



OPEN ACCESS

EDITED BY

Lei Bu,
New York University School of
Medicine, United States

REVIEWED BY

Andrew Williams,
Health Canada, Canada
Sayed Haidar Abbas Raza,
Northwest A&F University, China
Vadim Fraifeld,
Ben-Gurion University of the Negev,
Israel

*CORRESPONDENCE

Lin Yang,
✉ linyang821@163.com
Lina Chen,
✉ chenlina_78@163.com

SPECIALTY SECTION

This article was submitted to Genetics of
Common and Rare Diseases,
a section of the journal
Frontiers in Genetics

RECEIVED 03 July 2022

ACCEPTED 30 November 2022

PUBLISHED 06 January 2023

CITATION

Wu Y, Zhong L, Qiu L, Dong L, Yang L
and Chen L (2023), A potential three-
gene-based diagnostic signature for
idiopathic pulmonary fibrosis.
Front. Genet. 13:985217.
doi: 10.3389/fgene.2022.985217

COPYRIGHT

© 2023 Wu, Zhong, Qiu, Dong, Yang
and Chen. This is an open-access article
distributed under the terms of the
[Creative Commons Attribution License
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or
reproduction in other forums is
permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original
publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or
reproduction is permitted which does
not comply with these terms.

A potential three-gene-based diagnostic signature for idiopathic pulmonary fibrosis

Yi Wu^{1,2,3}, Lin Zhong^{1,2,3}, Li Qiu^{1,2,3}, Liqun Dong^{1,2,3}, Lin Yang^{1,2,3*}
and Lina Chen^{1,2,3*}

¹Division of Pediatric Pulmonology and Immunology, West China Second University Hospital, Sichuan University, Chengdu, China, ²Key Laboratory of Birth Defects and Related Diseases of Women and Children (Sichuan University), Ministry of Education, Chengdu, China, ³NHC Key Laboratory of Chronobiology (Sichuan University), Chengdu, China

Background: Idiopathic pulmonary fibrosis (IPF) is a life-threatening disease whose etiology remains unknown. This study aims to explore diagnostic biomarkers and pathways involved in IPF using bioinformatics analysis.

Methods: IPF-related gene expression datasets were retrieved and downloaded from the NCBI Gene Expression Omnibus database. Differentially expressed genes (DEGs) were screened, and weighted correlation network analysis (WGCNA) was performed to identify key module and genes. Functional enrichment analysis was performed on genes in the clinically significant module. Then least absolute shrinkage and selection operator (LASSO) logistic regression and support vector machine-recursive feature elimination (SVM-RFE) algorithms were run to screen candidate biomarkers. The expression and diagnostic value of the biomarkers in IPF were further validated in external test datasets (GSE110147).

Results: 292 samples and 1,163 DEGs were screened to construct WGCNA. In WGCNA, the blue module was identified as the key module, and 59 genes in this module correlated highly with IPF. Functional enrichment analysis of blue module genes revealed the importance of extracellular matrix-associated pathways in IPF. IL13RA2, CDH3, and COMP were identified as diagnostic markers of IPF via LASSO and SVM-RFE. These genes showed good diagnostic value for IPF and were significantly upregulated in IPF.

Conclusion: This study indicates that IL13RA2, CDH3, and COMP could serve as diagnostic signature for IPF and might offer new insights in the underlying diagnosis of IPF.

Abbreviations: AE-IPF, acute exacerbation of IPF; AUC, area under the receiver operating characteristic curve; COMP, Cartilage oligomeric matrix protein; CDH3, cadherin three; DEGs, differentially expressed genes; ECM, extracellular matrix; GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; PCA, principal component analysis; IPF, idiopathic pulmonary fibrosis; WGCNA, weighted correlation network analysis; LASSO, least absolute shrinkage and selection operator; SVM-RFE, support vector machine-recursive feature elimination; IL13RA2, interleukin-13 receptor alpha two; MMP, matrix metalloproteinase; TGF- β , transforming growth factor beta; ROC, receiver operating characteristic.

KEYWORDS

idiopathic pulmonary fibrosis, diagnostic biomarkers, WGCNA, bioinformatics analysis, differentially expressed genes

Introduction

Idiopathic pulmonary fibrosis (IPF) is a chronic, progressive, fibrotic lung disease of unknown cause (Spagnolo et al., 2021). The global annual incidence of IPF is estimated to range from 0.2 to 93.7 cases per 100,000 people (Hutchinson et al., 2015), and the median survival time from diagnosis is 4.5 years (Kaunisto et al., 2019). Substantial progress has been made in the clinical management of IPF. Two effective agents, nintedanib and pirfenidone, benefit physiological deterioration and progression-free survival (Noble et al., 2016; Raghu et al., 2017). Timely diagnosis may improve outcomes for patients in the future, but the coexistence of other disorders and the lack of specific symptoms delay diagnosis, which has worsened mortality (Lamas et al., 2011). Current efforts are directed at identifying IPF early, potentially relying on combinations of

biomarkers (Martinez et al., 2017), and identifying key biomarkers that may direct more personalized medicine to improve long-term prognosis (Barratt et al., 2018).

Major advances recognizing biological mechanisms and biomarkers have occurred over the past decade with the development of transcriptome analysis (Casamassimi et al., 2017; Raza et al., 2022). Numerous studies have focused on circulating molecular markers or lung-specific sampling to improve the diagnosis of IPF (Ley et al., 2014). These biomarkers include extracellular matrix (ECM)-modifying enzymes (Kropski et al., 2015), matrix metalloproteinase (MMP)-degraded proteins (Jenkins et al., 2015), inflammatory proteins (Li et al., 2022), and the transcriptomic signature (Bauer et al., 2015). However, knowledge of molecular biomarkers for IPF remains in its infancy (Ley et al., 2014).

In this study, we used multiple bioinformatics methods to identify key gene co-expression modules, functional pathways,

