Respiratory Research

RESEARCH

Open Access

Plasma proteomic and machine learning models for differentiating idiopathic pulmonary fibrosis and connective tissue disease–associated interstitial lung disease: findings from a prospective cohort

Chen-Shiou Wu[1], De-Wei Lai[1,2], Yi-Ching Chen[3], Yi-Hsuan Yu[4], Lan-Yan Yang[3†] and Pin-Kuei Fu[1,3,4,5*†]

Abstract

Background Idiopathic pulmonary fibrosis (IPF) is a progressive fibrotic interstitial lung disease (ILD) with limited treatment options and poor prognosis. Differentiating IPF from connective tissue disease–associated ILD (CTD-ILD) is clinically challenging due to overlapping features, and reliable circulating biomarkers are lacking. Recent studies suggest that multi-marker proteomic models combined with machine learning may enhance diagnostic precision and prognostic assessment in fibrotic ILDs.

Methods We prospectively analyzed plasma samples from Taiwanese patients with fibrotic ILDs (IPF, $n=22$; CTD-ILD, $n=66$) using the Olink inflammation panel (92 proteins). Differentially expressed proteins were identified and subjected to integrative network analyses. Predictive classification models were developed using generalized linear modeling (GLM), decision tree, and random forest approaches. Prognostic relevance was evaluated with Kaplan–Meier and Cox regression analyses, and findings were validated in public transcriptomic datasets.

Results Among 92 proteins profiled, 23 showed significant differences between IPF and CTD-ILD. Four candidates—MMP-10, FGF-19, ADA, and TWEAK (TNFSF12)—consistently emerged as key discriminatory markers. The GLM model incorporating FGF-19, ADA, and TWEAK achieved the highest diagnostic accuracy (AUC 0.870; sensitivity 0.97; specificity 0.82), outperforming decision tree and random forest models. Transcriptomic validation confirmed TWEAK downregulation in ILD lung tissues and in TGF-β1–stimulated fibroblasts, linking it to canonical profibrotic signaling. Survival analysis showed significantly worse outcomes in IPF versus CTD-ILD (log-rank $p<0.001$), with MMP-10 associated with poor prognosis (HR 2.08, $p=0.007$) and TWEAK with favorable prognosis (HR 0.04, $p<0.001$).

Conclusions This study identifies distinct plasma proteomic signatures that differentiate IPF from CTD-ILD and highlights TWEAK as both a diagnostic and prognostic biomarker. A multi-marker GLM model demonstrated excellent

†Lan-Yan Yang and Pin-Kuei Fu contributed equally to this work.

*Correspondence:
Pin-Kuei Fu
yetquen@gmail.com

Full list of author information is available at the end of the article

diagnostic performance, supporting the clinical utility of plasma proteomics combined with machine learning to improve disease classification and risk stratification in fibrotic ILDs.

Clinical implication Integrating plasma proteomics with machine learning enables development of a multi-marker model that enhances diagnostic accuracy and prognostic evaluation in fibrotic interstitial lung diseases, supporting more precise disease classification and risk stratification in clinical practice.

Key Point

Plasma proteomics combined with machine learning improves IPF vs CTD-ILD classification and identifies prognostic biomarkers, advancing precision diagnosis in fibrotic interstitial lung disease.

Keywords Idiopathic pulmonary fibrosis, Connective tissue disease–associated interstitial lung disease, Plasma proteomics, Multi-marker predictive models, Machine learning, TWEAK

Introduction

Idiopathic pulmonary fibrosis (IPF) is a chronic, progressive interstitial lung disease (ILD) characterized by irreversible architectural distortion and excessive extracellular matrix deposition in the lungs [1]. IPF carries a dismal prognosis, with median survival ranging from 3 to 5 years, and its incidence is increasing worldwide, creating a substantial burden on patients and healthcare systems [2]. Early and accurate diagnosis is critical for optimizing management and timely initiation of antifibrotic therapy [3]. However, differentiating IPF from other fibrotic ILDs, particularly connective tissue disease–associated ILD (CTD-ILD), remains a major clinical challenge because of overlapping clinical and radiological features. Misclassification can delay appropriate treatment, affect prognosis, and complicate enrollment into clinical trials. This underscores an urgent need for reliable, non-invasive biomarkers that can support diagnostic precision and enable risk stratification in fibrotic ILDs.

The pathogenesis of IPF involves complex interactions among epithelial injury, immune dysregulation, and aberrant fibrotic repair [3]. While transforming growth factor-β (TGF-β) is a central driver of fibrogenesis, multiple cytokines, chemokines, and growth factors—including interleukin (IL)-6, tumor necrosis factor (TNF)-α, and platelet-derived growth factor (PDGF)—also contribute to disease progression [4–6]. Despite decades of investigation, single biomarkers such as KL-6, surfactant protein-A (SP-A), surfactant protein-D (SP-D), and matrix metalloproteinase-7 (MMP-7) have shown only modest diagnostic or prognostic performance [7–10]. Their limited predictive value likely reflects the biological heterogeneity of IPF and the multifactorial pathways underlying fibrogenesis.

With the advent of high-throughput proteomics, multi-marker approaches have emerged as a more robust strategy. Recent studies using plasma proteomic profiling have identified biomarker panels associated with ILD progression and survival [11–14]. For example, Maher et al. reported that circulating protein signatures predict disease progression in progressive fibrosing ILDs in a large multicenter cohort [13]. Similarly, Oldham et al. demonstrated proteomic biomarkers of survival in IPF, emphasizing the prognostic potential of circulating proteins [14]. Importantly, Huang et al. recently applied machine learning to plasma proteomics and successfully classified ILD subtypes, highlighting the clinical utility of combining proteomic signatures with computational modeling [15]. Collectively, these studies suggest that integrative proteomic and machine learning approaches can enhance disease classification beyond conventional clinical or imaging assessments [15–17].

In this context, we hypothesized that targeted plasma proteomics combined with machine learning could identify clinically relevant biomarkers to distinguish IPF from CTD-ILD and provide prognostic information. We prospectively analyzed plasma samples from Taiwanese patients with fibrotic ILDs using the Olink inflammation panel and applied multiple modeling strategies to develop predictive diagnostic algorithms. We further validated candidate biomarkers in public transcriptomic datasets and assessed their association with survival outcomes. Our findings highlight distinct proteomic signatures in IPF and propose a multi-marker predictive model with potential clinical application in precision diagnosis and risk stratification of fibrotic ILDs.

Methods

Study design, participants and ethical approval

This prospective cohort study was conducted at Taichung Veterans General Hospital, Taiwan, and approved by the institutional review board (IRB numbers CE18325B and CG24056C). Written informed consent was obtained from all participants. Patients with fibrotic ILDs were consecutively recruited and classified as IPF or CTD-ILD based on multidisciplinary discussion and established diagnostic criteria. Clinical characteristics and follow-up outcomes were collected at enrollment.

Plasma collection and proteomic profiling

Based on the established role of immune dysregulation and chronic inflammation in IPF and CTD-ILD, we selected the Olink® Target 96 Inflammation panel to capture clinically relevant immune–fibrotic signals in circulation. This focused panel enables sensitive detection of inflammation-related proteins with direct relevance to fibrotic lung disease, facilitating the identification of discriminatory biomarkers between IPF and CTD-ILD.

