere.[29] All experiments were acquired using the Bruker In Vitro Diagnostics research (IVDr) methods. For each sample prepared in 5 mm tubes, two experiments were completed at 310 K in automation mode, amounting to a total of 8.3 min acquisition time per sample: a standard one-dimensional (1D) experiment with solvent presaturation (32 scans, 98k data points, a spectral width of 30 ppm) and a Carr−Purcell−Meiboom−Gill (CPMG) spin-echo experiment (32 scans, 74k data points, a spectral width of 20 ppm), which filters the spectrum by differential T_2 relaxation, removing signals from fast relaxing protons on large molecules. For each sample prepared in 3 mm tubes, two experiments were completed in automation mode (28 min total experimental acquisition time), a standard 1D experiment with solvent presaturation (128 scans, 98k data points, a spectral width of 30 ppm) and a T_2-filtered (spin-echo) experiment (128 scans, 74k data points, a spectral width of 20 ppm). Data were processed in automation using Bruker Topspin 3.6.2 and ICON NMR automation to achieve phasing and baseline correction. In addition to the information extracted from the full spectral data from the standard 1D and T_2-filtered (spin-echo) spectra, a total of 112 lipoprotein parameters for each sample were generated using the Bruker IVDr lipoprotein subclass analysis (B.I.-LISA) method whereby the $-(CH_2)_n$ at δ = 1.25 and $-CH_3$ at δ = 0.80 peaks of the 1D spectrum after normalization to the Bruker QuantRef manager within Topspin were quantified using a PLS-2 regression model. B.I.LISA data consist of the total plasma lipid analyte cholesterol, free cholesterol, phospholipids, triglycerides, apolipoproteins A1/A2/B100 and the B100/A1 ratio, and distributions of analytes in different density classes of plasma lipoproteins: high-density lipoprotein (HDL, density 1.063−1.210 kg/L), intermediate-density lipoprotein (IDL, density 1.006−1.019 kg/L), low-density lipoprotein (LDL, density 1.09−1.63 kg/L), and very low-density lipoprotein (VLDL, 0.950−1.006 kg/L). The main lipoprotein classes HDL, LDL, and VLDL were subdivided into different density subclasses. LDL subdivisions included: LDL-1: 1.019−1.031 kg/L, LDL-2: 1.031−1.034 kg/L, LDL-3: 1.034−1.037 kg/L, LDL-4: 1.037−1.040 kg/L, LDL-5: 1.040−1.044 kg/L, and LDL-6: 1.044−

https://doi.org/10.1021/acs.jproteome.1c00458
J. Proteome Res. 2021, 20, 4139−4152

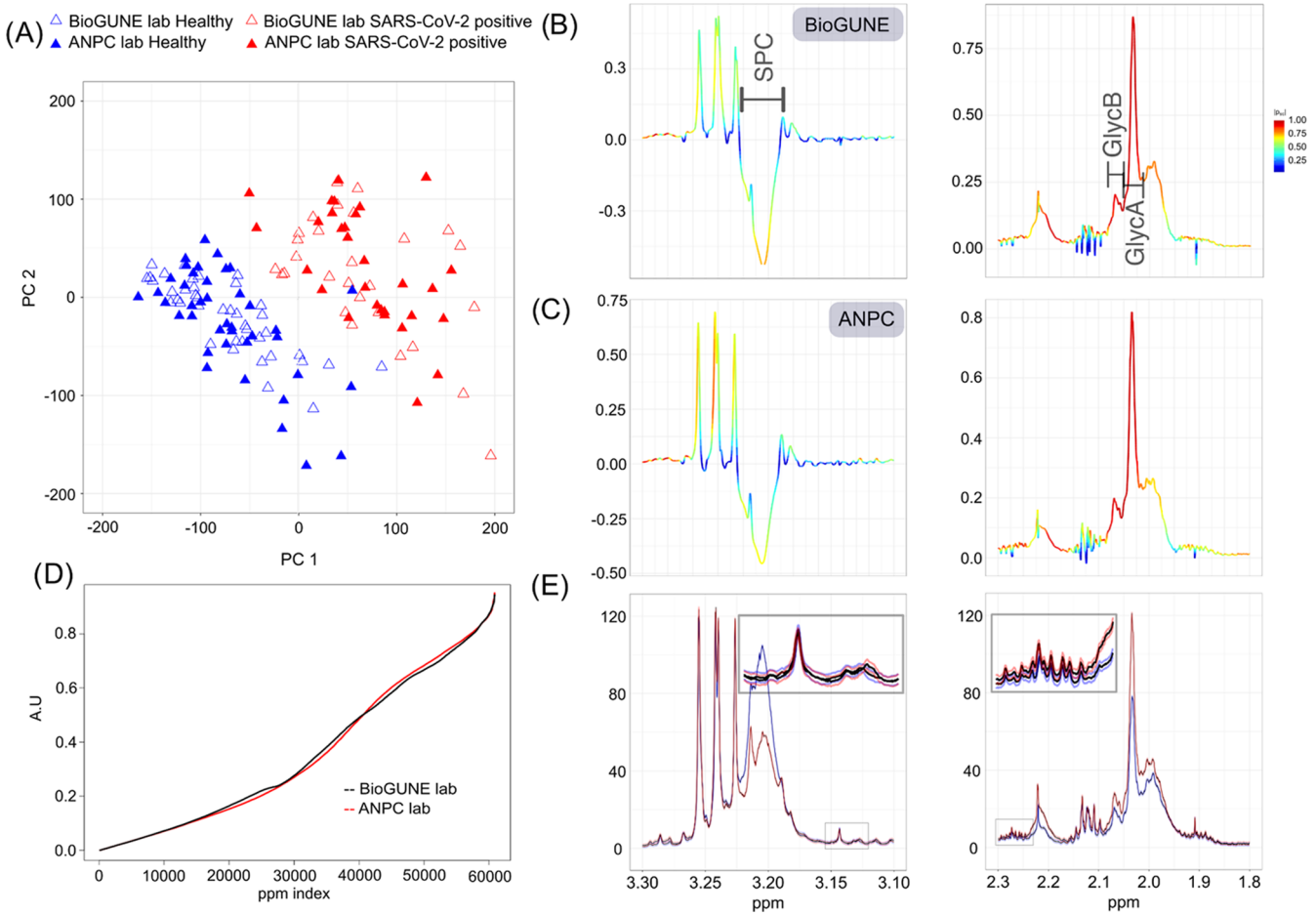


Figure 2. (A) PCA scores plot of the spin-echo NMR data for the Spanish cohort measured in the CIC bioGUNE lab (open triangles) overlaid with equivalent data generated in the ANPC lab (filled triangles) (BioGUNE lab PC1 = 19.7%, PC2 = 8.0% and ANPC lab PC1 = 20.0%, PC2 = 8.4%). Selected regions of the statistically reconstructed first principal component (PC1) measured (B) at CIC bioGUNE and (C) at ANPC. The trace and the color code represent the covariance and the correlation of the scores with the data matrix. (D) Ranked index of the correlation between the scores and the data matrix to highlight the similarity between PC1 (B, C) measured in both laboratories (black line = BioGUNE and red line = ANPC). (E) Median trace for each group (red = SARS-CoV-2, n = 33; blue = control, n = 33) plotted together with the corresponding experimental error for each group.

1.063 kg/L. HDL subfractions were also assigned to four density classes: HDL-1: 1.063−1.100 kg/L, HDL-2: 1.100−1.125 kg/L, HDL-3: 1.125−1.175 kg/L, and HDL-4: 1.175−1.210 kg/L, and the VLDL subfractions were divided into five density classes. A list of all of the 112 lipoprotein subfractions and parameter annotations are shown in Table S4. In addition to the 112 lipoprotein parameters, 23 low-molecular-weight metabolite concentrations were obtained from the Bruker IVDr quantification in plasma/serum B.I.Quant-PS (acetic acid, acetoacetic acid, acetone, alanine, citric acid, creatine, creatinine, formic acid, glucose, glutamic acid, glutamine, glycine, histidine, D-3-hydroxybutyric acid, isoleucine, lactic acid, leucine, lysine, methionine, phenylalanine, pyruvic acid, tyrosine, valine). Metabolites that were below the limits of detection were excluded from the analyses.

NMR Data Modeling. The electronic reference spike[30] was calibrated against a sample of known concentration and each sample was then calibrated against the electronic spike to account for variation in performance between spectrometers. This was achieved by dividing both the standard, 1D with water suppression and the T_2-filtered (spin-echo) spectra by the eretic factor in an elementwise fashion. Both the standard 1D and T_2-filtered (spin-echo) NMR spectral data sets were calibrated to the α-anomeric proton signal of glucose at δ5.23 ppm. Each spectrum was baseline corrected using an asymmetric least-squares routine, spectral regions corresponding to the residual water resonance signals (4.60−4.85 ppm) or regions predominantly containing noise (<0.5 and >9.5 ppm) were excluded from analyses.[31] Data were mean-centered and scaled to unit-variance prior to multivariate modeling. Principal component analysis (PCA) was used to assess the main sources of structured variation within each data set (lipoprotein classes and subclasses, low-MW metabolites, and 1D and T_2-filtered spectra), while the orthogonal projection to latent structures-discriminant analysis (OPLS-DA)[32] method was used to model the infection-related variance in the data and to extract discriminating features. The optimal number of orthogonal components for each model was determined using the area under the receiver operator characteristic curve (AUROC) calculated from predictive component scores, generated using an internal sevenfold cross-validation (CV) procedure. All models are shown in Figures 2−6 and S1−S8.

