



OPEN ACCESS

EDITED BY

Ahmed Fahim,
Royal Wolverhampton Hospitals NHS
Trust, United Kingdom

REVIEWED BY

Paola de Candia,
University of Naples Federico II, Italy
Mathew John,
Jubilee Mission Medical College and
Research Institute, India

*CORRESPONDENCE

Hong-Long Ji
✉ hji@luc.edu;
✉ Hong-Long.Ji@luhs.org

[†]These authors have contributed equally
to this work

RECEIVED 05 January 2026

REVISED 25 March 2026

ACCEPTED 30 March 2026

PUBLISHED 17 April 2026

CITATION

Zhao R, Xi NM, Gabby L, Gilbert ER,
Vanarsa K, Qiao M, Zhang D, Zhang J,
Mohan C, Judson MA, Koth LL and
Ji H-L (2026) New proteomic
biomarkers identified in plasma
extracellular vesicles in sarcoidosis: a
case-control matched study.
Front. Immunol. 17:1779835.
doi: 10.3389/fimmu.2026.1779835

COPYRIGHT

© 2026 Zhao, Xi, Gabby, Gilbert, Vanarsa,
Qiao, Zhang, Zhang, Mohan, Judson, Koth
and Ji. This is an open-access article
distributed under the terms of the
[Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/)
(CC BY). The use, distribution or
reproduction in other forums is
permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication
in this journal is cited, in accordance
with accepted academic practice. No
use, distribution or reproduction is
permitted which does not comply with
these terms.

New proteomic biomarkers identified in plasma extracellular vesicles in sarcoidosis: a case-control matched study

Runzhen Zhao ^{1,2†}, Nan Miles Xi ^{1,3†}, Lea Gabby⁴,
Emily R. Gilbert⁵, Kamala Vanarsa ⁴, Mark Qiao¹, Dee Zhang¹,
Jiwang Zhang ⁶, Chandra Mohan ⁴, Marc A. Judson ⁷,
Laura L. Koth ⁸ and Hong-Long Ji ^{1,2,9*}

¹Department of Surgery, Stritch School of Medicine, Loyola University Chicago Health Sciences
Division, Maywood, IL, United States, ²Burn and Shock Trauma Research Institute, Stritch School of
Medicine, Loyola University Chicago Health Sciences Division, Maywood, IL, United States,

³Department of Mathematics and Statistics, Loyola University Chicago, Chicago, IL, United States,

⁴Biomedical Engineering & Medicine, University of Houston, Houston, TX, United States, ⁵Department
of Medicine, Loyola University Chicago, Maywood, IL, United States, ⁶Department of Cancer Biology,
Loyola University Chicago, Maywood, IL, United States, ⁷Department of Medicine, Albany Einstein
College of Medicine, Albany, NY, United States, ⁸Department of Medicine, University of California, San
Francisco, San Francisco, CA, United States, ⁹Department of Microbiology and Immunology, Stritch
School of Medicine, Loyola University Chicago Health Sciences Division, Maywood, IL, United States

Background: Sarcoidosis is a heterogeneous disease with unknown mechanisms, nonspecific therapies, and multiple etiologies. The role of blood extracellular vesicles (EVs) in the diagnosis and pathogenesis of sarcoidosis remains obscure. AIMS. This study aims to test the hypothesis that the EV proteins in the blood can serve as phenotypic biomarkers of sarcoidosis.

Methods: We combined EV proteomics with machine learning algorithms to identify and prioritize biomarkers, enrich their functions, and cluster networks in case-control matched ACCESS patients.

Results: In total, 278 plasma EV proteins were significantly upregulated or downregulated in 40 sarcoidosis patients compared with 40 matched healthy controls. We identified 97 proteins that could serve as biomarkers with an AUC > 0.75. Of these, the AUC was > 0.90 for 13 proteins. 62 differentially expressed EV proteins strongly correlated with 20 clinical variables of severity, chest X-ray findings, and/or laboratory results. Functional annotation and network analysis suggest that these differentially expressed proteins regulate endocytosis, host responses to external stimuli, and transcription processes. Moreover, the top three ranked pathways were clathrin-mediated endocytosis, Hsp90 chaperone cycle, and spliceosome.

Conclusions: This study demonstrates that plasma EV proteins can serve as biomarkers of various clinical phenotypes of the disease.

KEYWORDS

biomarker, extracellular vesicles, machine learning, proteome wide, sarcoidosis

Introduction

Inflammatory sarcoidosis is characterized by the formation of granulomas in involved tissues, with the chest being the most commonly affected (~90%) (1, 2). The severity of lung disease can be classified according to the Scadding stage. Besides the lungs, organs that are commonly involved with sarcoidosis include the skin, eyes, liver, lymph nodes, salivary glands, bones, joints, muscles, spleen, nervous system, kidneys, sinuses, and heart. Approximately 20% of patients are progressive and tend to develop lung fibrosis (1, 3–5). The leading cause of death from sarcoidosis is respiratory failure associated with fibrotic lungs. The etiologies of sarcoidosis remain unclear despite its identification over a century ago. The prevalence, symptoms, potential triggers, and prognosis of the condition can vary significantly. Using traditional non-omics approaches (laboratory tests) and comparing statistical analysis, the following diagnostic biomarkers for sarcoidosis have been identified and validated: serum soluble interleukin-2 receptor (sIL-2R) (6–13), urinary U-8-OHdG (14–16), serum angiotensin-converting enzyme (ACE) (7–9, 13), serum chitotriosidase (17–19), serum KL-6 (20, 21), serum CRP (9, 21), and serum BNP (22, 23). IL12, IL18, MMP14, CTSS, amyloid A, ZNF688, ARFGAP1, CD14, LBP, α -2chain of haptoglobin, and PHA in serum and BAL may be useful biomarkers to distinguish sarcoidosis from healthy controls. In addition to serving as diagnostic biomarkers, some have also been suggested to predict lung function, inflammation, multiple organ involvement, organ failures, chronicity, and response to therapy. However, no study has evaluated the feasibility of identifying biomarkers in blood extracellular vesicles (EVs) for sarcoidosis.

EVs can be isolated from as little as 5 microliters of human liquid samples (24). Utilizing the purified EVs for omics studies offers the following advantages. First, the excessive plasma or serum proteins (e.g., albumin) can be removed without depleting the top 20 high-abundant proteins. Second, EV proteins exhibit long-term stability compared to bulk proteins, resulting in high detectability in EV proteins compared to cell-free plasma or serum (25–29). Third, EVs are a novel approach to identifying biomarkers, studying disease mechanisms, and clustering phenotypes of heterogeneous diseases (26, 28, 30–38). These studies combine advanced omics and machine learning algorithms. As recently reviewed, the feasibility of using EV proteomics as a strategy to study biomarker identification and molecular pathology in sarcoidosis remains an unanswered question (39). Regarding the application of plasma EV proteomic to biomarker discovery for sarcoidosis, it is still at its early stage. A

targeted serum EV proteomic study identified LBP and CD14 could be diagnostic biomarkers for sarcoidosis (13). Very recently, Kraaijvanger and colleagues reported that serum EV CHI3L1 and CPA1 were predictors for the response to prednisone and methotrexate, respectively, in pulmonary sarcoidosis (40). Reduction in serum EV serpin C1 may be a sign for the response to methotrexate too (41).

This study aimed to identify differentiated biomarkers from healthy controls and their associations with clinical sarcoidosis phenotypes through the integration of high-throughput unbiased proteomics of plasma EVs and advanced machine learning algorithms. Our findings illustrate that, for the first time, novel biomarkers associated various phenotypes of the disease have been discovered through the analysis of ACCESS patients.

Materials and methods

Patient cohorts

The participants were selected from the ACCESS (A Case Controlled Etiologic Study of Sarcoidosis) clinical trial (42–46). ACCESS was a case-control study of sarcoidosis where the clinical phenotype of sarcoidosis cases was described in detail and the control subjects were extremely well-matched to the cases. The cohort comprised of 40 sarcoidosis patients and 40 healthy controls matched for age (within 5 years) sex, race, and socioeconomic status/place of residence. The controls were all the same sex, same age within 5 years, and same race as the cases. In addition, random digit dialing was used to include matched controls. With random digit dialing, each case's area code and prefix part of the phone number (the next 3 digits) were fixed and then the last 4 digits were dialed. For example, if the case's phone number was (212)-787-3498, then a random number of the last 4 digits of the phone number would be called: (212)-787-XXXX, where the last 4 digits of the phone number were randomly selected. If someone answered the phone, the person would be asked if anyone in their house was the same sex, same age (within 5 years), and same race as the patient. If there was such a person, they would be asked to participate. On average, it took almost 100 calls to get a match of controls. This protocol used to control to some extent for socioeconomic status.

The inclusion and exclusion criteria mirrored those of the ACCESS trial. The severity of sarcoidosis and dyspnea was defined by Scadding stages while the severity of dyspnea was scaled with a standardized 5-point grading system. The plasma samples and clinical dataset collected by the ACCESS trial were obtained through BioLINCC (Biologic Specimen and Data Repository Information Coordinating Center). The use of the plasma samples and de-identified clinical dataset was approved by the Institutional Review Board (IRB) of Loyola University Chicago (LU#216964). Patients were categorized based on their Scadding stage, organ involvement, chest radiographic findings, dyspnea score, pulmonary function, and blood tests. The collection, storage, and shipping of plasma samples were conducted according to the standard

Abbreviations: ACE, angiotensin-converting enzyme; ACCESS, A Case Controlled Etiologic Study of Sarcoidosis; AUC, area under the curve; BALF, bronchoalveolar lavage fluid; BUN, blood urea nitrogen; CRP, C-reactive protein; CTSS, cathepsin S; DEP, differentially expressed proteins; EGF, epidermal growth factor; EV, extracellular vesicles; FEV1, forced expiratory volume in 1 second; FEVPRD, FEV1% predicted; FVC, forced vital capacity; FVCPRD, forced vital capacity predicted; LBP, lipopolysaccharide-binding protein; LC-MS, liquid chromatography–mass spectrometry; MMP14, matrix metalloproteinase 14; NPV, negative predictive value; PERFEV, FEV-1% predicted; PERFVC, FVC % predicted; PLK2, polo-like kinase 2; PPV, positive predictive value; ROC, receiver operating characteristics.

procedures recommended by the National Institutes of Health (NIH).

Extraction of plasma extracellular vesicles

EVs from plasma samples were captured and processed by Tymora Analytical Operations (West Lafayette, IN) using magnetic EVtrap beads as previously described (47). The isolated and dried EV samples were lysed to extract proteins using the phase-transfer surfactant (PTS)-aided procedure (48). The proteins were reduced and alkylated by incubating them in 10 mM TCEP and 40 mM CAA for 10 min at 95 °C. The samples were then diluted fivefold with 50 mM triethylammonium bicarbonate and digested with Lys-C (Wako) at a 1:100 (wt/wt) enzyme-to-protein ratio for 3 h at 37 °C. Trypsin was added at a final 1:50 (wt/wt) enzyme-to-protein ratio for overnight digestion at 37 °C. To remove the PTS surfactants from the samples by acidification, trifluoroacetic acid (TFA) and an ethyl acetate solution were added to a final concentration of 1% TFA and at a 1:1 ratio, respectively. The mixture was vortexed for 2 min and then centrifuged at $16,000 \times g$ for 2 min to separate the aqueous and organic phases. The organic phase (top layer) was removed, and the aqueous phase was collected. This step was repeated once more. The samples were dried in a vacuum centrifuge and desalted using Top-Tip C18 tips (Glygen) according to the manufacturer's instructions. A portion of each sample was used to determine the peptide concentration using the Pierce Quantitative Colorimetric Peptide Assay. Finally, the samples were dried completely in a vacuum centrifuge and stored at -80 °C.

