MICROBiome-based biomarkers to PREDICT decompensation of liver cirrhosis and treatment response

D3.1 Report on Acquire Proteomics Data

Project acronym:	MICROB-PREDICT
Grant Agreement:	825694
Project Duration:	01 January 2019 – 31 March 2025 (75 months)
Version:	V1
Date:	30/03/2021
WP Leader:	Matthias Mann (07 MPG)
Authors:	Philipp E. Geyer, Peter V. Treit, Matthias Mann (7 MPG), Jonel Trebicka (1 EFCLIF)
Due date of deliverable	Month 27
Actual submission date	30/03/2021
Dissemination level	PU

Table of Contents

1. Executive Summary 2
2. Background and rationale 2
3. Samples 2
4. Experimental methods 3
5. Data analysis 5
6. Quality control 5
7. Results 6
8. Discussion 10
9. Outlook 10
10. Conclusion 11
11. Acknowledgement and Disclaimer 11
12. References 11

1. Executive Summary

The work described in this report fulfils the requirements for the deliverable. Herein, we acquired MS-based proteomics data for all samples in the MUCOSA-PREDICT study. We deposited the data on MICROB-PREDICT´s data hub. Additionally, the quality of the acquired data has been assessed in depth by evaluating single samples and the total study towards preanalytical variation. We further report on quantitative reproducibility of the measurements and the biological variation across all sample types. Moreover, we already generated a first insightful analysis by identifying potential biomarker candidates and biological remodeling in ACLS by investigating the plasma proteome of MUCOSA-PREDICT samples.

2. Background and rationale

Patients with acutely decompensated cirrhosis, who do not have organ failures are considered as having acute decompensation (AD). The PREDICT study described three different clinical courses of patients with AD after hospital admission (Trebicka *et al*., 2020). The first clinical course (pre-ACLF) includes patients who develop ACLF and has high probability of death. These patients are characterized by high-grade systemic inflammation. The second clinical course (unstable decompensated cirrhosis, UDC) includes patients requiring frequent hospitalizations unrelated with ACLF, show low-grade systemic inflammation but suffer characteristically from complications related to severe portal hypertension. They present lower risk of mortality than patients with pre-ACLF. Finally, the third clinical course (stable decompensated cirrhosis, SDC), includes two-thirds of all patients admitted hospital with AD. They do not present severe systemic inflammation or frequent complications related with portal hypertension, rarely require hospital admissions and present an extremely low 1-year mortality risk.

Therefore, Pre-ACLF should be suspected in patients hospitalized for AD with prior unstable clinical course, very high levels of inflammatory markers and liver failure or liver or kidney dysfunction. Unfortunately, we were unable to design new specific tools that improve the accuracy of the CLIF-C AD and MELD-Sodium scores for predicting ACLF development.

Proteins control and execute the vast majority of biological processes and mass spectrometry (MS)-based proteomics is the technology of choice to investigate the entirety of all proteins in a biological system – its proteome (Aebersold and Mann, 2016). MS-based proteomics has continuously developed over the past 20 years, enabling the holistic, detailed and quantitative investigation of diverse biological systems. Discovery of new and potentially better biomarkers and biomarker panels is facilitated by these rapid developments in MS-based proteomics, promising widespread medical applications (Geyer *et al.*, 2017).

3. Samples

MPG has received 309 plasma, 309 urine, 309 stool and 362 mucosa samples from the MUCOSA-PREDICT cohort, which have been drawn from 309 patients.

4. Experimental methods

Sample preparation for proteomics analysis

Plasma sample preparation was carried out according to MPG´s Plasma Proteome Profiling pipeline with an automated setup on an Agilent Bravo Liquid Handling Platform. In brief, 1 µl plasma samples were diluted 1:10 with ddH2O and 10 µl of the sample was mixed with 10 µl PreOmics lysis buffer (P.O. 00001, PreOmics GmbH) for reduction of disulfide bridges, cysteine alkylation, and protein denaturation at 95°C for 10 min (Geyer *et al.*, 2016). Trypsin and LysC were added to the mixture at a ratio of 1:100 micrograms of enzyme to micrograms of protein. Digestion was performed at 37°C for 1 h. An amount of 200 ng of peptides were loaded on an Evotip according to (Bache *et al.*, 2018) for analysis.