Analytical Validation. To compare analytical reproducibility between laboratories (Figure 1A), eight OPLS-DA models that differentiate SARS-CoV-2 positive participants from healthy controls were built for the Spanish cohort using the lipoprotein

https://doi.org/10.1021/acs.jproteome.1c00458
J. Proteome Res. 2021, 20, 4139−4152

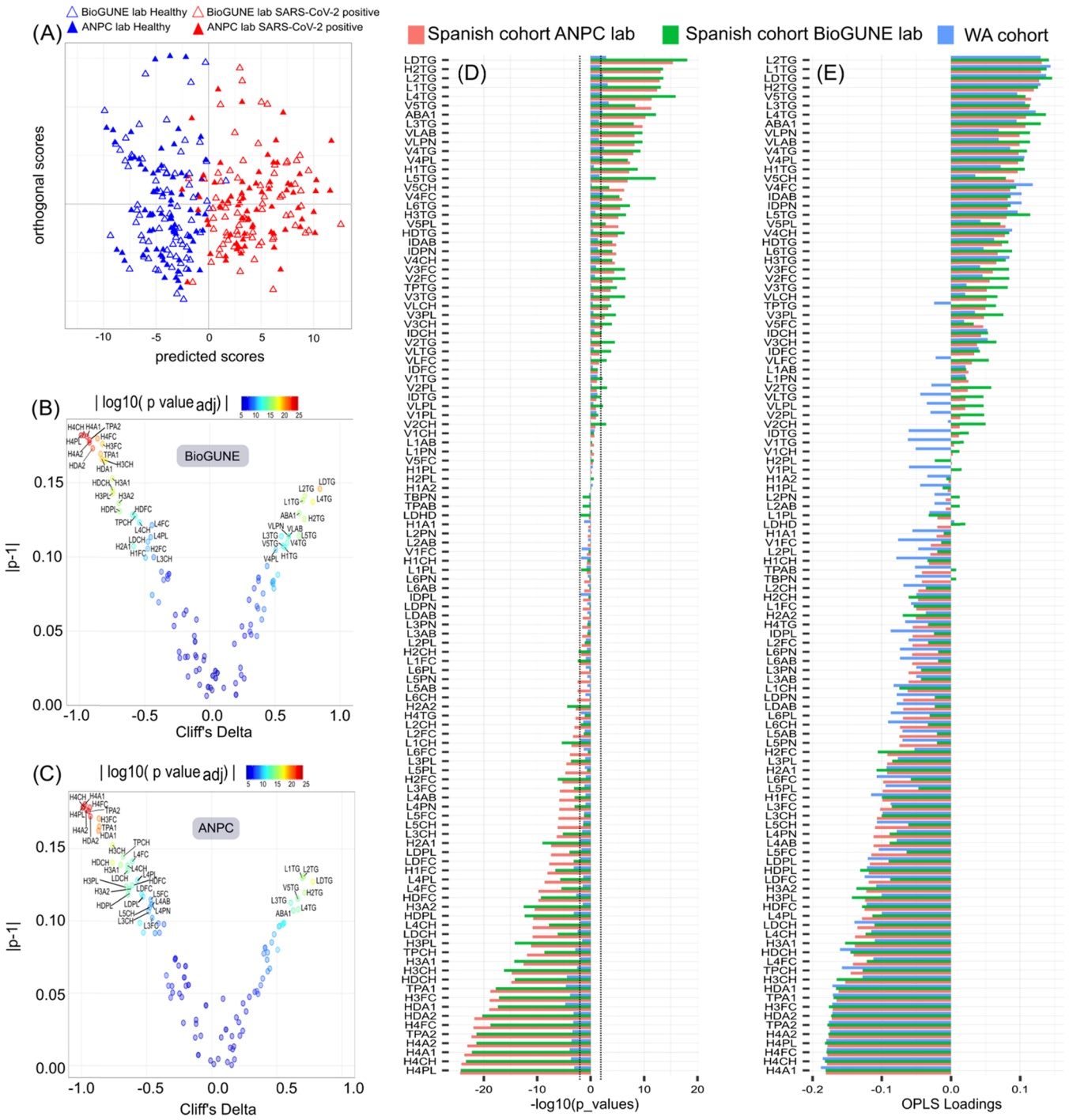


Figure 3. OPLS-DA scores and loadings plots of the lipoproteins of the Spanish cohort measured in the CIC bioGUNE and ANPC laboratories. (A) OPLS-DA scores plot of the lipoprotein parameters from CIC bioGUNE (open triangles; AUROC = 0.99) overlaid with those generated at the ANPC (closed triangles; AUROC = 1.00); (B) eruption plot of the lipoprotein data acquired in the CIC bioGUNE laboratory; and (C) eruption plot of the lipoprotein data acquired in the ANPC laboratory. (D) Lipoproteins −log 10 (p-values) between healthy controls and SARS-CoV-2 positive patients of the Spanish cohort run at the CIC bioGUNE and ANPC labs and the Western Australia cohort run at the ANPC lab. (E) Lipoproteins OPLS-DA loadings between healthy controls and SARS-CoV-2 positive patients. Only unadjusted p-values are shown in the figure for comparison, while tables of the OPLS-DA loadings, Cliff's delta, and associated p-values (unadjusted and adjusted for multiple hypothesis testing with the false discovery rate (FDR) correction) for all of the lipoproteins can be found in the Supporting Information. The two gray vertical dashed lines delineate the confidence interval.

parameters, extracted metabolites, and the standard 1D and T_2-filtered spectral data sets. Each model was constructed using a training sample set that comprised a single time point from PCR-confirmed SARS-CoV-2 patients (n = 72) and healthy control participants (n = 75).

The Cliff's delta statistic, a nonparametric effect size measure that quantifies the group differences of a variable, was calculated for all models. Absolute Cliff's delta values range from 1 that denotes the maximum difference, to zero that denotes no differences between the two groups. The arithmetic sign

https://doi.org/10.1021/acs.jproteome.1c00458
J. Proteome Res. 2021, 20, 4139−4152