LC-MS/MS analysis

Each dried peptide sample was dissolved at 0.1 µg/µL in 0.05% trifluoroacetic acid with 3% (vol/vol) acetonitrile. 10 µL of each sample was injected into an Ultimate 3000 nano UHPLC system (Thermo Fisher Scientific). Peptides were captured on a 2-cm Acclaim PepMap trap column and separated on a heated 50-cm column packed with ReproSil Saphir 1.8 µm C18 beads. The mobile phase buffer consisted of 0.1% formic acid in ultrapure water (buffer A) with an eluting buffer of 0.1% formic acid in 80% (vol/vol) acetonitrile (buffer B) run with a linear 60-min gradient of 6–30% buffer B at a flow rate of 300 nL/min. The UHPLC was coupled online with a Q-Exactive HF-X mass spectrometer (Thermo Fisher Scientific). The mass spectrometer was operated in the data-dependent mode, in which a full-scan MS (from m/z 375 to 1,500 with a resolution of 60,000) was followed by MS/MS of the 15 most intense ions (30,000 resolution; normalized collision energy - 28%; automatic gain control target (AGC) - $2E4$, a maximum injection time - 200 ms; 60-sec exclusion).

Data processing. The raw files were searched directly against the human UniProt database without redundant entries, using Byonic (Protein Metrics) and Sequest search engines loaded into Proteome Discoverer 2.3 software (Thermo Fisher Scientific). MS1 precursor mass tolerance was set at 10 ppm, and MS2 tolerance was set at 20 ppm. Search criteria included a static carbamidomethylation of cysteines (+57.0214 Da), variable modifications of oxidation (+15.9949 Da) on methionine residues, and acetylation (+42.011

Da) at the N-terminus of proteins. The search was performed with full trypsin/P digestion, allowing a maximum of two missed cleavages on the peptides analyzed from the sequence database. The false discovery rates (FDR) of proteins and peptides were set at 0.01. All protein and peptide identifications were grouped, and any redundant entries were removed. Unique peptides and unique master proteins were reported.

Data preprocessing and profiling

The dataset of protein abundance generated from LC-MS contained a data matrix where each row represented a protein and each column a participant. Before statistical and machine learning analyses, the following preprocessing procedures were conducted on the data matrix to create a clean dataset: removing proteins that were not expressed in either the patient or control group, normalizing protein abundance through variance stabilization normalization (VSN) using the justvs package in R (49), and imputing missing values using the random forest method with the R package missForest (50). The Wilcoxon rank sum test was used to compare the differences in abundance between patients and controls using the R function `wilcox.test`. The Benjamini-Hochberg (BH) method was utilized to control the FDR with the R function `p.adjust`. The fold change (FC) in the mean protein abundance between patients and controls was calculated. The differentially expressed proteins (DEP) were visualized in a heatmap with hierarchical clustering, volcano plot, and PCA plot using R packages ComplexHeatmap (51), EnhancedVolcano, and ggplot2 (52).

Machine learning approaches for identifying biomarkers

To identify EV biomarkers, we trained a univariate logistic regression model using individual DEP to predict sarcoidosis patients and healthy controls. The logistic regression model was implemented using the R function `glm`. To assess the predictive performance of each protein, a five-fold cross-validation approach was repeated 100 times to calculate the area under the receiver operating characteristic (AUC), accuracy, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV). The probability threshold was set at 0.5. The differences in the protein expression levels of biomarker candidates between patients and controls were assessed using the Wilcoxon rank sum test. The visualization of AUC and protein expression was created using R package ggplot2 (52).

Correlation analysis for the clinical relevance of biomarker candidates

To determine the clinical significance of the identified diagnostic biomarkers, their correlations with critical clinical variables were analyzed. The clinical variables include spirometry, Scadding stage, dyspnea score, blood tests, and chest X-ray readouts. Pearson correlation coefficients, p-values, and the slope of the linear regression line for continuous variables were calculated using the R function `cor.test`. The p-values were adjusted using the BH method to control the FDR. A significant correlation was considered if the adjusted p-value was < 0.05 .

or the Pearson correlation coefficient was > 0.4 and p -value < 0.05 . The Wilcoxon rank sum test was conducted on the FC value of mean protein abundance to analyze the correlations between categorical variables and the identified biomarkers. The criteria for determining the significant correlations between biomarkers and categorical clinical variables were either an adjusted p -value < 0.05 or an absolute FC > 2.0 along with a p -value < 0.05 . The linear regression was visualized using the `geom_smooth` function in the R package `ggplot2` (52).

Pathway enrichment and network analysis

As previously described (53), the R package `pathfindR` was used to enrich pathways for the identified biomarkers with an AUC > 0.750 (54). `PathfindR` considered the significance levels of individual biomarker proteins and utilized a protein-protein interaction network to enhance pathway enrichment outcomes. The `run_pathfindR` function in `pathfindR` package was executed to identify enriched pathways as defined in each database. To integrate the enriched pathways from different databases, pathway terms with an adjusted p -value of < 0.05 and at least two protein hits were included. The enriched pathways were hierarchically clustered by executing the `cluster_enriched_terms` function in the `pathfindR` package. The optimal number of clusters was determined by maximizing the average silhouette width. The pathway term with the lowest adjusted p -value in each cluster was selected as the representative pathway for that cluster and was included in the final enriched pathway result. The `term_gene_graph` function in `pathfindR` package was applied to create a network plot of the top-ranked pathways linked with associated biomarker proteins. The `Heatmap` function in the R package `ComplexHeatmap` was executed to perform hierarchical clustering of the biomarker proteins associated with the representative pathways (51).

Determination of tissue and cellular origin

TopGene was utilized to search the tissue and cellular origin of the DEPs (55). The hits of subcellular populations and tissues/organs were quantified and graphed as a function of their adjusted p -values.

Sample size determination. The power analysis and calculation of the sample size were conducted using G power (version 3.1.9.7) (56). A two-tailed Wilcoxon-Mann-Whitney test was used to compare between-group differences in protein abundance and clinical variables. The expected effect size was set to 0.84 with a power of 95% and an FDR of 5%. Consequently, the sample size for sarcoidosis patients was set at 40 and for healthy controls at 40, to achieve the required power.

Results

Baseline characteristics of participants

There were 80 participants (40 sarcoidosis cases and 40 controls) in the cohort (Figure 1). There were no significant differences in demographic and baseline variables between the controls and the

patient group (Table 1). The patients were represented at all Scadding stages (from stage 0 to stage 4). The predominant comorbidities included disorders of the heart, lungs, kidney, liver, rheumatologic system, neurologic system, endocrine system, and cancer. These comorbidities did not show significant differences, as eliminated by case-control matching and inclusion criteria.

Proteomic profile of the discovery cohort

The high-throughput LC-MS pipeline successfully identified 2,312 proteins in the plasma samples. Following the removal of proteins with low abundance, along with data normalization and imputation, a total of 1,659 high-quality proteins remained. Among them, 278 proteins exhibited notable differences in abundance (adjusted p -value < 0.05 , $|FC| > 1.2$), named as differentially expressed proteins (DEPs) (Figure 2A). Of these, 118 were upregulated, and 160 were downregulated compared to the controls. The upregulated DEPs were hierarchically clustered into 4 subsets on the heatmap, while the downregulated DEPs were clustered into 6 subgroups. As depicted on the volcano plot (Figure 2B), the most downregulated protein in sarcoidosis was calpain-1 catalytic subunit (P07384, $\log_2FC$, -4.9 and $\log_{10}P$, 14.9), followed by acyl-CoA-binding domain-containing protein 6 (Q9BR61, $\log_2FC$, -3.6 and $\log_{10}P$, 6.5). The most increased proteins in sarcoidosis were uncharacterized protein C6orf132 (Q5T0Z8, $\log_2FC$, 1.7 and $-\log_{10}P$, 11) and zinc finger protein 607 (Q96SK3, $\log_2FC$, 2.5 and $-\log_{10}P$, 7.0). The first and second principal components explained approximately 22.3% and 10.4% of the total variance, respectively (Figure 2C). The lack of overlap between the two groups indicates that these proteins hold great promise as diagnostic biomarkers to differentiate sarcoidosis patients from healthy controls.

Tracking tissue and cellular origin

EVs can be released by various tissues and cells into the bloodstream. To track the tissue origin of DEPs in EVs, we searched the TopGene database (Figure 3). The top-10 contributing tissues to the identified DEPs were peripheral blood mononuclear cells (PBMC), bronchoalveolar lavage fluid (BALF), epithelium, spleen, lymph node, endothelium, small intestine, airway, lung, and blood. Similarly, the top-ranked cells included platelets, megakaryocytes, dendritic cells, neutrophils, myeloids, club cells, monocytes, CD14⁺ cells, macrophages, and aerocytes. It appears that the majority of EVs in the plasma could be released by circulating inflammatory cells and injured alveolar epithelial cells.

Identification of diagnostic biomarkers for sarcoidosis

To identify EV biomarkers that differentiate sarcoidosis patients from healthy controls, we calculated the area under the curve (AUC) of receiver operating characteristics (ROC) curve for the 278 DEPs individually. The AUC values ranged from 0.583 to 0.971, with 97 proteins having an AUC value above 0.750. Of note, the AUC exceeded 0.900 for the top-13 ranked biomarker candidates (Table 2). These 13 biomarkers exhibited high predictive performance for diagnosis, as reflected by their corresponding sensitivity,

