Data in Brief

Abstract

Elucidation of molecular markers related to the mounted immune response is crucial for understanding the disease pathogenesis. In this article, we present the mass-spectrometry-based metabolomic and proteomic data of blood plasma of COVID-19 patients collected at two-time points, which showed a transition from non-severe to severe conditions during these time points. Metabolites were extracted and subjected to mass spectrometric analysis using the Q-Exactive mass spectrometer. For proteomic analysis, depleted plasma samples were tryptic digested and subjected to mass spectrometry analysis. The expression of a few significant proteins was also validated by employing the targeted proteomic approach of multiple reaction monitoring (MRM). Integrative pathway analysis was performed with the significant proteins to obtain biological insights into disease severity. For discussion and more information on the dataset creation, please refer to the related full-length article (Suvarna et al., 2021).

© 2022 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)

Keywords:
Proteomics
Metabolomics
Mass spectrometry
Workflow
Longitudinal
Plasma
COVID-19

Received 6 August 2022
Revised 31 October 2022
Accepted 16 November 2022
Available online 22 November 2022

Dataset link: A longitudinal study of COVID-19 patients using proteomics approach (Reference data)
Dataset link: A Multi-omics Longitudinal Study Reveals Alteration of the Leukocyte Activation Pathway in COVID-19 Patients (Reference data)
Dataset link: Validation of proteins showing significant dysregulation from shotgun proteomics experiments in longitudinal plasma samples of COVID-19 patients (Reference data)

ARTICLE INFO

Multiomics data analysis workflow to assess severity in longitudinal plasma samples of COVID-19 patients

Sakshi Rajoria1, Mehar Un Nissa1, Kruthi Suvarna, Harsh Khatri, Sanjeeva Srivastava*

Department of Biosciences and Bioengineering, Indian Institute of Technology Bombay, Mumbai 400076, India

Corresponding author.
E-mail address: sanjeeva@iitb.ac.in (S. Srivastava).

1 Equal Contribution.

https://doi.org/10.1016/j.dib.2022.108765

2352-3409/© 2022 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)
Specifications Table

| Subject | Biology, Clinical Proteomics |
|---------|-----------------------------|
| Specific subject area | Plasma proteomics |
| Type of data | Table, Figure |
| How the data were acquired | Data was acquired via liquid chromatography-tandem mass spectrometry (LC-MS/MS) from Orbitrap Fusion Tribrid mass spectrometer connected to an Easy-nLC 1200 nano-flow liquid chromatography system. A TSQ Altis Triple Quadrupole Mass spectrometer (Thermo) coupled to an HPLC system (Thermo Vanquish) was used for obtaining the SRM/MRM data. |
| Data format | Compound Discoverer (v 3.0) and MaxQuant (v1.6.6.0) were used to analyze the data acquired from mass spectrometer while metaboanalyst was majorly used for statistical data analysis. |
| Description of data collection | The data included plasma samples from 13 RT-PCR tested COVID-19 positive patients, collected at two time points. For metabolomic data, extracted metabolites were run in Q-Exactive mass spectrometer. For proteomic data, depleted plasma samples were processed for mass spectrometric data acquisition using discovery and targeted workflow. |
| Data source location | Institution: Indian Institute of Technology Bombay. City: Mumbai - 400076 Country: India. |
| Data accessibility | Metabolomics data availability: Repository name: MetaboLights Data identification number: MTBLS2469 Direct URL to data: Proteomics data availability (Discovery based proteomics) Repository name: Proteome Xchange Consortium via the PRIDE partner repository Data identification number: PXD023521 Direct URL to data: https://www.ebi.ac.uk/pride/archive/projects/PXD023521 Proteomics data availability (Targeted proteomics) Repository name: Panorma Data identification number: PXD026848 Direct URL to data: https://panoramaweb.org/longitudinal_plasma_MRM.url |
| Related research article | Suvarna, K., Salkar, A., Palanivel, V., Bankar, R., Banerjee, N., Gayathri J Pai, M., Srivastava, A., Singh, A., Khatri, H., Agrawal, S. and Shrivastav, O., 2021. A multi-omics longitudinal study reveals alteration of the leukocyte activation pathway in COVID-19 patients. Journal of Proteome Research, 20(10), pp.4667-4680. |

Value of the Data

- This data belongs to the first longitudinal study of COVID-19 patients from the Indian population that will help in understanding the features of a patient’s immune response to infection.
- This study provides an extensive data for the altered proteins and metabolites in plasma of COVID-19 patients showing a transition from non-severe to severe condition.
- This data highlights the role of leukocyte activation and arginine metabolism in the progression of COVID-19 disease.
- This data also provides putative metabolic pathway targets for COVID-19 therapeutics.

