Proteome Science

RESEARCH

Open Access

Plasma proteome analysis of rheumatic patients reveals differences in fingerprints based on cardiovascular history: a pilot study

Romy Hansildaar[1,2], Max van Velzen[3], Eduard W. J. van der Vossen[3], Gertjan Kramer[4], Michael T. Nurmohamed[1] and Johannes H. M. Levels[3*]

Abstract

The risk of cardiovascular disease (CVD) in patients with rheumatoid arthritis (RA) is much higher than that in the general population. As its etiology is not fully understood, we performed a pilot study using a shotgun proteomic approach to investigate whether the plasma signature in RA patients with CVD might show an altered profile. Subjects with RA were compared to a group of RA patients with a previous cardiovascular event (CVE). The cohort consisted of an RA control group ($n=10$) and a group ($n=10$) of RA patients with a history of CVD. Samples were collected at least 6 months before the CVE and 3–6 months after the CVE. All subjects were matched to controls for age, sex, and medication use. Plasma depletion of the 14 most abundant proteins was followed by bottom-up shotgun proteomics analysis (LC–MS/MS). Relative changes in protein/peptide abundance were investigated using classical statistical analyses with Perseus and XG-Boost machine learning to compare between groups and to determine the relative importance of identified proteins, respectively. Principal component analysis (PCA) revealed no difference in the global protein and peptide signatures between the control and CVE groups. A total of 150, 239 and 74 protein ID's showed in comparison between Post Event vs. controls, Event vs. no Event and Pre event vs. Post Event respectively a statistically difference in relative abundance ($p<0.05$). Remarkedly a total of 236 proteins ID's showed a statistical significant difference in relative abundance in the PRE-Event group compared to the control group which could also be confirmed by XGboost machine learning. Here, we demonstrated potential differences in the plasma proteome signature of rheumatic patients with cardiovascular events. Interestingly, this signature may be present prior to CVE's. However the conclusions must be drawn with caution, since this is a pilot study and further investigation with larger cohorts is warranted to identify potential risk markers that may predict the relative risk of CVEs in rheumatic diseases.

Keywords RA, Plasma proteomics, Machine learning

*Correspondence:
Johannes H. M. Levels
h.levels@amsterdamumc.nl
[1]Amsterdam, Rheumatology and Immunology Center, Reade, Amsterdam, The Netherlands
[2]Amsterdam, Rheumatology and Immunology Center, Department of Rheumatology, Amsterdam UMC, Amsterdam, The Netherlands
[3]Department of Experimental Vascular Medicine, G1-142, Academic Medical Center of the University of Amsterdam, Meibergdreef 15, Amsterdam, AZ 1105, The Netherlands
[4]Laboratory for Mass Spectrometry of Biomolecules, Swammerdam Institute for Life Science, University of Amsterdam, Amsterdam, The Netherlands

Introduction

Inflammatory rheumatic diseases, particularly rheumatoid arthritis (RA), are recognized as autoimmune and/or inflammatory disorders with more than 200 known manifestations [1, 2]. Initially, RA causes inflammation in the joints, resulting in pain, stiffness, and swelling, but other parts of the body, such as the eyes, lungs, and heart, can also be affected, making it a serious systemic disease [3]. To date, rheumatic diseases affect approximately 20 million patients worldwide, and a higher incidence of cardiovascular events (CVEs) affecting and damaging the heart muscle and vasculature around the heart has been detected [4]. Despite the use of a number of markers for predicting RA disease progression/activity, such as C-reactive protein (CRP) [5], the erythrocyte sedimentation rate (ESR), rheumatoid factor (RF) [6], anti-cyclic citrullinated peptide (anti-CCP) [7, 8], and matrix metalloproteinases (MMPs) [9], the exact relationships between rheumatic patients and CVEs are still unclear [10]. The ratio of platelets to lymphocytes, an inflammatory marker in rheumatic diseases, may serve as a measure of the inflammatory status of a patient [11]. In combination with other predictors, including common proinflammatory cytokines, such as interleukin-6 (IL-6), tumor necrosis factor alpha (TNF-alpha), and IL-10 and lipid homeostasis [12–14], these inflammatory markers may serve as surrogate markers for predicting atherosclerotic risk [15]. Several epidemiological studies have reported an overall increase in cardiovascular mortality in patients with RA. However, the causality can still not be fully explained by traditional risk factors [4, 6]. Several studies have suggested that the baseline CRP concentration at baseline is an important predictor of CVD mortality in patients with new-onset inflammatory polyarthritis [16, 17]. Thus, studying the relationship between rheumatic diseases and CVEs should contribute to a better understanding of the etiology of CVD in patients with a rheumatic background. This may eventually allow us to fine-tune its treatment [18].

The plasma proteome is a vast and complex, readily accessible cocktail of proteins and potential markers, that may vary in abundance depending on the disease status of a patient with RA [19–22]. Systems biology approaches have proven to be a promising valuable tool for studying the plasma proteome/metabolome and could provide new insights in the search for associations between CVD risk and potential markers of interest [20, 23–26]. These markers of interest can originate from different locations on the body. Primarily, they may be directly derived from plasma proteins and/or metabolites but might also originate from damaged surrounding tissues that leak into the blood circulation [4, 6, 27].

Here, we performed a pilot plasma proteomic study in a unique, well-matched RA patient group to evaluate feasibility and identify pathways/markers that may lead to a better prediction of CVD risk.

Materials & methods

Patient samples/study design

This retrospective pilot study was conducted on plasma from patients diagnosed with RA, consisting of a group that developed CVEs and an age- and sex-matched RA control group that was considered healthy with respect to CVEs.

Patients were recruited using Reade's patient medical records and biobank database, which consists of data from subjects who had previously participated in other RA trials and signed an informed consent form.

First, we looked for a recorded CVE with a precise date in the medical history and/or medical letters. To verify this, we also used the pharmacy history medication list to check whether the patient actually received medication and whether the date was correct. We searched for P2Y12 inhibitors, such as Ticagrelor, Perstantin, and Clopidogrel (which are used for the treatment of myocardial infarction or cerebral infarction). This resulted in approximately 279 patients. These 279 patients were then searched in the biobank database for samples from at least 6 months BEFORE and 3–6 months AFTER the event. This resulted in 10 patients who met the above criteria. We then matched the 10 RA patients who were diagnosed with a cardiovascular event by age and sex with 10 control RA patients who did not have an event. The medical records of the latter group were similarly reviewed, with no evidence of an experienced cardiovascular event, which was further verified by medication use. All blood samples were drawn into conventional tubes containing ethylenediaminetetraacetic acid (EDTA)/K2EDTA and were stored in the biobank at -80 °C until further processing. The study was conducted in accordance with the Declaration of Helsinki and approved by the local medical ethics committee of Slotervaart Hospital. All patients provided written informed consent.

Plasma depletion (top 14 most abundant proteins)

Plasma depletion of highly abundant proteins is essential for increasing total protein coverage prior to LC–MS/MS analysis [28]. Liquid chromatography (Äkta Explorer 10 s, GE, Mijdrecht, The Netherlands) was performed with a Multi Affinity Removal Column (Human-14HC 4.6 × 50 mm, Agilent Technologies, Santa Clara, US). The column contained immobilized antibodies to capture the top 14 most abundant plasma proteins, which accounted for more than 90–95% of the total protein mass. Prior to injection, the system was rinsed with the equil/load/wash buffer A (pH 7.0) (Agilent Technologies, Santa Clara, US). A plasma sample was diluted 1:4 with buffer A, and 100 µL of sample was injected at a flow rate of 0.125 ml/min

of buffer A. The flow through (containing the depleted plasma) was collected in different fractions in a 96-well plate (250 µL/well). Next, a buffer switch (Buffer B pH 3.0) (Agilent Technologies, Santa Clara, US) combined with a flow rate increase of up to 1 mL/min was applied. This resulted in the release of the 14 most abundant proteins captured (Supplemental Fig. 1). The wells containing the flow-through proteins were pooled (4 wells per run, 2 runs per sample) for further processing.

