U.PORTO

DOUTORAMENTO

MEDICINA

FIBROSING INTERSTITIAL LUNG DISEASES:

PATHOGENIC MARKERS AND THEIR PROGNOSTIC INFLUENCE

Patrícia Caetano Gomes da Mota

D

2024

Supervisor

António Morais, MD, PhD

Faculty of Medicine of the University of Porto

Department of Medicine, Pulmonology

Co-Supervisor

Miguel Luz Soares, PhD

Laboratory of Support to Research in Molecular Medicine

Department of Biomedicine

Faculty of Medicine of the University of Porto

JÚRI

PRESIDENTE

DOUTOR JOSÉ MANUEL ESTEVÃO DA COSTA,

PROFESSOR CATEDRÁTICO DA FACULDADE DE MEDICINA DA UNIVERSIDADE DO PORTO

VOGAIS

DOUTOR CARLOS MANUEL DA SILVA ROBALO CORDEIRO,

PROFESSOR CATEDRÁTICO DA FACULDADE DE MEDICINA DA UNIVERSIDADE DE COIMBRA

DOUTOR HILÁRIO LOURENÇO NUNES,

PROFESSOR CATEDRÁTICO DA UNIVERSITÉ SORBONNE PARIS NORD, FRANÇA

DOUTOR ANTÓNIO ALEXANDRE PINTO BUGALHO DE ALMEIDA,

PROFESSOR CATEDRÁTICO DA FACULDADE DE CIÊNCIAS MÉDICAS DA UNIVERSIDADE NOVA DE LISBOA

FILIPA ABREU GOMES DE CARVALHO,

PROFESSORA CATEDRÁTICA DA FACULDADE DE MEDICINA DA UNIVERSIDADE DO PORTO

ANTÓNIO MANUEL MARTINS DE MORAIS,

PROFESSOR AUXILIAR CONVIDADO DA FACULDADE DE MEDICINA DA UNIVERSIDADE DO PORTO

MARTA SUSANA DRUMMOND FREITAS,

PROFESSORA AUXILIAR CONVIDADA DA FACULDADE DE MEDICINA DA UNIVERSIDADE DO PORTO

VENCESLAU JOSÉ COELHO PINTO ESPANHOL,

PROFESSOR APOSENTADO DA FACULDADE DE MEDICINA DA UNIVERSIDADE DO PORTO

ACKNOWLEDGEMENTS

All the stages culminating in this doctoral dissertation were driven by the tireless work and effort of a research team, to whom I express my deep gratitude, today and always. Individually, I would like to thank:

Professor António Morais, supervisor of this thesis, founder of the Diffuse Lung Diseases Group, and current head of the Pulmonology Department at the Local Health Unit of São João (ULS São João), previous Centro Hospitalar Universitário de São João (CHUSJ), a key figure in my medical and academic career, for his constant encouragement and challenges, loyalty and sharing, and his unwavering confidence in me and my work as a doctor and researcher.

Professor Miguel Luz Soares, co-supervisor of this thesis, coordinator of the Laboratory for Support to Research in Molecular Medicine (LAIMM) at the Department of Biomedicine of the Faculty of Medicine of the University of Porto (FMUP), for his complete availability from the conception to the conclusion of this dissertation, his invaluable collaboration, work ethic, scientific rigor, constant enthusiasm, and motivation. My thanks extend to his entire research group for their exceptional collaboration on this project.

Professor Agostinho Marques, the first supervisor of this thesis, for recognizing my medical and academic skills and for promoting this research project.

Professor Venceslau Hespanhol, for his trust, professional recognition, and dedication to my growth as a pulmonologist and faculty member at FMUP.

The Diffuse Lung Diseases Group at ULS São João, for their professionalism, unity, and capacity of sharing knowledge and teachings, with a common goal of updating and growing the group's clinical and scientific components. My thanks extend to all my colleagues in Pulmonology, Radiology - Dr. André Carvalho, Dr. José Miguel Pereira, and Dr. Rui Cunha, Pathology - Dr. Susana Guimarães and Dr. Conceição Souto-Moura, members of the Immunology Laboratory at FMUP - Dr. Oksana Sokhatska, Dr. Marília Beltrão, and Professor Luís Delgado, and the nurses in the outpatient clinic and Bronchology unit.

Dr. Natália Melo, the current coordinator of the Diffuse Lung Diseases Group at ULS São João, for her genuine friendship, loyalty, boundless generosity, unconditional support, and for never letting me give up.

Professor Hélder Novais e Bastos and Professor Rita Santos, researchers of the FIBRALUNG trial (Fibrosing ILD Biomarkers that Rule Acceleration), for their motivation and collaboration, especially in the final phase of patient recruitment, laboratory support, and critical analysis of the reported results.

All the other co-authors for their contributions.

My colleagues and friends, Dr. Carla Damas, Dr. Anabela Marinho, Dr. Adelina Amorim, Professor Gabriela Fernandes, Dr. Ana Rosa Santos, Dr. Adriana Magalhães, and Dr. Isabel Gomes, for their example, inspiration, and friendship.

Professor Doctor Marta Drummond, my supervisor of Pulmonology Residence, for her affection and support through the years.

Dr. Rita Boaventura and other medical colleagues, from ULS São João and other hospital institutions, for their generosity and constant encouragement for scientific updating.

All the other professionals of the Pulmonology Department who, directly or indirectly, collaborated and believed in this project.

The institutions, ULS São João, FMUP, LAIMM, and the Institute for Research and Innovation in Health (i3S), and the funding entities that made this project possible.

The patients, for their complete availability and collaboration.

My family, my parents and brother, for their unconditional love and support, for believing and celebrating my achievements, demystifying my frustrations, always with infinite pride in my personal and professional journey.

Nuno and our daughters, Matilde and Benedita, for making me a better person every day.

AGRADECIMENTOS

Todas as etapas que culminaram nesta dissertação de doutoramento foram pautadas pelo trabalho incansável e esforço de uma equipa de investigação, a quem presto, hoje e sempre, a minha profunda gratidão.

De forma individual, agradeço:

Ao Sr. Professor Doutor António Morais, orientador desta tese, fundador do Grupo de Doenças Pulmonares Difusas e atual diretor do Serviço de Pneumologia da Unidade Local de Saúde de São João (ULS São João), anterior Centro Hospitalar Universitário de São João (CHUSJ), referência na minha carreira médica e académica, pelo estímulo e desafios constantes, pela lealdade e partilha, e pela confiança extrema que deposita na minha pessoa e no meu trabalho como médica e investigadora.

Ao Sr. Professor Doutor Miguel Luz Soares, coorientador desta tese, coordenador do Laboratório de Apoio à Investigação em Medicina Molecular (LAIMM), do Departamento de Biomedicina da Faculdade de Medicina da Universidade do Porto (FMUP), pela total disponibilidade desde a concetualização até à conclusão desta dissertação, colaboração inexcedível, ética de trabalho, estímulo e rigor científicos, entusiamo e motivação constantes. Agradecimento extensível a todo o seu grupo de trabalho, pela forma exímia como colaborou neste projeto.

Ao Sr. Professor Doutor Agostinho Marques, primeiro orientador desta tese, pelo reconhecimento das minhas competências médicas e académicas e por impulsionar este projeto de investigação.

Ao Sr. Professor Doutor Venceslau Hespanhol, pela confiança, pelo reconhecimento profissional, e por todo o seu empenho no meu crescimento como médica pneumologista e docente da FMUP.

Ao Grupo de Doenças Pulmonares Difusas da ULS São João, pelo profissionalismo, pela união e capacidade de partilha de ensinamentos e conhecimentos, pelo objetivo comum de atualização e crescimento das componentes clínica e científica do grupo. Agradecimento extensível a todos os colegas médicos de Pneumologia, Radiologia – Dr. André Carvalho, Dr. José Miguel Pereira e Dr. Rui Cunha, Anatomia Patológica – Dra. Susana Guimarães e Dra. Conceição Souto-Moura, elementos do Laboratório de

Imunologia da FMUP – Dra. Oksana Sokhatska, Dra. Marília Beltrão e Professor Doutor Luís Delgado, e enfermeiras da consulta e da unidade de Broncologia.

À Dra. Natália Melo, atual coordenadora do Grupo de Doenças Pulmonares Difusas da ULS São João, pela amizade genuína, lealdade, generosidade infinita, apoio incondicional, e por nunca me ter deixado desistir.

Ao Sr. Professor Doutor Hélder Novais e Bastos e Sra. Professora Doutora Rita Santos, investigadores do ensaio FIBRALUNG (*Fibrosing ILD Biomarkers that Rule Acceleration*), pela motivação e colaboração, sobretudo na fase final de recrutamento de doentes, apoio laboratorial e análise crítica dos resultados reportados.

A todos os restantes coautores pelo seu contributo.

Às colegas e amigas, Dra. Carla Damas, Dra. Anabela Marinho, Dra. Adelina Amorim, Sra. Professora Doutora Gabriela Fernandes, Dra. Ana Rosa Santos, Dra. Adriana Magalhães e Dra. Isabel Gomes pelo exemplo, inspiração e amizade.

À Sra. Professora Doutora Marta Drummond, minha orientadora de Formação Específica em Pneumologia, pelo carinho e apoio ao longo de todos estes anos.

À Dra. Rita Boaventura e demais colegas médicos, da ULS São João e de outras instituições hospitalares, pela generosidade e incentivo constante de atualização científica.

A todos os restantes profissionais do Serviço de Pneumologia que, direta ou indiretamente, colaboraram e acreditaram neste projeto.

Às instituições, ULS São João, FMUP, LAIMM e Instituto de Investigação e Inovação em Saúde (i3S) e às entidades financiadoras que proporcionaram a realização deste projeto.

Aos doentes, pela total disponibilidade e colaboração.

À minha família, aos meus pais e irmão, pelo amor e apoio incondicionais, por acreditarem e vibrarem com as minhas realizações, desmistificarem as minhas frustrações, sempre com um orgulho infinito no meu percurso pessoal e profissional. Ao Nuno e às nossas filhas, Matilde e Benedita, por todos os dias fazerem de mim uma pessoa melhor.

INDEX

ACKNOWLEDGEMENTS **5**

AGRADECIMENTOS **7**

LIST OF PUBLICATIONS **11**

ABBREVIATIONS **15**

ABSTRACT **21**

RESUMO **23**

INTRODUCTION **25**

DIFFUSE INTERSTITIAL LUNG DISEASES: GENERAL CONSIDERATIONS 25

STATE-OF-THE-ART OF FIBROSING INTERSTITIAL LUNG DISEASES 28

GENETICS IN SUSCEPTIBILITY AND PROGNOSIS OF FIBROSING INTERSTITIAL LUNG DISEASES 36

COMMON GENETIC VARIANTS IN IDIOPATHIC PULMONARY FIBROSIS AND FIBROTIC HYPERSENSITIVITY PNEUMONITIS 44

THESIS AIMS **53**

PUBLICATIONS **55**

FIRST STUDY 57

SECOND STUDY 81

THIRD STUDY 103

DISCUSSION **147**

CONCLUSIONS AND FUTURE PERSPECTIVES **161**

REFERENCES **163**

LIST OF PUBLICATIONS

In compliance with the provisions of Decree-Law No. 388/70, Article 8, paragraph 2, I clarify that I carried out the planning and execution of the experiments, observation of the material, and analysis of the results and drafted the first versions of the following publications, which are an integral part of this dissertation:

1. Mota PC, Soares ML, Vasconcelos CD, Ferreira AC, Lima BA, Manduchi E, Moore JH, Melo N, Novais-Bastos H, Pereira JM, Guimarães S, Moura CS, Marques JA, Morais A. Predictive value of common genetic variants in idiopathic pulmonary fibrosis survival. J Mol Med (Berl). 2022 Sep;100(9):1341-1353. doi: 10.1007/s00109-022-02242-y. Epub 2022 Aug 19. PMID: 35986225.

2. Mota PC, Soares ML, Ferreira AC, Santos RF, Rufo JC, Vasconcelos D, Carvalho A, Guimarães S, Vasques-Nóvoa F, Cardoso C, Melo N, Alexandre AT, Coelho D, Novais-Bastos H, Morais A. Polymorphisms and haplotypes of *TOLLIP* and *MUC5B* are associated with susceptibility and survival in patients with fibrotic hypersensitivity pneumonitis. Pulmonology. 2024 Feb 2:S2531-0437(24)00007-2. doi: 10.1016/j.pulmoe.2024.01.002. Epub ahead of print. PMID: 38309995.