Mucosa samples were added to 50 µl of PreOmics lysis buffer, heated at 95°C for 10 min and sonified in a Covaris instrument. Trypsin and LysC were added to the mixture at a ratio of 1:100 micrograms of enzyme to micrograms of protein. Digestion was performed at 37°C for 4 h. 20 µg of peptides were loaded on two 14-gauge StageTip plugs, followed by consecutive purification steps with 1% TFA in isopropanol and 0.2% TFA in water. Finally, peptides were eluted with 1% ammonia in 80% ACN. The StageTips were centrifuged using an in-house 3D-printed StageTip centrifugal device at 1,500 g. The collected material was completely dried using a SpeedVac centrifuge at 60°C (Eppendorf, Concentrator plus). Peptides were suspended in buffer A* [2% acetonitrile (v/v), 0.1% formic acid (v/v)] and sonicated (Branson Ultrasonics, Ultrasonic Cleaner Model 2510).

Stool samples: 40 µg of raw material were dissolved in 200 µl of PreOmics lysis buffer, heated at 95°C for 10 min, frozen at -80°C and sonified in a biorupter instrument. This process was repeated. Trypsin and LysC were added to the mixture at a ratio of 1:100 micrograms of enzyme to micrograms of protein. Digestion was performed at 37°C for 4 h. 20 µg of peptides were loaded on two 14-gauge StageTip plugs, followed by consecutive purification steps with 1% TFA in isopropanol and 0.2% TFA in water. Finally, peptides were eluted with 1% ammonia in 80% ACN. The StageTips were centrifuged using an in-house 3D-printed StageTip centrifugal device at 1,500 g. The collected material was completely dried using a SpeedVac centrifuge at 60°C (Eppendorf, Concentrator plus). Peptides were suspended in buffer A* [2% acetonitrile (v/v), 0.1% formic acid (v/v)] and sonicated (Branson Ultrasonics, Ultrasonic Cleaner Model 2510).

Urine samples were prepared using MStern Blot protocol as described previously (Berger *et al.*, 2015). Briefly, 100 µl of urine was first diluted in 300 µl of urea sample solution (8 M urea in 50 mM ammonium bicarbonate (ABC)) and subsequently mixed with 30 µl of 150 mM dithiothreitol (DTT) solution (150 mM DTT, 8 M urea, 50 mM ABC) in a 96-well plate. The resulting solution was incubated for 20 min at room temperature. Reduced cysteine side chains were alkylated by adding 30 µL of iodoacetamide (IAA) solution (700 mM IAA, 8 M urea, 50 mM ABC) and incubated for 20 min in the dark. During incubation, each well of the 96-well PVDF membrane plates (MSIPS4510, Merck Millipore) was activated and equilibrated with 150 µl of 70% ethanol/water and urea sample solution, respectively. The urine samples were transferred through the PVDF membranes using a vacuum manifold (MSVMHTS00, Merck Millipore). Adsorbed proteins were washed two times with 150 µl of 50 mm ABC. Digestion was performed at 37°C for 2 hours by adding 100 µl digestion buffer (5% v/v acetonitrile (ACN)/50 mm ABC) containing 0.35 µg per well of each protease trypsin and LysC. After incubation in a humidified incubator, the resulting peptides were collected by applying vacuum and remaining peptides were eluted twice with 75 µl of 40%/0.1%/59.9% (v/v) acetonitrile/formic acid/water. The pooled peptide solutions were dried in a vacuum centrifuge and 200 ng were loaded

on an Evotip for LC-MS analysis.
Pools for each of the four sample types were generated from all individuals and prepared according to the procedure above. The peptides were fractionated using the high pH reversed-phase "Spider fractionator" into 24 fractions as described previously to generate deep proteomes (Kulak, Geyer and Mann, 2017).
LC-MS/MS analysis was performed on an Evosep One system for plasma and urine samples and on an EASY-nLC 1200 for stool and mucosa samples. The LC systems were coupled to a Q Exactive HF-X Orbitrap mass spectrometer via a nano-electrospray ion source (all Thermo Fisher Scientific). Purified peptides were separated at 60°C on 50 cm columns with an inner diameter of 75 µm packed in-house with ReproSil-Pur C18-AQ 1.9 µm resin (Dr. Maisch GmbH). For plasma samples the Evosep 60 samples/day method, for urine the Evosep 30 samples/day method, for mucosa and stool samples 100 min gradients were used. Mobile phase for the Evosep One were 99.9/0.1% water/formic acid (v/v) and 99.9/0.1% acetonitrile/formic acid (v/v/v). Mobile phase for the EASY-nLC 1200 were 99.9/0.1% water/formic acid (v/v) and 80/20/0.1% acetonitrile/water/formic acid (v/v/v).

Plasma, urine and mucosa samples were analyzed by data independent acquisition and stool samples with data dependent acquisition.