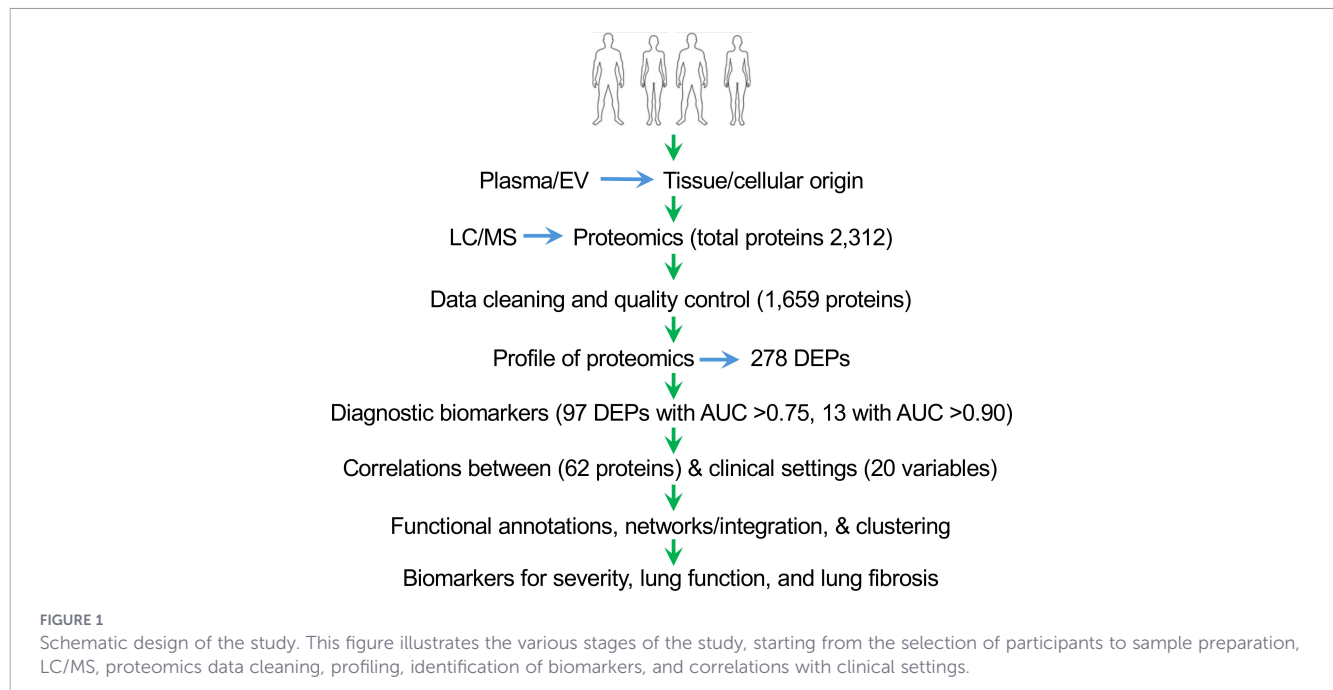


TABLE 1 Baseline characteristics of participants.

variable	Controls (%)	Cases (%)
Total	40	40
Male	14 (35)	14 (35)
Female	26 (65)	26 (65)
Age(y)		
< 30	5 (12.5)	6 (15)
30-39	11 (27.5)	10 (25)
40-49	14 (35)	14 (35)
50-59	7 (17.5)	7 (17.5)
> 60	3 (7.5)	3 (7.5)
White	20 (50)	20 (50)
Black	20 (50)	20 (50)
Family history	0 (0)	4 (10)
Never smoke	15 (37.5)	18 (45)
Former smoker	25 (62.5)	22 (55)
Current smoker	11 (27.5)	5 (12.5)
Comorbidities		
Cardiovascular system	11 (27.5)	14 (35)
Pulmonary system	4 (10)	9 (22.5)
Cancer	4 (10)	3 (7.5)
Endocrine	4 (10)	4 (10)
Kidney	0 (0)	1 (2.5)
Liver	0 (0)	2 (5)
Rheumatologic	21 (52.5)	20 (50)
Neurologic	0 (0)	0 (0)

(Continued)

TABLE 1 Continued

variable	Controls (%)	Cases (%)
Scadding stage		
0	-	5
1	-	20
2	-	8
3	-	4
4	-	3

specificity, accuracy, positive predictive values (PPV), and negative predictive values (NPV). Their functions include anti-infection, host inflammatory responses, cell proliferation, and glucose metabolism. [Figure 4](#) illustrates the ROC curves of the top-10 candidate biomarkers, including calpain-1 catalytic subunit (P07384, AUC 0.971, sensitivity 100%, specificity 92.5%), probable maltase-glucoamylase 2 (Q2M2H8, 0.962, 95.0%, 92.5%), and synaptotagmin-like protein 4 (Q96C24, 0.947, 92.5%, 95.0%). The protein abundance of these biomarker candidates exhibited a significant difference between patients and controls. Except for toll-like receptor 8 (Q9NR97) and dynein axonemal heavy chain 8 (Q96JB1), the other 8 biomarker proteins, including Calpain-1 catalytic subunit (P07384), Probable maltase-glucoamylase 2 (Q2M2H8), Synaptotagmin-like protein 4 (Q96C24), Protein argonaute-3 (Q9H9G7), Metal-response element-binding transcription factor 2 (Q9Y483), Protein WWC3 (Q9ULE0), Selenoprotein P (P49908), Fukutin related protein (M0QYV8), Uncharacterized protein C6orf132 (Q5T0Z8), Lysozyme C (P61626), and Testin (Q9UGI8), showed a marked reduction ($p < 0.0001$) in the patient group. Our study suggests that both upregulated and downregulated EV proteins in the plasma could serve as diagnostic biomarkers.

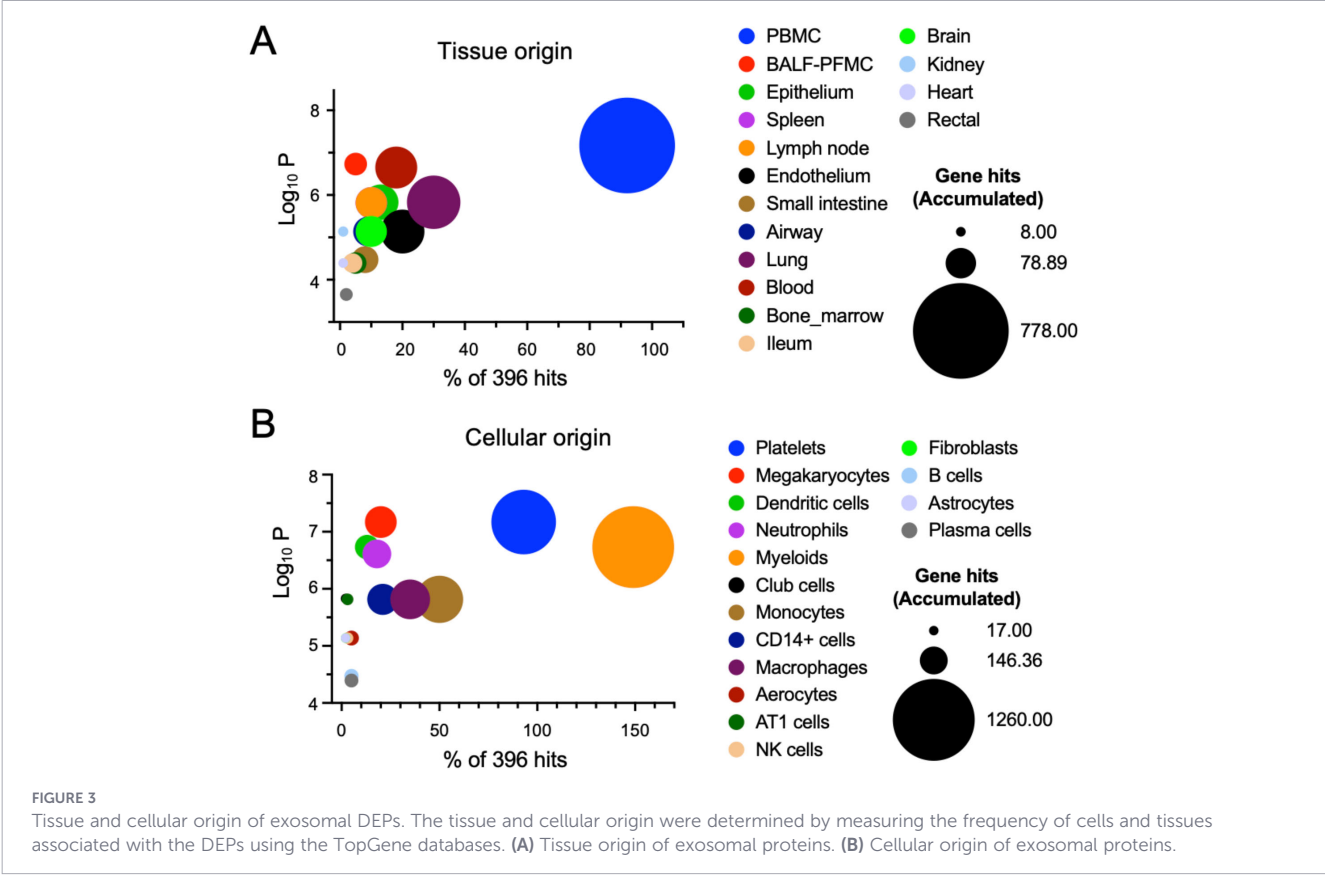
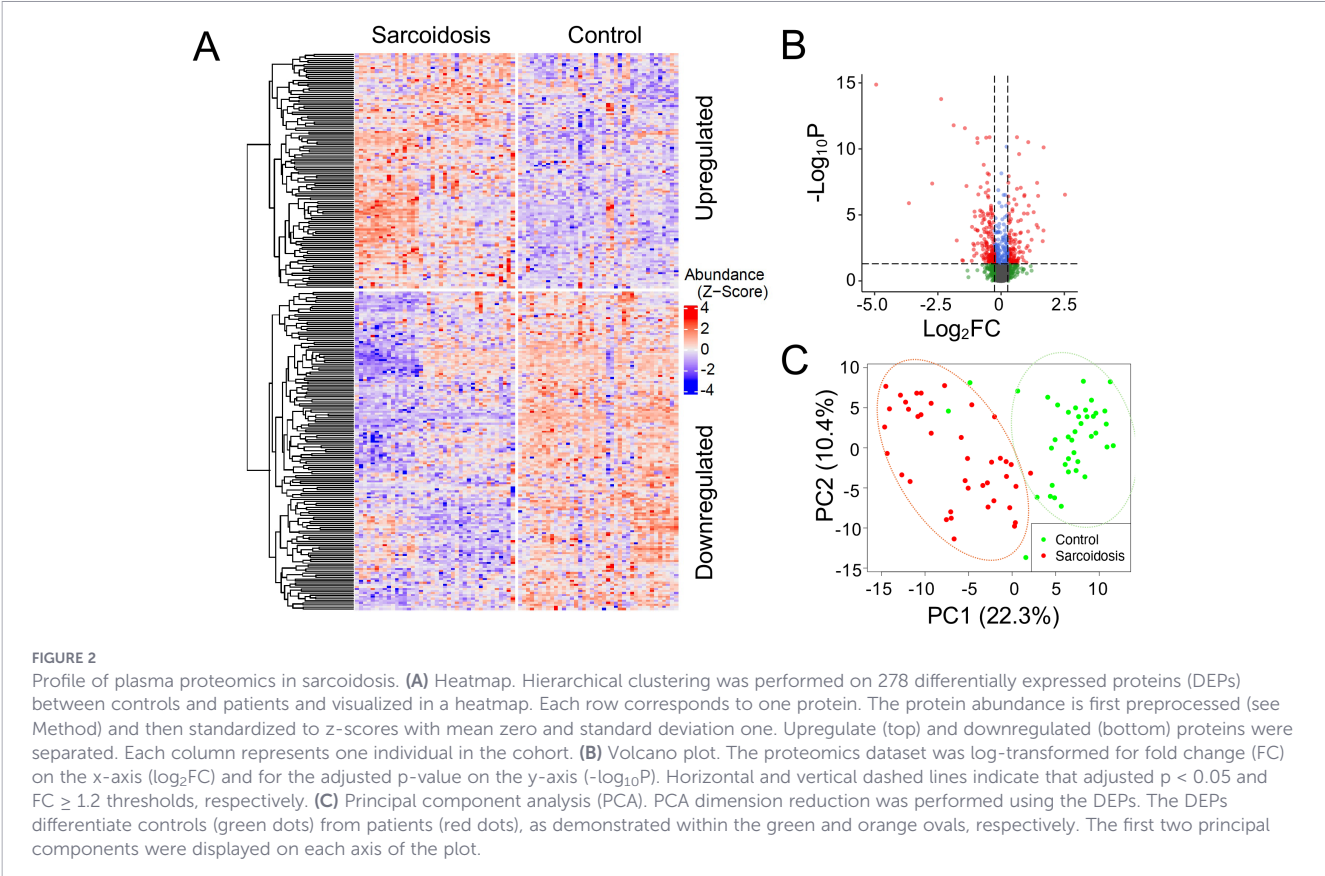


TABLE 2 Prediction performance of top 13 potential biomarkers.