3. Mota PC, Soares ML, Ferreira AC, Santos RF, Dias CC, Rufo JC, Vasconcelos D, Carvalho A, Guimarães S, Vasques-Nóvoa F, Cardoso CG, Melo N, Alexandre AT, Coelho D, Sokhatska O, Beltrão M, Novais-Bastos H, Morais A. *MUC5B* and *TOLLIP* polymorphisms in the differential diagnosis between idiopathic pulmonary fibrosis and fibrotic hypersensitivity pneumonitis: a case-control study.
   Submitted to publication in Respiratory Research.

The full-text versions of the published papers will be made available according to the publishing licensing and permissions of each journal (copyrights).

Other related publications:

- Freitas C, Lima B, Melo N, Mota P, Novais-Bastos H, Alves H, Sokhatska O, Delgado L, Morais A. Distinct TNF-alpha and HLA polymorphisms associate with fibrotic and non-fibrotic subtypes of hypersensitivity pneumonitis. Pulmonology. 2023 Dec;29 Suppl 4:S63-S69. doi: 10.1016/j.pulmoe.2021.08.013. Epub 2021 Oct 8. PMID: 34629327.
- Freitas C, Lima B, Martins N, Melo N, Mota P, Novais-Bastos H, Alves H, Sokhatska O, Delgado L, Morais A. Cytokine gene polymorphisms in Pigeon Breeder's disease expression. Sarcoidosis Vasc Diffuse Lung Dis. 2020;37(3):e2020004. doi: 10.36141/svdld.v37i3.9342. Epub 2020 Sep 30. PMID: 33264379; PMCID: PMC7690056.
- Santos G, Fabiano A, Mota PC, Rodrigues I, Carvalho D, Melo N, Novais-Bastos H, Alexandre AT, Moura CS, Guimarães S, Pereira JM, Carvalho A, Morais A. The impact of nintedanib and pirfenidone on lung function and survival in patients with idiopathic pulmonary fibrosis in real-life setting. Pulm Pharmacol Ther. 2023 Dec;83:102261. doi: 10.1016/j.pupt.2023.102261. Epub 2023 Sep 25. PMID: 37758002.
- Raimundo S, Pimenta AC, Cruz-Martins N, Rodrigues MC, Melo N, Mota PC, Sokhatska O, Bastos HN, Beltrão M, Guimarães S, Moura CS, Cunha R, Pereira JM, Morais A. Insights on chronic hypersensitivity pneumonitis' treatment: Factors associated with a favourable response to azathioprine. Life Sci. 2021 May 1;272:119274. doi: 10.1016/j.lfs.2021.119274. Epub 2021 Feb 20. PMID: 33621590.
- Terras Alexandre A, Martins N, Raimundo S, Melo N, Caetano Mota P, Novais-Bastos H, Pereira JM, Cunha R, Guimarães S, Souto Moura C, Morais A. Impact of azathioprine use in chronic hypersensitivity pneumonitis patients. Pulm Pharmacol Ther. 2020 Feb;60:101878. doi: 10.1016/j.pupt.2019.101878. Epub 2019 Dec 17. PMID: 31862300.
- Mendonça Almeida L, Fernandes AL, Gouveia Cardoso C, Lima B, Neves I, Novais-Bastos H, Caetano Mota P, Melo N, Souto Moura C, Guimarães S, Carvalho A, Cunha R, Pereira JM, Morais A. Mortality risk prediction with ILD-GAP index in a

fibrotic hypersensitivity pneumonitis cohort. Ther Adv Respir Dis. 2022 Jan-Dec;16:17534666221135316. doi: 10.1177/17534666221135316. PMID: 36476249; PMCID: PMC9742694.

- Santos V, Martins N, Sousa C, Jacob M, Padrão E, Melo N, Mota PC, Bastos HN, Guimarães S, Moura CS, Sokhatska O, Cunha R, Pereira J, Morais A. Hypersensitivity pneumonitis: Main features characterization in a Portuguese cohort. Pulmonology. 2020 May-Jun;26(3):130-137. doi: 10.1016/j.pulmoe.2019.09.004. Epub 2019 Oct 29. PMID: 31672592.
- Nogueira R, Melo N, Novais-Bastos H, Martins N, Delgado L, Morais A, Mota PC. Hypersensitivity pneumonitis: Antigen diversity and disease implications. Pulmonology. 2019 Mar-Apr;25(2):97-108. doi: 10.1016/j.pulmoe.2018.07.003. Epub 2018 Aug 17. PMID: 30126802.

ABBREVIATIONS

ABCA3, ATP binding cassette subfamily A member 3

AECs, Alveolar epithelial cells

AFOP, Acute fibrinous and organizing pneumonia

AIP, Acute interstitial pneumonia

BAL, Bronchoalveolar lavage

CHUSJ, Centro Hospitalar Universitário de São João

COP, Cryptogenic organizing pneumonia

CPFE, Combined pulmonary fibrosis and emphysema

CTD-ILDs, Connective tissue disease-associated interstitial lung diseases

DC, Dyskeratosis congenita

DIP, Desquamative interstitial pneumonia

DKC1, Dyskerin pseudouridine synthase 1

DLCO, Diffusing capacity of the lung for carbon monoxide

DPIf, Doenças pulmonares intersticiais fibrosantes

ERS, European Respiratory Society

fHP, Fibrotic hypersensitivity pneumonitis

fILDs, Fibrosing interstitial lung diseases

FMUP, Faculdade de Medicina da Universidade do Porto

FPI, Fibrose pulmonar idiopática

FPF, Familial pulmonary fibrosis

FVC, Forced vital capacity

GWAS, Genome-wide association studies

HLA, Human leucocyte antigen

HP, Hypersensitivity pneumonitis

HR, Hazard ratio

HRCT, High-resolution computed tomography

IFN-β, Interferon-beta

IFN-γ, Interferon-gamma

IgG, Immunoglobulin G

IGF, Insulin-like growth factor

IPF, Idiopathic pulmonary fibrosis

IIP, Idiopathic interstitial pneumonia

IL, Interleukin

ILAs, Interstitial lung abnormalities

ILDs, Interstitial lung diseases

i3S, Instituto de Investigação e Inovação em Saúde

IPAF, Interstitial pneumonia with auto-immune features

LAM, Lymphangioleiomyomatosis

LAIMM, Laboratório de Apoio à Investigação em Medicina Molecular

LCH, Langerhans cell histiocytosis

LIP, Lymphoid interstitial pneumonia

MAF, Minor allele frequency

MCTD, Mixed connective tissue disease

MDR, Multifactor dimensionality reduction

MHC, Major histocompatibility complex

mRNA, messenger ribonucleic acid

MUC5B, Mucin 5b

NAC, N-acetylcysteine

NAF1, Nuclear assembly factor 1 ribonucleoprotein

NGS, Next-generation sequencing

NHP2, NHP2 ribonucleoprotein

NOP10, NOP10 ribonucleoprotein

NSIP, Nonspecific interstitial pneumonia

OR, Odds ratio

PAP, Pulmonary alveolar proteinosis

PARN, Poly(A)-specific ribonuclease

PDGF, Platelet-derived growth factor

PHf, Pneumonite de hipersensibilidade fibrótica

PLAU, Plasminogen activator urokinase

$PM_{2.5}$, Particulate matter <2.5 µm in aerodynamic diameter

PPF, Progressive pulmonary fibrosis

PPFE, Pleuroparenchymal fibroelastosis

PRS, Polygenic risk score

RA-ILD, Rheumatoid arthritis-associated interstitial lung disease

RBILD, Respiratory bronchiolitis-associated interstitial lung disease

RR, Relative risk

RTEL1, Regulator of telomere elongation helicase

SERPINB1, Serpin family B member 1

SFTPA, Surfactant protein A

SFTPC, Surfactant protein C

SNP, Single-nucleotide polymorphism

SRGs, Surfactant-related genes

SSc-ILD, Systemic-sclerosis associated interstitial lung disease

TBLC, Transbronchial lung cryobiopsy

TERC, Telomerase RNA component

TERT, Telomerase reverse transcriptase

TGF-β, Transforming growth factor-beta

Th, T-helper

TINF2, TERF1 interacting nuclear factor 2

TL, Telomere length

TLR, Toll-like receptor

TNF-α, Tumor necrosis factor-alpha

TOLLIP, Toll-interacting protein

TRGs, Telomere-related genes

uILD, Unclassifiable interstitial lung disease

UIP, Usual interstitial pneumonia

ULS São João, Unidade Local de Saúde de São João

VUS, Variant of uncertain significance

WGS, Whole genome sequencing

ZCCHC8, Zinc finger CCHC-type containing 8

ABSTRACT

Fibrosing interstitial lung diseases (fILDs) represent a subset of conditions characterized by self-sustaining lung fibrosis, leading to progressive and often fatal outcomes. Idiopathic Pulmonary Fibrosis (IPF), the most recognized fILD, carries a poor prognosis with a median untreated life expectancy of 3–5 years. Similar outcomes are observed in other non-IPF fILDs, like fibrotic Hypersensitivity Pneumonitis (fHP). Accurate differentiation between IPF and fHP is crucial, especially in cases with indeterminate clinical and radiological presentations for both diseases, where pathogenic biomarkers may offer diagnostic guidance.

Genetic studies on IPF and fHP have primarily focused on research of rare and common variants in several genes, in particularly on single nucleotide polymorphisms (SNPs) within *MUC5B* and *TOLLIP*, two neighboring loci located on chromosome 11. However, these studies have been based on the study of the variants independently. A broader approach examining how these variants are arranged relative to each other within the genomic landscape, defining haplotypes, has not been commonly addressed in current literature.

This thesis consists of a three-step investigation to assess the independent impact of *MUC5B* and neighboring *TOLLIP* SNPs, and their respective haplotypes, in IPF and in fHP. The study entailed the analyses of sporadic IPF and fHP patients versus healthy controls and direct comparisons between patients of both conditions. The main objective was to identify genetic associations that can either distinguish or reveal similarities between the two fibrotic diseases, potentially providing new insights into their underlying molecular mechanisms.

Analyses of each SNP individually showed that the *MUC5B* rs35705950 was associated with susceptibility to the two conditions. The minor allele of rs35705950 was significantly more frequent in IPF and fHP patients than controls. In contrast, *TOLLIP* SNPs rs3750920, rs111521887, and rs5743894 were associated with susceptibility to fHP only, with a significantly higher frequency of their minor alleles in fHP patients compared to controls. Direct comparisons between IPF and fHP patients did not reveal significant differences in the allelic and genotypic distributions of these SNPs.

The analysis of haplotypes, defined by two specific *MUC5B-TOLLIP* haplotype blocks (rs35705950-rs111521887-rs5743894-rs5743854 and rs35705950-rs3750920-rs111521887-rs5743894), revealed associations with disease susceptibility and survival in and between the two conditions. Haplotype T-C-T-C of the former block, for instance, emerged as an independent protective factor in IPF survival (adjusted HR=0.37, 95% CI 0.17–0.78, *P*=0.009). On the other end, haplotype G-T-G-C of the latter block showed a significant decrease in fHP survival (adjusted HR=6.92, 95% CI 1.73-27.64, *P*=0.006). In addition, in a direct comparison between IPF and fHP patients, the G-C-C-T haplotype, composed of the ancestral alleles of each SNP, was associated with a better survival in fHP carriers than in IPF carriers (adjusted HR=2.06, 95% CI 1.07-3.97, *P*=0.031).

While revealing genetic similarities between IPF and fHP, the work presented in this thesis also unveiled distinctive genetic features with impact on risk and survival, as well as on lung function and radiographic features, highlighting the influence of genetic factors in these disorders. Altogether, these findings indicate that *TOLLIP* and *MUC5B* variants and haplotypes may serve as valuable tools for risk assessment, prognosis, and patient stratification in IPF and fHP.

Keywords**:** idiopathic pulmonary fibrosis; fibrotic hypersensitivity pneumonitis; common genetic variants; haplotypes

RESUMO

As doenças pulmonares intersticiais fibrosantes (DPIf) representam um subgrupo de entidades caraterizadas por um processo de fibrose autossustentável, associado a uma evolução progressiva e frequentemente fatal. A Fibrose Pulmonar Idiopática (FPI) é a DPIf mais reconhecida e com um prognóstico reservado, com uma sobrevivência mediana, quando não tratada, de 3-5 anos. Um comportamento semelhante é observado noutras DPIf não-FPI, como a Pneumonite de Hipersensibilidade fibrótica (PHf). A diferenciação precisa entre FPI e PHf é crucial, especialmente em casos com apresentações clínica e radiológica indeterminadas, onde biomarcadores patogénicos poderão contribuir na orientação diagnóstica.

Os estudos genéticos realizados no âmbito da FPI e PHf têm-se focado principalmente na pesquisa de variantes raras e comuns em vários genes, em particular nos polimorfismos de nucleotídeo único (*Single Nucleotide Polymorphisms* – SNPs) da *MUC5B* e *TOLLIP*, dois *loci* adjacentes no cromossoma 11. No entanto, esses estudos têm sido baseados na análise das variantes de forma independente. Uma abordagem mais ampla, que examine como essas variantes estão dispostas umas em relação às outras no contexto genómico, definindo haplótipos, não tem sido frequentemente abordada na literatura atual.

