ORIGINAL ARTICLE

Transcriptomic profiling of collagen gene expression in lung fibroblasts from patients with idiopathic pulmonary fibrosis

Eun-Jeong Seo[1], Seung-Lee Park[1], Min-Kyung Kim[1], Jae-Sung Choi[2], Hun-Gyu Hwang[3], Jung-Hyun Kim[4], Hun-Soo Chang[1], Choon-Sik Park[5], Jong-Uk Lee[1,5]

[1]Department of Interdisciplinary Program in Biomedical Science Major, Graduate School, Soonchunhyang University, Asan, Republic of Korea; [2]Division of Respiratory Medicine, Department of Internal Medicine, Soonchunhyang University Cheonan Hospital, Cheonan, Republic of Korea; [3]Soonchunhyang University Bucheon Hospital, Bucheon, Republic of Korea; [4]Division of Allergy and Clinical Immunology, Department of Internal Medicine, Soonchunhyang University Hospital, Seoul, Republic of Korea; [5]Department of Internal Medicine, Soonchunhyang University Bucheon Hospital, Bucheon, Republic of Korea

ABSTRACT

Background and aim: Idiopathic Pulmonary Fibrosis (IPF) is a progressive interstitial lung disease characterized by aberrant extracellular matrix remodeling. While type I and III collagens are known contributors, the broader transcriptional landscape of diverse collagen genes in IPF fibroblasts and their potential associations with clinical outcomes remain incompletely understood. This study aimed to comprehensively characterize gene expression of collagen genes in IPF fibroblasts and to explore its relationship with patient survival.

Methods: RNA sequencing was conducted on primary lung fibroblasts from 33 IPF patients and 10 non-fibrotic controls. Differential expression of collagen genes was analyzed using EdgeR's exact test, with significance defined as false discovery rate-adjusted $p < 0.05$. Association with mortality was assessed using Cox regression analysis.

Results: Among 17 collagen genes, 14 genes — including *COL1A1, COL1A2, COL3A1, COL4A4, COL4A5, COL4A6, COL5A2, COL6A6, COL8A1, COL11A1, COL14A1, COL18A1, COL24A1,* and *COL27A1* — were significantly upregulated in IPF fibroblasts, while *COL5A3, COL13A1*, and *COL16A1* were downregulated. *COL3A1, COL4A*4, and *COL24A1* were associated with reduced survival in both Cox regression analysis and Kaplan–Meier plots. Among them, *COL4A4* remained association with mortality in multivariable Cox model.

Received: 14 October 2025 | Accepted: 5 January 2026

Correspondence: Jong-Uk Lee, PhD / Department of Internal Medicine, Soonchunhyang University Bucheon Hospital, Bucheon, Republic of Korea / E-mail: lajoal@nate.com
ORCID: 0000-0002-0232-5381

Conclusions: This study delineates fibroblast-specific dysregulation across multiple collagen families in IPF, providing mechanistic insight into extracellular matrix remodeling. Although *COL4A4* remained associated with mortality in multivariable Cox models, this finding should be interpreted as exploratory, given the limited number of events and the absence of experimental validation. Overall, our results highlight collagen transcript signatures as hypothesis-generating markers that warrant further validation to clarify their potential clinical relevance in IPF.

Key words: idiopathic pulmonary fibrosis, collagen, fibroblast, extracellular matrix, rna-seq

Introduction

Idiopathic Pulmonary Fibrosis (IPF) is a chronic, progressive interstitial lung disease of unknown cause, characterized by irreversible scarring of lung tissue and associated with high morbidity and mortality (1, 2). A hallmark of IPF is the aberrant accumulation of extracellular matrix (ECM) components, which disrupts alveolar architecture and impairs gas exchange (3). This fibrotic remodeling is driven by persistently activated fibroblasts and myofibroblasts, which secrete excessive amounts of ECM proteins, including fibronectin and multiple collagen genes (3-5). Collagen, the most abundant ECM protein, plays a central role in maintaining tissue structure and mechanical integrity. To date, 28 collagen types (Type I–XXVIII) have been identified, all sharing a characteristic triple-helical Gly-X-Y repeat structure, where X and Y are often proline and 4-hydroxyproline, respectively (6, 7). These types are encoded by distinct collagen genes, often composed of multiple genes (e.g., *COL1A1, COL1A2* for Type I collagen). In humans, at least 40 collagen genes are known, many of which correspond to or compose the 28 collagen types (6). These collagens are classified into several families based on their structure and function: fibril-forming (types I, II, III, V, XI), fibril-associated collagens with interrupted triple helices (types IX, XII, XIV), network-forming (types IV, VIII, X), anchoring fibrils (type VII), transmembrane (types XIII, XVII, XXIII, XXV), and basement membrane collagens (types IV, XVIII). Among these, type I and type III collagens are the most abundant fibrillar collagens in the lung and are critical for the biomechanical properties of bronchi, alveolar septa, and interstitial connective tissue (8). In IPF, these collagens are significantly upregulated, contributing to matrix stiffening and loss of lung compliance. Their accumulation also disrupts epithelial–mesenchymal signaling and promotes a profibrotic feedback loop that accelerates disease progression (1). However, the broader transcriptional landscape of other collagen genes in IPF fibroblasts and their relationship to clinical outcomes remain poorly characterized except type I, III, VI (1, 9). To address these limitations, we performed total RNA sequencing on primary lung fibroblasts isolated from IPF patients and non-fibrotic controls. Our analysis aimed to comprehensively profile the expression of collagen genes, encompassing both well-characterized fibrillar collagens and less-explored collagen genes with potential clinical relevance. We further examined the relationship between these transcriptional patterns and patient survival to explore candidate collagen genes associated with clinical outcomes.