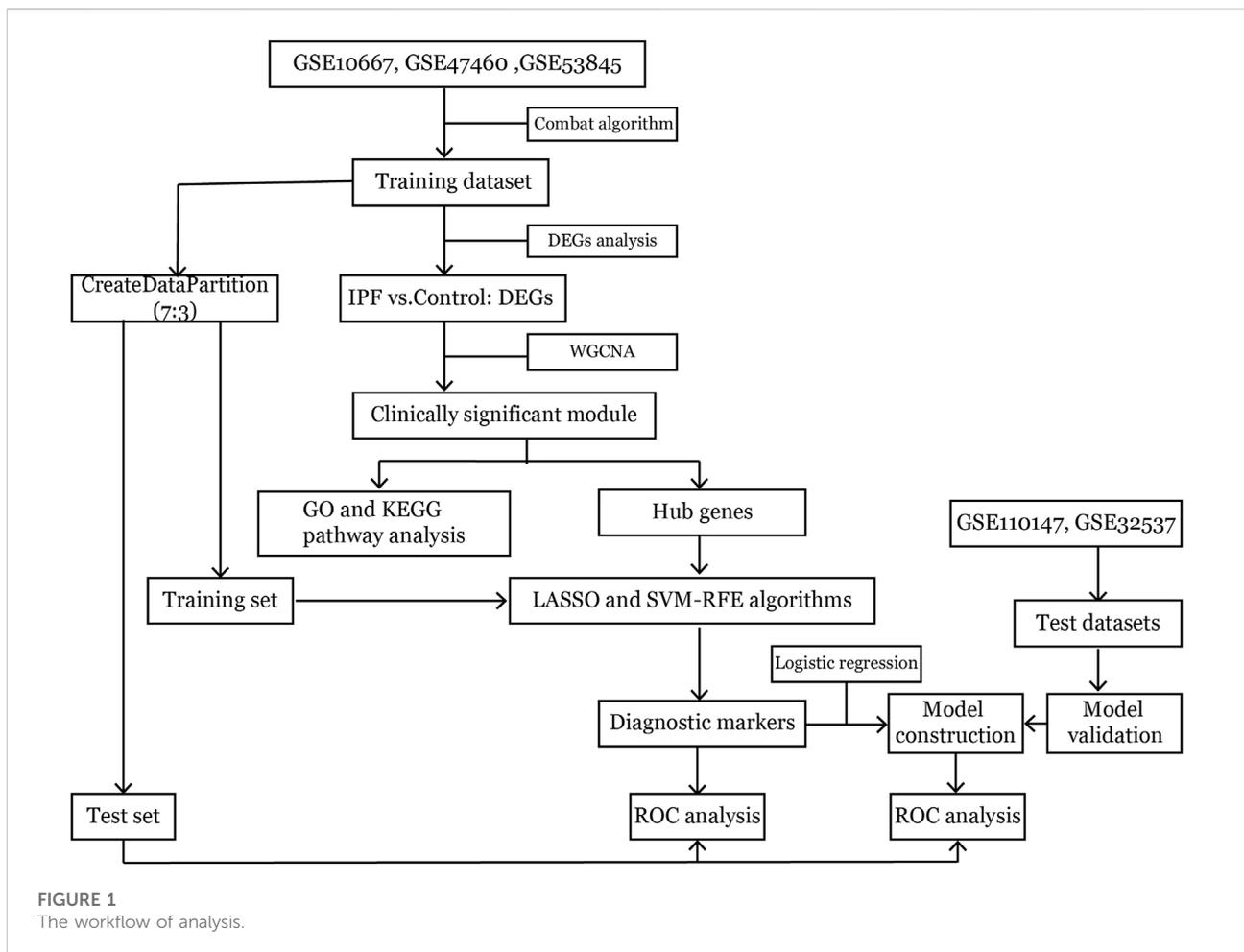


TABLE 1 Microarray datasets included in this study.

Geo accession	Publication year	Sample size (IPF/Control)	Source	Platform
GSE110147	2018	22/11	Lung tissue	Affymetrix Human Gene 1.0 ST Array
GSE53845	2014	40/8	Lung tissue	Agilent-014850 Whole Human Genome Microarray 4 × 44K G4112F
GSE47460	2013	122/91	Lung tissue	Agilent-028004 SurePrint G3 Human GE 8 × 60K Microarray
GSE32537	2013	119/50	Lung tissue	Affymetrix Human Gene 1.0 ST Array
GSE10667	2009	31/15	Lung tissue	Agilent-014850 Whole Human Genome Microarray 4 × 44K G4112F

and significant diagnostic biomarkers in IPF, and constructed a potential diagnostic signature for IPF.

Materials and methods

Acquisition of microarray data

The workflow of this study is shown in [Figure 1](#). Microarray datasets of IPF-related gene expression profiles were retrieved and downloaded from the NCBI Gene Expression Omnibus public database (<https://www.ncbi.nlm.nih.gov/geo/>). The included datasets met several selection criteria: 1) the datasets published in the past 15 years; 2) samples had been extracted from lung tissue; 3) the total sample size in two groups was at least 30; and 4) raw data were available in the datasets. Five datasets that met the selection criteria and the diagnostic criteria for IPF were analyzed in this study ([American Thoracic Society, 2000](#); [Holmes et al., 2006](#); [Raghu et al., 2011](#)), and information about them was presented in [Table 1](#). Ethical approval for this study was not required because these involved datasets were freely available in public datasets.

Data preprocessing and differential expression analysis

Raw data from the GSE10667, and datasets were normalized with the normexp function and the between-array-normalization quantile in Limma R package ([Ritchie et al., 2015](#)) and merged into a training dataset. These three datasets were selected for integrated analysis because they had the same platform, which is crucial when combining different datasets. Then SVA R package was used to remove batch effects using the combat algorithm ([Leek et al., 2012](#)). Two-dimensional principal component analysis (PCA) was used to evaluate whether the batch effect had been removed. Outlier samples were detected and removed by hierarchical cluster analysis with average linkage. Limma R package was used to investigate the DEGs in the training dataset, and genes with a false discovery rate <0.05 and absolute log₂ fold change >0.5 were considered DEGs.

Construction of Co-Expression network

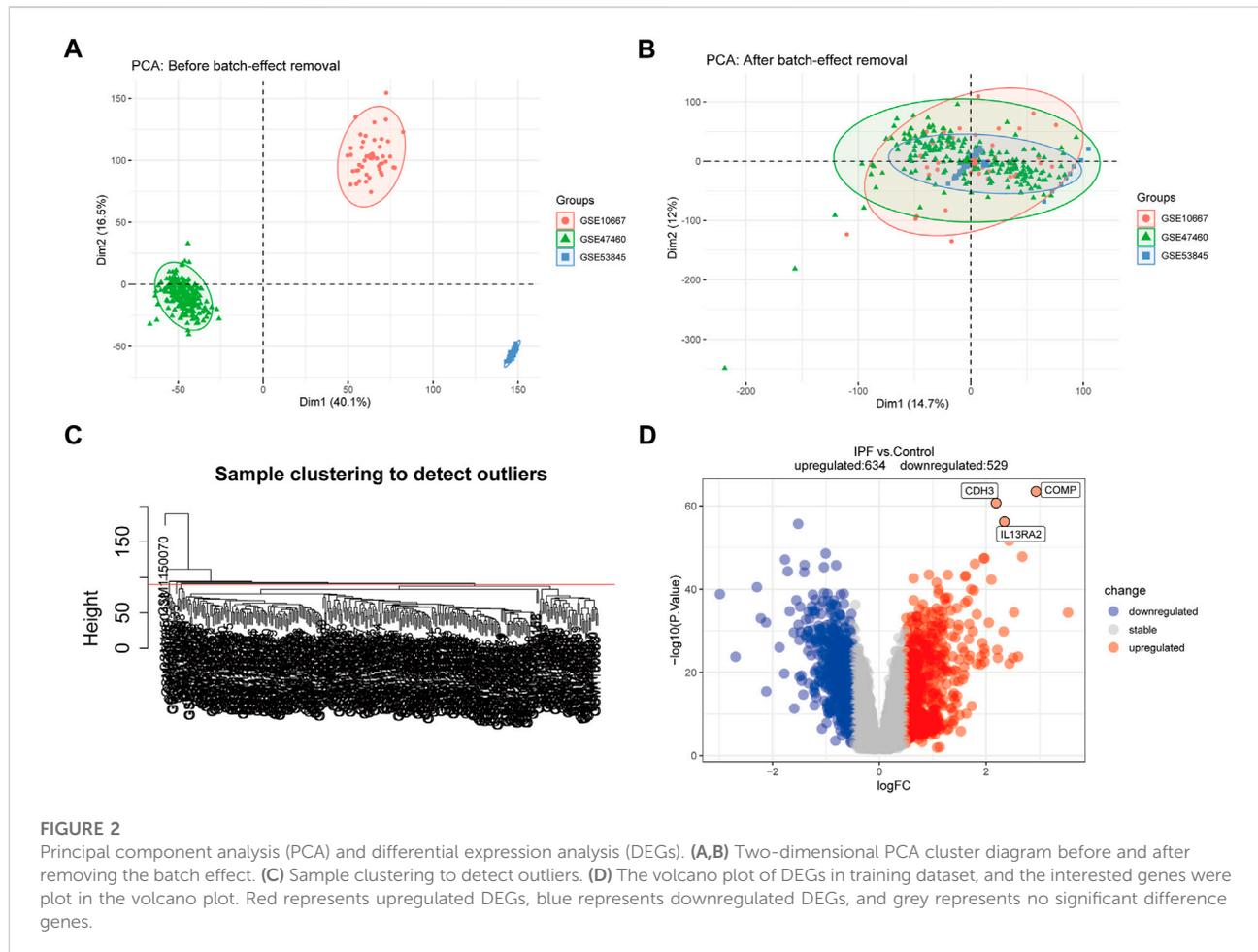
We used DEGs to construct a gene co-expression network *via* WGCNA R package ([Langfelder and Horvath, 2008](#)). We constructed a scale-free network ($R^2 = 0.85$) based on the following criteria: the soft-thresholding power β was set as 20, the minimum number of genes in the modules was 30, and the threshold for cut height to merge possible similar modules was 0.25. To further analyze the module, we calculated the module eigengene, which represents the expression of all genes in a given module. The correlation between the module eigengene and the genes was defined as the module membership, and the correlation between the genes and clinical traits was defined as the gene significance. Module eigengene values were correlated with control and IPF groups by Pearson's correlation. Finally, the module most highly associated with IPF groups was selected for further analysis.