Esta tese consiste numa investigação realizada em três etapas destinada a avaliar o impacto independente dos SNPs da *MUC5B* e *TOLLIP* adjacentes, e seus respetivos haplótipos, na FPI e na PHf. O estudo envolveu a análise de doentes com formas esporádicas de FPI e PHf versus controlos saudáveis e comparações diretas entre ambas as doenças. O principal objetivo foi identificar associações genéticas que possam distinguir ou revelar semelhanças entre as duas DPIf, potencialmente fornecendo novos *insights* sobre os mecanismos moleculares subjacentes às mesmas.

A análise individual de cada SNP mostrou que o *MUC5B* rs35705950 associou-se à suscetibilidade tanto para FPI como para PHf. O alelo *minor* foi significativamente mais frequente nos doentes com FPI e PHf do que nos controlos. Em contraste, os SNPs do *TOLLIP* rs3750920, rs111521887 e rs5743894 associaram-se apenas à suscetibilidade para PHf, com uma frequência significativamente maior dos seus alelos *minor* nos doentes com PHf em comparação com os controlos. Comparações diretas entre os

doentes com FPI e PHf não revelaram diferenças significativas na distribuição alélica e genotípica desses SNPs.

A análise de haplótipos, definida por dois blocos específicos *MUC5B-TOLLIP* (rs35705950-rs111521887-rs5743894-rs5743854 e rs35705950-rs3750920-rs111521887-rs5743894), revelou associações com a suscetibilidade à doença e a sobrevivência nas duas condições e entre elas. O haplótipo T-C-T-C do primeiro bloco, por exemplo, emergiu como um fator de proteção independente na sobrevivência da FPI (HR ajustado=0,37, IC 95% 0,17–0,78, *P*=0,009). Por outro lado, o haplótipo G-T-G-C do segundo bloco mostrou uma diminuição significativa na sobrevivência da fHP (HR ajustado=6,92, IC 95% 1,73-27,64, *P*=0,006). Além disso, numa comparação direta entre doentes com FPI e PHf, o haplótipo G-C-C-T, composto pelos alelos ancestrais de cada SNP, associou-se a uma melhor sobrevivência nos doentes com PHf do que nos doentes com FPI portadores desse haplótipo (HR ajustado=2,06, IC 95% 1,07-3,97, *P*=0,031).

Apesar das semelhanças genéticas entre FPI e PHf, o trabalho apresentado nesta tese também revelou características genéticas distintas com impacto no risco e na sobrevivência, bem como na função pulmonar e nas características radiográficas, destacando a influência dos fatores genéticos nestas doenças. No seu conjunto, estes resultados indicam que as variantes e haplótipos do *TOLLIP* e *MUC5B* podem constituir ferramentas valiosas para avaliação de risco, prognóstico e estratificação de doentes na FPI e PHf.

Palavras-chave**:** fibrose pulmonar idiopática; pneumonite de hipersensibilidade fibrótica; variantes genéticas comuns; haplótipos

INTRODUCTION

DIFFUSE INTERSTITIAL LUNG DISEASES: GENERAL CONSIDERATIONS

Diffuse lung diseases, also known as interstitial lung diseases (ILDs), encompass a broad spectrum of over 200 disorders that mostly affect the pulmonary interstitium (1, 2). Damage to the lung is mediated by varying degrees of inflammation and fibrosis. These disorders are very heterogeneous regarding their etiology, clinical presentation, radiological and pathological features, as well as survival (3).

Available epidemiological data show a large diversity of ILDs incidence according to age, gender, ethnicity, and geographical location. Moreover, access to these data is biased by differences in study methodologies, access to health care, type of health records and disease coding. Among seventeen methodologically heterogeneous studies that examined the incidence, prevalence, and relative frequencies of ILDs, the incidence of ILD ranged from 1 to 31.5 per 100 000 person-years and prevalence ranged from 6.3 to 71 per 100 000 persons (4). Recently, the PERSEIDS study aimed at estimating incidence and prevalence of ILDs in general and in specific disorders, in six European countries, including Portugal. The overall (average of all countries) incidence and prevalence for ILDs in general ranged between 20.0 and 42.5 per 100 000 person-years and 72.1 and 164.2 per 100 000 persons, respectively (5).

ILDs can be classified into five broad categories: idiopathic (idiopathic pulmonary fibrosis, IPF), autoimmune-related (e.g. rheumatoid arthritis-associated interstitial lung disease, RA-ILD); systemic sclerosis-associated interstitial lung disease, SSc-ILD; others), exposure-related (e.g. hypersensitivity pneumonitis, HP), interstitial lung diseases with cysts or airspace filling and sarcoidosis (**Figure 1**) (1, 3) .

Multidisciplinary integration and discussion of clinical, radiological, and sometimes pathological data are the cornerstone of ILDs diagnosis and classification (6). However, about 15–20% of cases of fibrosing ILDs (fILDs) have an unclassifiable ILD (uILD), when current diagnostic criteria do not apply, even following extensive investigation and multidisciplinary discussion (1).

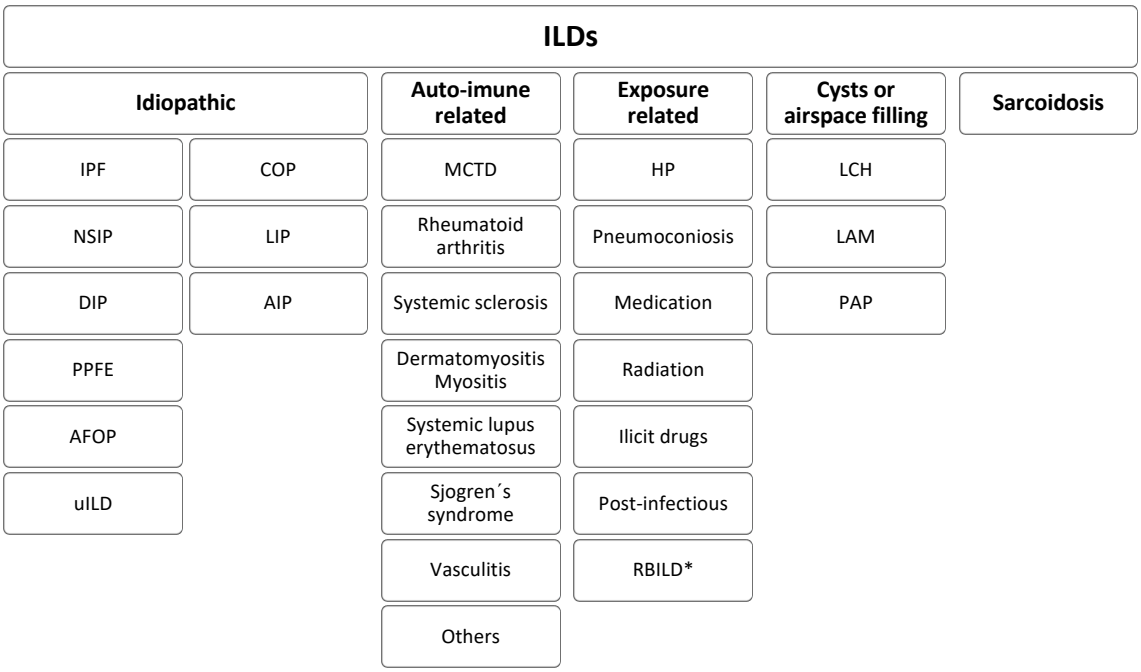


Figure 1. Classification of ILDs. AFOP, acute fibrinous and organizing pneumonia; AIP, acute interstitial pneumonia; COP, cryptogenic organizing pneumonia; DIP, desquamative interstitial pneumonia; HP, hypersensitivity pneumonitis; IPF, idiopathic pulmonary fibrosis; ILDs, interstitial lung diseases; LAM, lymphangioleiomyomatosis; LCH, Langerhans cell histiocytosis; LIP, lymphoid interstitial pneumonia; MCTD, mixed connective tissue disease; NSIP, nonspecific interstitial pneumonia; PAP, pulmonary alveolar proteinosis; PPFE, pleuroparenchymal fibroelastosis; RBILD, respiratory bronchiolitis-associated interstitial lung disease; uILD, unclassifiable interstitial lung disease. *Despite the clear association with smoking, RBILD was officially classified within the spectrum of idiopathic interstitial pneumonias, however there are propositions from certain authors advocating for its reconsideration and potential reclassification within exposure-related entities. Adapted from references 2 and 3.

The diagnosis work-up of this type of diseases, beyond the clinical description of specific pulmonary and extra-pulmonary symptoms and signs, requires a thorough investigation of domiciliary and/or occupational exposure to various antigens as well as pneumotoxic drugs or dusts, that can contribute to exposure-related disorders (7). Additionally, the possibility of a familial history must be considered due to the influence of genetic factors attributed to these diseases, mainly those with a fibrotic component. Thus, genetic testing may integrate the diagnosis, despite not being addressed in recent guidelines (2, 8). Nonetheless, a recent statement has emphasized the importance of conducting genetic sequencing, particularly in select cases and performed by highly specialized centers (9).

High-resolution computed tomography (HRCT) of the chest is the key diagnostic test for ILDs. In more than two-thirds of the patients, HRCT pattern, when paired with clinical observations, is often adequate to establish a definitive diagnosis and offers insights into the prognosis (3). The usual interstitial pneumonia (UIP) pattern is the hallmark of pulmonary fibrosis, observed frequently in IPF, in RA–ILD, and in advanced disease irrespective of the underlying condition. This pattern bears negative implications for disease prognosis (10, 11). Lung function and autoimmune serological tests also integrate the initial approach to diagnose ILDs.

The two invasive diagnostic tests available for ILD diagnosis include bronchoscopy with bronchoalveolar lavage (BAL) and lung biopsy. The cellular analysis of BAL can be useful in certain suspected conditions, such as HP (characterized by lymphocytosis), eosinophilic pneumonia or drug-induced lung injury (characterized by eosinophilia) or in some specific and rare diagnoses, such as pulmonary alveolar proteinosis or Langerhans cell histiocytosis (12, 13).

In cases lacking a clear diagnosis after evaluation of clinical, radiological and BAL data, histopathological assessment by surgical lung biopsy or, more recently, transbronchial lung cryobiopsy (TBLC), should be considered (14-17). Surgical lung biopsy remains the gold standard of the diagnosis; however, it carries inherent risks of postoperative complications, with a 30-day post-procedure mortality rate ranging between 1.5% to 2.4% (14, 15). TBLC has emerged as a less invasive alternative to surgical lung biopsy, showing a lower incidence of complications and a comparable level of diagnostic accuracy. This method is particularly effective when performed in specialized centers with expertise in executing the technique and interpreting pathological observations (18).

Clinically distinct ILDs exhibit considerable heterogeneity both in their underlying pathophysiology and clinical progression under different therapeutic approaches. Their evolution ranges from completely reversible self-limiting or to progressive, irreversible, and often fatal despite initial management (1, 10, 19). Evidence-based treatment recommendations are available for only a few ILDs and off-label use of drugs is common. Major advances have been made in the treatment of IPF with the advent of antifibrotics, pirfenidone and nintedanib, in detriment of immunosuppressive therapy (20-22). By

contrast, inflammatory ILDs (e.g. connective tissue disease-associated interstitial lung diseases (CTD-ILDs), HP, and sarcoidosis) often benefit from immunosuppressive therapy or other disease-specific therapies (23).

STATE-OF-THE-ART OF FIBROSING INTERSTITIAL LUNG DISEASES

Definition

Irrespective of the ILD, a subset of patients, in a variable proportion, experiences the development of self-sustaining lung fibrosis, leading to progressive disease marked by substantial morbidity and mortality. Recent literature has labeled this phenomenon as "the progressive fibrotic phenotype" (24), “progressive fibrotic interstitial lung disease” (25) and “progressive pulmonary fibrosis” (PPF) (2).

IPF is the archetype of idiopathic disorders and the most common fILD, defined by the histopathologic and/or radiologic pattern of UIP, with a poor prognosis reported by a median untreated life expectancy from diagnosis of 3–5 years (2, 8, 26). However, other non-IPF fILDs will ultimately manifest these features, with a subsequent outcome similar to untreated IPF patients (2, 10, 25, 27-32). Those include fibrotic hypersensitivity pneumonitis (fHP), idiopathic nonspecific interstitial pneumonia (iNSIP), CTD-ILDs, pneumoconiosis, drug-induced ILDs, uILD, sarcoidosis, and pleuroparenchymal fibroelastosis (PPFE) (**Figure 1**) (2).