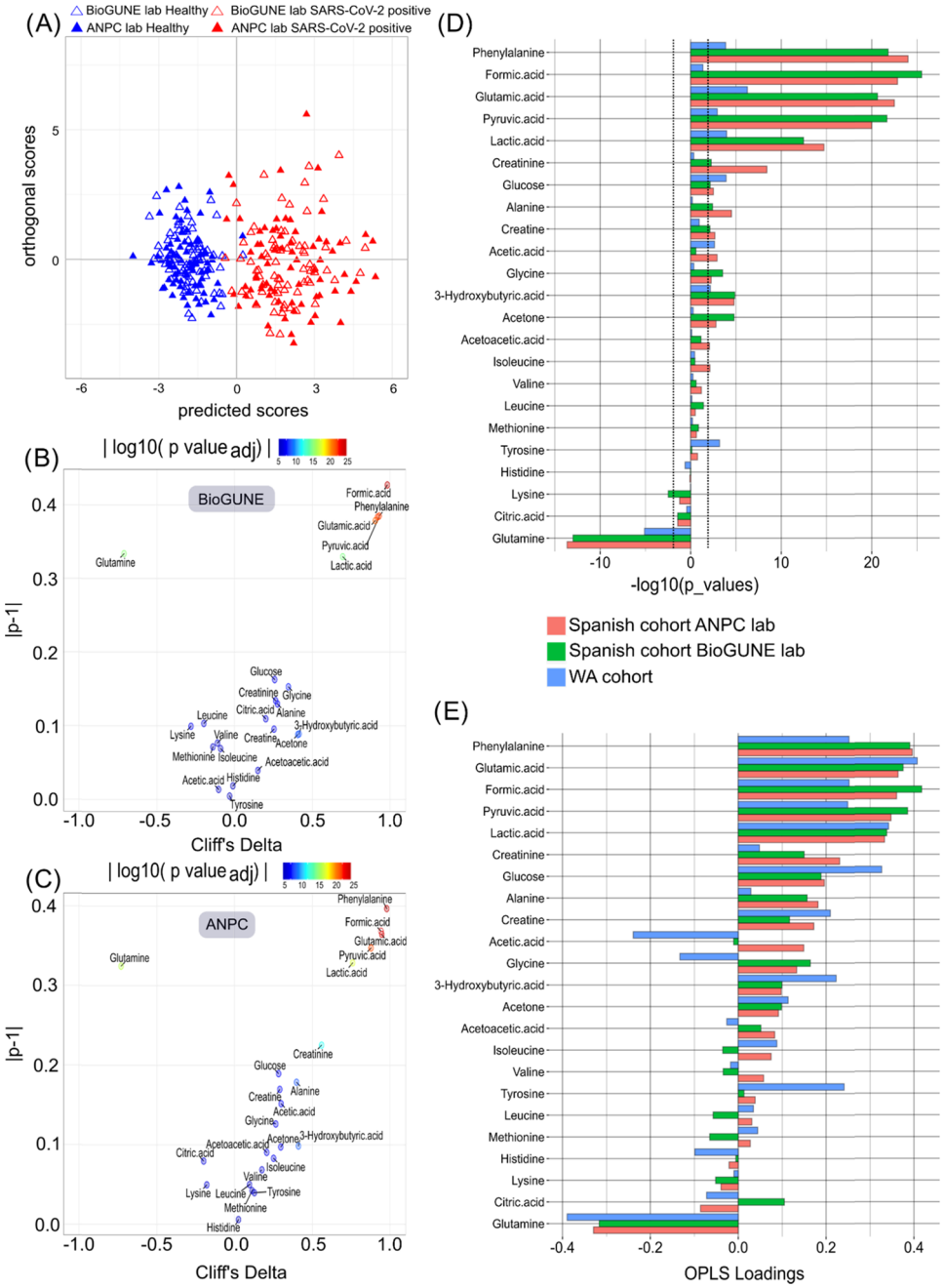


Figure 4. OPLS-DA scores and loadings plots of the low-MW metabolites measured in the CIC bioGUNE and ANPC laboratories. (A) OPLS-DA scores plot of the extracted metabolites from CIC bioGUNE (open triangles; AUROC = 1.0) overlaid with those generated at the ANPC (closed triangles; AUROC = 1.00); (B) eruption plot of the metabolite data acquired in the CIC bioGUNE lab; and (C) eruption plot of the small-molecule data acquired in the ANPC lab. (D) IVDr-extracted low-MW metabolites −log 10 (*p*-values) between healthy controls and SARS-CoV-2 positive patients of the Spanish cohort run at CIC bioGUNE and the ANPC labs and the Western Australian cohort run at the ANPC. (E) IVDr-extracted low-MW metabolite OPLS-DA loadings between healthy controls and the SARS-CoV-2 positive patients. Only unadjusted *p*-values are shown in the figure for comparison, while tables of the OPLS-DA loadings, Cliff's delta, and associated *p*-values (unadjusted and adjusted for multiple hypothesis testing with the false discovery rate (FDR) correction) for all of the metabolites can be found in the Supporting Information. The two gray vertical dashed lines delineate the confidence interval.

indicates an elevation or decrease of that variable with reference to the control group. To test for statistical significance, a two-tailed paired Wilcoxon rank sum was applied and *p*-values adjusted using the false discovery rate (FDR) correction of multiple hypothesis testing. The statistical significance level was set to $\alpha = 0.05$. The description of each parameter, Cliff's delta, and *p*-values can all be found in the Supporting Information (Tables S4−S9 for IVDr-extracted lipoproteins and Tables S10 and S12 for low-MW metabolites).

Biological Validation. The cohorts from Spain and Western Australia were used to cross validate each other, i.e., one cohort was used to build a model for SARS-CoV-2 infection versus healthy controls while the remaining one was used as a validation set and was predicted into the model (Figure 1B). The models were built, as described in the previous section, and the

https://doi.org/10.1021/acs.jproteome.1c00458
J. Proteome Res. 2021, 20, 4139−4152

validation data set was used to compute a confusion matrix. Thus, an orthogonal projection to latent structures-discriminant analysis (OPLS-DA) model was constructed using a training sample set comprising a single time point from PCR-confirmed SARS-CoV-2 patients (n = 35, only including 5 mm tubes) and healthy control participants (randomly sampled from the full set n = 40). This model was then used as a predictor for the WA data set (n = 17 SARS-CoV-2 positive samples and n = 20 healthy controls). A confusion matrix calculated using a validation set prepared from the remaining Spanish samples was used for comparison (with equal sample size) and the sensitivity and specificity reported. This procedure was repeated for each data set of the Spanish cohort, generating four models. The models can be found in Figures 5 and S8.

Fusion of Cohort Data. The two cohorts were combined and modeled as an integrated data set to ascertain whether the predictive strength of the candidate biomarkers identified for the individual cohorts improved with increasing sample size (Figure 1C). The data from both the Spanish and Western Australian cohorts were concatenated and modeled with OPLS-DA, as described in the "NMR Data Modeling" section.

All computation and data visualization were performed using R and RStudio IDE with the open-source R package metabom8 (version 0.4.2), available from GitHub (github.com/tkimhofer/metabom8), while the figures were integrated and prepared using Inkscape (https://inkscape.org, version 1.0.2).

RESULTS AND DISCUSSION

Cross-Validation of NMR Spectral Profile, Lipoprotein, and Low-Molecular-Weight Metabolite Data for Laboratory Interoperability

The split samples from the Spanish cohort measured in the CIC bioGUNE and ANPC laboratories were comapped in the PCA scores space for the T_2-filtered experiment (Figure 2A), illustrating the analytical robustness of the sample preparation and NMR experimental methods. The clear differentiation of the SARS-CoV-2 infected and control groups, reflecting the biological variance between the infected and control groups, was greater than any variation attributed to analytical sources. As with the scores, the reconstructed loadings for the first component (PC1) were found to be in excellent agreement for both laboratories and both NMR experiment types (as illustrated for selected windows of the NMR serum spectra Figure 2B,C). This is also illustrated in Figure 2D where the variable importance for the full spectral range (represented by the color code of the traces in Figure 2B,C) is provided as a ranked index for both laboratories. Furthermore, the median trace for each group (SARS-CoV-2 and control) is displayed with its corresponding experimental errors (Figure 2E) estimated by summing the difference between each replicate pair depicted in red (SARS-CoV-2 infected) and blue (control). The experimental error is largely constant across all spectral regions and overall is of much lower magnitude than the biological effect of interest. For example, the magnitude of the signals for GlycA and GlycB far exceed the analytical variation across the two laboratories. A similar result was obtained for the standard water-suppressed 1D NMR data (Figure S7). Thus, based on the above data the two laboratories generated analytically equivalent results. It should be noted that only the samples with enough volume to be prepared in 5 mm tubes in both laboratories were included in these comparisons using full spectra, i.e., 33 (two samples were excluded) of the SARS-CoV-2 positive samples. Consequently, the number of healthy controls was adjusted to balance the two populations.

It was possible to incorporate samples from both 5 mm outer diameter tubes and 3 mm tubes within the same statistical model for the quantitated lipoprotein parameters measured across the two laboratories since the IVDr method generates absolute concentrations and this accommodates the difference in the experimental conditions between the tube types, e.g., compensating for signal to noise discrepancies. The OPLS-DA models for the two IVDr quantified data sets for lipoproteins and low-molecular-weight molecules (Figures 3 and 4, respectively) again show good agreement between the measurements acquired in the two laboratories. The superimposed scores plots (Figures 3A and 4A) and the corresponding eruption plots stratified at the laboratory (Figures 3B,C and 4B,C) show that the presence/absence of infection accounts for the majority of the variance, with the eruption plots[4] manifesting similar geometries. To better illustrate this similarity, the log of the significance (Figures 3D and 4D) of each extracted parameter in differentiating the infected and control groups for the Spanish cohort was displayed side-by-side (with bioGUNE denoted in orange and ANPC in green). The order of variable significance is largely preserved for the parameters that exert significant weighting in the models, for example, the LDL triglycerides and apolipoprotein A1 (ABA1) (higher in COVID-19) and HDL-related parameters (associated with the healthy group) are ranked in similar order, whereas a greater disparity in the rank order is apparent for the nonsignificant or "noisy" parameters. Within the highly discriminatory lipoprotein parameters, L5TG, L4TG, and ABA1 showed a slight difference in rank-ordered significance between the two laboratories. The magnitudes of the statistically constructed loadings (Figures 3E and 4E) for the data set generated in each laboratory are ranked and displayed side-by-side. Some of the least stable parameters in the models, for example, VLTG, V1TG, IDTG, V1FC, V2FC, and IDFC have been shown to be susceptible to freeze−thaw cycles.[28,33] In the case of the low-MW metabolites, citrate, acetate, and acetoacetate have also been reported to be strongly influenced by freeze−thaw cycles,[33] which may account for the instability in these parameters between the two measurement sites, since the infected samples underwent one more freeze−thaw cycle at the ANPC laboratory than bioGUNE.

Combining the 3 and 5 mm IVDr data sets for lipoproteins and low-MW metabolites in the current study allowed an increase in the group size from 40 to 75 in the control and from 35 to 72 in the infected groups, thereby substantially increasing the statistical power of the analysis. Nevertheless, although the IVDr quantification enabled increased numbers of samples to be included in the model through combination of different tube sizes, the T_2-filtered experiment on lower sample numbers still marginally outperformed the quantified extracted parameters' OPLS-DA models in terms of the sensitivity and specificity values. (Figure S8). This indicates that the full spectra contain extra systematic variation, which contributes to the model that is not contained in the lipoprotein data.