1. Data Description

Data shown in this article includes multi-omics study on longitudinal samples collected at two time points from COVID-19 patients. We found that 5 metabolites were significantly dysregulated at the severe time points and data for the same is provided in Table 1. We have highlighted the role of arginine metabolism and its importance in therapeutics. Our finding is also supported by the literature summarized in Table S5 from the main manuscript. Fig. 1 describes the summary of the data from the main manuscript. Proteomic analysis showed a significant dysregulation of 10 proteins and 32 peptides. Dysregulated proteins were further used to do the
Table 1
Data for major metabolites and proteins mapped with arginine metabolism and leukocyte pathway.

| Proteins/Metabolites       | Fold change | p-Value   |
|----------------------------|-------------|-----------|
| Metabolites                |             |           |
| L-Arginine                 | 1.9273      | 0.00002   |
| Creatine                   | 3.0884      | 0.01      |
| 3-hydroxyoctanoylcarnitine| 0.48947     | 6.98E-07  |
| 3, 5-tetradecadiencarnitine| 0.49595   | 0.02      |
| 1-stearoyl-2-hydroxy-sn-glycero-3-PE | 0.51045 | 0.02      |
| Proteins                   |             |           |
| RAP1B                      | 1.42        | 6.30377E-05 |
| HSPA8                      | 1.61        | 0.010839226|
| PDIA3                      | 2.45        | 0.002180285|

Fig. 1. Summary of the acquired data. Plasma was collected at two-time points from the patients showing a transition from non-severe (N) to severe (S). Proteomic and metabolomic analysis was done for the collected samples through liquid chromatography-tandem mass spectrometry. Acquired data were used to identify significant proteins and metabolites using Metaboanalyst. Plots for the representative proteins and metabolites are showing the differential expression among the two-time points (Black lines show the medians and green lines represent individual data points for the expression of metabolites). Further analysis showed the significant pathways dysregulation between two-time points highlighting the importance of these proteins in disease pathogenesis and therapeutics.

pathway analysis and the results revealed the alteration of complement and coagulation cascade, platelet aggregation, myeloid leukocyte activation pathway and arginine metabolism (Fig. 1(B)). Myeloid leukocyte activation was the major focus of the proteomic analysis of the current study and data for the proteins mapped with this pathway is provided in Table 1. Moreover, for a detailed understanding of the results regarding metabolites and proteins, we suggest to Refer Figs. 2–4 of the main manuscript. Some of the proteins identified were also validated using multiple reaction monitoring approach (MRM). For detailed MRM results, please refer Fig. 5 of the main manuscript. We have provided the list of reagents used for the study (Table S1) and parameters (Tables 2–4) used for the analysis as supplementary information [1].

2. Experimental Design, Materials and Methods

Reagent used in the experimental procedure has been given in supplementary information (Table S1).
2.1. Plasma Sample Collection

Plasma samples were procured from Kasturba Hospital for Infectious Diseases, Mumbai with the approval of the Institute Ethics Committee, IIT Bombay, and Kasturba Hospital for Infectious Diseases Institutional Review Board. All the samples included in the study were RT-PCR positive. Clinical parameters of the patients like SpO2 levels, severe respiratory symptoms, and need for mechanical ventilation were further used to categorize the patients into severe and non-severe groups (Table S1). Collected blood was centrifuged at 3000 rpm for 10 min and plasma was separated for further processing.

2.2. Metabolomics Analysis

The schematic workflow adopted for metabolomic analysis is described in Fig. 2.