Peripheral blood samples were obtained at baseline. Plasma was separated within 2 h of collection by centrifugation at 1,500 × g for 15 min, aliquoted, and stored at −80 °C. Samples were analyzed in a single batch using the Olink® Target 96 Inflammation panel, which quantifies 92 inflammation-related proteins through proximity extension assay (PEA). Protein expression levels were normalized and reported as normalized protein expression (NPX) values on a log2 scale. Further details of the PEA technology are available in previous publications [13].

Differential protein analysis and network exploration

Group differences in protein expression between IPF and CTD-ILD were assessed using the Wilcoxon rank-sum test, with significance defined as false discovery rate (FDR)–adjusted $p < 0.05$. To explore regulatory mechanisms, transcription factor (TF) enrichment and pathway analyses were performed with Enrichr (https://maayanlab.cloud/Enrichr/), while protein–protein interaction (PPI) and co-expression networks were constructed using GeneMANIA (https://genemania.org).

Public transcriptomic validation

Candidate biomarkers were further examined in three publicly available transcriptomic datasets (GSE47460, GSE225982, and GSE225159). Expression patterns were compared across ILD, COPD, and control groups, as well as under TGF-β1 stimulation. Statistical analyses employed two-tailed Student's t-tests or one-way ANOVA with Tukey's post hoc tests, as appropriate.

Enzyme-Linked Immunosorbent Assay (ELISA)

Plasma concentrations of candidate proteins were quantified using commercially available ELISA kits according to the manufacturers' instructions. Human plasma samples were collected, processed under standardized conditions, and stored at −80 °C until analysis. Prior to measurement, samples were thawed slowly on ice. Plasma levels of MMP-10 were measured using the Human Total MMP-10 ELISA kit (DY910, R&D Systems). FGF-19 concentrations were determined using the Human FGF-19 ELISA kit (DY969, R&D Systems), and TWEAK levels were quantified using the Human TWEAK ELISA kit (DY1090, R&D Systems). For these colorimetric ELISA assays, absorbance was measured using a microplate reader (EnSpire®, PerkinElmer) at 450 nm, with 570 nm used as a reference wavelength for background correction after termination of the color development reaction. Plasma adenosine deaminase (ADA) levels were quantified using a chemiluminescence-based ELISA kit (NBP2-66434, Novus Biologicals). Relative light unit (RLU) values were measured immediately after signal development according to the manufacturer's protocol. For all assays, standard curves were generated using recombinant standards provided with each kit, and protein concentrations were calculated based on the corresponding standard curves following the manufacturers' instructions.

Statistical analysis

Gene expression data from public transcriptomic datasets were analyzed using two-tailed unpaired Student's t-tests for pairwise comparisons. For comparisons involving more than two groups, one-way ANOVA followed by Tukey's post hoc test was applied. Analyses were performed in GraphPad Prism, and statistical significance was defined as $p < 0.05$ ($p < 0.05$ *, $p < 0.01$**, $p < 0.001$***).

For plasma proteomic data, group comparisons between IPF and CTD-ILD were conducted using the non-parametric Wilcoxon rank-sum test. The same non-parametric approach (Mann–Whitney U test) was applied for ELISA-based plasma biomarker validation. Differentially expressed proteins were defined as those with FDR–adjusted $p < 0.05$ and carried forward for model development.

To construct predictive models for IPF classification, we applied three approaches: (1) decision tree, (2) random forest, and (3) generalized linear model (GLM) with stepwise selection based on Akaike information criterion (AIC). Model performance was assessed using repeated k-fold cross-validation, with accuracy, sensitivity, specificity, and Cohen's kappa coefficient as evaluation metrics.

For prognostic analyses, Kaplan–Meier survival curves were generated to compare overall survival across biomarker-defined subgroups or model-based predictions, with statistical significance tested by the log-rank method. Cox proportional hazards regression was applied to estimate hazard ratios (HR) and 95% confidence intervals (CI). All statistical analyses were conducted in R (version 4.3.2, R Foundation for Statistical Computing) using the survival (version 3.5-7) and survminer (version 0.4.9) packages.

Results

Patient selection and baseline characteristics

Of 358 patients with ILD discussed in multidisciplinary team meetings, 88 were included for proteomic analysis

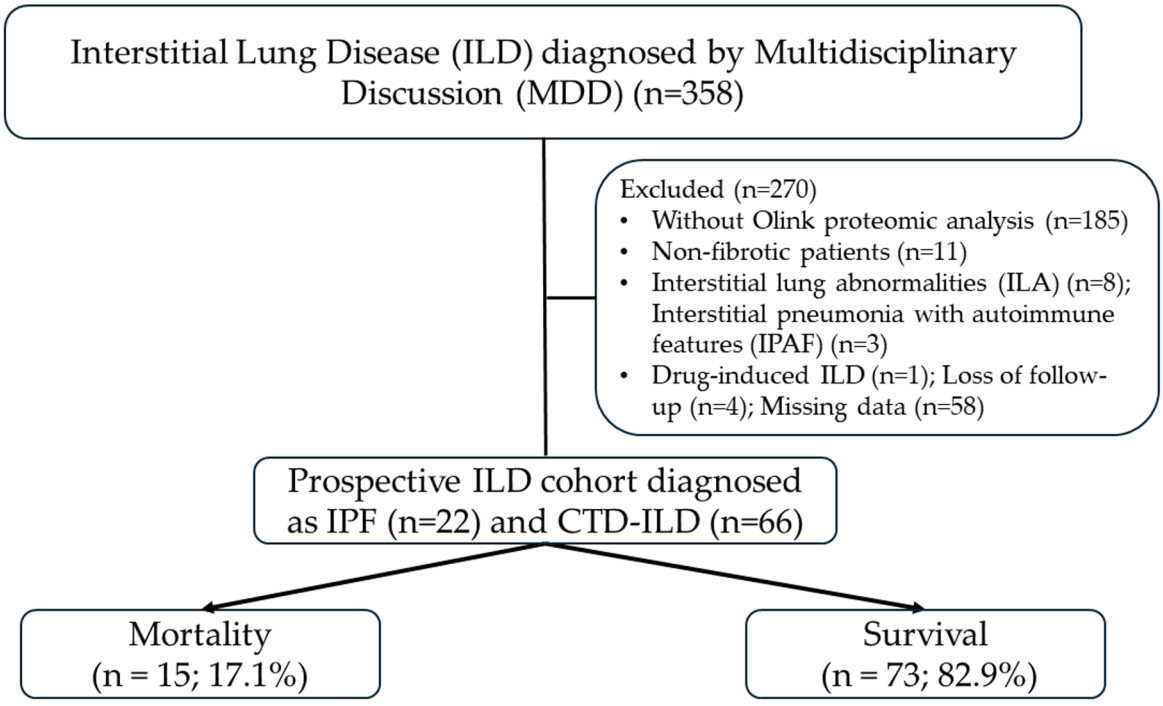


Fig. 1 Patient selection and study flow. A total of 358 interstitial lung disease (ILD) cases were screened, and 88 were included for biomarker analysis (IPF: 22; CTD-ILD: 66). Outcomes focused on diagnostic and prognostic evaluation using plasma proteins. IPF: idiopathic pulmonary fibrosis; CTD-ILD: connective tissue disease–associated ILD

Table 1 Demographic and clinical characteristics of the study participants

	Median	(min, max)
Age (years)	68	(28, 93)
Sex (Male, %)	44 (50%)	
Smoker (n, %)	42	(47.73%)
CCI	2	(0, 15)
SGRQ score	30.5	(0, 89)
SF-36 score	61	(13, 98)
Pulmonary function test		
FVC (% predicted)	76	(26, 153)
FEV1 (% predicted)	76.5	(24, 133)
FEV1/FVC (%)	83	(26, 99)
DLCO (% predicted)	67	(17, 141)
Use of anti-fibrotic drugs (n, %)	32	(36.36%)
Overall mortality (n, %)	15	(17.05%)
Follow-up time (month)	36.5	(1.9, 60)

CCI Charlson Comorbidity Index, *FVC* Forced vital capacity, *FEV* Forced expiratory volume, *DLCO* Diffusion capacity of carbon monoxide

after applying exclusion criteria (IPF, $n = 22$; CTD-ILD, $n = 66$) (Fig. 1). Baseline demographic and clinical characteristics are summarized in Table 1. The study focused on identifying plasma proteomic differences between IPF and CTD-ILD and evaluating their diagnostic and prognostic value.

