

Plasma extracellular vesicle proteins as promising noninvasive biomarkers for diagnosis of idiopathic pulmonary fibrosis

Raju S. R. Adduri¹ | Kai Cai² | Karen Velasco-Alzate¹ | Ravikiran Vasireddy¹ | Jeffrey W. Miller³ | Sergio Poli de Frías^{4,5} | Fernando Poli de Frías^{4,6} | Yasushi Horimasu⁷ | Hiroshi Iwamoto⁷ | Noboru Hattori⁷ | Yingze Zhang⁸ | Kevin F. Gibson⁸ | Anoop K. Pal⁹ | Zhe Chen¹⁰ | Daniela Nicastro² | Li Li¹¹ | Sujith Cherian¹² | Lynette M. Sholl¹³ | Sreerama Shetty¹ | Harrison Ndetan¹⁴ | Anthony H. Maeda⁴ | Maria A. Planchart Ferretto⁴ | Gary M. Hunninghake⁴ | David A. Schwartz¹⁵ | Daniel J. Kass⁸ | Ivan O. Rosas^{4,6} | Nagarjun V. Konduru¹

¹Department of Cellular and Molecular Biology, University of Texas Health Science Center at Tyler, Tyler, Texas, USA

²Departments of Cell Biology and Biophysics, University of Texas Southwestern Medical Center, Dallas, Texas, USA

³Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA

⁴Pulmonary Critical Care Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, USA

⁵Internal Medicine, Mount Sinai Medical Center, Miami Beach, Florida, USA

⁶Pulmonary, Critical Care, and Sleep Medicine, Baylor College of Medicine, Houston, Texas, USA

⁷Department of Molecular and Internal Medicine, Graduate School of Biomedical and Health Sciences, Hiroshima University, Hiroshima, Japan

⁸Division of Pulmonary, Allergy and Critical Care Medicine, University of Pittsburgh, Pittsburgh, Pennsylvania, USA

⁹Izon Science US LLC, Medford, Massachusetts, USA

¹⁰Department of Biophysics, University of Texas Southwestern Medical Center, Dallas, Texas, USA

¹¹Division of Environmental and Occupational Health Sciences, Division of Pulmonary Sciences and Critical Care Medicine, Department of Medicine, School of Medicine, University of Colorado Denver, Denver, Colorado, USA

¹²Department of Internal Medicine, McGovern Medical School, University of Texas Health Science Center at Houston, Houston, Texas, USA

¹³Department of Pathology, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts, USA

¹⁴Department of Biostatistics, School of Health Professions, University of Texas Health Science Center at Tyler, Tyler, Texas, USA

¹⁵Department of Medicine, University of Colorado School of Medicine, Aurora, Colorado, USA

Correspondence

Nagarjun V. Konduru, Department of Cellular and Molecular Biology, University of Texas Health Science Center at Tyler, Tyler, Texas 75708, USA.
Email: Nagarjun.KonduruVenkata@uttyler.edu

Funding information

UTHSCT; UT System Rising STARS award; National institute of health, Grant/Award Numbers: R01 GM083122, R01 HL111024, R01 HL135142, R01 HL130974; Cancer Prevention and

Abstract

High-resolution computed tomography (HRCT) imaging is critical for diagnostic evaluation of Idiopathic Pulmonary Fibrosis (IPF). However, several other interstitial lung diseases (ILDs) often exhibit radiologic pattern similar to IPF on HRCT making the diagnosis of the disease difficult. Therefore, biomarkers that distinguish IPF from other ILDs can be a valuable aid in diagnosis. Using mass spectrometry, we performed proteomic analysis of plasma extracellular vesicles (EVs) in patients diagnosed with IPF, chronic hypersensitivity pneumonitis, nonspecific interstitial pneumonitis, and healthy subjects. A five-protein signature was identified by lasso regression and was

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs License](https://creativecommons.org/licenses/by-nc-nd/4.0/), which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2023 The Authors. *Journal of Extracellular Biology* published by Wiley Periodicals, LLC on behalf of the International Society for Extracellular Vesicles.

Research Institute of Texas, Grant/Award Number: RR140082; gifts from the families of David Herlihy and George Spiliotis

validated in an independent cohort using ELISA. The five-protein signature derived from mass spectrometry data showed an area under the receiver operating characteristic curve of 0.915 (95%CI: 0.819–1.011) and 0.958 (95%CI: 0.882–1.034) for differentiating IPF from other ILDs and from healthy subjects, respectively. Stepwise backwards elimination yielded a model with 3 and 2 proteins for discriminating IPF from other ILDs and healthy subjects, respectively, without compromising diagnostic accuracy. In summary, we discovered and validated EV protein biomarkers for differential diagnosis of IPF in independent cohorts. Interestingly, the biomarker panel could also distinguish IPF and healthy subjects with high accuracy. The biomarkers need to be evaluated in large prospective cohorts to establish their clinical utility.

KEYWORDS

and nonspecific interstitial pneumonia (NSIP), biomarker, chronic hypersensitivity pneumonitis (CHP), differential diagnosis, ELISA, extracellular vesicles, idiopathic pulmonary fibrosis (IPF), interstitial lung diseases, lasso regression, mass spectrometry, noninvasive diagnosis, proteomics, usual interstitial pneumonia (UIP)

1 | INTRODUCTION

Idiopathic pulmonary fibrosis (IPF) is a chronic and progressively fibrosing interstitial pneumonia of unknown etiology that often leads to respiratory failure (Wuyts et al., 2014). With a median survival of 3.8 years, IPF appears to be more lethal than many cancer types (Wuyts et al., 2014). Diagnosis at early stage and therapeutic intervention before the lung function is severely impaired could potentially improve treatment response, and thereby prolong survival (Flaherty et al., 2004).

High-resolution computed tomography (HRCT) is a crucial diagnostic component in clinical diagnosis of IPF. A clear morphologic pattern of usual interstitial pneumonia (UIP) on HRCT indicates IPF. However, UIP is not synonymous with IPF as other interstitial lung diseases (ILDs) including chronic hypersensitivity pneumonitis (CHP), and nonspecific interstitial pneumonia (NSIP), etc., may also exhibit a similar pattern (Wuyts et al., 2014). Therefore, accurate diagnosis of IPF involves exclusion of other ILDs such as CHP and NSIP which necessitate inquiry into patient's history and sometimes acquisition of surgical lung biopsy or broncho-alveolar lavage fluid.

There is an unmet need to identify biomarkers for discerning IPF from other ILDs and to aid in IPF diagnosis. Several studies have reported biomarkers for differential diagnosis of IPF (Greene et al., 2002; Ishii et al., 2003; Morais et al., 2015; Onishi et al., 2020; White et al., 2016). Majority of these studies used the classical approach of biomarker identification wherein genes/proteins with a putative role in the pathology of the IPF were evaluated for their efficacy as diagnostic markers, but their diagnostic efficacy has been unsatisfactory. Notably, different ILDs with fibrosis may share similar molecular landscapes between themselves. Therefore, high-throughput analyses may help in discovering molecular changes unique to each ILD.

High-throughput analyses have recently attracted attention for identification of disease biomarkers owing to their ability to quantify thousands of molecules in a single screening. The advantage of this approach is that (1) the biomarkers could be selected from a large pool of genes/proteins and not limited to those with an evidence of disease involvement, and (2) will enable the use of robust machine learning algorithms to select robust biomarkers from large pool of genes/proteins.

Extracellular vesicles (EVs) involved in intercellular signaling are a unique biological matrix comprising different biomolecules (DNA, RNA, proteins, and metabolites) and their compositions reflect the molecular and physiological status of the parental cell (Kahlert et al., 2014; Kolonics et al., 2020; Phan et al., 2022; Thakur et al., 2014). Due to their high abundance in accessible body fluids (2015), in the past decade, EVs have become a reliable source for biomarker discovery (Boukouris & Mathivanan, 2015; Huda et al., 2021). In this study, we explored small EVs (exosomes) as a source of potential biomarkers for IPF diagnosis. We focused on small EVs due to their higher stability for longer periods in circulation than large EVs (Record et al., 2018).