Biomarker	Description	AUC	SNS (%)	SPC (%)	Accuracy	PPV	NPV	Function
P07384 CAPN1	Calpain-1 catalytic subunit	0.971	100.0	92.5	0.963	0.930	1.000	Ca-regulated non-lysosomal thiol-protease, which catalyzes substrates involved in cytoskeletal remodeling and signal transduction
Q2M2H8 MGAM2	Probable maltase-glucoamylase 2	0.962	95.0	92.5	0.938	0.927	0.949	D-glucose metabolism
Q96C24 SYTL4	Synaptotagmin-like protein 4	0.947	92.5	95.0	0.938	0.949	0.927	Modulates exocytosis of dense-core granules and secretion of hormones in the pancreas and the pituitary
Q9H9G7 AGO3	Protein argonaute-3	0.930	82.5	87.5	0.850	0.868	0.833	Required for RNA-mediated gene silencing (RNAi)
Q9Y483 MTF2	Metal-response element-binding transcription	0.928	92.5	85.0	0.888	0.860	0.919	transcriptional networks, methylation activity
Q9ULE0 WWC3	Protein WWC3	0.925	87.5	82.5	0.850	0.833	0.868	Hippo signaling, cell fate
Q9NR97 TLR8	Toll-like receptor 8	0.924	80.0	90.0	0.850	0.889	0.818	innate and adaptive immunity
P49908 SELENOP	Selenoprotein P	0.924	90.0	92.5	0.913	0.923	0.902	responsible for extracellular antioxidant defense of selenium
Q96JB1 DNAH8	Dynein axonemal heavy chain 8	0.919	95.0	82.5	0.888	0.844	0.943	produces force towards the minus ends of microtubules, sperm motility, ATPase activity
M0QYV8 RKR	Fukutin related protein	0.914	92.5	85.0	0.888	0.860	0.919	n/a
Q5T0Z8 C6orf132	Uncharacterized protein C6orf132	0.904	70.0	90.0	0.900	0.882	0.783	n/a
P61626 LYZ	Lysozyme C	0.903	82.5	82.5	0.825	0.825	0.825	bacteriolytic function of monocytes and macrophages
Q9UGI8 TES	Testin	0.900	82.5	87.5	0.875	0.868	0.833	cell adhesion, cell spreading, cell proliferation, actin cytoskeleton reorganization, tumor suppression

Area under the curve (AUC), sensitivity (SNS), specificity (SPC), accuracy, PPV (positive predictive value), and NPV (negative predictive value). The information on these proteins, including ID, description, and function, is retrieved from the UniProt database.

Correlations between biomarker candidates and clinical variables

The clinical assessment of sarcoidosis relies on symptoms, chest imaging, lung functional tests, electrophysiological tests, and blood tests. We hypothesize that the identified EV biomarker candidates correlate with these clinical tests. Pearson correlation coefficient was calculated between the protein biomarkers and the following clinical variables: blood tests, chest X-rays, spirometry, radiographic Scadding stage, and dyspnea scale (Figure 5). Three proteins, namely Pulmonary surfactant-associated protein B (P07988), DnaJ homolog subfamily A member 2 (O60884), and Immunoglobulin heavy variable 6-1 (A0A0B4J1U7), were found to correlate with the Scadding stage ($p < 0.05$, $|R| > 0.4$). Serine/threonine-protein kinase PLK2 (Q9NYY3) correlated with the dyspnea scale. Regarding the lung functional tests, 16 proteins were correlated with 7 clinical variables, including FEV1 (correlated with 3 proteins), FEVPRD (4 proteins), FVC (3 proteins), FVCPRD (8 proteins), PERFVC (1 protein), PERFVC (1 protein), and FEV/FVC ratio (3 proteins). Then, we validated these correlations by analyzing the links between chest X-rays and biomarker candidates. In total, there were 39 proteins correlated with 6 X-ray readouts ($p < 0.05$, $|FC| > 1.4$). These readouts included lymphadenopathy (corrected with 7 proteins), alveolar infiltrates (9 proteins), parenchymal bulla or blebs (4 proteins), hilar retraction (17 proteins), interstitial infiltrate (1

protein), and importantly, pulmonary fibrosis (3 proteins). Third, blood tests were correlated with 8 proteins ($p < 0.05$, $|R| > 0.5$). These blood tests were creatinine level (4 proteins), WBC (1 protein), eosinophils (1 protein), potassium content (1 protein), and blood urea nitrogen (BUN, 1 protein). In total, 54 EV proteins were identified that correlated with the analyzed clinical variables (Figure 5D). Among them, 8 proteins were shared by the three subgroups of clinical variables. More than half of the correlated proteins (29 out of 54) were identified as potential biomarkers ($AUC > 0.750$ as a broadly used cutoff criterion), with 2 of them, namely, protein WWC3 (Q9ULE0) and uncharacterized protein C6orf132 (Q5T0Z8) being top-ranked candidates ($AUC > 0.90$). These two proteins were correlated with alveolar infiltrates and hilar retraction, respectively. Furthermore, the correlations between the selected 10 biomarkers ($AUC > 0.750$) and continuous clinical variables were visualized using linear regression (Figure 6A). These 9 proteins were serine/threonine-protein kinase PLK2 (Q9NYY3), small ribosomal subunit protein eS8 (P62241), kinesin-1 heavy chain (P33176), chitinase-3-like protein 1 (P36222), peptidyl-prolyl cis-trans isomerase D (Q08752), DNAJ homolog subfamily A member 2 (O60884), cytoplasmic tryptophan-tRNA ligase (P23381), thrombospondin-4 (P35443), and uncharacterized protein C16orf46 (Q6P387). Two-thirds of the correlations were negative, while one-third were positive. Among the correlations, the slope of 6 plots was above 1.0 and up to 7.0. These results suggest a strong link between differential proteins

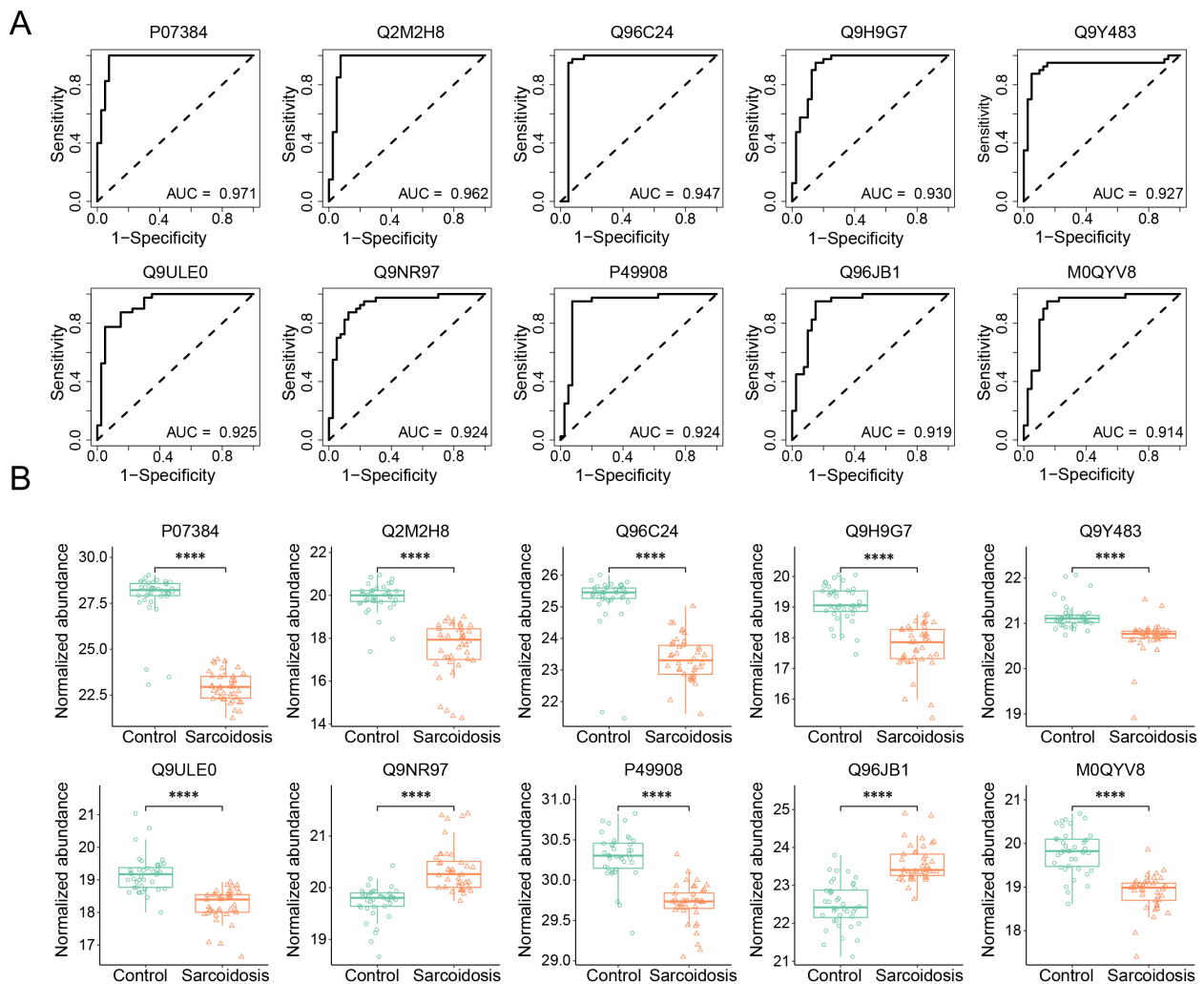


FIGURE 4

Identification of diagnostic biomarkers and comparison of their expression levels. **(A)** Top 10 biomarker candidates with an AUC value >0.900. Performance metrics were averaged over 100 repetitions of the five-fold cross-validation (CV) using univariable logistic regression. The protein biomarkers were graphed in descending order of AUC value from left to right. PPV, positive predictive value. NPV, negative predictive value.

(B) Expression levels of the top-10 identified biomarkers. The box and whisker plots showed the differences in protein abundance between controls and patients. The first quartile, median, third quartile, range, and normalized protein abundance **** adjusted p < 0.0001, Wilcoxon rank-sum test.

and clinical variables, suggesting these EV proteins could function as clinical biomarkers.

