Am J Transl Res 2025;17(1):585-602
www.ajtr.org /ISSN:1943-8141/AJTR0160177

Original Article

Identifcation of candidate biomarkers for polyarteritis nodosa using data-independent acquisition mass spectrometry

Huimin Ma[1,2*], Xintian Cai[1*], Delian Zhang[1*], Qing Zhu[1], Ting Wu[1], Xiayire Aierken[1], Ayguzaili Ahmat[1], Shasha Liu[1], Nanfang Li[1]

[1]Hypertension Center of People's Hospital of Xinjiang Uygur Autonomous Region, Xinjiang Hypertension Institute, NHC Key Laboratory of Hypertension Clinical Research, Key Laboratory of Xinjiang Uygur Autonomous Region, Hypertension Research Laboratory, Xinjiang Clinical Medical Research Center for Hypertension (Cardio-Cerebrovascular) Diseases, Urumqi, Xinjiang Uygur Autonomous Region, China; [2]Xinjiang Medical University, Urumqi, Xinjiang Uygur Autonomous Region, China. [*]Equal contributors.

Received August 29, 2024; Accepted December 17, 2024; Epub January 15, 2025; Published January 30, 2025

Abstract: Objectives: Polyarteritis nodosa (PAN) is a rare autoimmune disease that can cause severe functional impairment. Early diagnosis and timely intervention are essential to reduce disease severity and improve outcomes. Methods: Serum proteins from PAN patients and healthy controls were analyzed using data-independent acquisition mass spectrometry (DIA-MS), identifying 55 differentially expressed proteins. Validation was conducted on an independent set of 35 serum samples (10 healthy controls, 15 disease controls, and 10 PAN patients) to evaluate the diagnostic potential of selected biomarkers. Results: Eighteen proteins showed significantly altered expression in PAN patients compared to controls. A diagnostic panel of seven proteins - AZGP1, F13B, LBP, RBP4, SERPINF1, PGLYRP2, and PPBP - was identified using the least absolute shrinkage and selection operator (LASSO) binary logistic regression model. This panel achieved an area under the receiver operating characteristic (ROC) curve of 0.994, effectively distinguishing PAN patients from controls. Conclusion: By combining DIA-MS technology with the LASSO regression model, this study developed a 7-protein diagnostic panel, providing a highly accurate and efficient tool for PAN diagnosis.

Keywords: Data independent acquisition, polyarteritis nodosa, protein biomarkers, serum, mass spectrometry

Introduction

Polyarteritis nodosa (PAN) was first identified and described by Kussmaul and Maier in 1866 [1]. In the field of rheumatology, PAN is considered an orphan disease, a term used for diseases diagnosed in fewer than 50 out of 100,000 patients [1, 2]. The annual incidence of PAN is estimated at 2 to 9 cases per million people. This systemic necrotizing vasculitis predominantly affects medium-sized arteries in visceral organs and soft tissues [2]. While small arteries may also be involved, arterioles, capillaries, and venules are generally spared [3]. In PAN, arterial inflammation causes ischemic damage to the organs served by the affected vessels, manifesting as constitutional symptoms, mononeuritis multiplex, livedo reticularis, purpura, renal dysfunction, hypertension, abdominal pain, and orchitis [4]. The diagnosis of PAN is extremely challenging without specific laboratory tests. The most frequent clinical manifestations of PAN encompass general physical symptoms, neurological and dermatological signs, abdominal pain, and elevated blood pressure [5]. Occasionally, PAN and anti-neutrophil cytoplasmic antibody (ANCA)-associated vasculitis can occur with necrotizing inflammation of medium and small arteries. These conditions are difficult to distinguish clinically and pathologically [6]. Moreover, specific serological markers for PAN are limited. Therefore, the diagnosis of PAN depends on angiography or histological findings of medium-sized arteries, ANCA negativity, and the absence of mucocutaneous lymph node syndrome and glomerulonephritis [1, 2]. However, because angiography or percutaneous renal biopsy is

https://doi.org/10.62347/QSRI6160

invasive and difficult to perform, clinicians have to rely on clinical examination and less sensitive indicators (such as ESR and CRP) to diagnose PAN. Consequently, a large number of patients with PAN are misdiagnosed [7].

Existing studies have highlighted the importance of proteomics [8]. Proteins are essential components responsible for performing various physiological and cellular functions [9]. Monitoring proteomic changes may help us understand the roles of proteins in various diseases [10]. Analyzing serum protein specimens with liquid chromatography - tandem mass spectrometry (LC-MS/MS) presents significant challenges [11]. The wide dynamic range of protein concentrations often hinders the detection and quantification of low-abundance proteins, particularly in data-dependent acquisition (DDA) mode [12]. Additionally, the random selection of precursor ions in DDA tends to prioritize the identification of highly abundant species. To address these limitations, the data-independent acquisition (DIA) mode has recently been developed and successfully implemented [13]. Therefore, the aim of this study is to employ data-independent acquisition mass spectrometry (DIA-MS) to perform an in-depth serum proteomic analysis of patients with PAN, with the goal of identifying potential diagnostic biomarkers. This research not only fills a gap in PAN diagnostic biomarker studies but also provides new directions for future clinical diagnosis and treatment of this rare and serious disease [14].

Methods

Serum sample collection

For the initial experiment, we recruited 5 patients with PAN and 5 healthy individuals. For the follow-up validation experiment, 10 patients with PAN, 10 healthy controls (HC), and 15 patients with other autoimmune diseases (disease controls [DC], diagnosed using gold-standard methods) were enrolled from the People's Hospital of Xinjiang Uygur Autonomous Region. Fasting blood samples were collected from all participants following hospital guidelines, with written informed consent obtained prior to participation. The samples were centrifuged at 5000 × g for 10 minutes at 4°C, and the resulting serum (supernatant) was stored at -80°C for future analysis. All patients, including those with PAN and other autoimmune diseases, presented with constitutional symptoms, mononeuritis multiplex, livedo reticularis, purpura, renal insufficiency, and hypertension at initial admission. PAN diagnoses were confirmed through histological evidence of vasculitis in medium-sized arteries and/or angiographic findings of multiple microaneurysms in visceral organs. Clinical details of all participants are provided in Tables S1 and S2.

Sample preparation and protein processing

Serum samples were depleted of abundant proteins using the Agilent Human-14 Multiple Affinity Removal Column, followed by desalination and concentration with a 10-kDa ultrafiltration tube. Proteins were treated with sodium dodecyl sulfate (SDT) buffer, boiled, centrifuged, and quantified with a BCA protein assay kit before storage at -80°C. For LC-MS/MS analysis, proteins were separated by SDS-PAGE and visualized with Coomassie blue staining. Filter-aided sample preparation (FASP) was used to remove low-molecular-weight components, block cysteine residues, and digest proteins with trypsin, producing peptides that were quantified spectrophotometrically.

LC-MS/MS analysis

Peptides were fractionated using reversed-phase chromatography on an Agilent HPLC system with a gradient of ammonium formate and acetonitrile buffers. For data-dependent acquisition (DDA), peptide separation was achieved on an Easy-nLC 1200 system coupled to a Q-Exactive HF-X mass spectrometer, with specific MS1 and dd-MS2 settings. Raw data were processed using Spectronaut Pulsar X software, applying a 1% FDR and searching against the UniProt Homo sapiens database. Data-independent acquisition (DIA) was performed under similar conditions, using a variable isolation window method and processed with Spectronaut software, maintaining a 1% FDR threshold. Detailed sample preparation methods are provided in the Supplementary Materials.

Bioinformatics analyses

To identify differentially expressed proteins (DEPs) in PAN serum samples compared to con-

trols, proteins with an average fold-change of ≥1.2 or ≤0.83 were selected. Statistical analysis was performed using Student's t-test, with the Benjamini-Hochberg method for multiple testing correction and a significance threshold of P<0.05. Hierarchical clustering was used to classify samples, and bioinformatics analyses explored the biological functions of altered proteins. Gene Ontology (GO) annotation was conducted with Blast2GO [15], while pathway analysis utilized the KEGG database [16]. Fisher's Exact Test identified enriched pathways by comparing DEPs to total proteins in each pathway. PCA, volcano plots, and hierarchical clustering were generated in R (version 3.6.3) to visualize protein expression patterns. The protein-protein interaction (PPI) network of DEPs was analyzed using STRING (version 11.0), with interactions scoring >0.4 considered significant [17].

Data analyses and statistics

All statistical analyses were performed using R software (version 3.6.3). Baseline characteristics were presented as counts and percentages (n [%]) for categorical variables, mean ± standard deviation (SD) for normally distributed continuous variables, and median with interquartile range (IQR) for non-normally distributed continuous variables. Group comparisons were conducted using the chi-square test for categorical variables and either one-way ANOVA or the Kruskal-Wallis test for continuous variables, depending on the data distribution. Receiver Operating Characteristic (ROC) curves and Area Under the Curve (AUC) values were calculated with 1000 bootstrap resamples to assess protein sensitivity and specificity [18]. To identify the optimal diagnostic protein panel, LASSO penalized logistic regression was applied using the glmnet package in R [19]. LASSO selected candidate predictors by shrinking non-informative variable coefficients to zero, with 10-fold cross-validation repeated 1000 times [20]. A binomial generalized linear model with a logit link function was then fitted using the selected proteins as predictors [21]. Boxplots were created in GraphPad Prism (version 7.0, GraphPad Software, San Diego, CA, USA). A two-tailed test with P<0.05 was considered statistically significant for all analyses.

Results

Workflow of quantitative proteomics and diagnostic biomarker screening

We have established a comprehensive workflow to identify reproducible biomarkers with high diagnostic value from serum samples. **Figure 1** presents a schematic overview of this entire multi-step process.

Proteomic profiling of patients with PAN using SDS-PAGE

Sodium dodecyl sulfate - polyacrylamide gel electrophoresis (SDS-PAGE) was employed to perform a semi-quantitative comparison of serum proteomes between the PAN and HC groups. Gel-based fractionation techniques, such as SDS-PAGE, are particularly valuable in proteomic profiling when dealing with samples like serum, where protein concentration spans a wide dynamic range. By separating proteins into distinct bands based on their molecular weights, these methods enable highly sensitive detection. Additionally, gel-based approaches are widely utilized for serum proteomic profiling and biomarker identification. In this study, SDS-PAGE was used to compare serum samples from five PAN patients with those from five sex-matched healthy controls. As illustrated in **Figure 2**, the results demonstrated strong consistency between the PAN and HC groups.

Identification of differentially expressed proteins

LC-MS/MS analysis identified 293 unique proteins in serum samples from five patients with PAN and five healthy individuals. All proteins were subjected to statistical evaluation, and proteins with a *p*-value less than 0.05 were considered significantly differentially expressed. Among these, 55 proteins showed significant differential expression between the PAN and HC groups, with 39 proteins upregulated and 16 downregulated (P<0.05), as detailed in **Table 1**. To assess the differences between the groups, an unsupervised multivariate method - two-dimensional principal component analysis (PCA) - was utilized. As shown in **Figure 3A**, the first principal component, accounting for 42.7% of the variance, clearly differentiated between the HC and PAN groups. This suggests that mul-

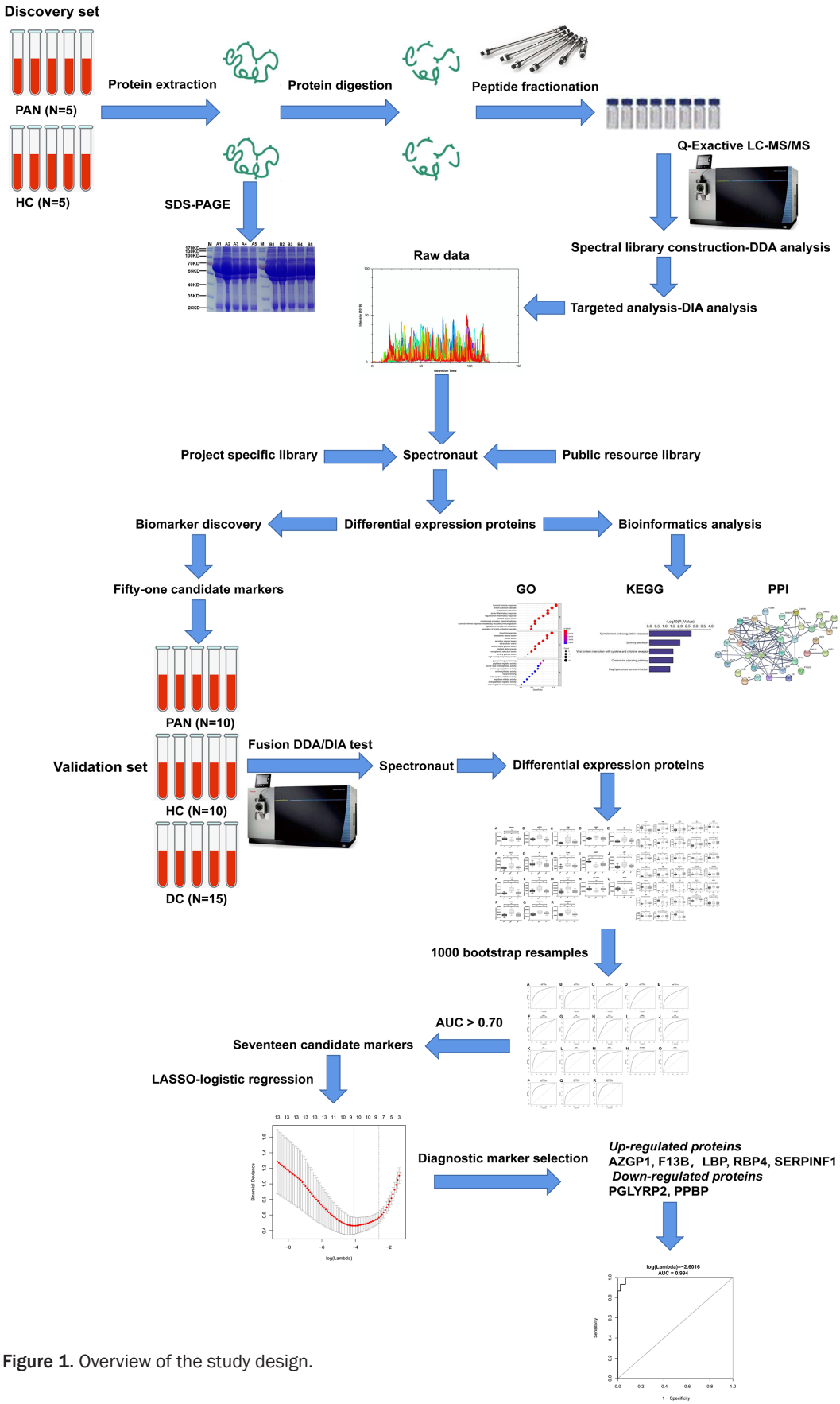


Figure 1. Overview of the study design.