Comparison of the Metabolic Consequences of SARS-CoV-2 Infection Measured in Two Independent Cohorts

A more sophisticated level of interlaboratory investigation concerns the direct comparison of the different population cohorts for biological similarity using identical analytical platforms (Figure 1B). We used OPLS-DA to build a training set model for the Spanish patient cohort analyzed at the ANPC

https://doi.org/10.1021/acs.jproteome.1c00458
J. Proteome Res. 2021, 20, 4139−4152

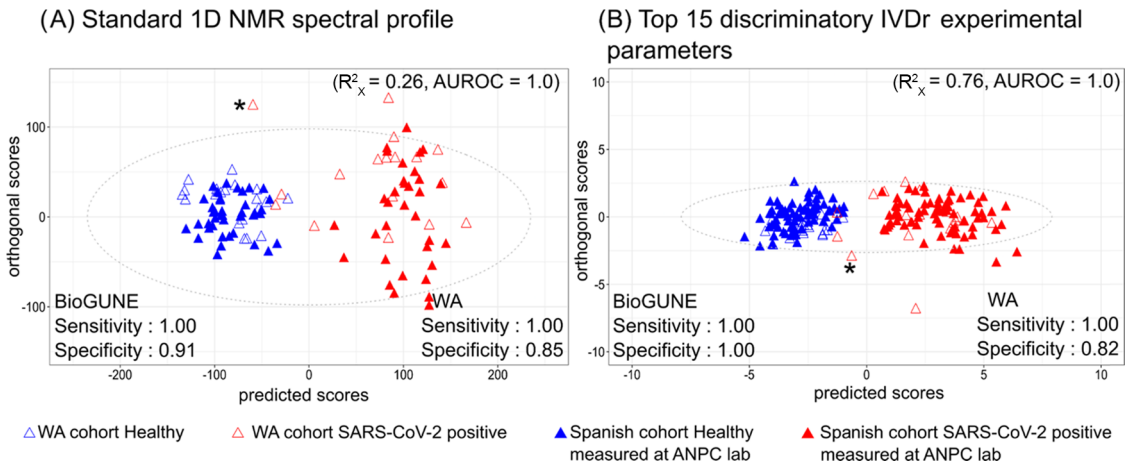


Figure 5. OPLS-DA training models generated using the healthy controls (blue closed triangles) and SARS-CoV-2 positive patients (red closed triangles) from the Spanish sample cohort analyzed in the ANPC lab are plotted for the Western Australian test set for the healthy (blue open triangles) and SARS-CoV-2 positive samples (red open triangles). (A) 1D ^{1}H NMR with the presaturation OPLS-DA Spanish cohort training model with the WA cohort projected in. (B) Ten most significant IVDr lipoprotein parameters combined with the five most significant IVDr low-MW metabolites obtained from modeling of the Spanish cohort with the Western Australian cohort projected as a validation set.

(healthy controls versus active COVID-19 cases) and then used the WA cohort as a completely independent test set to estimate classification accuracy. The OPLS-DA scores plots of the 1D NMR and the top 15 discriminatory IVDr parameter data sets (Figure 5A,B) show that all of the Western Australia healthy controls are correctly predicted into the healthy control model for the Spanish cohort. This projection of the basic measured WA NMR data (with no data alignment) showed extremely similar distributions of both SARS-CoV-2 infected groups. All but three of the WA SARS-CoV-2 positive controls are classified correctly in the 1D NMR spectral model. It should be noted that one of the three individuals that did not classify correctly were from a patient with a high BMI and diabetes (marked with an asterix). It can be clearly seen that these samples misclassify as healthy controls in the 1D, T_2-filtered, and lipoprotein models but not in the low-MW metabolite model (Figures 5 and S8). An analogous comparison of the IVDr data (Figure S8B,C) indicated that the models based on full spectral data were more predictive of the external validation cohort than the models based on quantified parameters. One of the reasons for this observation is that while both the lipoproteins and extracted low-MW metabolite are major contributors to the differential molecular signature of SARS-CoV-2 infection, other molecules that are not quantified by either of the IVDr methods also contributed highly in the model (Figure S8B,C) such as GlycA and GlycB, indices of inflammation that have been shown to be key "biomarkers" of the infection.[4,13]

The models relating to all spectral data sets and the IVDr-extracted parameters demonstrated high sensitivity (≥98% internal validation with the bioGUNE data sets and external WA data set), with the exception of the IVDr low-molecular-weight internal (bioGUNE) cross-validation data, which yielded a model with slightly lower sensitivity (95%). The specitivities of the models were slightly lower ranging from 85 to 97% for the internal cross-validation of the bioGUNE data sets. Cross-validation against the WA cohort caused a drop in sensitivity of 6.6 and 15%, respectively, for the 1D and T_2-filtered spectral data sets. Again the IVDr-extracted parameter models for the lipoprotein and low-MW metabolite sets, respectively.

The two cohorts investigated here, although both Caucasian, were from different ethnic backgrounds with differences in the diet and culture. The Spanish cohort is predominantly from the Basque region, which in itself is culturally and genetically different from the main Spanish population.[34] Although we do not have detailed genetic or dietary data for either of the cohorts, there is a reasonable expectation that there will be some systematic differences between the populations at the genetic and environmental/lifestyle levels.[35,36] Furthermore, there may also be differences in the range of severity of infection between the two patient groups, and so it is highly pertinent to question whether or not such diverse and largely uncontrolled groups can be cross-modeled in integrated data sets. Indeed, a PCA model comparing the control groups from the Spanish and WA cohorts showed inherent differentiation of the two populations with higher concentrations of serum triglycerides (data not shown) defining the WA cohort and higher concentrations of lactate and glycine being characteristic of the Spanish cohort (Figure S5). This reveals small but consistent metabolic variations between individuals from different countries and accounts for why the controls do not perfectly cross-model using the WA and Spanish cohorts, consistent with previous studies.[37−39]

The molecular signature of SARS-CoV-2 infection derived from standard 1D and T_2-filtered experiments for both the Spanish and WA cohorts was consistent with the previous literature reporting on the WA cohort (compare Lodge et al.[11] Figures 1 and 2 with Figure 2B,C). The lipoprotein and low-MW metabolite data for the WA cohort (Figures 3D,E and 4D,E, blue bars) ranked in similar order to those of the Spanish cohort denoted by the orange and green bars representing measurements made at the ANPC and bioGUNE laboratories, respectively. As expected, the statistical significance of the parameters in the WA cohort (Figures 3D and 4D blue bars) is lower than those derived from the Spanish cohort, due to the substantially smaller sample numbers in the WA cohort (n = 37 WA; n = 147 Spanish). Although a high degree of similarity in the ranked significance of the lipoprotein parameters was evident, some differences in the lipoprotein chemistries between the two cohorts were observed. For example, TPTG, V2TG, VLPL, VLTG, and VLFC showed good technical stability but were more influential in the Spanish cohort.

https://doi.org/10.1021/acs.jproteome.1c00458
J. Proteome Res. 2021, 20, 4139−4152

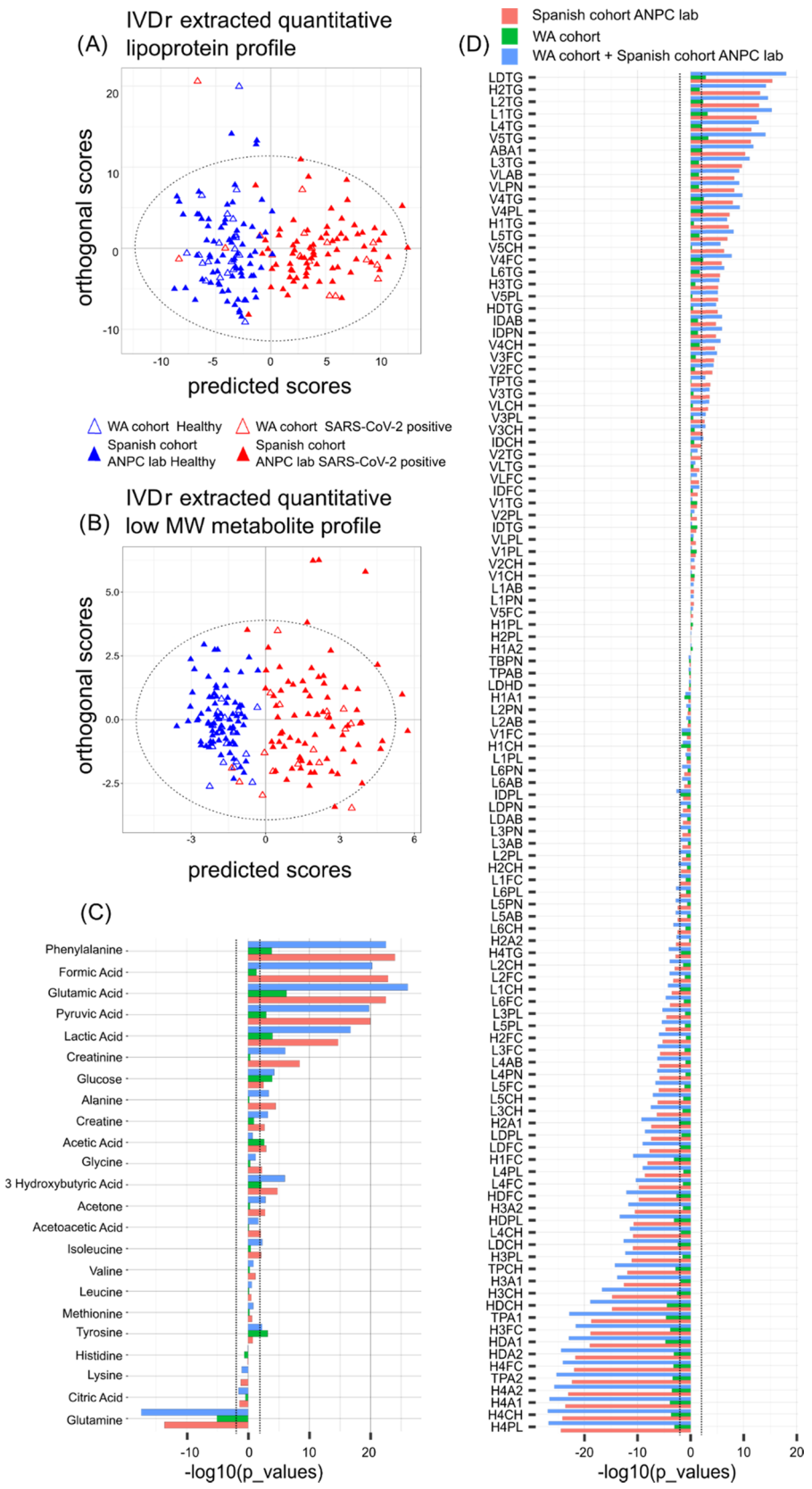


Figure 6. OPLS-DA scores and loadings plots of the comodels. (A) OPLS-DA scores plots of the Western Australian cohort comodeled with the Spanish sample cohort analyzed at the BioGUNE laboratory (AUROC = 0.98) and the (B) corresponding Cliff's delta eruption plot. (C) OPLS-DA scores plots of the Western Australian cohort comodeled with the Spanish sample cohort analyzed at the ANPC lab (AUROC = 0.97) and (D) corresponding eruption plot. The two gray vertical dashed lines delineate the confidence interval.