Functional enrichment analysis

For further understanding the function of the genes in the most related module for IPF, we conducted Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses using ClusterProfiler R package ([Wu et al., 2021](#)). The ontology of the GO analysis contains three categories: molecular function, biological process, and cellular component. Adjusted $p < 0.05$ was set as the cutoff to identify significantly enriched pathways.

Screening of candidate diagnostic biomarker

To screen candidate diagnostic biomarkers, we identified genes with high within-module connectivity of the modules ($| \text{module membership} | > 0.8$) and $| \text{gene significance} | > 0.6$ as hub genes and used them to build the least absolute shrinkage and selection operator (LASSO) and support vector machine-recursive feature elimination (SVM-RFE) algorithms ([Huang et al., 2014](#)). The LASSO algorithm, with penalty parameter



tuning conducted by 10-fold cross-validation, was applied with the glmnet R package. The SVM-RFE algorithm was established by e1071 R package to search for lambda with the smallest classification error to select appropriate features, and the k-fold cross-validation was set as 10. The overlapping genes between the two algorithms were selected as candidate diagnostic biomarkers, and their expression was further validated by the validation datasets (GSE110147, GSE32537).

Verification of diagnostic markers

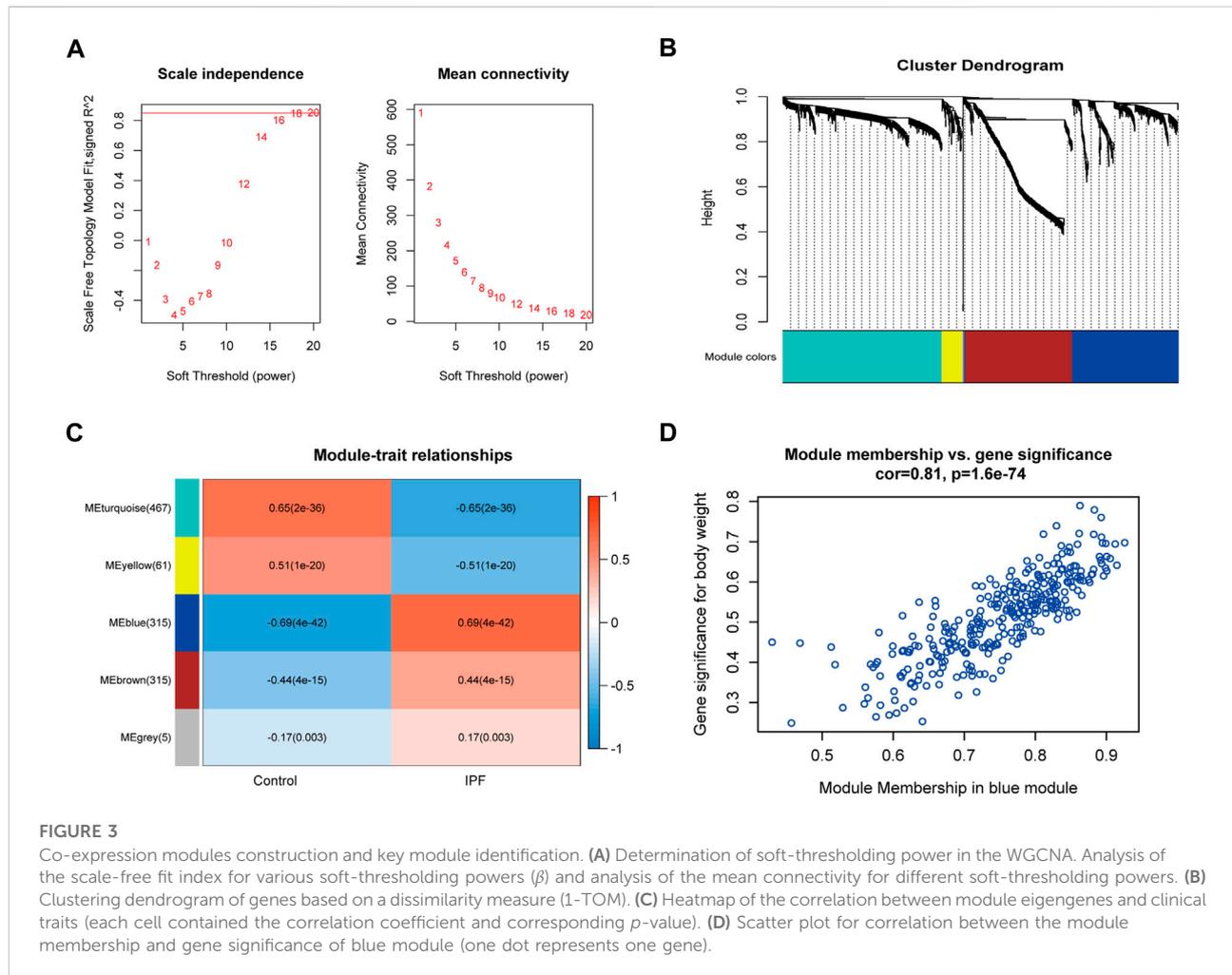
The area under the curve and the receiver operating characteristic (ROC) curve were plotted with the pROC package to evaluate the capability of candidate diagnostic biomarkers between controls and IPF. We randomly divided samples from the training dataset into a training set and a test set (7:3) using the createDataPartition function in the caret R package and constructed the

diagnostic model using the logistic regression algorithm in the training set. We used the test set and validation datasets to evaluate the capability of the diagnostic model between IPF and controls *via* ROC analysis.