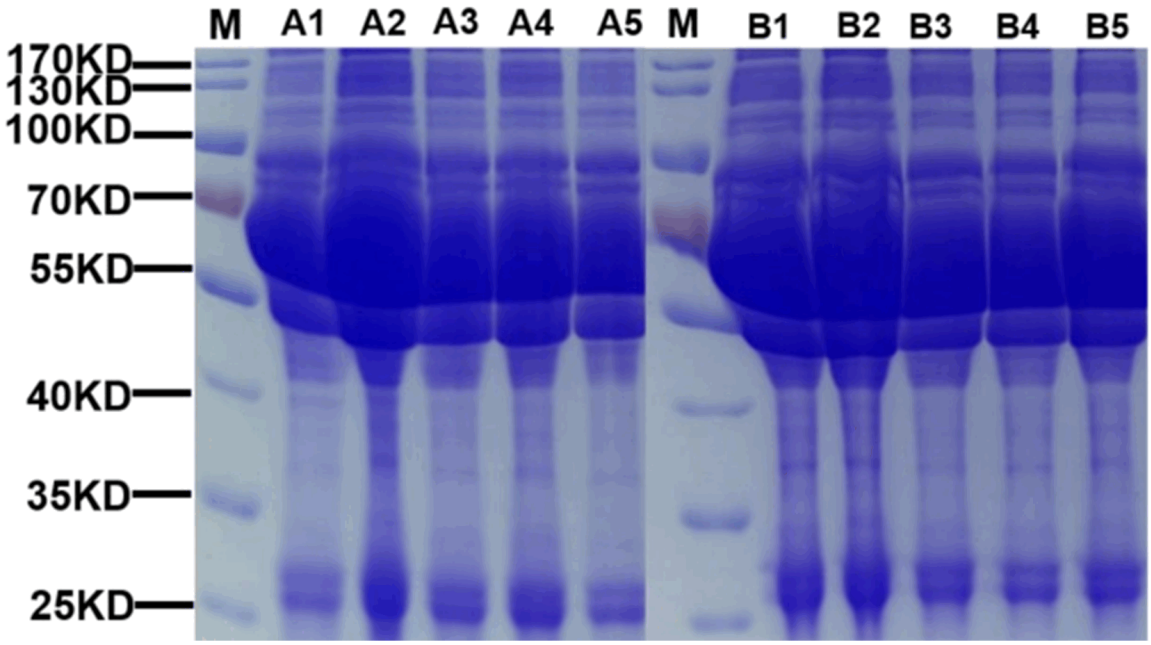


Figure 2. Electrophoresis results by group. Note: KD refers to the ratio between the rate of antibody dissociation from its antigen (k_{off}) and the rate of antibody-antigen association (k_{on}).

Table 1. Expressions of proteins between PNA and HC

Protein accessions	Gene name	Protein descriptions	Fold-change	*p*-value
O43790	KRT86	Keratin, type II cuticular Hb6	4.110583494	<0.001
P01877	IGHA2	Immunoglobulin heavy constant alpha 2	3.097185824	0.004
P0DJI8	SAA1	Serum amyloid A-1 protein	2.488267981	<0.001
P61626	LYZ	Lysozyme C	2.132142438	<0.001
P61769	B2M	Beta-2-microglobulin	2.019985965	<0.001
P36980	CFHR2	Complement factor H-related protein 2	1.946876947	<0.001
P41222	PTGDS	Prostaglandin-H2 D-isomerase	1.945689704	0.018
P01034	CST3	Cystatin-C	1.919056586	<0.001
P00746	CFD	Complement factor D	1.834006455	<0.001
P04264	KRT1	Keratin, type II cytoskeletal 1	1.7993401	<0.001
P13645	KRT10	Keratin, type I cytoskeletal 10	1.792959486	<0.001
P19652	ORM2	Alpha-1-acid glycoprotein 2	1.754184085	<0.001
P18428	LBP	Lipopolysaccharide-binding protein	1.738724359	<0.001
P35908	KRT2	Keratin, type II cytoskeletal 2 epidermal	1.732466345	<0.001
P02760	AMBP	Protein AMBP	1.695489907	<0.001
P02753	RBP4	Retinol-binding protein 4	1.590197077	<0.001
A0A0A0MS15	IGHV3-49	Immunoglobulin heavy variable 3-49	1.555290298	<0.001
P36955	SERPINF1	Pigment epithelium-derived factor	1.513067983	<0.001
P02763	ORM1	Alpha-1-acid glycoprotein 1	1.512636008	<0.001
A0A0J9YX35	IGHV3-64D	Immunoglobulin heavy variable 3-64D	1.481650549	<0.001
Q86UD1	OAF	Out at first protein homolog	1.466191946	0.036
P02748	C9	Complement component C9	1.4543037	<0.001
P05160	F13B	Coagulation factor XIII B chain	1.429550692	<0.001
P55056	APOC4	Apolipoprotein C-IV	1.420036951	<0.001
Q02985	CFHR3	Complement factor H-related protein 3	1.396535878	<0.001

P01011	SERPINA3	Alpha-1-antichymotrypsin	1.356221363	<0.001
P0D0X2		Immunoglobulin alpha-2 heavy chain	1.355796542	<0.001
P07225	PROS1	Vitamin K-dependent protein S	1.317088766	<0.001
P20851	C4BPB	C4b-binding protein beta chain	1.312200476	<0.001
P0C0L4	C4A	Complement C4-A	1.296290722	<0.001
Q03591	CFHR1	Complement factor H-related protein 1	1.292032876	<0.001
P25311	AZGP1	Zinc-alpha-2-glycoprotein	1.287153248	<0.001
P05156	CFI	Complement factor I	1.279644298	<0.001
P04003	C4BPA	C4b-binding protein alpha chain	1.262439525	<0.001
O00187	MASP2	Mannan-binding lectin serine protease 2	1.256131601	<0.001
P12259	F5	Coagulation factor V	1.249170137	<0.001
P00740	F9	Coagulation factor IX	1.243006439	<0.001
P03951	F11	Coagulation factor XI	1.210028745	<0.001
P08603	CFH	Complement factor H	1.206611818	<0.001
P02787	TF	Serotransferrin	0.800918246	<0.001
Q96PD5	PGLYRP2	N-acetylmuramoyl-L-alanine amidase	0.80016675	<0.001
P02652	APOA2	Apolipoprotein A-II	0.789515934	<0.001
P0D0X5		Immunoglobulin gamma-1 heavy chain	0.776155194	<0.001
P01780	IGHV3-7	Immunoglobulin heavy variable 3-7	0.733398449	<0.001
Q15582	TGFBI	Transforming growth factor-beta-induced protein ig-h3	0.732410871	<0.001
P01871	IGHM	Immunoglobulin heavy constant mu	0.726079814	<0.001
P55058	PLTP	Phospholipid transfer protein	0.7204789	<0.001
P22352	GPX3	Glutathione peroxidase 3	0.698887316	<0.001
P02751	FN1	Fibronectin	0.624396362	<0.001
P01743	IGHV1-46	Immunoglobulin heavy variable 1-46	0.621475185	<0.001
P02775	PPBP	Platelet basic protein	0.620708356	<0.001
P02776	PF4	Platelet factor 4	0.61934211	<0.001
P68032	ACTC1	Actin, alpha cardiac muscle 1	0.611898933	0.001
P13473	LAMP2	Lysosome-associated membrane glycoprotein 2	0.586815863	<0.001
P07996	THBS1	Thrombospondin-1	0.486332912	<0.001

Abbreviations: PNA, polyarteritis nodosa; HC, healthy controls.

tiple variables contribute to the differences in protein expression between the two groups. Hierarchical clustering was subsequently employed to visualize the protein expression profiles (**Figure 3B**). The volcano plot in **Figure 3C** further illustrates that 55 proteins were differentially expressed between the PAN and HC groups ($P<0.05$ and fold change >1.2), with 39 proteins significantly upregulated and 16 significantly downregulated in the PAN group. Detailed patient information for the discovery group is provided in Table S1.

GO functional analysis

To investigate the biological functions of the 55 differentially expressed proteins, we performed Gene Ontology (GO) analysis (**Figure 4**). In the category of biological processes (BPs), these proteins were predominantly associated with developmental processes, biological regulation, and immune system processes. Regarding cellular components (CCs), the proteins were mainly linked to protein-containing complexes, the extracellular region, and cells. In terms of molecular functions (MFs), most of the differentially expressed proteins were involved in activities such as binding, catalytic functions, regulation of molecular functions, transporter activity, and molecular transducer activity.

KEGG pathway analysis of differentially expressed proteins

Given that the signaling pathways involving the differentially expressed proteins may be relat-

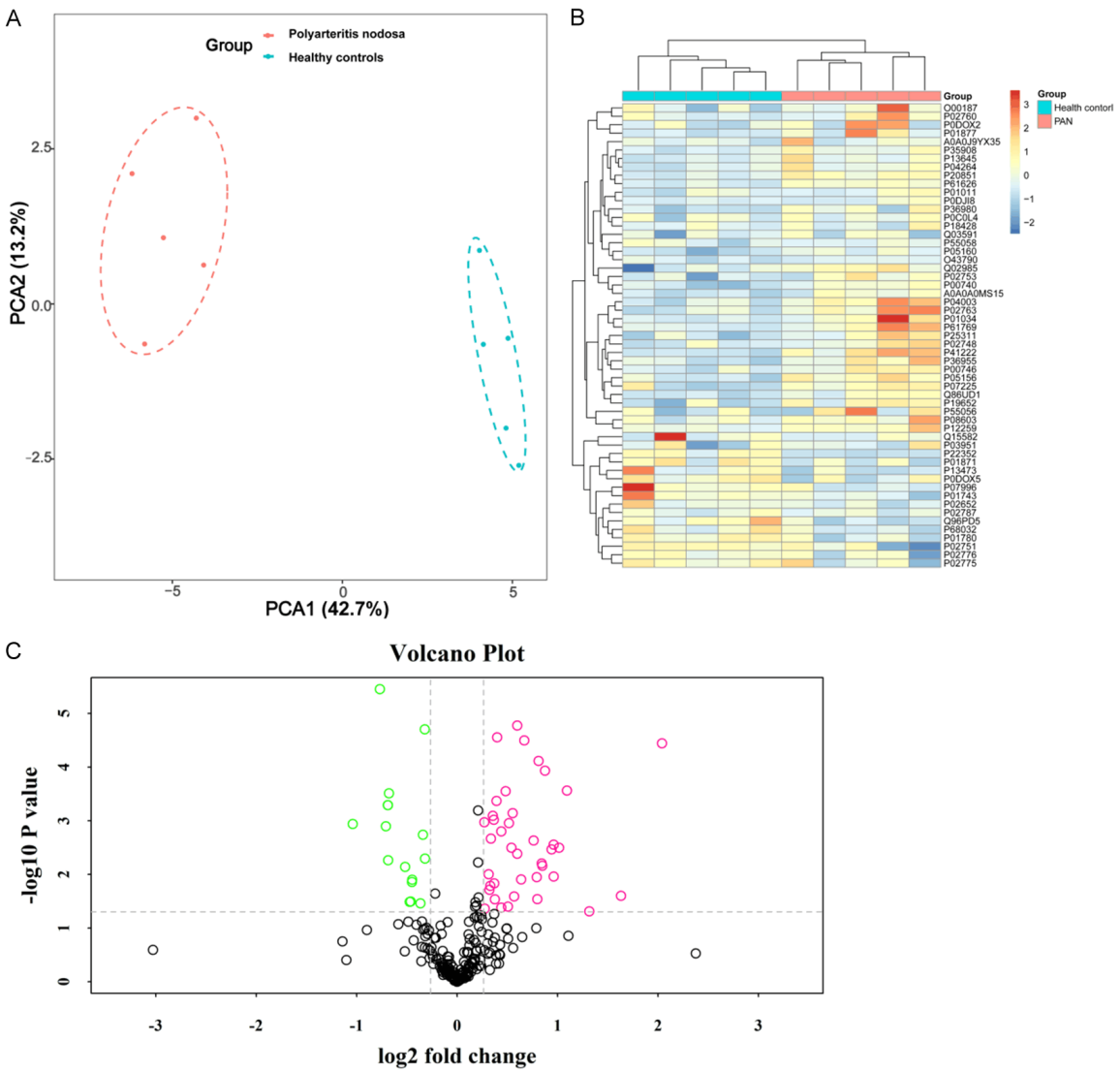


Figure 3. Statistical analysis of serum proteome in PAN patients. A. PCA of 293 proteins from serum proteins of PAN patients and HC. B. Hierarchical clustering of 55 proteins from 10 samples (5 PAN patients and 5 HC). Lines represent proteins, and color is related to their abundance (red indicates more abundant; blue indicates less abundant). C. Volcano plot of statistical significance analysis assessing protein level differences between PAN and HC serum samples. Red dots indicate proteins significantly up-regulated in PAN and green dots indicate proteins significantly down-regulated in PAN. Abbreviations: PAN, polyarteritis nodosa; PCA, principal components analysis; HC, healthy controls.