Differential plasma protein expression

Targeted profiling of 92 inflammation-related proteins (Olink panel) identified 23 differentially expressed proteins between IPF and CTD-ILD (Wilcoxon rank-sum test, FDR-adjusted $p < 0.05$). Of these, 22 were upregulated in IPF, while TWEAK was significantly downregulated. Pairwise correlation heatmaps revealed a distinct co-expression landscape: in IPF, several candidates (MMP-10, FGF-19, ADA, TWEAK, IL-24) showed weak correlations with other proteins, suggesting relative independence and potential specificity for classification (Fig. 2A–B). In contrast, CTD-ILD samples exhibited more diffuse and interconnected protein–protein correlations.

Regulatory and pathway analysis

Transcription factor enrichment highlighted FOXP3, STAT4, and TBX21 as significant regulators [18], implicating T-cell–related immune pathways in IPF (Fig. 2C). Reactome pathway enrichment revealed cytokine signaling, IL-4/IL-13 signaling, and non-canonical NF-κB signaling as dominant functional pathways (Fig. 2D). A PPI network identified central hubs, including matrix metalloproteinases (MMP-1, MMP-9) and chemokines (CXCL5, CXCL8), which are known to drive extracellular matrix remodeling and chronic inflammation (Fig. 2E).

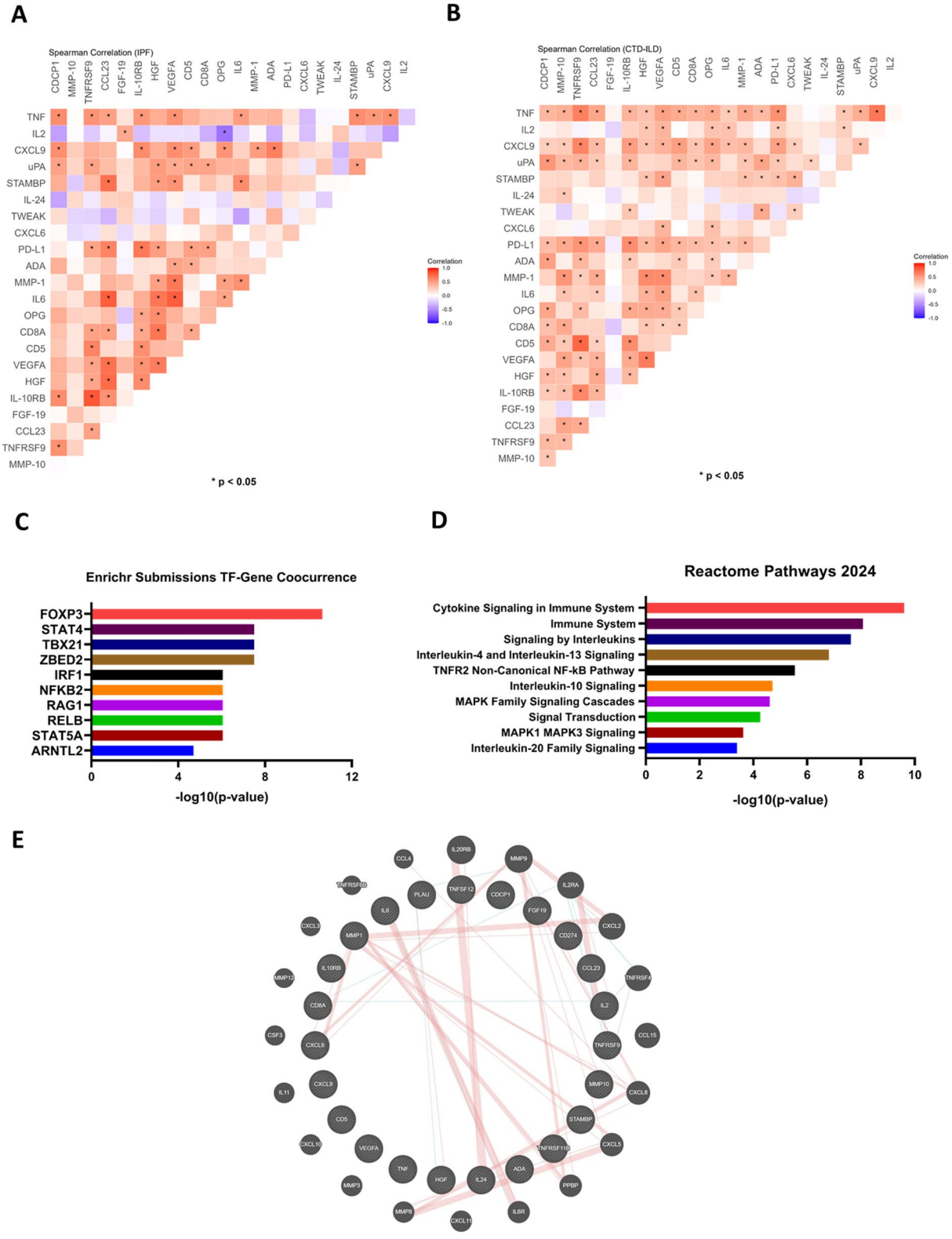


Fig. 2 (See legend on next page.)

(See figure on previous page.)

Fig. 2 Integrative analysis of differentially expressed plasma proteins in IPF and CTD-ILD. **A**–**B** Coexpression analysis of 23 candidate proteins in IPF (**A**) and CTD-ILD (**B**) samples. Red indicates positive correlations and purple indicates negative correlations, with color intensity reflecting correlation strength (Spearman coefficients). **C** Transcription factor–gene co-occurrence analysis using Enrichr. **D** Pathway enrichment analysis with Reactome 2024, highlighting cytokine, interleukin-4/13, and non-canonical NF-κB signaling. **E** Protein–protein interaction network of the 23 proteins, with pink edges denoting physical interactions and light blue edges indicating shared pathways

These findings underscore the immune-fibrotic interplay underlying IPF pathogenesis [19–21].

Diagnostic model development

Three classification strategies were applied: decision tree, random forest, and GLM. In both tree-based approaches, MMP-10 emerged as the primary classifier (cutoff: 10.521), with FGF-19 as a secondary node (cutoff: 8.975) (Supplementary Fig. 1). The GLM incorporating FGF-19, ADA, and TWEAK achieved the highest performance (AUC 0.870; sensitivity 0.97; specificity 0.82), surpassing the decision tree and random forest models (Table 2; Supplementary Fig. 2).