PPF denotes fILDs displaying signs of clinical, radiological, or physiological progression despite optimal management directed at the underlying fILD. Approximately 18% to 32% of individuals diagnosed with non-IPF fILDs witness the onset of PPF within 61 to 80 months from the appearance of the initial symptoms (33).

Risk factors for progression can be categorized as general, such as the presence of a UIP pattern (23), or disease-specific, as exemplified by a lower baseline forced vital capacity (FVC), a decline in FVC of at least 10%, or the absence of specific antigen identification in fHP (34).

There have been numerous descriptions of PPF in recent years, leading to a lack of uniformity in the approach to these disorders. To address this, Raghu et al. (2022) recently proposed a diagnosis definition of a patient with ILD, of either known or unknown etiology other than IPF, demonstrating radiological evidence of pulmonary

fibrosis and presenting with at least two of the following three criteria within the past year (with no alternative explanation): 1) worsening respiratory symptoms; 2) physiological progression (absolute decline in FVC ≥ 5% of predicted value, or absolute decline in diffusing capacity of the lung for carbon monoxide (DLCO) ≥ 10% of predicted value); 3) radiological progression (2).

Epidemiology

Despite the negative impact of fibrosis on prognosis, epidemiological data available for fILDs are scarce (35-37). The PERSEIDS study, mentioned above, was the first to assess the epidemiology and progression of fILDs in parallel across six European countries, including Portugal (5). The overall (mean for all countries) incidence ranged between 2.1 and 6.3 per 100 000 person-years for IPF and 14.7 and 33.9 per 100 000 person-years for non-IPF fILDs. The overall prevalence ranged between 7.8 and 24.3 per 100 000 persons for IPF and 56.2 and 132.0 per 100 000 persons for non-IPF fILDs. In general, numbers for non-IPF fILDs surpassed those of IPF, perhaps due to the currently lower misdiagnosis of non-IPF fILDs as IPF, following the refinement of IPF guidelines, the development of guidelines for diagnosis of other fILDs and the improvement of diagnostic procedures in recent years (2, 16, 34, 38, 39). A third of non-IPF fILD cases showed a progressive fibrotic behavior, and nearly half of those (43%) had a UIP-like pattern. In this category, the subtype more prone to progression was fHP (5). These results were generally consistent with those previously reported (36, 37, 40), including in the PROGRESS study, a French retrospective and single-center study, that provided data in a real-word setting, which reported that 44% of patients had non-IPF fILDs and, among them, nearly 30% had a progressive fibrotic behavior. Concerning etiology, the largest groups affected were autoimmune ILDs (46.7%), uILD (31.5%), fHP (8.5%) and idiopathic interstitial pneumonia (IIP) (7.3%) (40).

According to PERSEIDS, in 2018, sarcoidosis (36.1%) and fHP (22.2%) were the most frequent fILDs in Portugal, the latter partly explained by the dimension of the cork industry and the widespread practice of pigeon breeding. However, these results may not represent the overall Portuguese reality, due to the low number and regional asymmetry of the included centers (2 from the North of Portugal, 1 from Coimbra and 1

from Lisbon) (5). Although data from national registries are not available, Santos et al. (2020) reported that, between 2012–2016 and in the North of Portugal, HP contributed for more than 20% of ILD diagnosis and was always the most frequently diagnosed ILD in all the years considered. From our experience, we believe that the Portuguese situation is analogous to that of those countries with higher incidence of HP, and fHP is probably the most frequent fILD representing the complexity of the differential diagnosis between IPF and fHP in our routine clinical practice (39, 41).

Idiopathic pulmonary fibrosis and fibrotic hypersensitivity pneumonitis: diagnostic, treatment, and prognostic particularities

Distinguishing between IPF and fHP is crucial despite the presence of overlapping clinical, radiological, and pathological features. Although evidence-based guidelines have been established for their diagnosis, the similarities between these entities make their differentiation a significant challenge (**Table 1**) (2, 34, 38, 39, 42). Marinescu et al. (2022) proposed an integrated approach of these guidelines to a single patient, based on algorithms, when both conditions are prominent diagnostic considerations (39). They attribute an indeterminate profile to those situations where both disorders are approximately equally likely in the clinical, radiological, and pathological domains.

From a clinical point of view, the identification of a causative antigen is crucial for fHP diagnosis (7, 43). However, exposition per se does not always mean HP, as it can be observed in other fILDs without a causative role. Moreover, a lack of exposure does not rule out the diagnosis of fHP, as in about 30-60% of cases the antigen remains hidden or unrecognized, negatively affecting disease prognosis (7, 44).

Lymphocytosis in BAL (especially > 30%) increases the likelihood of fHP compared with IPF, but this finding may lack sensitivity in patients with more established fibrosis. Patolia et al. (2020) reported that a threshold of 20% BAL fluid lymphocytes distinguished fHP from IPF with a sensitivity and specificity of 69% and 61%, respectively (12).

Regardless of the clinical phenotype, it is critical that all features are contextualized and integrated with radiologic findings. Although a UIP pattern, typified by basal-predominant subpleural reticulation, traction bronchiectasis, and honeycombing is

typical of IPF, it may also appear in fHP. On the other hand, the presence of small airways abnormalities (centrilobular nodularity or mosaic attenuation) in a fibrotic pattern background, with a mid or upper lung predominance, can also be seen in fILDs other than fHP, such as CTD-ILDs or IPF with obstructive airways disease. Thus, in an appropriate clinical context, imaging consistent with definite or probable UIP (without honeycombing) and typical of or compatible with fHP can lead to IPF and fHP diagnoses, respectively (2, 34, 38). However, in both diseases, we can have situations with an indeterminate radiologic pattern, where features are not suggestive of a specific disease or containing a mixture of features and/or distribution, which are more likely to be clarified by histopathologic analysis (39). When definite or probable UIP patterns are identified, there's a high likelihood of corresponding to a histological UIP pattern (2, 8). As these patterns often indicate features that negatively impact on the prognosis of fILDs (41, 45), some experts prefer to adopt this radiological classification for all fILDs, regardless of baseline disorder (46).

In the pathological context, identifying a UIP pattern in the biopsy is indicative of IPF but lacks specificity. A UIP pattern suggestive of IPF differs from a background UIP pattern with additional features indicating fHP, like notable peribronchiolar involvement, peribronchiolar metaplasia, and granulomas. Although UIP pattern generally exhibit more fibroblast foci per square centimeter compared to fHP pattern, no specific threshold is necessary as a diagnostic discriminator (47).

Randomized controlled trials in IPF patients have led to the development of antifibrotic drugs, as the mainstay of treatment, and have supported the notion that previous combined therapy with n-acetylcysteine (NAC) and immunosuppressants (prednisone and azathioprine) was deleterious for this disease (20-22). Santos et al. (2023) corroborated, through a real-life observational study, the positive impact of antifibrotics on the IPF clinical course and survival, with no difference in efficacy seen between nintedanib or pirfenidone (48). In fHP, antigen eviction represents the key initial intervention followed by immunosuppression. In the fHP context, evidence about efficacy of immunosuppressants is scarce and not so robust, although in recent years azathioprine and mycophenolate mofetil have been sustainably prescribed in these patients (49-51).

	IPF	Intermediate between IPF or fHP *(Indeterminate for both)*	fHP
Clinical features	Male Older age (> 60 years) Smoker or former smoker Restrictive physiology No identifiable antigen	Family history of fibrosis and/or predisposing genetic factors Dyspnea and cough Insidious onset of respiratory symptoms and absence of multisystemic manifestations No identifiable or indeterminate antigen Restrictive, obstructive, or mixed ventilatory pattern Negative or weakly positive autoimmunity BAL neutrophilia and absence of lymphocytosis	No sex predominance Any age Variable smoking habits Identified antigen (Worsening with re-exposure, may stabilize or improve with antigen avoidance) Restrictive physiology BAL lymphocytosis
Radiological features	*UIP pattern* Basal and subpleural predominant Reticular pattern and traction bronchiectasis ± honeycombing *AND* Absence of signs of small airways disease	Subtle reticulation not suggestive of a specific cause or a UIP pattern with atypical features (peribronchovascular involvement, sparing of costophrenic angles, ground glass) *AND* Not enough signs of small airways diseases to suggest fHP	*Fibrotic HP pattern* Mid or upper lung predominant with peribronchovascular involvement Reticular pattern and traction bronchiectasis ± honeycombing *AND* Presence of signs of small airways disease
Histopathological features	*UIP pattern* Dense fibrosis with architectural distortion Patchy Fibroblast foci Predominantly subpleural/paraseptal distribution *AND* No significant features of fHP	Features favoring a pattern other than UIP/IPF or a fibrosing process UIP-*like* in setting of other non-IPF diseases *AND* not enough ancillary features of fHP	A background fibrotic pattern (UIP, NSIP) *AND* Ancillary features of fHP ▪ Predominantly peribronchiolar fibrosis ▪ Peribronchiolar metaplasia > 50% bronchioles ▪ Poorly formed granulomas

Table 1. Clinical, radiological, and histopathological features typical of IPF and fHP and intermediate between both diseases. BAL, bronchoalveolar lavage; fHP, fibrotic hypersensitivity pneumonitis; IPF, idiopathic pulmonary fibrosis; NSIP, nonspecific interstitial pneumonia; UIP, usual interstitial pneumonia. Adapted from reference 39.

Despite fHP being historically associated with comparatively better outcomes, increasing evidence suggests that this entity may have similar rates of progression as IPF, and in this setting antifibrotics should be considered (28, 32, 52). Factors associated with a poor prognosis can be related to the antigen (type, time, and intensity of exposure), advanced age, smoking, and a radiological and/or histological pattern with fibrosis (7, 45, 53, 54).

The interpretation and amalgamation of clinical, radiological, and pathological data in these conditions pose considerable challenges, even after multidisciplinary discussion. Indeterminate clinical presentations and atypical morphological patterns remain particularly problematic, situations in which diagnostic approaches with pathogenic biomarkers might offer valuable insights.

Pathogenesis

The pathogenesis of pulmonary fibrosis, notably in IPF, results from a complex interplay between external factors (such as environmental exposure, smoking, infections, and mechanical stress) and individual predispositions (genetic variation) (55). Alveolar epithelial cells (AECs) and fibroblasts play central roles in this process. Chronic damage to the AECs triggers an epithelial-driven activation of fibroblasts, leading to the formation of fibroblast/myofibroblast clusters and an excessive deposition of extracellular matrix. Specific cytokines and pro-fibrotic growth factors, such as tumor necrosis factor-alpha (*TNF-*α), platelet-derived growth factor (*PDGF*), insulin-like growth factor (*IGF*), and transforming growth factor-beta (*TGF-*β), released by the activated or injured AECs, are crucial in differentiating fibroblasts into myofibroblasts and in promoting epithelial-mesenchymal transition, which suggests the transformation of epithelial cells into fibroblast-like cells followed by their activation. An imbalance between pro-fibrotic and anti-fibrotic factors contributes to an abnormal and excessive healing response in the lung tissue (55-57).

Non-IPF fILDs, in contrast to IPF, exhibit an initial inflammatory phase that varies among conditions, transitioning into a fibrotic phase that tends to overlap in later stages, shaping the disease progression (**Figure 2**) (10). The initial phase involves lymphocyte activation and differentiation, heightened immune responses in immune-related

disorders (CTD-ILDs), and persistent granulomatous inflammation triggered by various external agents (e.g. HP, drug-induced reactions, or sarcoidosis).