https://doi.org/10.1021/acs.jproteome.1c00458
J. Proteome Res. 2021, 20, 4139−4152

Comodeling Independent SARS-CoV-2 Disease Cohorts and Characterization of the Metabolic Effects of Infection

In addition to testing the interoperability between laboratories, the Western Australian sample cohort was comodeled (Figure 1C) with the Spanish cohort analyzed in both the CIC bioGUNE and the ANPC laboratories. The comparison and fusion of the data sets were enabled by normalization of the spectra by adjusting to the eretic factor of each sample.[30] The performance of all spectrometers differs slightly, but using IVDr methods and normalizing subsequent spectra using the eretic factor for each sample, the "batch" effect of the sample run can be eliminated.

The OPLS-DA models generated for the lipoproteins and metabolite IVDr data generated similar scores and eruption plots (Figure 6) to those generated for the individual cohorts (Figures 3 and 4). While the model quality for the combined cohort model and models for the individual cohorts were similar, the statistical power was increased. However, despite this higher statistical power with the combination of the cohorts, there was little change in either the eruption plots or the ranked loadings (Figure S9). The identities of the COVID-19 discriminating lipoproteins matched those reported previously. The most discriminatory low-MW metabolites across both populations were glutamate, lactate, pyruvate, formate, phenylalanine (higher concentrations associated with SARS-CoV-2), and glutamine (lower concentrations associated with SARS-CoV-2), of which glutamic acid and glutamine were the most significant across both the individual and the combined populations. For most metabolites, the significance of the variable was proportional to the number of samples used to generate the model.

Biological Significance of the Combined Data Sets

The Spanish cohort lipoprotein and metabolic data effectively confirm the findings of the earlier, smaller studies on the WA data alone.[4] Certain important metabolic features still stand out such as the glutamine:glutamic acid ratio and the high phenylalanine levels. Formic acid emerges as a strong biomarker for SARS-CoV-2 infection, and this may relate to the immunological stimulation of the tryptophan−kynurenine pathway, best measured by mass spectrometry.[5,12] Formic acid is a byproduct of the breakdown of *N*-formylkynurenine, which in turn, is formed from tryptophan via immune-stimulated indoleamine 2,3-dioxygenase activity.[40]

We observed significant shifts in the HDL/LDL profiles, as previously reported,[4] with significant depletion of multiple HDL parameters. This has also been observed by other scientists and indeed low HDL/LDL status has been reported to relate to disease severity.[41] In the present study, the control group was defined by higher concentrations of HDL cholesterol, total plasma lipoproteins A1 and A2, and HDL phospholipids. The SARS-CoV-2 group in both data sets was driven by high concentrations of L1−L4 triglycerides, VLDL phospholipids, and a high ratio of apolipoprotein B100 to apolipoprotein A1 (ABA1). The ABA1 ratio is elevated in type 1 diabetes[42] and is a marker of increased atherogenicity[43] and cardiovascular risk.[44] This observation is potentially of serious long-term concern as hundreds of millions of patients have been exposed to the virus and, therefore, have potentially enhanced disease cardiovascular risks associated with the elevated ABA1 ratio. In a recent study on a small cohort of mildly affected patients measured 3 months after the acute COVID-19 phase we found that the ABA1 had substantially normalized as part of the phenoreversion process.[18] However, these data cannot be extrapolated to patients who have experienced more severe acute episodes of the disease and this remains to be tested. We also previously noted that depletion of HDL apolipoproteins A1 and A2 and phospholipids is associated with carotid artery calcification and intima thickening.[45]

It is of particular note that the HDL fraction 4 apolipoprotein A1 and apolipoprotein A2 components, and cholesterol were the most severely depleted in the SARS-CoV-2 infected patients (Figure 3D) and were the most significant four biomarkers in the combined data set. We previously found that the HDL4 phospholipid component being highly correlated with a novel NMR-detected supramolecular phospholipid composite (SPC) peak was observable in the plasma diffusion and relaxation editing NMR spectra and this compartmental biomarker was profoundly depleted in SARS-CoV-2 infected patients.[13] Furthermore, reduction of HDL4 phospholipids apolipoprotein A1 and apolipoprotein A2 were the most significant changes observed in patients with pulmonary hypertension (PHT) and reduced levels of these components were significantly related to associated PHT mortality.[46] Detailed proteomic and network enrichment analysis revealed that HDL4 parameters relate to a series of proteins that regulate fibrinolysis and depletion of HDL4 results in negative regulation of blood coagulation, hemostasis, fibrinolysis, and thrombosis microembolism are also recognized side effects of SARS-CoV-2 pathology. Further insights into the mechanistic processes may be gained by more detailed lipoprotein to lipidomic cross-modeling in future studies.

CONCLUSIONS

We demonstrated that 1D ^{1}H NMR with presaturation, spin-echo, and quantitative IVDr lipoprotein NMR methods for lipoprotein and low-MW metabolite quantification all give substantially equivalent cross-validation models for SARS-CoV-2 infection biomarkers when samples from a population cohort are analyzed independently in two laboratories. Furthermore, OPLS-DA training set models of the Spanish cohort correctly predict the classification of independent population cohorts with a different ethnic and lifestyle background. Further work is required to assess the general interethnic variation in biomarkers, but these data indicate that the metabolic phenoconversion responses caused by SARS-CoV-2 infection greatly outweigh any differences in base population metabolism and metabolic phenotypes and that comodeling studies lead to deeper mechanistic insights. By combining the two data sets some biomarker combinations emerged more strongly as discriminatory, including several HDL parameters. In particular HDL subclass 4 phospholipids, apolipoprotein A1 and apolipoprotein A2 were reduced, which has previously been associated with a range of cardiovascular risk factors and PHT mortality, and may have mechanistic significance with respect to blood haemodynamic and coagulation changes found in SARS-CoV-2 infections.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jproteome.1c00458.

Table S1: Full cohort demographic data for the Spanish sample cohort; healthy control samples and SARS-CoV-2 positive patients; Table S2: full cohort demographic data

https://doi.org/10.1021/acs.jproteome.1c00458
J. Proteome Res. 2021, 20, 4139−4152

for the Western Australian sample cohort; healthy control samples and SARS-CoV-2 positive patients; Table S3: symptom presentation in Western Australian sample cohort SARS-CoV-2 positive patients; Table S4: annotation of the keys used by the Bruker IVDr lipoprotein subclass analysis (B.I.-LISA) method; Table S5: lipoprotein parameters ranked in the order of Cliff's delta for the Spanish samples run in Spain; Table S6: lipoprotein parameters ranked in the order of Cliff's delta for the Spanish samples run at the ANPC; Table S7: lipoprotein parameters ranked in the order of Cliff's delta for the Australia samples run at the ANPC; Table S8: lipoprotein parameters ranked in the order of Cliff's delta for the combined Spain run samples and the WA cohort; Table S9: lipoprotein parameters ranked in the order of Cliff's delta for the combined ANPC run samples and the WA cohort; Table S10: IVDr low-MW metabolite quantification ranked in the order of Cliff's delta for the Spanish samples run in Spain; Table S11: IVDr low-MW metabolite quantification ranked in the order of Cliff's delta for the Spanish samples run in the ANPC; Table S12: IVDr low-MW metabolite quantification ranked in the order of Cliff's delta for the Western Australian cohort run at the ANPC; Figure S1: principal components analysis and loadings plots of the lipoproteins of the Spanish cohort; Figure S2: principal component modeling of lipoproteins for healthy control groups of both Spanish and WA cohorts; Figure S3: principal component modeling of lipoproteins for SARS-CoV-2 positive groups of both Spanish and WA cohorts; Figure S4: principal component analysis of the IVDr low-MW metabolite quantification of the Spanish cohort of samples; Figure S5: principal component modeling of low-MW metabolites for healthy control groups of both Spanish and WA cohorts; Figure S6: principal component modeling of low-MW metabolites for SARS-CoV-2 positive groups of both Spanish and WA cohorts; Figure S7: Spanish Cohort modeled using standard 1D with presaturation experiment. Comparison between the measurements obtained at the BioGUNE and ANPC laboratories; Figure S8: OPLS-DA models of a training set from the Spanish cohort and projections of the validation set from both the Spanish cohort and WA cohort, for A) spin-echo spectra, B) IVDr-extracted lipoproteins, and C) IVDr-extracted low-MW metabolites; and Figure S9: Lipoprotein OPLS-DA scores and loadings plots of the comodels and the corresponding eruption plot (PDF)