Materials and Methods

Fibroblast culture

Primary lung fibroblasts were isolated from surgical biopsy specimens from 33 patients with histologically confirmed usual interstitial pneumonia, and from histologically normal lung tissue of 10 patients who underwent resection for stage I or II lung cancer. Fibroblast culture protocols followed previously published methods (10). Briefly, tissue samples were minced and cultured in 150 cm² flasks containing tissue culture medium composed of Dulbecco's Modified

Eagle Medium (Lonza Walkersville, MD, USA), supplemented with 10% fetal bovine serum (Thermo Fisher Scientific, Rockford, IL, USA), 2 mmol/L glutamine, and 1% penicillin-streptomycin-amphotericin B (Lonza, Basel, Switzerland). Cultures were maintained at 37°C in a 5% CO_2 incubator and serially passaged to obtain a morphologically homogeneous population of adherent fibroblasts.

Total RNA sequencing data processing

We used GSE301181 containing Total RNA sequencing data of primary lung fibroblasts from 33 patients with IPF and 10 controls whose samples were obtained from histologically normal lung tissue adjacent to resected tumors in lung cancer surgery. Comprehensive clinical evaluation, including medical history, chest radiography, high-resolution computed tomography, and pulmonary function tests were obtained through the biobank of Soonchunhyang University Hospital, Bucheon, Korea (KBN4-A06). The study protocol was approved by the hospital's Institutional Review Board (IRB Nos. SCHCA-IRB-2018-10-034 and 201910-BR-058). Collagen vascular diseases were excluded on the basis of clinical or serological evidence. IPF diagnoses were established according to the 2011 and 2018 ATS/ERS/JRS/ALAT guidelines (11, 12).

Differential expression analysis

The analysis was performed using the Biomedical Analysis Workspace (hyper.schedulerju.com), a clinical bioinformatics analysis platform, with the edgeR package. Of a total of 46,427 genes, genes with fewer than 10 read counts (n = 31,175) were excluded, leaving 15,252 genes, including 30 collagen genes. Count data were normalized using the trimmed mean of M-values method. Differentially expressed genes were identified using the exact test implemented in edgeR, with significance defined as an FDR-adjusted p-value < 0.05. Gene expression values were reported as log2 counts per million.

Statistical analysis

Analyses were performed using the Biomedical Analysis Workspace (hyper.schedulerju.com), a clinical statistical analysis platform built on SciPy, statsmodels, lifelines, and scikit-learn. The normality of distributions was assessed using the Shapiro-Wilk test. Continuous variables are presented as the mean ± standard deviation or median with interquartile range, as appropriate. Between-group comparisons were performed using independent t-tests or Mann-Whitney U tests for continuous variables and chi-square tests for categorical variables. Receiver operating characteristic curves and the Youden index were used to determine optimal cut-off values. Kaplan-Meier survival analysis and Cox proportional hazards regression analyses were performed in a cohort of 28 patients with IPF who had a follow-up period of more than one year. A two-tailed p-value < 0.05 was considered statistically significant.

Results

Patient characteristics

The clinical characteristics of the study population are presented in Table 1. A total of 43 subjects were enrolled, including 33 patients with IPF and 10 controls. Although the mean age was higher in the IPF group compared to controls, the difference did not reach statistical significance (p = 0.081). The distribution of sex (p = 0.761) and smoking status (p = 0.489) were also comparable between groups. In contrast, pulmonary function tests demonstrated marked impairment among IPF patients. The median predicted forced vital capacity was significantly reduced in IPF

Table 1. Clinical characteristics of the study subjects

Variable	Control	IPF	*p*-value
No.	10	33	-
Age (year)	58.4±9.2	64.3±9.2	0.081
Sex (male/female)	4/6	15/18	0.761
Smoke (NS/ES/SM)	7/1/2	16/6/11	0.489
FVC (% pred.)	89.5±18.4	66.4±30	0.01
DL_{CO} (% pred.)	81.1±32.8	42.8±34	0.002
Death/Survival	2/8	13/20	0.26

Values are presented as mean ± standard error of the mean. *Abbreviations:* IPF: idiopathic pulmonary fibrosis; NS/ES/SM: non-smokers, ex-smokers, smokers; FVC: forced vital capacity; DL_{CO}: diffusion capacity of the lung for carbon monoxide

group relative to controls (p = 0.010), and the median predicted diffusion capacity for carbon monoxide was likewise substantially lower (p = 0.002).

Differential expression of collagen genes in human lung-derived fibroblasts

Based on total RNA sequencing of lung fibroblasts, approximately 15,000 genes were detected, including 30 collagen genes. Of these, 17 passed the exact test with multiple correction using FDR (p < 0.05), showing significant changes in IPF compared to control fibroblasts (Figure 1, Table S1). Specifically, 14 genes — including *COL1A1*, *COL1A2*, *COL3A1*, *COL4A4*, *COL4A5*, *COL4A6*, *COL5A2*, *COL6A6*, *COL8A1*, *COL11A1*, *COL14A1*, *COL18A1*, *COL24A1*, and *COL27A1* — were significantly upregulated in IPF fibroblasts, whereas *COL5A3*, *COL13A1*, and *COL16A1* were more highly expressed in control-derived fibroblasts. Among these, *COL1A2* exhibited the highest expression level, followed by *COL3A1* (Figure 2, Table S1).