The numbers of measured samples for each sample type as well as the numbers of fractions are displayed in Fig.1.

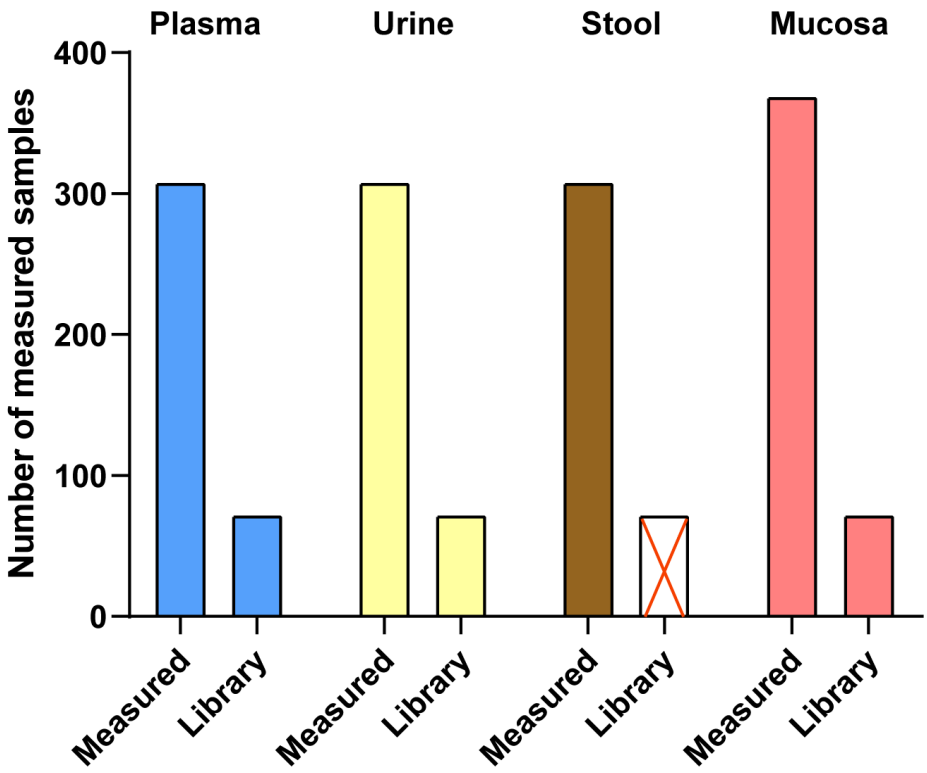


Fig. 1: Overview of acquired samples. The total number of measured single samples (Measured) as well as the number of fractionations to generate a library are displayed.

5. Data analysis

Mass spectrometry data processing: For plasma, urine and mucosa samples the MS data of the fractionated and the single-shot subject samples were used to generate a DDA library and directDIA library, respectively, which were computationally merged into two cohort-specific hybrid libraries using Spectronaut. Peptide lists were searched against the human UniProt FASTA database.

For stool samples MS raw files were analyzed by MaxQuant software, version 1.5.6.8, (Cox & Mann, 2008), against the human integrated gene catalogue database. A contaminant database generated by the Andromeda search engine (Cox et al, 2011) was configured with cysteine carbamidomethylation as a fixed modification and N-terminal acetylation and methionine oxidation as variable modifications. We set the false discovery rate (FDR) to 0.01 for protein and peptide levels with a minimum length of 7 amino acids for peptides, and the FDR was determined by searching a reverse database. Enzyme specificity was set as C-terminal to arginine and lysine as expected using trypsin and LysC as proteases. A maximum of two missed cleavages were allowed. Peptide identification was performed with an initial precursor mass deviation up to 7 ppm and a fragment mass deviation of 20 ppm. The “match between run algorithm” in the MaxQuant quantification (Nagaraj et al, 2012) was disabled. All proteins and peptides matching to the reversed database were filtered out. Label-free protein quantitation (LFQ) was performed with a minimum ratio count of 2 (Cox et al, 2014).

Bioinformatics data analysis: The Perseus software package versions 1.6.0.7 and 1.6.1.3 (Tyanova *et al*, 2016) and GraphPad Prism version 7.03 were used for the data analysis.

6. Quality control

Performance of LC-MS instrumentation was evaluated constantly as measurements proceeded, after and weekly during the measurements. This has been done by investigation of standard pressure and flow-rate parameters for the LC system as well as analysis of a standard peptide sample based on a HeLa cell line. The performance of the MS instrument has been evaluated by a maintenance procedure using calibration mixtures and based on a standard peptide sample based on a HeLa cell line.