![Fig. 2. Schematic representation of the metabolomics-based study of plasma of COVID-19-infected patients. Longitudinal blood samples were collected at two-time points from COVID-19-positive patients who showed transitions from severe (T1) to non-severe (T2). Plasma was separated from the collected blood and samples were processed to extract metabolite using ethanol and methanol-based extraction. Metabolites were analyzed using Q-Exactive liquid chromatography mass spectrometry to obtain the raw data. The acquired data were analyzed using different tools such as Compound Discoverer, Human metabolome database, and Metaboanalyst to overview the differential trends of metabolites in the experimental groups.](image)

2.2.1. Sample Preparation

The metabolite sample preparation was done using ethanol-based method [2]. For this 100 μL of plasma was taken in a fresh sterile tube from four COVID-19 patients. To the sample 200 μL of absolute ethanol (chilled) was added. It was followed by incubation of sample inside the biosafety cabinet for 1.5 h so that ethanol can be evaporated. After this incubation, absolute methanol (chilled) was added to the semi-dried sample in four times volume of that of original sample i.e., 400 μL. Sample was then vortexed and stored at -20°C for overnight incubation. After the incubation, sample was centrifuged (4°C, 12,000 g) and supernatant was collected for future use. Further processing was done after transporting samples to Proteomics Lab, IIT Bombay. 250 μL of the supernatant was vacuum concentrated to a final volume of 80 μL. Further, for Mass spectrometric (MS) based metabolomic analysis, 50 μL of concentrated sample was taken forward in a glass vial. To analyze the instrument variability throughout the experiment, 0.5 μL of a standard metabolite reserpine (stock concentration 10 μg/mL) was added to each metabolite sample just before the MS run.

2.2.2. Data Acquisition Using Ultra-High-Performance Liquid Chromatography Tandem Mass Spectrometry

The metabolite extract with spiked in internal standard were analysed using Q Exactive mass spectrometer (Thermo Fisher Scientific, USA) coupled with ultra-high-performance liquid chro-
matography (UHPLC) system (Ultimate 3000, Thermo). For this analysis, the instrument was operated in only positive ion electrospray ionization (ESI) mode using HESI (heated ESI) source. Orbitrap mass analyser was employed and a resolution of 140,000 was set for full MS. For ddMS2, the resolution was kept as 17,500. Mass scanning was performed with an m/z range of 100 to 700. Sample were loaded to the C18 hypersil gold column (100 × 2.1 mm, 1.9 μm particle size, Thermo Fisher Scientific, USA). Parameter file for the analysis is given as supplementary information (Table S2). Solvent system used consist of solvent A as water with 0.1% formic acid (FA) and solvent B with absolute methanol.

The sheath gas flow was set as 42 and the auxiliary gas rate was set as 10 and a temperature of 34°C was set for capillary. The samples were run over a gradient of 20 min with a flow rate of 0.350 mL/min (Table S3). For every batch of samples, a blank (50% methanol) was run at the beginning and end of the batch for a particular patient. Both at the MS and MS2 levels, all the samples were run in technical triplicates. A quality check pooled sample was run in each batch of samples after every five sets of technical triplicates. This QC pool sample of a batch was prepared from the individual samples run in that particular batch.

2.2.3. Statistical Analysis

The Compound Discoverer (CD) 3.0 software (Thermo Fisher) was used to analyse the raw mass spectrometric data. This analysis primarily focused on chromatography peak alignment, metabolite identification and quantification, mass spectrum visualization and additional statistical analysis.

The workflow adopted for CD consisted of a signal-to-noise (S/N) ratio of 3 for compound identification, minimum peak intensity threshold of 10⁶ for peak alignment and compound identification using database search. The database search was performed against ChemSpider [3] which consisted of Human Metabolome Database (HMDB), KEGG and BioCyc and a mass tolerance of 5 ppm was selected. Metabolika and mzCloud were also employed as data sources for compound annotation in addition to ChemSpider. Additionally, the web-based software MetaboAnalyst [4] was primarily used for statistical analysis. The additional normalization techniques were chosen using several batches of QC pools and internal standards. To ensure the accuracy of the data, Spearman rank correlation analyses were utilized. Samples with an R² of at least 0.5 were then taken into consideration. For each cohort, the features having missing values in less than 30% of the samples were imputed with k-nearest neighbor algorithm. The missing value imputed data of both the cohorts was merged and taken for median-normalization and log-transformation in Metaboanalyst tool. This data was further taken for two-tailed unpaired t-test with false discovery rate (FDR) adjusted values to identify the differentially expressed metabolites. A log2 fold-change cut off of 1.5 and p-value threshold < 0.05 were considered as a significant criterion. Comparative abundance of significant metabolite candidates was then visualized using heatmap and Principal component analysis options in Metaboanalyst.