Fraction concentration and trypsin gold digestion

After concentration, the sample was subjected to protein digestion using the rapigest-SF method with trypsin gold A total of 50 µg of total protein was pipetted into a 0.5 mL Eppendorf low-binding tube to a final volume of 116 µL using ULC/ms grade water (Biosolve BV, Valkenswaard, the Netherlands). Subsequently, 15 µL of 10% acetonitrile (ACN) (Biosolve BV, Valkenswaard, Netherlands), 15 µL of 1% Rapigest SF Surfactant (Waters, Milford, US) diluted in 50 mM ammonium bicarbonate (Sigma–Aldrich, Saint Louis, US), and 1.5 µL of 0.5 M dithiothreitol (DTT) (Sigma Aldrich, Saint Louis, US) were added. Next, 3 µL of 1 mg/ml alcohol dehydrogenase (Yeast, Sigma Aldrich, and Saint Louis, US) dissolved in saline was added as the internal standard.

This mixture was then incubated at 60 °C for 30 min to allow denaturation of the protein disulfide bonds. DTT was added to prevent reformation of the disulfide bonds. After incubation, 1.5 µL of 1 M iodoacetamide (IAA) (Sigma Aldrich, Saint Louis, US) was added, and the mixture was incubated in the dark at room temperature for 30 min to prevent reformation of disulfide bonds. After incubation, 1.0 µL of trypsin gold (Promega, Madison, US) was added to digest the protein content at 37 °C overnight, and after incubation, 7.9 µL of 10% trifluoroacetic acid (TFA) (Sigma Aldrich, Saint Louis, US) was added to lower the pH, thereby inactivating the trypsin. Finally, all samples were centrifuged at 14,750×g for 2 min, and the supernatant containing the peptide mixture was pipetted into a new 0.5 mL low binding tube for further purification.

C18 pipette-based solid phase extraction (SPE) and peptide quantification

To initially activate the pipette-based C18 column (Agilent Bond Elut OMIX Pipette-based SPE, Agilent Technologies, Santa Clara, US), 100 µL of 50% acetonitrile (ACN) (Biosolve, Valkenswaard, Netherlands) was pipetted up and down and finally eluted into a waste tube. After activation, the C18 tip was rinsed three times with 100µL of 0.1% formic acid LC−MS grade (FA) (Thermo Fisher Scientific, Waltham, US). Then, 100 µL of sample mixture was loaded onto the C18 tip and eluted into another low-binding tube. After loading, the C18 tip was "washed" 3 times, and the purified peptides were eluted into a clean low-binding tube by loading 100 µL of 75% acetonitrile (ACN) (Biosolve, Valkenswaard, Netherlands). Quantification of all eluted peptides was performed by using the Pierce^{tm} Quantitative Colorimetric Peptide Assay (Thermo Scientific, Rockford, US) in a 96-well plate. Subsequently, all samples were freeze-dried overnight and were resolubilized in 10% acetonitrile (ACN) (Biosolve, Valkenswaard, Netherlands)/0.1% trifluoroacetic acid (TFA) (Sigma Aldrich, Saint Louis, US)/ULC/89.9% ULC/MS grade water (Biosolve BV, Valkenswaard, the Netherlands) at such a volume to reach a final peptide concentration of 100 ng/µL.

LC–MS/MS analysis

Mass spectrometric analysis was carried out on a TIMS-TOF Pro (Bruker, Bremen, Germany) instrument equipped with an Ultimate 3000 nanoRSLC UHPLC system (Thermo Scientific, Germeringen, Germany) [29]. Specifically, a total of 200 ng (2 µL) of peptide per sample was injected onto a C18 column (75 m, 250 mm, 1.6 m particle size; Aurora, Ionopticks, Fritzroy, Australia) heated at 50 °C.

The sample was loaded at 400 nl/min for 2 min in 3% solvent B and separated using a multistep gradient: 6% solvent B for 55 min, 21% solvent B for 21 min, 31% solvent B for 12 min, 42.5% solvent B for 3 min and 99% solvent B for 7 min (solvent A: 0.1% formic acid in water, solvent B: 0.1% formic acid in acetonitrile). Analysis of the eluted peptides was performed using a time-of-flight mass spectrometer with a collision energy of 20–59 eVa. The precursor scan ranged from 100 to 1700 m/z, and the TIMS range was 0.6–1.6 V.s/cm2 in PASEF mode [29].

To monitor the accuracy of the mass spectrometer, a quality control (QC) sample was prepared by pooling 2 µL of all patient samples in one vial. After every 10 sample injections, the system was injected with this QC sample, allowing any changes in the absolute peak intensities to be observed over the time it took for the system to measure the 60 patient samples and 7 QC samples, thereby controlling for technical variation.

Normalization, data analysis and statistics

Validation of the dataset is critical prior to analysis. A quality control sample was included in the TIMS-TOF analysis to correct for machine bias, and a sample spike-in peptide (ADH) was added to correct for inter- and intra-peak variation. The tracing of the intensities of the QC samples over time was monitored (Supplemental Fig. 2). The mathematical function created to monitor intensities over time was manually set to a second-order polynomial function trend line, which allowed normalization of the entire dataset (Supplemental Fig. 3).

Spectral analysis was performed using SearchGUI (version 4.1.1) and a peptide shaker (version 2.2.0). Mass spectrometry protein identification was performed using MaxQuant software (version 1.6.14.0) and the Human Proteome Database from UniProt (© 2002–2023 UniProt consortium), which was used in combination with reversed decoy protein sequences for false discovery rate (FDR) estimation of protein identification [30].

The settings in MaxQuant were as follows: trypsin/P digestion enzyme, allowing for a maximum of 2 missed cleavages. The variable modifications were set to oxidation (M), and the fixed modifications were set to carbamidomethyl (C). Matching between runs was enabled with a matching time window of 12 s and a matching ion mobility window of 0.05 indices as the default setting for the TIMS-DDA. For label-free quantification, both iBAQ and LFQ were enabled.

The resulting intensities were normalized using RStudio (version 2021.09.0 Build 351 with R (version 4.1.1) to implement and normalize the matrix by the calculated function. Spike protein normalization was also performed in R-studio. The code used to normalize these files can be found in supplementary file 1. Plot visualization and t tests were performed using Perseus (version 2.0.11). The Benjamini-Hochberg FDR method was used for correction for multiple testing and was set to 0.05. Pathway analysis was performed using the pathway analysis tool of the Reactome database [31].

Machine learning analysis

The eXtreme Gradient Boosting (XGBoost) algorithm was used to select a panel of proteins that showed the best ability to predict the difference between the control group and the CVD group that experienced a CVD event. To reduce the complexity of the analysis, a filtering process (using the ANOVA F value) was applied to select the top 50 most important proteins. A stability selection process was then applied to prevent overfitting and ensure the robustness of the results. The complete dataset was randomly divided into 20 different subsets. Each of the subsets consisted of 85% of the data. Within each random subset, leave-one-out cross-validation was applied. The training set included all samples but one sample, and the sample that was left out was included in the test set. The hyperparameters of the XGBoost classifier were optimized via a randomized search with triple cross-validation. For each random search, 10 random parameter settings were tested. The training set consisted of 70% of the data, and the validation set contained the remaining 30%. The splits were made using a stratified shuffle split. The final performance of the model was evaluated using the area under the curve (AUC) metric. The importance of each feature in the model was determined by calculating the mean decrease in impurity, in which the most important feature was scaled to 100%. The other feature importance levels are relative to the most important feature. All of these steps were performed using Python (v3.7.7) and the scikit-learn package (v0.23.1) (Program code available on request). Pathway analysis was performed by using Gene Ontology and Reactome applications [32].

Results

Patients

In this study, we investigated the plasma proteome of patients with RA or RA in addition to CVEs, and we hypothesized that differences in the plasma fingerprint would predict the cardiovascular outcome of the subjects based on their cardiovascular history. A total of 20 RA patients were enrolled: 10 patients with a recent (3–6 months) history of CVEs were matched with patients of sex and age. The baseline table (Table 1) shows that the ages of both groups were the same (63.3 years). Half of both groups consisted of women. The disease duration for the control group, with a median of 15.7, was much higher compared to the CVE group (9.7). In addition, other disease characteristics, such as anti-citrullinated protein autoantibodies (ACPA) and rheumatoid factor (RF), as well as the use of RA-related medications, were more or less similar in both groups. In the CVE group, half of the patients had dyslipidemia and were treated with statins. Notably, half of the patients in the non-CVE group used platelet aggregation inhibitors. Due to missing data in the patients' records, smoking status, alcohol consumption, BMI and physical activity level could not be included in this table.