The quality of plasma and urine samples has been further evaluated using quality marker panels for coagulation, platelet contamination and erythrocyte lysis according to (Geyer *et al.*, 2019). The majority of all samples passed the quality control cut-offs to have very good quality.

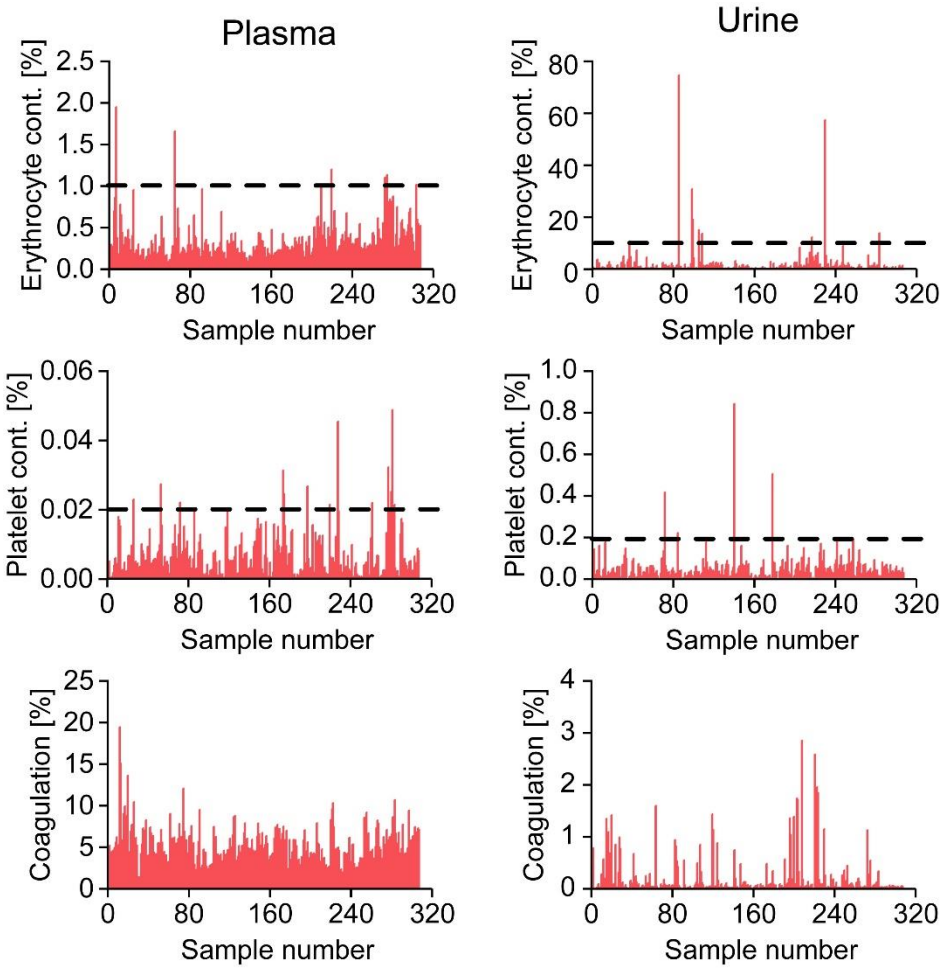


Fig. 2: Quality assessment. Plasma and urine samples have been evaluated with regards to the indicated quality marker panels.

7. Results

MPG has measured 309 plasma, 309 urine, 309 stool and 362 mucosa samples from the MUCOSA-PREDICT cohort and obtained proteomes of all samples (100% of samples have been successfully analyzed).

In total, we quantified 614, 3018, 7441 and 16,996 proteins in plasma, urine, mucosa and stool samples, respectively (Fig. 3A/B). The distribution of the numbers of proteins quantified in each of the sample types is as expected. In plasma samples proteins have a distribution across a very large concentration range, resulting, in the lowest number of quantified proteins. In contrast, the proteins within the mucosa samples show a narrower, equally distributed concentration range.

The high number of proteins in stool samples is due to the various individual-specific bacterial species present in each sample. The consistency of identified proteins for the four different tissues can be explained by the nature of these samples.

Plasma and mucosa samples contain proteins that must be present according to the nature of these samples in all individuals (Fig. 3A/B). In both sample types 50% of all proteins were quantified in around 75% of all samples. Therefore, the biological variation of protein intensities across all samples was also very low in plasma and mucosa with an average coefficient of variation (CV) of 66% and 51% (Fig. 3C).