However, most of the studies examining lung diseases have focused on the miRNA cargo within EVs for identifying biomarkers (Njock et al., 2019). Recently there has been heightened interest in exploiting EV proteins as diagnostic biomarkers (Hinestrosa et al., 2022; Li et al., 2021; Melo et al., 2015; Tian et al., 2021). There have been few studies exploring serum/plasma proteins as biomarkers for differential diagnosis of IPF although with low accuracy (Greene et al., 2002; Morais et al., 2015; White et al., 2016). The discovery of biomarkers on serum/plasma is confronted with inherent challenges. Serum/plasma contain high abundant proteins while the valuable biomarkers are generally present several orders of magnitude lower in concentrations than the high abundant proteins (Merrell et al., 2004). Recently, methods to deplete high abundant proteins have been developed but this approach potentially leads to less desirable experimental variations often presenting no added value (Tu et al., 2010). Furthermore, most of the earlier studies involving plasma/serum proteins analyzed the efficacy of a limited number of proteins with a known function in the pathophysiology of IPF (Greene et al., 2002; Morais et al., 2015; White et al., 2016). Consequently,

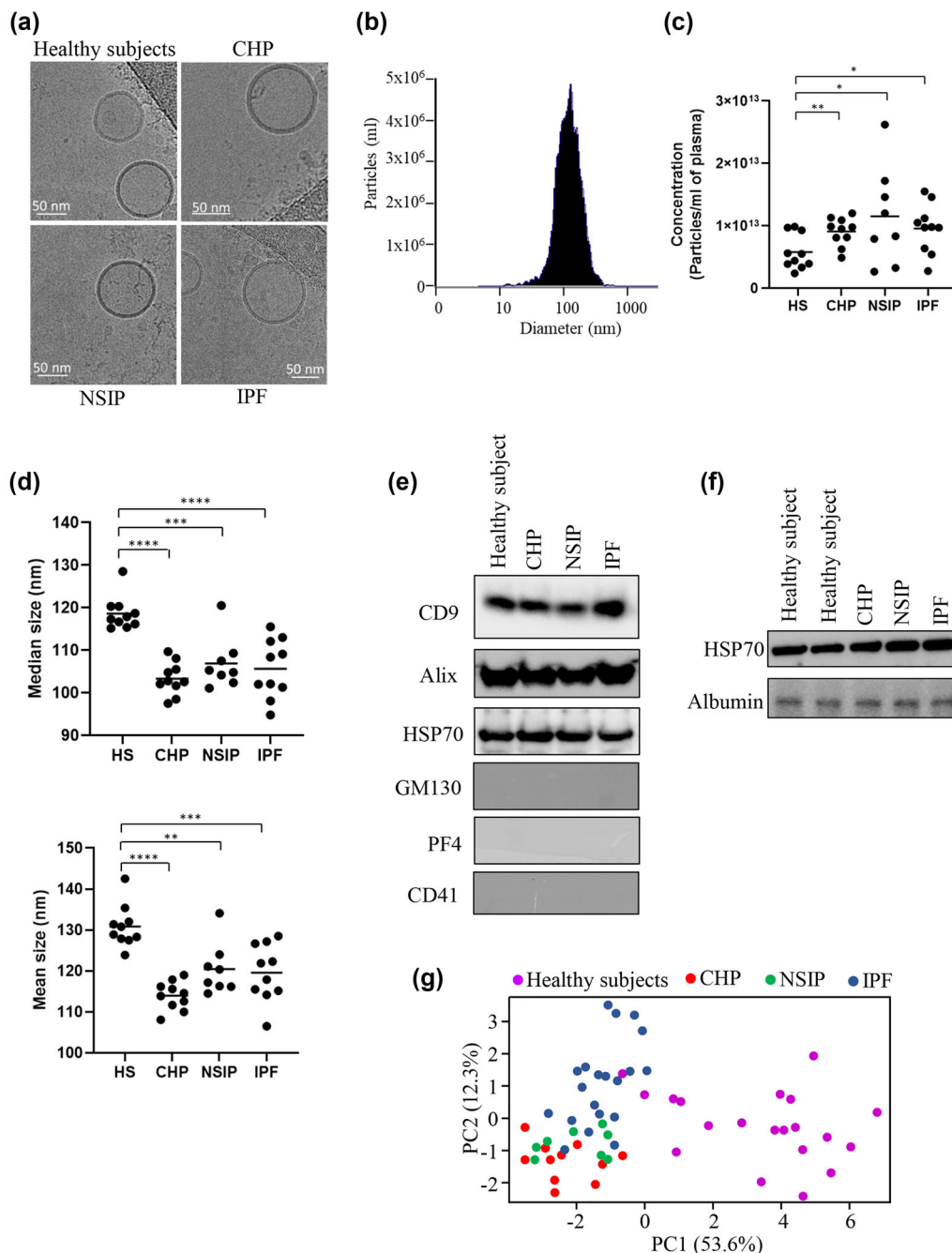


FIGURE 1 EV characterization, protein measurements, principal components, and pathway analyses of EV proteomes. (a) Representative cryo-electron micrographs of EVs isolated from peripheral blood plasma of participants in the study. (b) A representative nanoparticle tracking analysis (NTA) showing size and quantification of plasma EVs from a healthy subject (c) Plasma EV concentrations in healthy subjects, CHP and IPF ($n = 10$ each); and NSIP ($n = 8$) measured using NTA. (d) Mean and median size of EVs isolated from plasma samples of healthy subjects and different diseases. (e) Detection of common EV markers using western blot for representative samples from healthy subjects, CHP, NSIP, and IPF. Presence of EV membrane markers (CD9), internal markers (HSP70 and Alix), and absence of Golgi complex marker (GM130) and platelet markers CD41 and PF4 are shown. Representative western blot (Panel f) showing similar levels of albumin in EV isolates across samples. (g) Principal components analysis of proteomic profiles of 20 IPF, 19 other-ILDs [consisting of 11 chronic hypersensitivity pneumonitis (CHP) and 8 Nonspecific interstitial pneumonia (NSIP)] and 15 healthy subjects. Two principal components explaining highest variation were plotted on X and Y axis, respectively. HS, healthy subjects.

proteins with a confirmed role in the pathology of IPF were tested as biomarkers. However, the performance of all these biomarkers in differential diagnosis has remained suboptimal (Raghu et al., 2018). On the other hand, biomarker discovery on EVs circumvents the problems posed by high abundant blood proteins. We hypothesized that high throughput proteomics analysis of EVs would enable the systematic discovery of a biomarker signature with higher efficacy than previous biomarkers. Here, we aimed to identify a protein signature in EVs that distinguishes IPF from other ILDs and healthy subjects.

2 | MATERIALS AND METHODS

2.1 | Patient cohort

The study was conducted on a total of 163 samples including 54 IPF, 38 CHP, 27 NSIP and 44 healthy subjects from four different ILD centers/Biobanks. Biomarker discovery was performed on cohort-I comprising 20 IPF, 11 CHP and 8 NSIP and 20 healthy subjects from University of Pittsburgh, Pittsburgh, USA and Brigham and Women's hospital, Boston, USA. Diagnosis of IPF or other ILDs was made according to consensus criteria (Wuyts et al., 2014). 47.8% of other ILD patients and 54.2% IPF patients required surgical biopsy for confirmation of disease diagnosis. The biomarker signature was validated on cohort-II comprising 34 IPF, 27 CHP, 19 NSIP and 24 healthy subjects from Hiroshima University, Hiroshima, Japan; National heart, lung, and blood institute (NHLBI) (<http://www.ltrcpublic.com/>), USA; and Brigham and Women's hospital, Boston, USA. The plasma samples from these patients have been extensively used for prior studies (Herazo-Maya et al., 2017; Noth et al., 2013; Richards et al., 2012; Rosas et al., 2008; Zhang et al., 2011, 2019). The study was approved by institutional review boards of University of Pittsburgh (IRB# STUDY19040326 and STUDY20030223), Brigham and Women's hospital (IRB#2012P000840), Hiroshima University (IRB#M326), and University of Texas Health Science Center at Tyler (IRB #20-019 and #0000370). The institutional review boards of participating centers approved the LTRC study and patients' consent permitted further use of de-identified data. Plasma samples were collected at the time of diagnosis or during one of the follow-up clinical visits. There is no significant difference between other ILDs and IPF with respect to age, gender, and lung function parameters. Plasma samples from healthy volunteers with no history of lung diseases were also obtained in the same geographic location as of the patients. A screening interview was administered to participants enquiring about their general health and their medical history. The healthy subjects were matched for demographic variables. Demographic and clinical features of the subjects participating in the study are shown in Tables S1.