Data handling and analysis

After LC–MS/MS analysis and subsequent protein identification, the raw intensities and peptides were summarized in a matrix file (Available on request). The protein ID file included a total of 429 identified proteins, and the peptide file revealed 5335 identified sequences corresponding to a total of 580 unique proteins (Data available on GitHub, Link: https://github.com/jhlevelsGIT/Plasma-proteome-analysis-of-RA-patients-reveals-CVA-history-dependent-differences-in-fingerprint). All the matrices were first cleaned and subsequently analyzed by using Rstudio. With the use of the "cleaning" codes (Supplemental attachment 1 and 2), all peak intensity files were normalized and transformed for further analysis in Perseus.

To visualize an overview of potential differences between the CVE and non-CVE groups for this complex dataset, principal component analysis (PCA) was performed to generate 2D plots. These plots combine the intensities into components that best describe the locations of many different intensities (Fig. 1). The centroids

Table 1 Baseline characteristics

Patient Characteristics	All patients (*n*=20)	CVE (*n*=10)	Controls (*n*=10)
Demographics			
Age, years, mean (sd)	63,3± 11,3	63,4± 11,9	63,1± 11,8
Woman no (%)	10 (50%)	5 (50%)	5 (50%)
RA disease characteristics			
Disease duration, months	11,2 (7,5-16,7)	9,7 (5,7-16,1)	15,7 (8,5-16,4)
RF positive, n (%)	16 (80%)	8 (80%)	8 (80%)
ACPA positive, n (%)	16 (80%)	8 (80%)	8 (80%)
NSAID	7 (35%)	4 (40%)	3 (30%)
csDMARD	19 (95%)	10 (100%)	9 (90%)
bDMARD	19 (95%)	9 (90%)	10 (100%)
CV Risk factors			
Hypertension	10 (50%)	5 (50%)	5 (50%)
Antihypertensive agents	10 (50%)	5 (50%)	5 (50%)
Dyslipidaemia	6 (30%)	5 (50%)	1 (10%)
Statins	6 (30%)	5 (50%)	1 (10%)
Diabetus mellitus	3 (10%)	0 (0%)	2 (20%)
Antidiabetics	1 (5%)	0 (0%)	1 (10%)
Platelet aggregation inhibitors	15 (75%)	10(100%)	5 (50%)
Dual antiplatelet therapy *(acetylsalicylic acid **AND** P2Y12 inhibitor)*	10(50%)	10(100%)	0(0%)

Data are presented as mean ± SD, median (IQR: interquartile range) or number (percentage) where appropriate. ACPA: Anti-citrullinated protein autoantibodies, RF: rheumatoid factor, DMARD: disease-modifying anti-rheumatic drug, bDMARD: biological DMARD, csDMARD: conventional synthetic DMARD, NSAIDs: non-steroidal anti-inflammatory drugs

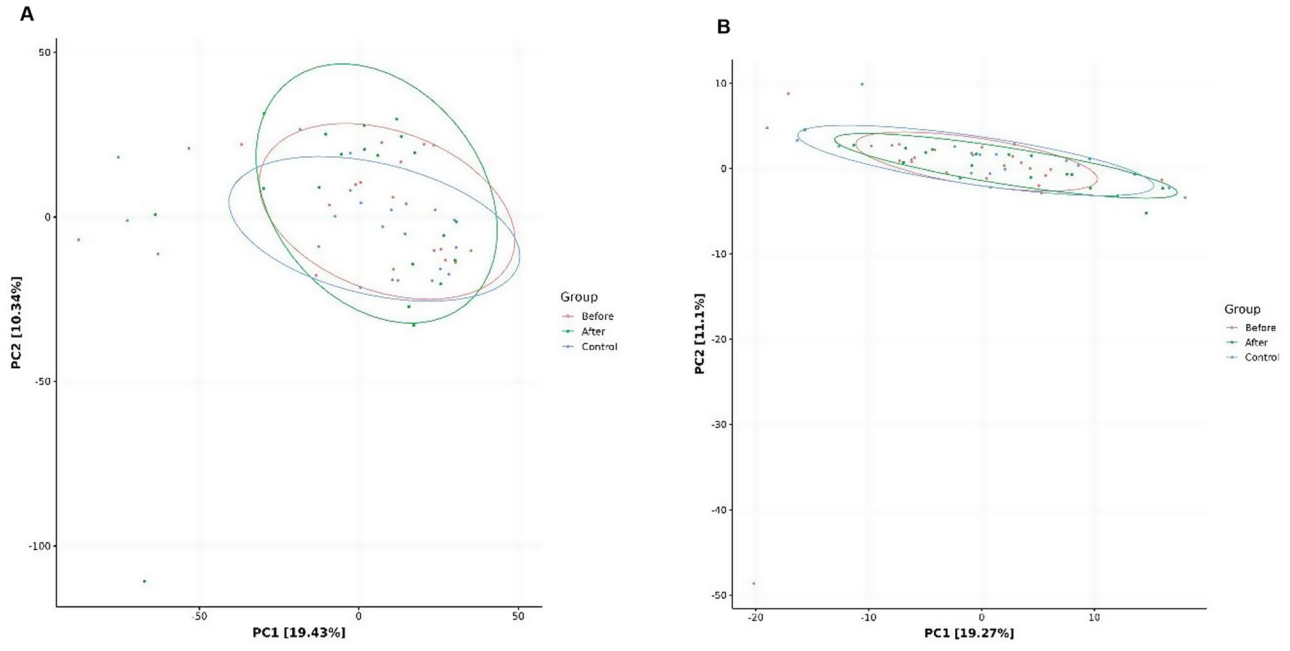

Fig. 1 PCA plots showing centroids of the (**A**) peptide database matrix and (**B**) protein database matrix before (red), after (green) the CVE and the controls (blue). The centroids show no trend in the difference in distribution of each group (Adonis function, permutational ANOVA applied) resulting in a p-value of (**A**) 0.6982 and (**B**) 0.853 respectively

represent the total potential differences between the groups based on the peptide intensities (A) and protein identifications (B). The analysis of variance of the PCA plots is visualized using distance matrices, resulting in an ANOVA table. Although this function produces no significant values, the plots show a potential difference between the CVE and non-CVE groups in both the peptide and protein datasets. PCA plot visualization is only able to provide a general overview and lacks detailed information on the individual identified proteins. To

obtain a more detailed view of the dataset, further analysis by examining individual intensities using multivariate analysis was evaluated as a next step.

Patient study statistical analysis

More detailed visualization at the protein level could be achieved after cleaning and normalization of both the protein and peptide matrices, which were then read, transformed and visualized by the Perseus application. The transformation of intensities to a log2(x) scale allows for much better behavior in statistical tests due to the normal distribution of the data. With these transformed and normally distributed data, paired and unpaired t tests were used to compare the means of 2 groups. These tests were then visualized in a volcano plot showing the log2-fold change on the x-axis, which is similar to the difference in intensity measured between the groups (Fig. 2). In all four of the comparison plots, a substantial amount of proteins demonstrated a difference in abundance in the depicted comparisons. This indicates that the conditions of the patients appear to have an effect on protein expression and/or protein levels in the plasma. A comprehensive summary of all protein ID's with a p value < 0.05 in the assessed comparison analyses of the 4 different conditions has been provided in supplementary file 3.