Validation of candidate biomarkers

Among the four selected biomarkers (MMP-10, FGF-19, ADA, and TWEAK), all showed significant differential plasma levels between IPF and CTD-ILD. Notably, given that previous studies have reported reduced circulating FGF-19 in IPF, we further performed ELISA-based validation to confirm the direction of FGF-19 regulation in the present cohort (Supplementary Fig. 3). Plasma levels of MMP-10, FGF-19, and ADA were elevated in IPF, whereas TWEAK was downregulated (Fig. 3). Validation in the public GSE47460 dataset ($n = 582$) showed higher gene expression of MMP-10 and ADA in ILD compared with COPD and controls, while TNFSF12 (encoding gene of TWEAK) was significantly reduced (Fig. 4A, C and D). FGF-19 did not show significant differences (Fig. 4B). These results confirm the disease-specific regulation of TWEAK, MMP-10, and ADA, consistent across independent datasets.

Independent cohort validation of candidate biomarkers at the protein level

To further assess the robustness of the candidate biomarkers at the protein level, we performed ELISA-based validation using an independent cohort comprising 38 patients with CTD-ILD and 22 patients with IPF. Consistent with the discovery proteomics and transcriptomic analyses, plasma levels of MMP-10 and ADA were significantly elevated in IPF, whereas TWEAK levels were significantly reduced, compared with CTD-ILD (Fig. 5A–C). All three biomarkers demonstrated highly significant differences between groups ($p < 0.001$, Mann–Whitney U test). These findings confirm the reproducibility and stability of TWEAK, MMP-10, and ADA at the protein level across independent patient populations, supporting their potential utility as circulating biomarkers for distinguishing IPF from CTD-ILD.

Association with TGF-β1 signaling

To investigate biological regulation, we assessed expression in TGF-β1–stimulated fibroblasts (GSE225982, GSE225159). TNFSF12 was significantly suppressed by TGF-β1 in both primary human lung fibroblasts and a fibroblast cell line (Supplementary Fig. 4). Other candidates (MMP-10, FGF-19, ADA) showed variable or cell type–specific responses, suggesting regulation by alternative sources or pathways.

Prognostic analysis

Kaplan–Meier analysis demonstrated significantly poorer survival in IPF compared with CTD-ILD (log-rank $p < 0.001$) (Fig. 6A). Among candidate biomarkers, elevated MMP-10 was associated with increased mortality (HR = 2.08; 95% CI: 1.22–3.54; $p = 0.007$), whereas higher TWEAK levels strongly predicted improved survival (HR = 0.04; 95% CI: 0.01–0.22; $p < 0.001$) (Fig. 6B). ADA showed a non-significant trend toward improved outcomes, while FGF-19 had no significant association. Collectively, MMP-10 and TWEAK emerged as the most robust prognostic biomarkers in this cohort.

Discussion

In this prospective study, we applied targeted plasma proteomics and integrative machine learning approaches to identify biomarkers that differentiate IPF from CTD-ILD and predict clinical outcomes. Our findings demonstrate that IPF is characterized by distinct proteomic signatures, and that multi-marker predictive models can improve diagnostic accuracy and enable risk stratification. Among the candidate biomarkers, MMP-10 and TWEAK emerged as particularly informative, providing both diagnostic and prognostic value. These observations align with recent work showing that plasma protein panels predict progression or survival across fibrotic ILDs and that machine learning can classify ILD subtypes using proteomic signatures [12, 13, 15].

Traditional single biomarkers such as KL-6, SP-A, SP-D, and MMP-7 have shown only modest performance in IPF diagnosis and prognosis, reflecting the disease's biological heterogeneity and multifactorial pathogenesis [7–10]. By contrast, high-throughput proteomics enables construction of multi-marker models with superior clinical utility [12, 13, 15]. In large, multicenter cohorts [12],

Table 2 Performance comparison of predictive models based on selected plasma proteomic biomarkers

Model	Protemics			Accuracy (95%CI)	Kappa (95%CI)	Sensitivity (95%CI)	Specificity (95%CI)
Decision Tree	MMP-10	FGF-19		0.85 (0.76,0.92)	0.54 (0.33,0.76)	0.50 (0.31,0.69)	0.97 (0.90,0.99)
Random Forest	MMP-10	FGF-19		0.77 (0.67,0.86)	0.35 (0.13,0.58)	0.45 (0.27,0.65)	0.88 (0.78,0.94)
GLM	FGF-19	ADA	TWEAK	0.93 (0.86,0.97)	0.81 (0.67,0.96)	0.97 (0.90,0.99)	0.82 (0.61,0.93)

proteomic signatures predicted progressive fibrosing ILD phenotypes and outperformed clinical models for IPF outcomes; similarly, machine learning classifiers trained on plasma proteomes improved ILD subtype discrimination [15]. Consistent with these reports, our GLM-based model incorporating FGF-19, ADA, and TWEAK achieved excellent diagnostic accuracy (AUC 0.870), outperforming tree-based models. These results underscore the advantage of combining machine learning with proteomic profiling to capture complementary molecular signals and uncover biologically meaningful biomarker panels.

Building on the classification models evaluated in our results, we identified four candidate proteins (FGF-19, MMP-10, ADA, and TWEAK) that consistently contributed to distinguishing IPF from CTD-ILD. Among them, FGF-19 was the most robust feature, being consistently selected by all three algorithms, supporting its potential as a central discriminatory biomarker. FGF-19 exhibited consistently higher plasma levels in IPF patients compared to CTD-ILD patients. As a member of the endocrine FGF subfamily, FGF-19 is involved in several fibrotic and proliferative signaling pathways [22], and its elevated expression may reflect active fibrogenic signaling in IPF [23]. MMP-10 was prioritized in tree-based models and associated with worse prognosis in our data. Prior work localized MMP-10 to alveolar epithelium/macrophages [24], linked higher circulating levels to impaired lung function [24], and identified MMP-10 as a predictor of short-term deterioration and overall survival in IPF—supporting its role as a marker of active tissue remodeling [24].

TWEAK emerged as a clinically relevant biomarker, showing significant downregulation in IPF and association with improved survival. Mechanistically, TWEAK-Fn14 signaling limits fibroblast activation and recruits pro-regenerative macrophages; loss of signaling exacerbates fibrosis in vivo [25, 26]—consistent with our observation that TGF-β1–stimulated fibroblasts suppress TNFSF12 expression. Moreover, TWEAK-Fn14 signaling has been closely associated with IPF disease progression and has been proposed as a potential therapeutic target [25]. Collectively, these data support TWEAK as both a protective biomarker and a potential therapeutic target at the immune–fibrotic interface.

Although ADA did not validate uniformly at the transcript level, its inclusion in the GLM panel is biologically plausible given the adenosine/ADORA2B pathway's role in lung fibrosis [27, 28]. ADA deficiency or adenosine elevation promotes pulmonary fibrosis in vivo [29], and ADORA2B activation induces hyaluronan synthases and glycosaminoglycan accumulation [28]—linking purinergic signaling to matrix remodeling and fibrotic progression.