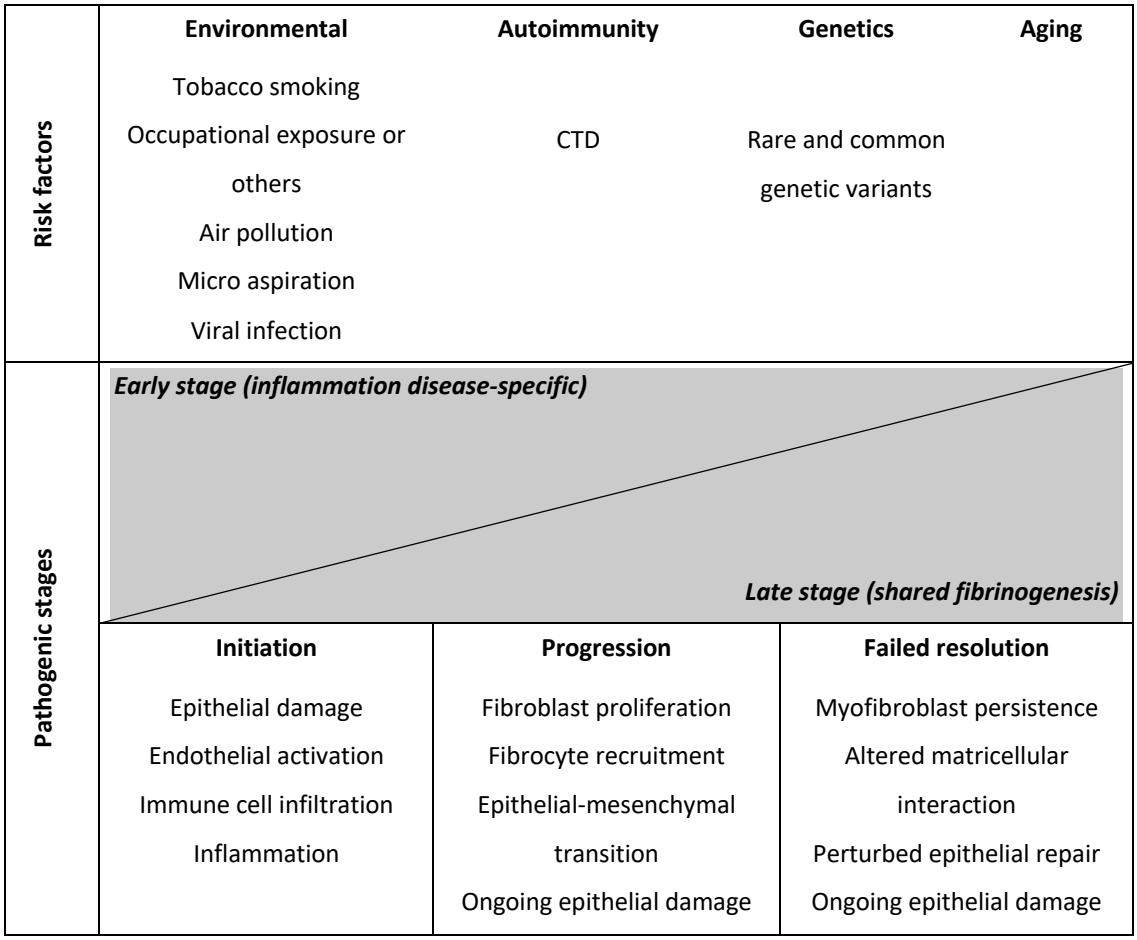


Figure 2. Pathogenesis of non-IPF fILDs. CTD, connective tissue disease; non-IPF fILDs, non-idiopathic pulmonary fibrosis fibrosing interstitial lung diseases.

As seen in IPF, multiple environmental factors contributing to repeated injury to alveolar cells, alongside intrinsic factors such as aging, genetic predispositions, and epigenetic changes, are also involved in the development and progression of non-IPF fILDs (3, 10). The initial inflammatory responses across these conditions may trigger a pro-fibrotic environment, initiating a later phase marked by shared mechanisms, regardless of the initial diagnosis. This phase involves the production of extracellular matrix by fibroblasts and remodeling of lung tissue, mirroring the processes observed in IPF. Tissue stiffness and hypoxia arising from this remodeling further enhance pro-fibrotic cytokine pathways and the activation of myofibroblasts, thus perpetuating the process of fibrosis

(58). Indeed, fibrosis arises as a lung response to both internal and external agents, guided by shared mechanisms despite distinct initial triggers. Yet, the precise factors that differentiate normal wound repair from the progression to fibrosis remain not fully understood (58).

In more detail, HP is induced by a wide variety of antigens, which can be divided into six broad categories: bacteria, fungi, mycobacteria, animal and plant proteins, chemicals, and metals, represented by disease prototypes (7). At baseline, it is a typical inflammatory disease, where presence and activity of T and B lymphocytes are crucial for progression from sensitization to antigen to disease development. It requires the accumulation of CD4+ T-helper 1 (Th1) cells in the lung, creating a pro-inflammatory microenvironment. As the suppressive activity of regulatory T cells is impaired, amplification of the inflammatory response occurs. Interferon-gamma (*IFN-γ*) and *TNF-*α promote the accumulation, activation, and aggregation of macrophages, resulting in the development of granulomatous inflammation. Also, immune complex-mediated lung injury with specific immunoglobulin G (IgG) antibodies may contribute to the inflammatory response (59). Progression to fibrosis occurs in the presence of factors that hamper the resolution of inflammation, such as maintenance of antigen exposure, cigarette smoking, genetic susceptibility, and other unidentified factors (7, 44, 60). The fibrotic phase of HP is characterized by a switch from a predominant Th1-like to a Th2-like phenotype. Consequently, a gradual increase in CD4+ T cells and CD4+/CD8+ ratio and decline of CD8+ T cells occurs. Th2 cells secrete, among others, interleukin-4 (*IL-4*) and mainly *IL-13* that contribute to a fibrotic milieu stimulating the *TGF-β1* signaling pathway and activating the expansion of the fibroblast population (61). Fibroblasts arrive at the injured areas and differentiate into myofibroblasts, which are responsible for the accumulation of extracellular matrix, in a process similar to IPF (59, 62).

Genetics can definitely play a crucial role in the pathogenesis of both IPF and non-IPF fILDs, such as fHP. Identifying these genetic factors could significantly enhance the accuracy of diagnosis and improve management strategies for these diseases, mainly in those with indeterminate clinical and radiological profiles.

GENETICS IN SUSCEPTIBILITY AND PROGNOSIS OF FIBROSING INTERSTITIAL LUNG DISEASES

The genetic component of pulmonary fibrosis involves a mix of rare and common variants, each with distinct impacts on disease susceptibility and prognosis. Recent years have seen significant advancements in the understanding of the role of genetic variants through diverse studies like genome-wide association studies (GWAS), candidate gene analyses, and familial inheritance research. These studies have revealed candidates involved in critical processes such as host defense, telomere maintenance, cell adhesion, and signaling, elucidating their role in the risk and outcome of fILDs, both in their familial and sporadic forms (2, 63-67) (**Tables 2 and 3**). However, while GWAS have pinpointed numerous loci linked to fILD susceptibility, they primarily identified common variants, leaving undiscovered rare alleles, non-coding variants, or sequence variants of unknown effect (68, 69).

Patients with a family history of fILD have a higher likelihood of inheriting rare pathogenic variants compared to those without such a family history (25% versus 5%) (70). These rare variants, characterized by a minor allele frequency (MAF) of < 1%, although much less frequent wield higher penetrance and a more pronounced effect size (resembling a monogenic-like presentation). They encompass variants in genes associated with telomere biology and those implicated in maintaining alveolar stability (**Table 2**) (71-73).

In contrast, common genetic variants (MAF > 1%) contribute to a larger portion of disease risk in patients with pulmonary fibrosis, albeit with generally smaller effect size than rare variants (70). The challenge with these common variants is that individually they may not provide enough information to accurately assess an individual's risk of developing disease. In complex diseases like IPF, it is either the cumulative effect of multiple risk variants or the interaction between genes and environment that contributes the most to disease susceptibility (74). To address this challenge, researchers have proposed the application of Polygenic Risk Scores (PRS), also for IPF (75, 76). Some of these common variants, notably the single-nucleotide polymorphism (SNP) rs35705950 within the mucin 5B (*MUC5B*) gene promoter and specific SNPs in the

toll-interacting protein (*TOLLIP*) gene, both on chromosome 11, have been associated with susceptibility and/or survival in IPF and non-IPF fILDs (**Table 3**) (64, 77-81).

The subsequent sections will delve into a more comprehensive examination of these variants and their influence on both familial and sporadic forms of fILDs, focusing particularly on IPF and fHP.

Genetics of familial and sporadic pulmonary fibrosis

Although there is no consensus definition, a recent statement has defined familial pulmonary fibrosis (FPF) as any instance with fILD in at least two blood related first or second degree family members (9). The presence of clear familial inheritance, associated with a single gene variant segregating with the presence of ILD, strongly supports the conclusion that ILD in these cases is a monogenic disorder. An autosomal dominant mode of inheritance with incomplete penetrance (not all carriers of the pathogenic variants developing disease) occurs in 80% of cases (69, 82). On the other hand, patients with fILDs without a known family history are deemed to have sporadic pulmonary fibrosis.

Patients with FPF can manifest a spectrum of pulmonary fibrosis phenotypes, mostly IPF (20-25%) but also non-IPF fILDs, such as fHP (14-17%), uILD (15%) and CTD-ILD (3-8%) and even relatives within the same family can manifest different pulmonary fibrosis subtypes (81, 83-85). Such heterogeneity suggests that different forms of fILDs might share common genetic origins and disease pathways, but additional genetic or environmental factors act on a common genetic background to drive the disease towards different phenotypic manifestations (86). FPF patients often exhibit similarities in clinical presentation, radiographic features, and histopathological findings to those seen in sporadic fILDs cases. However, FPF cases tend to manifest at a younger age, have a lower prevalence of males and smokers, and typically present with higher FVC at the time of diagnosis (87-89).

FPF is usually progressive and associated with a poor prognosis, irrespective of baseline disease. Median survival from diagnosis is estimated as 2.4-7.3 years (9, 84, 85, 87). Cutting et al. (2021) found a significantly worse survival in familial forms of IPF and non-IPF fILDs compared with their sporadic counterparts; those with familial IPF showed an

80% increased risk of death or transplantation (hazard ratio [HR], 1.8; 95% CI, 1.37-2.37; *P*<0.001), whereas those with familial non-IPF fILDs showed a twofold increased risk relative to sporadic disease (HR, 2.08; 95% CI, 1.46-2.96; *P*<0.001) (85). However, this prognostic difference was not supported by previous studies (87, 88).

The pathogenic genetic variants more commonly associated with FPF can be grouped according to the underlying affected physiological mechanism: telomere biology (25-35%), surfactant homeostasis (3-8%) and syndromic forms, such as Dyskeratosis congenita (DC) or *Hermansky-Pudlak* syndrome (3-5%). However, most suspected monogenic pulmonary fibrosis cases remain without an identified genetic cause (60%) (90, 91).

Rare telomere-related and surfactant-related genes variants

Telomeres consist of hexamer repeats (CCCTAA) located at the ends of chromosomes and serve as protective caps during cell replication. In FPF, rare variants within telomere-related genes (TRGs) were identified in telomerase reverse transcriptase (*TERT*) (15%), regulator of telomere elongation helicase (*RTEL1*) (5-10%), poly(A)-specific ribonuclease (*PARN*) (5%) and telomerase RNA component (*TERC*) (3%) genes and, more rarely (<1%), in dyskerin pseudouridine synthase 1 (*DKC1*), nuclear assembly factor 1 ribonucleoprotein (*NAF1*), TERF1 interacting nuclear factor 2 (*TINF2*), NHP2 ribonucleoprotein (*NHP2*), NOP10 ribonucleoprotein (*NOP10*) and zinc finger CCHC-type containing 8 (*ZCCHC8*) (**Table 2**) (9, 69, 70, 92-96). The penetrance in carriers depends on several factors, including age, gender, environmental exposures, and telomere length (TL) (97). The prototypic telomeropathy is the DC which has been associated with some TGRs variants, transmitted through various modes of inheritance (X-linked, autosomal recessive, and autosomal dominant), and extreme telomere shortening (98). Classically, it is characterized by abnormal skin pigmentation, oral leukoplakia, and nail dystrophy; bone marrow failure occurs in > 80% of DC cases and is the leading cause of death (99). Pulmonary fibrosis develops in approximately 20% of these patients, contrary to adult cases with heterozygous TRGs variants, in which it constitutes the most common manifestation. Extrapulmonary manifestations can also coexist with fibrosis in these patients or relatives, such as premature grey hair, anemia, macrocytosis,

thrombocytopenia, and elevated liver enzymes (100, 101). The pulmonary fibrosis phenotype may be variable, IPF being the most frequent (nearly 50%) followed by fHP (7%-12%), uILD (8%-20%), CTD-ILD (2%-3%), and other IIP (14%-18%) (83, 97, 102). In sporadic fILDs, rare TRGs variants are present in about 10% of IPF, fHP, and RA-ILD cases (73, 103, 104). Short age-adjusted TL is found in about 20% to 60% of IPF (105), 20% to 35% of fHP (81), and 26% of RA-ILD sporadic cases (84).

In patients with TRGs variants, TL shortens from generation to generation due to the transmission of short telomeres, independently of the transmission of the variant. This instance constitutes an epigenetic phenomenon, which explains the occurrence of phenocopies (an affected patient without the familial variant but with telomere shortening – 50% of cases) and genetic anticipation (earlier onset and more severe disease in each subsequent generation) (106, 107). Here, pulmonary fibrosis is an age-dependent phenotype, rare in carriers less than 40 years of age but found in 60% of 60-year-old men or older (107). Despite the heterogeneity of the pulmonary fibrosis phenotype, TRGs variant carriers experience uniformly progressive disease and poor survival (97). TL might prompt the screening for unknown TRGs variants or phenocopies when patients inherited very short telomeres from their parents (108). Shorter TL is associated with a poor outcome in IPF and other fILDs and may also influence the prognosis after pharmacological treatment or lung transplantation (81, 109-112).