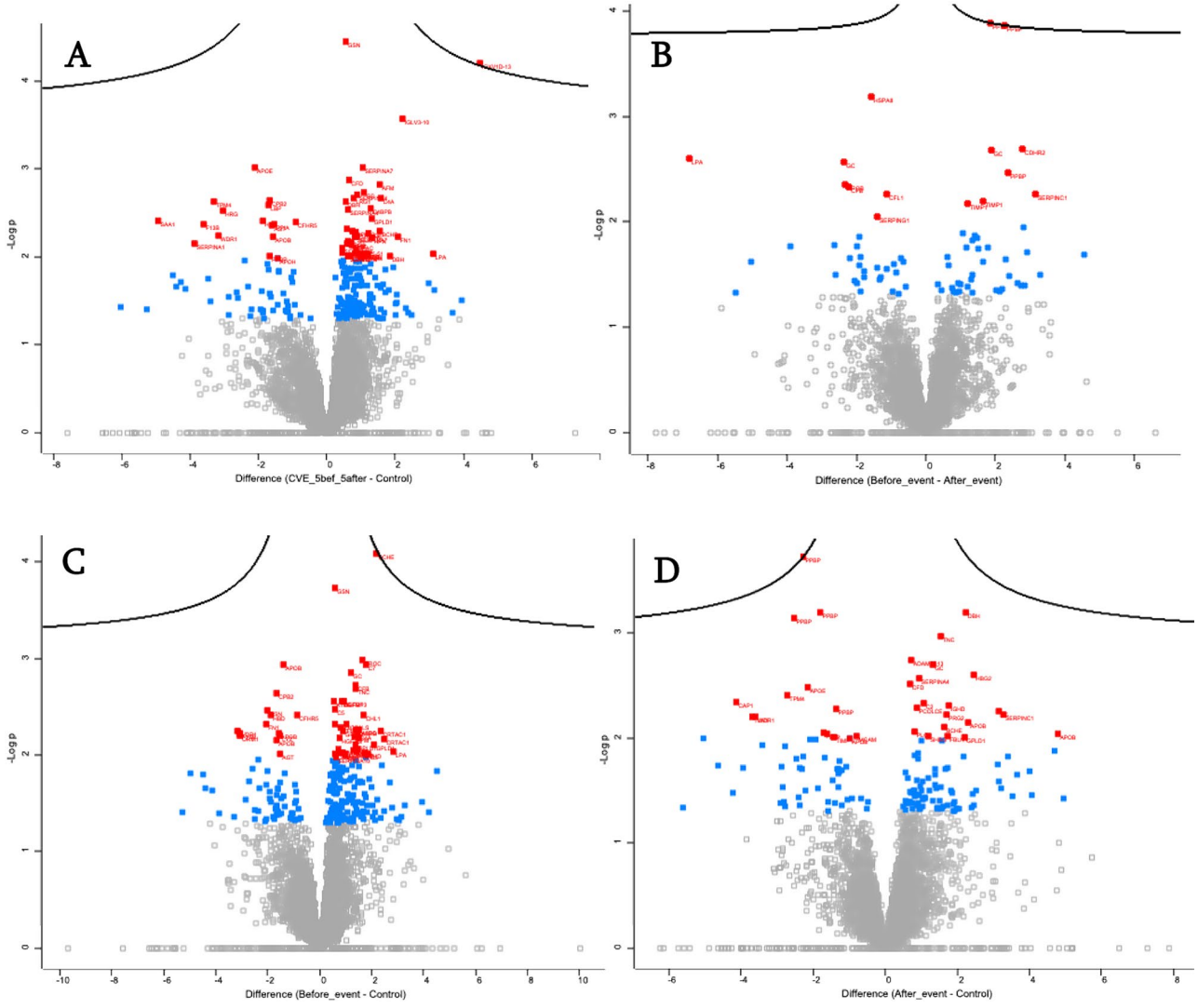

Fig. 2 Volcano plots of the–$\log_{10}$ (p-value) versus the $\log_2$ fold change of peptide intensity, to compare groups a two-sample Student's T-Test with a 0.05 false discovery rate (FDR) is used to indicate significance between given groups. Significant values (p-value < 0,05) are indicated as filled circles (blue) and (p-value < 0.01) filled (red) squares. The total overview of all significant proteins can be found in supplemental file 3. **(A)** T-test with CVD compared to without CVD (239 sign. proteins), **(B)** paired T-test Before CVD group compared to After CVD (74 sign. proteins), **(C)** T-test before CVD compared to control group (236 sign. proteins), **(D)** T-test After CVD compared to control group (150 significant proteins)

Pathway analysis

To put these results into a general perspective, a Reactome pathway analysis was performed with all of the significantly different proteins that emerged from the statistical analysis. The protein IDs can be read and traced back to a pathway and its significance in the list given (Table 2). This overrepresentation analysis of the significant proteins provided a list of proteins that could be linked to a set of significantly altered pathways. In addition, Gene Ontology analysis (GO analysis) was performed only on the significant genes identified in the groups with and without a proven CVE. The percentages of significant proteins and their corresponding biological processes are shown in Fig. 3. By comparing the GO analysis with the Reactome pathway analysis, most of the genes associated with the significant proteins found were hits related to stress response and defense processes. It is also noteworthy that more than half of the identified proteins that are significantly different in samples that suffer from a CVE are related to the stress response.

Machine learning

Additionally a machine learning approach, the XGBoost model, was utilized to discriminate and confirm any significant differences between the 4 different comparisons as demonstrated in Fig. 2. Between the control group and the pre-CVE group the model showed discriminative performance (AUC 0.72; Fig. 4) as for all the other models no discrimate performance was reached (Not shown). A set of features identified through the Gini importance was then obtained (Supplemental Fig. 4), and a difference in the 15 most important features between the two groups was observed (Fig. 5). Univariate analyses of these 15 features revealed significant differences between the

Table 2 Pathway analysis using the reactome pathway analysis, these are the most significant pathways sorted by p-value. (A) significant proteins in CVD compared to no CVD, (B) significant proteins in Pre-CVE group compared to post-CVE, (C) significant proteins in the pre-CVE compared to the control group, (D) significant proteins in post-CVE group compared to the control group

A

Pathway name	Entities			
	found	ratio	p-value	FDR*
Platelet degranulation	12 / 185	0.009	4.41e-13	2.84e-10
Response to elevated platelet cytosolic Ca2+	12 / 250	0.012	1.41e-11	4.54e-09
Post-translational protein phosphorylation	8 / 109	0.005	1.72e-09	3.68e-07
Regulation of Insulin-like Growth Factor (IGF) transport and uptake by Insulin-like Growth Factor Binding Proteins (IGFBPs)	10 / 257	0.012	5.92e-09	9.53e-07
Activation of C3 and C5	5 / 42	0.002	2.14e-07	2.74e-05
Complement cascade	9 / 256	0.012	1.30e-05	0.001
Platelet activation, signaling and aggregation	14 / 890	0.041	1.33e-05	0.001

B

Pathway name	Entities			
	found	ratio	p-value	FDR*
Platelet degranulation	3 / 185	0.009	1.25e-05	0.005
Response to elevated platelet cytosolic Ca2+	3 / 250	0.012	3.07e-05	0.006
Post-translational protein phosphorylation	2 / 109	0.005	3.81e-04	0.053
Platelet activation, signaling and aggregation	5 / 890	0.041	0.001	0.081
Hemostasis	6 / 2,401	0.112	0.002	0.081
Regulation of Insulin-like Growth Factor (IGF) transport and uptake by Insulin-like Growth Factor Binding Proteins (IGFBPs)	2 / 257	0.012	0.002	0.081
Interleukin-4 and Interleukin-13 signaling	2 / 351	0.016	0.004	0.081

C

Pathway name	Entities			
	found	ratio	p-value	FDR*
Complement cascade	11 / 256	0.012	8.11e-12	4.20e-09
Regulation of Complement cascade	10 / 232	0.011	7.79e-11	2.02e-08
Post-translational protein phosphorylation	6 / 109	0.005	1.54e-07	2.64e-05
Terminal pathway of complement	3 / 9	4.19e-04	1.18e-06	1.50e-04
Regulation of Insulin-like Growth Factor (IGF) transport and uptake by Insulin-like Growth Factor Binding Proteins (IGFBPs)	8 / 257	0.012	1.45e-06	1.50e-04
Activation of C3 and C5	6 / 42	0.002	2.44e-06	2.10e-04
Platelet degranulation	6 / 185	0.009	3.24e-06	2.40e-04
Response to elevated platelet cytosolic Ca2+	6 / 250	0.012	1.78e-05	0.001