AUTHOR INFORMATION

Corresponding Authors

Julien Wist − *Australian National Phenome Centre, Health Futures Institute, Murdoch University, Perth, WA 6150, Australia; Centre for Computational and Systems Medicine, Health Futures Institute, Murdoch University, Perth, WA 6150, Australia; Chemistry Department, Universidad del Valle, 76001 Cali, Colombia;* orcid.org/0000-0002-3416-2572; Email: julien.wist@murdoch.edu.au

Oscar Millet − *CIC bioGUNE, Asociación Centro de Investigación Cooperativa en Biociencias, 48160 Derio, Bizkaia, Spain;* orcid.org/0000-0001-8748-4105; Email: omillet@cicbiogune.es

Jeremy K. Nicholson − *Australian National Phenome Centre, Health Futures Institute, Murdoch University, Perth, WA 6150, Australia; Centre for Computational and Systems Medicine, Health Futures Institute, Murdoch University, Perth, WA 6150, Australia; Institute of Global Health Innovation, Imperial College London, London SW7 2NA, U.K.;* orcid.org/0000-0002-8123-8349; Email: Jeremy.nicholson@murdoch.edu.au, j.nicholson@imperial.ac.uk

Authors

Reika Masuda − *Australian National Phenome Centre, Health Futures Institute, Murdoch University, Perth, WA 6150, Australia; Centre for Computational and Systems Medicine, Health Futures Institute, Murdoch University, Perth, WA 6150, Australia*

Samantha Lodge − *Australian National Phenome Centre, Health Futures Institute, Murdoch University, Perth, WA 6150, Australia; Centre for Computational and Systems Medicine, Health Futures Institute, Murdoch University, Perth, WA 6150, Australia;* orcid.org/0000-0001-9193-0462

Philipp Nitschke − *Australian National Phenome Centre, Health Futures Institute, Murdoch University, Perth, WA 6150, Australia*

Manfred Spraul − *Bruker Biospin GmbH, Ettlingen 76275, Germany*

Hartmut Schaefer − *Bruker Biospin GmbH, Ettlingen 76275, Germany*

Sze-How Bong − *Australian National Phenome Centre, Health Futures Institute, Murdoch University, Perth, WA 6150, Australia;* orcid.org/0000-0002-3313-5097

Torben Kimhofer − *Australian National Phenome Centre, Health Futures Institute, Murdoch University, Perth, WA 6150, Australia; Centre for Computational and Systems Medicine, Health Futures Institute, Murdoch University, Perth, WA 6150, Australia;* orcid.org/0000-0001-7158-9930

Drew Hall − *Australian National Phenome Centre, Health Futures Institute, Murdoch University, Perth, WA 6150, Australia*

Ruey Leng Loo − *Australian National Phenome Centre, Health Futures Institute, Murdoch University, Perth, WA 6150, Australia; Centre for Computational and Systems Medicine, Health Futures Institute, Murdoch University, Perth, WA 6150, Australia;* orcid.org/0000-0001-5307-5709

Maider Bizkarguenaga − *CIC bioGUNE, Asociación Centro de Investigación Cooperativa en Biociencias, 48160 Derio, Bizkaia, Spain*

Chiara Bruzzone − *CIC bioGUNE, Asociación Centro de Investigación Cooperativa en Biociencias, 48160 Derio, Bizkaia, Spain;* orcid.org/0000-0003-4252-8180

Rubén Gil-Redondo − *CIC bioGUNE, Asociación Centro de Investigación Cooperativa en Biociencias, 48160 Derio, Bizkaia, Spain*

Nieves Embade − *CIC bioGUNE, Asociación Centro de Investigación Cooperativa en Biociencias, 48160 Derio, Bizkaia, Spain;* orcid.org/0000-0001-9878-3290

José M. Mato − *CIC bioGUNE, Asociación Centro de Investigación Cooperativa en Biociencias, 48160 Derio, Bizkaia, Spain*

Elaine Holmes − *Australian National Phenome Centre, Health Futures Institute, Murdoch University, Perth, WA 6150, Australia; Centre for Computational and Systems Medicine, Health Futures Institute, Murdoch University, Perth, WA*

https://doi.org/10.1021/acs.jproteome.1c00458
J. Proteome Res. 2021, 20, 4139−4152

6150, Australia; Section for Nutrition Research, Department of Metabolism, Nutrition and Reproduction, Faculty of Medicine, Imperial College London, London SW7 2AZ, U.K.;
orcid.org/0000-0002-0556-8389

Complete contact information is available at:
https://pubs.acs.org/10.1021/acs.jproteome.1c00458

Author Contributions

○R.M., S.L., and P.N. contributed equally.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors thank the support provided by the Department of Industry, Tourism, and Trade of the Government of the Autonomous Community of the Basque Country (Elkartek BG2019) and the Severo Ochoa Excellence A. The authors thank the Spinnaker Health Research Foundation, WA., the McCusker Foundation, WA, the Western Australian State Government, and the MRFF for funding the Australian National Phenome Centre for this and related work. The authors thank the UK MRC for funding (S.-H.B.), and the Department of Jobs, Tourism, Science, and Innovation, Government of Western Australian Premier's Fellowship for funding R.L.L. and E.H.; and ARC Laureate Fellowship funding for E.H. The authors would also like to acknowledge the Western Australian Covid Research Response team (https://research-au.net/covid-research-response/), Giuliana D'Aulerio, Kelly Beer, Rolee Kumar, Doug Robb, Joseph Miocevich, Dominic Mallonic, Michael Epis, Merrilee Needham, Daniel Fatovich, Aron Chakera, Thomas Gilbert, Nathanael Foo, @STRIVE WA, Candice Peel, Sheeraz Mohd, and Ali Alishum for the coordination, sampling, and biobanking of patient samples and clinical metadata; and Anuradha Sooda and Emily Thomas for their technical assistance.

■ REFERENCES

(1) Rubino, F.; Amiel, S. A.; Zimmet, P.; Alberti, G.; Bornstein, S.; Eckel, R. H.; Mingrone, G.; Boehm, B.; Cooper, M. E.; Chai, Z.; Del Prato, S.; Ji, L.; Hopkins, D.; Herman, W. H.; Khunti, K.; Mbanya, J. C.; Renard, E. New-Onset Diabetes in Covid-19. *N. Engl. J. Med.* **2020**, *383*, 789−790.

(2) Zhang, C.; Shi, L.; Wang, F. S. Liver Injury in COVID-19: Management and Challenges. *Lancet Gastroenterol. Hepatol.* **2020**, *5*, 428−430.

(3) Ellul, M. A.; Benjamin, L.; Singh, B.; Lant, S.; Michael, B. D.; Easton, A.; Kneen, R.; Defres, S.; Sejvar, J.; Solomon, T. Neurological Associations of COVID-19. *Lancet Neurol.* **2020**, *19*, 767−783.

(4) Kimhofer, T.; Lodge, S.; Whiley, L.; Gray, N.; Loo, R. L.; Lawler, N. G.; Nitschke, P.; Bong, S. H.; Morrison, D. L.; Begum, S.; Richards, T.; Yeap, B. B.; Smith, C.; Smith, K. C. G.; Holmes, E.; Nicholson, J. K. Integrative Modelling of Quantitative Plasma Lipoprotein, Metabolic and Amino Acid Data Reveals a Multi-Organ Pathological Signature of SARS-CoV-2 Infection. *J. Proteome Res.* **2020**, *19*, 4442−4454.

(5) Lawler, N. G.; Gray, N.; Kimhofer, T.; Boughton, B.; Gay, M.; Yang, R.; Morillon, A.-C.; Chin, S.-T.; Ryan, M.; Begum, S.; Bong, S. H.; Coudert, J. D.; Edgar, D.; Raby, E.; Pettersson, S.; Richards, T.; Holmes, E.; Whiley, L.; Nicholson, J. K. Systemic Perturbations in Amine and Kynurenine Metabolism Associated with Acute SARS-CoV-2 Infection and Inflammatory Cytokine Responses. *J. Proteome Res.* **2021**, *20*, 2796−2811.

(6) Shah, R. R.; Smith, R. L. Inflammation-Induced Phenoconversion of Polymorphic Drug Metabolizing Enzymes: Hypothesis with Implications for Personalized Medicine. *Drug Metab. Dispos.* **2015**, *43*, 400−410.

(7) Nicholson, J. K.; Holmes, E.; Kinross, J. M.; Darzi, A. W.; Takats, Z.; Lindon, J. C. Metabolic Phenotyping in Clinical and Surgical Environments. *Nature* **2012**, *491*, 384−392.

(8) Nicholson, J. K.; Buckingham, M. J.; Sadler, P. J. High Resolution 1H N.m.r. Studies of Vertebrate Blood and Plasma. *Biochem. J.* **1983**, *211*, 605−615.

(9) Otvos, J. D.; Jeyarajah, E. J.; Bennett, D. W. Quantification of Plasma Lipoproteins by Proton Nuclear Magnetic Resonance Spectroscopy. *Clin. Chem.* **1991**, *37*, 377−386.