As already mentioned, antifibrotics are the first-line treatment for IPF and immunosuppressants have been a common therapeutic option for non-IPF fILDs. However, there have been scientific data that point out to the potentially harmful effects of these drugs in carriers of TRGs variants or shortening TL. Adegunsoye et al. (2021) showed that mycophenolate mofetil was associated with a mortality benefit for patients with fHP and longer TL, which was not reported in those with short TL (113). Zhang et al. (2023) explored the association between time-dependent exposure to mycophenolate or azathioprine and transplant-free survival in non-IPF fILD patients stratified by age-adjusted TL in percentiles. Immunosuppressant exposure was associated with higher mortality risk in fHP and uILD patients with age-adjusted TL $<10^{th}$ percentile, a subgroup where this treatment should be cautiously prescribed (112).

Gene	*Locus*	Function	Diseases	References
Telomere-related genes				
TERT	5p15.33	Encodes a component of telomerase (the reverse transcriptase)	IPF, IPAF, HP, CTD-ILD, PPFE, NSIP, DIP, CPFE	(68, 69, 73, 97, 102, 114-116)
TERC	3q26.2	Encodes a component of telomerase (RNA)	IPF, HP, CTD-ILD, PPFE, CPFE	(69, 97, 102, 114, 115, 117)
DKC1	Xq28	Telomerase complex template stabilization	IPF	(69, 114, 115, 118)
RTEL1	20q13.33	DNA helicase activity	IPF, IPAF, HP, PPFE	(68, 69, 97, 102, 114, 119, 120)
PARN	16p13.12	Stability of mRNA	IPF, IPAF, HP	(69, 97, 114, 121)
TINF2	14q12	Maintenance of telomeres	IPF	(69, 114, 122)
NAF1	4q32.2	Maintenance of telomeres	CPFE	(92)
NOP10	15q14	Maintenance of telomeres	IPF	(93)
ZCCHC8	12q24.31	Maintenance of telomeres	IPF	(94)
Surfactant-related genes				
SFTPC	8p21.3	Encodes a protein of surfactant C	IPF	(69, 91, 115, 123)
SFTPA1	10q22.3	Encodes a protein of surfactant A1	IPF	(69, 115, 124, 125)
SFTPA2	10q22.3	Encodes a protein of surfactant A2	IPF	(69, 91, 115, 125, 126)
ABCA3	16p13.3	Lipid transport across plasma membrane	IPF, CPFE, DIP	(69, 91, 115, 127-129)

Table 2. Rare genetic variants associated with fILDs. CPFE, combined pulmonary fibrosis and emphysema; CTD-ILD, connective tissue disease-associated interstitial lung disease; DIP, desquamative interstitial pneumonia; fILDs, fibrosing interstitial lung diseases; HP, hypersensitivity pneumonitis; IPAF, interstitial pneumonia with autoimmune features; IPF, idiopathic pulmonary fibrosis; mRNA, messenger ribonucleic acid; NSIP, nonspecific interstitial pneumonia; PPFE, pleuroparenchymal fibroelastosis.

Thus, in patients with short TL or rare variants in TRGs, whether presenting with IPF or a non-IPF fILD, the use of antifibrotic therapies constitutes a safe and effective option to slow progression of the disease (130). Considering the young age of most patients, lung transplantation is often an option. In reported series of patients with *TERT* and *TERC*

variants submitted to lung transplantation, almost all patients required adjustment of immunosuppression due to haematological toxicity (111, 131, 132). In these cases, immunosuppressive drugs must be more closely monitored and tapered, and azathioprine and mycophenolate mofetil should be avoided (90). Rare variants within surfactant-related genes (SRGs) have been described in surfactant protein C (*SFTPC*) (1-25%) and less commonly in *SFTPA1/2* (< 1%) and ATP binding cassette subfamily A member 3 (*ABCA3*) (<1%) of FPF cases. These variants are rarely encountered in sporadic IPF (**Table 2**) (69, 83, 91, 95, 96, 123, 126).

From a clinical perspective, a unique hallmark of families that harbor a rare SRGs variant is the variable age of disease onset, from infancy to late adulthood. As surfactant production occurs in the lung, extrapulmonary manifestations are absent. *SFTPA1* and *SFTPA2* are associated with both pulmonary fibrosis and lung adenocarcinoma and their presence should prompt diligent lung cancer surveillance (125, 133, 134). The fibrosis phenotype is also variable; unclassifiable fibrosis, cysts and ground glass opacities are the most frequent radiological features, and UIP, NSIP and DIP are the pathologic patterns observed more often (91, 125, 135).

No aggregated data on disease evolution and treatment outcome have been presented for these SRGs. Treatment with corticosteroids, azithromycin or hydroxychloroquine may prove beneficial (9, 136). Recently, Klay et al. (2023) compared a SRGs mutation group with sporadic IPF and FPF and reported a similar median FVC decline 6 months after diagnosis and median transplant-free survival among them. Moreover, patients with a SRGs variant treated for 6 months with antifibrotics showed stable lung function, whereas patients with an SRGs mutation treated with immunosuppression displayed a variable response dependent on the gene involved, speculating about the potential benefit of antifibrotics in this subtype of variants (137).

Genetic testing

Genomics-based testing options encompass genetic sequencing and TL measurement (95). Genetic sequencing aims to identify specific variants within risk genes, achieved through methods like whole genome, whole exome, or panel sequencing. The latter focuses on sequencing a specific subset of genes previously associated with FPF, making

it a preferred option due to its cost-effectiveness and efficiency. Therefore, clinical genetic testing in FPF kindred focuses primarily on identifying rare heritable TRGs and SRGs variants (9). Genetic variants identified through gene sequencing are reported according to standards outlined by the American College of Medical Genetics: 1) pathogenic; 2) likely pathogenic; 3) uncertain significance (VUS); 4) likely benign; and 5) benign (138). These designations are based on various factors, including the functional consequence of the variant, its prevalence within the population, previous associations with disease, and its co-segregation with disease in families. This test is associated with a low positivity rate (30%) which may be due to the result of multiple factors, including yet to be identified genetic variants that contribute to disease risk, additive effects of lower effect-size variants, or the effect of genetic plus certain environmental exposures (98). Other factors include inheritance patterns (autosomal dominant, autosomal recessive, or X-linked), incomplete penetrance, variable expressivity, uncertain genotype-phenotype correlations, potential phenocopies, and genetic anticipation resulting from telomere shortening (9). Its interpretation requires the involvement of clinicians with genetic expertise who can effectively explain the implications of the test results to both patients and their family members.

A recent European Respiratory Society (ERS) statement on FPF suggested offering genetic sequencing in specific clinical scenarios: 1) Patients with fILD having one or more first or second-degree family members with fILD; 2) Patients with a relative carrying a pathogenic/likely pathogenic variant known to cause ILD; 3) Patients suspected of having short telomere syndrome; and 4) Patients with an idiopathic fILD before the age of 50 (9).

The Local Health Unity of São João, in Porto, proposes genetic testing utilizing next-generation sequencing (NGS) of a panel comprising 33 genes associated with FPF. This panel includes genes such as *ABCA3, AP3B1, ASAH1, CSF2RA, CSF2RB, DKC1, ELMOD2, ERF, FAM111B, FARSB, FOXF1, GBA, HPS1, HPS4, IL1RN, ITGA3, MARS1, NKX2-1, NSMCE3, PARN, PLCG2, RTEL1, SFTPA1, SFTPA2, SFTPB, SFTPC, SFTPD, SLC34A2, STAT3, STING1, TERC, TERT, and TINF2*. Upon obtaining test results, an approach aligning with international recommendations is adopted (9). In the event of a pathogenic variant being identified, genetic counseling is extended to all first-degree family members.

Moreover, genetic testing is offered to these family members to facilitate comprehensive assessment and management (**Figure 3**).

Pairing genetic testing with TL measurement is considered an ideal approach, especially in suspected telomere-related diseases. This combination enables a more comprehensive assessment of potential telomere dysfunction by offer additional insight into the potential pathogenicity of an identified telomere-related VUS, for instance (9, 108). Clinical TL measurements typically involve the assessment of peripheral blood mononuclear cells or DNA extracted from these cells using various laboratory methods. Among these, flow-fluorescence in situ hybridization (flow-FISH) is often considered the most reproducible. TL is commonly expressed as a relative length compared to a control group, with adjustments made for age in order to account for age-related changes (139).

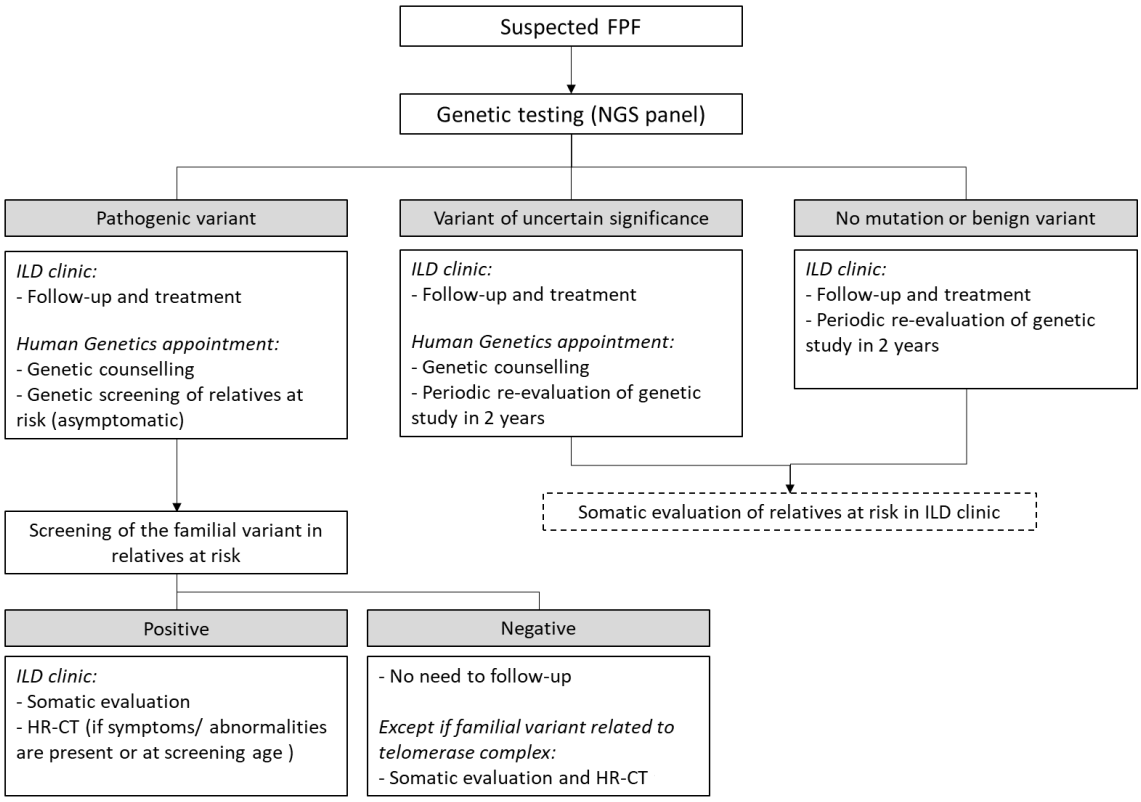


Figure 3. Clinical algorithm for management of patients and relatives in the setting of suspected FPF applied in the Local Health Unit of São João. FPF, familial pulmonary fibrosis; HRCT, high-resolution computed tomography; ILD, interstitial lung disease; NGS, next generation sequencing.

Genotyping of common genetic variants is not currently part of the standard genetic diagnostic workup for FPF. However, some groups advocate screening for the *MUC5B* promoter variant in FPF cases, though its high prevalence and greater size effect (despite a common variant), especially when variants typically associated with FPF as a monogenic disease (e.g. telomerase/surfactant-related) are not identified (90).

Zhang et al. (2022) reported the potential multidimensional value of whole genome sequencing in detecting rare and common variants and estimate TL. It can delineate genetic-mediated risk for pulmonary fibrosis, evaluate the functional impact of TRGs variants by estimating TL, and predict clinically relevant disease outcomes. As costs for this technology decrease, its accessibility and potential utility in clinical settings are expected to grow in the future, making it an increasingly attractive option for understanding and managing fILDs (70).