D

Pathway name	Entities			
	found	ratio	p-value	FDR*
Regulation of Complement cascade	29 / 232	0.011	1.11e-16	3.01e-14
Complement cascade	30 / 256	0.012	1.11e-16	3.01e-14
Platelet degranulation	29 / 185	0.009	1.11e-16	3.01e-14
Response to elevated platelet cytosolic Ca2+	29 / 250	0.012	1.11e-16	3.01e-14
Post-translational protein phosphorylation	16 / 109	0.005	2.22e-15	4.82e-13
Regulation of Insulin-like Growth Factor (IGF) transport and uptake by Insulin-like Growth Factor Binding Proteins (IGFBPs)	19 / 257	0.012	8.50e-12	1.54e-09
Initial triggering of complement	15 / 153	0.007	6.35e-11	9.84e-09

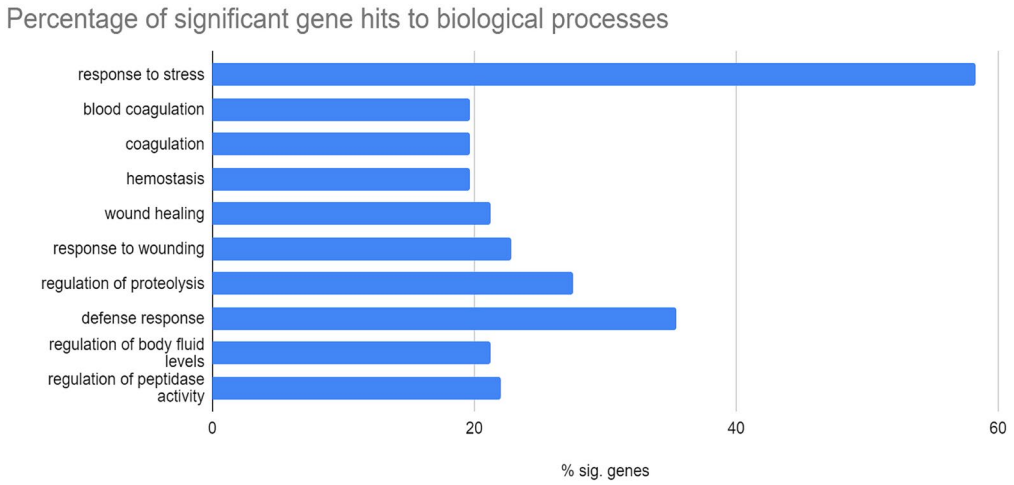

Fig. 3 Gene Ontology analysis [32] of significantly different genes between samples that suffered a cardiovascular event and samples that did not. The percentage relates to the % of significant proteins as identified in the data set which can be associated to the different biological processes as indicated

groups (Fig. 6). Notably, 11 of the 15 DEGs were lower in the pre-CVE group than in the matched controls.

Discussion

In this pilot study, patients with RA were compared with a group of RA patients who experienced a cardiovascular event (CVE) in addition to their rheumatic background. A shotgun proteomic approach was used to investigate whether the plasma signature in the CVE group differed from that in the non-CVE group before and after CVE.

An important aspect of this pilot study is that we aimed to predict cardiovascular events by selecting patients based on treatment with dual antiplatelet therapy (acetylsalicylic acid and P2Y12 inhibitor), as P2Y12 inhibitors are initiated only after a cardiovascular event. comprehensive assessment of the medical records was conducted to verify the occurrence of a cardiovascular event, ensuring that the patient selection process was highly accurate However, in the control group, 50% had hypertension without evidence of a previous cardiovascular event, and this group was also using acetylsalicylic acid for primary prevention. A key strength of the study lies in the meticulous selection of patients through pharmacy data and medical records, which did not interfere with the primary objective of predicting cardiovascular events in RA patients.

Principal component analysis (PCA) revealed that the overall centroids tended to be at different positions. There may already be a difference in the signature between the CVE group and the non-CVE group, although this difference reached no statistical significance (Permanova pvalue > 0.1). It should be noted that in a complex system such as the plasma proteome, there may still be significant differences between individual proteins or sets of proteins. These nuanced differences cannot be captured by PCA, as this method aims to maximize the variance within the dataset on a global scale. To provide a more detailed view of the behavior of all identified individual proteins, a univariate analysis of relative abundance in the different comparisons was used. The plots in Fig. 4 show that many peptides together with their corresponding identified proteins appeared to be significantly different between the groups. Interestingly, the group of patients who suffered from a cardiovascular event showed an increase in platelet basic protein compared to the controls and demonstrated significantly greater intensity compared to the pre-CVE group. However, these findings are in line with the literature and are expected given the etiology of the development of a CVE, where platelet activation plays an initial and key role in the thrombus formation of atherothrombosis and the onset of a cardiac event [33, 34].

As shown in Fig. 4A and C, cholinesterase was more abundant in the non-CVE group than in the control group, but cholinesterase was still more abundant in the non-CVE group than in the “before CVE” group. The cholinesterase-matched peptides appeared to be more intense in the group of patients who did not suffer a cardiovascular event and may provide meaningful information for predicting a CVE. It would be interesting to further investigate the role of this enzyme in relation to the anti-inflammatory pathways with respect to cardiovascular risk [35].

Pathway analysis revealed a list of the top 10 most significant biological processes in which the identified protein set appeared to be involved. In all assays, the platelet degranulation pathway was remarkably abundant, even in non-CVE samples. In all the samples, significant pathways involving the complement cascade, platelet regulation and insulin growth factor regulation were detected. Notably, more than half of the significant genes were related to the response to stress. This may indicate that