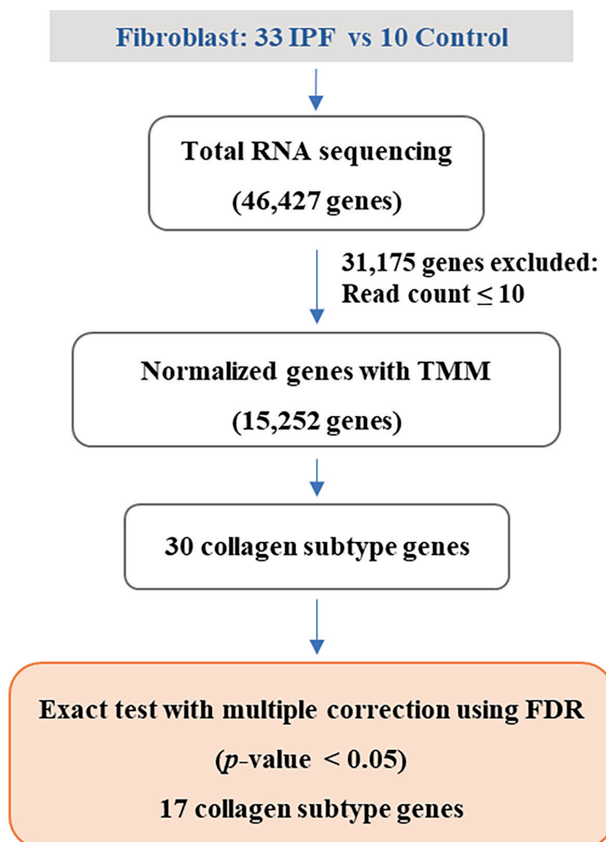

Figure 1. Workflow for selection of 17 collagen genes from lung fibroblast total RNA sequencing data. *Abbreviations:* IPF: idiopathic pulmonary fibrosis; TMM: trimmed mean of M-values; FDR: false discovery rate.

Survival analysis based on collagen gene expression in IPF patients

Survival analyses were performed using continuous Cox proportional hazards models. In the initial unadjusted analyses, *COL3A1*, *COL4A4*, *COL6A6*, and *COL24A1* showed association with overall mortality, although none remained significant after FDR correction (Table 2). These four genes were therefore included in the subsequent multivariable analyses. To assess whether these collagen genes were associated with mortality independently of clinical factors, a backward selection multivariable Cox model including age, FVC, DL_{CO}, and four candidate collagen genes - *COL3A1*, *COL4A4*, *COL6A6*, and *COL24A1*- was constructed. Within this model, *COL4A4* remained significantly associated with mortality (HR = 2.39, 95% CI: 1.25–4.58, p = 0.009, Table 2). Kaplan–Meier analyses were generated solely for exploratory visualization and were not used for statistical inference. Youden index–derived cut-off values were generated for the same four collagen genes (Table S2), and the corresponding Kaplan–Meier curves are presented in Figure S1 as exploratory illustrations. Although cut off values were significant in *COL3A1* and *COL24A1*, all of the cut-off values were selected for Kaplan–Meier survival plots. Kaplan–Meier survival curves showed that in *COL3A1*, *COL4A4*, and *COL24A1*, statistical significance was found via the log-rank test.

Discussion

In this study, we conducted transcriptomic evaluations of collagen gene expression in primary lung fibroblasts from patients with IPF and healthy controls. We detected 17 collagen genes that were differentially expressed, with 14 showing increased and three showing decreased expression in IPF fibroblasts. Cox regression analysis revealed associations between *COL3A1*, *COL4A4*, *COL6A6*, *COL24A1* and mortality. *COL4A4* showed a consistent association with