Although our results provide real-world evidence that plasma proteomic biomarkers can distinguish IPF from CTD-ILD and inform risk stratification—complementing recent multicenter proteomic studies in PF-ILD and IPF—several limitations remain. First, discrepancies with prior reports—for example, studies describing lower circulating FGF-19 in IPF [22]—underscore the need for external validation in larger, multicenter, and ethnically diverse cohorts; broader enrollment across care settings will improve generalizability. Second, although ADA and MMP-10 showed diagnostic and prognostic signals, prospective longitudinal studies are required to establish their value for monitoring disease trajectory and treatment response, including clinically meaningful thresholds and within-patient variability. Third, the mechanistic roles of several candidates, notably FGF-19 and TWEAK, remain incompletely defined; targeted experimental work is needed to delineate their pathways in fibrogenesis and to evaluate their suitability for integration into precision diagnostic and therapeutic strategies.

This prospective study demonstrates that plasma proteomic profiling combined with machine learning can effectively distinguish IPF from CTD-ILD and provide prognostic insights. Among candidate biomarkers, MMP-10 and TWEAK (TNFSF12) emerged as the most informative, with MMP-10 reflecting active tissue remodeling and poor outcomes, and TWEAK serving as a protective marker linked to immune–fibrotic regulation. The multi-marker GLM model incorporating FGF-19, ADA, and TWEAK achieved strong diagnostic accuracy, highlighting the advantage of integrative biomarker strategies over single-analyte approaches. Together, these findings support the clinical utility of plasma proteomics in enhancing early diagnosis, risk stratification, and potentially guiding therapeutic development in fibrotic ILDs.

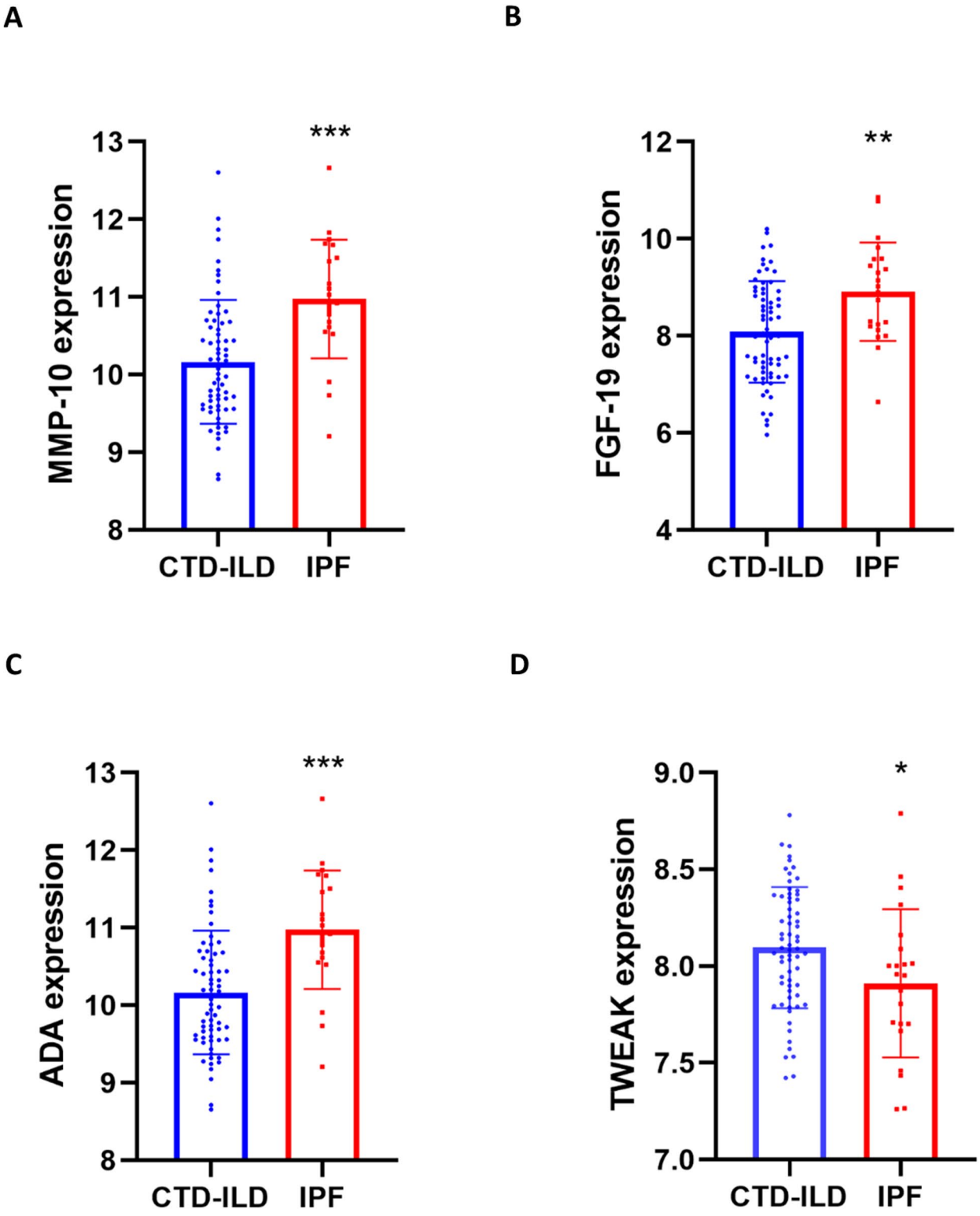


Fig. 3 Differential plasma expression of candidate protein biomarkers in CTD-ILD and IPF. Expression levels of (**A**) MMP-10, (**B**) FGF-19, (**C**) ADA, and (**D**) TWEAK were compared between CTD-ILD and IPF patients. Statistical comparisons were performed using unpaired two-tailed t-tests. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$