2.2 | Plasma collection

Peripheral blood was collected from participants in Ethylenediaminetetraacetic acid (EDTA) tubes or Citrated Cell Preparation tubes (Citrated CPT) tubes after informed consent. The samples from the University of Pittsburgh were collected in sodium citrate tubes while the rest were collected in EDTA tubes. Plasma was separated from blood no later than 1 h from the time of collection. Blood samples were centrifuged at $1100 \times g$ for 10 min at room temperature. Platelet free plasma was obtained by centrifugation at $2500 \times g$ for 15 min.

2.3 | Preparation of extracellular vesicle samples

Size exclusion chromatography (SEC) was performed for isolating EVs using qEV original 35 nm pore size columns and automated fraction collector V1 setup (Izon Science US Ltd, MA, USA), from 150 μL of plasma, as per manufacturer's instructions. Plasma was clarified by centrifugation at $10,000 \times g$ for 10 min at 4°C . SEC columns were equilibrated with phosphate buffered saline and clarified plasma was loaded onto the column. After discarding 3 mL of void volume, 2 mL of EVs were collected and concentrated using 300 KDa centrifugal filters (Pall corp, NY, USA) at $3500 \times g$ at 4°C .

2.4 | Cryo-electron microscopy of extracellular vesicles

Four microliters of the EV solution were added to Lacey carbon grids (300-mesh; Electron Microscopy Sciences) that were negatively glow-discharged for 80s at 30 mA. Excess sample was removed by blotting once for 3.5 s with Whatman filter paper and then the grid was plunge-frozen in liquid ethane cooled by liquid nitrogen using a Vitrobot plunge-freezer (Thermo-Fisher Scientific, Hillsboro, OR, USA). The vitrified samples were imaged using a Talos Arctica 200 kV transmission electron microscope (Thermo-Fisher Scientific, Hillsboro, OR, USA). The SerialEM software was used to collect 2D images under low-dose conditions with dose fractionation. Images were recorded at $79,000\times$ magnification on a K3 Summit direct electron detector (Gatan Inc,

Pleasanton CA, USA) with an effective pixel size of 1.1 Å in dose fractionation mode. For each image, 40 frames were recorded over 2 s exposure time at a dose rate of ~20 electrons/pixel/s. The movie frames were aligned using SerialEM.

2.5 | Nanoparticle tracking analysis (NTA) of extracellular vesicles

EV concentration and size were determined using Zetaview Quatt instrument (Particle Metrix, Germany) in scatter mode with 520 nm laser and sCMOS camera. The instrument was calibrated using 100 nm fluorescent polystyrene beads. EVs were diluted using filtered PBS. Data acquisition was performed at the following parameters: Sensitivity = 80; Shutter = 100; Track length = 15; Minimum brightness = 20; Cycles/position = 2/11.

2.6 | Lysis of extracellular vesicles, protein estimation and western blot

The EV fractions from SEC were concentrated down to 50 µL in vacuum at 30°C, lysed in RIPA buffer (Sigma Aldrich, USA) at 95°C for 10 min and sonicated in ultrasonic sonicator bath (Branson Ultrasonics Corp, CT, USA) for 2 min. The supernatants containing the EV proteins were collected after centrifugation at 30,000 × *g* for 15 min at 4°C. The protein concentration was estimated using Pierce BCA protein assay kit (Thermo Fisher Scientific, MA, USA) using bovine serum albumin as standard. Western blot was performed following SDS PAGE using standard protocols involving CD9 (Cat# 13174S) antibody and Alix (Cat #2171T) from Cell Signaling Technologies; HSP70 (Cat# SC-24), albumin (Cat# SC-271605) and GM130 (Cat# SC-55591) antibodies from Santa Cruz Biotechnology, CA, USA; Platelet factor 4 (PF4) (Cat #PA5-120547) and platelet glycoprotein IIb (CD41) (Cat# PA5-79527) from Invitrogen. Proteins were detected by chemiluminescence using Clarity ECL substrate (Bio-Rad Laboratories, CA, USA) on Chemidoc-MP Gel imaging system (Bio-Rad Laboratories).

2.7 | Liquid chromatography–mass spectrometry/mass spectrometry analysis

Dried peptide samples were dissolved in 4.8 µL of 0.25% formic acid with 3% (vol/vol) acetonitrile and 4 µL of each sample was injected into an EasyLC 1000 (Thermo Fisher Scientific). Peptides were separated on a 45-cm in-house packed column (360 µm OD×75 µm ID) containing C18 resin (2.2 µm, 100 Å; Michrom Bioresources, CA, USA). 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 flow rate of 250 nL/min. The Easy-nLC 1000 was coupled online with a hybrid high-resolution LTQ-Orbitrap Velos Pro mass spectrometer (Thermo Fisher Scientific). The mass spectrometer was operated in the data-dependent mode, in which a full-scan MS (from *m/z* 300–1500 with the resolution of 30,000 at *m/z* 400) was performed, followed by MS/MS of the 10 most intense ions [normalized collision energy – 30%; automatic gain control (AGC) – 3E4, maximum injection time – 100 ms; 90 s exclusion]. The raw files were searched directly against the human Uniprot database version downloaded August 2017 with no redundant entries, using Byonic search engine (Protein Metrics, CA, USA) loaded into Proteome Discoverer 2.2 software (Thermo Fisher Scientific). Initial precursor mass tolerance was set at 10 ppm, the final tolerance was set at 6 ppm, and ion trap mass spectrometry (ITMS) MS/MS tolerance was set at 0.6 Da. Search criteria included a static carbamidomethylation of cysteines (+57.0214 Da), and variable modifications of oxidation (+15.9949 Da) on methionine residues and acetylation (+42.011 Da) at N terminus of proteins. Search was performed with full trypsin/P digestion and allowed a maximum of two missed cleavages on the peptides analyzed from the sequence database. The false-discovery rates of proteins and peptides were set at 0.01. All protein and peptide identifications were grouped, and any redundant entries were removed. Only unique peptides and unique master proteins were reported.

2.8 | Biomarker discovery

Protein abundance values were obtained from mass spectrometry analysis. These values were obtained by summing the peak intensities of all peptides mapped to the protein and represent the abundance of the individual proteins in each sample. Analysis of protein abundance values is well-suited for comparing each protein across samples and has been used for biomarker discovery (Adduri et al., 2022; Behrmann et al., 2020; Ramirez-Martinez et al., 2021; Zhao et al., 2020). The raw protein abundance values were $\log_2(n + 1)$ transformed prior to differential expression analysis. Proteins detected in less than 10% of the samples were removed from further analysis. A *t*-test was used for identifying differentially expressed genes between different classes of samples, using an FDR cut off value of 0.05 to correct for multiple hypothesis testing. Proteins were considered differentially expressed if both (A) the *t*-test indicates significant differential expression and (B) $|\log_2 \text{fold change}| \geq 0.585$ (equivalent to 50% increase or decrease in expression). Glmnet package (Friedman et al., 2010) was used for performing LASSO regression. Five-fold

cross validation was performed using `cv.glmnet` function to estimate cross validation error at varying values of tuning parameter (λ). Features included in the binomial model at λ_{1se} were selected as protein signature. LASSO risk score was calculated using the intercept and coefficients obtained in LASSO. The generalizability of the classifier was analyzed using leave one out cross validation (LOOCV) using Caret package in a binomial model (Kuhn, 2008).

2.9 | Biomarker validation

Biomarker validation was performed using Enzyme-linked immunosorbent assay (ELISA). ELISA was performed on standard sandwich ELISA kits (Aviva Systems Biology, CA, USA) as per manufacturer's instructions and the microplates were read on Synergy microplate reader (BioTek Instruments Ltd, VT, USA). SPSS package was used for constructing binomial logistic regression models and ROC curves. The logistic regression classifier was further evaluated by plotting calibration curves and calculating the concordance index (Jr, 2015). Individual risk scores were calculated for each subject in the logistic regression model by subtracting 0.5 from predicted probability value of the subject.

2.10 | Statistical analysis

All statistical analyses were performed in the R environment (R4.0.3) if not stated otherwise. The student t test was used to compare the means of two groups. P value < 0.05 was considered statistically significant.

3 | RESULTS

3.1 | Clinical features of IPF, other-ILDs and healthy subjects

Clinical characteristics of participants diagnosed with IPF, CHP and NSIP, and healthy subjects are summarized in Table S1. Majority of participants are non-Hispanic whites (USA) and East Asians (Japan), respectively, and are smokers over 50 years of age. CHP and NSIP together are referred as 'other ILDs' hereafter. Diagnosis of IPF and other ILDs was made according to the consensus criteria (Raghu et al., 2011).