ed to the pathogenesis of PAN, we subjected the 55 proteins to analysis using the Kyoto Encyclopedia of Genes and Genomes (KEGG) database. As shown in **Figure 5**, these proteins were significantly enriched in pathways associated with complement and coagulation cascades, salivary secretion, interactions of viral proteins with cytokines and cytokine receptors, chemokine signaling, and *Staphylococcus* aureus infection.

Proteins interaction analysis

Biological processes are regulated through complex interactions among the components of different pathways. To obtain deeper insights into the pathogenesis of PAN, we constructed a protein - protein interaction (PPI) network of the differentially expressed proteins using STRING. As shown in **Figure 6**, the network had 178 edges and 47 nodes. The nodes represented

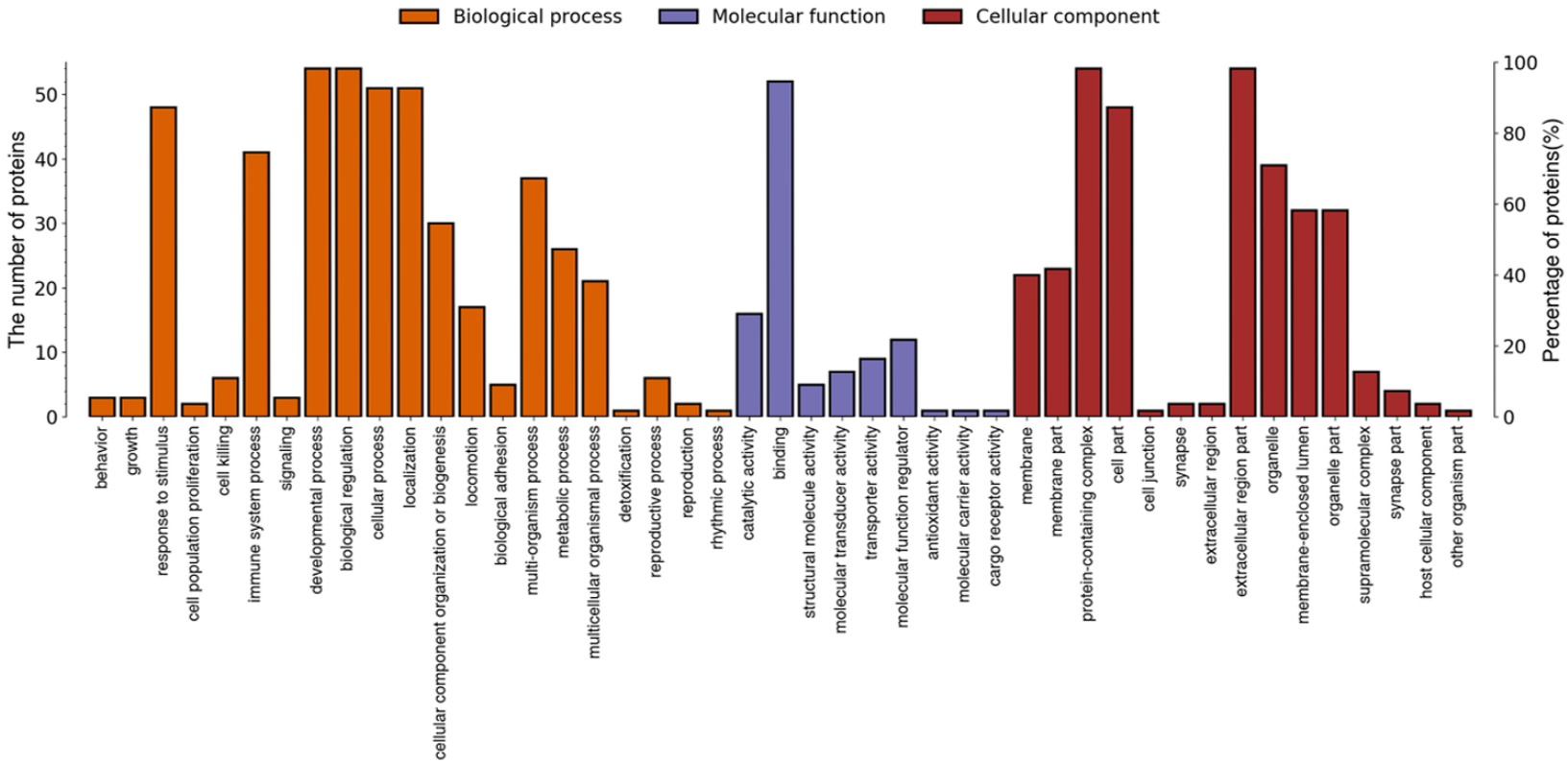


Figure 4. Column chart of the GO enrichment analysis concept map. The abscissa represents enriched GO function classifications, which are divided into three major categories: biological process (BP), molecular function (MF) and cellular component (CC). The vertical axis represents the number of different proteins and the percentage of the total number of proteins included in each entry.

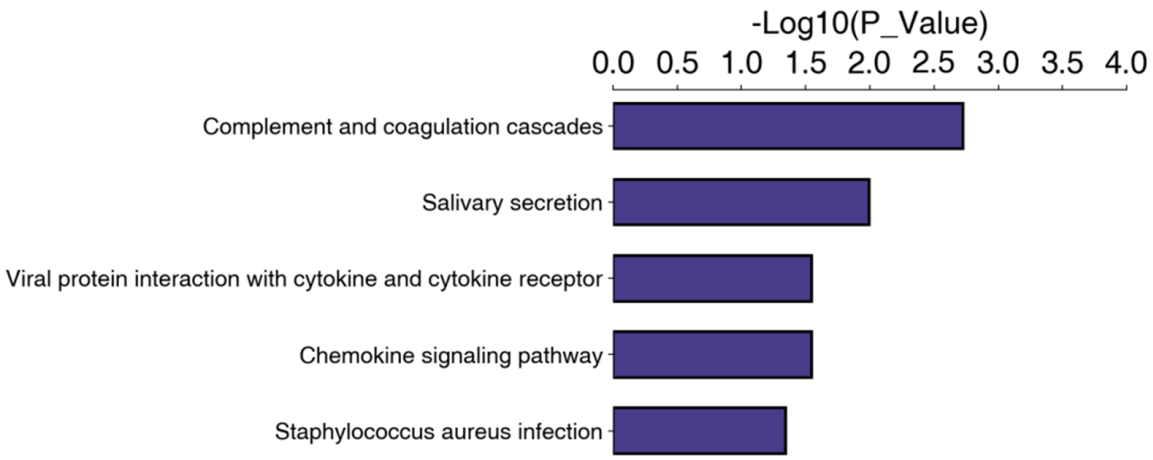


Figure 5. A horizontal histogram showing the top 5 significantly enriched pathways according to the KEGG analysis. The signal paths are displayed on the vertical axis, and the horizontal axis is log *p* value.

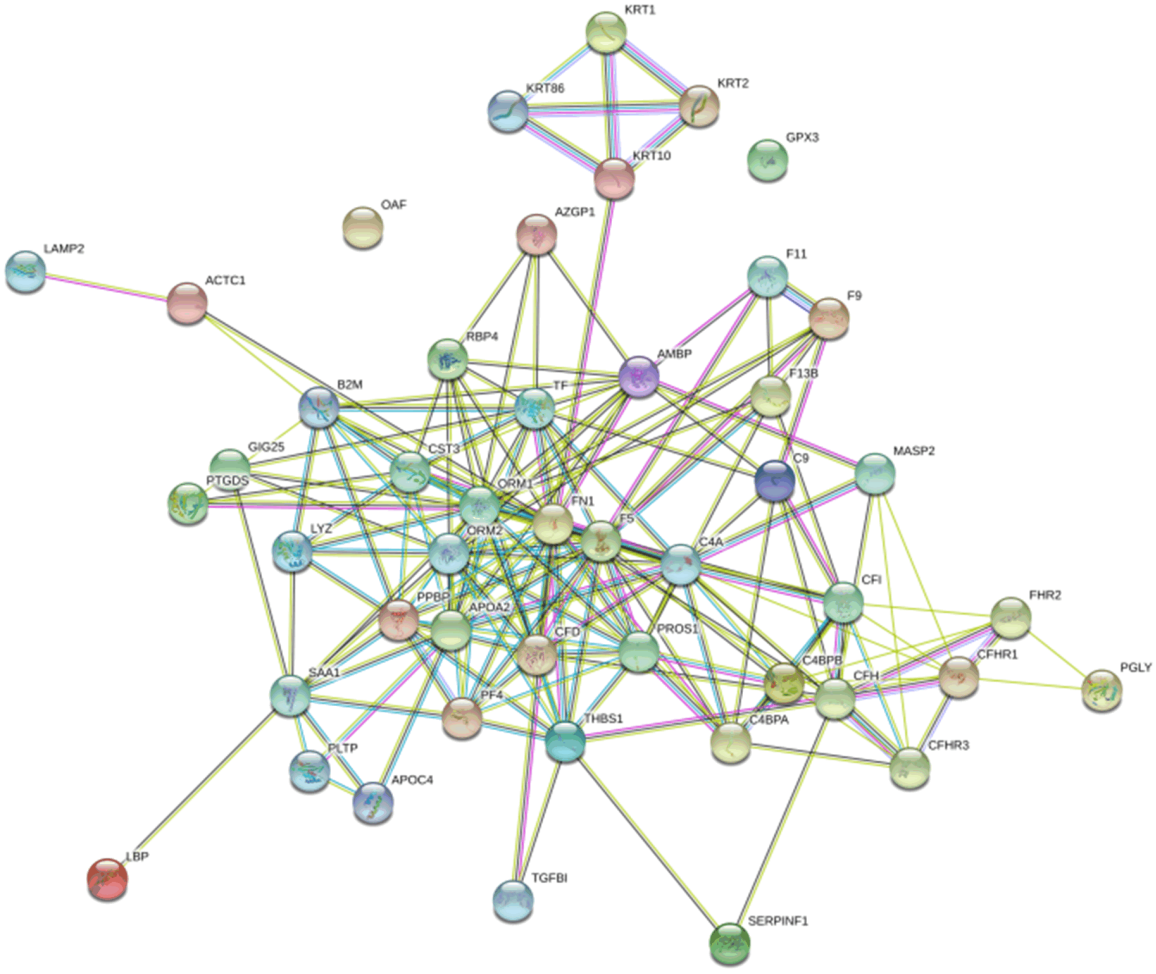


Figure 6. Interaction network analysis of differential expression of proteins. In the protein interaction network, nodes represent proteins, lines represent protein - protein interactions, and different line colors represent the types of evidence for the predicted functional association. A red line indicates the presence of fusion evidence; a green line indicates neighborhood evidence; a blue line indicates co-occurrence evidence; a purple line indicates experimental evidence; a yellow line indicates text mining evidence; a light blue line indicates database evidence; a black line indicates co-expression evidence.

the differentially expressed proteins, whereas the edges represented the interactions between these proteins. The core proteins in the network diagram included RBP4, PPBP, AZGP1, and CST3, which may play an important role in the development of PAN.