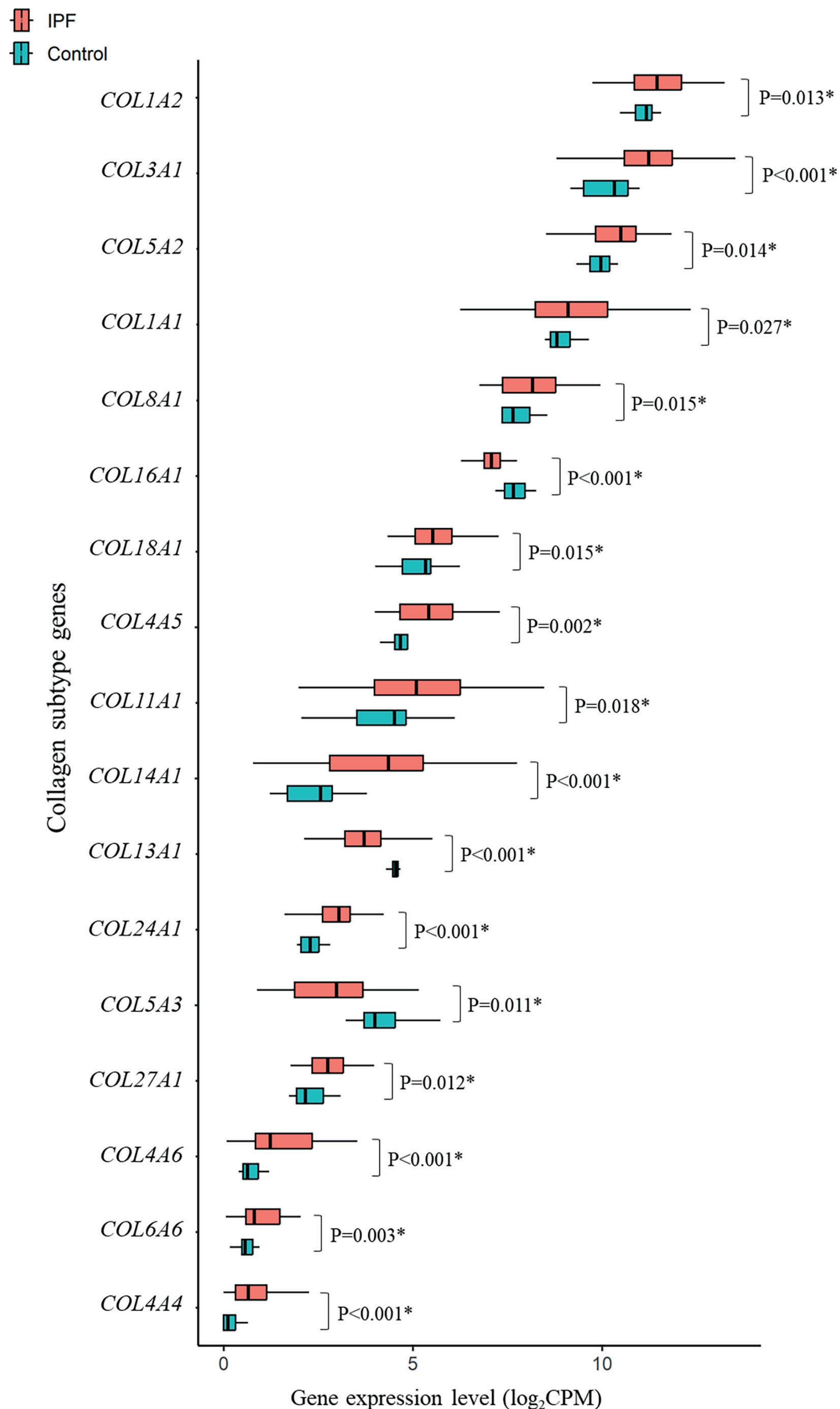

Figure 2. Differential expression of 17 collagen subtype genes between 33 IPF and 10 control fibroblasts. Box plots depicting expression levels of 17 collagen subtype genes in lung fibroblasts from IPF patients and controls. Each box represents the median and interquartile range. * Compared with the control, $p < 0.05$. *Abbreviations:* IPF: idiopathic pulmonary fibrosis, CPM: count per million.

Table 2. Associations between collagen gene expression and mortality in 28 IPF patients

Univariate analysis				
Parameter	**HR**	**95% CI**	***p*-value**	**FDR**
Age	1.06	0.95–1.17	0.296	0.724
Sex (male/female)	0.65	0.18–2.28	0.501	0.861
Smoke (NS/ES/SM)	0.84	0.43–1.64	0.615	0.861
FVC (% pred.)	1.00	0.98–1.02	0.791	0.861
DL_{CO} (% pred.)	1.00	0.98–1.02	0.777	0.861
COL1A1	1.25	0.79–1.98	0.347	0.740
COL1A2	1.92	0.86–4.29	0.114	0.358
*COL3A1**	1.85	1.06–3.22	0.032	0.176
*COL4A4**	2.39	1.25–4.58	0.009	0.176
COL4A5	1.11	0.62–1.98	0.729	0.861
COL4A6	1.15	0.76–1.72	0.510	0.861
COL5A2	2.06	0.91–4.66	0.086	0.319
COL5A3	1.05	0.63–1.73	0.861	0.861
*COL6A6**	1.77	1.08–2.89	0.023	0.176
COL8A1	0.92	0.47–1.80	0.814	0.861
COL11A1	1.20	0.81–1.78	0.370	0.740
COL13A1	0.94	0.51–1.73	0.850	0.861
COL14A1	1.34	0.96–1.86	0.087	0.319
COL16A1	1.62	0.32–8.34	0.565	0.861
COL18A1	1.06	0.57–1.98	0.858	0.861
*COL24A1**	3.94	1.15–13.54	0.030	0.176
COL27A1	0.49	0.16–1.53	0.222	0.611
Backward multivariate analysis				
Parameters	**HR**	**95% CI**	***p*-value**	
*COL4A4**	2.39	1.25–4.58	0.009	

Univariate Cox proportional hazards regression analyses were performed to evaluate the association between individual collagen gene expression levels and all-cause mortality in IPF patients over a follow-up period of 1 to 10 years. Additional univariate analyses were conducted using clinical parameters. Multivariate Cox regression analysis using a backward selection method was conducted including four significant genes from the univariate analysis as well as clinical data such as age, FVC, and DL_{CO}. * $p < 0.05$. *Abbreviations:* IPF: idiopathic pulmonary fibrosis; HR: hazard ratio; CI: confidence interval; NS/ES/SM: non-smokers, ex-smokers, smokers; FVC: forced vital capacity; DL_{CO}: diffusion capacity of the lung for carbon monoxide; FDR: false discovery rate

mortality after adjustment for age and lung function parameters. As expected, *COL1A1, COL1A2*, and *COL3A1* — encoding major fibrillar collagens — were markedly upregulated, consistent with their known roles in ECM accumulation and fibrosis progression (3, 4). *COL5A2*, a minor fibrillar collagen, was also significantly upregulated in IPF fibroblasts. It co-assembles with type I collagen and regulates fibril diameter and spacing (13). In single-cell transcriptomic datasets of bleomycin-induced fibrotic lung tissue, *COL5A2* was among the genes significantly upregulated in the activated fibroblast population (14). These findings highlight that activated fibroblasts overproduce fibrillar collagens—especially type I, III, and V collagens—which stiffen the extracellular matrix, reduce alveolar flexibility, and promote profibrotic cellular response,

