|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Author name | Year | Study | Study Participants Notes | Processing Notes |
| Ahern JD et al. | 2022 | A blood atlas of COVID-19 defines hallmarks of disease severity and specificity | None | Received data from the authors. The dataset has been processed as described in the paper: “The proteomics dataset was processed as follows: (1) Protein filtering such that proteins with at least 50% of valid values in one group were kept; (2) Sample filtering such that samples with more than 50% of missing values were removed from the dataset; (3) Data normalization with log2 transformation and median centering of the dataset. Imputation of missing values was performed using a mixed model that combines a K-Nearest Neighbor approach (KNN) when at least 60% of valid values are present, otherwise a Minimum probability approach is used where missing values are randomly drawn from a Gaussian distribution (shift = 1.8, nstd = 0.3). The resulting data matrix contains 353 samples and 105 proteins. Thirteen samples were further excluded from analysis for malignancy, immunosuppression, or being alternative samples.” We used this dataset and extracted only the COVID-19 patients who had an acute disease (positive SARS-CoV-2 PCR test) and compared them to the group of the healthy volunteers (negative SARS-CoV-2 test). |
| Babacic H et al. | 2022 | Comprehensive proteomics- and meta-analysis of COVID-19 host response detects elevated proteasomal proteins in blood traceable to SARS-CoV-2 infection | This study | This study |
| Byeon SK et al. | 2022 | Development of a multiomics model for identification of predictive biomarkers for COVID-19 severity: a retrospective cohort study | Paired samples from the same individuals, before and after COVID-19. Received information from the authors about which samples match to which patient in the three TMT-16 sets | Downloaded tables from PRIDE, ID: PXD029376, including proteome quantifications from sets 1, 2, and 3, which included paired samples (pre- and post- COVID-19) from 8, 8, and 5 individuals, respectively. Extracted a list of proteins identified in the study (UniProt IDs) and made lists of quantifications of all individuals for each protein. Values of 0 were assigned as NA, followed by log2-transformation. Then, the UniProt protein IDs were mapped to gene names with the ID mapper in the UniProt database and recalculated average values for proteins mapping to the same gene names. |
| Ciccosanti F et al. | 2022 | Proteomic analysis identifies a signature of disease severity in the plasma of COVID-19 pneumonia patients associated to neutrophil, platelet and complement activation | The authors have indicated HD, non-ICU and ICU patients in the dataset file; a) and b) were annotations for technical replicates and referred to the same patient. | Received data from authors. First sheet was log2 iBaq normalized on the median of total iBAQ. The second sheet was the dataset obtained after filtering and computation of empty values by Perseus software; this second sheet is the one used for the analyses by the authors. We used this dataset for analyses. We have calculated the average values for each protein where the participants were analysed in replicates. We recalculated average values for proteins mapping to the same gene names. |
| Di B et al. | 2020 | Identification and validation of predictive factors for progression to severe COVID-19 pneumonia by proteomics | None | Downloaded from Supplementary Data S1. log2-transformed the intensity values. The UniProt protein IDs were mapped to gene names with the ID mapper in the UniProt database and recalculated average values for proteins mapping to the same gene names. |
| Feng Z et al. | 2022 | Screening and Analysis of Serum Protein Biomarkers Infected by Coronavirus Disease 2019 (COVID-19) | None | Data available in Supplementary Table S1. Sheet “Quantifiable protein”. log2-transformed the intensity values. We recalculated average values for proteins mapping to the same gene names. |
| Geyer EP et al. | 2021 | High-resolution longitudinal serum proteome trajectories in COVID-19 reveal patients-specific seroconversion | None | Received a log10 normalised dataset from the authors. Converted it back to non-transformed values and then log2-transformed the values. The UniProt IDs were mapped to gene names with the ID mapper in the UniProt database for proteins that had no matching gene names. We recalculated average values for proteins mapping to the same gene names. |
| Messner CB et al. | 2020 | Ultra-High-Throughput Clinical Proteomics Reveals Classifiers of COVID-19 Infection. | None | Received datasets on discovery and validation cohort from the authors. log2-transformed the values. |
| Mohammed Y et al. | 2022 | Longitudinal Plasma Proteomics Analysis Reveals Novel Candidate Biomarkers in Acute COVID-19 | None | Downloaded a protein table from PRIDE, ID: PXD029437. log2-transformed the values. The UniProt protein IDs were mapped to gene names with the ID mapper in the UniProt database. |