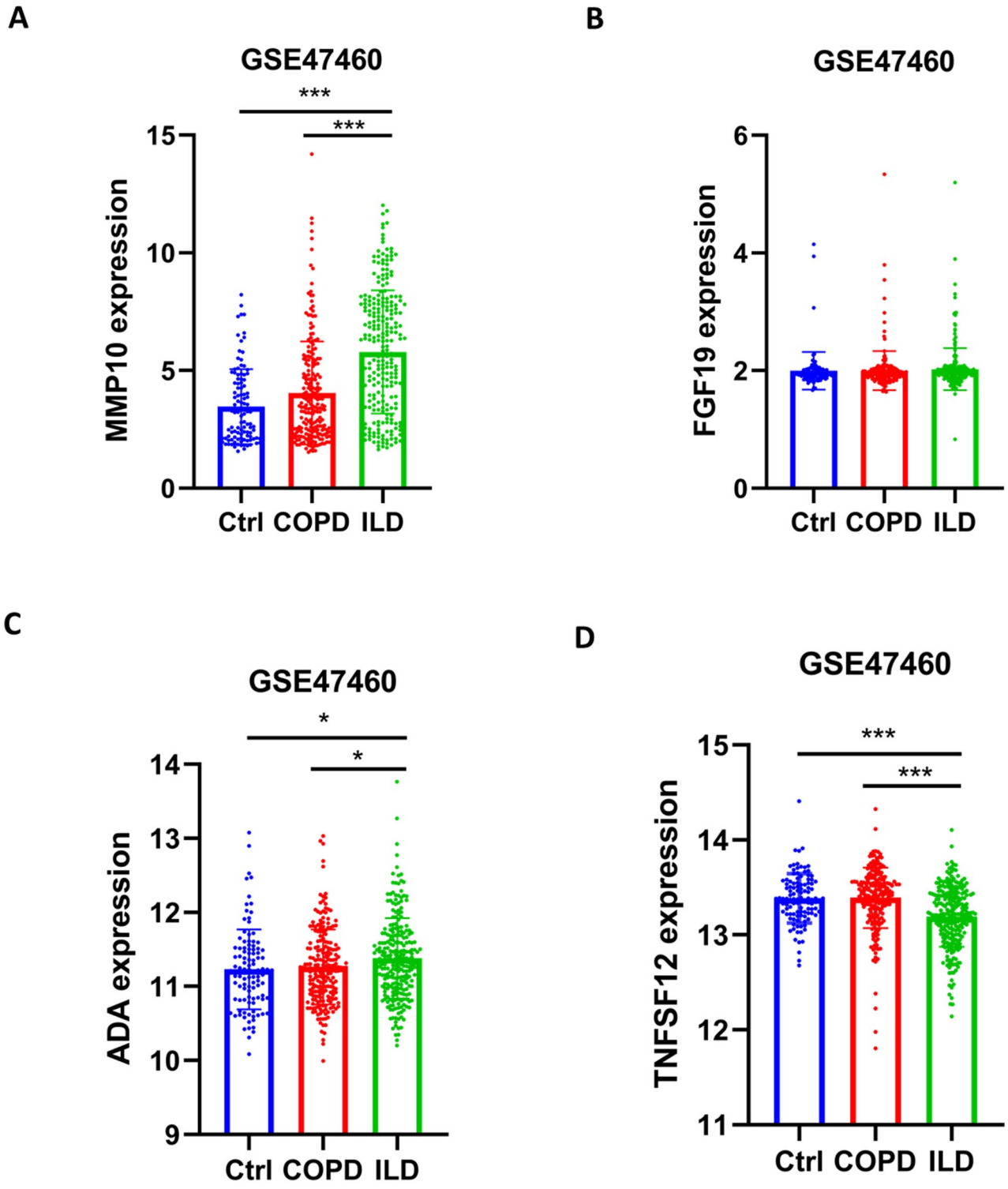


Fig. 4 Gene expression of candidate biomarkers in control, COPD, and ILD samples (GSE47460 dataset). Transcriptomic data ($n = 582$) were analyzed for (**A**) MMP-10, (**B**) FGF-19, (**C**) ADA, and (**D**) TNFSF12 (TWEAK). Groups included control (Ctrl), chronic obstructive pulmonary disease (COPD), and ILD. Each dot represents one sample; bars show mean ± SD. Statistical significance was assessed by one-way ANOVA with Tukey's post hoc test. *$p < 0.05$, ***$p < 0.001$

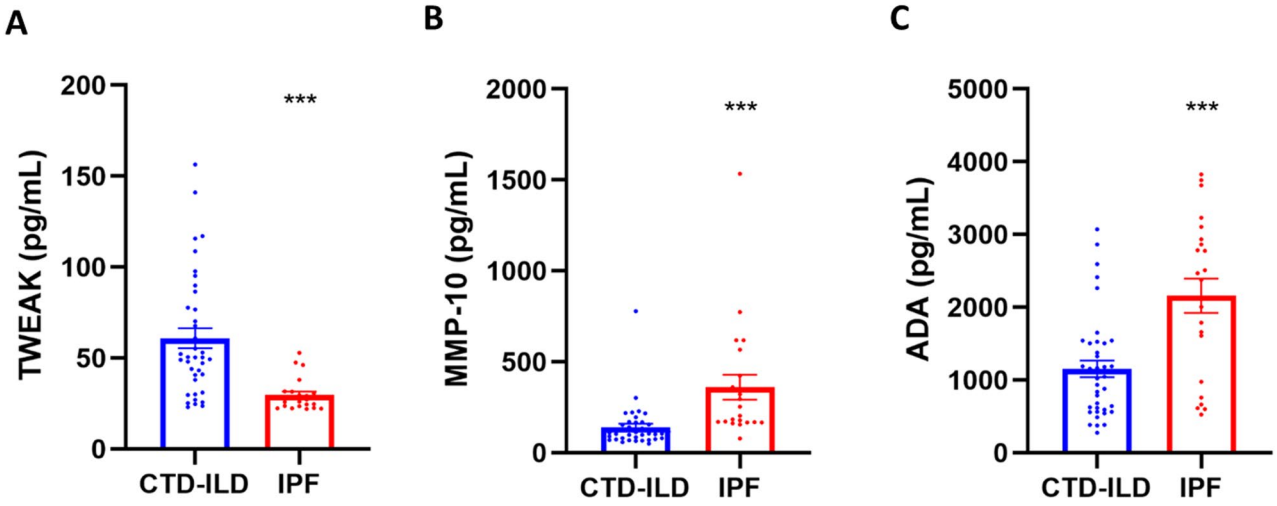
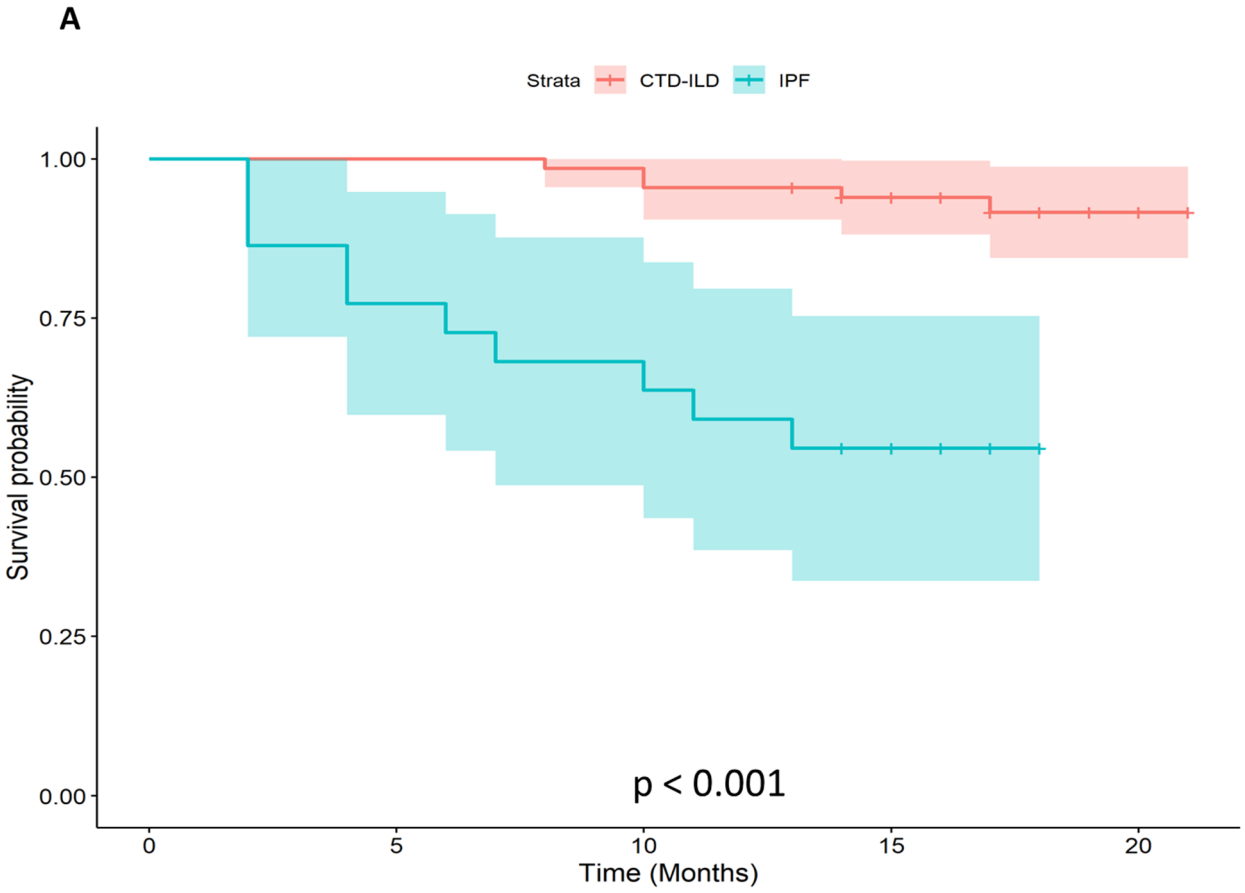