In contrast, urine and stool can be seen as the medium used for the disposal of distinct molecules. Therefore, it is not surprising that the variation of identified proteins and their respective concentrations varied considerably across these sample types. Therefore, only a minority of all identified proteins were present completely in all samples in stool and urine.

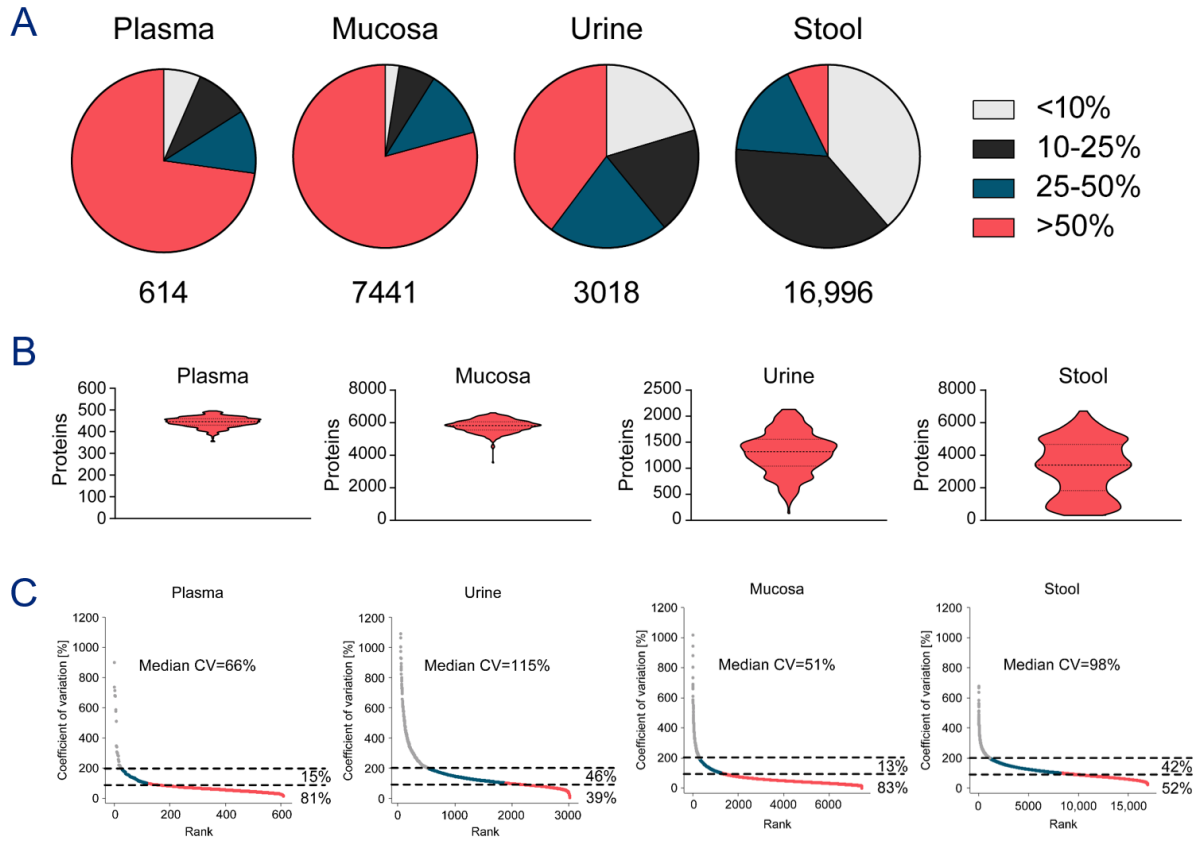


Fig. 3: Overview of quantified proteins per sample. (A) The indicated number of proteins have been identified across the samples. The consistency of proteins found in the study is indicated. (B) Distribution of quantified proteins per sample is displayed as violine plots. (C) Distribution of proteins quantified with a distinct CV of quantified proteins across the complete cohort. The generated data has been transferred to the data hub and WP4 for further analyses.