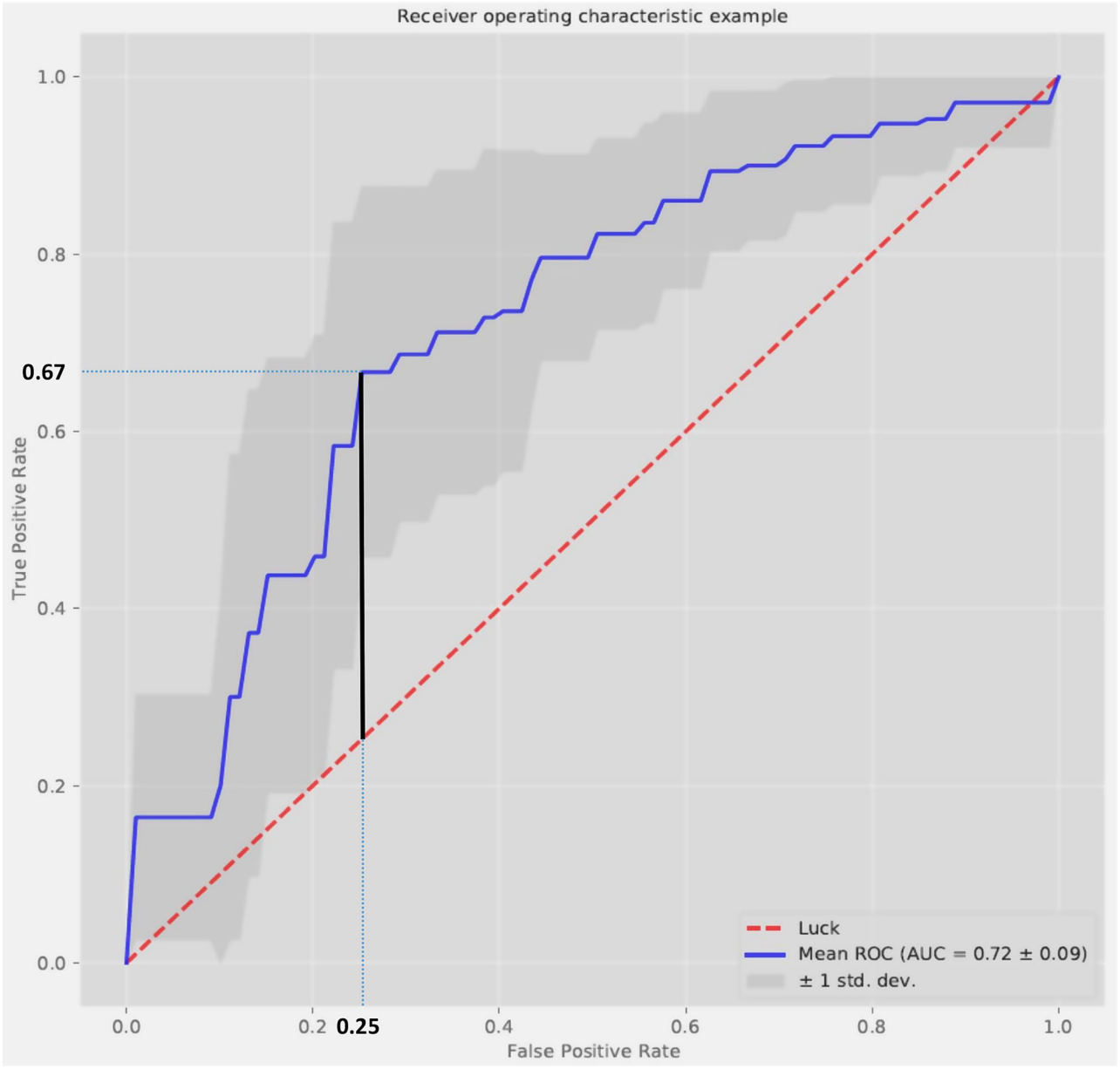

Fig. 4 ROC curve of the Control vs. the pre CVE group to predict a potential difference in plasma fingerprint between the control group and the pre-CVD group (AUC = 0.72 ± 0.09) as generated using R. At a treshhold of a J index of 0.42 (black line) a 67% Sensitivity and a 75% was reached

even before CVEs, a number of proteins are differentially expressed based on subclinical damage to the vasculature and/or surrounding tissues in terms of endothelial dysfunction, vessel wall inflammation, oxidative stress, apoptosis, coagulation and thrombosis [36].

Machine learning is a form of artificial intelligence in the growing field of data science and is necessary for the unbiased future study of (prote)omic datasets [37]. Here, we applied the XGBoost model, a type analysis that is slowly finding its way into the rheumatology research field as well [38]. Besides for confirmation this also allowed us to identify patterns which can not be seen by multivariate analysis and classical statistical tests used previously, such as the Perseus application used here.

In all the different comparisons, a top 15 most important protein sets could be identified, showing differences in fingerprints between the groups. As shown in the radar plots in the non-CVE group vs. the CVE group before the event, profound differences in protein fingerprints were observed which also revealed to have some potential predictive power as shown in the generated ROC curve. This actually confirms the findings determined by classical statistical approaches and is likely a good indication that there may be new markers or sets of markers that have

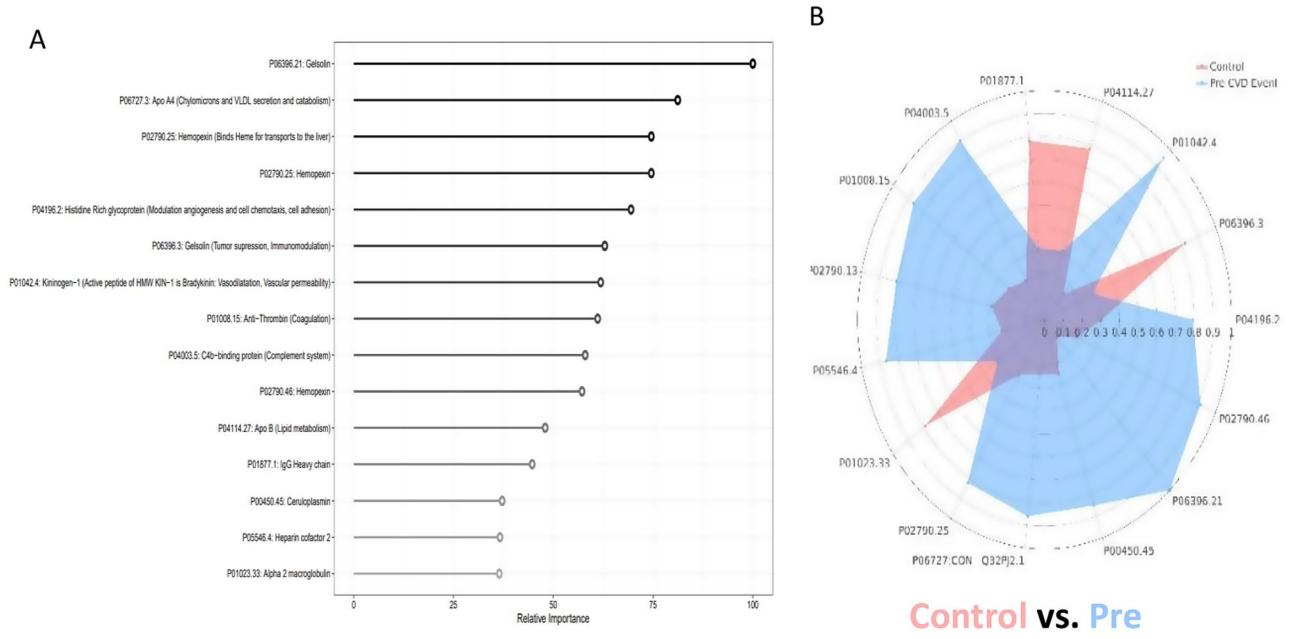

Fig. 5 **(A)** summary of the relative importance (X-axis in %) of the top 15 identified proteins (Y-axis) in the entire dataset determining the difference in fingerprint between the controls and the pre-CVE group. All based on Gini importance. **(B)** RADAR plot visualizing the analysis comparing the control vs. pre-CVE group showing the top 15 determining proteins of relative importance. The model used: XGboost leave one out

predictive value with regard to the risk of developing a CVE in the (near) future in an individual. This was nicely demonstrated in the presented ROC curve (AUC of 0.72). Notably, one of the top 10 protein IDs included was GSN. This protein has previously been associated with low abundance and increased disease activity in individuals with RA [39–41]. Among a variety of biological functions, including apoptosis, the GSN protein is a member of the GSN family and has important functions in cutting and sealing actin filaments. Recently, increasing evidence has shown that GSN is also closely related to atherosclerosis, which involves lipid metabolism, inflammation and thrombosis [42].

Some limitations of this study must be mentioned. Plasma depletion affects the completeness of the total number of proteins identified. There is a risk of losing proteins during sample preparation that form a complex with the depleted proteins in the first column step. On the other hand, after analysis of the depleted protein fraction by LC–MS/MS, a total of 20 confident proteins were identified, including the targeted top 14 (data not shown, available on request), implying that any excessive loss of nontargeted proteins seems to be very limited [28].

Despite the small overall sample size, the matched control design, accounting for age, medication use, and disease state, ensures that the comparisons drawn are valid and statistically robust. However, due to the limited sample size, these findings must be interpreted with caution. As this is a pilot study, further investigation with a larger cohort is crucial, but the results from this trial offer a promising direction for future research. It is essential to emphasize that these findings, derived from a small sample, should be considered exploratory, and definitive conclusions should be drawn from larger cohorts or an expanded sample size within the current study.

It should be recognized that comparing the groups with respect to different relative protein abundances and different protein patterns does not provide direct insight into causality. It will be of great interest to further investigate these findings in an independent new group of patients for confirmation by using a targeted approach of selecting the protein set deduced from this study.

In conclusion, we suggest that the increased risk of developing CVEs in RA patients may be reflected in the altered plasma fingerprint prior to CVEs and that better treatment should be achieved in a preventive, personalized manner. Our findings in this study may be a valuable contribution to the development of a more efficient personalized clinical treatment regarding the prevention of cardiovascular risk in the field of rheumatology.