disrupting normal lung architecture (1, 3-5). In addition to these fibrillar collagens, our data highlight elevated expression of other collagen genes encoding a basement membrane collagen, and its upregulation may reflect extensive remodeling of the alveolar interstitium, which has not traditionally emphasized in IPF. Our analysis revealed the upregulation of *COL4A4*, *COL4A5*, and *COL4A6* subunits of type IV collagen typically associated with basement membranes. While the role of type IV collagen in IPF remains less well-defined, emerging evidence suggests that abnormal basement membrane remodeling may influence epithelial integrity and facilitate fibroblast invasion into the alveolar space (15, 16). Taken together, the upregulation of both basement membrane–associated and fibrillar collagens supports a broader disruption of ECM architecture that extends beyond classical interstitial collagen remodeling. Furthermore, *COL11A1, COL14A1*, *COL18A1* and *COL27A1* were significantly elevated in IPF fibroblasts. *COL11A1* and *COL14A1* have been implicated in regulating collagen fiber formation and interfibrillar interactions (17, 18), while *COL18A1* encodes endostatin, a proteolytic fragment with anti-angiogenic properties (19). *COL27A1* is expressed during skeletal development and contributes to extracellular matrix organization at the growth plate (20). These expression changes may reflect distinct aspects of fibrotic remodeling, including abnormal collagen fiber assembly (*COL11A1, COL14A1*), impaired vascular homeostasis (*COL18A1*), and reactivation of developmental matrix pathways (*COL27A1*). Conversely, downregulated genes such as *COL5A3, COL13A1*, and *COL16A1* may play regulatory or structural roles that counteract fibrosis, although their specific functions in lung fibroblasts warrant further investigation. In survival analyses using Cox proportional hazards models, *COL3A1, COL4A4, COL6A6*, and *COL24A1* showed associations with mortality. Among these, *COL4A4* remained associated with mortality after adjustment for age and lung function parameters in a multivariable model. For illustrative purposes, exploratory Kaplan–Meier curves using expression thresholds derived for these genes are provided in the Figure S1. However, the present study was not designed to establish validated clinical biomarkers, and the interpretation of these survival associations should therefore be considered exploratory. In particular, the lack of experimental validation, such as RT-qPCR or protein-level assessment, limits definitive conclusions regarding clinical utility. In this context, the observed associations may offer insight into potential biological contexts rather than definitive clinical implications. *COL4A4* is of particular interest given its typical expression in epithelial and endothelial basement membranes (21, 22). Its elevated expression in fibroblasts may reflect changes in basement membrane remodeling, potentially related to aberrant epithelial–mesenchymal crosstalk or changes in fibroblast activation states. Previous studies have implicated type IV collagen fragments as circulating biomarkers in fibrotic diseases (23), but their cell-specific expression and mechanistic roles in IPF remain incompletely understood. While increased turnover of type VI collagen has been associated with disease progression in IPF (9, 24), the specific contribution of the *COL6A6* chain itself has not been well characterized in this disease. *COL24A1*, a relatively understudied collagen gene involved in skeletal development and matrix mineralization (25), was also transcriptionally upregulated in IPF fibroblasts. Higher *COL24A1* expression was associated with worse survival in our cohort, and this observation is consistent with independent proteomic studies reporting that elevated *COL24A1* levels in blood are linked to increased risk of death or transplantation in IPF (26). Together, these findings suggest that fibroblasts may contribute to circulating *COL24A1* signals, although direct experimental validation will be required to clarify its cellular origin and functional role in fibrotic remodeling. A key strength of this study is the use of lung-derived fibroblasts from a relatively large IPF cohort, enabling high-resolution analysis of cell-specific gene expression. Our RNA-seq methodology captured a broad dynamic range of collagen gene expression, surpassing the limitations of prior studies that relied on immunohistochemistry or small transcript panels. Collagen genes in IPF fibroblasts were often assessed one at a time by RT-PCR or protein assays, documenting upregulation of type I and III collagen (1) but not exploring wider patterns. Similarly, clinical biomarker studies in IPF have concentrated on circulating collagen fragments such as C3M, C6M (9). Our unbiased RNA-seq approach provides a more