Results

Removal of batch effects and identification of DEGs

The training dataset sample distributions before and after the removal of batch effects were visualized in a two-dimensional PCA cluster diagram (Figures 2A,B), and the scatter plot of PCA based on normalized expression showed that batch processing effects were eliminated (Figure 2B). 15 outlier samples were removed by hierarchical cluster analysis (cutHeight = 90; Figure 2C), and a total of 1,163 DEGs—529 downregulated and 634 upregulated



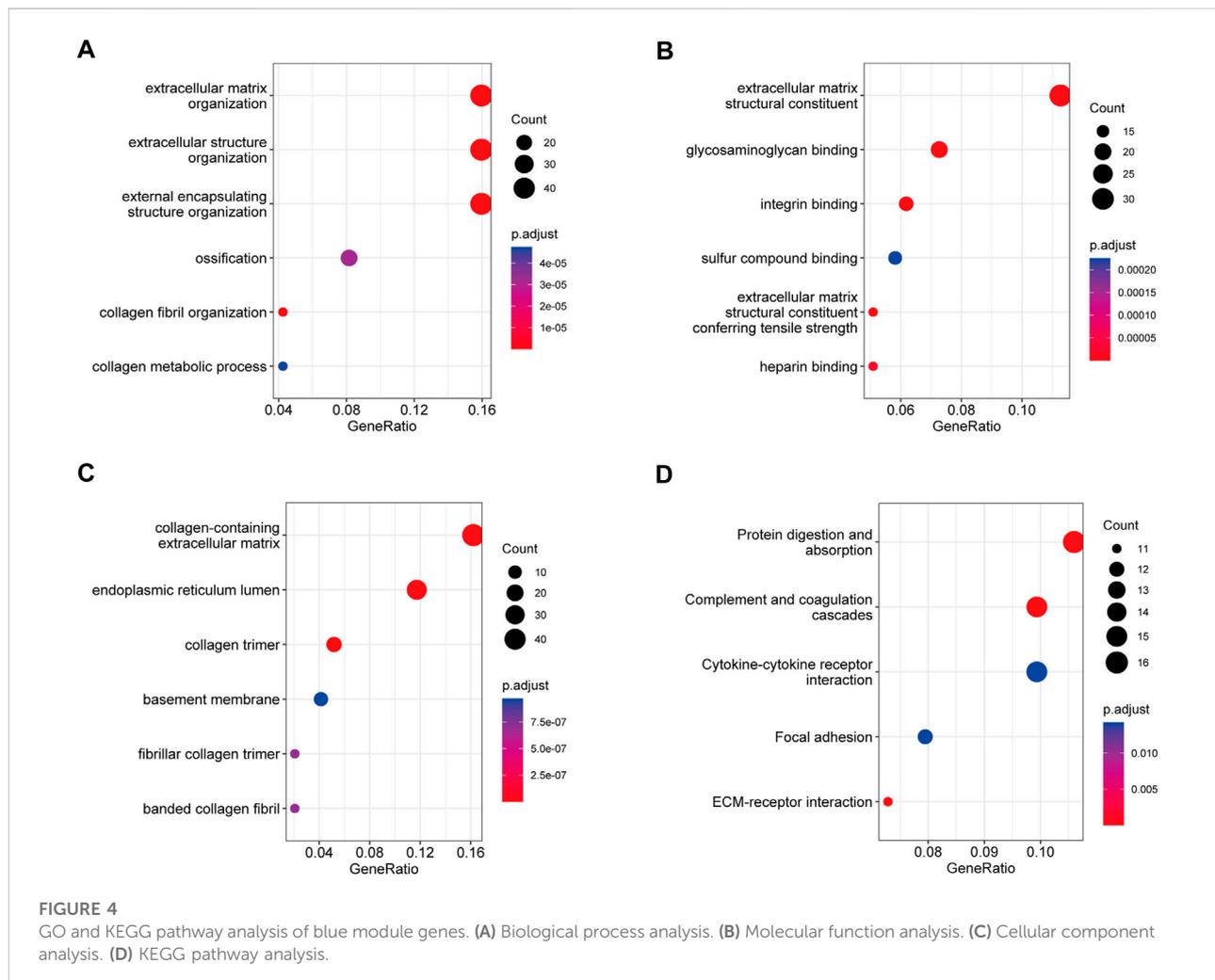
genes—were identified in the training dataset (Figure 2D and Supplementary Table 1).

Construction of the weighted Co-expression network and identification of key modules

A total of 292 samples and 1,163 DEGs were used to construct the co-expression network analysis. The power of $\beta = 20$ (scale-free $R^2 = 0.85$) was selected as the correlation coefficient threshold to identify the module-trait relationship (Figure 3A). A total of five modules were identified (Figure 3B). The blue module, which included 315 genes, had the strongest association with IPF (Figure 3C), and 59 genes with high connectivity in the blue module were selected for further analysis based on the cutoff criteria ($|\text{gene significance}| > 0.6$ and $|\text{module membership}| > 0.8$; Figure 3D and Supplementary Table 2).

GO and KEGG pathway analyses

In the biological process category (Figure 4A), the genes were mainly enriched in ECM organization, extracellular structure organization, and external encapsulating structure organization. In the molecular function category (Figure 4B), the genes were mainly enriched in ECM structural constituent, endopeptidase activity, and glycosaminoglycan binding. In the cellular component category (Figure 4C), the genes were mainly enriched in collagen-containing ECM, endoplasmic reticulum lumen, and collagen trimer. KEGG pathway analysis revealed that protein digestion and absorption, complement and coagulation cascades, and ECM-receptor interaction were the most significant pathways (Figure 4D). These results indicate that blue module genes might impact the development of IPF by influencing these biological processes and signaling pathways.



Screening and verification of diagnostic markers

A total of 23 genes were identified as signature genes by the LASSO algorithm (Figures 5A,B), and three genes were identified by the SVM-RFE algorithm (Figure 5C). Interleukin-13 receptor alpha 2 (IL13RA2), cadherin 3 (CDH3), and Cartilage oligomeric matrix protein (COMP) were selected as signature genes by both the LASSO and SVM-RFE algorithms (Figure 5D), and they were identified as candidate diagnostic markers. The expression of the three genes was tested in the validation datasets, and it was notably higher in IPF lung tissue ($p < 0.001$; Figures 5E,F).