Different expression levels of biomarkers correlated with X-ray readouts

Abnormal X-ray readouts were correlated with specific biomarkers. Notably, a significant difference existed in the expression levels of correlated protein biomarkers between patients with and without abnormal X-ray findings (Figure 6B). These 10 biomarker proteins exhibited an AUC value > 0.774, namely, protein WWC3 (Q9ULE0), uncharacterized protein C6orf132 (Q5T0Z8), EGF-containing fibulin-like extracellular matrix protein 1 (Q12805), Rho-associated protein kinase 1 (Q13464), serine/threonine-protein kinase PLK2 (Q9NYY3), NEDD4-binding protein 2 (Q86UW6), protein disulfide-isomerase A4 (P13667), apolipoprotein A-IV (P06727), coagulation factor VIII (P00451), and protein

argonaute-2 (Q9UKV8). Half of these biomarker proteins were upregulated in patients with abnormal chest X-ray reports: 2 for pulmonary fibrosis, 2 for hilar retraction, and 1 for lymphadenopathy. In comparison, the expression of 6 biomarker proteins was downregulated in patients with alveolar infiltrates ((protein WWC3 (Q9ULE0), Rho-associated protein kinase 1 (Q13464), and apolipoprotein A-IV (P06727)), lymphadenopathy ((EGF-containing fibulin-like extracellular matrix protein 1 (Q12805) and NEDD4-binding protein 2 (Q86UW6)), and hilar retraction (protein disulfide-isomerase A4, P13667). Protein Q86UW6 was correlated with lung fibrosis, hilar retraction, and lymphadenopathy. Pulmonary fibrosis showed increased expression of serine/threonine-protein kinase PLK2 (Q9NYY3) and NEDD4-binding protein 2 (Q86UW6), while alveolar infiltrates exhibited reduced expression of protein WWC3 (Q9ULE0), Rho-associated protein kinase 1 (Q13464), and apolipoprotein A-IV (P06727). These biomarkers would hold the potential to predict pulmonary sarcoidosis.

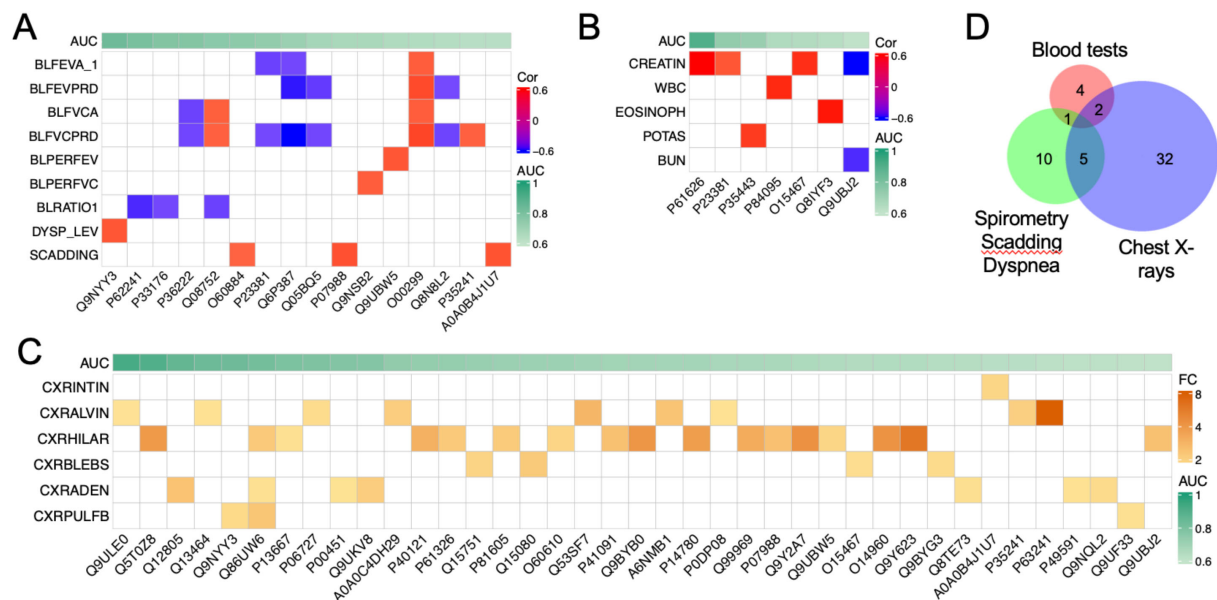


FIGURE 5

Correlations between identified biomarkers and three subgroups of clinical variables. Correlations between biomarker candidates and clinical variables were sorted by their AUC values. (A) Correlation matrix between spirometry and biomarker candidates. Sixteen proteins correlated with 7 spirometry variables, Scadding stage (SCADDING), and dyspnea scale (DYSP_LEV). Seven spirometry variables were pre-bronchodilators FEV-1 (BLFEVA_1), FEV-1 predicted (BLFEVPRD), pre-bronchodilators FVC (BLFVCA), FVC predicted (BLFVCPRD), FEV-1% predicted (BLPERFEV), FVC-1% predicted (BLPERFVC), FEV-1/FVC (BLRATIO1). p -value < 0.05, Pearson correlation coefficient > 0.4 (shown by colors). (B) Correlations between blood tests and biomarker candidates. Seven proteins correlated with 5 variables of blood tests. The correlated blood tests included creatinine level (CREATIN, mg/dL), white blood cells (WBC, $\times 10^3/\text{mm}^3$), eosinophils (EOSINOPH, %), plasma potassium (POTAS, mEq/L), and blood urea nitrogen level (BUN, mg/dL). P -value < 0.005, Pearson correlation coefficient > 0.5. (C) Correlation matrix between biomarker candidates and chest X-ray readouts. Six image readouts and 39 biomarker candidates were correlated. The six x-ray readouts were interstitial infiltrates (CXRLIN), alveolar infiltrates (CXRLVIN), Hilar retraction (CXRHILAR), Bullae or blebs (CXRBLEBS), adenopathy (CXRADEN), and pulmonary fibrosis (CXRPULFB). P -values < 0.05, FC > 2.0 for categorical variables. (D) Venn plot. Unique and shared biomarker proteins correlated to the three subgroups of clinical variables are illustrated.

Functions and pathway enrichment networks of biomarker candidates

To enrich the functions and network of the biomarker candidates with an AUC > 0.750, hierarchical clustering of the functional pathways, visualization of top-ranked biomarker candidates, and network analysis were performed (Figure 7). The top three pathways were clathrin-mediated endocytosis, Hsp90 chaperone cycle for steroid hormone receptors in the presence of ligand, and spliceosome (Figure 7A). The top 30 pathways could be grouped into three subgroups: endocytosis, responses to external stimuli, and transcription, which were enriched by 9, 13, and 5 biomarker proteins, respectively (Figure 7B). Moreover, both upregulated and downregulated biomarker proteins were analyzed for their network interactions (Figure 7C). There were 2 pathways in the group for endocytosis: clathrin-mediated endocytosis and endocytosis, 6 pathways in the group of response to external stimuli, and 2 in the group for transcription. Interactions of the “endocytosis” and “response to external stimuli” were mediated by 4 biomarker proteins, namely, F-actin-capping protein subunit alpha-2 (P47755), HLA class I histocompatibility antigen B alpha chain (D9J307), HLA class I antigen (Q53Z42), and E3 ubiquitin-protein ligase CBL (A0A0U1RQX8). These proteins were shared by 2 endocytosis pathways and 3 pathways belonging to the group of “response to external stimuli”: Hsp90 chaperone cycle, antigen

processing and presentation, and the VEGFR1 pathway. In contrast, 2 pathways in the “transcription” group did not interact with each other or with the pathways from the other two groups. The functional enrichment and network analysis suggest that clathrin-mediated endocytosis, Hsp90, and spliceosome could play a major role in the development of sarcoidosis.

Discussion

The utilization of extracellular vesicles (EVs) in conjunction with proteomics facilitates the identification of potential diagnostic plasma biomarkers for sarcoidosis. Our application of EV proteomics represents an innovative method for the discovery of novel biomarkers. In addition, these biomarkers validated previously identified biomarkers of disease severity in sarcoidosis. The tissue and cellular origins of plasma EV proteins are systemic. The identified EV biomarkers exhibit a strong correlation with essential clinical variables related to disease severity, laboratory findings, pulmonary function test findings, and the severity of dyspnea. The signaling pathways and networks associated with the identified EV biomarkers may provide mechanistic insights of the disease. This study represents the initial demonstration of the efficacy of EV omics as a high-throughput approach for biomarker discovery in sarcoidosis.