comprehensive view, confirming known alterations while additionally revealing novel collagen gene dysregulation. These transcriptomic findings raise the question of whether fibroblast-derived collagen profiles may be relevant to survival-related clinical outcomes. Unlike circulating biomarkers derived from blood or bronchoalveolar lavage fluid (BALF), however, fibroblast-based profiling requires lung tissue acquisition, primary cell culture, limiting direct clinical applicability. However, circulating collagen fragments such as PRO-C3 and PRO-C6 have previously been shown to predict IPF progression (24), underscoring the relevance of collagen remodeling to disease course. While such biomarkers offer accessibility, they lack cellular specificity. In contrast, fibroblast-derived transcriptomic signatures provide mechanistic insight by directly reflecting the profibrotic activity. Previous work has shown that markers predominantly expressed in activated fibroblasts, such as fibroblast activation protein-α, are associated with disease progression in pulmonary fibrosis (27). Together, these observations support the concept that profiling disease-relevant cell types can yield biologically informative signatures, even when tissue access is required. Future studies should aim to determine whether fibroblast-derived collagen gene signatures correlate with serum or BALF-based markers, thereby facilitating their integration into clinically accessible biomarker frameworks. However, several limitations should be noted. Our sample size, while moderate (33 IPF, 10 controls); however, the limited number of survival events reduced statistical power, which likely contributed to the absence of FDR-significant associations. Accordingly, survival results based on unadjusted p-values and dichotomized Kaplan–Meier plots should be interpreted as exploratory. Secondly, we studied cultured bulk fibroblasts, which may not fully recapitulate the in vivo lung environment; the process of cell isolation and culture could alter gene expression. The bulk-RNA approach can't discern the heterogeneity of fibroblast subpopulations recently uncovered by single-cell analyses. Third, the control lung samples were obtained from histologically normal regions of stage I or II lung cancer patients. Although these tissues were non-fibrotic, the possibility that they were indirectly influenced by the subtle molecular alterations or tumor microenvironment cannot be excluded. In conclusion, this study reveals distinct transcriptional alterations in collagen genes in IPF fibroblasts, including upregulation of both fibrillar and basement membrane collagens. *COL4A4* showed a consistent association with mortality in multivariable analyses, highlighting its potential relevance to fibroblast-mediated extracellular matrix remodeling. Rather than establishing validated biomarkers, our findings provide mechanistic insight into collagen dysregulation in IPF and generate hypotheses regarding survival-related molecular signatures. Further experimental validation and multi-omics studies will be required to clarify the biological roles of these collagens and their potential relevance to disease progression.

Acknowledgements: Biospecimens and data used in this study were provided by the Biobank of Soonchunhyang University Bucheon Hospital (KBN4_A06), a member of the Korea Biobank Network. We sincerely thank Myung-Ran Lee, Young-Gil Lee, Jeong-Eun Lee, and Da-Jeong Jang for their valuable assistance in the collection and management of biospecimens and clinical data.

Funding: This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (RS202300240858, RS202400462943).

Conflict of Interests: Each author declares that he or she has no commercial associations (e.g., consultancies, stock ownership, equity interest, patent/licensing arrangement etc.) that might pose a conflict of interest in connection with the submitted article.

Data Availability: The RNA-seq datasets analyzed during the current study have been deposited in the NCBI Gene Expression Omnibus (GEO) under accession number GSE301181. Additional data are available upon reasonable request.

Ethics Approval: This study was performed in line with the approval from the Institutional Review Board (IRB numbers: SCHCA-IRB-2018-10-034 and 201910-BR-058).

Consent to Participate: Informed consent was obtained from all individual participants included in the study.

Consent to Publish: Informed written consent was obtained from all participants.

Authors' Contributions: SEJ, LJU, PCS, CHS: Conceptualization; SEJ, LJU, KMK: Data curation; LJU, PCS: Funding acquisition; LJU, PCS, CHS, CJS: Methodology; SEJ, LJU: Validation; SEJ, LJU, PSL: Visualization; SEJ, LJU, HHG, KJH, PCS, CHS: Writing – original draft; all authors: Writing – review and editing.

Declaration on the Use of AI: None.