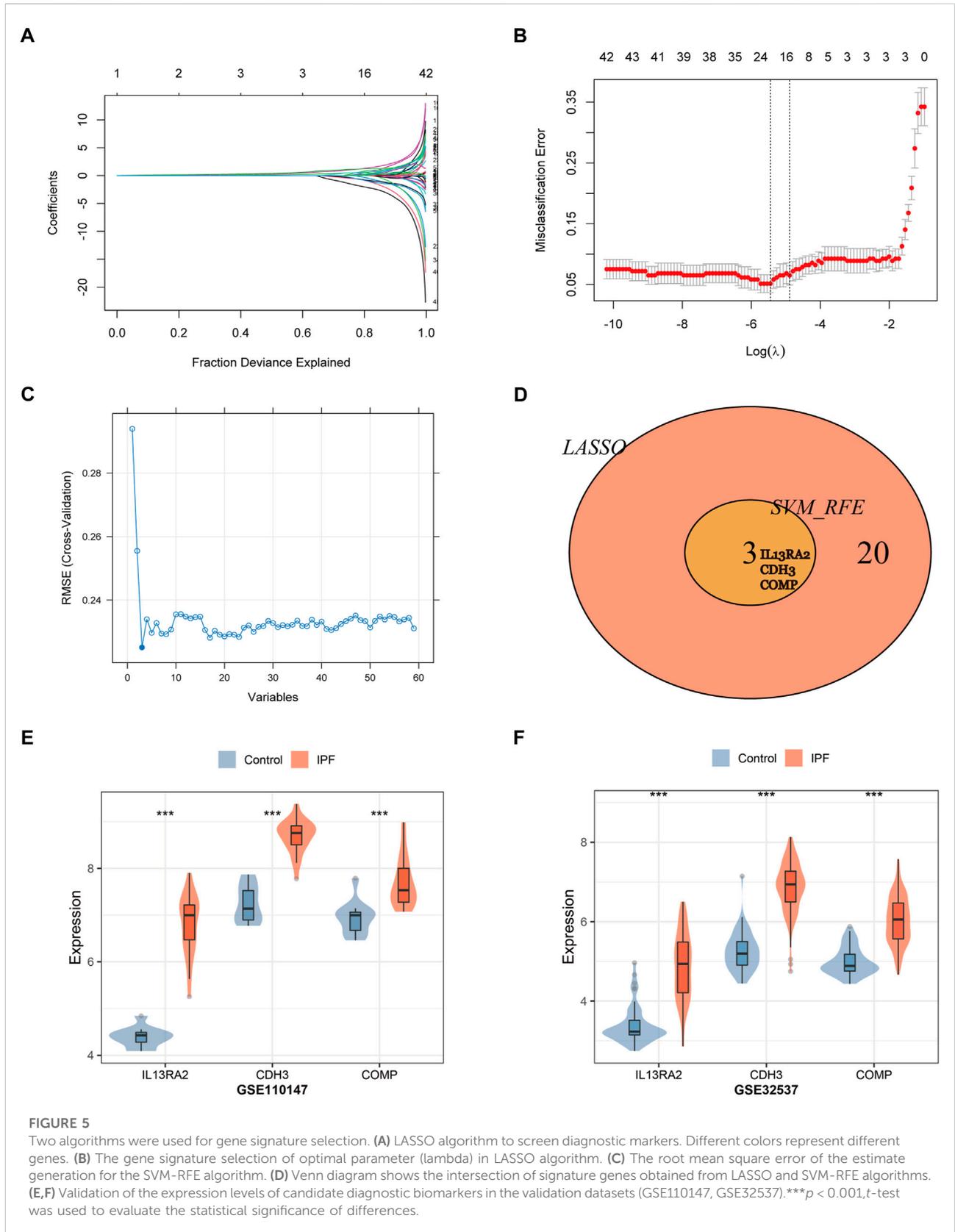
Construction and verification of the IPF diagnostic model

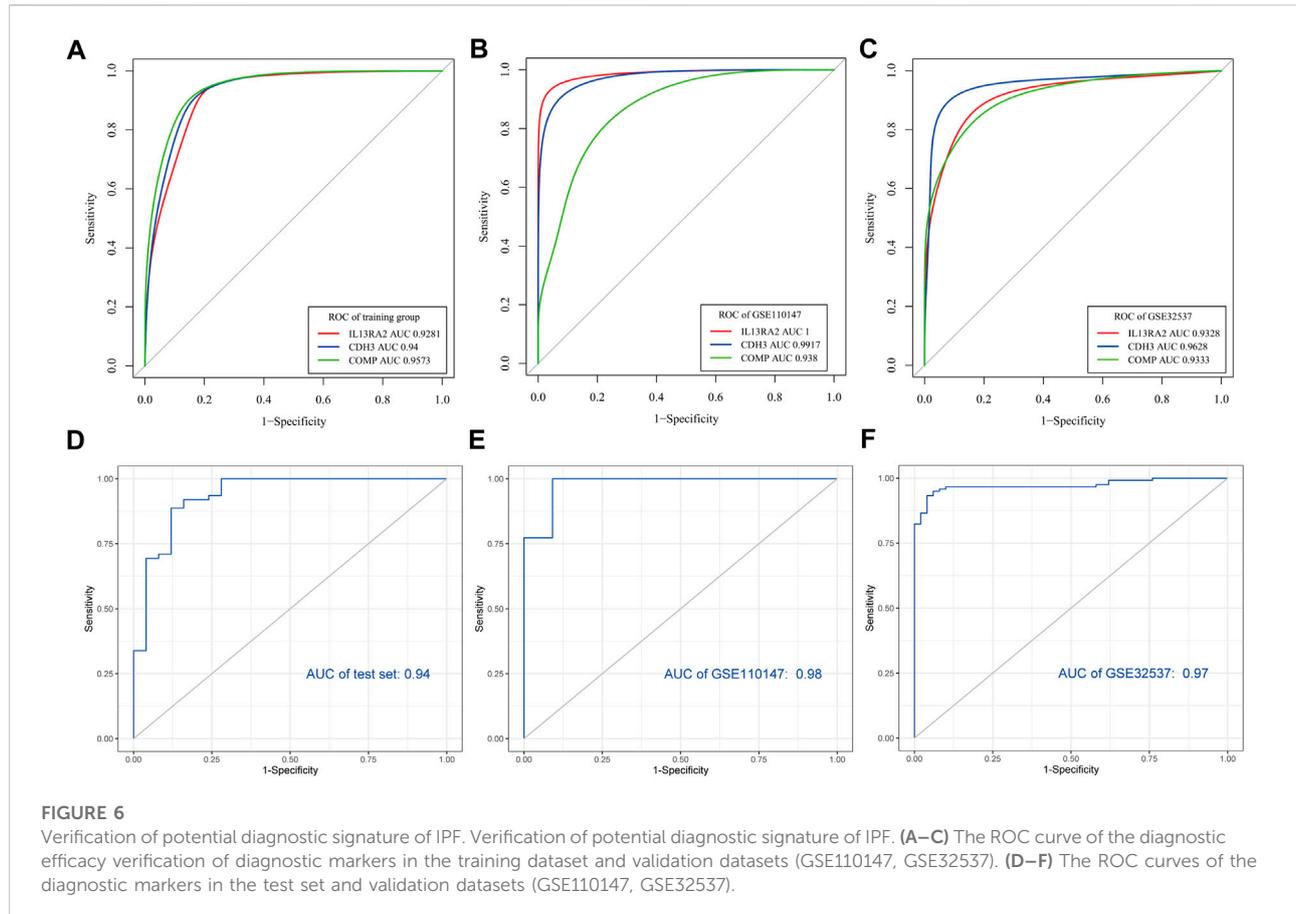
Every area under the ROC curve (AUC) of IL13RA2, CDH3, and COMP was greater than 0.9 in training and validation

datasets (GSE110147) (Figures 6A–C). When the three genes were combined into one variable, the area under the curve in the test set was 0.94 (Figure 6D). It is important to note that we also used external test datasets to validate the diagnostic value of the three-gene signature. The area under the curve for the three-gene signature was 0.98 in GSE110147 (Figure 6E) and 0.97 in GSE32537 (Figure 6F), which indicated that the three-gene signature had high diagnostic value.

Discussion

Numerous studies have found that dysregulated genes are involved in the progression of IPF and might be potential diagnostic biomarkers of IPF (Ley et al., 2014; Martinez et al., 2017). However, most of these studies have focused on the diagnostic value of a single biomarker, and small sample sizes may limit the clinical application of these biomarkers. Integrating multiple biomarkers and increasing sample sizes with developing





high-throughput technologies and bioinformatics may improve predictive accuracy. In this study, we constructed a novel gene signature for diagnosing IPF by integrating multiple datasets and combining multiple statistical methods. Three genes (IL13RA2, CDH3, and COMP) were identified as independent diagnostic biomarkers. An integrated three-gene signature was built based on their regression coefficients and expression profiles, ROC analysis indicated that the signature had high diagnostic accuracy, as confirmed by two external test datasets.