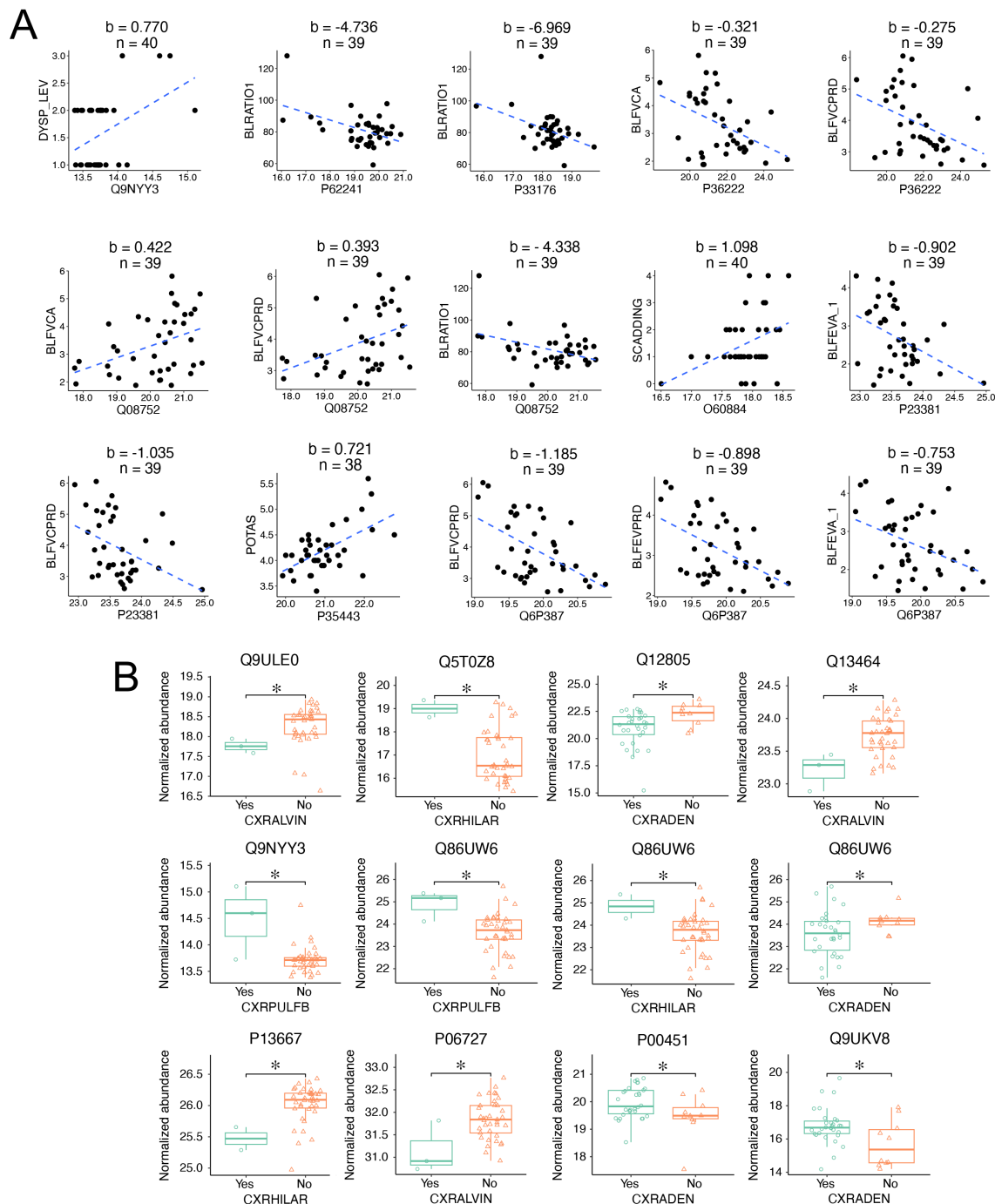
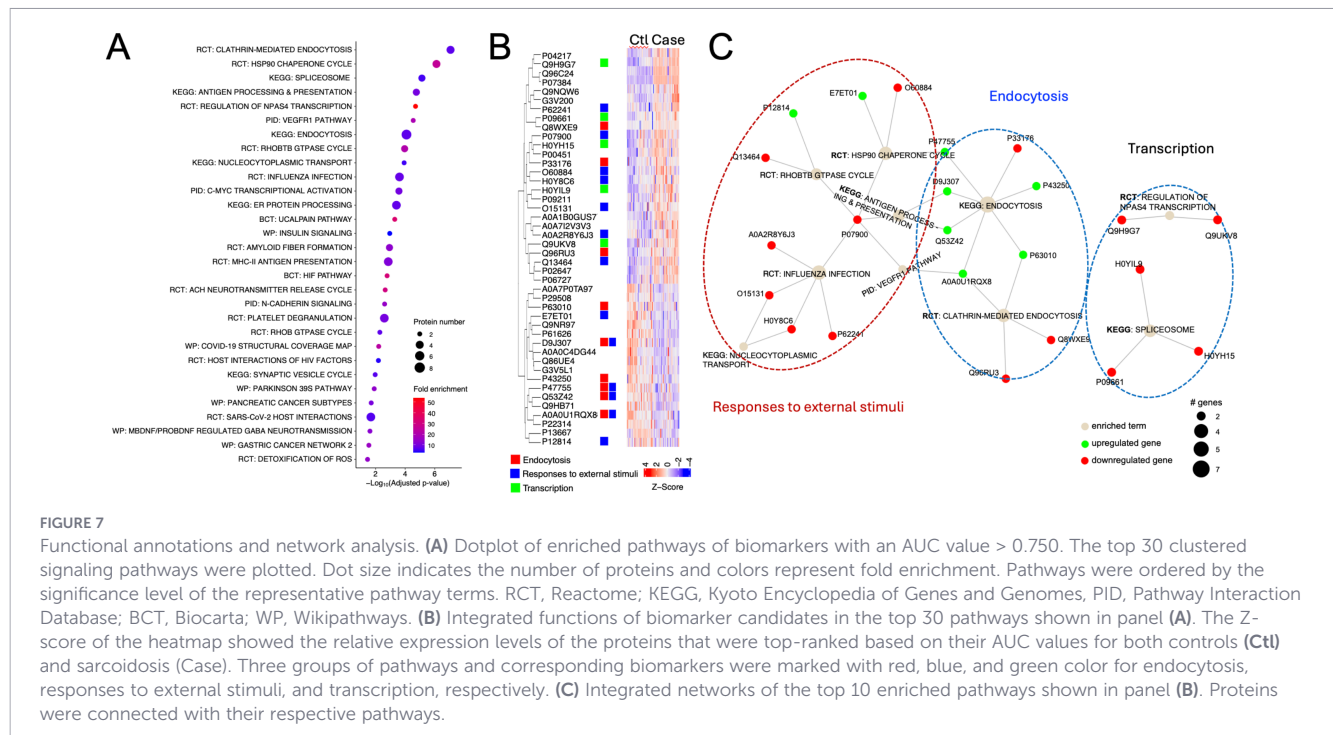


FIGURE 6

Linear and binary correlations between biomarker candidates and the top-correlated clinical variables. (A) Linear regression of the top-ranked continuous variables. The plots were sorted by their AUC values descendingly. Clinical variables were plotted as a function of normalized protein abundance. The blue dashed lines were generated by regressing the clinical variables with the corresponding normalized protein abundance. The slope (b) and sample size (n) were on the top of the plots. The correlated clinical variables included dyspnea scale (DYSPLV), FEV-1/FVC (BLRATIO1), pre-bronchodilators FVC (BLFVCA), FVC predicted (BLFVCPRD), Scadding stage (SCADDING), pre-bronchodilators FEV-1 (BLFEVA_1), and potassium content (POTAS, mEq/L). (B) Boxplots of protein abundance between the binary outcomes of the top-ranked chest X-ray variables. The plots were sorted by their AUC values descendingly. The normalized protein abundance was compared between negative (No, orange triangle) and positive patients (Yes, aqua circle) for the correlated variables. The boxplot shows the first quartile, median, third quartile, range, and normalized protein abundance. The X-ray readouts were alveolar infiltrates (CXRALVIN), Hilar retraction (CXRHILAR), adenopathy (CXRADEN), and pulmonary fibrosis (CXRPULFB). *p < 0.05, Wilcoxon rank-sum test.

Seven of the 278 biomarkers identified in this study have been previously reported. Among the top 13 biomarkers with an AUC > 0.90, plasma lysozyme (P61626, encoded by the LYZ gene) and toll-like receptor 8 (Q9NR97, encoded by the TLR8 gene) have been

identified as diagnostic biomarkers specifically for pulmonary sarcoidosis (57) and cardiac sarcoidosis (58), respectively. In addition to the aforementioned top 13 biomarkers, a previous study identified AP2B1 as a potential biomarker for sarcoidosis



through proteomic analysis of alveolar macrophages (59). In addition, membrane-associated HGFA, ILF3, RAB7A, and RPL18 in alveolar macrophages were proposed as diagnostic biomarker candidates for sarcoidosis (60, 61).

Numerous proteins that we identified have been previously associated with sarcoidosis. Apolipoprotein A1 (P02647) could be a biomarker that has been associated with chronic obstructive pulmonary disease (COPD), tuberculosis (TB), and sarcoidosis (62). Complement C5 (P01031) and C8a (P07357) are actively synthesized in alveolar macrophages in sarcoidosis (63). Whole exome sequencing identified that the protein DNAH11 may contribute to the formation of the characteristic lesion through regulating G-proteins in pediatric sarcoidosis (64). Moreover, GSTP1 has been recognized as a central node in the functional enrichment analysis of proteomics in pulmonary sarcoidosis (62). The serum concentration of C-C motif chemokine ligand 18 (CCL18) was observed to be markedly increased in individuals diagnosed with intrathoracic sarcoidosis. In contrast, the detection of increased levels of CCL16 (O1546) in the bronchoalveolar lavage (BAL) of individuals diagnosed with sarcoidosis has shown inconsistent results (65, 66).

We found evidence that some of our EV biomarkers correlate with lung fibrosis of sarcoidosis. The identified EV biomarkers CCL18 could be implicated in the prediction of sarcoidosis-associated fibrosis. CCL18 has been suggested as a marker for identifying patients at a higher risk of developing pulmonary fibrosis or progressive disease (67). However, CCL18 also serves as a biomarker for interstitial lung disease (ILD) due to its elevated levels in serum, bronchoalveolar lavage (BAL), and alveolar macrophages (68). Of note, significant myofibroblasts in neuromuscular sarcoidosis is correlated with increased expression of CCL18 in M2 macrophage phenotype (69). These studies suggest that some of the EV biomarkers could be shared with other fibrotic diseases.

We found that serine/threonine-protein kinase PLK2 (Q9NYY3) protein level in plasma EVs was highly correlated with the occurrence of lung fibrosis in sarcoidosis. Genetic knockout of the corresponding gene PLK2 leads to the manifestation of a lung fibrosis phenotype (70). The second lung fibrosis-correlated EV protein, Q9UF33 (encoded by the EPHA6 gene), is implicated in the severity of bleomycin-induced pulmonary fibrosis in mice (71). Moreover, EPHA3⁺ lung cells from IPF patients induce lung fibrosis in mice, and antibody-mediated depletion of these cells ameliorates fibrosis (72).

Several of the top thirty-ranked signaling pathways enriched by the identified EV biomarkers have been reported previously. The upregulation of two phagocytotic pathways, namely Fcγ receptor-mediated phagocytosis and clathrin-mediated endocytosis, has been documented in alveolar macrophages in sarcoidosis (59, 61). These pathways are associated with the functionality of alveolar macrophages. A robust correlation has been reported between HSP-70 and uveitis in patients diagnosed with sarcoidosis (73). Furthermore, the expression level of adenosine diphosphate-ribosylation factor GTPase activating protein 1 is notably elevated in sarcoidosis compared to asthma (74). Of note, the causality of CCL18, PLK2, and these signaling pathways in the development of sarcoidosis is unknown.

There are some limitations to this study. Although the plasma samples were collected by a multi-centered clinical trial - ACCESS, all these patients were from the United States and may not be representative of world-wide sarcoidosis. An independent validation cohort could be applied to confirm the identified EV biomarkers. This is a single omics study based on the analysis of plasma EV proteomes. Other omics datasets from the plasma and other liquid samples (e.g., urine, BAL), including plasma EV RNA profiling, EV DNA analysis, and EV metabolomics, could improve the accuracy

of the biomarkers identified in this study. Long-term storage of the ACCESS samples could be a concern. However, the considerable stability of EV proteins compensates for the alterations associated with long-term storage.

In summary, this study demonstrates that hundreds of EV biomarkers can be identified by combining unbiased high-throughput proteomics and machine-learning algorithms. Many of the biomarkers that we identified were associated with previously known clinical findings, biomarkers and immunologic pathways that have been thought to be important in sarcoidosis. This suggests that our discovered biomarkers have great clinical relevance. Our results provide novel biomarker candidates for designing prospective clinical trials to identify biomarkers for phenotype-specific, organ-specific, predictive, and treatment responses in sarcoidosis. Further studies await to specific biomarkers for the responses to treatment, organ involvements, various outcomes of the disease, and differentiating from other interstitial lung disease.

Data availability statement

Restrictions apply to the datasets: The datasets presented in this article are not readily available because of the request of Data Transfer Agreement by Loyola University Chicago. Requests to access the datasets should be directed to corresponding author.

Ethics statement

The studies involving humans were approved by Loyola University Chicago (LU#216964). The studies were conducted in accordance with the local legislation and institutional requirements. The human samples used in this study were acquired from The plasma samples and clinical dataset collected by the ACCESS trial were obtained through BioLINCC (Biologic Specimen and Data Repository Information Coordinating Center). Written informed consent for participation was not required from the participants or the participants' legal guardians/next of kin in accordance with the national legislation and institutional requirements.

Author contributions

RZ: Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Writing – original draft, Writing – review & editing. NX: Data curation, Formal analysis, Investigation, Methodology, Validation, Visualization, Writing – original draft. LG: Data curation, Formal analysis, Writing – original draft. EG: Resources, Writing – review & editing. KV: Data curation, Formal analysis, Writing – review &

editing. MQ: Data curation, Formal analysis, Writing – original draft. DZ: Resources, Validation, Writing – review & editing. JZ: Writing – review & editing. CM: Writing – review & editing. MJ: Writing – review & editing. LK: Funding acquisition, Writing – review & editing. H-LJ: Conceptualization, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Writing – original draft, Writing – review & editing.