(10) Wolak-Dinsmore, J.; Gruppen, E. G.; Shalaurova, I.; Matyus, S. P.; Grant, R. P.; Gegen, R.; Bakker, S. J. L.; Otvos, J. D.; Connelly, M. A.; Dullaart, R. P. F. A Novel NMR-Based Assay to Measure Circulating Concentrations of Branched-Chain Amino Acids: Elevation in Subjects with Type 2 Diabetes Mellitus and Association with Carotid Intima Media Thickness. *Clin. Biochem.* **2018**, *54*, 92−99.

(11) Lodge, S.; Nitschke, P.; Kimhofer, T.; Coudert, J. D.; Begum, S.; Bong, S.-H.; Richards, T.; Edgar, D.; Raby, E.; Spraul, M.; Schaefer, H.; Lindon, J. C.; Loo, R. L.; Holmes, E.; Nicholson, J. K. NMR Spectroscopic Windows on the Systemic Effects of SARS-CoV-2 Infection on Plasma Lipoproteins and Metabolites in Relation to Circulating Cytokines. *J. Proteome Res.* **2021**, *20*, 1382−1396.

(12) Gray, N.; Lawler, N. G.; Yang, R.; Morillon, A.-C.; Gay, M. C. L.; Bong, S.-H.; Holmes, E.; Nicholson, J. K.; Whiley, L. A Simultaneous Exploratory and Quantitative Amino Acid and Biogenic Amine Metabolic Profiling Platform for Rapid Disease Phenotyping via UPLC-QToF-MS. *Talanta* **2021**, *223*, No. 121872.

(13) Lodge, S.; Nitschke, P.; Kimhofer, T.; Wist, J.; Bong, S.-H.; Loo, R. L.; Masuda, R.; Begum, S.; Richards, T.; Lindon, J. C.; Bermel, W.; Reinsperger, T.; Schaefer, H.; Spraul, M.; Holmes, E.; Nicholson, J. K. Diffusion and Relaxation Edited Proton NMR Spectroscopy of Plasma Reveals a High-Fidelity Supramolecular Biomarker Signature of SARS-CoV-2 Infection. *Anal. Chem.* **2021**, *93*, 3976−3986.

(14) Bruzzone, C.; Bizkarguenaga, M.; Gil-Redondo, R.; Diercks, T.; Arana, E.; García de Vicuña, A.; Seco, M.; Bosch, A.; Palazón, A.; San Juan, I.; Laín, A.; Gil-Martínez, J.; Bernardo-Seisdedos, G.; Fernández-Ramos, D.; Lopitz-Otsoa, F.; Embade, N.; Lu, S.; Mato, J. M.; Millet, O. SARS-CoV-2 Infection Dysregulates the Metabolomic and Lipidomic Profiles of Serum. *iScience* **2020**, *23*, No. 101645.

(15) Schmelter, F.; Foeh, B.; Mallagaray, A.; Rahmoeller, J.; Ehlers, M.; Lehrian, S.; Kopylow, V.; von Künsting, I.; Lixenfeld, A. S.; Martin, E.; Ragab, M.; Borsche, M.; Balck, A.; Vollstedt, E. J.; Meyer-Saraei, R.; Kreutzmann, F.; Eitel, I.; Taube, S.; Klein, C.; Katalinic, A.; Rupp, J.; Jantzen, E.; Graf, T.; Sina, C.; Günther, U. L. Metabolic Markers Distinguish COVID-19 from Other Intensive Care Patients and Show Potential to Stratify for Disease Risk, 2021. DOI: 10.1101/2021.01.13.21249645.

(16) Dos Santos Junior, G. C.; Pereira, C. M.; da Silva Fidalgo, T. K.; Valente, A. P. Saliva NMR-Based Metabolomics in the War Against COVID-19. *Anal. Chem.* **2020**, *92*, 15688−15692.

(17) Nalbandian, A.; Sehgal, K.; Gupta, A.; Madhavan, M. V.; McGroder, C.; Stevens, J. S.; Cook, J. R.; Nordvig, A. S.; Shalev, D.; Sehrawat, T. S.; Ahluwalia, N.; Bikdeli, B.; Dietz, D.; Der-Nigoghossian, C.; Liyanage-Don, N.; Rosner, G. F.; Bernstein, E. J.; Mohan, S.; Beckley, A. A.; Seres, D. S.; Choueiri, T. K.; Uriel, N.; Ausiello, J. C.; Accili, D.; Freedberg, D. E.; Baldwin, M.; Schwartz, A.; Brodie, D.; Garcia, C. K.; Elkind, M. S. V.; Connors, J. M.; Bilezikian, J. P.; Landry, D. W.; Wan, E. Y. Post-Acute COVID-19 Syndrome. *Nat. Med.* **2021**, *27*, 601−615.

(18) Holmes, E.; Wist, J.; Masuda, R.; Lodge, S.; Nitschke, P.; Kimhofer, T.; Loo, R. L.; Begum, S.; Boughton, B.; Yang, R.; Morillon, A.-C.; Chin, S.-T.; Hall, D.; Ryan, M.; Bong, S.-H.; Gay, M.; Edgar, D. W.; Lindon, J. C.; Richards, T.; Yeap, B. B.; Pettersson, S.; Spraul, M.; Schaefer, H.; Lawler, N. G.; Gray, N.; Whiley, L.; Nicholson, J. K. Incomplete Systemic Recovery and Metabolic Phenoreversion in Post-Acute-Phase Nonhospitalized COVID-19 Patients: Implications for Assessment of Post-Acute COVID-19 Syndrome. *J. Proteome Res.* **2021**, *20*, 3315−3329.