The mouse bleomycin-induced fibrosis model is generally viewed as the standard in modeling pulmonary fibrosis (Jenkins et al., 2017). Recent studies revealed that genes related to the mouse bleomycin-induced fibrosis model and human IPF have much in common (Bauer et al., 2015; Toren et al., 2021). Interleukin (IL)-13 is a type 2 cytokine with important roles in inflammatory and fibrotic diseases (Gieseck et al., 2018; Bieber, 2020). IL-13 receptor $\alpha 2$ (IL-13R $\alpha 2$), which has great affinity for IL-13, acts as a non-signaling decoy receptor (Chandriani et al., 2014; Ulzii et al., 2019). IL-13 and its receptors are increased in the blood and lung tissue of IPF patients (Murray et al., 2008; Chandriani et al., 2014). TGF- β is upregulated and activated in fibrotic diseases and is considered a central mediator of fibrogenesis (Györfi et al., 2018; Frangogiannis, 2020).

Previous studies have found that IL-13 signals through IL-13R $\alpha 2$ to induce transforming growth factor beta (TGF- β) and fibrosis progression (Fichtner-Feigl et al., 2008), and silencing of IL-13R $\alpha 2$ reduced TGF- β production and lung fibrosis (Fichtner-Feigl et al., 2006). In contrast, Robert V et al. found that overexpression of IL-13R $\alpha 2$ inhibited the IL-13 induction of fibrotic markers and bleomycin-induced pulmonary fibrosis (Lumsden et al., 2015). IL-13R $\alpha 2$ deficiency had been shown to lead to increases in collagen deposition (Wilson et al., 2007). Given these contrasting possibilities, it is vital to interpret the meaning of dysregulated IL-13R $\alpha 2$ expression in IPF and the function of IL-13R $\alpha 2$. The expression of IL13RA2 was significantly higher in IPF and acute exacerbation of IPF (AE-IPF) compared with control samples, but there was no significant difference between IPF and AE-IPF (Konishi et al., 2009). These results indicate that IL13RA2 may serve as a signature for IPF, but can't predict the onset of AE-IPF.

CDH3, also known as placental cadherin, is a classic cell-to-cell adhesion molecule that regulates multiple cellular homeostatic processes in normal tissue (Vieira and Paredes, 2015). High CDH3 expression is associated with tumor progression in invasive epithelial tumors (Ribeiro et al., 2010) and non-small-cell lung cancer (Imai et al., 2018). A well-known

mechanism associated with CDH3-induced cancer cell invasion is MMP activation (Ribeiro et al., 2010; Imai et al., 2018). MMPs have been implicated in the pathogenesis of IPF, and most MMPs promote the development of IPF (Craig et al., 2015; Mahalanobish et al., 2020). CDH3 was expressed in aberrant basaloid cells in the IPF lung (Adams et al., 2020; Travaglini et al., 2020) and overexpressed in all three stages of IPF (early-, progressive, and end-stage IPF) (Ghandikota et al., 2022), but was not found to be differentially expressed in bleomycin-treated animals (Bauer et al., 2015). It was not further investigated due to lack of translational relevance.

COMP is an ECM glycoprotein that plays a role in fibrillogenesis and collagen secretion (Rockey et al., 2015; Posey et al., 2018). Dysregulated COMP expression is involved in numerous diseases (Posey et al., 2018), including fibrosis (Agarwal et al., 2013; Vuga et al., 2013). COMP expression was increased in skin fibroblasts and lung epithelial cell induction by TGF- β , silencing COMP expression in human lung fibroblasts was associated with a reduction in TGF- β 1 (Agarwal et al., 2013; Vuga et al., 2013). Once fibrosis is initiated, COMP and TGF- β constitute a mutual positive regulation loop (Agarwal et al., 2013; Cui and Zhang, 2022). Serum COMP was upregulated in IPF patients and correlated with declines in force vital capacity, indicating that it is a potential biomarker for disease activity (Vuga et al., 2013). K. Miller, et al. found that the expression of COMP was significantly increased in bleomycin-treated mice, while tissue elastance and lung function testing showed significant differences compared to control group. However, there was no significant difference between bleomycin treated WT and COMP-KO mice. It indicated COMP maybe a biomarker for pulmonary fibrosis but not a causative factor in bleomycin-induced fibrosis (Miller et al., 2019).

IPF is a chronic, progressive lung disease characterized by the progressive deposition of ECM proteins. Available evidence suggests that the ECM plays a central role in the pathogenesis of IPF (Hewlett et al., 2018). WGCNA is a bioinformatics application for exploring the relationships between clinical traits and co-expression modules (Langfelder and Horvath, 2008). In the present study, one module mostly associated with IPF was found by WGCNA. Functional enrichment analysis revealed that the module's genes involved in multiple biological processes, including ECM organization, extracellular structure organization, and collagen metabolism by protein digestion and absorption signaling pathways. These results suggest that dysregulated expression of these genes could play important roles in orchestrating the development of IPF by influencing ECM-associated pathways.

Although the three-gene signature constructed here appears to be a potential diagnostic signature for IPF, this study has some limitations. First, the expression and exact mechanisms of the biomarkers should be further investigated experimentally. Second, the diagnostic value of the three-gene signature was only tested in microarray datasets in this study and should be further validated with clinical data. Third, other

genes in the clinically significant module were not studied in this study.

In conclusion, we revealed the involvement of the key gene co-expression module and functional pathways in the pathogenesis of IPF. In addition, we identified IL13RA2, CDH3, and COMP as potential biomarkers of IPF and constructed an integrated three-gene signature of IPF. This work provides further insights into the underlying molecular mechanisms and diagnosis of IPF.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary Material.

Author contributions

YW and LY designed the study. LZ, LQ, and LD performed data analysis. YW wrote the manuscript. LC provided suggestions and guidance of this study, and revised the manuscript. All the authors read and approved the final manuscript.

Funding

This study was partially supported by the grant 22ZDZX0014 and 2020YFS0105 from the Science & Technology department of Sichuan Province.

Acknowledgments

We would like to thank Xiaoli Zhang for the figure technology support. We also thank Dr. Jianming Zeng (University of Macau), and all the members of his bioinformatics team, biotrainee, for generously sharing their experience and codes.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their

affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Adams, T. S., Schupp, J. C., Poli, S., Ayaub, E. A., Neumark, N., Ahangari, F., et al. (2020). Single-cell RNA-seq reveals ectopic and aberrant lung-resident cell populations in idiopathic pulmonary fibrosis. *Sci. Adv.* 6 (28), eaba1983. doi:10.1126/sciadv.aba1983
- Agarwal, P., Schulz, J. N., Blumbach, K., Andreasson, K., Heinegård, D., Paulsson, M., et al. (2013). Enhanced deposition of cartilage oligomeric matrix protein is a common feature in fibrotic skin pathologies. *Matrix Biol.* 32 (6), 325–331. doi:10.1016/j.matbio.2013.02.010
- American Thoracic Society (2000). American thoracic society. Idiopathic pulmonary fibrosis: Diagnosis and treatment. International consensus statement. American thoracic society (ATS), and the European respiratory society (ERS). *Am. J. Respir. Crit. Care Med.* 161 (1), 646–664. doi:10.1164/ajrccm.161.2.ats3-00
- Barratt, S. L., Creamer, A., Hayton, C., and Chaudhuri, N. (2018). Idiopathic pulmonary fibrosis (IPF): An overview. *J. Clin. Med.* 7 (8), 201. doi:10.3390/jcm7080201
- Bauer, Y., Tedrow, J., de Bernard, S., Birker-Robaczewska, M., Gibson, K. F., Guardela, B. J., et al. (2015). A novel genomic signature with translational significance for human idiopathic pulmonary fibrosis. *Am. J. Respir. Cell Mol. Biol.* 52 (2), 217–231. doi:10.1165/rcmb.2013-0310OC
- Bieber, T. (2020). Interleukin-13: Targeting an underestimated cytokine in atopic dermatitis. *Allergy* 75 (1), 54–62. doi:10.1111/all.13954
- Casamassimi, A., Federico, A., Rienzo, M., Esposito, S., and Ciccocicola, A. (2017). Transcriptome profiling in human diseases: New advances and perspectives. *Int. J. Mol. Sci.* 18 (8), 1652. doi:10.3390/ijms18081652
- Chandriani, S., DePianto, D. J., N'Diaye, E. N., Abbas, A. R., Jackman, J., Bevers, J., 3rd, et al. (2014). Endogenously expressed IL-13Ra2 attenuates IL-13-mediated responses but does not activate signaling in human lung fibroblasts. *J. Immunol.* 193 (1), 111–119. doi:10.4049/jimmunol.1301761
- Craig, V. J., Zhang, L., Hagood, J. S., and Owen, C. A. (2015). Matrix metalloproteinases as therapeutic targets for idiopathic pulmonary fibrosis. *Am. J. Respir. Cell Mol. Biol.* 53 (5), 585–600. doi:10.1165/rcmb.2015-0020TR
- Cui, J., and Zhang, J. (2022). Cartilage oligomeric matrix protein, diseases, and therapeutic opportunities. *Int. J. Mol. Sci.* 23 (16), 9253. doi:10.3390/ijms23169253
- Fichtner-Feigl, S., Strober, W., Kawakami, K., Puri, R. K., and Kitani, A. (2006). IL-13 signaling through the IL-13alpha2 receptor is involved in induction of TGF-beta1 production and fibrosis. *Nat. Med.* 12 (1), 99–106. doi:10.1038/nm1332
- Fichtner-Feigl, S., Young, C. A., Kitani, A., Geissler, E. K., Schlitt, H. J., and Strober, W. (2008). IL-13 signaling via IL-13R alpha2 induces major downstream fibrogenic factors mediating fibrosis in chronic TNBS colitis. *Gastroenterology* 135 (6), 2003–2013. e2001-2007. doi:10.1053/j.gastro.2008.08.055
- Frangogiannis, N. (2020). Transforming growth factor- β in tissue fibrosis. *J. Exp. Med.* 217 (3), e20190103. doi:10.1084/jem.20190103
- Ghandikota, S., Sharma, M., Ediga, H. H., Madala, S. K., and Jegga, A. G. (2022). Consensus gene Co-expression network analysis identifies novel genes associated with severity of fibrotic lung disease. *Int. J. Mol. Sci.* 23 (10), 5447. doi:10.3390/ijms23105447
- Gieseck, R. L., 3rd, Wilson, M. S., and Wynn, T. A. (2018). Type 2 immunity in tissue repair and fibrosis. *Nat. Rev. Immunol.* 18 (1), 62–76. doi:10.1038/nri.2017.90
- Györfi, A. H., Matei, A. E., and Distler, J. H. W. (2018). Targeting TGF- β signaling for the treatment of fibrosis. *Matrix Biol.* 68–69, 8–27. doi:10.1016/j.matbio.2017.12.016
- Hewlett, J. C., Kropski, J. A., and Blackwell, T. S. (2018). Idiopathic pulmonary fibrosis: Epithelial-mesenchymal interactions and emerging therapeutic targets. *Matrix Biol.* 71–72, 112–127. doi:10.1016/j.matbio.2018.03.021
- Holmes, D., III, Bartholmai, B., Karwoski, R., Zavaletta, V., and Robb, R. (2006). “The Lung Tissue Research Consortium: An extensive open database containing histological, clinical, and radiological data to study chronic lung disease,” in Insight J (2006 MICCAI Open Science Workshop).
- Huang, M. L., Hung, Y. H., Lee, W. M., Li, R. K., and Jiang, B. R. (2014). SVM-RFE based feature selection and Taguchi parameters optimization for multiclass SVM classifier. *ScientificWorldJournal.* 2014, 795624. doi:10.1155/2014/795624
- Hutchinson, J., Fogarty, A., Hubbard, R., and McKeever, T. (2015). Global incidence and mortality of idiopathic pulmonary fibrosis: A systematic review. *Eur. Respir. J.* 46 (3), 795–806. doi:10.1183/09031936.00185114
- Imai, S., Kobayashi, M., Takasaki, C., Ishibashi, H., and Okubo, K. (2018). High expression of P-cadherin is significantly associated with poor prognosis in patients with non-small-cell lung cancer. *Lung Cancer* 118, 13–19. doi:10.1016/j.lungcan.2018.01.018
- Jenkins, R. G., Moore, B. B., Chambers, R. C., Eickelberg, O., Königshoff, M., Kolb, M., et al. (2017). An official American thoracic society workshop report: Use of animal models for the preclinical assessment of potential therapies for pulmonary fibrosis. *Am. J. Respir. Cell Mol. Biol.* 56 (5), 667–679. doi:10.1165/rcmb.2017-0096ST
- Jenkins, R. G., Simpson, J. K., Saini, G., Bentley, J. H., Russell, A. M., Braybrooke, R., et al. (2015). Longitudinal change in collagen degradation biomarkers in idiopathic pulmonary fibrosis: An analysis from the prospective, multicentre PROFILE study. *Lancet. Respir. Med.* 3 (6), 462–472. doi:10.1016/s2213-2600(15)00048-x
- Kaunisto, J., Salomaa, E. R., Hodgson, U., Kaarteenaho, R., Kankaanranta, H., Koli, K., et al. (2019). Demographics and survival of patients with idiopathic pulmonary fibrosis in the FinnishIPF registry. *ERJ Open Res.* 5 (3), 00170. doi:10.1183/23120541.00170-2018
- Konishi, K., Gibson, K. F., Lindell, K. O., Richards, T. J., Zhang, Y., Dhir, R., et al. (2009). Gene expression profiles of acute exacerbations of idiopathic pulmonary fibrosis. *Am. J. Respir. Crit. Care Med.* 180 (2), 167–175. doi:10.1164/rccm.200810-1596OC
- Kropski, J. A., Pritchett, J. M., Zoz, D. F., Crossno, P. F., Markin, C., Garnett, E. T., et al. (2015). Extensive phenotyping of individuals at risk for familial interstitial pneumonia reveals clues to the pathogenesis of interstitial lung disease. *Am. J. Respir. Crit. Care Med.* 191 (4), 417–426. doi:10.1164/rccm.201406-1162OC
- Lamas, D. J., Kawut, S. M., Bagiella, E., Philip, N., Arcasoy, S. M., and Lederer, D. J. (2011). Delayed access and survival in idiopathic pulmonary fibrosis: A cohort study. *Am. J. Respir. Crit. Care Med.* 184 (7), 842–847. doi:10.1164/rccm.201104-0668OC
- Langfelder, P., and Horvath, S. (2008). Wgcna: an R package for weighted correlation network analysis. *BMC Bioinforma.* 9, 559. doi:10.1186/1471-2105-9-559
- Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E., and Storey, J. D. (2012). The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* 28 (6), 882–883. doi:10.1093/bioinformatics/bts034
- Ley, B., Brown, K. K., and Collard, H. R. (2014). Molecular biomarkers in idiopathic pulmonary fibrosis. *Am. J. Physiol. Lung Cell. Mol. Physiol.* 307 (9), L681–L691. doi:10.1152/ajplung.00014.2014
- Li, Y., He, Y., Chen, S., Wang, Q., Yang, Y., Shen, D., et al. (2022). S100A12 as biomarker of disease severity and prognosis in patients with idiopathic pulmonary fibrosis. *Front. Immunol.* 13, 810338. doi:10.3389/fimmu.2022.810338
- Lumsden, R. V., Worrell, J. C., Boylan, D., Walsh, S. M., Cramton, J., Counihan, I., et al. (2015). Modulation of pulmonary fibrosis by IL-13Ra2. *Am. J. Physiol. Lung Cell. Mol. Physiol.* 308 (7), L710–L718. doi:10.1152/ajplung.00120.2014
- Mahalanobish, S., Saha, S., Dutta, S., and Sil, P. C. (2020). Matrix metalloproteinase: An upcoming therapeutic approach for idiopathic pulmonary fibrosis. *Pharmacol. Res.* 152, 104591. doi:10.1016/j.phrs.2019.104591
- Martinez, F. J., Collard, H. R., Pardo, A., Raghu, G., Richeldi, L., Selman, M., et al. (2017). Idiopathic pulmonary fibrosis. *Nat. Rev. Dis. Prim.* 3 (1), 17074. doi:10.1038/nrdp.2017.74