Validation of candidate biomarkers using DIA-MS

DIA-MS was employed to validate the 55 differentially expressed proteins, of which 51 were selected for further analysis. Protein abundance differences among the PAN, HC, and DC groups were visualized using box plots. The abundance of each candidate protein was normalized to the total protein abundance within each sample and displayed as 5-95% box plots (**Figures 7** and S1). The serum levels of AZGP1, B2M, C4BPB, C9, CST3, F9, F13B, LBP, LYZ, OAF, ORM2, RBP4, SERPINA3, and SERPINF1 were significantly higher in the PAN group compared to the HC and DC groups. Conversely, the serum levels of APOA2, LAMP2, PGLYRP2, and PPBP were significantly lower in the PAN group than those in the HC and DC groups. However, no significant differences ($P>0.05$) were observed in the serum levels of APOC4, C4A, or F5 between the PAN and HC groups, or in the levels of ACTC1, AMBP, or C4BPA between the PAN and DC groups.

ROC curve analysis

DIA-MS results indicated that proteins with significantly different abundances between the PAN and control (HC or DC) groups could serve as potential diagnostic biomarkers for PAN. To assess the diagnostic performance of each marker in distinguishing PAN from control samples, ROC curve analysis was conducted using 1000 bootstrap resamples to evaluate their specificity and sensitivity. The area under the ROC curve (AUC) was interpreted as follows: 0.9-1, excellent; 0.8-0.9, good; 0.7-0.8, fair; 0.5-0.7, poor; and <0.5, not useful. The specificity and sensitivity of each diagnostic biomarker are shown in **Table 2** and Figure S2. The ROC curve analysis revealed that the panel of seven candidate proteins achieved an AUC value of 0.994, indicating excellent diagnostic performance. This high AUC value suggests that the combined biomarker panel has significant potential for accurately distinguishing PAN patients from controls, which is crucial for early diagnosis and intervention.

LASSO binary logistic regression model

To identify a panel of proteins that could be used to distinguish between PAN and control samples, we performed LASSO penalized logistic regression analysis on serum proteomic data. The λ value was 0.0742. Of all relevant variables, 18 proteins were reduced to 7 proteins on the basis of the validation cohort. The 7 proteins with non-zero coefficients in the LASSO logistic regression model, namely, AZGP1, F13B, LBP, RBP4, SERPINF1, PGLYRP2, and PPBP, were used in the final model (**Figure 8**). This panel of 7 proteins exhibited the optimal ability to discriminate between PAN and control samples, with an AUC value of 0.994 (**Figure 9**).

Discussion

PAN is a rare systemic necrotizing vasculitis with different clinical manifestations, treatment methods, and disease courses [22, 23]. The diagnosis of PAN relies on histopathological confirmation of necrotizing inflammation in small- and medium-sized arteries. For patients ineligible for biopsy, angiographic detection of microaneurysms is commonly used as a diagnostic aid [24, 25]. However, both histopathological examination and angiography are invasive and difficult to perform. As PAN poses a serious threat to human life and health, early diagnosis and treatment are crucial for improving patient survival. Early evaluation is essential to assess progression risk and develop an effective treatment plan. However, techniques for early diagnosis of PAN are currently lacking [25]. Proteomic analysis based on DIA-MS is a recently developed technique that is used to identify diagnostic biomarkers and therapeutic targets for diseases [26]. Compared with traditional proteomics, DIA-MS provides more sensitive and accurate quantitative protein data with excellent reproducibility [27]. In this study, we used proteomic analysis based on DIA-MS to identify diagnostic markers for PAN. In the discovery cohort, 55 differentially expressed proteins were identified between the PAN and HC groups. Subsequently, we used an independent cohort as the validation dataset to re-examine these 55 proteins. Of the 55 proteins, 51 pro-

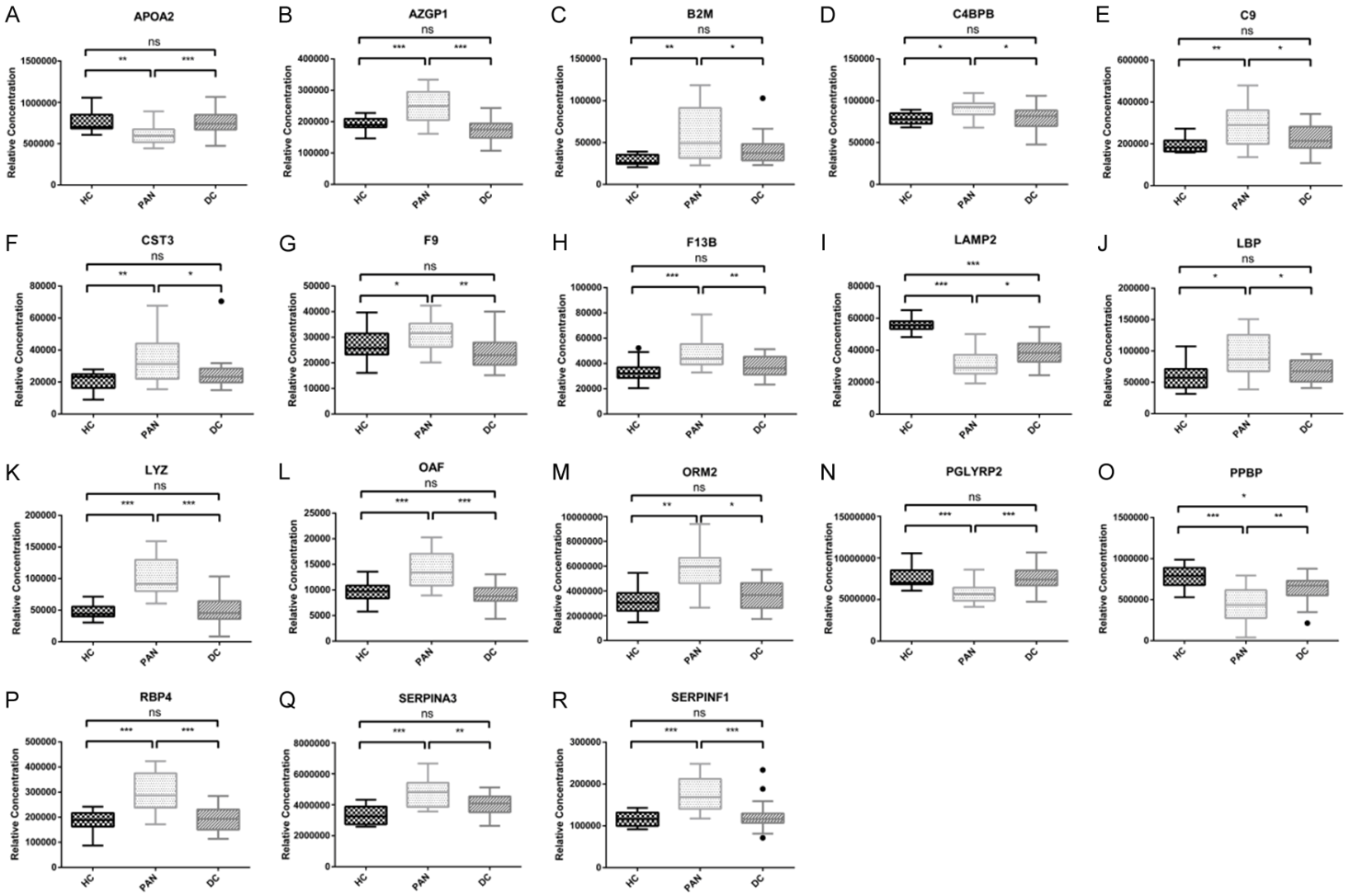


Figure 7. Validation results for the candidate biomarkers in controls (healthy controls, disease controls) and patients with PAN. A-R. The 18 biomarker candidates were measured in controls (healthy controls, disease controls) and patients with PAN. Box plots indicate the individual protein abundance of each group. Independent t-tests were used to determine statistical significance. PAN represents the group of polyarteritis nodosa, HC represents the healthy control group, and DC represents the group of disease control. *P<0.05; **P<0.01; ***P<0.001; ns: non-significant.

Table 2. ROC analysis results of 18 selected candidate biomarkers

Protein	ROC area (AUC)	95% CI low	95% CI up	Specificity	Sensitivity
LYZ	0.887	0.777	0.959	0.897	0.786
RBP4	0.879	0.753	0.96	0.705	0.933
SERPINF1	0.878	0.77	0.959	0.8	0.867
OAF	0.875	0.727	0.963	0.912	0.750
PGLYRP2	0.874	0.755	0.959	0.864	0.800
ORM2	0.855	0.708	0.933	1	0.667
AZGP1	0.84	0.697	0.943	0.814	0.800
LBP	0.824	0.663	0.934	0.932	0.600
APOA2	0.82	0.671	0.936	0.844	0.733
PPBP	0.809	0.651	0.935	0.767	0.857
LAMP2	0.79	0.636	0.907	0.727	0.800
SERPINA3	0.771	0.613	0.913	0.977	0.500
F9	0.763	0.618	0.898	0.568	0.929
F13B	0.763	0.621	0.879	0.651	0.846
C4BPB	0.758	0.595	0.876	0.571	0.929
CST3	0.746	0.568	0.902	0.933	0.643
C9	0.715	0.537	0.858	0.889	0.533
B2M	0.673	0.486	0.831	0.837	0.533

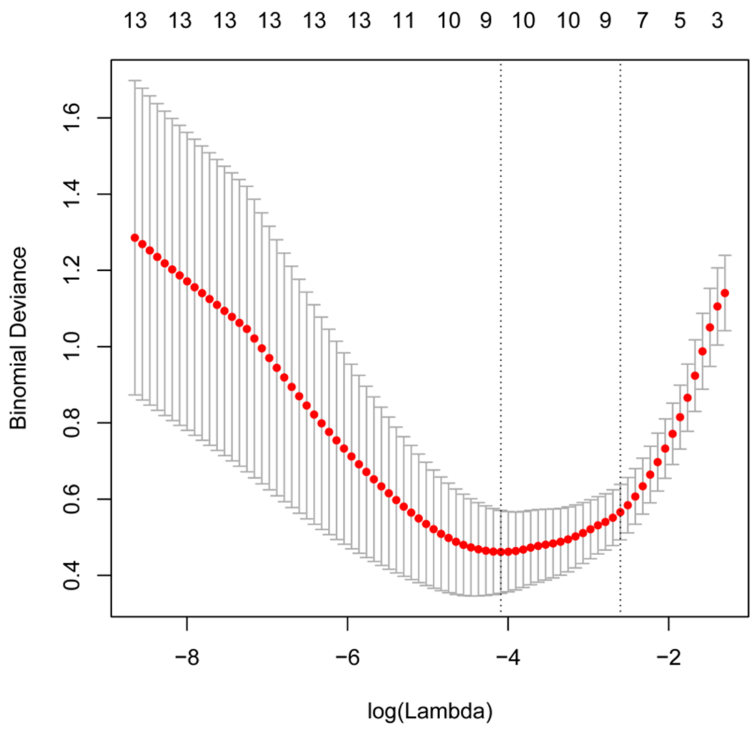


Figure 8. Texture feature selection using the LASSO binary logistic regression model. Ten-fold cross-validation via minimum criteria was used to select the Tuning parameter (λ) in the LASSO model. Dotted vertical lines were drawn at the optimal values with the use of the minimum criteria and the one standard error of the minimum criteria (the 1-SE criteria).

teins were found eligible for further analysis. Eventually, 18 differentially expressed proteins were identified between the PAN and control (HC and DC) groups. Previous studies have shown that diagnostic performance can be improved by combining multiple proteins into a single panel. We analyzed the serum proteomic data of the 18 proteins using the LASSO binary logistic regression model and established a panel of 7 proteins, namely, AZGP1, F13B, LBP, RBP4, SERPINF1, PGLYRP2, and PPBP. These 7 proteins exhibited relatively good diagnostic performance for PAN, with an AUC value of 0.994. By reducing down the candidate proteins in the diagnostic panel, we lowered the potential costs of applying this panel in a multiplex format for large multi-center clinical trials. These findings demonstrate that DIA-MS is a fast and efficient approach for detecting proteins in complex biological matrices like serum.

GO and KEGG analyses showed that most differentially expressed proteins were primarily enriched in biological processes such as developmental processes, biological regulation, and immune system processes. They were also associated with pathways including complement and coagulation cascades, viral protein interactions with cytokines and receptors, and chemokine signaling. These results suggest that PAN pathogenesis may be closely linked to infection and immunity, aligning with findings from previous studies. Our study successfully identified differentially expressed proteins in blood serum,

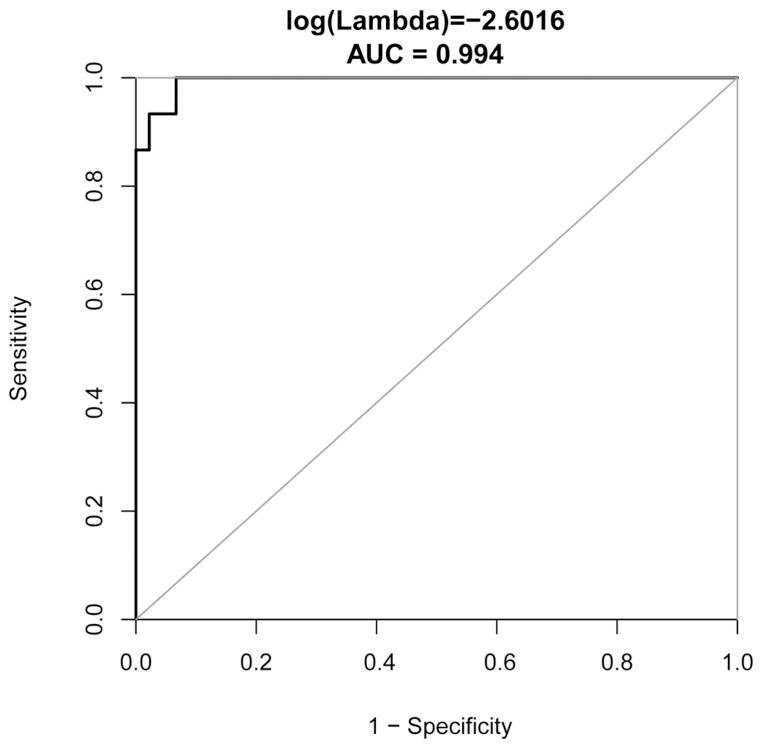


Figure 9. ROC curve for the determined panel of seven proteins.

potentially reflecting altered release or consumption due to the pathological processes of PAN. Thus, these proteins may have physiological or pathological roles in inter-tissue and organ communication.