Funding

The author(s) declared that financial support was received for this work and/or its publication. NIH HL134828, HL157533. This work was supported by NHLBI: grant NIH grant HL87017 (to H-LJ), HL157533 (to LK), and institutional funds (to H-LJ).

Conflict of interest

MJ has received grants for his institution from Xentria Pharmaceuticals and aTyr Pharmaceuticals and is a consultant for Merck, Priovant Pharmaceuticals, and Sparrow Pharmaceuticals.

The remaining author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author H-LJ declared that they were an editorial board member of *Frontiers*, at the time of submission. This had no impact on the peer review process and the final decision.

Generative AI statement

The author(s) declared that generative AI was not used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by *Frontiers* with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Spagnolo P, Rossi G, Trisolini R, Sverzellati N, Baughman RP, Wells AU. Pulmonary sarcoidosis. *Lancet Respir Med.* (2018) 6:389–402. doi: 10.1055/a-2768-2520
- Grunewald J, Grutters JC, Arkema EV, Saketkoo LA, Moller DR, Müller-Quernheim J. Sarcoidosis. *Nat Rev Dis Primers.* (2019) 5:45. doi: 10.1097/mcp.0b013e3283043de7
- Nardi A, Brillet PY, Letoumelin P, Girard F, Brauner M, Uzunhan Y, et al. Stage IV sarcoidosis: comparison of survival with the general population and causes of death. *Eur Respir J.* (2011) 38:1368–73. doi: 10.1183/09031936.00187410
- Patterson KC, Strek ME. Pulmonary fibrosis in sarcoidosis. Clinical features and outcomes. *Ann Am Thorac Soc.* (2013) 10:362–70. doi: 10.1513/annalsats.201303-069fr
- Gupta R, Kim JS, Baughman RP. An expert overview of pulmonary fibrosis in sarcoidosis. *Expert Rev Respir Med.* (2023) 17:119–30. doi: 10.1080/17476348.2023.2183193
- Rothkrantz-Kos S, Van Dieijen-Visser MP, Mulder PG, Drent M. Potential usefulness of inflammatory markers to monitor respiratory functional impairment in sarcoidosis. *Clin Chem.* (2003) 49:1510–7. doi: 10.1373/49.9.1510
- Bons JA, Drent M, Bouwman FG, Mariman EC, Van Dieijen-Visser MP, Wodzig WK. Potential biomarkers for diagnosis of sarcoidosis using proteomics in serum. *Respir Med.* (2007) 101:1687–95. doi: 10.1016/j.rmed.2007.03.002
- Thi Hong Nguyen C, Kambe N, Kishimoto I, Ueda-Hayakawa I, Okamoto H. Serum soluble interleukin-2 receptor level is more sensitive than angiotensin-converting enzyme or lysozyme for diagnosis of sarcoidosis and may be a marker of multiple organ involvement. *J Dermatol.* (2017) 44:789–97. doi: 10.1111/1346-8138.13792
- Uysal P, Durmus S, Sozer V, Gelisgen R, Seyhan EC, Erdenen F, et al. YKL-40, soluble IL-2 receptor, angiotensin converting enzyme and C-reactive protein: Comparison of markers of sarcoidosis activity. *Biomolecules.* (2018) 8:84. doi: 10.3390/biom8030084
- Miyata J, Ogawa T, Tagami Y, Sato T, Nagayama M, Hirano T, et al. Serum soluble interleukin-2 receptor level is a predictive marker for EBUS-TBNA-based diagnosis of sarcoidosis. *Sarcoidosis Vasc Diffuse Lung Dis.* (2020) 37:8–16.
- Otto C, Wengert O, Unterwalder N, Meisel C, Ruprecht K. Analysis of soluble interleukin-2 receptor as CSF biomarker for neurosarcoidosis. *Neurol Neuroimmunol Neuroinflamm.* (2020) 7:e725. doi: 10.1212/nxi.0000000000000725
- Schimmelpennink MC, Quanjer M, Vorselaars A, Wiertz I, Veltkamp M, Van Moorsel C, et al. Value of serum soluble interleukin-2 receptor as a diagnostic and predictive biomarker in sarcoidosis. *Expert Rev Respir Med.* (2020) 14:749–56. doi: 10.1080/17476348.2020.1751614
- Futami Y, Takeda Y, Koba T, Narumi R, Nojima Y, Ito M, et al. Identification of CD14 and lipopolysaccharide-binding protein as novel biomarkers for sarcoidosis using proteomics of serum extracellular vesicles. *Int Immunol.* (2022) 34:327–40. doi: 10.1093/intimm/dxac009
- Kobayashi S, Myoren T, Oda S, Inari M, Ishiguchi H, Murakami W, et al. Urinary 8-hydroxy-2'-deoxyguanosine as a novel biomarker of inflammatory activity in patients with cardiac sarcoidosis. *Int J Cardiol.* (2015) 190:319–28. doi: 10.1016/j.ijcard.2015.04.144
- Myoren T, Kobayashi S, Oda S, Nanno T, Ishiguchi H, Murakami W, et al. An oxidative stress biomarker, urinary 8-hydroxy-2'-deoxyguanosine, predicts cardiovascular-related death after steroid therapy for patients with active cardiac sarcoidosis. *Int J Cardiol.* (2016) 212:206–13. doi: 10.1016/j.ijcard.2016.03.003
- Ishiguchi H, Kobayashi S, Myoren T, Kohno M, Nanno T, Murakami W, et al. Urinary 8-hydroxy-2'-deoxyguanosine as a myocardial oxidative stress marker is associated with ventricular tachycardia in patients with active cardiac sarcoidosis. *Circ Cardiovasc Imaging.* (2017) 10:e006764. doi: 10.1161/circimaging.117.006764
- Bargagli E, Bennett D, Maggiorcelli C, Di Sipio P, Margollicci M, Bianchi N, et al. Human chitotriosidase: a sensitive biomarker of sarcoidosis. *J Clin Immunol.* (2013) 33:264–70. doi: 10.1007/s10875-012-9754-4
- Harlander M, Salobir B, Zupančič M, Terčelj M. Bronchoalveolar lavage chitotriosidase activity as a biomarker of sarcoidosis. *Sarcoidosis Vasc Diffuse Lung Dis.* (2016) 32:313–7.
- Bennett D, Cameli P, Lanzarone N, Carobene L, Bianchi N, Fui A, et al. Chitotriosidase: a biomarker of activity and severity in patients with sarcoidosis. *Respir Res.* (2020) 21:6. doi: 10.1186/s12931-019-1263-z
- Bergantini L, Bianchi F, Cameli P, Mazzei MA, Fui A, Sestini P, et al. Prognostic biomarkers of sarcoidosis: a comparative study of serum chitotriosidase, ACE, lysozyme, and KL-6. *Dis Markers.* (2019) 2019:8565423. doi: 10.1155/2019/8565423
- Bergantini L, D'alessandro M, Vietri L, Rana GD, Cameli P, Acerra S, et al. Utility of serological biomarker panels for diagnostic accuracy of interstitial lung diseases. *Immunol Res.* (2020) 68:414–21. doi: 10.1007/s12026-020-09158-0
- Handa T, Nagai S, Ueda S, Chin K, Ito Y, Watanabe K, et al. Significance of plasma NT-proBNP levels as a biomarker in the assessment of cardiac involvement and pulmonary hypertension in patients with sarcoidosis. *Sarcoidosis Vasc Diffuse Lung Dis.* (2010) 27:27–35.
- Kiko T, Yoshihisa A, Kanno Y, Yokokawa T, Abe S, Miyata-Tatsumi M, et al. A multiple biomarker approach in patients with cardiac sarcoidosis. *Int Heart J.* (2018) 59:996–1001. doi: 10.1536/ihj.17-695
- Chen Y, Zhu Q, Cheng L, Wang Y, Li M, Yang Q, et al. Exosome detection via the ultrafast-isolation system: EXODUS. *Nat Methods.* (2021) 18:212–8. doi: 10.1038/s41592-020-01034-x
- Gerszten RE, Accurso F, Bernard GR, Caprioli RM, Klee EW, Klee GG, et al. Challenges in translating plasma proteomics from bench to bedside: update from the NHLBI Clinical Proteomics Programs. *Am J Physiol Lung Cell Mol Physiol.* (2008) 295:L16–22. doi: 10.1152/ajplung.00044.2008
- Lam SM, Zhang C, Wang Z, Ni Z, Zhang S, Yang S, et al. A multi-omics investigation of the composition and function of extracellular vesicles along the temporal trajectory of COVID-19. *Nat Metab.* (2021) 3:909–22. doi: 10.1038/s42255-021-00425-4
- Gongye X, Tian M, Xia P, Qu C, Chen Z, Wang J, et al. Multi-omics analysis revealed the role of extracellular vesicles in hepatobiliary & pancreatic tumor. *J Control Release.* (2022) 350:11–25. doi: 10.1016/j.jconrel.2022.08.010
- Setua S, Thangaraju K, Dzieciatkowska M, Wilkerson RB, Nemkov T, Lamb DR, et al. Coagulation potential and the integrated omics of extracellular vesicles from COVID-19 positive patient plasma. *Sci Rep.* (2022) 12:22191. doi: 10.1038/s41598-022-26473-8
- Shaba E, Vantaggiato L, Governini L, Haxhiu A, Sebastiani G, Fignani D, et al. Multi-omics integrative approach of extracellular vesicles: A future challenging milestone. *Proteomes.* (2022) 10:12. doi: 10.3390/proteomes10020012
- Altadill T, Campoy I, Lanau L, Gill K, Rigau M, Gil-Moreno A, et al. Enabling metabolomics based biomarker discovery studies using molecular phenotyping of exosome-like vesicles. *PloS One.* (2016) 11:e0151339. doi: 10.1371/journal.pone.0151339
- Sequeiros T, Rigau M, Chiva C, Montes M, Garcia-Grau I, Garcia M, et al. Targeted proteomics in urinary extracellular vesicles identifies biomarkers for diagnosis and prognosis of prostate cancer. *Oncotarget.* (2017) 8:4960–76. doi: 10.18632/oncotarget.13634
- Herrera-Van Oostdam AS, Salgado-Bustamante M, López JA, Herrera-Van Oostdam DA, López-Hernández Y. Placental exosomes viewed from an 'omics' perspective: implications for gestational diabetes biomarkers identification. *biomark Med.* (2019) 13:675–84. doi: 10.2217/bmm-2018-0468
- Song JW, Lam SM, Fan X, Cao WJ, Wang SY, Tian H, et al. Omics-driven systems interrogation of metabolic dysregulation in COVID-19 pathogenesis. *Cell Metab.* (2020) 32:188–202.e185. doi: 10.1016/j.cmet.2020.06.016
- Kugeratski FG, Hodge K, Lilla S, Mcandrews KM, Zhou X, Hwang RF, et al. Quantitative proteomics identifies the core proteome of exosomes with syntenin-1 as the highest abundant protein and a putative universal biomarker. *Nat Cell Biol.* (2021) 23:631–41. doi: 10.1038/s41556-021-00693-y
- Nielsen JE, Honoré B, Vestergård K, Maltesen RG, Christiansen G, Bøge AU, et al. Shotgun-based proteomics of extracellular vesicles in Alzheimer's disease reveals biomarkers involved in immunological and coagulation pathways. *Sci Rep.* (2021) 11:18518. doi: 10.1038/s41598-021-97969-y
- Saliccia S, Capriotti AL, Laganà A, Fais S, Logozzi M, De Berardinis E, et al. Biomarkers in prostate cancer diagnosis: From current knowledge to the role of metabolomics and exosomes. *Int J Mol Sci.* (2021) 22:4367. doi: 10.3390/ijms22094367
- Ge X, Guo M, Li M, Zhang S, Qiang J, Zhu L, et al. Potential blood biomarkers for chronic traumatic encephalopathy: the multi-omics landscape of an observational cohort. *Front Aging Neurosci.* (2022) 14:1052765. doi: 10.3389/fnagi.2022.1052765
- Taşlı NP, Gönen ZB, Kırbaş OK, Gökdemir NS, Bozkurt BT, Bayraktar B, et al. Preclinical studies on convalescent human immune plasma-derived exosome: Omics and antiviral properties to SARS-CoV-2. *Front Immunol.* (2022) 13:824378.
- Ji HL, Xi NMS, Mohan C, Yan X, Jain KG, Zang QS, et al. Biomarkers and molecular endotypes of sarcoidosis: lessons from omics and non-omics studies. *Front Immunol.* (2023) 14:1342429. doi: 10.3389/fimmu.2023.1342429
- Kraaijvanger R, Janssen Bonás M, Paspali I, Grutters JC, Veltkamp M, De Kleijn DPV, et al. Targeted proteomics in extracellular vesicles identifies biomarkers predictive for therapeutic response in sarcoidosis. *ERJ Open Res.* (2025) 11(2):00672–2024. doi: 10.1183/23120541.00672-2024
- Kraaijvanger R, Janssen Bonás M, Grutters JC, Paspali I, Veltkamp M, De Kleijn DPV, et al. Decreased serpin C1 in extracellular vesicles predicts response to methotrexate treatment in patients with pulmonary sarcoidosis. *Respir Res.* (2024) 25:166. doi: 10.1186/s12931-024-02809-y
- Judson MA, Baughman RP, Teirstein AS, Terrin ML, Yeager H. Defining organ involvement in sarcoidosis: the ACCESS proposed instrument. ACCESS Research Group. A Case Control Etiologic Study of Sarcoidosis. *Sarcoidosis Vasc Diffuse Lung Dis.* (1999) 16:75–86. doi: 10.1164/rccm.200402-249oc