3.2 | Characterization and proteomic landscape of plasma extracellular vesicles

Extracellular vesicles were isolated from blood plasma in cohort-I consisting 20 IPF, other ILDs (11 CHP and 8 NSIP) and 20 healthy subjects. Cryo-electron microscopy imaging of plasma EV preparations revealed the presence of intact spherical vesicles consisting of a lipid bi-layer and were on average 100 nm in diameter (Figure 1a). We did not observe any morphological differences among EVs from different conditions (Figure 1a). We further performed nanoparticle tracking analysis to determine the size distribution of EV preparations and observed that the majority EVs were in the range of 50–200 nm in diameter (Figure 1b). We performed NTA on 10 healthy subjects, 10 IPF, 10 CHP and 8 NSIP patients. The EV concentrations in plasma of ILD patients were significantly higher than healthy subjects (Figure 1c). However, there was no difference in EV concentrations between the different ILDs (Figure 1c). Interestingly, the mean and median size of EVs was slightly lower within all ILDs compared to healthy subjects (Figure 1d). In addition, the tetraspanin protein CD9 (Figure 1e) and exosome luminal proteins such as Alix (Figure 1e) and HSP70 (Figure 1e) were detected. Further, 130 KDa Cis-Golgi Matrix Protein (GM130), a Golgi apparatus protein, was absent in EV preparation suggesting that our EV isolated were not contaminated with other cellular organelles (Figure 1e). Altogether, these results demonstrate the qualitative characteristics of EVs and suggest that the plasma EV preparations were enriched with exosomes. EVs from blood platelets are known to contaminate plasma EV preparations. We tested for the presence of two platelet specific proteins, namely CD41 and platelet factor 4 in our EV preparations. Their absence in our EV samples suggests that platelet contamination was minimal in our EV preparations (Figure 1e). Plasma EV purification steps and subsequent proteomic analysis are affected by contamination with high-abundance blood proteins such as albumin, lipoproteins, and immunoglobulins. Albumin (the most abundant plasma/serum protein) was used as a surrogate marker for plasma protein contamination while HSP70 was used as internal control EV protein. Similar levels of albumin contamination suggest that impurities were even across the EV isolates, thereby, unlikely to differentially affect the protein detection across the comparison groups and confound statistical comparisons (Figure 1f).

Mass spectrometry based proteomic profiling of EVs from 20 IPF, 19 other ILDs (11 CHP and 8 NSIP) and 20 healthy subjects quantified the expression of 520 proteins. Principle component analysis revealed that the protein profiles of EVs from healthy subjects, CHP, NSIP and IPF clustered according to the lung conditions (Figure 1g). Healthy subjects appeared distinct from all the ILDs. While the three ILDs clustered separately, they exhibited major overlap, underscoring the similarities in the proteome profiles among different ILDs.

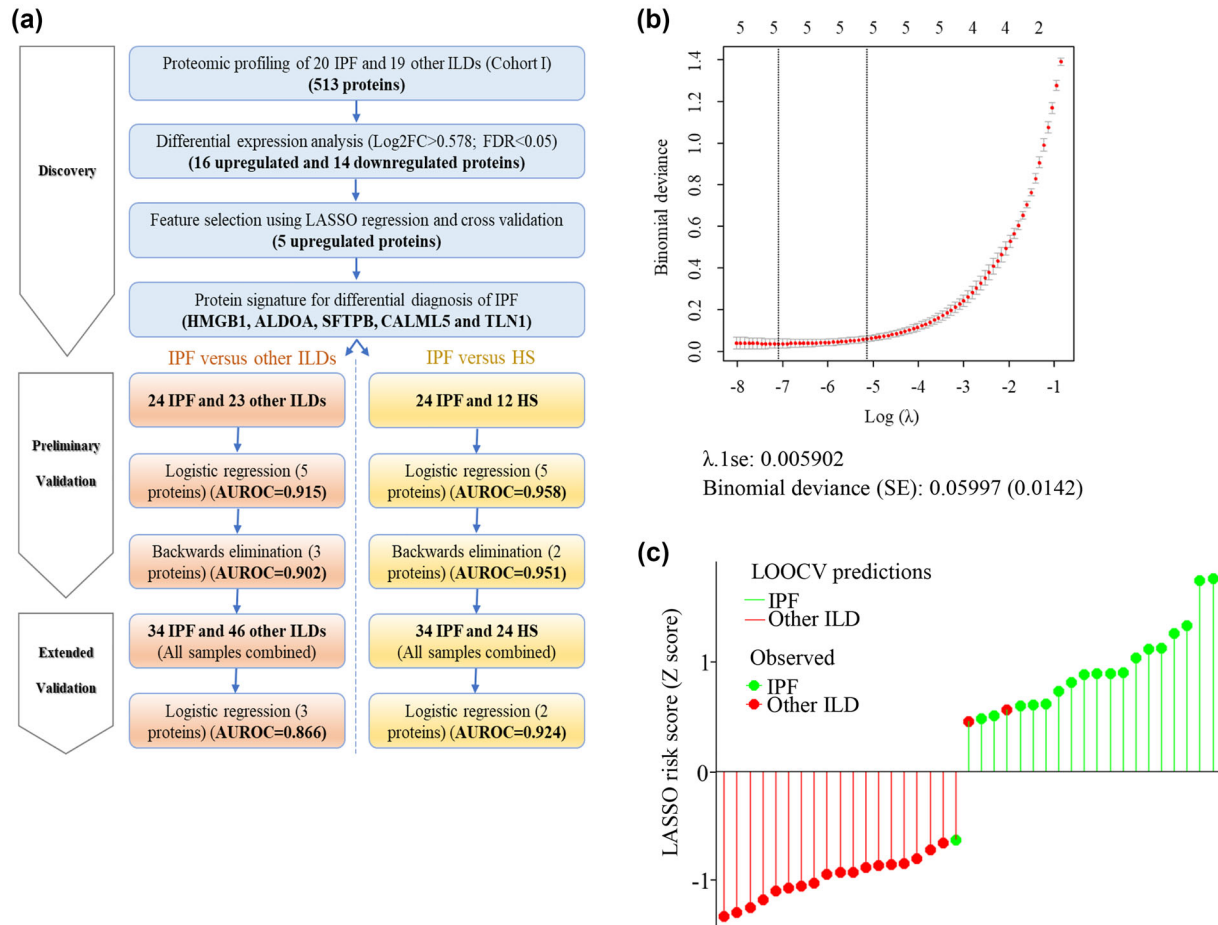


FIGURE 2 Study design and biomarker discovery. (a) Study design of systematic discovery of EV protein biomarker signature. Mass spectrometry based proteomic analysis was used for quantifying plasma EV proteins. Least Absolute Shrinkage and Selection Operator coupled with five-fold cross-validation was used for identifying a protein signature that discriminates IPF from other-ILDs. As part of preliminary validation, the signature was validated in an independent set of samples. To minimize the number of proteins in the signature, stepwise backwards elimination was used. The minimal signature was further tested on additional samples in extended validation where additional samples as well as samples used for preliminary validation were combined to increase statistical power. The protein signature identified in discovery phase was also tested to discriminate IPF from healthy subjects in preliminary and extended validations identical to the comparison of IPF and other-ILDs. (b) Plot showing cross validation error at different values of tuning parameter (λ) in lasso regression in cohort-I. Red dotted curve shows the estimated binomial deviance based on 5-fold cross-validation while the whiskers represent \pm one standard error, respectively. The vertical dotted lines show the locations of $\lambda_{.min}$ and $\lambda_{.1se}$. The numbers across the top show the number of nonzero coefficient estimates. SE, Standard error. $\lambda_{.min}$ is the value of λ at which cross validation error is the least and $\lambda_{.1se}$ is the largest value of λ within 1 standard error of $\lambda_{.min}$. (c) Lollipop plot showing the performance of leave one out cross validation (LOOCV) of the 5-protein signature. The color of the lines represents observed classes, and the circles represent predicted class in LOOCV. LASSO risk scores were converted to Z-scores to make the values interpretable.