Alpha-2-glycoprotein 1 (AZGP1), also referred to as zinc α2-glycoprotein (ZAG), is a 41-kDa soluble protein with a folding structure similar to that of major histocompatibility complex-1 (MHC-1) [28]. It was first identified and purified from human serum in 1961 [29]. AZGP1 is an important protein because it is involved in many essential processes in the body, including immune regulation, lipid mobilization, and fertilization [30]. After the discovery of this protein, in the past 60 years, various studies have investigated its structure and function; however, it is considered a protein of unknown function to date [31]. The expression of AZGP1 is regulated by glucocorticoids. Owing to its high homology with lipid mobilization-related factors and because downregulation of its expression can suppress lipolysis and lead to lipid accumulation, AZGP1 is considered a novel type of fat-related factor [32]. The structural organization and folding of AZGP1 are similar to those of MHC class I antigen-presenting molecules [33]. Therefore, AZGP1 may play a role in the generation of immune responses [33, 34]. Guo et al. found that the β3-AR/PKA/CREB pathway mediates the role of AZGP1 in reducing inflammation [34]. In this study, the serum expression of AZGP1 was found to be high in patients with PAN. Therefore, we speculate that AZGP1 may serve as a key factor in the pathogenesis of PAN.

F13B is produced and secreted by hepatocytes [35]. During secretion, its 20-amino-acid leader sequence is cleaved [36]. The mature protein comprises 641 amino acids and 8.5% carbohydrates, with a molecular mass of approximately 80 kDa [37]. Although F13B, an important subunit of coagulation factor XIII, does not have catalytic activity, it is thought to stabilize the A subunit and regulate the production rate of glutamine transaminase by thrombin [38]. As a multifunctional protein, F13B is involved in not only hemostasis but also wound healing and angiogenesis [35]. Studies suggest that F13B plays a role in maintaining vascular permeability, stabilizing and mineralizing the extracellular matrix in bone and cartilage, and providing cardioprotective effects [39]. Additionally, it functions as an intracellular enzyme with specific roles in platelets and monocytes/macrophages [36, 37]. However, research on the role of F13B in PAN remains scarce.

Lipopolysaccharide-binding protein (LBP) is a 58-kDa glycosylated acute-phase protein that is mainly synthesized by hepatocytes [40]. Human LBP belongs to the lipid-binding protein family, which also includes the bactericidal/permeability-enhancing protein, phospholipid ester transfer protein, and cholesterol ester transfer protein [41]. The concentration of LBP in serum is low under physiological conditions but increases several times under inflammatory conditions, such as infection [42]. This

increase is induced by IL-6 and IL-1. LBP can extract lipopolysaccharides (LPSs) from bacterial membranes and enhance the response to LPSs mediated by CD14 [43]. Upon monomerization by CD14, LPSs enter the TLR4MD-2 complex, which activates multiple signaling components and produces proinflammatory cytokines [44]. Although a low concentration of LBP can enhance the cellular response to microorganisms, some researchers have found that the high concentration of LBP observed in the acute phase of inflammation inhibits the immune stimulation of bacterial components [45].

Retinol-binding protein 4 (RBP4) is a low-molecular-weight protein belonging to the lipoprotein superfamily [46]. Its name comes from its ability to transport retinol (vitamin A), which allows the hydrofluoric acid vitamin to be distributed to surrounding tissues. There are several subtypes expressed in different tissues [47]. However, the subtype present in plasma is known as RBP4. The accumulation of visceral fat is initially linked to the onset of inflammation, which is subsequently followed by hormonal dysfunction in adipose tissue, including excessive production of RBP4 [48]. Elevated RBP4 secretion can enhance the expression of adhesion molecules in endothelial cells, contributing to the progression of atherosclerosis and arterial hypertension [49]. Population studies have shown a correlation between circulating RBP4 levels and the severity of atherosclerosis, as well as an increased risk of cardiovascular events and type 2 diabetes mellitus [50]. Thus, circulating RBP4 levels may serve as a biomarker for chronic vascular injury [50, 51].

Serpin peptidase inhibitor, clade F, member 1 (SERPINF1), is a 50-kDa glycoprotein belonging to the serine protease inhibitor (serpin) family, though it lacks inhibitory activity [52]. It was first identified in cultured human retinal epithelial cells. SERPINF1 plays a crucial role in various physiological and pathophysiological processes, including neuroprotection, angiogenesis, inflammatory responses, and fibrosis [53]. It is involved in the regulation of angiogenesis and inflammation through its synergistic effects with vascular endothelial growth factor (VEGF) and other angiogenic and inflammatory factors [54]. Changes in the synergistic activity and dysfunction of VEGF and SERPINF1 regulation are associated with the development of various angiogenesis-related diseases [55]. These changes are usually related to the upregulation of VEGF and the downregulation of SERPINF1, resulting in the aggravation of angiogenesis and inflammation [56].

Peptidoglycan recognition proteins (PGLYRP1-4) form a distinct family of antibacterial proteins. While PGLYRP1, PGLYRP3, and PGLYRP4 directly kill bacteria by triggering a stress response [57]. In contrast, PGLYRP2 functions differently as an N-acetyl-murico-1-alanine amidase. Instead of directly killing bacteria, PGLYRP2 interacts with host factors to induce bacterial cell death and digests pro-inflammatory peptidoglycans into biologically active fragments, potentially playing a key role in preventing excessive inflammation [58]. It is primarily expressed in the liver, where it is secreted into the bloodstream, but is also found in other cells, such as epithelial cells, including those in the intestine [59]. PGLYRP2 is present in the skin and gastrointestinal mucosa as an anti-inflammatory factor and protects the body from excessive inflammation by hydrolyzing proinflammatory peptidoglycans [60]. However, recent studies in mouse models suggest that PGLYRP2 contributes to peptidoglycan-induced acute inflammation and arthritis by playing a pro-inflammatory role [61, 62]. This pro-inflammatory effect is mediated by the NOD2-dependent induction of PGLYRP2 expression in local non-immune cells [63]. Thus, beyond its enzymatic activity, PGLYRP2 also functions as an alarm molecule, similar to antibacterial peptides like β-defensin [64]. These peptides not only exhibit antibacterial activity but also enhance immune responses and inflammation.

Pro-platelet basic protein (PPBP), also known as CXCL7, belongs to the angiogenic ELR+CXC chemokine family [65]. It is predominantly synthesized in megakaryocytes, stored in platelet alpha granules, and released during platelet activation [66]. Secreted CXCL7 plays a vital role in inflammation regulation by activating the CXC chemokine receptor 2 (CXCR2) and interacting with sulfated glycosaminoglycans (GAGs) to modulate receptor activity [67]. Furthermore, it promotes angiogenesis through the activation of the Ras/Raf/mitogen-activated protein kinase (MAPK) and PI3K/AKT/mTOR signaling pathways [68]. PPBP is critical for neutrophil

recruitment to tissues, with dysregulation of this process associated with inflammatory conditions such as rheumatoid arthritis, acute lung injury, COPD, and various cancers [69, 70]. Another key function of PPBP is facilitating neutrophil - platelet crosstalk during vascular injury, where it is released in high concentrations from activated platelets to guide neutrophil migration to the injury site [65, 70].

The results of this study suggest that AZGP1, F13B, LBP, RBP4, SERPINF1, PGLYRP2, and PPBP are candidate diagnostic biomarkers for PAN.

However, this study has some limitations that should be acknowledged. We recruited 10 patients with PAN, 10 healthy individuals, and 15 patients with other autoimmune disease for DIA-MS. However, a larger cohort study is warranted to verify the diagnostic efficacy of the 7-biomarker set in PAN. Because PAN is an extremely rare disease with a very low incidence, we plan to increase the sample size by collaborating with other medical centers in mainland China for further research.

Conclusions

In conclusion, this study utilized DIA-MS to identify a panel of seven candidate biomarkers (AZGP1, F13B, LBP, RBP4, SERPINF1, PGLYRP2, and PPBP) with high diagnostic potential for PAN. These biomarkers not only offer promising tools for the early diagnosis of PAN but also provide insights into the disease's underlying mechanisms. Future studies involving larger, multi-center cohorts are necessary to validate these findings.

Acknowledgements

This study was supported by the "Tianshan Talents" medical and health high-level personnel training program (No. TSYC202301A019).

Disclosure of conflict of interest

None.

Address correspondence to: Nanfang Li and Delian Zhang, Hypertension Center of People's Hospital of Xinjiang Uygur Autonomous Region, Xinjiang Hypertension Institute, NHC Key Laboratory of Hypertension Clinical Research, Key Laboratory of Xinjiang Uygur Autonomous Region, Hypertension Research Laboratory, Xinjiang Clinical Medical Research Center for Hypertension (Cardio-Cerebrovascular) Diseases, No. 91 Tianchi Road, Urumqi 830001, Xinjiang Uygur Autonomous Region, China. E-mail: lnanfang2016@sina.com (NFL); delian53@sina.com (DLZ)

References

[1] Iudici M, Quartier P, Pagnoux C, Merlin E, Agard C, Aouba A, Roblot P, Cohen P, Terrier B, Mouthon L, Guillevin L and Puéchal X; French Vasculitis Study Group. Childhood- versus adult-onset polyarteritis nodosa results from the french vasculitis study group registry. Autoimmun Rev 2018; 17: 984-989.

[2] Karadag O and Jayne DJ. Polyarteritis nodosa revisited: a review of historical approaches, subphenotypes and a research agenda. Clin Exp Rheumatol 2018; 36 Suppl 111: 135-142.

[3] Peters B, von Spiczak J, Ruschitzka F, Distler O, Manka R and Alkadhi H. Cardiac manifestation of polyarteritis nodosa. Eur Heart J 2018; 39: 2603.

[4] Ebersberger U, Rieber J, Wellmann P, Goebel C and Gansera B. Polyarteritis nodosa causing a vast coronary artery aneurysm. J Am Coll Cardiol 2015; 65: e1-2.

[5] Parperis K and Rast F. Inner peace: cutaneous polyarteritis nodosa. Am J Med 2017; 130: 796-798.

[6] Segelmark M and Selga D. The challenge of managing patients with polyarteritis nodosa. Curr Opin Rheumatol 2007; 19: 33-38.

[7] Hernández-Rodríguez J, Alba MA, Prieto-González S and Cid MC. Diagnosis and classification of polyarteritis nodosa. J Autoimmun 2014; 48-49: 84-89.

[8] Aslam B, Basit M, Nisar MA, Khurshid M and Rasool MH. Proteomics: technologies and their applications. J Chromatogr Sci 2017; 55: 182-196.

[9] Altelaar AF, Munoz J and Heck AJ. Next-generation proteomics: towards an integrative view of proteome dynamics. Nat Rev Genet 2013; 14: 35-48.

[10] Li X, Wang W and Chen J. Recent progress in mass spectrometry proteomics for biomedical research. Sci China Life Sci 2017; 60: 1093-1113.

[11] Lundberg E and Borner GHH. Spatial proteomics: a powerful discovery tool for cell biology. Nat Rev Mol Cell Biol 2019; 20: 285-302.