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2022.985217/full#supplementary-material>

- Miller, K., Wilson, C., Collum, S., Hecht, J., Karmouty-Quintana, H., and Alcorn, J. (2019). "Global deletion of COMP does not protect from bleomycin-induced pulmonary fibrosis," in *C65. Diffuse parenchymal lung diseases: Basic and clinical studies* (New York, United States: American Thoracic Society), A5421.
- Murray, L. A., Argentieri, R. L., Farrell, F. X., Bracht, M., Sheng, H., Whitaker, B., et al. (2008). Hyper-responsiveness of IPF/UIP fibroblasts: Interplay between TGFbeta1, IL-13 and CCL2. *Int. J. Biochem. Cell Biol.* 40 (10), 2174–2182. doi:10.1016/j.biocel.2008.02.016
- Noble, P. W., Albera, C., Bradford, W. Z., Costabel, U., du Bois, R. M., Fagan, E. A., et al. (2016). Pirfenidone for idiopathic pulmonary fibrosis: Analysis of pooled data from three multinational phase 3 trials. *Eur. Respir. J.* 47 (1), 243–253. doi:10.1183/13993003.00026-2015
- Posey, K. L., Coustry, F., and Hecht, J. T. (2018). Cartilage oligomeric matrix protein: COMPathies and beyond. *Matrix Biol.* 71–72, 161–173. doi:10.1016/j.matbio.2018.02.023
- Raghu, G., Collard, H. R., Egan, J. J., Martinez, F. J., Behr, J., Brown, K. K., et al. (2011). An official ATS/ERS/JRS/ALAT statement: Idiopathic pulmonary fibrosis: Evidence-based guidelines for diagnosis and management. *Am. J. Respir. Crit. Care Med.* 183 (6), 788–824. doi:10.1164/rccm.2009-040GL
- Raghu, G., Wells, A. U., Nicholson, A. G., Richeldi, L., Flaherty, K. R., Le Maulf, F., et al. (2017). Effect of nintedanib in subgroups of idiopathic pulmonary fibrosis by diagnostic criteria. *Am. J. Respir. Crit. Care Med.* 195 (1), 78–85. doi:10.1164/rccm.201602-0402OC
- Raza, S. H. A., Khan, R., Cheng, G., Long, F., Bing, S., Easa, A. A., et al. (2022). RNA-Seq reveals the potential molecular mechanisms of bovine KLF6 gene in the regulation of adipogenesis. *Int. J. Biol. Macromol.* 195, 198–206. doi:10.1016/j.ijbiomac.2021.11.202
- Ribeiro, A. S., Albergaria, A., Sousa, B., Correia, A. L., Bracke, M., Seruca, R., et al. (2010). Extracellular cleavage and shedding of P-cadherin: A mechanism underlying the invasive behaviour of breast cancer cells. *Oncogene* 29 (3), 392–402. doi:10.1038/ncr.2009.338
- Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., et al. (2015). Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 43 (7), e47. doi:10.1093/nar/gkv007
- Rockey, D. C., Bell, P. D., and Hill, J. A. (2015). Fibrosis--a common pathway to organ injury and failure. *N. Engl. J. Med.* 372 (12), 1138–1149. doi:10.1056/NEJMra1300575
- Spagnolo, P., Kropski, J. A., Jones, M. G., Lee, J. S., Rossi, G., Karamitsakos, T., et al. (2021). Idiopathic pulmonary fibrosis: Disease mechanisms and drug development. *Pharmacol. Ther.* 222, 107798. doi:10.1016/j.pharmthera.2020.107798
- Toren, D., Yanai, H., Abu Taha, R., Bunu, G., Ursu, E., Ziesche, R., et al. (2021). Systems biology analysis of lung fibrosis-related genes in the bleomycin mouse model. *Sci. Rep.* 11 (1), 19269. doi:10.1038/s41598-021-98674-6
- Travaglini, K. J., Nabhan, A. N., Penland, L., Sinha, R., Gillich, A., Sit, R. V., et al. (2020). A molecular cell atlas of the human lung from single-cell RNA sequencing. *Nature* 587 (7835), 619–625. doi:10.1038/s41586-020-2922-4
- Ulzii, D., Kido-Nakahara, M., Nakahara, T., Tsuji, G., Furue, K., Hashimoto-Hachiya, A., et al. (2019). Scratching counteracts IL-13 signaling by upregulating the decoy receptor IL-13ra2 in keratinocytes. *Int. J. Mol. Sci.* 20 (13), 3324. doi:10.3390/ijms20133324
- Veira, A. F., and Paredes, J. (2015). P-cadherin and the journey to cancer metastasis. *Mol. Cancer* 14, 178. doi:10.1186/s12943-015-0448-4
- Vuga, L. J., Milosevic, J., Pandit, K., Ben-Yehudah, A., Chu, Y., Richards, T., et al. (2013). Cartilage oligomeric matrix protein in idiopathic pulmonary fibrosis. *PLoS One* 8 (12), e83120. doi:10.1371/journal.pone.0083120
- Wilson, M. S., Elnekave, E., Mentink-Kane, M. M., Hodges, M. G., Pesce, J. T., Ramalingam, T. R., et al. (2007). IL-13Ralpha2 and IL-10 coordinately suppress airway inflammation, airway-hyperreactivity, and fibrosis in mice. *J. Clin. Invest.* 117 (10), 2941–2951. doi:10.1172/jci31546
- Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., et al. (2021). clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation*. 2 (3), 100141. doi:10.1016/j.xinn.2021.100141