| Overmyer KA et al. | 2021 | Large-Scale Multi-omic Analysis of COVID-19 Severity | None | Received log2-normalised dataset from authors. It contained also imputed values, as described in the paper: “We imputed values if only if we detected the protein in 50% of the samples. Imputation was done by randomly drawing values from the left tail of the normal distribution of all measured protein abundance values. We also filtered based on RSDs of our instrument control sample (pooled human plasma); proteins with control RSDs greater than 30% were removed”. The UniProt IDs were mapped to gene names with the ID mapper in the UniProt database for proteins that had no matching gene names. We recalculated average values for proteins mapping to the same gene names. For proteins that were mapping to the same gene name, we calculated the average quantification value. |
| Sahin A et al. | 2022 | Plasma proteomics identify potential severity biomarkers from COVID-19 associated network | The data consisted of four sheets containing protein quantification values for 6 critical-, 3 severe-, 4 moderate- COVID-19 patients and 5 healthy individuals, respectively. Most samples from COVID-19 patients were run in replicates. | Received data from authors, which included individual normalized PSM values of patients and healthy subjects. The normalization was applied by taking the mean of the proteins with a PSM value of more than 500 and dividing every PSM value of the proteins by the mean. We log2-transformed the values. For individuals with replicated measurements, we calculated the average values. We the recalculated the average values for proteins mapping to the same gene names. |
| Sameh M et al. | 2023 | Integrated multiomics analysis to infer COVID-19 biological insights. | None | Data and code available at https://gitlab.com/prolab11/Covid-19. We used the same code by the authors to obtain the normalized log2-transformed data using probabilistic quotient normalization (PQN). The latter transforms proteins/metabolites spectra according to an overall estimation on the most probable dilution. The UniProt IDs were mapped to gene names with the ID mapper in the UniProt database. We recalculated average values for proteins mapping to the same gene names. |
| Shen B et al. | 2020 | Proteomic and Metabolomic Characterization of COVID-19 Patient Sera | None | Data available from github. Used the authors code to extract the dataset and log2-normalised it. |
| Shu T | 2020 | Plasma Proteomics Identify Biomarkers and Pathogenesis of COVID-19 | In the discovery cohort, we used only the quantification values for the samples T1, which corresponded to the first samples from the fatal subgroup and the samples at the disease peak for the subgroups of moderate and severe disease. | Data on two cohorts available in Table S3 of the manuscript. log2-transformed the data. Removed samples with 100% missing values. In the validation cohort, removed those with 80% missing values before processing, because they had insufficient observations to calculate the necessary parameters for SMD. |
| Spick M et al. | 2022 | Multi-Omics Reveals Mechanisms of Partial Modulation of COVID-19 Dysregulation by Glucocorticoid Treatment | None | Data available in Supplementary Table S2. The UniProt IDs were mapped to gene names with the ID mapper in the UniProt database. We recalculated average values for proteins mapping to the same gene names. |
| Sullivan KD et al. | 2021 | The COVIDome Explorer researcher portal | None | Data downloaded from: https://covidome.shinyapps.io. log2-transformed the values. |
| Suvarna K et al. | 2021 | Proteomics and Machine Learning Approaches Reveal a Set of Prognostic Markers for COVID-19 Severity With Drug Repurposing Potential. | The attached files contain data for 71 samples, although the authors have run 74 samples. Rationale: From the Result section of the Manuscript: "A sample-wise correlation analysis of 74 samples was performed to understand the data quality. Of these, three samples were removed as they did not meet the standards. Proteomic data of 71 samples were taken forward." | Downloaded a meta file and quantification file from PRIDE: PXD022296. Used the LFQ[.]intensity values and log2-normalised them. The authors have performed additional filtering: "We have identified around 1200+ proteins but filtered proteins based on 50% Missing Value criteria to negate false identification of DEPs. However, we have further validated our markers using Targeted Proteomics to gain confidence." |
| Tepasse P et al. | 2022 | The Dysregulation of the Renin-Angiotensin System in COVID-19 Studied by Serum Proteomics: Angiotensinogen Increases with Disease Severity | Included only COVID-19 patients with acute disease available in the dataset and compared them to PCR-negative controls. Convalescent samples were not included in the analysis. | Received normalised data from authors. NAs were assigned to values of 0 and the data were log2-transformed. Extracted participants' information from Table S2. |
| Zhang Y et al. | 2022 | Potential Use of Serum Proteomics for Monitoring COVID-19 Progression to Complement RT-PCR Detection. | Selected the patients with course 1 and 2., which refer to samples taken during stages 1 and 2, i.e., the nucleic acid positive (NCP) stage. The remaining samples were excluded because the patients were PCR negative at the time of analysis. | Data available in Supplementary Table S3. The UniProt IDs were mapped to gene names with the ID mapper in the UniProt database. We recalculated average values for the proteins in instances where the participants were analysed in replicates. |