[12] Barkovits K, Linden A, Galozzi S, Schilde L, Pacharra S, Mollenhauer B, Stoepel N, Steinbach S, May C, Uszkoreit J, Eisenacher M and Marcus K. Characterization of cerebrospinal

fluid via data-independent acquisition mass spectrometry. J Proteome Res 2018; 17: 3418-3430.
[13] Muntel J, Kirkpatrick J, Bruderer R, Huang T, Vitek O, Ori A and Reiter L. Comparison of protein quantification in a complex background by DIA and TMT workflows with fixed instrument time. J Proteome Res 2019; 18: 1340-1351.
[14] Cho KC, Clark DJ, Schnaubelt M, Teo GC, Leprevost FDV, Bocik W, Boja ES, Hiltke T, Nesvizhskii AI and Zhang H. Deep proteomics using two dimensional data independent acquisition mass spectrometry. Anal Chem 2020; 92: 4217-4225.
[15] Götz S, García-Gómez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, Robles M, Talón M, Dopazo J and Conesa A. High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic Acids Res 2008; 36: 3420-3435.
[16] Kanehisa M, Furumichi M, Tanabe M, Sato Y and Morishima K. KEGG: new perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Res 2017; 45: D353-D361.
[17] Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, Simonovic M, Doncheva NT, Morris JH, Bork P, Jensen LJ and Mering CV. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res 2019; 47: D607-D613.
[18] Hafeman DM, Merranko J, Goldstein TR, Axelson D, Goldstein BI, Monk K, Hickey MB, Sakolsky D, Diler R, Iyengar S, Brent DA, Kupfer DJ, Kattan MW and Birmaher B. Assessment of a person-level risk calculator to predict new-onset bipolar spectrum disorder in youth at familial risk. JAMA Psychiatry 2017; 74: 841-847.
[19] Vasquez MM, Hu C, Roe DJ, Halonen M and Guerra S. Measurement error correction in the least absolute shrinkage and selection operator model when validation data are available. Stat Methods Med Res 2019; 28: 670-680.
[20] Zhang Y, Li H, Zhang W, Che Y, Bai W and Huang G. LASSO-based Cox-PH model identifies an 11-lncRNA signature for prognosis prediction in gastric cancer. Mol Med Rep 2018; 18: 5579-5593.
[21] Cook G, Royle KL, Pawlyn C, Hockaday A, Shah V, Kaiser MF, Brown SR, Gregory WM, Child JA, Davies FE, Morgan GJ, Cairns DA and Jackson GH. A clinical prediction model for outcome and therapy delivery in transplant-ineligible patients with myeloma (UK Myeloma Research Alliance Risk Profile): a development and validation study. Lancet Haematol 2019; 6: e154-e166.
[22] Liu S, Li N, Zhu Q, Zhu B, Wu T, Wang G, Liu S and Luo Q. Increased serum MCP-1 levels in systemic vasculitis patients with renal involvement. J Interferon Cytokine Res 2018; 38: 406-412.
[23] Wu T, Zhu B, Zhu Q, Tursun D, Liu S, Liu S, Hu J and Li N. Study on serum Pentraxin-3 levels in vasculitis with hypertension. J Interferon Cytokine Res 2019; 39: 522-530.
[24] Cai X, Zhu Q, Wu T, Zhu B, Liu S, Liu S, Aierken X, Ahmat A and Li N. Association of circulating resistin and adiponectin levels with Kawasaki disease: a meta-analysis. Exp Ther Med 2020; 19: 1033-1041.
[25] Li N, Zhu B, Zhu Q, Heizati M, Wu T, Wang G, Yao X, Luo Q, Liu S and Liu S. Serum lysosomal-associated membrane protein-2 levels are increased in small and medium-vessel vasculitis, especially in polyarteritis nodosa. Clin Exp Rheumatol 2019; 37 Suppl 117: 79-85.
[26] Malmström L, Bakochi A, Svensson G, Kilsgård O, Lantz H, Petersson AC, Hauri S, Karlsson C and Malmström J. Quantitative proteogenomics of human pathogens using DIA-MS. J Proteomics 2015; 129: 98-107.
[27] Searle BC, Swearingen KE, Barnes CA, Schmidt T, Gessulat S, Küster B and Wilhelm M. Generating high quality libraries for DIA MS with empirically corrected peptide predictions. Nat Commun 2020; 11: 1548.
[28] Hassan MI, Waheed A, Yadav S, Singh TP and Ahmad F. Zinc alpha 2-glycoprotein: a multidisciplinary protein. Mol Cancer Res 2008; 6: 892-906.
[29] Liu J, Han H, Fan Z, El Beaino M, Fang Z, Li S and Ji J. AZGP1 inhibits soft tissue sarcoma cells invasion and migration. BMC Cancer 2018; 18: 89.
[30] Sidaway P. Prostate cancer: AZGP1 expression predicts favourable outcomes. Nat Rev Urol 2017; 14: 391.
[31] Zhang AY, Grogan JS, Mahon KL, Rasiah K, Sved P, Eisinger DR, Boulas J, Vasilaris A, Henshall SM, Stricker PD, Kench JG and Horvath LG. A prospective multicentre phase III validation study of AZGP1 as a biomarker in localized prostate cancer. Ann Oncol 2017; 28: 1903-1909.
[32] Xue Y, Yu F, Yan D, Cui F, Tang H, Wang X, Chen J, Lu H, Zhao S and Peng Z. Zinc-α-2-glycoprotein: a candidate biomarker for colon cancer diagnosis in Chinese population. Int J Mol Sci 2014; 16: 691-703.
[33] Na HS, Kwon JE, Lee SH, Jhun J, Kim SM, Kim SY, Kim EK, Jung K, Park SH and Cho ML. Th17 and IL-17 cause acceleration of inflammation and fat loss by inducing α2-glycoprotein 1 (AZGP1) in rheumatoid arthritis with high-fat diet. Am J Pathol 2017; 187: 1049-1058.

[34] Guo J, Li Y, Zhao R and Yang X. Adipokine zinc-α2-glycoprotein alleviates lipopolysaccharide-induced inflammatory responses through the β3-AR/PKA/CREB pathway. Cytokine 2019; 123: 154742.
[35] Gemmati D, Occhionorelli S, Tisato V, Vigliano M, Longo G, Gonelli A, Sibilla MG, Serino ML and Zamboni P. Inherited genetic predispositions in F13A1 and F13B genes predict abdominal adhesion formation: identification of gender prognostic indicators. Sci Rep 2018; 8: 16916.
[36] Keenan TD, Toso M, Pappas C, Nichols L, Bishop PN and Hageman GS. Assessment of proteins associated with complement activation and inflammation in maculae of human donors homozygous risk at chromosome 1 CFH-to-F13B. Invest Ophthalmol Vis Sci 2015; 56: 4870-4879.
[37] Thomas A, Biswas A, Dodt J, Philippou H, Hethershaw E, Ensikat HJ, Ivaskevicius V and Oldenburg J. Coagulation factor XIIIA subunit missense mutations affect structure and function at the various steps of factor XIII action. Hum Mutat 2016; 37: 1030-1041.
[38] Mezei ZA, Bereczky Z, Katona É, Gindele R, Balogh E, Fiatal S, Balogh L, Czuriga I, Ádány R, Édes I and Muszbek L. Factor XIII B subunit polymorphisms and the risk of coronary artery disease. Int J Mol Sci 2015; 16: 1143-1159.
[39] Guo ZP, Hou HT, Jing R, Song ZG, Liu XC and He GW. Plasma protein profiling in patients undergoing coronary artery bypass grafting surgery and clinical significance. Oncotarget 2017; 8: 60528-60538.
[40] Escribano BM, Medina-Fernández FJ, Aguilar-Luque M, Agüera E, Feijoo M, Garcia-Maceira FI, Lillo R, Vieyra-Reyes P, Giraldo AI, Luque E, Drucker-Colín R and Túnez I. Lipopolysaccharide binding protein and oxidative stress in a multiple sclerosis model. Neurotherapeutics 2017; 14: 199-211.
[41] Ayyappan P, Harms RZ, Seifert JA, Bemis EA, Feser ML, Deane KD, Demoruelle MK, Mikuls TR, Holers VM and Sarvetnick NE. Heightened levels of antimicrobial response factors in patients with rheumatoid arthritis. Front Immunol 2020; 11: 427.
[42] Smiljanovic B, Grützkau A, Sörensen T, Grün JR, Vogl T, Bonin M, Schendel P, Stuhlmüller B, Claussnitzer A, Hermann S, Ohrndorf S, Aupperle K, Backhaus M, Radbruch A, Burmester GR and Häupl T. Synovial tissue transcriptomes of long-standing rheumatoid arthritis are dominated by activated macrophages that reflect microbial stimulation. Sci Rep 2020; 10: 7907.
[43] Moreno-Navarrete JM, Jové M, Padró T, Boada J, Ortega F, Ricart W, Pamplona R, Badimón L, Portero-Otín M and Fernández-Real JM. Adipocyte lipopolysaccharide binding protein (LBP) is linked to a specific lipidomic signature. Obesity (Silver Spring) 2017; 25: 391-400.
[44] Wen W, Li Y, Cheng Y, He J, Jia R, Li C, Guo J, Sun X and Li Z. Lipopolysaccharide-binding protein is a sensitive disease activity biomarker for rheumatoid arthritis. Clin Exp Rheumatol 2018; 36: 233-240.
[45] Moreno-Navarrete JM, Blasco G, Puig J, Biarnés C, Rivero M, Gich J, Fernández-Aranda F, Garre-Olmo J, Ramió-Torrentà L, Alberich-Bayarri Á, García-Castro F, Pedraza S, Ricart W and Fernández-Real JM. Neuroinflammation in obesity: circulating lipopolysaccharide-binding protein associates with brain structure and cognitive performance. Int J Obes (Lond) 2017; 41: 1627-1635.
[46] Fedders R, Muenzner M, Weber P, Sommerfeld M, Knauer M, Kedziora S, Kast N, Heidenreich S, Raila J, Weger S, Henze A and Schupp M. Liver-secreted RBP4 does not impair glucose homeostasis in mice. J Biol Chem 2018; 293: 15269-15276.
[47] Ortega-Senovilla H, de Oya M and Garcés C. Relationship of NEFA concentrations to RBP4 and to RBP4/retinol in prepubertal children with and without obesity. J Clin Lipidol 2019; 13: 301-307.
[48] Karunanithi S, Levi L, DeVecchio J, Karagkounis G, Reizes O, Lathia JD, Kalady MF and Noy N. RBP4-STRA6 pathway drives cancer stem cell maintenance and mediates high-fat diet-induced colon carcinogenesis. Stem Cell Reports 2017; 9: 438-450.
[49] Li G, Esangbedo IC, Xu L, Fu J, Li L, Feng D, Han L, Xiao X, Li M, Mi J, Li M, Gao S and Willi SM. Childhood retinol-binding protein 4 (RBP4) levels predicting the 10-year risk of insulin resistance and metabolic syndrome: the BCAMS study. Cardiovasc Diabetol 2018; 17: 69.
[50] Zhou W, Ye SD, Chen C and Wang W. Involvement of RBP4 in diabetic atherosclerosis and the role of vitamin D intervention. J Diabetes Res 2018; 2018: 7329861.
[51] Domingos MAM, Queiroz M, Lotufo PA, Benseñor IJ and Titan SMO. Serum RBP4 and CKD: association with insulin resistance and lipids. J Diabetes Complications 2017; 31: 1132-1138.
[52] Wang JY, Li LJ, Zhang Q, Liu Y, Lv F, Xu XJ, Song YW, Wang O, Jiang Y, Xia WB, Xing XP and Li M. Extremely low level of serum pigment epithelium-derived factor is a special biomarker of Chinese osteogenesis imperfecta patients with SERPINF1 mutations. Clin Chim Acta 2018; 478: 216-221.
[53] Geyer PE, Wewer Albrechtsen NJ, Tyanova S, Grassl N, Iepsen EW, Lundgren J, Madsbad S,

Holst JJ, Torekov SS and Mann M. Proteomics reveals the effects of sustained weight loss on the human plasma proteome. Mol Syst Biol 2016; 12: 901.

[54] Eslani M, Putra I, Shen X, Hamouie J, Afsharkhamseh N, Besharat S, Rosenblatt MI, Dana R, Hematti P and Djalilian AR. Corneal mesenchymal stromal cells are directly antiangiogenic via PEDF and sFLT-1. Invest Ophthalmol Vis Sci 2017; 58: 5507-5517.

[55] Wang P, Smit E, Brouwers MC, Goossens GH, van der Kallen CJ, van Greevenbroek MM and Mariman EC. Plasma pigment epithelium-derived factor is positively associated with obesity in Caucasian subjects, in particular with the visceral fat depot. Eur J Endocrinol 2008; 159: 713-718.

[56] Schlereth SL, Karlstetter M, Hos D, Matthaei M, Cursiefen C and Heindl LM. Detection of pro- and antiangiogenic factors in the human sclera. Curr Eye Res 2019; 44: 172-184.

[57] Wolf AJ and Underhill DM. Peptidoglycan recognition by the innate immune system. Nat Rev Immunol 2018; 18: 243-254.

[58] Dabrowski AN, Conrad C, Behrendt U, Shrivastav A, Baal N, Wienhold SM, Hackstein H, N'Guessan PD, Aly S, Reppe K, Suttorp N and Zahlten J. Peptidoglycan recognition protein 2 regulates neutrophil recruitment into the lungs after streptococcus pneumoniae infection. Front Microbiol 2019; 10: 199.

[59] Yang Z, Feng J, Xiao L, Chen X, Yao Y, Li Y, Tang Y, Zhang S, Lu M, Qian Y, Wu H and Shi M. Tumor-derived peptidoglycan recognition protein 2 predicts survival and antitumor immune responses in hepatocellular carcinoma. Hepatology 2020; 71: 1626-1642.