Concerning the sporadic forms of fILDs, despite sharing common and rare variants with FPF, there aren't large multicenter studies supporting the clinical value of such screening in this subgroup of patients. The multi-ancestry of the variations found, the polygenic behaviour of these diseases and the absence of clearly defined clinical decision-making guidelines based on genetic results pose barriers to the widespread adoption of genetic testing in the routine care of this subgroup of patients. Nonetheless, genotyping of some polymorphisms, namely of genes *MUC5B* and the adjacent *TOLLIP*, could prove beneficial and clinically relevant, as it could enable stratification based on the risk of developing lung fibrosis and provide crucial prognostic information, namely in IPF and fHP.

COMMON GENETIC VARIANTS IN IDIOPATHIC PULMONARY FIBROSIS AND FIBROTIC HYPERSENSITIVITY PNEUMONITIS

In fILDs, the *MUC5B* gene promoter rs35705950 SNP has been an extensively studied variant, gathering substantial attention in research extending beyond IPF to other diseases. It has been linked to an increased risk of interstitial lung abnormalities (ILAs) among the general population, potentially aiding in the identification of individuals with preclinical forms of fILD (76, 140). This variant is also associated with other fILD types,

including RA-ILD (141), sporadic fHP (142), and asbestosis (143), whereas it hasn´t been associated with scleroderma-ILD, myositis-ILD or sarcoidosis (80, 144). Moreover, IPF patients carrying the *MUC5B* risk allele are less likely to present with TRGs variants (70, 145). Second in the list are several SNPs in *TOLLIP* (namely rs5743890, rs5743894, rs111521887 and rs3750920), which have been another focal point of investigation, particularly in the context of IPF, where the bulk of available data resides (64, 69). *MUC5B* encodes one of several mucin-producing genes, which facilitate airway clearance and function to maintain immune homeostasis (77), whether *TOLLIP* encodes toll-interacting protein, which inhibits toll-like receptor (TLR) signaling and acts as a critical mediator of airway host defense (146) (**Figure 4**) (**Table 3**).



Figure 4. NCBI Genome Data Viewer representation of the studied SNPs in the *MUC5B* and *TOLLIP* genes region on Chr. 11 (NC_000011.10:1,217,210 - 1,315,278) Assembly: GRCh38p.14 (GCF_000001405.40)

Idiopathic pulmonary fibrosis

The majority of GWAS have been performed in patients with IPF, revealing SNPs within at least 23 loci associated with disease susceptibility (95). Among these, the *MUC5B* gene stands out as the strongest and most consistently replicated genetic risk factor (64-66, 147-149). Seibold et al. (2011), first evaluated the *MUC5B* promoter rs35705950 polymorphism as a risk factor for IPF, showing that the minor allele T was present in 38% of IPF patients, 34% of FPF patients, but only in 9% of healthy controls. This allele was associated with a 14-fold increased expression of *MUC5B* in the lung and had the most substantial impact on pulmonary fibrosis risk. Homozygotes for the T allele had 20-times increased risk of developing both sporadic and familial disease, whereas in heterozygotes the odds ratio (OR) was 9 for IPF and 7 for FPF (77). Thereafter, numerous types of studies were carried out corroborating the higher proportion of this variant, mainly in Caucasian IPF patients (68, 80, 150-154).

Despite the robustness of the *MUC5B* association with sporadic and familial forms of the disease, only a small proportion of carriers of this variant in the general population will develop fibrotic disease, suggesting that genetic predisposition is probably necessary, yet not sufficient, to underlie the pathogenetic pathways. Interestingly, even individuals without the *MUC5B* variant show increased *MUC5B* production in pulmonary fibrosis cases. It is conceivable that unexplored genetic variants, potentially within the less accessible repetitive mucin regions, closely linked with rs35705950, could be the actual genetic contributors. However, the precise mechanism by which increased *MUC5B* expression contributes to the disease remains unclear. It is speculated that excessive mucin production might hinder mucus clearance, prevent alveolar repair, and interfere with the host's defense mechanisms (77).

Notably, *TOLLIP* gene variants have also been linked to IPF susceptibility. Noth et al. (2013) identified three distinct SNPs - rs111521887, rs5743894, and rs5743890 - in the genomic region of *TOLLIP* showing different frequencies between IPF patients and control groups. Specifically, the minor alleles of intronic SNPs rs111521887 and rs5743894 were associated with increased susceptibility to IPF, while the minor allele of rs5743890 was observed to mitigate this risk (64).

TOLLIP is mainly expressed in the lung by alveolar type II cells, macrophages, and basal cells, where it can prevent damage from reactive oxygen species induced by mitochondrial damage. Moreover, it can protect bronchial cells and modulate inflammation and autoimmunity via different pathways (TLR, *TNF-*α, *IL-1*, *IFN-*β, *IL-13*) (155). The genetic variants within the *TOLLIP* gene could potentially have an impact on its function or expression, disrupting its regulatory role in the immune response within the lung tissue. The discernible reduction in *TOLLIP* expression among carriers of these specific *TOLLIP* SNPs underscores the potential relevance of this gene in the context of this disease (146).

In addition to influencing IPF susceptibility, SNPs within *MUC5B* and *TOLLIP* seem to have an impact on disease survival as well.

Gene	SNP	*Locus*	Function	Diseases	References
Airway mucin production					
MUC5B	*rs35705950*	11p15.5	Encodes the major component of mucus secretions	IPF, HP, RA-ILD, ILAs	(64, 65, 69, 77, 78, 81, 140-142, 145, 150, 151, 153, 154, 156-159)
MUC2	*rs7934606*	11p15.5	Mucin production	IPF	(65, 69, 159)
Toll-like receptor signaling					
TOLLIP	*rs111521887* *rs5743894* *rs5743890*	11p15.5	Encodes a protein involved in the signaling pathway of IL1 and TLR	IPF	(64, 69)
TLR3	*rs3775291*	4q35	Influences recognition of pathogens and activation of innate immunity	IPF	(69, 160)
ATP11A	*1278769*	13q34	Phospholipid translocation	IPF	(64-66, 69)
Cell-cell adhesion					
DSP	*rs2076295*	6p24	Cell adhesion, integrity, and mechanotransduction	IPF	(64, 65, 69)
DPP9	rs12610495	19p13.3	Cell adhesion, integrity, and mechanotransduction	IPF	(65, 66, 69)
Cytokine/growth factor signaling					
IL1RN	*rs408392* *rs419598*	2q14.1	Inhibits proinflammatory effect of IL-1a and IL-1b	IPF	(69, 114, 115)
IL8	*rs4073* *rs2227307*	4q13	Proinflammatory cytokine	IPF	(69, 114, 115)
TGB1	*rs1800470*	19q13.2	Controls proliferation and differentiation of diverse cell types	IPF	(69, 114, 115)
Cell cycle regulation					
TP53	*rs12951053* *rs12602273*	17p13.1	Production of tumor suppressor protein p53	IPF	(114, 115)
CDKN1A	*rs2395655*	6p21.2	Critical in cell cycle and response to DNA damage	IPF	(114, 115)
Telomere maintenance					
TERT	*rs2736100*	5p15.33	Telomerase complex enzyme that maintains telomere length	IPF, CPFE	(65, 69, 114, 115)
OBFC1	*rs11191865*	10q25.1	Stimulates DNA polymerase-alpha-primase activity	IPF	(64, 65, 69, 114, 115)
Others					
HLA-DRB1	*rs2395655*	6p21.2	Major histocompatibility complex, critical to immune system	IPF, HP, CTD-ILD	(10, 68, 69, 115, 161, 162)

Table 3. Common genetic variants associated with fILDs. CPFE, combined pulmonary fibrosis and emphysema; CTD-ILD, connective tissue disease-associated interstitial lung disease; fILDs, fibrosing interstitial lung diseases; HP, hypersensitivity pneumonitis; ILAs, interstitial lung abnormalities; IL, interleukin; IPF, idiopathic pulmonary fibrosis; NSIP, nonspecific interstitial pneumonia; RA-ILD, rheumatoid arthritis- associated interstitial lung disease; SNP, single nucleotide polymorphism; TLR, toll-like receptor.

Peljto et al. (2013) showed that the *MUC5B* rs35705950 minor allele was paradoxically associated with a twofold decrease in mortality risk, while increasing the IPF susceptibility risk (79). This survival effect is controversial with some recent studies replicating this better outcome (163, 164), while others showing the opposite effect or neither association (153, 165). On the other hand, the carriers of the minor allele C of *TOLLIP* rs5743890 SNP reported an increased mortality, despite the lower risk of disease development (64, 165). In addition, Oldham et al. (2015) showed that *TOLLIP* rs3750920 was associated with a favorable response to NAC, as carriers of the TT genotype (25% of IPF) treated with NAC had a significantly reduced composite endpoint risk, including death, hospitalization, and FVC decline (166). Results from the "Prospective Treatment Efficacy in IPF Using Genotype for NAC Selection trial (PRECISIONS)", (https://clinicaltrials.gov/study/NCT04300920), which seeks to address whether NAC has a differential effect on lung fibrosis progression depending on *TOLLIP* gene variants, are eagerly awaited in order to shed light on this possible genetic-treatment interaction. These opposite effects in risk and survival align with the results of a recent GWAS performed by Allen et al. (2023), which revealed that a genetic variant located in an antisense RNA gene for protein kinase N2 was associated with IPF progression, but not with the risk of disease development.

This suggests that genetic factors conferring susceptibility to IPF may be different from those driving its prognosis. This concept also applies to *MUC5B* rs35705950, wherein the minor T allele is associated with both an increased risk of IPF and a more favorable prognosis (167).

Fibrotic hypersensitivity pneumonitis

Studies assessing genetic susceptibility in fHP are scarce and genomic studies have failed to completely explain heritability of the disease (168). However, evidence suggests that Human Leucocyte Antigen (*HLA*) gene polymorphisms, particularly within the major histocompatibility complex (*MHC*) class II, play a role in HP susceptibility. Notably, *HLA-DR3* has been implicated in Caucasians, and specific *HLA-DR7* and *HLA-DRB1/DQB1* haplotypes have been implicated in Hispanic patients with pigeon breeder's disease (161, 169). Freitas et al. (2023) reported distinct profiles of *TNF-α* and *HLA* polymorphisms among non-fibrotic and fHP subtypes, potentially aiding in predicting disease course and refining treatment strategies. Among fHP patients, increased frequencies of *HLA A**02 (46.7% vs 25.7%; OR=2.53, *P*=0.02) and HLA *DRB1**14 (10.0% vs 0.7%; OR=15.44, *P*=0.02) were observed compared to controls. Non-fibrotic/inflammatory HP patients exhibited significantly higher frequencies of the *TNF-α* GG genotype compared to patients with fibrotic presentations (88% vs 60%; relative risk (RR)=0.44; *P*=0.04) and to control individuals (88% vs 63%, OR 4.33, *P*=0.037). Additionally, the former patients displayed a markedly increased frequency of the G allele (94.0% vs 73.3%, RR=0.44, *P*=0.01), while fHP patients predominantly presented the A allele (26.7% vs 6.0%, RR=2.28, *P*=0.01) (170).

The *MUC5B* promoter variant and TL have been identified as determinants of susceptibility and survival in fHP (81, 103, 171). A study carried out by Ley et al. (2017) analyzed the *MUC5B* rs35705950 and *TOLLIP* rs5743890 SNPs in two independent HP cohorts. Results showed a significantly higher MAF of the *MUC5B* SNP in both HP cohorts (24.4% and 32.3%) compared to healthy control subjects (10.7%; *P*<0.0001). Notably, unlike in IPF, the *MUC5B* promoter SNP was associated with a higher risk of mortality among individuals with HP. Shorter TL has been correlated with various degrees of fibrosis and reduced survival among HP patients (81). Conversely, no discernible association was found between the *TOLLIP* SNP rs5743890 and either susceptibility or survival (81). While no study has specifically investigated the involvement of different *TOLLIP* SNPs, previously associated with IPF, in the context of fHP, recent research has focused on the impact of this gene in disease presentation. For instance, Katayanagi et al. (2022) reported that among Japanese fHP patients genotyped for *TOLLIP* rs3750920

and rs5743899, the GG (minor allele) genotype of the latter was associated with rapid FVC deterioration over time. However, the relationship between rs3750920 genotypes and FVC at baseline was not specified (172).

An intriguing aspect lies in the seemly independent and controversial roles that the *MUC5B* and *TOLLIP* genes play in conferring susceptibility and/or impacting on the prognosis of both IPF and fHP, despite residing in very close proximity on the short arm of chromosome 11. While their individual contributions have been explored, with heterogeneous results, the potential for interaction between these genes hasn't been ruled out. The challenge with these common variants is that individually they may not provide enough information to accurately assess an individual's risk of developing disease. To address this critical issue, it is reasonable to explore the cumulative effect of these variants together and their interaction with environment. As already mentioned, researchers have already developed the PRS, including for IPF (75, 76). It is essentially an aggregated measure of an individual's genetic risk based on multiple common variants associated with the disease. This aggregated genetic risk, quantified by PRS, can then be used to justify early diagnostic screening and other genetic tests for individuals at higher risk of developing IPF. It helps in identifying individuals who may benefit from closer monitoring or preventive measures even before symptoms of the disease manifest.