43. Rybicki BA, Iannuzzi MC, Frederick MM, Thompson BW, Rossman MD, Bresnitz EA, et al. Familial aggregation of sarcoidosis. A case-control etiologic study of sarcoidosis (ACCESS). *Am J Respir Crit Care Med.* (2001) 164:2085–91. doi: 10.1164/ajrccm.164.11.2106001
44. Newman LS, Rose CS, Bresnitz EA, Rossman MD, Barnard J, Frederick M, et al. A case control etiologic study of sarcoidosis: environmental and occupational risk factors. *Am J Respir Crit Care Med.* (2004) 170:1324–30. doi: 10.1093/bmb/ldg021
45. Barnard J, Rose C, Newman L, Canner M, Martyny J, Mccammon C, et al. Job and industry classifications associated with sarcoidosis in a case-control etiologic study of sarcoidosis (ACCESS). *J Occup Environ Med.* (2005) 47:226–34. doi: 10.1097/01.jom.0000155711.88781.91
46. Semenzato G. ACCESS: A case control etiologic study of sarcoidosis. *Sarcoidosis Vasc Diffuse Lung Dis.* (2005) 22:83–6. doi: 10.1128/9781555817879.ch10
47. Wu X, Li L, Iliuk A, Tao WA. Highly efficient phosphoproteome capture and analysis from urinary extracellular vesicles. *J Proteome Res.* (2018) 17:3308–16. doi: 10.1021/acs.jproteome.8b00459
48. Zhao R, Hadisurya M, Ndetan H, Xi NM, Adduri S, Konduru NV, et al. Regenerative signatures in BAL of acute respiratory distress syndrome. *Am J Respir Cell Mol Biol.* (2024) 71:740–2. doi: 10.1165/rcmb.2024-0193le
49. Huber W, Von Heydebreck A, Sültmann H, Poustka A, Vingron M. Variance stabilization applied to microarray data calibration and to the quantification of differential expression. *Bioinformatics.* (2002) 18 Suppl 1:S96–S104. doi: 10.1093/bioinformatics/18.suppl_1.s96
50. Stekhoven DJ, Bühlmann P. MissForest–non-parametric missing value imputation for mixed-type data. *Bioinformatics.* (2012) 28:112–8. doi: 10.1093/bioinformatics/btr597
51. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics.* (2016) 32:2847–9. doi: 10.1093/bioinformatics/btw313
52. H. W. *ggplot2: elegant graphics for data analysis*. New York: Springer-Verlag. (2016).
53. Duijvelaar E, Gisby J, Peters JE, Bogaard HJ, Aman J. Longitudinal plasma proteomics reveals biomarkers of alveolar-capillary barrier disruption in critically ill COVID-19 patients. *Nat Commun.* (2024) 15:744. doi: 10.1038/s41467-024-44986-w
54. Ulgen E, Ozisik O, Sezerman OU. pathfindR: An R package for comprehensive identification of enriched pathways in omics data through active subnetworks. *Front Genet.* (2019) 10:858. doi: 10.3389/fgene.2019.00858
55. Chen J, Bardes EE, Aronow BJ, Jegga AG. ToppGene suite for gene list enrichment analysis and candidate gene prioritization. *Nucleic Acids Res.* (2009) 37:W305–11. doi: 10.1093/nar/gkp427
56. Faul F, Erdfelder E, Lang AG, Buchner A. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav Res Methods.* (2007) 39:175–91. doi: 10.3758/bf03193146
57. Ashitani J, Matsumoto N, Nakazato M. Elevated alpha-defensin levels in plasma of patients with pulmonary sarcoidosis. *Respirology.* (2007) 12:339–45. doi: 10.1111/j.1440-1843.2007.01061.x
58. Lassner D, Kühl U, Siegmund CS, Rohde M, Elezkurtaj S, Escher F, et al. Improved diagnosis of idiopathic giant cell myocarditis and cardiac sarcoidosis by myocardial gene expression profiling. *Eur Heart J.* (2014) 35:2186–95. doi: 10.1093/eurheartj/ehu101
59. Silva E, Souchelnyskyi S, Kasuga K, Eklund A, Grunewald J, Wheelock ÅM. Quantitative intact proteomics investigations of alveolar macrophages in sarcoidosis. *Eur Respir J.* (2013) 41:1331–9. doi: 10.1183/09031936.00178111
60. Beirne P, Pantelidis P, Charles P, Wells AU, Abraham DJ, Denton CP, et al. Multiplex immune serum biomarker profiling in sarcoidosis and systemic sclerosis. *Eur Respir J.* (2009) 34:1376–82. doi: 10.1183/09031936.00028209
61. Kjellin H, Silva E, Branca RM, Eklund A, Jakobsson PJ, Grunewald J, et al. Alterations in the membrane-associated proteome fraction of alveolar macrophages in sarcoidosis. *Sarcoidosis Vasc Diffuse Lung Dis.* (2016) 33:17–28.
62. Landi C, Bargagli E, Carleo A, Bianchi L, Gagliardi A, Cillis G, et al. A functional proteomics approach to the comprehension of sarcoidosis. *J Proteomics.* (2015) 128:375–87. doi: 10.1016/j.jpro.2015.08.012
63. Pettersen HB, Johnson E, Mollnes TE, Garred P, Hetland G, Osen SS. Synthesis of complement by alveolar macrophages from patients with sarcoidosis. *Scand J Immunol.* (1990) 31:15–23. doi: 10.1111/j.1365-3083.1990.tb02738.x
64. Calender A, Rollat Farnier PA, Buisson A, Pinson S, Bentaher A, Lebecque S, et al. Whole exome sequencing in three families segregating a pediatric case of sarcoidosis. *BMC Med Genomics.* (2018) 11:23. doi: 10.1186/s12920-018-0338-x
65. Arakelyan A, Kriegova E, Kubistova Z, Mrazek F, Kverka M, Du Bois RM, et al. Protein levels of CC chemokine ligand (CCL)15, CCL16 and macrophage stimulating protein in patients with sarcoidosis. *Clin Exp Immunol.* (2009) 155:457–65. doi: 10.1111/j.1365-2249.2008.03832.x
66. Nureki S, Miyazaki E, Usagawa Y, Ueno T, Ando M, Takenaka R, et al. Elevated concentrations of liver-expressed chemokine/CC chemokine ligand 16 in bronchoalveolar lavage fluid from patients with eosinophilic pneumonia. *Int Arch Allergy Immunol.* (2009) 150:282–90. doi: 10.1159/000222681
67. Ramos-Casals M, Retamozo S, Sisó-Almirall A, Pérez-Alvarez R, Pallarès L, Brito-Zerón P. Clinically-useful serum biomarkers for diagnosis and prognosis of sarcoidosis. *Expert Rev Clin Immunol.* (2019) 15:391–405. doi: 10.1080/1744666x.2019.1568240
68. Cai M, Bonella F, He X, Sixt SU, Sarria R, Guzman J, et al. CCL18 in serum, BAL fluid and alveolar macrophage culture supernatant in interstitial lung diseases. *Respir Med.* (2013) 107:1444–52. doi: 10.1016/j.rmed.2013.06.004
69. Prokop S, Heppner FL, Goebel HH, Stenzel W. M2 polarized macrophages and giant cells contribute to myofibrosis in neuromuscular sarcoidosis. *Am J Pathol.* (2011) 178:1279–86. doi: 10.1016/j.ajpath.2010.11.065
70. Kant TA, Neue M, Winter L, Hoffmann M, Kämmerer S, Klapproth E, et al. Genetic deletion of polo-like kinase 2 induces a pro-fibrotic pulmonary phenotype. *Cells.* (2021) 10:617. doi: 10.3390/cells10030617
71. Leem AY, Shin MH, Douglas IS, Song JH, Chung KS, Kim EY, et al. All-trans retinoic acid attenuates bleomycin-induced pulmonary fibrosis via downregulating EphA2-EphrinA1 signaling. *Biochem Biophys Res Commun.* (2017) 491:721–6. doi: 10.1016/j.bbrc.2017.07.122
72. Hohmann MS, Habel DM, Espindola MS, Huang G, Jones I, Narayanan R, et al. Antibody-mediated depletion of CCR10+EphA3+ cells ameliorates fibrosis in IPF. *JCI Insight.* (2021) 6:e141061. doi: 10.1172/jci.insight.141061
73. Spagnolo P, Sato H, Marshall SE, Antoniou KM, Ahmad T, Wells AU, et al. Association between heat shock protein 70/Hom genetic polymorphisms and uveitis in patients with sarcoidosis. *Invest Ophthalmol Vis Sci.* (2007) 48:3019–25. doi: 10.1167/iops.06-1485
74. Häggmark A, Hamsten C, Wiklundh E, Lindskog C, Mattsson C, Andersson E, et al. Proteomic profiling reveals autoimmune targets in sarcoidosis. *Am J Respir Crit Care Med.* (2015) 191:574–83.