[60] Duerr CU, Salzman NH, Dupont A, Szabo A, Normark BH, Normark S, Locksley RM, Mellroth P and Hornef MW. Control of intestinal Nod2-mediated peptidoglycan recognition by epithelium-associated lymphocytes. Mucosal Immunol 2011; 4: 325-334.

[61] Gowda RN, Redfern R, Frikeche J, Pinglay S, Foster JW, Lema C, Cope L and Chakravarti S. Functions of peptidoglycan recognition proteins (Pglyrps) at the ocular surface: bacterial keratitis in gene-targeted mice deficient in Pglyrp-2, -3 and -4. PLoS One 2015; 10: e0137129.

[62] Yao X, Gao M, Dai C, Meyer KS, Chen J, Keeran KJ, Nugent GZ, Qu X, Yu ZX, Dagur PK, McCoy JP and Levine SJ. Peptidoglycan recognition protein 1 promotes house dust mite-induced airway inflammation in mice. Am J Respir Cell Mol Biol 2013; 49: 902-911.

[63] Jing X, Zulfiqar F, Park SY, Núñez G, Dziarski R and Gupta D. Peptidoglycan recognition protein 3 and Nod2 synergistically protect mice from dextran sodium sulfate-induced colitis. J Immunol 2014; 193: 3055-3069.

[64] Saha S, Jing X, Park SY, Wang S, Li X, Gupta D and Dziarski R. Peptidoglycan recognition proteins protect mice from experimental colitis by promoting normal gut flora and preventing induction of interferon-gamma. Cell Host Microbe 2010; 8: 147-162.

[65] Shusterman A, Munz M, Richter G, Jepsen S, Lieb W, Krone B, Hoffman P, Laudes M, Wellmann J, Berger K, Kocher T, Offenbacher S, Divaris K, Franke A, Schreiber S, Dommisch H, Weiss E, Schaefer AS, Houri-Haddad Y and Iraqi FA. The PF4/PPBP/CXCL5 gene cluster is associated with periodontitis. J Dent Res 2017; 96: 945-952.

[66] Su C, Li H, Peng Z, Ke D, Fu H and Zheng X. Identification of plasma RGS18 and PPBP mRNAs as potential biomarkers for gastric cancer using transcriptome arrays. Oncol Lett 2019; 17: 247-255.

[67] Eicher JD, Wakabayashi Y, Vitseva O, Esa N, Yang Y, Zhu J, Freedman JE, McManus DD and Johnson AD. Characterization of the platelet transcriptome by RNA sequencing in patients with acute myocardial infarction. Platelets 2016; 27: 230-239.

[68] Tan F, Guio-Aguilar PL, Downes C, Zhang M, O'Donovan L, Callaway JK and Crack PJ. The σ 1 receptor agonist 4-PPBP elicits ERK1/2 phosphorylation in primary neurons: a possible mechanism of neuroprotective action. Neuropharmacology 2010; 59: 416-424.

[69] Moodley YP, Corte TJ, Oliver BG, Glaspole IN, Livk A, Ito J, Peters K, Lipscombe R, Casey T and Tan DBA. Analysis by proteomics reveals unique circulatory proteins in idiopathic pulmonary fibrosis. Respirology 2019; 24: 1111-1114.

[70] Zhang Y, Ma KL, Gong YX, Wang GH, Hu ZB, Liu L, Lu J, Chen PP, Lu CC, Ruan XZ and Liu BC. Platelet microparticles mediate glomerular endothelial injury in early diabetic nephropathy. J Am Soc Nephrol 2018; 29: 2671-2695.

Supplementary Materials

Sample preparation

Serum samples were depleted of most abundant proteins using the Agilent Human-14 Multiple Affinity Removal Column according to the manufacturer's protocol (Agilent Technologies). A 10-kDa ultrafiltration tube (Sartorius) was used for desalination and concentration of low-abundance components. One volume of sodium dodecyl sulfate (SDT) buffer was added; boiled for 15 min; and centrifuged at 14,000 × g for 20 min. The protein in the supernatant was quantified using a BCA protein assay kit (Bio-Rad, USA), and the supernatant was stored at -80°C until further use.

LC-MS/MS analysis

SDS-PAGE

A total of 20 μg of protein from each sample was mixed with 5 × loading buffer, and the mixture was boiled for 5 min. The proteins were separated on a 12.5% sodium dodecyl sulfate - polyacrylamide gel, and protein bands were visualized using Coomassie blue R-250 staining.

Filter-aided sample preparation

A total of 200 μg of protein from each sample was mixed with 30 μL of SDT buffer (4% SDS, 100-mM DTT, and 150-mM Tris-HCl; pH 8.0). The detergent, DTT, and other low-molecular-weight components were removed via repeated ultrafiltration using UA buffer (8-M urea and 150-mM Tris-HCl; pH 8.5) (Sartorius, 30 kD). Subsequently, 100 μL of iodoacetamide (IAA) (100-mM IAA in UA buffer) was added to block reduced cysteine residues and the samples were incubated for 30 min in the dark. Filters were washed with 100 μL of UA buffer three times and 100 μL of 0.1 M TEAB two times. The protein suspensions were digested with 4-μg trypsin (Promega) in 40 μL of 0.1 M TEAB overnight at 37°C, and the resulting peptides were collected as a filtrate. The peptide content was estimated by measuring the UV light spectral density at 280 nm using an extinction coefficient of 1.1 for a 0.1% (g/L) solution that was calculated based on the frequency of tryptophan and tyrosine in the proteins.

Peptide fractionation with reversed phase chromatography

Peptide fractionation was performed using reversed-phase (RP) chromatography on an Agilent 1260 Infinity II HPLC system. The peptide mixture was diluted in buffer A (10 mM ammonium formate, 5% acetonitrile, pH 10.0) and loaded onto an XBridge Peptide BEH C18 column (130 Å, 5 μm, 4.6 mm × 100 mm). Elution was carried out at a flow rate of 1 mL/min using a gradient of buffer B (10 mM ammonium formate, 85% acetonitrile, pH 10.0) as follows: 0-7% buffer B over 5 minutes, 7-40% from 5 to 40 minutes, 40-100% from 45 to 50 minutes, and maintained at 100% from 50 to 65 minutes. Elution was monitored at 214 nm, and fractions were collected every minute between 5 and 50 minutes. The collected fractions were concentrated via vacuum centrifugation at 45°C, re-dissolved in 0.1%.

DDA analysis

Two microliters (2 μL) of 10 × iRT peptides were added to six microliters (6 μL) of the peptide solution. The mixture was injected into an Easy-nLC 1200 nano-LC system coupled to a Q-Exactive HF-X mass spectrometer (Thermo Fisher Scientific). Peptide separation was achieved using a gradient of buffer B (80% acetonitrile and 0.1% formic acid) at a flow rate of 300 nL/min. The gradient was programmed as follows: 1% buffer B for 0-5 minutes; 1% to 28% buffer B from 5 to 95 minutes; 28% to 38% buffer B from 95 to 110 minutes; 38% to 100% buffer B from 110 to 115 minutes; and held at 100% buffer B from 115 to 120 minutes. Mass spectrometer parameters were configured as follows: MS1 settings: Scan range m/z 350-1500; resolution 60,000; AGC target 3×10^6; maximum injection time 50 ms; charge states 2-7; dynamic exclusion duration 30 s. dd-MS2 settings: Isolation window 1.6 m/z; resolu-

tion 15,000; AGC target 1×10^5; maximum injection time 45 ms; normalized collision energy (NCE) 28%. Raw data were processed using Spectronaut Pulsar X software (version 12, Biognosys AG) with the following parameters: Maximum missed cleavages set to 2. Static modification of carbamidomethylation on cysteine residues. Dynamic modifications, including methionine oxidation and N-terminal acetylation. False discovery rate (FDR) thresholds of 1% for both PSM and peptide levels. The data were searched against the UniProt Homo sapiens database (20,386 entries, downloaded in September 2018 from http://www.uniprot.org).

DIA analysis

DIA acquisition was conducted using the same instrument and chromatographic conditions as the DDA analysis. A variable isolation window DIA method was developed specifically for serum peptide samples. The isolation window list was constructed based on the distribution of precursor ions within each window (Table S3). Two microliters (2 μL) of 10 × iRT peptides were added to six microliters (6 μL) of the peptide solution, and the mixture was injected into the Easy-nLC 1200 nano-LC system coupled to a Q-Exactive HF-X mass spectrometer (Thermo Fisher Scientific). The gradient and analytical column were identical to those used in the DDA analysis. The DIA method included one MS1 scan and 42 DIA scans with the following settings: MS1 settings: Scan range m/z 350-1500; resolution 60,000; AGC target 3×10^6; maximum injection time 50 ms. DIA settings: Resolution 30,000; AGC target 1×10^6; maximum injection time set to "auto"; normalized collision energy (NCE) 28%. To validate the analysis of other sample groups, DIA acquisition was also performed on an Orbitrap Fusion mass spectrometer using identical LC conditions and parameters. Raw data were processed using Spectronaut Pulsar X software (version 12, Biognosys AG) with default BGS Factory Settings. An FDR cutoff of 1% was applied at the precursor level.

Table S1. Baseline characteristics of patients for discovery

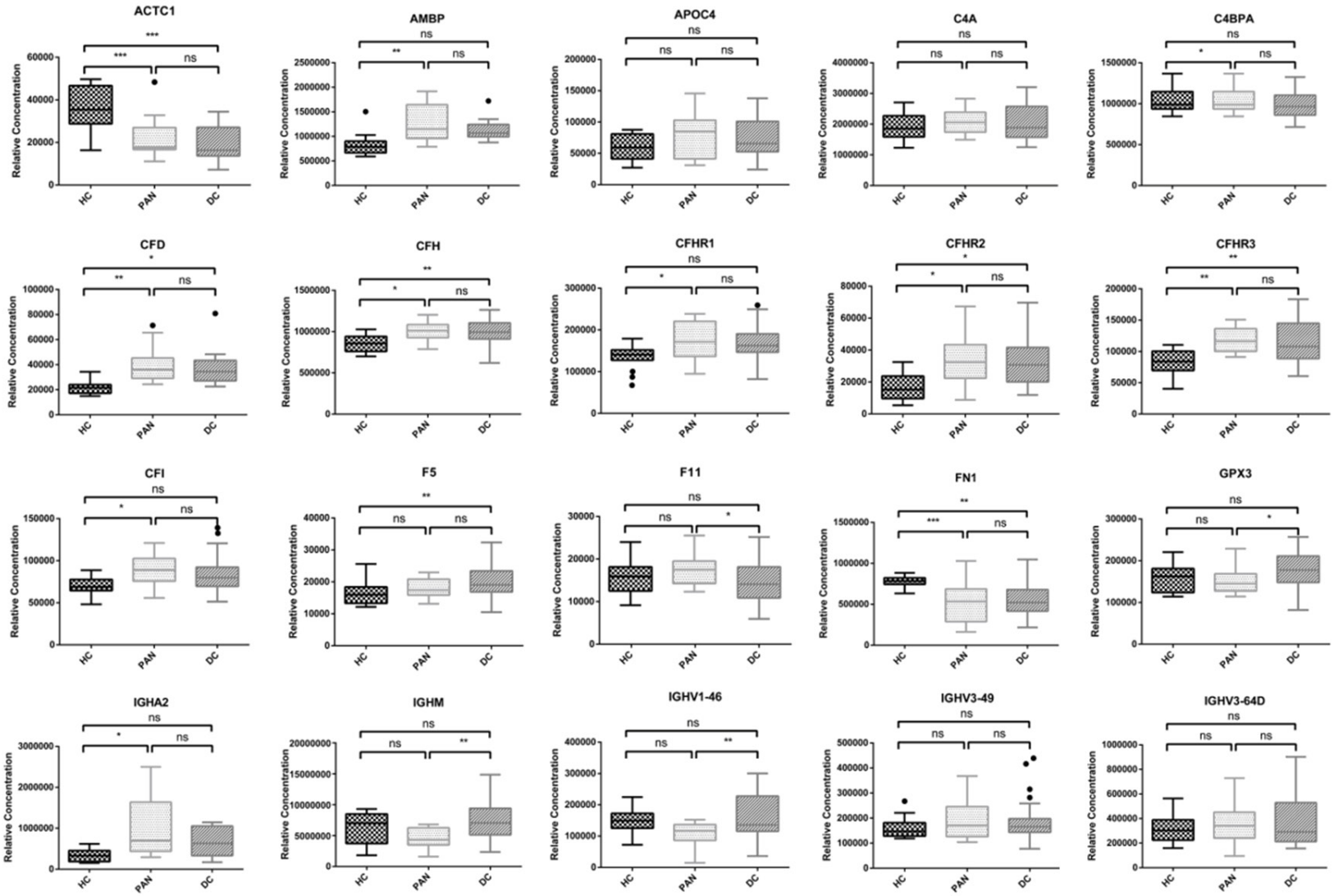
Characteristics	PAN (n=5)	Healthy (n=5)
Smoking n (%)		
No	3 (60.00%)	5 (100.00%)
Yes	2 (40.00%)	0 (0.00%)
Drinking n (%)		
No	2 (40.00%)	5 (100.00%)
Yes	3 (60.00%)	0 (0.00%)
BMI (kg/m^2)	25.60±5.15	23.58±2.24
BUN (mmol/L)	7.54±2.26	4.99±1.36
Cr (μmol/L)	137.12±37.27	85.80±6.57
Glu (mmol/L)	5.92±3.53	5.11±0.55
AST (U/L)	27.77±11.35	19.74±12.21
ALT (U/L)	21.61±7.40	19.76±5.16
WBC ($10^9/L$)	7.41±0.95	6.07±0.96
RBC ($10^9/L$)	5.04±0.44	5.34±0.32
HB ($10^9/L$)	148.20±12.28	160.40±4.34
PLT ($10^9/L$)	194.80±52.07	292.80±31.34
NEUT ($10^9/L$)	4.99±0.81	3.24±0.55
LY ($10^9/L$)	1.73±0.20	2.10±0.24
MONO ($10^9/L$)	0.53±0.19	0.43±0.09
NEUT%	67.31±3.40	53.42±2.60
LY%	23.67±4.58	34.94±3.26
MONO%	6.99±1.59	7.12±0.75

BMI, body mass index; BUN, blood urea nitrogen; Cr, creatinine; Glu, glucose; AST, aspartate transaminase; ALT, alanine transaminase; WBC, white blood cell; RBC, red blood cell; HB, hemoglobin; PLT, platelet; NEUT, neutrophil; LY, lymphocyte; MONO, monocyte.