At enrolment and likewise during each subsequent visit throughout the course of the study, patients were stratified into four groups that represented different clinical courses of liver diseases, namely a group of patients with liver disease and no complications, liver disease and complications, liver disease patients that develop ACLF within this study and patients that have already ACLF at the time of recruitment. Patients in the No-ACLF stable group showed signs of stable decompensated cirrhosis but didn't develop ACLF and hence showed a lower 1-year mortality risk. Patients within the No-ACLF unstable group showed signs of unstable decompensated cirrhosis and typically suffered from complications of severe portal hypertension, however they still didn't develop ACLF. Patients within the pre-ACLF group developed ACLF within 90 days and showed pertinent signs of systemic inflammation and suffered from a higher risk of mortality. Patients in the ACLF group had ACLF at the onset visit, and subsequently showed high mortality within the following 30-day period. These patients tend to suffer from multi-organ failure throughout the course of their hospitalization. Clinical metadata was collected for each patient and supplied to the MPG, along with samples that were collected at

inclusion, and at various further time points throughout the course of disease. Samples received by MPG were all collected at the University of Debrecen.

The main goal of the proteomics work was to find a multi-protein biomarker panel that can be used to track ACLF progression, eventually also connected to the results from other omics technologies present in the consortium. By doing so, we hoped to be able to stratify patients into the aforementioned clinical courses based on patterns of alternate regulation in the proteome. In order to draw these comparisons, we first separated the study time points, and compared each patient group within that time point against each other. We started with comparing the ACLF group with the other three groups that have a lower disease severity. This might allow us to identify differentially expressed proteins that have the potential to also predict the disease course as they are differently expressed across disease states. Figure 4A exemplifies the analysis for plasma samples, where the ACLF group at the time point of inclusion was compared to all three other groups. Interestingly, the ACLF group had distinct proteins increased and decreased compared to all other groups. In more detail, we identified in the first analysis 27 statistically significantly regulated proteins, consisting of 15 immunoglobulins that were increased in the ACLF group, three complement factors and four very consistently regulated proteins (F9, B2M, RBP4, PTGS). Proteins significantly increased or decreased between the ACLF and the other groups were further investigated towards the fold-change of these proteins (Fig. 4B). Herein, we especially want to highlight the extensive involvement of the immune-system shown by the upregulation of immunoglobulins.

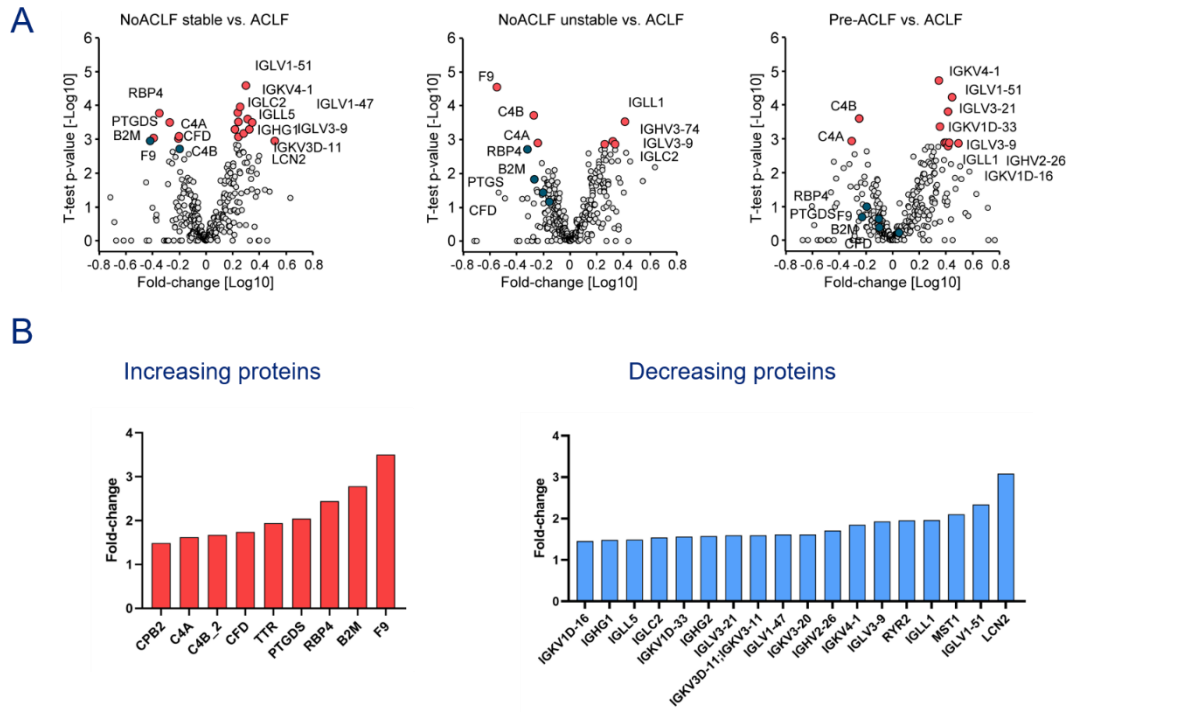


Fig. 4: Comparison of the four disease groups in MUCOSA-Predict. (A) Volcano plots showing results for the plasma samples for the two-sample t-test at the inclusion time point (S00). The fold-change in protein levels is depicted on the x-axis and the -log10 t-test p-value on the y-axis. Significantly different proteins are highlighted in red and labeled. (B) Proteins increasing and decreasing ranked according to the fold-changes of the proteins between the ACLF and the other three groups.