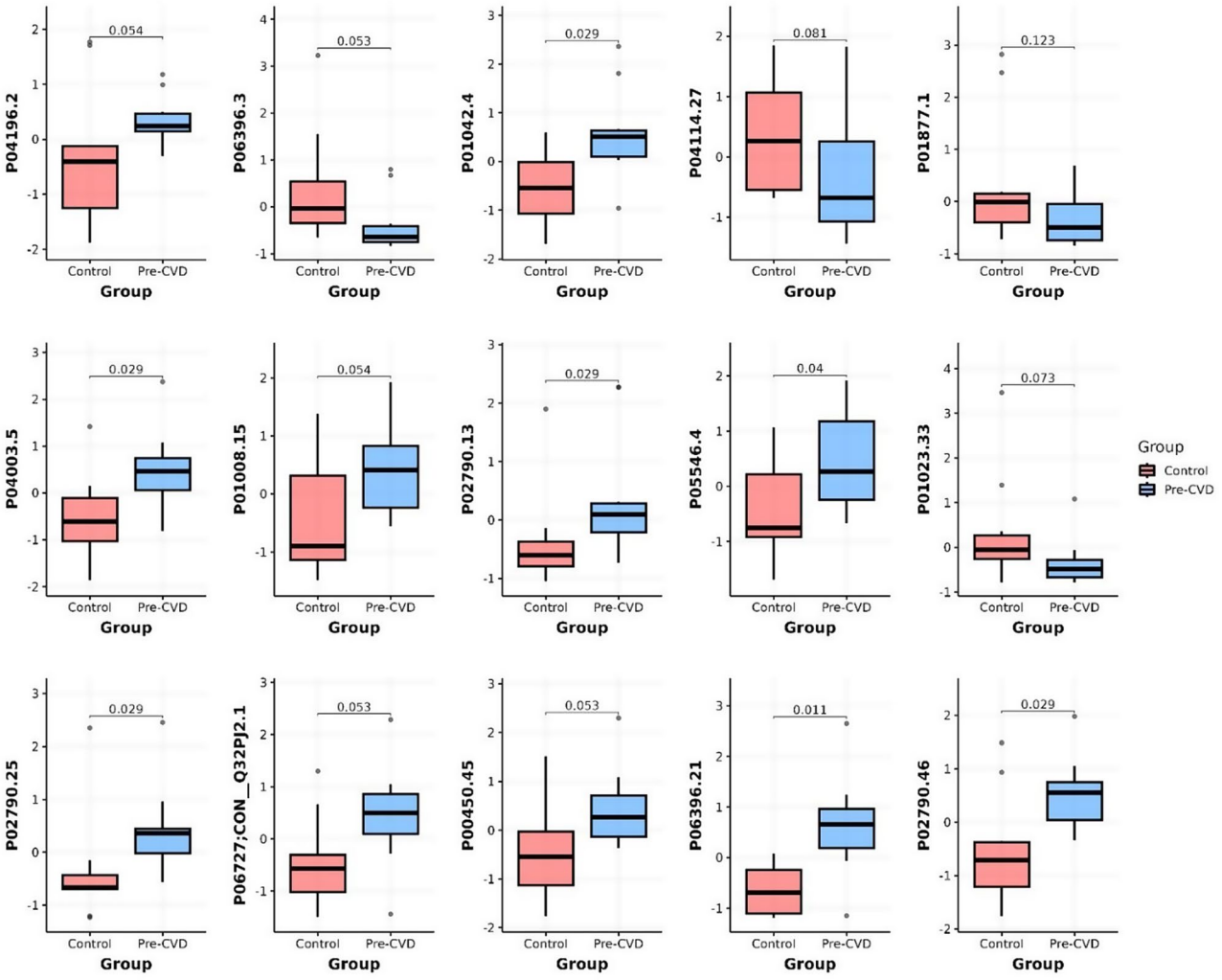

Fig. 6 Boxplots of the relative abundance of the 15 individual proteins ($n = 10$ per group) contributing to the difference in fingerprint between the control group versus the pre-CVD group. The Wilcoxon test was used and a correction for multiple measurements was applied to all calculated p values which are indicated above the box plots. The protein codes (y-axis) per box plot correspond with the protein list as summarized in Fig. 5A

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12953-025-00243-6.

Supplementary Material 1

Supplementary Material 2

Supplementary Material 3

Supplementary Material 4

Author contributions

MV, GJK were mainly involved in the experimental design and investigation. JHML and MTN were involved involved in the experimental design, resources and supervising the study. RH contributed in providing the resources, data curation and investigation. EWJV contributed to both the formal analysis, data curation and machine learning analysis. All authors contributed to writing and reviewing and revising the manuscript.

Funding

Not applicable.

Data availability

Data have been provided within the manuscript or supplementary information files or will be available on request as stated in the manuscript.

Declarations

Ethical approval

The study was conducted in accordance with the Declaration of Helsinki and received approval by the local medical ethics committee of the Slotervaart Hospital in Amsterdam, Netherlands. Furthermore, all patients provided written informed consent for this Rheumatology Registry which is an observational daily clinical practice inception cohort. It was initiated with the objective of capturing long-term data from patients using biologicals and targeted synthetic disease-modifying antirheumatic drugs (b/tsDMARDs) in inflammatory arthritis. This registry was started in 2004, and comprises RA, PsA and AS patients starting on a new b/tsDMARDs therapy. (NL59549.048.16).

Competing interests

The authors declare no competing interests.

Received: 22 August 2024 / Accepted: 26 March 2025
Published online: 11 April 2025