(19) Jiménez, B.; Holmes, E.; Heude, C.; Tolson, R. F.; Harvey, N.; Lodge, S. L.; Chetwynd, A. J.; Cannet, C.; Fang, F.; Pearce, J. T. M.; Lewis, M. R.; Viant, M. R.; Lindon, J. C.; Spraul, M.; Schafer, H.; Nicholson, J. K. Quantitative Lipoprotein Subclass and Low Molecular Weight Metabolite Analysis in Human Serum and Plasma by (1)H NMR Spectroscopy in a Multilaboratory Trial. *Anal. Chem.* **2018**, *90*, 11962−11971.
(20) Monsonis Centelles, S.; Hoefsloot, H. C. J.; Khakimov, B.; Ebrahimi, P.; Lind, M. V.; Kristensen, M.; de Roo, N.; Jacobs, D. M.; van Duynhoven, J.; Cannet, C.; Fang, F.; Humpfer, E.; Schafer, H.; Spraul, M.; Engelsen, S. B.; Smilde, A. K. Toward Reliable Lipoprotein Particle Predictions from NMR Spectra of Human Blood: An Interlaboratory Ring Test. *Anal. Chem.* **2017**, *89*, 8004−8012.
(21) Martin, J.-C.; Maillot, M.; Mazerolles, G.; Verdu, A.; Lyan, B.; Migné, C.; Defoort, C.; Canlet, C.; Junot, C.; Guillou, C.; Manach, C.; Jabob, D.; Bouveresse, D. J.-R.; Paris, E.; Pujos-Guillot, E.; Jourdan, F.; Giacomoni, F.; Courant, F.; Favé, G.; Le Gall, G.; Chassaigne, H.; Tabet, J.-C.; Martin, J.-F.; Antignac, J.-P.; Shintu, L.; Defernez, M.; Philo, M.; Alexandre-Gouaubau, M.-C.; Amiot-Carlin, M.-J.; Bossis, M.; Triba, M. N.; Stojilkovic, N.; Banzet, N.; Molinié, R.; Bott, R.; Goulitquer, S.; Caldarelli, S.; Rutledge, D. N. Can We Trust Untargeted Metabolomics? Results of the Metabo-Ring Initiative, a Large-Scale, Multi-Instrument Inter-Laboratory Study. *Metabolomics* **2015**, *11*, 807−821.
(22) Chen, Y.-M.; Zheng, Y.; Yu, Y.; Wang, Y.; Huang, Q.; Qian, F.; Sun, L.; Song, Z.-G.; Chen, Z.; Feng, J.; An, Y.; Yang, J.; Su, Z.; Sun, S.; Dai, F.; Chen, Q.; Lu, Q.; Li, P.; Ling, Y.; Yang, Z.; Tang, H.; Shi, L.; Jin, L.; Holmes, E. C.; Ding, C.; Zhu, T.-Y.; Zhang, Y.-Z. Blood Molecular Markers Associated with COVID-19 Immunopathology and Multi-Organ Damage. *EMBO J.* **2020**, *39*, No. e105896.
(23) Meoni, G.; Ghini, V.; Maggi, L.; Vignoli, A.; Mazzoni, A.; Salvati, L.; Capone, M.; Vanni, A.; Tenori, L.; Fontanari, P.; Lavorini, F.; Peris, A.; Bartoloni, A.; Liotta, F.; Cosmi, L.; Luchinat, C.; Annunziato, F.; Turano, P. Metabolomic/lipidomic Profiling of COVID-19 and Individual Response to Tocilizumab. *PLoS Pathog.* **2021**, *17*, No. e1009243.
(24) Lodge, S.; Nitschke, P.; Loo, R. L.; Kimhofer, T.; Bong, S.-H.; Richards, T.; Begum, S.; Spraul, M.; Schaefer, H.; Lindon, J. C.; Holmes, E.; Nicholson, J. K. Low Volume in Vitro Diagnostic Proton NMR Spectroscopy of Human Blood Plasma for Lipoprotein and Metabolite Analysis: Application to SARS-CoV-2 Biomarkers. *J. Proteome Res.* **2021**, *20*, 1415−1423.
(25) Holmes, E.; Loo, R. L.; Stamler, J.; Bictash, M.; Yap, I. K. S.; Chan, Q.; Ebbels, T.; De Iorio, M.; Brown, I. J.; Veselkov, K. A.; Daviglus, M. L.; Kesteloot, H.; Ueshima, H.; Zhao, L.; Nicholson, J. K.; Elliott, P. Human Metabolic Phenotype Diversity and Its Association with Diet and Blood Pressure. *Nature* **2008**, *453*, 396−400.
(26) Holmes, E.; Wilson, I. D.; Nicholson, J. K. Metabolic Phenotyping in Health and Disease. *Cell* **2008**, *134*, 714−717.
(27) Sagner, M.; McNeil, A.; Puska, P.; Auffray, C.; Price, N. D.; Hood, L.; Lavie, C. J.; Han, Z.-G.; Chen, Z.; Brahmachari, S. K.; McEwen, B. S.; Soares, M. B.; Balling, R.; Epel, E.; Arena, R. The P4 Health Spectrum - A Predictive, Preventive, Personalized and Participatory Continuum for Promoting Healthspan. *Prog. Cardiovasc. Dis.* **2017**, *59*, 506−521.
(28) Loo, R. L.; Lodge, S.; Kimhofer, T.; Bong, S. H.; Begum, S.; Whiley, L.; Gray, N.; Lindon, J. C.; Nitschke, P.; Lawler, N. G.; Schafer, H.; Spraul, M.; Richards, T.; Nicholson, J. K.; Holmes, E. Quantitative In-Vitro Diagnostic NMR Spectroscopy for Lipoprotein and Metabolite Measurements in Plasma and Serum: Recommendations for Analytical Artefact Minimization with Special Reference to COVID-19/SARS-CoV-2 Samples. *J. Proteome Res.* **2020**, *19*, 4428−4441.
(29) Dona, A. C.; Jimenez, B.; Schafer, H.; Humpfer, E.; Spraul, M.; Lewis, M. R.; Pearce, J. T.; Holmes, E.; Lindon, J. C.; Nicholson, J. K. Precision High-Throughput Proton NMR Spectroscopy of Human Urine, Serum, and Plasma for Large-Scale Metabolic Phenotyping. *Anal. Chem.* **2014**, *86*, 9887−9894.
(30) Akoka, S.; Barantin, L.; Trierweiler, M. Concentration Measurement by Proton NMR Using the ERETIC Method. *Anal. Chem.* **1999**, *71*, 2554−2557.
(31) Dieterle, F.; Ross, A.; Schlotterbeck, G.; Senn, H. Probabilistic Quotient Normalization as Robust Method to Account for Dilution of Complex Biological Mixtures. Application in 1H NMR Metabonomics. *Anal. Chem.* **2006**, *78*, 4281−4290.
(32) Bylesjö, M.; Rantalainen, M.; Cloarec, O.; Nicholson, J. K.; Holmes, E.; Trygg, J. OPLS Discriminant Analysis: Combining the Strengths of PLS-DA and SIMCA Classification. *J. Chemom.* **2006**, *20*, 341−351.
(33) Wang, F.; Debik, J.; Andreassen, T.; Euceda, L. R.; Haukaas, T. H.; Cannet, C.; Schäfer, H.; Bathen, T. F.; Giskeødegård, G. F. Effect of Repeated Freeze-Thaw Cycles on NMR-Measured Lipoproteins and Metabolites in Biofluids. *J. Proteome Res.* **2019**, *18*, 3681−3688.
(34) Flores-Bello, A.; Bauduer, F.; Salaberria, J.; Oyharçabal, B.; Calafell, F.; Bertranpetit, J.; Quintana-Murci, L.; Comas, D. Genetic Origins, Singularity, and Heterogeneity of Basques. *Curr. Biol.* **2021**, *31*, 2167−2177.
(35) Garagnani, P.; Laayouni, H.; González-Neira, A.; Sikora, M.; Luiselli, D.; Bertranpetit, J.; Calafell, F. Isolated Populations as Treasure Troves in Genetic Epidemiology: The Case of the Basques. *Eur. J. Hum. Genet.* **2009**, *17*, 1490−1494.
(36) Veeramah, K. R.; Novembre, J. Demographic Events and Evolutionary Forces Shaping European Genetic Diversity. *Cold Spring Harbor Perspect. Biol.* **2014**, *6*, No. a008516.
(37) Swann, J. R.; Spagou, K.; Lewis, M.; Nicholson, J. K.; Glei, D. A.; Seeman, T. E.; Coe, C. L.; Goldman, N.; Ryff, C. D.; Weinstein, M.; Holmes, E. Microbial-Mammalian Cometabolites Dominate the Age-Associated Urinary Metabolic Phenotype in Taiwanese and American Populations. *J. Proteome Res.* **2013**, *12*, 3166−3180.
(38) Lenz, E. M.; Bright, J.; Wilson, I. D.; Hughes, A.; Morrisson, J.; Lindberg, H.; Lockton, A. Metabonomics, Dietary Influences and Cultural Differences: A 1H NMR-Based Study of Urine Samples Obtained from Healthy British and Swedish Subjects. *J. Pharm. Biomed. Anal.* **2004**, *36*, 841−849.
(39) Yap, I. K. S.; Brown, I. J.; Chan, Q.; Wijeyesekera, A.; Garcia-Perez, I.; Bictash, M.; Loo, R. L.; Chadeau-Hyam, M.; Ebbels, T.; De Iorio, M.; Maibaum, E.; Zhao, L.; Kesteloot, H.; Daviglus, M. L.; Stamler, J.; Nicholson, J. K.; Elliott, P.; Holmes, E. Metabolome-Wide Association Study Identifies Multiple Biomarkers That Discriminate North and South Chinese Populations at Differing Risks of Cardiovascular Disease: INTERMAP Study. *J. Proteome Res.* **2010**, *9*, 6647−6654.
(40) Wang, Q.; Liu, D.; Song, P.; Zou, M.-H. Tryptophan-Kynurenine Pathway Is Dysregulated in Inflammation, and Immune Activation. *Front. Biosci.* **2015**, *20*, 1116−1143.
(41) Masana, L.; Correig, E.; Ibarretxe, D.; Anoro, E.; Arroyo, J. A.; Jericó, C.; Guerrero, C.; Miret, M.; Näf, S.; Pardo, A.; Perea, V.; Pérez-Bernalte, R.; Plana, N.; Ramírez-Montesinos, R.; Royuela, M.; Soler, C.; Urquizu-Padilla, M.; Zamora, A.; Pedro-Botet, J.; STACOV-XULA Research Group. Low HDL and High Triglycerides Predict COVID-19 Severity. *Sci. Rep.* **2021**, *11*, No. 7217.
(42) Llauradó, G.; Amigó, N.; Cano, A.; Ballesta, S.; Albert, L.; Mazarico, I.; Fernández-Veledo, S.; Pedro-Botet, J.; Vendrell, J.; González-Clemente, J.-M. Specific Nuclear Magnetic Resonance Lipoprotein Subclass Profiles and Central Arterial Stiffness in Type 1 Diabetes Mellitus: A Case Control Study. *J. Clin. Med.* **2019**, *8*, No. 1875.
(43) Kaneva, A. M.; Potolitsyna, N. N.; Bojko, E. R.; Odland, J. Ø. The Apolipoprotein B/apolipoprotein A-I Ratio as a Potential Marker of Plasma Atherogenicity. *Dis. Markers* **2015**, *2015*, No. 591454.
(44) Walldius, G. The apoB/apoA-I Ratio Is a Strong Predictor of Cardiovascular Risk. In *Lipoproteins*; Frank, S.; Kostner, G., Eds.; IntechOpen: Rijeka, 2012.
(45) Tzoulaki, I.; Castagne, R.; Boulange, C. L.; Karaman, I.; Chekmeneva, E.; Evangelou, E.; Ebbels, T. M. D.; Kaluarachchi, M. R.; Chadeau-Hyam, M.; Mosen, D.; Dehghan, A.; Moayyeri, A.; Ferreira, D. L. S.; Guo, X.; Rotter, J. I.; Taylor, K. D.; Kavousi, M.; de Vries, P. S.;

https://doi.org/10.1021/acs.jproteome.1c00458
J. Proteome Res. 2021, 20, 4139−4152

Lehne, B.; Loh, M.; Hofman, A.; Nicholson, J. K.; Chambers, J.; Gieger, C.; Holmes, E.; Tracy, R.; Kooner, J.; Greenland, P.; Franco, O. H.; Herrington, D.; Lindon, J. C.; Elliott, P. Serum Metabolic Signatures of Coronary and Carotid Atherosclerosis and Subsequent Cardiovascular Disease. *Eur. Heart J.* **2019**, *40*, 2883−2896.

(46) Harbaum, L.; Ghataorhe, P.; Wharton, J.; Jiménez, B.; Howard, L. S. G.; Gibbs, J. S. R.; Nicholson, J. K.; Rhodes, C. J.; Wilkins, M. R. Reduced Plasma Levels of Small HDL Particles Transporting Fibrinolytic Proteins in Pulmonary Arterial Hypertension. *Thorax* **2019**, *74*, 380−389.

https://doi.org/10.1021/acs.jproteome.1c00458
J. Proteome Res. 2021, 20, 4139−4152