Proteins that are statistically different between the ACLF group and the other samples were further investigated. For this purpose, we first displayed the overall expression of the protein levels in plasma across the four disease groups. We calculated the median level of each significantly altered protein, Z-scored the protein levels and color-coded their levels in a heatmap (Fig. 5A). Indeed, most proteins were gradually increasing across the disease level shown by the four groups.

Next, we assumed that a potential marker protein should not only be significantly different between the ACLF group and the other groups but should also be longitudinally increased in disease progression within the pre-ACLF group (Fig. 5B). Herein, most proteins were increasing or decreasing longitudinally, further indicating their potential to be used as a biomarker for disease progression and the development of ACLF. Interestingly, we found that proteins started to increase and decrease at different time points in the disease course. This further indicates the potential that distinct proteins could be used to further characterize the "disease course time point" or "state" of a patient that is progressing to ACLF. Moreover, the strong involvement of the adaptive immune system in ACLF as indicated in previous tests is further confirmed with this analysis. It seems that the immunoglobulins are systematically depleted from the plasma samples, which could indicate a collapse of the adaptive immune system.

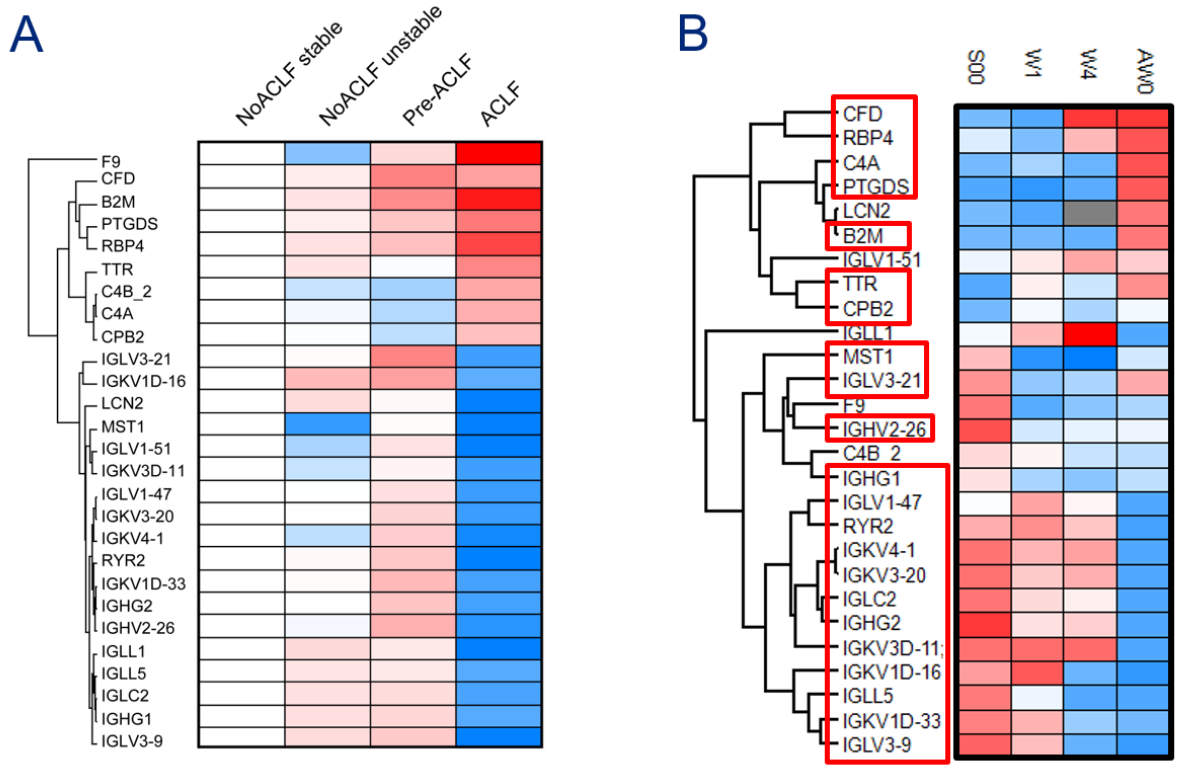


Fig. 5: Evaluation of candidate proteins. (A) Median Z-scored protein intensities are color coded and displayed for the four disease groups. (B) Protein levels in the Pre-ACLF group across the disease course. S00 is the time point of inclusion in the study. W1 and W4 are one and four weeks after inclusion. AW0 is the time point of ACLF onset. Protein levels displayed in red are increased and in blue are decreased.