Table S2. Baseline characteristics of patients for validation

Characteristics	PAN (n=10)	Healthy (n=10)	Disease control (n=30)
Smoking n (%)			
No	2 (20.00%)	10 (100.00%)	25 (83.33%)
Yes	8 (80.00%)	0 (0.00%)	5 (16.67%)
Drinking n (%)			
No	5 (50.00%)	9 (90.00%)	21 (70.00%)
Yes	5 (50.00%)	1 (10.00%)	9 (30.00%)
BMI (kg/m^2)	26.53±2.28	22.25±3.54	24.93±2.08
BUN (mmol/L)	6.93±3.54	4.05±1.09	5.01±2.06
Cr (μmol/L)	120.75±46.61	76.57±13.56	68.17±33.42
Glu (mmol/L)	4.76±1.27	4.54±0.31	4.44±0.54
AST (U/L)	20.25±6.47	14.73±9.04	22.47±26.02
ALT (U/L)	16.82±1.99	17.01±3.43	22.08±23.06
WBC ($10^9/L$)	6.86±2.15	6.34±0.87	5.36±1.69
RBC ($10^9/L$)	4.68±0.56	4.93±0.38	4.64±0.51
HB ($10^9/L$)	139.00±13.20	146.10±16.70	136.00±19.87
PLT ($10^9/L$)	181.60±49.41	277.40±51.11	240.93±61.08
NEUT ($10^9/L$)	4.34±1.95	3.69±0.81	3.03±1.02
LY ($10^9/L$)	1.86±0.41	2.12±0.56	1.74±0.68
MONO ($10^9/L$)	0.43±0.21	0.39±0.09	0.41±0.13
NEUT%	61.06±10.02	57.96±8.78	56.57±6.40
LY%	29.37±10.40	33.69±8.26	32.31±6.25
MONO%	6.28±2.24	6.11±1.12	7.92±1.92

BMI, body mass index; BUN, blood urea nitrogen; Cr, creatinine; Glu, glucose; AST, aspartate transaminase; ALT, alanine transaminase; WBC, white blood cell; RBC, red blood cell; HB, hemoglobin; PLT, platelet; NEUT, neutrophil; LY, lymphocyte; MONO, monocyte.


ACTC1
AMBP
APOC4
C4A
C4BPA
CFD
CFH
CFHR1
CFHR2
CFHR3
CFI
F5
F11
FN1
GPX3
IGHA2
IGHM
IGHV1-46
IGHV3-49
IGHV3-64D
Relative Concentration
HC
PAN
DC
ns

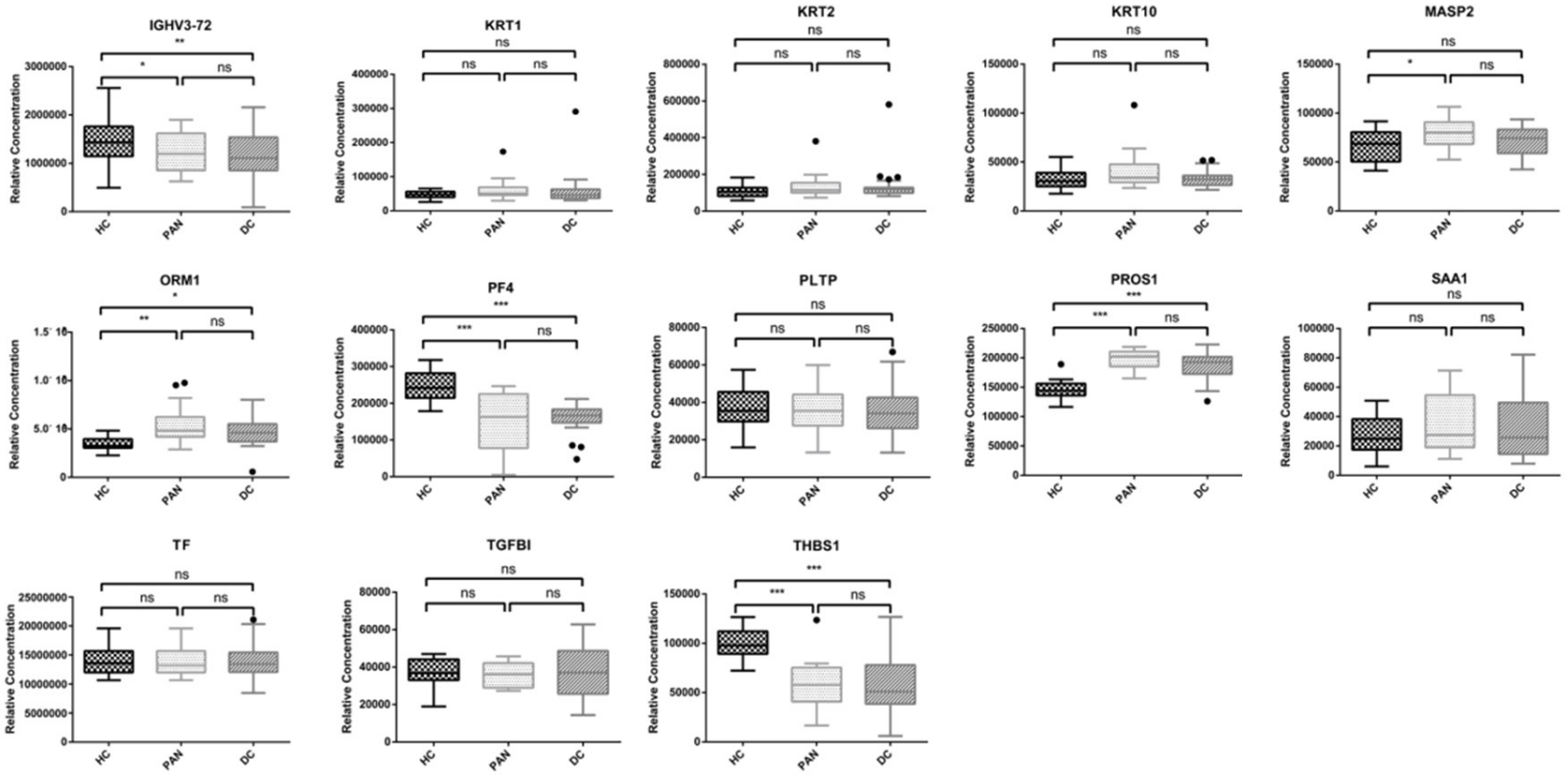


Figure S1. Validation results for the candidate biomarkers in controls (healthy controls, disease controls) and patients with PAN. The 33 biomarker candidates were measured in controls (healthy controls, disease controls) and patients with PAN. Box plots indicate the individual protein abundance of each group. Independent t-tests were used to determine statistical significance. *P<0.05; **P<0.01; ***P<0.001; ns: non-significant.

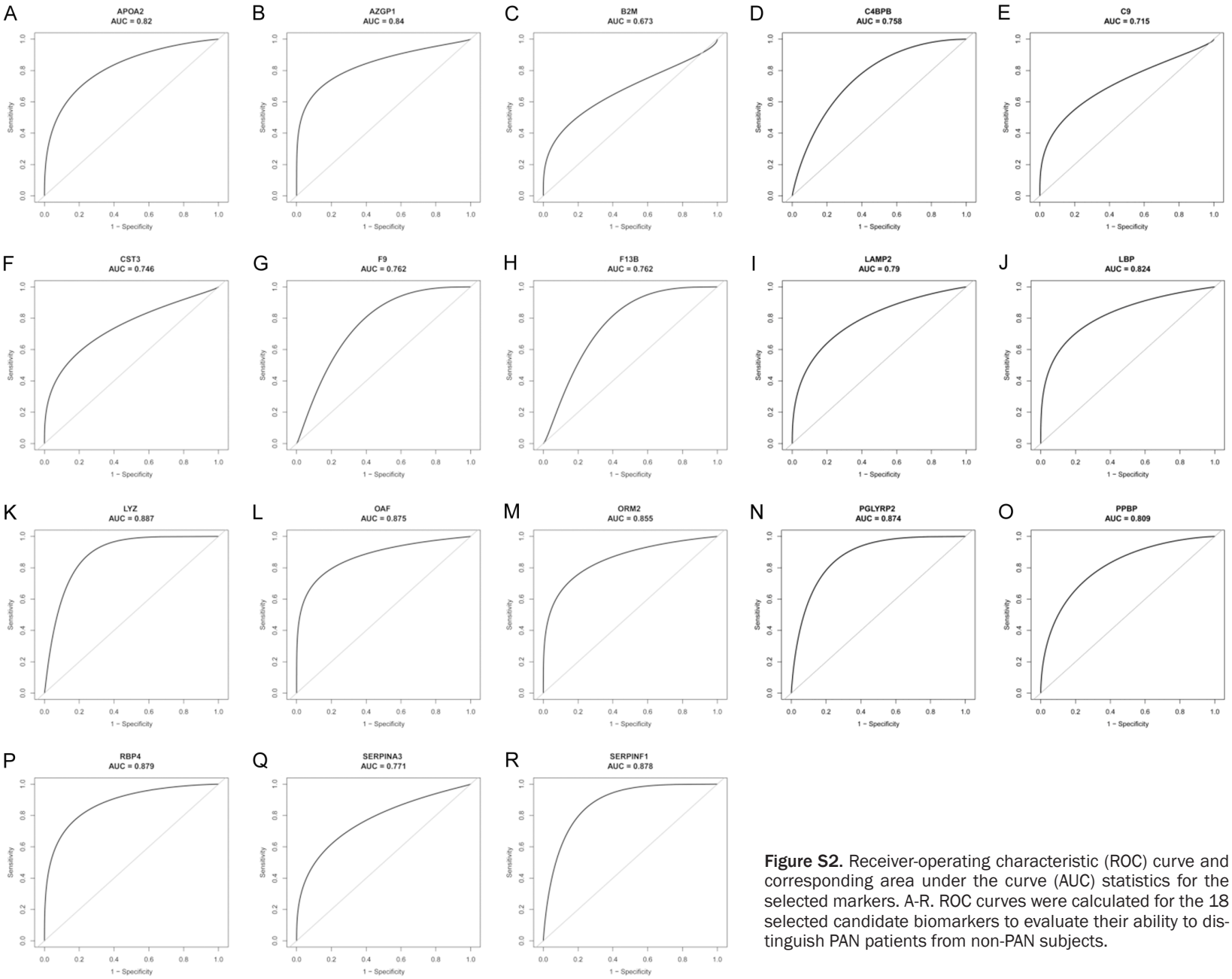


Figure S2. Receiver-operating characteristic (ROC) curve and corresponding area under the curve (AUC) statistics for the selected markers. A-R. ROC curves were calculated for the 18 selected candidate biomarkers to evaluate their ability to distinguish PAN patients from non-PAN subjects.

Table S3. The window list of the DIA method

DIA window	Start m/z	End m/z	Width
1	350	405	55
2	405	429	24
3	429	453	24
4	453	466	13
5	466	479	13
6	479	492	13
7	492	505	13
8	505	518	13
9	518	531	13
10	531	544	13
11	544	557	13
12	557	570	13
13	570	583	13
14	583	596	13
15	596	609	13
16	609	622	13
17	622	635	13
18	635	648	13
19	648	661	13
20	661	674	13
21	674	687	13
22	687	700	13
23	700	713	13
24	713	726	13
25	726	739	13
26	739	752	13
27	752	765	13
28	765	778	13
29	778	796	18
30	796	814	18
31	814	832	18
32	832	855	23
33	855	878	23
34	878	901	23
35	901	924	23
36	924	947	23
37	947	981	34
38	981	1015	34
39	1015	1063	48
40	1063	1111	48
41	1111	1189	78
42	1189	1500	311