3.3 | EV protein biomarkers for non-invasive differential diagnosis of IPF

3.3.1 | Biomarker discovery

A schematic of biomarker discovery and validation pipeline is shown in Figure 2a. To identify biomarkers that distinguish IPF from other ILDs, we performed differential expression analysis on the proteome profiles of EVs isolated from plasma of 20 IPF and a total of 19 other ILDs patients. Other ILDs patients included 11 CHP and 8 NSIP samples exhibiting fibrosis that served as ILD as reference dataset to identify proteins specific to IPF. A total of 30 differentially expressed genes were identified at FDR and $|\log_2$ fold change| cut offs of 0.05 and 0.585, respectively (Table S2). Based on the premise that upregulated proteins as biomarkers are more suitable for diagnostic use in clinical settings and with the aim to develop an easy-to-adopt assay, we prioritized upregulated proteins for further analysis. We applied Least Absolute Shrinkage and Selection Operator (LASSO) on mass-spectrometry protein abundances of upregulated proteins, coupled with 5-fold cross-validation to minimize prediction error to identify robust biomarker panel. At $\lambda_{.1se}$ (0.000837), a five-protein signature comprising High mobility group box protein 1 (HMGB1), surfactant protein B (SFTPB), Aldolase A (ALDOA), calmodulin like 5 (CALML5) and Talin-1 (TLN1) discriminated IPF from other ILDs with minimum cross validation error (Figure 2b). We performed leave one out cross validation (LOOCV)

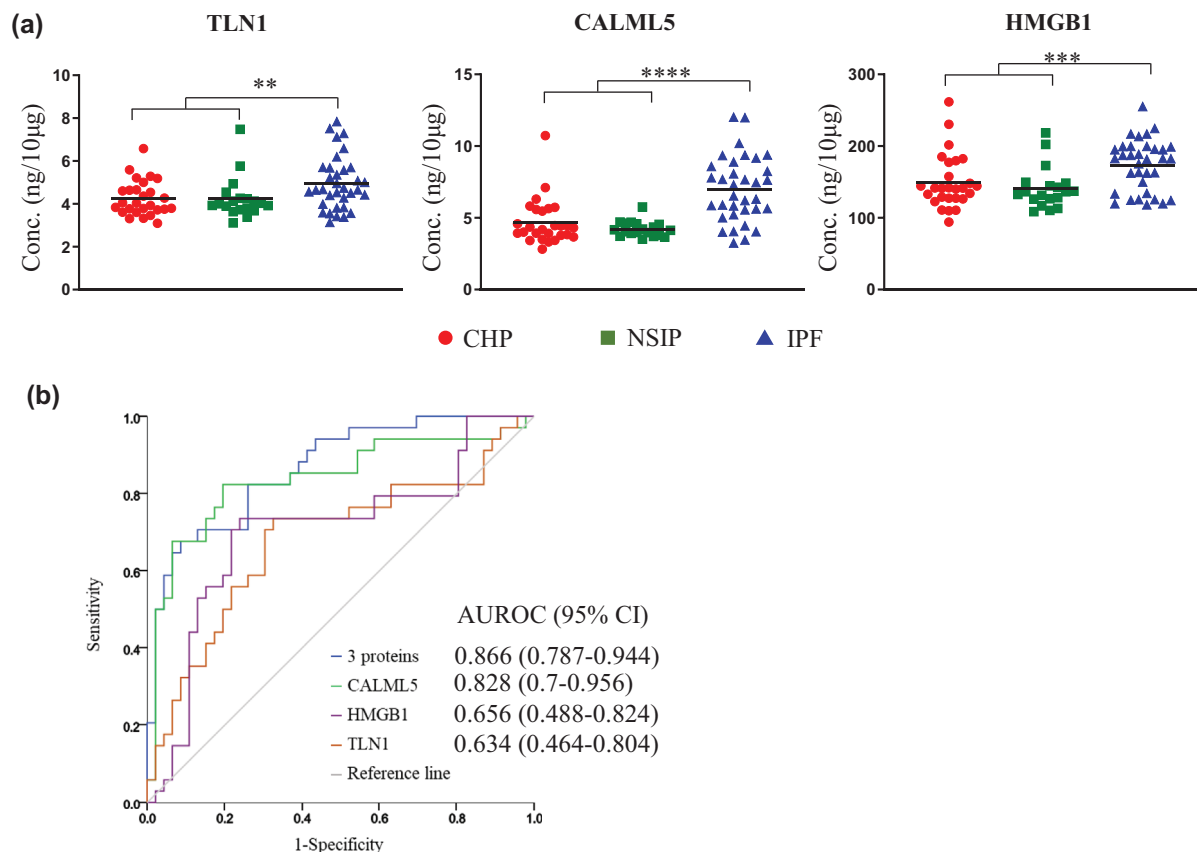


FIGURE 3 Extended validation of EV protein biomarker signature for differential diagnosis of IPF from other-ILDs. (a) As part of extended validation, expression levels of three proteins in plasma EVs of 34 IPF and 46 other-ILDs were measured using sandwich ELISA with purified recombinant human protein as standard. The asterisks *, ** and **** denote $p < 0.05$, $p < 0.001$ and $p < 0.0001$, respectively. (b) Receiver operating characteristic (ROC) curves for logistic regression model generated for two proteins on 34 IPF patients and 46 other-ILDs.

of the 5-protein signature on the proteomics data. This signature displayed an accuracy of 0.92 and Kappa of 0.85 in LOOCV and predicted 3 misclassifications out of 39 samples tested (Figure 2c). This suggests that the signature is less likely to overfit the training dataset.

3.3.2 | Preliminary validation

Encouraged by our findings in the discovery phase, we first estimated the levels of our protein biomarkers in plasma EVs of 24 IPF, 23 other ILDs and 12 healthy subjects from an independent validation cohort using ELISA. The expression levels of HMGB1, ALDOA and CALML5 were different between IPF and other ILDs in an independent cohort (Figure S1A). To evaluate the efficacy of the five EV proteins for differential diagnosis of IPF, we performed logistic regression and constructed ROC curves (Figure S1B). Our protein signature classified IPF and other ILDs with excellent efficacy in independent samples (AUROC = 0.915, 95% CI: 0.819–1.011; Figure S1B). To check whether the number of proteins in the classifier could be minimized, we applied a backward stepwise elimination approach and selected a sparse model comprising CALML5, HMGB1 and TLN1 which exhibited an AUROC of 0.902 (95% CI: 0.805–1) (Figure S1B).

3.3.3 | Extended validation

Next, we performed an extended validation of this three-protein signature on additional samples comprising 10 IPF, 15 CHP and 8 NSIP (Figure 3a). We combined both preliminary and extended validation samples to increase the sample size for logistic regression. In the combined dataset the three-protein signature showed an AUROC of 0.866 (0.787–0.944) (Figure 3b, Figure S2, Table 1). Taken together, these results suggest that our EV protein biomarker panel has good efficacy in differential diagnosis of IPF. These markers need to be validated in large-scale prospective cohort studies to establish their usability in clinics.

TABLE 1 Summary of the logistic regression model generated for different classifiers for discriminating IPF from other ILDs and healthy subjects.

Classifier	AUROC (95% CI)	Sensitivity	Specificity	PPV	NPV	DOR
Classifiers for discriminating IPF from other ILDs						
CALML5	0.828 (0.7–0.956)	0.618	0.913	0.840	0.764	16.962
TLN1	0.634 (0.464–0.804)	0.382	0.848	0.650	0.650	3.449
HMGB1	0.656 (0.488–0.824)	0.588	0.783	0.667	0.720	5.143
3 proteins	0.866 (0.787–0.944)	0.647	0.913	0.846	0.778	19.250
Classifiers for discriminating IPF from other healthy subjects						
CALML5	0.703 (0.65–0.784)	0.765	0.625	0.743	0.652	5.417
HMGB1	0.795 (0.693–0.872)	0.735	0.792	0.833	0.679	10.556
2 proteins	0.924 (0.858–0.99)	0.824	0.875	0.903	0.778	32.667

Abbreviations: DOR, diagnostic odds-ratio; NPV, negative predictive value; PPV, positive predictive value.

3.4 | EV protein biomarkers for discriminating IPF from healthy subjects

3.4.1 | Preliminary validation (IPF vs. healthy subjects)

Further, three out of five proteins, namely CALML5, HMGB1, and SFTPB exhibited significantly high expression in IPF compared to healthy subjects (Figure S3A). Encouraged by this finding, we next investigated if this five-protein biomarker signature is suitable for differentiating IPF from healthy subjects as well. Logistic regression followed by ROC curves revealed that this 5-protein signature could distinguish IPF from healthy controls with good accuracy (AUROC = 0.958) (Figure S3B). Backward stepwise elimination approach yielded a model comprising CALML5 and HMGB1 which exhibited AUROC of 0.951 (95% CI:0.876–1.026) (Figure S3B).