Similar analysis for the other samples just started. Nevertheless, so far, we have encountered only one interesting protein in urine samples (GLRX), which was close to significance in the comparison of the ACLF group to the other groups.

8. Discussion

MS-based proteomics data have been generated for all samples and reported to MICROB-PREDICT´s data hub. Furthermore, the quality of plasma and urine samples has been assessed towards preanalytical processing and was considered to be high. There are no established markers for the preanalytical variation for mucosa or stool samples. Nevertheless, we assessed the analytical variability and reported the number of proteins quantified across samples. Furthermore, we gave an estimation of the biological variability in each sample type.

Next, we already started with the assessment of potential biomarker candidates. Finding biologically and medically meaningful data will be the ultimate confirmation of the data quality. For the inclusion time-point (S00), we have found several interesting biomarker candidates, especially amongst groups that compared the ACLF at admission group to the other, non-ACLF groups. We found proteins across various functions differently expressed in the ACLF group and the other disease groups as well as longitudinally regulated in the Pre-ACLF group. This finding is supported by the knowledge that ACLF typically leads to multi-organ failure, which is measured using the CLIF-OF metric. Organ failure is typically associated with a vast array of infections, made possible by the progressive deterioration of physiological barriers found within the body. The change in the regulation of certain immunoglobulins as can be seen already at the inclusion time point supports this fact, as patients in the ACLF group already show marked signs of systemic immune response when compared to the other groups. Other interesting proteins include members of the complement cascade, which are present throughout most time-points of the study. Interestingly, also the coagulation factor F9 was among the candidates pointing towards an involvement in the coagulation system in ACLF.

As the regulation of immunoglobulins is individual-specific, we did not assume that they will show up in the longitudinal regulation of Fig. 5B in which the median of the pre-ACLF group has been investigated longitudinally. This points towards a very strong involvement of the innate immune system. In one of the next steps, we will investigate the immunoglobulin profiles of each patient of the pre-ACLF group in more detail as recently shown by us (Geyer *et al.*, 2021). This in turn could further pinpoint towards a collapse of the adaptive immune system during the development of ACLF.

9. Outlook

As a next step, we will perform a more in-depth analysis of urine, mucosa and stool proteomics data. In addition to a first potential candidate in urine samples, we hope to find further interesting proteins that could be applied to predict ACLF onset. We will further use the method established in the plasma samples, in which we first try to find statistically significant differences between the ACLF and all other groups and follow up the regulation of these candidates in the Pre-ACLF group.

10. Conclusion

The work described in this report fulfils the requirements for the deliverable as all data have been acquired. We generated proteomics data for all proposed samples in Mucosa-Predict and uploaded the data on the MICROB-PREDICT´s data hub. We further assessed the quality of the acquired data in depth by evaluating single samples and the total study towards preanalytical variation. Additionally, we reported quantitative reproducibility of the mass spectrometry-based proteomics measurements and the biological variation across all sample types. Moreover, we could generate already first insightful results showing the high quality of the acquired data.

11. Acknowledgement and Disclaimer

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 825694.

This report reflects only the author's view and the Commission is not responsible for any use that may be made of the information it contains.

12. References

Aebersold, R. and Mann, M. (2016) 'Mass-spectrometric exploration of proteome structure and function', *Nature*, 537(7620), pp. 347–355. doi: 10.1038/nature19949.

Geyer, P. E. *et al.* (2017) 'Revisiting biomarker discovery by plasma proteomics', *Molecular Systems Biology*, 13(9), p. 942. doi: 10.15252/msb.20156297.

Geyer, P. E. *et al.* (2021) *High-resolution longitudinal serum proteome trajectories in COVID-19 reveal patients-specific seroconversion*. preprint. Infectious Diseases (except HIV/AIDS). doi: 10.1101/2021.02.22.21252236.

Trebicka, J. *et al.* (2020) 'The PREDICT study uncovers three clinical courses of acutely decompensated cirrhosis that have distinct pathophysiology', *Journal of hepatology*, 73(4). doi: 10.1016/j.jhep.2020.06.013.