2.3. Discovery Based Proteomics Analysis

Fig. 3 is the representation of the workflow adopted for proteomic analysis.

2.3.1. Sample Preparation

For proteomic analysis, 9 COVID-19 infected patients showing transition from non-severe to severe symptoms were considered. Workflow for sample preparation was adopted from our previous study [5]. Schematic representation is shown in Fig. 3. Plasma samples were heat inactivated at 56°C for 30 min and stored at −80°C until use. Depletion of the plasma samples was carried out to remove the highly abundant proteins using Pierce Top 12 Abundant Protein Depletion Spin Column as per the manufacturer’s instructions. Depleted plasma was further reduced four times of its initial volume by vacuum concentrator (Thermo Fisher Scientific). Protein quantification was done for the depleted plasma samples by Bradford assay and ~30 µg of protein was used for in-solution digestion. For the reduction, plasma proteins were incubated with TCEP
Fig. 3. **Mass spectrometry-based workflow for proteomic analysis of COVID-19 infected patients.** Blood samples were collected at two-time points T1 and T2. (A) The figure shows the sample preparation for mass spectrometry. Plasma was separated from whole blood followed by the depletion of the most abundant proteins. Proteins present in the depleted plasma were precipitated using three solvents (ethanol, acetone, and isopropanol) and digested with trypsin. Fragmented proteins were desalted and fed into a mass spectrometry instrument. (B) This panel shows shotgun proteomic analysis using multiple *in silico* tools for statistical and pathway studies. (C) MRM-based targeted validation of the proteins identified through shotgun proteomic analysis.

at final concentration of 20 mM, at 37°C for 1 h. Reduced proteins were alkylated using iodoacetamide at the final concentration of 40 mM by incubating the samples under dark for 15 min. pH of the samples was adjusted to around 8 using 50 mM ammonium bicarbonate before to ease the tryptic digestion. Enzyme/substrate ratio used for digestion was 1:30, for 16 h at 37°C. Digested peptides were desalted using C-18 columns and quantified by the Scopes method [6].

### 2.3.2. Data Acquisition Using nLC

One microgram of the peptides were injected into the Orbitrap Fusion Tribrid Mass Spectrometer (Thermo Fisher Scientific) with the EASY-nLC 1200 system. A gradient of 120 min was set using 80% acetonitrile (ACN) and 0.1% FA with the sample flow rate of 300 nL/min. Blank runs were performed after every run of the plasma sample. Mass spectrometric data was acquired in the data-dependent acquisition mode. Mass scan range used for the current analysis was 375–1700 m/z and mass resolution of 60,000. A mass tolerance of 10 ppm was set with a dynamic exclusion of 40 s. High energy collision dissociation (HCD) method was used for the fragmentation of ions. Thermo Xcalibur software version 4.0 was used for the data acquisition (Table S4).

### 2.3.3. Statistical Analysis

Raw data obtained from mass spectrometry were analyzed using MaxQuant (v1.6.6.0), a freely available online tool [7]. The data were searched through Andromeda, an in-built search engine in MaxQuant against the Human Swiss Prot database (downloaded on 09/07/2020), that consisted of 20,353 proteins. During analysis, parameters were set for label free quantification (LFQ) and setting for label type was kept as ‘standard’ with a multiplicity of 1. Digestion was performed using trypsin with a maximum of two missed cleavage. The carbamidomethylation of cysteine (+57.021464 Da) and the oxidation of methionine (+15.994915 Da) were set as the static and variable modifications respectively. FDR value was set to 1% to obtain only high confidence protein and peptides with a reverse decoy mode. An output file was generated by MaxQuant that reflects the intensity of each protein sample-wise and was taken forward for statistical analysis. All the statistical analysis was performed using metaboanalyst [4]. Correlation map was made for all the samples to assess the quality of data. Missing values were imputed us-
ing the k-nearest neighbor (KNN) at a threshold of 50%. Significance for the differential proteins was calculated using Welch’s t test with a $p$ value $<0.05$ and log2 fold-change $>1.2$.