References

1. Quinteros A, Leal M, Vasquez D, Espindola M, Medina M, Garcia G, Mabroyrinis R. Reumatoid arthritis: education, a strategy for living better. Jcr-Journal Clin Rheumatol. 2010;16:S88–9.
2. Rada JAR, Garcia L, Puente LR. Evaluation of cardiovascular risk through the algorithms of Framingham 2008, score and Qrisk3 in patients with reumatoid arthritis (Evark). Jcr-Journal Clin Rheumatol. 2020;26:92–92.
3. Qin Q, Song R, Du P, Gao C, Yao Q, Zhang JA. Systemic Proteomic Analysis Reveals Distinct Exosomal Protein Profiles in Rheumatoid Arthritis. *J Immunol Res* 2021, 2021:9421720.
4. Lumngwena EN, Skatulla S, Blackburn JM, Ntusi NAB. Mechanistic implications of altered protein expression in rheumatic heart disease. Heart Fail Rev. 2022;27:357–68.
5. Kalani R, Bartz TM, Psaty BM, Elkind MSV, Floyd JS, Gerszten RE, Shojaie A, Heckbert SR, Bis JC, Austin TR, et al. Plasma proteomic associations with incident ischemic stroke in older adults: the cardiovascular health study. Neurology. 2023;100:e2182–90.
6. Arts EE, Popa CD, Den Broeder AA, Donders R, Sandoo A, Toms T, Rollefstad S, Ikdahl E, Semb AG, Kitas GD, et al. Prediction of cardiovascular risk in rheumatoid arthritis: performance of original and adapted SCORE algorithms. Ann Rheum Dis. 2016;75:674–80.
7. Elsawy NA, Ghazala RA, Elnemr R. Cartilage and bone loss in premenopausal women with rheumatoid arthritis: radiological and laboratory assessments. Int J Rheum Dis. 2023;26:2195–205.
8. Raslan HM, Attia HR, Hamed Ibrahim M, Mahmoud Hassan E, Salama II, Ismail S, Abdelmotaleb E, El Menyawi MM, Amr KS. Association of anti-cyclic citrullinated peptide antibodies and rheumatoid factor isotypes with HLA-DRB1 shared epitope alleles in Egyptian rheumatoid arthritis patients. Int J Rheum Dis. 2020;23:647–53.
9. Liu J, Li Y, Jiang Z, Liu Y, Wei Z. Protein arginine methyltransferase 1 upregulates matrix metalloproteinase-2/9 expression via Zeste homolog 2 to promote human rheumatoid arthritis fibroblast-like synovial cell survival and metastasis. Int J Rheum Dis. 2023;26:88–98.
10. Cuppen B, Fritsch-Stork R, Eekhout I, de Jager W, Marijnissen AC, Bijlsma J, Custers M, van Laar JM, Lafeber F, Welsing P. All society for rheumatology research Utrecht i: proteomics to predict the response to tumour necrosis factor-alpha inhibitors in rheumatoid arthritis using a supervised cluster-analysis based protein score. Scand J Rheumatol. 2018;47:12–21.
11. Gasparyan AY, Ayvazyan L, Mukanova U, Yessirkepov M, Kitas GD. The Platelet-to-Lymphocyte ratio as an inflammatory marker in rheumatic diseases. Ann Lab Med. 2019;39:345–57.
12. Rauber S, Luber M, Weber S, Maul L, Soare A, Wohlfahrt T, Lin NY, Dietel K, Bozec A, Herrmann M, et al. Resolution of inflammation by interleukin-9-producing type 2 innate lymphoid cells. Nat Med. 2017;23:938–44.
13. Kasher M, Cherny SS, Group CIW, Livshits G. Exploring potential shared genetic influences between rheumatoid arthritis and blood lipid levels. Atherosclerosis. 2022;363:48–56.
14. Charles-Schoeman C, Gugiu GB, Ge H, Shahbazian A, Lee YY, Wang X, Furst DE, Ranganath VK, Maldonado M, Lee T, Reddy ST. Remodeling of the HDL proteome with treatment response to abatacept or adalimumab in the AMPLE trial of patients with rheumatoid arthritis. Atherosclerosis. 2018;275:107–14.
15. Robinson WH, Mao R. Biomarkers to guide clinical therapeutics in rheumatology? Curr Opin Rheumatol. 2016;28:168–75.
16. Pope JE, Choy EH. C-reactive protein and implications in rheumatoid arthritis and associated comorbidities. Semin Arthritis Rheum. 2021;51:219–29.
17. Kaplan MJ. Cardiovascular disease in rheumatoid arthritis. Curr Opin Rheumatol. 2006;18:289–97.
18. Raadsen R, Agca R, Boers M, van Halm VP, Peters MJL, Smulders Y, Beulens JWJ, Blom MT, Stehouwer CDA, Voskuyl AE, et al. In RA patients without prevalent CVD, incident CVD is mainly associated with traditional risk factors: A 20-year follow-up in the CARRE cohort study. Semin Arthritis Rheum. 2023;58:152132.
19. Deutsch EW, Omenn GS, Sun Z, Maes M, Pernemalm M, Palaniappan KK, Letunica N, Vandenbrouck Y, Brun V, Tao SC, et al. Advances and utility of the human plasma proteome. J Proteome Res. 2021;20:5241–63.
20. Mun S, Lee J, Park A, Kim HJ, Lee YJ, Son H, Shin M, Lim MK, Kang HG. Proteomics approach for the discovery of rheumatoid arthritis biomarkers using mass spectrometry. Int J Mol Sci 2019, 20.
21. Birkelund S, Bennike TB, Kastaniegaard K, Lausen M, Poulsen TBG, Kragstrup TW, Deleuran BW, Christiansen G, Stensballe A. Proteomic analysis of synovial fluid from rheumatic arthritis and spondyloarthritis patients. Clin Proteom. 2020;17:29.
22. Lind L, Titova O, Zeng R, Zanetti D, Ingelsson M, Gustafsson S, Sundstrom J, Arnlov J, Elmstahl S, Assimes T, Michaelsson K. Plasma protein profiling of incident cardiovascular diseases: A multisample evaluation. Circ Genom Precis Med. 2023;16:e004233.
23. He B, Huang Z, Huang C, Nice EC. Clinical applications of plasma proteomics and peptidomics: towards precision medicine. Proteom Clin Appl. 2022;16:e2100097.
24. Palstrom NB, Matthiesen R, Rasmussen LM, Beck HC. Recent developments in clinical plasma Proteomics-Applied to cardiovascular research. Biomedicines. 2022;10:162.
25. Zoanni B, Brioschi M, Mallia A, Gianazza E, Eligini S, Carini M, et al. Novel insights about albumin in cardiovascular diseases: focus on heart failure. Mass Spectrom Rev. 2021;42:1113–1128:e21743.
26. Park YJ, Chung MK, Hwang D, Kim WU. Proteomics in rheumatoid arthritis research. Immune Netw. 2015;15:177–85.
27. Ferreira MB, Fonseca T, Costa R, Marinhoc A, Carvalho HC, Oliveira JC, Zannad F, Rossignol P, Gottenberg JE, Saraiva FA, et al. Prevalence, risk factors and proteomic bioprofiles associated with heart failure in rheumatoid arthritis: the RA-HF study. Eur J Intern Med. 2021;85:41–9.
28. Kaur G, Poljak A, Ali SA, Zhong L, Raftery MJ, Sachdev P. Extending the depth of human plasma proteome coverage using simple fractionation techniques. J Proteome Res. 2021;20:1261–79.
29. Huang Y, Swarge BN, Roseboom W, Bleeker JD, Brul S, Setlow P, Kramer G. Integrative metabolomics and proteomics allow the global intracellular characterization of Bacillus subtilis cells and spores. J Proteome Res. 2024;24:596–608.
30. Barsnes H, Vaudel M. SearchGUI: A highly adaptable common interface for proteomics search and de Novo engines. J Proteome Res. 2018;17:2552–5.
31. Rothfels K, Milacic M, Matthews L, Haw R, Sevilla C, Gillespie M, Stephan R, Gong C, Ragueneau E, May B, et al. Using the reactome database. Curr Protoc. 2023;3:e722.
32. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, et al. Gene ontology: tool for the unification of biology. The gene ontology consortium. Nat Genet. 2000;25:25–9.
33. Asada Y, Yamashita A, Sato Y, Hatakeyama K. Thrombus formation and propagation in the onset of cardiovascular events. J Atheroscler Thromb. 2018;25:653–64.
34. Liu Y, Jiang H, Kang T, Shi X, Liu X, Li C, Hou X, Li M. Platelets-related signature based diagnostic model in rheumatoid arthritis using WGCNA and machine learning. Front Immunol. 2023;14:1204652.
35. Yin Q, Wu YJ, Pan S, Wang DD, Tao MQ, Pei WY, Zuo J. Activation of cholinergic Anti-Inflammatory pathway in peripheral immune cells involved in therapeutic actions of alpha-Mangostin on Collagen-Induced arthritis in rats. Drug Des Devel Ther. 2020;14:1983–93.
36. Badimon L, Suades R, Arderiu G, Pena E, Chiva-Blanch G, Padro T. Microvesicles in atherosclerosis and angiogenesis: from bench to bedside and reverse. Front Cardiovasc Med. 2017;4:77.
37. Mann M, Kumar C, Zeng WF, Strauss MT. Artificial intelligence for proteomics and biomarker discovery. Cell Syst. 2021;12:759–70.
38. Matsuo H, Kamada M, Imamura A, Shimizu M, Inagaki M, Tsuji Y, Hashimoto M, Tanaka M, Ito H, Fujii Y. Machine learning-based prediction of relapse in rheumatoid arthritis patients using data on ultrasound examination and blood test. Sci Rep. 2022;12:7224.
39. Osborn TM, Verdrengh M, Stossel TP, Tarkowski A, Bokarewa M. Decreased levels of the Gelsolin plasma isoform in patients with rheumatoid arthritis. Arthritis Res Ther. 2008;10:R117.
40. Bucki R, Kulakowska A, Byfield FJ, Zendzian-Piotrowska M, Baranowski M, Marzec M, Winer JP, Ciccarelli NJ, Gorski J, Drozdowski W, et al. Plasma Gelsolin modulates cellular response to sphingosine 1-phosphate. Am J Physiol Cell Physiol. 2010;299:C1516–1523.
41. Mun S, Lee J, Lim MK, Lee YR, Ihm C, Lee SH, Kang HG. Development of a Novel Diagnostic Biomarker Set for Rheumatoid Arthritis Using a Proteomics Approach. *Biomed Res Int* 2018, 2018:7490723.
42. Zhang Q, Wen XH, Tang SL, Zhao ZW, Tang CK. Role and therapeutic potential of Gelsolin in atherosclerosis. J Mol Cell Cardiol. 2023;178:59–67.

Publisher's note