3.4.2 | Extended validation

The two proteins were quantified in additional 10 IPF and 12 healthy subjects (Figure 4a). We combined both preliminary and extended validation samples to increase the sample size for logistic regression. In the combined dataset, the two proteins showed an AUROC of 0.924 (0.858–0.99) (Figure 4b, Figure S4, Table 1). Taken together, these results suggest that the EV biomarkers have excellent efficacy in discriminating IPF from healthy subjects. These markers need to be validated in large-scale prospective cohort studies to establish their usability in clinics.

4 | DISCUSSION

Definitive diagnosis of IPF is often challenging because few other fibrotic lung diseases also exhibit pathological features similar to IPF. A vast majority of studies investigated the diagnostic efficacy of individual proteins for differential diagnosis of IPF, that showed poor sensitivity and specificity (Greene et al., 2002; Morais et al., 2015; Rosas et al., 2008; White et al., 2016). For instance, Greene and coworkers (Greene et al., 2002) reported elevated surfactant proteins A and D in sera of IPF patients compared to healthy subjects. But these proteins were also found to be elevated in other lung conditions consistently. In another study, Morais and coworkers (Morais et al., 2015) proposed that matrix metalloproteinases 1 and 7 in serum would be suitable for differentiating IPF from other ILDs but the combined accuracy of these two serum proteins is low (0.74). Similarly, White and coworkers (2016) identified three plasma protein biomarkers—Plasma Surfactant Protein-D, Matrix Metalloproteinase-7, and Osteopontin (White et al., 2016). The combined accuracy of these proteins for differential diagnosis was reported to be 0.77 which is lower than the accuracy of our biomarker signature. Due to etiological and molecular complexity of IPF and other ILDs, use of a single biomarker to differentiate IPF from other closely resembling ILDs has not been successful. Notably, the current ATS/ERS/JRS/ALAT clinical practice guidelines strongly recommend against use of serum/plasma biomarkers such as MMP9, SFTPD, CCL18 and KL-6 for the purpose of distinguishing IPF from other ILDs, owing mainly to their poor efficacy (Raghu et al., 2018). In this study, we identified a five-protein signature, none of them overlap with the above-mentioned biomarkers, for differential diagnosis of IPF from other ILDs with greater specificity and sensitivity. This study describes a pipeline for mass spectrometry and ELISA based systematic discovery and development of plasma EV protein biomarkers that can be applied for other disease conditions as well. We used a robust machine learning based feature selection methodology for biomarker panel selection which is expected to account for molecular heterogeneity of the disease and hence may alleviate the drawbacks of a single biomarker. Though

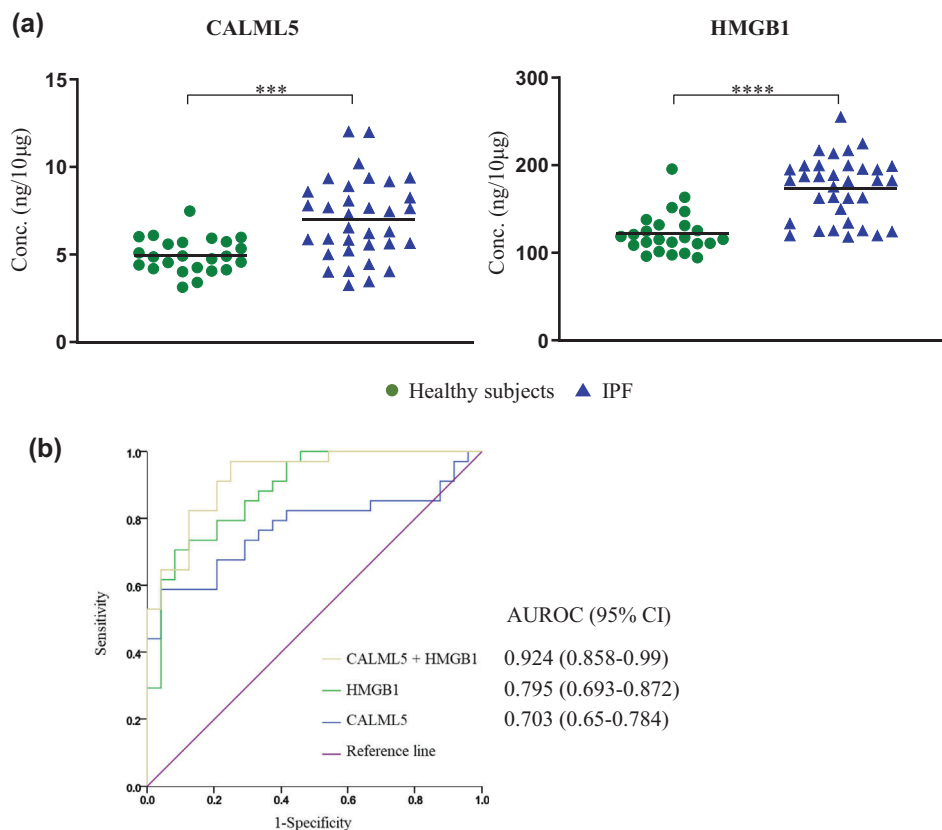


FIGURE 4 Extended validation of EV protein biomarker signature for differentiating IPF from healthy subjects. (a) As part of extended validation, the levels of two proteins (CALML5 and HMGB1) in plasma EVs of 34 IPF and 24 healthy subjects were assessed using sandwich ELISA with purified recombinant human protein as standard. The asterisks *, **, *** and **** denote $p < 0.05$, $p < 0.01$, $p < 0.001$ and $p < 0.0001$, respectively. (b) ROC curves for logistic regression model generated for two proteins on 34 IPF patients and 24 healthy subjects. This model achieved specificity of 53% at 99% sensitivity.

mass spectroscopy methods purposed for biomarker discovery currently identify a limited number of proteins per sample and generally quantify high abundant proteins, technical improvements in future may help in detecting very low abundant proteins as well.

Our protein signature comprised of SFTPB, ALDOA, HMGB1, CALML5 and TLN1. SFTPB levels in serum of IPF patients are known to be elevated (Kahn et al., 2018) and are associated with poor survival (Papaioannou et al., 2016). Similarly, proteomic analysis of broncho-alveolar lavage fluid (BALF) revealed upregulation of ALDOA in IPF patients presented with acute exacerbations (Carleo et al., 2020). In IPF patients, HMGB1 protein was elevated in inflammatory cells and in hyperplastic epithelial cells (Hamada et al., 2008). Serum HMGB1 levels have been reported to be high in IPF patients than in healthy subjects (Yamaguchi et al., 2020). In addition, higher levels of HMGB1 in IPF are associated with acute exacerbations and poor survival (Yamaguchi et al., 2020). This study focused on the discovery and development of biomarkers enriched in plasma EVs. Currently, the circulating levels of these biomarkers in IPF and other ILDs are unknown.

5 | CONCLUSIONS

The diagnostic work up of IPF and other ILDs is still evolving and there is risk of misdiagnosis. To the best of our knowledge, this is the first EV-based noninvasive protein signature for IPF diagnosis. In addition, the protein signature was discovered and validated in cohorts from two different geographic locations. However, we also acknowledge that the two retrospective cohorts used in this study are of small sample size. Therefore, the efficacy of the protein signature needs to be validated in large prospective cohorts.

AUTHOR CONTRIBUTIONS

N.V.K. conceived the study. N.V.K. and R.S.R.A. designed the experiments with inputs from K.C., K.V.A., R.V., A.K.P., S.P.d.F., H.I., Y.Z., D.A.S., A.H.M., L.L., S.C., D.J.K. R.S.R.A., K.C., K.V.A., R.V. and A.K.P. performed experiments and analyzed data together with N.V.K., D.N. R.S.R.A., H.N. and J.W.M. performed statistical analysis. S.P.d.F., F.P.d.F., Y.H., H.I., N.H., Y.Z., K.F.G., L.M.S., M.A.P.F., G.M.H., D.J.K., I.O.R. provided specimens and collected patient data. Z.C. carried out the cryoEM experiments and analyzed data with N.V.K. and R.S.R.A. N.V.K. and R.S.R.A. wrote the original draft. All authors contributed into writing,

reviewing, and editing. N.V.K., I.O.R., S.S., G.M.H. and D.A.S. provided resources. N.V.K. acquired funding and supervised the study.