Fig. 5 Independent protein-level validation of candidate biomarkers in CTD-ILD and IPF. **A** Plasma levels of TWEAK, (**B**) MMP-10, and (**C**) ADA were quantified by ELISA in an independent validation cohort (CTD-ILD, $n = 38$; IPF, $n = 22$). Statistical significance was assessed using the Mann–Whitney U test (***$p < 0.001$)



B

Variable	N	Hazard Ratio (95% CI)	p
MMP-10	88	2.08 (1.22, 3.54)	0.007
FGF-19	88	1.21 (0.76, 1.93)	0.430
ADA	88	0.51 (0.18, 1.49)	0.220
TWEAK	88	0.04 (0.01, 0.22)	<0.001

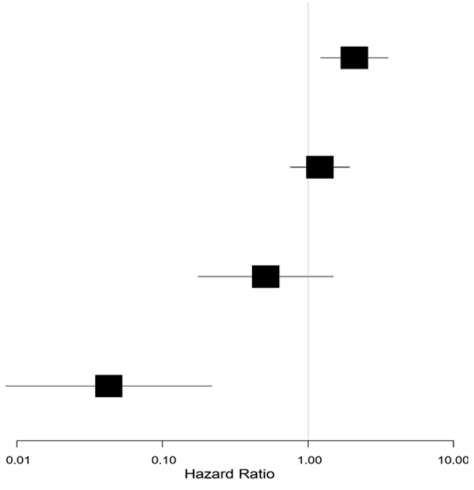


Fig. 6 Prognostic value of disease subgroup and biomarker expression in fibrotic ILD. **A** Kaplan–Meier survival curves comparing IPF ($n = 22$) and CTD-ILD ($n = 66$). The IPF group showed significantly lower survival over 600 days (log-rank $p < 0.001$). Shaded areas denote 95% CI; numbers at risk are shown below. **B** Forest plot of hazard ratios (HR) and 95% CI for selected biomarkers ($n = 88$). TWEAK was strongly protective (HR = 0.04, $p < 0.001$)

Abbreviations

IPF	Idiopathic pulmonary fibrosis
ILD	Interstitial lung disease
CTD-ILD	Connective tissue disease–associated ILD
GLM	Generalized linear modeling
TGF-β	Transforming growth factor-β
IL-6	Interleukin-6
TNF-α	Tumor necrosis factor-α
PDGF	Platelet-derived growth factor
SP-A	Surfactant protein-A
SP-D	Surfactant protein-D
MMP-7	Matrix metalloproteinase-7
PEA	Proximity extension assay
NPX	Normalized protein expression
FDR	False discovery rate
TF	Transcription factor
PPI	Protein–protein interaction
AIC	Akaike information criterion
HR	Hazard ratios
CI	Confidence intervals

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12931-026-03596-4.

Supplementary Material 1.

Acknowledgements

We thank the Biostatistics Task Force of Taichung Veterans General Hospital for statistical support and acknowledge funding from the Ministry of Health and Welfare, Taiwan (MOHW114-TDU-B-211-124008). The technical support provided by the Instrument Center of Department of Medical Research of Taichung Veterans General Hospital is gratefully acknowledged. Editorial assistance was provided by ChatGPT (OpenAI) to improve grammar and clarity; the authors reviewed all suggestions and take full responsibility for the final content.

Authors' contributions

Conceptualization: CSW, PKF; Methodology: CSW, LYY, PKF; Formal analysis: DWL, YHC; Investigation: YHY, LYY, PKF; Data curation: CSW, DWL; Visualization: CSW, LYY, PKF; Writing—original draft: CSW, LYY, PKF; Writing—review & editing: all authors; Funding acquisition: PKF; Supervision: PKF.

Funding

This study was supported by the Taichung Veterans General Hospital (TCVGH-1143910B; TVGH-1143916 C; TCVGH-1143909D; TCVGH-1114401 C) and the National Science and Technology Council of Taiwan (NSTC 112-2314-B-075 A-003-MY3).

Data availability

The plasma proteomic datasets generated and analyzed during the current study are available in the Figshare repository at https://doi.org/10.6084/m9.figshare.31061860. Additional deidentified participant data, data dictionary, protocol, and statistical analysis plan will be available upon reasonable request to the corresponding author after publication for 36 months. Access will be granted to qualified researchers for noncommercial, hypothesis-driven analyses under a data-use agreement.

Declarations

Ethics approval and consent to participate

We prospectively enrolled consecutive patients at Taichung Veterans General Hospital, Taiwan, from January 1 to December 31, 2024. The study was approved by the Institutional Review Board (CE18325B, CG24056C) and registered at ClinicalTrials.gov (NCT06476470). All eligible, evaluable cases were included. The study adhered to the Declaration of Helsinki and Good Clinical Practice guidelines. Written informed consent was obtained from all participants prior to enrollment. Patient privacy and confidentiality were strictly maintained in accordance with institutional guidelines.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

[1]Department of Medical Research, Taichung Veterans General Hospital, 1650 Taiwan Boulevard, Sect. 4, Taichung 407219, Taiwan
[2]Department of Health Business Administration, HungKuang University, Taichung 433304, Taiwan
[3]Division of Clinical Research, Department of Medical Research, Taichung Veterans General Hospital, Taichung 407219, Taiwan
[4]Integrated Care Center of Interstitial Lung Disease, Taichung Veterans General Hospital, Taichung 407219, Taiwan
[5]Department of Post-Baccalaureate Medicine, College of Medicine, National Chung Hsing University, Taichung 402202, Taiwan

Received: 20 September 2025 / Accepted: 19 February 2026
Published online: 06 March 2026