References

1. Jessen H, Hoyer N, Prior TS, Frederiksen P, Karsdal MA, Leeming DJ, et al. Turnover of type I and III collagen predicts progression of idiopathic pulmonary fibrosis. Respir Res. 2021 Jul 15;22(1):205. doi: 10.1186/s12931-021-01801-0.
2. Ko SJ, Choi SM, Han KD, Lee CH, Lee J. All-cause mortality of patients with idiopathic pulmonary fibrosis: a nationwide population-based cohort study in Korea. Sci Rep. 2021;11(1):15145. doi: 10.1038/s41598-021-94655-x.
3. Gimenez A, Duch P, Puig M, Gabasa M, Xaubet A, Alcaraz J. Dysregulated Collagen Homeostasis by Matrix Stiffening and TGF-beta1 in Fibroblasts from Idiopathic Pulmonary Fibrosis Patients: Role of FAK/Akt. Int J Mol Sci. 2017;18(11). doi: 10.3390/ijms18112431.
4. Hardie WD, Glasser SW, Hagood JS. Emerging concepts in the pathogenesis of lung fibrosis. Am J Pathol. 2009;175(1):3-16. doi: 10.2353/ajpath.2009.081170.
5. Staab-Weijnitz CA. Fighting the Fiber: Targeting Collagen in Lung Fibrosis. Am J Respir Cell Mol Biol. 2022;66(4):363-81. doi: 10.1165/rcmb.2021-0342TR.
6. Ricard-Blum S. The collagen family. Cold Spring Harb Perspect Biol. 2011;3(1):a004978. doi: 10.1101/cshperspect.a004978.
7. Gelse K, Poschl E, Aigner T. Collagens--structure, function, and biosynthesis. Adv Drug Deliv Rev. 2003;55(12):1531-46. doi: 10.1016/j.addr.2003.08.002.
8. Liu L, Stephens B, Bergman M, May A, Chiang T. Role of Collagen in Airway Mechanics. Bioengineering (Basel). 2021;8(1). doi: 10.3390/bioengineering8010013.
9. Hoyer N, Jessen H, Prior TS, Sand JMB, Leeming DJ, Karsdal MA, et al. High turnover of types III and VI collagen in progressive idiopathic pulmonary fibrosis. Respirology. 2021;26(6):582-9. doi: 10.1111/resp.14056.
10. Lee JU, Cheong HS, Shim EY, Bae DJ, Chang HS, Uh ST, et al. Gene profile of fibroblasts identify relation of CCL8 with idiopathic pulmonary fibrosis. Respir Res. 2017;18(1):3. doi: 10.1186/s12931-016-0493-6.
11. Raghu G, Remy-Jardin M, Myers JL, Richeldi L, Ryerson CJ, Lederer DJ, et al. Diagnosis of Idiopathic Pulmonary Fibrosis. An Official ATS/ERS/JRS/ALAT Clinical Practice Guideline. Am J Respir Crit Care Med. 2018;198(5):e44-e68. doi: 10.1164/rccm.201807-1255ST.
12. Raghu G, Collard HR, Egan JJ, Martinez FJ, Behr J, Brown KK, et al. An official ATS/ERS/JRS/ALAT statement: idiopathic pulmonary fibrosis: evidence-based guidelines for diagnosis and management. Am J Respir Crit Care Med. 2011;183(6):788-824. doi: 10.1164/rccm.2009-040GL.
13. Birk DE, Fitch JM, Babiarz JP, Doane KJ, Linsenmayer TF. Collagen fibrillogenesis in vitro: interaction of types I and V collagen regulates fibril diameter. J Cell Sci. 1990;95 (Pt 4):649-57. doi: 10.1242/jcs.95.4.649.
14. Peyser R, MacDonnell S, Gao Y, Cheng L, Kim Y, Kaplan T, et al. Defining the Activated Fibroblast Population in Lung Fibrosis Using Single-Cell Sequencing. Am J Respir Cell Mol Biol. 2019;61(1):74-85. doi: 10.1165/rcmb.2018-0313OC.
15. Urushiyama H, Terasaki Y, Nagasaka S, Terasaki M, Kunugi S, Nagase T, et al. Role of alpha1 and alpha2 chains of type IV collagen in early fibrotic lesions of idiopathic interstitial pneumonias and migration of lung fibroblasts. Lab Invest. 2015;95(8):872-85. doi: 10.1038/labinvest.2015.66.
16. Chen H, Qu J, Huang X, Kurundkar A, Zhu L, Yang N, et al. Mechanosensing by the alpha6-integrin confers an invasive fibroblast phenotype and mediates lung fibrosis. Nat Commun. 2016;7:12564. doi: 10.1038/ncomms12564.
17. Li Y, Lacerda DA, Warman ML, Beier DR, Yoshioka H, Ninomiya Y, et al. A fibrillar collagen gene, Col11a1, is essential for skeletal morphogenesis. Cell. 1995;80(3):423-30. doi: 10.1016/0092-8674(95)90492-1.
18. Grassel S, Bauer RJ. Collagen XVI in health and disease. Matrix Biol. 2013;32(2):64-73. doi: 10.1016/j.matbio.2012.11.001.
19. Li Q, Olsen BR. Increased angiogenic response in aortic explants of collagen XVIII/endostatin-null mice. Am J Pathol. 2004;165(2):415-24. doi: 10.1016/S0002-9440(10)63307-X.
20. Hjorten R, Hansen U, Underwood RA, Telfer HE, Fernandes RJ, Krakow D, et al. Type XXVII collagen at the transition of cartilage to bone during skeletogenesis. Bone. 2007;41(4):535-42. doi: 10.1016/j.bone.2007.06.024.
21. Loscertales M, Nicolaou F, Jeanne M, Longoni M, Gould DB, Sun Y, et al. Type IV collagen drives alveolar epithelial-endothelial association and the morphogenetic movements of septation. BMC Biol. 2016;14:59. doi: 10.1186/s12915-016-0281-2.
22. Abreu-Velez AM, Howard MS. Collagen IV in Normal Skin and in Pathological Processes. N Am J Med Sci. 2012;4(1):1-8. doi: 10.4103/1947-2714.92892.
23. Leeming DJ, Sand JM, Nielsen MJ, Genovese F, Martinez FJ, Hogaboam CM, et al. Serological investigation of the collagen degradation profile of patients with chronic obstructive pulmonary disease or idiopathic pulmonary fibrosis. Biomark Insights. 2012;7:119-26. doi: 10.4137/BMI.S9415.
24. Organ LA, Duggan AR, Oballa E, Taggart SC, Simpson JK, Kang'ombe AR, et al. Biomarkers of collagen synthesis predict progression in the PROFILE idiopathic pulmonary fibrosis cohort. Respir Res. 2019 Jul 12;20(1):148. doi: 10.1186/s12931-019-1118-7.

25. Matsuo N, Tanaka S, Yoshioka H, Koch M, Gordon MK, Ramirez F. Collagen XXIV (Col24a1) gene expression is a specific marker of osteoblast differentiation and bone formation. Connect Tissue Res. 2008;49(2):68-75. doi: 10.1080/03008200801913502.
26. Oldham JM, Huang Y, Bose S, Ma SF, Kim JS, Schwab A, et al. Proteomic Biomarkers of Survival in Idiopathic Pulmonary Fibrosis. Am J Respir Crit Care Med. 2024; 209(9):1111-20. doi: 10.1164/rccm.202301-0117OC.
27. Lavis P, Garabet A, Cardozo AK, Bondue B. The fibroblast activation protein alpha as a biomarker of pulmonary fibrosis. Front Med (Lausanne). 2024;11:1393778. doi: 10.3389/fmed.2024.1393778.