ACKNOWLEDGEMENTS

We thank Dr. Daria Filanov, Alpha Nano Tech LLC, Durham, NC, USA for helping with EV size analysis. Mass spectrometry analysis of EV samples was performed by Dr. Andrew Lemoff at the Proteomics Core at UT Southwestern Medical center, Dallas, TX. We thank Dr. Priyanka Sharma, MD Anderson Cancer center, Houston for critical reading and suggestions during manuscript preparation. We thank the Structural Biology Laboratory at UT Southwestern Medical Center which is partially supported by grant RP220582 from the Cancer Prevention & Research Institute of Texas (CPRIT) for cryo-EM studies.

CONFLICT OF INTEREST STATEMENT

A.K.P. is employed with Izon Science US Ltd. The company has no role in design of the study and acquisition of experimental data and interpretation. All other authors have no conflict of interests.

DATA AVAILABILITY STATEMENT

All data relevant to this study are included in the manuscript and its supplementary information files.

REFERENCES

- Adduri, R. S. R., Vasireddy, R., Mroz, M. M., Bhakta, A., Li, Y., Chen, Z., Miller, J. W., Velasco-Alzate, K. Y., Gopalakrishnan, V., Maier, L. A., Li, Li., & Konduru, N. V. (2022). Realistic biomarkers from plasma extracellular vesicles for detection of beryllium exposure. *International Archives of Occupational and Environmental Health*, 95(8), 1785–1796.
- Behrmann, A., Zhong, D., Li, Li., Cheng, Su. Li., Mead, M., Ramachandran, B., Sabaeifard, P., Goodarzi, M., Lemoff, A., Kronenberg, H M., & Towler, D A. (2020). PTH/PTHrP receptor signaling restricts arterial fibrosis in diabetic LDLR(-/-) mice by inhibiting myocardin-related transcription factor relays. *Circulation Research*, 126(10), 1363–1378.
- Boukouris, S., & Mathivanan, S. (2015). Exosomes in bodily fluids are a highly stable resource of disease biomarkers. *Proteomics Clinical Applications*, 9(3-4), 358–367.
- Carleo, A., Landi, C., Prasse, A., Bergantini, L., D'alessandro, M., Cameli, P., Janciauskiene, S., Rottoli, P., Bini, L., & Bargagli, E. (2020). Proteomic characterization of idiopathic pulmonary fibrosis patients: Stable versus acute exacerbation. *Monaldi Archives for Chest Disease*, 90(2), 180–190.
- Flaherty, K. R., King, T. E., Raghu, G., Lynch, J. P., Colby, T. V., Travis, W. D., Gross, B. H., Kazerooni, E. A., Toews, G. B., Long, Q., Murray, S., Lama, V. N., Gay, S. E., & Martinez, F. J. (2004). Idiopathic interstitial pneumonia: What is the effect of a multidisciplinary approach to diagnosis? *American Journal of Respiratory and Critical Care Medicine*, 170(8), 904–910.
- Friedman, J., Hastie, T., & Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, 33(1), 1–22.
- Greene, K. E., King, T. E., Kuroki, Y., Bucher-Bartelson, B., Hunninghake, G. W., Newman, L. S., Nagae, H., & Mason, R. J. (2002). Serum surfactant proteins-A and -D as biomarkers in idiopathic pulmonary fibrosis. *The European Respiratory Journal*, 19(3), 439–446.
- Hamada, N., Maeyama, T., Kawaguchi, T., Yoshimi, M., Fukumoto, J., Yamada, M., Yamada, S., Kuwano, K., Nakanishi, Y., & N. T. T. M. (2008). The role of high mobility group box1 in pulmonary fibrosis. *American Journal of Respiratory Cell and Molecular Biology*, 39(4), 440–447.
- Herazo-Maya, J. D., Sun, J., Molyneaux, P. L., Li, Q., Villalba, J. A., Tzouveleakis, A., Lynn, H., Juan-Guardela, B. M., Risquez, C., Osorio, J. C., Yan, X., Michel, G., Aurelien, N., Lindell, K. O., Klesen, M. J., Moffatt, M. F., Cookson, W. O., Zhang, Y., Garcia, J. G. N., ... Kaminski, N. (2017). Validation of a 52-gene risk profile for outcome prediction in patients with idiopathic pulmonary fibrosis: An international, multicentre, cohort study. *Lancet Respiratory Medicine*, 5(11), 857–868.
- Hinestrosa, J. P., Kurzrock, R., Lewis, J. M., Schork, N. J., Schroeder, G., Kamat, A M., Lowy, A. M., Eskander, R. N., Perrera, O., Searson, D., Rastegar, K., Hughes, J. R., Ortiz, V., Clark, I., Balcer, H. I., Arakelyan, L., Turner, R., Billings, P. R., Adler, M. J., ... Krishnan, R. (2022). Early-stage multi-cancer detection using an extracellular vesicle protein-based blood test. *Communications Medicine (London)*, 2, 29.
- Huda, Md. N., Nafiujjaman, M. d., Deaguero, I. G., Okonkwo, J., Hill, M. L., Kim, T., & Nurunnabi, Md. (2021). Potential use of exosomes as diagnostic biomarkers and in targeted drug delivery: Progress in clinical and preclinical applications. *ACS Biomaterials Science & Engineering*, 7(6), 2106–2149.
- Ishii, H. (2003). High serum concentrations of surfactant protein A in usual interstitial pneumonia compared with non-specific interstitial pneumonia. *Thorax*, 58(1), 52–57.
- Jr, F. E. H. (2015). *Regression Modeling Strategies*. 2nd ed.: Springer.
- Kahlert, C., Melo, S. A., Protopopov, A., Tang, J., Seth, S., Koch, M., Zhang, J., Weitz, J., Chin, L., Futreal, A., & Kalluri, R. (2014). Identification of double-stranded genomic DNA spanning all chromosomes with mutated KRAS and p53 DNA in the serum exosomes of patients with pancreatic cancer. *The Journal of Biological Chemistry*, 289(7), 3869–3875.
- Kahn, N., Rossler, A. K., Hornemann, K., Muley, T., Grünig, E., Schmidt, W., Herth, F. J. F., & Kreuter, M. (2018). C-proSP-B: A possible biomarker for pulmonary diseases? *Respiration*, 96(2), 117–126.
- Kolonics, F., Szeifert, V., Timár, C. I., Ligeti, E., & Lőrincz, Á. M. (2020). The functional heterogeneity of neutrophil-derived extracellular vesicles reflects the status of the parent cell. *Cells*, 9(12), 2718.
- Kuhn, M. (2008). Building predictive models in R using the caret package. *Journal of Statistical Software*, 28(5), 1–26.
- Li, D., Lai, W., & Fan, D., Fang, Q. (2021). Protein biomarkers in breast cancer-derived extracellular vesicles for use in liquid biopsies. *American Journal of Physiology*, 321(5), C779–C797.
- Melo, S. A., Luecke, L. B., Kahlert, C., Fernandez, A. F., Gammon, S. T., Kaye, J., Lebleu, V. S., Mittendorf, E. A., Weitz, J., Rahbari, N., Reissfelder, C., Pilarsky, C., Fraga, M. F., Piwnicka-Worms, D., & Kalluri, R. (2015). Glypican-1 identifies cancer exosomes and detects early pancreatic cancer. *Nature*, 523(7559), 177–182.
- Merrell, K., Southwick, K., Graves, S. W., Esplin, M. S., Lewis, N. E., & Thulin, C. D. (2004). Analysis of low-abundance, low-molecular-weight serum proteins using mass spectrometry. *The Journal of Biomolecular Techniques*, 15(4), 238–248.
- Morais, A., Beltrão, M., Sokhatska, O., Costa, D., Melo, N., Mota, P., Marques, A., & Delgado, L. (2015). Serum metalloproteinases 1 and 7 in the diagnosis of idiopathic pulmonary fibrosis and other interstitial pneumonias. *Respiratory Medicine*, 109(8), 1063–1068.