References

1. Martinez FJ, Collard HR, Pardo A, Raghu G, Richeldi L, Selman M, Swigris JJ, Taniguchi H, Wells AU. Idiopathic pulmonary fibrosis. Nat Rev Dis Primers. 2017;3:17074.
2. Bartold K, Iskierko Z, Sharma PS, Lin HY, Kutner W. Idiopathic pulmonary fibrosis (IPF): Diagnostic routes using novel biomarkers. Biomed J. 2024;47:100729.
3. Heukels P, Moor CC, von der Thusen JH, Wijsenbeek MS, Kool M. Inflammation and immunity in IPF pathogenesis and treatment. Respir Med. 2019;147:79–91.
4. Vasakova M, Sterclova M, Kolesar L, Slavcev A, Pohunek P, Sulc J, Striz I. Cytokine gene polymorphisms and BALF cytokine levels in interstitial lung diseases. Respir Med. 2009;103:773–9.
5. Inui N, Sakai S, Kitagawa M. Molecular Pathogenesis of Pulmonary Fibrosis, with Focus on Pathways Related to TGF-beta and the Ubiquitin-Proteasome Pathway. Int J Mol Sci. 2021;22:6107.
6. Saha P, Talwar P. Idiopathic pulmonary fibrosis (IPF): disease pathophysiology, targets, and potential therapeutic interventions. Mol Cell Biochem. 2024;479:2181–94.
7. Zheng Z, Peng F, Zhou Y. Biomarkers in idiopathic pulmonary fibrosis: Current insight and future direction. Chin Med J Pulm Crit Care Med. 2024;2:72–9.
8. Majewski S, Szewczyk K, Zal A, Bialas AJ, Milkowska-Dymanowska J, Piotrowski WJ. Serial Measurements of Circulating KL-6, SP-D, MMP-7, CA19-9, CA-125, CCL18, and Periostin in Patients with Idiopathic Pulmonary Fibrosis Receiving Antifibrotic Therapy: An Exploratory Study. J Clin Med. 2021;10:3864.
9. He X, Ji J, Zheng D, Luo Z, Luo L, Guo L. Serum surfactant protein D as a significant biomarker for predicting occurrence, progression, acute exacerbation, and mortality in interstitial lung disease: a systematic review and meta-analysis. Front Immunol. 2025;16:1450798.
10. Wang L, Huang J, Tan G, Zhang Y, Xia M, Kang Z, Xiao Y, Lei M. Trends and hotspots in pulmonary fibrosis biomarker research: a bibliometric analysis. Front Med (Lausanne). 2025;12:1541364.
11. Wang L, Zhu M, Li Y, Yan P, Li Z, Chen X, Yang J, Pan X, Zhao H, Wang S, et al. Serum Proteomics Identifies Biomarkers Associated With the Pathogenesis of Idiopathic Pulmonary Fibrosis. Mol Cell Proteom. 2023;22:100524.
12. Oldham JM, Huang Y, Bose S, Ma SF, Kim JS, Schwab A, Ting C, Mou K, Lee CT, Adegunsoye A, et al. Proteomic Biomarkers of Survival in Idiopathic Pulmonary Fibrosis. Am J Respir Crit Care Med. 2024;209:1111–20.
13. Bowman WS, Newton CA, Linderholm AL, Neely ML, Pugashetti JV, Kaul B, Vo V, Echt GA, Leon W, Shah RJ, et al. Proteomic biomarkers of progressive fibrosing interstitial lung disease: a multicentre cohort analysis. Lancet Respir Med. 2022;10:593–602.
14. Maddali MV, Moore AR, Sinha P, Newton CA, Kim JS, Adegunsoye A, Ma SF, Strek ME, Chen CH, Linderholm AL, et al. Molecular Endotypes of Idiopathic Pulmonary Fibrosis: A Latent Class Analysis of Two Multicenter Observational Cohorts. Am J Respir Crit Care Med. 2024;210:455–64.
15. Huang Y, Ma SF, Oldham JM, Adegunsoye A, Zhu D, Murray S, Kim JS, Bonham C, Strickland E, Linderholm AL, et al. Machine Learning of Plasma Proteomics

Classifies Diagnosis of Interstitial Lung Disease. Am J Respir Crit Care Med. 2024;210:444–54.
16. Khan T, Dasgupta S, Ghosh N, Chaudhury K. Proteomics in idiopathic pulmonary fibrosis: the quest for biomarkers. Mol Omics. 2021;17:43–58.
17. Hou J, Yang Y, Han X. Machine Learning and Single-Cell Analysis Identify Molecular Features of IPF-Associated Fibroblast Subtypes and Their Implications on IPF Prognosis. Int J Mol Sci. 2023;25:94.
18. Artham S, Verma A, Alwhaibi A, Adil MS, Manicassamy S, Munn DH, Somanath PR. Delayed Akt suppression in the lipopolysaccharide-induced acute lung injury promotes resolution that is associated with enhanced effector regulatory T cells. Am J Physiol Lung Cell Mol Physiol. 2020;318:L750–61.
19. de Almeida LGN, Thode H, Eslambolchi Y, Chopra S, Young D, Gill S, Devel L, Dufour A. Matrix Metalloproteinases: From Molecular Mechanisms to Physiology, Pathophysiology, and Pharmacology. Pharmacol Rev. 2022;74:712–68.
20. Traber KE, Hilliard KL, Allen E, Wasserman GA, Yamamoto K, Jones MR, Mizgerd JP, Quinton LJ. Induction of STAT3-Dependent CXCL5 Expression and Neutrophil Recruitment by Oncostatin-M during Pneumonia. Am J Respir Cell Mol Biol. 2015;53:479–88.
21. Richter AG, Perkins GD, Chavda A, Sapey E, Harper L, Thickett DR. Neutrophil chemotaxis in granulomatosis with polyangiitis (Wegener's) and idiopathic pulmonary fibrosis. Eur Respir J. 2011;38:1081–8.
22. Justet A, Ghanem M, Boghanim T, Hachem M, Vasarmidi E, Jaillet M, Vadel A, Joannes A, Mordant P, Bonniaud P, et al. FGF19 Is Downregulated in Idiopathic Pulmonary Fibrosis and Inhibits Lung Fibrosis in Mice. Am J Respir Cell Mol Biol. 2022;67:173–87.
23. Ghanem M, Justet A, Jaillet M, Vasarmidi E, Boghanim T, Hachem M, Vadel A, Joannes A, Mordant P, Balayev A, et al. Identification of FGFR4 as a regulator of myofibroblast differentiation in pulmonary fibrosis. Am J Physiol Lung Cell Mol Physiol. 2024;327:L818–30.
24. Sokai A, Handa T, Tanizawa K, Oga T, Uno K, Tsuruyama T, Kubo T, Ikezoe K, Nakatsuka Y, Tanimura K, et al. Matrix metalloproteinase-10: a novel biomarker for idiopathic pulmonary fibrosis. Respir Res. 2015;16:120.
25. Liu L, Wu P, Wei Y, Lu M, Ge H, Wang P, Sun J, Horng T, Liu X, Shen X, et al. TWEAK-Fn14 signaling protects mice from pulmonary fibrosis by inhibiting fibroblast activation and recruiting pro-regenerative macrophages. Cell Rep. 2025;44:115220.
26. Liu X, Wang X, Yang Q, Luo L, Liu Z, Ren X, Lei K, Li S, Xie Z, Zheng G, et al. Th17 Cells Secrete TWEAK to Trigger Epithelial-Mesenchymal Transition and Promote Colorectal Cancer Liver Metastasis. Cancer Res. 2024;84:1352–71.
27. Zhou Y, Murthy JN, Zeng D, Belardinelli L, Blackburn MR. Alterations in adenosine metabolism and signaling in patients with chronic obstructive pulmonary disease and idiopathic pulmonary fibrosis. PLoS ONE. 2010;5:e9224.
28. Collum SD, Molina JG, Hanmandlu A, Bi W, Pedroza M, Mertens TCJ, Wareing N, Wei W, Wilson C, Sun W, et al. Adenosine and hyaluronan promote lung fibrosis and pulmonary hypertension in combined pulmonary fibrosis and emphysema. Dis Model Mech. 2019;12:dmm038711.
29. Chunn JL, Molina JG, Mi T, Xia Y, Kellems RE, Blackburn MR. Adenosine-dependent pulmonary fibrosis in adenosine deaminase-deficient mice. J Immunol. 2005;175:1937–46.

Publisher's note