Annex

Table S1. Expression levels of collagen genes in 33 IPF fibroblasts compared to 10 controls

Gene name	Control	IPF	Log_2FC	*p*-value
COL1A1	8.85±0.51	9.15±1.51	1.055	0.027
COL1A2	11.08±0.34	11.54±0.9	0.712	0.013
COL3A1	10.16±0.68	11.23±1.3	1.445	< 0.001
COL4A4	0.18±0.21	0.87±0.85	3.169	< 0.001
COL4A5	4.65±0.45	5.49±0.99	1.219	0.002
COL4A6	0.72±0.29	1.67±1.28	2.731	< 0.001
COL5A2	9.92±0.37	10.37±0.88	0.656	0.014
COL5A3	4.01±0.95	2.93±1.17	-0.952	0.011
COL6A6	0.59±0.23	1.11±0.98	2.061	0.003
COL8A1	7.75±0.43	8.21±0.92	0.711	0.015
COL11A1	4.03±1.67	5.04±1.69	1.332	0.018
COL13A1	4.61±0.31	3.69±0.92	-0.708	< 0.001
COL14A1	2.44±0.86	4.14±1.9	2.722	< 0.001
COL16A1	7.7±0.35	7.06±0.45	-0.597	< 0.001
COL18A1	5.19±0.66	5.75±0.97	0.879	0.015
COL24A1	2.3±0.31	2.95±0.62	0.899	< 0.001
COL27A1	2.32±0.49	2.83±0.66	0.717	0.012
COL4A1	9.48±0.45	9.23±0.86	-0.069	0.767
COL4A2	9.13±0.37	8.93±0.74	-0.039	0.844
COL5A1	6.86±0.38	6.74±1.19	0.419	0.288
COL6A1	8.23±0.37	8.2±0.67	0.096	0.673
COL6A2	8.96±0.54	9.01±0.56	0.078	0.696
COL6A3	11.43±0.15	11.34±0.63	0.044	0.848

Copyright: The Author(s), 2026. Licensee Mattioli 1885, Fidenza, Italy. This is an open-access article distributed under the terms of the Creative Commons Attribution NonCommercial License (CC BY-NC-4.0).
Disclaimer/Publisher's Note: The statements, opinions and data contained in this article are solely those of the author(s) and contributor(s) and do not necessarily reflect those of their affiliated organizations, the publisher, the editors or the reviewers. The publisher and the editors disclaim any responsibility for injury to people or property resulting from any ideas, methods, instructions or products mentioned in the content. Any product that may be evaluated in this article, or claim made by its manufacturer, is not guaranteed or endorsed by the publisher.

Gene name	Control	IPF	Log_2FC	*p*-value
COL7A1	5.73±0.46	5.58±0.99	0.098	0.765
COL8A2	2.89±0.39	2.62±0.84	-0.047	0.853
COL9A2	0.71±0.35	0.83±0.36	0.307	0.374
COL10A1	2.06±0.58	2.27±0.9	0.504	0.214
COL11A2	0.53±0.17	0.56±0.24	0.149	0.636
COL12A1	10.92±0.24	10.73±0.52	-0.107	0.537
COL15A1	3.28±1.67	3.9±1.81	0.690	0.280

Values are presented as mean ± standard error of the mean. *Abbreviations:* IPF: idiopathic pulmonary fibrosis, log_2FC: log_2 fold change of IPF/Control

Table S2. Determination of optimal cut-off values for collagen gene expression using ROC analysis and Youden's Index

Gene	Cut-off value	C-index	Sensitivity (%)	Specificity (%)	Overall accuracy (%)	*p*-value
*COL3A1**	11.86	0.78	60	88.9	78.6	0.002
COL4A4	1.27	0.65	40	88.9	71.4	0.197
COL6A6	0.90	0.62	70	61.1	64.3	0.310
*COL24A1**	3.42	0.76	50	94.4	78.6	0.009

This table presents the optimal cut-off values for collagen-related gene expression used in survival analyses, determined by Youden's Index from ROC analysis. C-index is equivalent to the area under the curve as the outcome was binary and uncensored. *$p < 0.05$. *Abbreviations:* ROC: receiver operating characteristic curve; C-index: concordance index

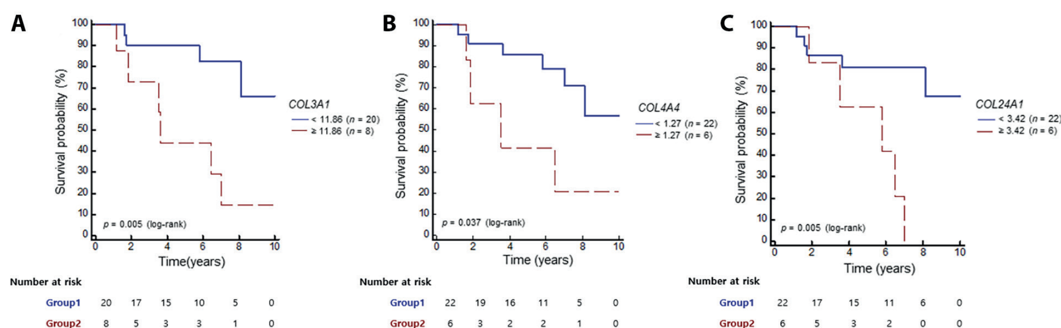

Figure S1. Kaplan-Meier plots of individual collagen subtype gene expression level illustrating survival rate of 28 IPF patients. (a-c) Kaplan–Meier plots over a follow-up period of 1 to 10 years shown with number at risk respectively. Patients with <1 year of follow-up (n = 5) were excluded, resulting in a final cohort of 28 IPF patients. Cut-off values were determined using Youden's Index from receiver operating characteristic curve analysis. *P*-values were calculated using the log-rank test. *Abbreviations:* IPF: idiopathic pulmonary fibrosis; HR: hazard ratio.