2.4. Targeted Proteomic Validation Using Multiple Reaction Monitoring

In this dataset, targeted proteomic approach of multiple reaction monitoring (MRM) was used to comparatively analyze the samples. For the selected list of proteins, the transition list required for setting the MRM method, was prepared using Skyline-daily (version 20.2.1.404) [8]. Under peptide setting in the Skyline, missed cleavages were set to 0 and human UniProt database was used as background proteome. All $+1$ and $+2$ product $y$ ions of both $+2$ and $+3$ precursors were considered. Pooled peptide sample from both the time points were utilised for optimization. The final data was acquired using seven longitudinal severe and non-severe plasma samples. Before injecting the peptide sample to mass spectrometric analysis, spiking of equal amount of heavy labelled synthetic peptide (THCLYTHVCDAIK) was done. The peak area intensity of the heavy labelled peptides was utilized for analyzing the consistency of the instrument throughout the experiment. The data was acquired using Triple quadrupole Altis mass spectrometer (Thermo Fisher Scientific, USA) coupled with a Vanquish UHPLC system (Thermo Fisher Scientific, USA) using parameters as shown in table S4. Buffer system used consisted of Solvent A as Milli Q water with 0.1% FA and Solvent B as 80% Acetonitrile with 0.1% FA. Samples were run over a gradient of 10 min with flow rate of 0.45 mL/min.

2.5. Pathway Analysis

Reactome (version 73) [9] and metaboanalyst [4] was used to map the significant proteins to the biological pathways. Integrative gene ontology enrichment analysis was done in Metascape [10] and proteins mapped with the interesting pathways were used for visualization. Protein-protein interaction analysis was performed using STRING (version 11) [11]. Knowledge of affected pathways can be further explored to identify the therapeutic targets and to improve our understanding on disease pathogenesis.

Ethics Statements

This work was approved by the Institute Ethics Committee, IIT Bombay, and Kasturba Hospital for Infectious Diseases Institutional Review Board (IITB-IEC/2020/030).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

A longitudinal study of COVID-19 patients using proteomics approach (PRIDE)
A Multi-omics Longitudinal Study Reveals Alteration of the Leukocyte Activation Pathway in COVID-19 Patients (metaboLights)
Validation of proteins showing significant dysregulation from shotgun proteomics experiments in longitudinal plasma samples of COVID-19 patients (Panorma).
CRediT Author Statement

Sakshi Rajoria: Writing – original draft, Investigation, Visualization; Mehar Un Nissa: Writing – original draft, Investigation, Visualization; Kruthi Suvarna: Writing – review & editing, Formal analysis; Harsh Khatri: Formal analysis; Sanjeeta Srivastava: Conceptualization, Project administration, Funding acquisition.

Acknowledgments

The study was supported through the Ministry of Science and Technology, Department of Biotechnology, Government of India (BT/PR41020/COT/142/14/2020), and a special COVID-19 seed grant (RD/0520-IRCCHC0-006) from IRCC, IIT Bombay, to S.S. MASSFIITB (Mass Spectrometry Facility, IIT Bombay) from the Department of Biotechnology (BT/PR13114/INF/22/206/2015) is gratefully acknowledged for the MS-based proteomics work. The authors thank to MERCK-COE (DO/2021-MLSP) for their extended support.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi: 10.1016/j.dib.2022.108765.

References

[1] K. Suvarna, A. Salkar, V. Palanivel, R. Bankar, N. Banerjee, M. Gayathri J Pai, A. Srivastava, A. Singh, H. Khatri, S. Agrawal, O. Shrivastav, J. Shastri, S. Srivastava, A multi-omics longitudinal study reveals alteration of the leukocyte activation pathway in COVID-19 patients. J. Proteome Res. 20 (2021) 4667–4680, doi: 10.1021/acs.jproteome.1c00215.