- Njock, M. S., Guiot, J., Henket, M. A., Nivelles, O., Thiry, M., Dequiedt, F., Corhay, J. L., Louis, R. E., & Struman, I. (2019). Sputum exosomes: Promising biomarkers for idiopathic pulmonary fibrosis. *Thorax*, *74*(3), 309–312.
- Noth, I., Zhang, Y., Ma, S. F., Flores, C., Barber, M., Huang, Y., Broderick, S. M., Wade, M. S., Hysi, P., Scuirba, J., Richards, T. J., Juan-Guardela, B. M., Vij, R., Han, M. K., Martinez, F. J., Kossen, K., Seiwert, S. D., Christie, J. D., Nicolae, D., ... Garcia, J. G. (2013). Genetic variants associated with idiopathic pulmonary fibrosis susceptibility and mortality: A genome-wide association study. *Lancet Respiratory Medicine*, *1*(4), 309–317.
- Onishi, Y., Kawamura, T., Higashino, T., Kagami, R., Hirata, N., & Miyake, K. (2020). Clinical features of chronic summer-type hypersensitivity pneumonitis and proposition of diagnostic criteria. *Respiratory Investigation*, *58*(1), 59–67.
- Papaoannou, A. I., Kostikas, K., Manali, E. D., Papadaki, G., Roussou, A., Spathis, A., Mazioti, A., Tomos, I., Papanikolaou, I., Loukides, S., Chainis, K., Karakitsos, P., Griese, M., & Papiris, S. (2016). Serum levels of surfactant proteins in patients with combined pulmonary fibrosis and emphysema (CPFE). *PLoS One*, *11*(6), e0157789.
- Phan, T. H., Kim, S. Y., Rudge, C., & Chrzanowski, W. (2022). Made by cells for cells – Extracellular vesicles as next-generation mainstream medicines. *Journal of Cell Science*, *135*(1), jcs259166.
- Raghu, G., Collard, H. R., Egan, J. J., Martinez, F. J., Behr, J., Brown, K. K., Colby, T. V., Cordier, J. F., Flaherty, K. R., Lasky, J. A., Lynch, D. A., Ryu, J. H., Swigris, J. J., Wells, A. U., Ancochea, J., Bouros, D., Carvalho, C., Costabel, U., Ebina, M., ... Schönemann, H. J. (2011). An official ATS/ERS/JRS/ALAT statement: Idiopathic pulmonary fibrosis: Evidence-based guidelines for diagnosis and management. *American Journal of Respiratory and Critical Care Medicine*, *183*(6), 788–824.
- Raghu, G., Remy-Jardin, M., Myers, J. L., Richeldi, L., Ryerson, C. J., Lederer, D. J., Behr, J., Cottin, V., Danoff, S. K., Morell, F., Flaherty, K. R., Wells, A., Martinez, F. J., Azuma, A., Bice, T. J., Bouros, D., Brown, K. K., Collard, H. R., Duggal, A., ... Wilson, K. C. (2018). Diagnosis of idiopathic pulmonary fibrosis. An official ATS/ERS/JRS/ALAT clinical practice guideline. *American Journal of Respiratory and Critical Care Medicine*, *198*(5), e44–e68.
- Ramirez-Martinez, A., Zhang, Y., Chen, K., Kim, J., Cenik, B. K., Mcanally, J. R., Cai, C., Shelton, J. M., Huang, J., Brennan, A., Evers, B. M., Mammen, P. P. A., Xu, L., Bassel-Duby, R., Liu, N., & Olson, E. N. (2021). The nuclear envelope protein Net39 is essential for muscle nuclear integrity and chromatin organization. *Nature Communication*, *12*(1), 690.
- Record, M., Silvente-Poirot, S., Poirot, M., & Wakelam, M. O. (2018). Extracellular vesicles: Lipids as key components of their biogenesis and functions. *Journal of Lipid Research*, *59*(8), 1316–1324.
- Richards, T. J., Kaminski, N., Baribaud, F., Flavin, S., Brodmerkel, C., Horowitz, D., Li, K., Choi, J., Vuga, L. J., Lindell, K. O., Klesen, M., Zhang, Y., & Gibson, K. F. (2012). Peripheral blood proteins predict mortality in idiopathic pulmonary fibrosis. *American Journal of Respiratory and Critical Care Medicine*, *185*(1), 67–76.
- Rosas, I. O., Richards, T. J., Konishi, K., Zhang, Y., Gibson, K., Lokshin, A. E., Lindell, K. O., Cisneros, J., Macdonald, S. D., Pardo, A., Scuirba, F., Dauber, J., Selman, M., Gochuico, B. R., & Kaminski, N. (2008). MMP1 and MMP7 as potential peripheral blood biomarkers in idiopathic pulmonary fibrosis. *PLoS Medicine*, *5*(4), e93.
- Thakur, B. K., Zhang, H., Becker, A., Matei, I., Huang, Y., Costa-Silva, B., Zheng, Y., Hoshino, A., Brazier, H., Xiang, J., Williams, C., Rodriguez-Barrueco, R., Silva, J. M., Zhang, W., Hearn, S., Elemento, O., Paknejad, N., Manova-Todorova, K., Welte, K., ... Lyden, D. (2014). Double-stranded DNA in exosomes: A novel biomarker in cancer detection. *Cell Research*, *24*(6), 766–769.
- Tian, F., Zhang, S., Liu, C., Han, Z., Liu, Y., Deng, J., Li, Y., Wu, X., Cai, L., Qin, L., Chen, Q., Yuan, Y., Liu, Y., Cong, Y., Ding, B., Jiang, Z., & Sun, J. (2021). Protein analysis of extracellular vesicles to monitor and predict therapeutic response in metastatic breast cancer. *Nature Communications*, *12*(1), 2536.
- Tu, C., Rudnick, P. A., Martinez, M. Y., Cheek, K. L., Stein, S. E., Slebos, R. J. C., & Liebler, D. C. (2010). Depletion of abundant plasma proteins and limitations of plasma proteomics. *Journal of Proteome Research*, *9*(10), 4982–4991.
- White, E. S., Xia, M., Murray, S., Dyal, R., Flaherty, C. M., Flaherty, K. R., Moore, B. B., Cheng, L., Doyle, T. J., Villalba, J., Dellaripa, P. F., Rosas, I. O., Kurtis, J. D., & Martinez, F. J. (2016). Plasma surfactant protein-D, matrix metalloproteinase-7, and osteopontin index distinguishes idiopathic pulmonary fibrosis from other idiopathic interstitial pneumonias. *American Journal of Respiratory and Critical Care Medicine*, *194*(10), 1242–1251.
- Wuyts, W. A., Cavazza, A., Rossi, G., Bonella, F., Sverzellati, N., & Spagnolo, P. (2014). Differential diagnosis of usual interstitial pneumonia: When is it truly idiopathic? *The European Respiratory Review*, *23*(133), 308–319.
- Yamaguchi, K., Iwamoto, H., Sakamoto, S., Horimasu, Y., Masuda, T., Miyamoto, S., Nakashima, T., Ohshimo, S., Fujitaka, K., Hamada, H., Hattori, N., & Y. T. (2020). Serum high-mobility group box 1 is associated with the onset and severity of acute exacerbation of idiopathic pulmonary fibrosis. *Respirology*, *25*(3), 275–280.
- Zhang, Y., Jiang, M., Nourai, M., Roth, M. G., Tabib, T., Winters, S., Chen, X., Sembrat, J., Chu, Y., Cardenes, N., Tuder, R. M., Herzog, E. L., Ryu, C., Rojas, M., Lafyatis, R., Gibson, K. F., McDyer, J. F., Kass, D. J., & Alder, J. K. (2019). GDF15 is an epithelial-derived biomarker of idiopathic pulmonary fibrosis. *American Journal of Physiology – Lung Cellular and Molecular Physiology*, *317*(4), L510–L521.
- Zhang, Y., Noth, I., Garcia, J. G. N., & Kaminski, N. (2011). A variant in the promoter of MUC5B and idiopathic pulmonary fibrosis. *New England Journal of Medicine*, *364*(16), 1576–1577.
- Zhao, L., Cong, X., Zhai, L., Hu, H., Xu, J. Yu., Zhao, W., Zhu, M., Tan, M., & Ye, B. Ce. (2020). Comparative evaluation of label-free quantification strategies. *Journal of Proteomics*, *215*, 103669.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Adduri, R. S. R., Cai, K., Velasco-Alzate, K., Vasireddy, R., Miller, J. W., de Frías, S. P., de Frías, F. P., Horimasu, Y., Iwamoto, H., Hattori, N., Zhang, Y., Gibson, K. F., Pal, A. K., Chen, Z., Nicastro, D., Li, Li., Cherian, S., Sholl, L. M., Shetty, S., ... Konduru, N. V. (2023). Plasma extracellular vesicle proteins as promising noninvasive biomarkers for diagnosis of idiopathic pulmonary fibrosis. *Journal of Extracellular Biology*, *2*, e98.

<https://doi.org/10.1002/jex2.98>