[2] S. Aggarwal, S. Parikhari, A. Banerjee, J. Roy, N. Banerjee, R. Bankar, S. Kumar, M. Choudhury, R. Shah, K. Bhojak, V. Palanivel, A. Salkar, S. Agrawal, O. Shrivastav, J. Shastri, S. Srivastava, Metabolome and exosome profiling of the biospecimens from COVID-19 patients in India. 98 (2021) 397–413. 10.3623/j0372-9311-161.

[3] H.E. Pence, A. Williams, ChemSpider: an online chemical information resource, J. Chem. Educ. 87 (2010) 1123–1124, doi: 10.1021/ed100697w.

[4] Z. Pang, J. Chong, C. Zhou, D.A. de Lima Morais, L. Chang, M. Barrette, C. Gauthier, P.-É. Jacques, S. Li, J. Xia, Metabo-Analyzer 5.0: narrowing the gap between raw spectra and functional insights, Nucleic Acids Res. 49 (2021) W388–W396, doi: 10.1093/nar/gkab382.

[5] K. Suvarna, D. Biswas, M.G.J. Pai, A. Acharjee, R. Bankar, V. Palanivel, A. Salkar, A. Verma, A. Mukherjee, M. Choudhury, S. Ghantasala, S. Ghosh, A. Singh, A. Banerjee, A. Badaya, S. Bihani, G. Loya, A. Mantri, A. Burli, J. Roy, A. Srivastava, S. Agrawal, O. Shrivastav, J. Shastri, S. Srivastava, Proteomics and machine learning approaches reveal a set of prognostic markers for COVID-19 severity with drug repurposing potential, Front. Physiol. 12 (2021) 652799, doi: 10.3389/fphys.2021.652799.

[6] V. Kumar, S. Ray, S. Ghantasala, S. Srivastava, An integrated quantitative proteomics workflow for cancer biomarker discovery and validation in plasma, Front. Oncol. 10 (2020) https://www.frontiersin.org/articles/10.3389/fonc.2020.543997, (accessed July 7, 2022).

[7] S. Tyanova, T. Temu, J. Cox, The MaxQuant computational platform for mass spectrometry-based shotgun proteomics, Nat. Protoc. 11 (2016) 2301–2319, doi: 10.1038/nprot.2016.136.

[8] B. MacLean, D.M. Tomazela, N. Shulman, M. Chambers, G.L. Finney, B. Frewen, R. Kern, D.L. Tabb, D.C. Liebler, M.J. MacCoss, Skyline: an open source document editor for creating and analyzing targeted proteomics experiments, Bioinformatics. Oxf. Engl. 26 (2010) 966–968, doi: 10.1093/bioinformatics/btq054.

[9] A. Fabregat, S. Jupe, L. Matthews, K. Sidiropoulos, M. Gillespie, P. Garapati, R. Haw, B. Jassal, F. Korninger, B. May, M. Milacic, C.D. Roca, K. Rotherfils, C. Sevilla, V. Shamovsky, S. Sharor, T. Varusai, G. Viteri, J. Weiser, G. Wu, L. Stein, H. Hermjakob, P. D’Eustachio, The reactome pathway knowledgebase, Nucleic Acids Res. 46 (2018) D649–D655, doi: 10.1093/nar/gkx1132.

[10] Y. Zhou, B. Zhou, L. Pache, M. Chang, A.H. Khodabakhshi, O. Tanaseichuk, C. Benner, S.K. Chanda, Metascape provides a biologist-oriented resource for the analysis of systems-level datasets, Nat. Commun. 10 (2019) 1523, doi: 10.1038/s41467-019-09234-6.

[11] D. Szklarczyk, A.L. Gable, D. Lyon, A. Junge, S. Wyder, J. Huerta-Cepas, M. Simonovic, N.T. Doncheva, J.H. Morris, P. Bork, L.J. Jensen, C. von Mering, STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets, Nucleic Acids Res. 47 (2019) D607–D613, doi: 10.1093/nar/gky1131.