Studies exploring a possible interplay or mutual influence between *MUC5B* and *TOLLIP* are really necessary for a comprehensive understanding of their effects on disease susceptibility and prognosis both in IPF and fHP.

In summary, the diagnosis of fILDs remains a complex and challenging task, despite heightened diagnostic awareness, enhanced quality of HRCT, the availability of new and safer diagnostic techniques, and the gold-standard of multidisciplinary discussion.

A precision medicine strategy, centered on identifying genetic biomarkers holds promise in improving early diagnosis, reducing misdiagnosis rates, offering mechanistic insights into the diseases, personalizing radiological and functional follow-up protocols, and shedding light on prognosis, thereby informing tailored management strategies. This approach is particularly crucial in non-IPF fILDs, such as fHP, where immunosuppression

remains pivotal in specific cases. With this rationale in mind, the work supporting this thesis was developed with specific aims and focusing on IPF and fHP within the realm of fILDs. Its main goal was to probe and identify potential genetic biomarkers of phenotypic expression in the sporadic forms of these both challenging disorders.

THESIS AIMS

The application of precision medicine in the clinical practice of fILDs remains an unmet challenge. Within our routine clinical practice, distinguishing between sporadic forms of IPF and fHP poses a frequent dilemma. In this context, the discovery of distinctive signatures of common genetic variants is not only necessary but imperative as they may offer invaluable information for differentiating both diseases and guiding treatment strategies.

This investigation aimed to explore the role of specific SNPs in IPF and fHP, by examining allelic, genotypic and haplotypic associations with susceptibility, prognosis, and clinical outcomes.

A three-step strategy was planned, each step aligned with a specific objective:

1. To identify potential associations of SNPs of four candidate genes (*MUC5B*, *TOLLIP*, *PLAU*, *SERPIN*) with susceptibility and survival in Portuguese patients with IPF (*first study, published*)
2. To identify potential associations of SNPs in *MUC5B* and *TOLLIP* with susceptibility, clinical presentation, and survival in Portuguese patients with fHP (*second study, published*)
3. To directly compare genetic signatures of *MUC5B* and *TOLLIP* SNPs among IPF and fHP patients, seeking to identify potential risk and survival associations that could discriminate the two fibrosing diseases or, conversely, to identify shared pathophysiologic mechanisms among them (*third study, submitted to publication*).

PUBLICATIONS

FIRST STUDY

PREDICTIVE VALUE OF COMMON GENETIC VARIANTS IN IDIOPATHIC PULMONARY FIBROSIS SURVIVAL

Mota PC, Soares ML, Vasconcelos CD, Ferreira AC, Lima BA, Manduchi E, Moore JH, Melo N, Novais-Bastos H, Pereira JM, Guimarães S, Moura CS, Marques JA, Morais A

J Mol Med (Berl). 2022 Sep;100(9):1341-1353. doi: 10.1007/s00109-022-02242-y. Epub 2022 Aug 19. PMID: 35986225

Journal of Molecular Medicine
https://doi.org/10.1007/s00109-022-02242-y

JMolMed

ORIGINAL ARTICLE

Predictive value of common genetic variants in idiopathic pulmonary fibrosis survival

Patrícia Caetano Mota[1,2,4] · Miguel Luz Soares[2,3,4] · Carlos Daniel Vasconcelos[2,3,4] · António Carlos Ferreira[2,3,4] · Bruno A. Lima[5] · Elisabetta Manduchi[6] · Jason H. Moore[7] · Natália Melo[1] · Hélder Novais-Bastos[1,2,4] · José Miguel Pereira[8] · Susana Guimarães[9] · Conceição Souto Moura[9] · José Agostinho Marques[1,2,4] · António Morais[1,2,4]

Received: 14 February 2022 / Revised: 13 July 2022 / Accepted: 2 August 2022
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

Abstract
Idiopathic pulmonary fibrosis (IPF) is a chronic, progressive, fibrosing interstitial pneumonia of unknown etiology. The role of genetic risk factors has been the focus of numerous studies probing for associations of genetic variants with IPF. We aimed to determine whether single-nucleotide polymorphisms (SNPs) of four candidate genes are associated with IPF susceptibility and survival in a Portuguese population. A retrospective case–control study was performed with 64 IPF patients and 74 healthy controls. Ten single-nucleotide variants residing in the MUC5B, TOLLIP, SERPINB1, and PLAU genes were analyzed. Single- and multi-locus analyses were performed to investigate the predictive potential of specific variants in IPF susceptibility and survival. Multifactor dimensionality reduction (MDR) was employed to uncover predictive multi-locus interactions underlying IPF susceptibility. The MUC5B rs35705950 SNP was significantly associated with IPF: T allele carriers were significantly more frequent among IPF patients (75.0% vs 20.3%, $P < 1.0 \times 10^{-6}$). Genotypic and allelic distributions of TOLLIP, PLAU, and SERPINB1 SNPs did not differ significantly between groups. However, the MUC5B-TOLLIP T-C-T-C haplotype, defined by the rs35705950-rs111521887-rs5743894-rs5743854 block, emerged as an independent protective factor in IPF survival (HR = 0.37, 95% CI 0.17–0.78, $P = 0.009$, after adjustment for FVC). No significant multi-locus interactions correlating with disease susceptibility were detected. MUC5B rs35705950 was linked to an increased risk for IPF, as reported for other populations, but not to disease survival. A haplotype incorporating SNPs of the MUC5B-TOLLIP locus at 11p15.5 seems to predict better survival and could prove useful for prognostic purposes and IPF patient stratification.

Key messages
- The MUC5B rs35705950 minor allele is associated with IPF risk in the Portuguese.
- No predictive multi-locus interactions of IPF susceptibility were identified by MDR.
- A haplotype defined by MUC5B and TOLLIP SNPs is a protective factor in IPF survival.
- The haplotype may be used as a prognostic tool for IPF patient stratification.

Keywords Idiopathic pulmonary fibrosis · TOLLIP · MUC5B promoter · Haplotypes · IPF survival · Common genetic variants

Introduction

IPF is a rare form of chronic, progressive, fibrosing interstitial pneumonia of unknown cause, defined by the histopathologic and/or radiologic pattern of usual interstitial pneumonia [1–3]. Current theory suggests that lung parenchyma destruction and alveolar replacement by dense fibrotic tissue are promoted by signaling to various cell types, predominantly fibroblasts and macrophages, following initial damage

Carlos Daniel Vasconcelos and António Carlos Ferreira contributed equally to this work.

✉ Patrícia Caetano Mota
patcmota@gmail.com

Extended author information available on the last page of the article

Published online: 19 August 2022

to the alveolar epithelium [1]. Other pathways underlying inflammation and oxidative stress have also emerged as playing a role in the fibrosing process [3–5].

The genetic component of IPF has been the focus of numerous studies. Genome-wide association and familial studies have uncovered candidates involved in host defense, telomere maintenance, cell–cell adhesion, and cell signaling, among other processes [6–9]. Rare genetic variants have been identified, including mutations in genes implicated in alveolar stability, and genes related to telomere biology [10–12]. Among common variants, a single-nucleotide polymorphism (SNP) in the promoter of the mucin 5B (MUC5B) gene (rs35705950) and specific SNPs of the toll-interacting protein (TOLLIP) gene, both on the short arm of chromosome 11, have been shown to associate with disease susceptibility and/or survival [7, 13–15]. Nonetheless, the variants identified thus far account for only about 30–35% of the genetic risk and additional contributing genetic determinants remain elusive [13, 16]. Importantly, it remains to be determined whether a combinatorial effect of specific variants may influence susceptibility to IPF and its severity.

The objective of the current study was to determine whether polymorphisms in two new candidate genes, urokinase-type plasminogen activator (PLAU) and serpin family B member 1 (SERPINB1), as well as the MUC5B and neighboring TOLLIP SNPs, are associated—independently or through a combined effect—with IPF susceptibility and survival in a Portuguese population.

Materials and methods

Participants and study design

A retrospective case–control study was undertaken with 64 IPF patients, followed from 2013 to 2018, in a tertiary Portuguese hospital in Porto. A control group with 74 healthy subjects (similar age distribution and gender proportions) was selected from EPIPorto, a population-based cohort of adults [17].

Patient diagnosis conformed to the criteria of established international guidelines at the time of inclusion [2, 18, 19]. Treatments were administered according to IPF practice guidelines available at the time of prescription and re-evaluated according to their updates [18–20].

Demographic, clinical, and lung function data (forced vital capacity, FVC; total lung capacity, TLC; and diffusion capacity of lung for carbon monoxide, DLCO; expressed as percentages of predicted normal values) were collected. Treatments were scored at the end of follow-up.

The study protocol was reviewed and approved by the Ethics Committee of the Institution and performed in accordance with the Helsinki declaration. Written informed consent was obtained from all participants.

Genotyping

Genomic DNA was extracted from 5 ml of peripheral blood using the Wizard Genomic DNA Purification Kit (Promega Corp., Madison, USA) or by standard phenol–chloroform extraction. Genotyping was performed by restriction fragment length polymorphism (RFLP) analysis of PCR amplicons. DNA sequences containing the polymorphic sites were amplified by PCR in 20 µl reactions containing 20 ng of genomic DNA, 1.5 mM of each primer, and Kapa Taq EXtra HotStart ReadyMix (Merck, Darmstadt, Germany). Cycling conditions were 95 °C denaturing for 5 min, 35 cycles of 30 s at 95 °C, 30 s at annealing temperature, and 30 s at 72 °C, with a final extension for 10 min at 72 °C. The amplicons were digested overnight in 15 µl reactions with 8 µl of PCR product and 3 units of restriction endonuclease. Primer sequences, annealing temperatures, restriction enzymes, and digestion product sizes are listed in Table S1 in Online Resource file.

Ten SNPs were analyzed: a missense variant of PLAU (rs2227564, exon 6); three variants within SERPINB1 [rs408265 and rs386713 (5'-regulatory region); rs57721448 (intronic)]; five variants within TOLLIP [rs5743854 (5'-regulatory region); rs5743890, rs5743894 and rs111521887 (intronic); rs3750920 (synonymous variant)]; and the variant in the MUC5B promoter (rs35705950). Alleles are reported in the forward strand orientation, except rs5743854 in conformity with the SNP database (dbSNP) minor allele designation.

Statistical analysis

Data are presented as mean (± standard deviation) if continuous or as frequencies (percentages) if categorical.

SNP allele and genotype frequencies were determined by direct counting. Goodness of fit of allele frequencies to Hardy–Weinberg equilibrium was analyzed using a chi-squared test. Chi-squared or Fisher exact tests, when appropriate, were used to assess allele and genotype association with disease susceptibility. Haplotype frequencies were estimated and cross-validated using PHASE v2.1.1 and Haploview v4.2 programs [21–23]. Pairwise linkage disequilibrium between SNPs was calculated using SNPAlyze v8.1.1 (Dynacom Co., Japan).

Survival was analyzed with time from diagnosis to death (in months) censured by end of follow-up. Kaplan–Meier curves were used for survival estimates and compared with a log-rank test. Univariate and multivariate cox proportional hazards regressions were used to identify variables that predict survival status. A final Cox regression model with forward conditional stepwise selection was obtained. The proportional hazards assumption for the final Cox regression model was also tested.

Two-sided *P*-values lower than 0.05 were considered statistically significant. *P*-values presented throughout the manuscript were not corrected for multiple comparisons, except

where noted. All analyses were performed within R Studio, a computing environment for R programming language.

With IPF susceptibility as the target outcome, 1-, 2-, and 3-way models were explored involving as features the 10 SNPs, gender, and smoker status using multifactor dimensionality reduction (MDR), a machine learning approach described elsewhere [24, 25]. Default settings, including tenfold cross-validation (CV), were employed. Balanced accuracy was used as a measure of test performance in the MDR analysis (default). To probe for interactions among the analyzed features, we utilized the permutation-based MDR explicit test for interactions described in Greene et al. [26].

Results

Study subjects

Demographic features of studied subjects and clinical and lung function data from IPF patients are shown in Table 1. No significant differences were observed between groups regarding gender, age, and smoking history. Overall mean follow-up duration, defined as time from IPF diagnosis to death from any cause or end of study, was 47.1 (±27.2) months. Thirty-six patients (56.2%) died during follow-up.

Table 2 lists the pharmacological agents administered to IPF patients. Antifibrotic (AF) treatment, with pirfenidone or nintedanib, was administered to 35 patients (54.7%), 28 (80%) of whom for at least one year. Five patients (7.8%) received N-acetylcysteine (NAC) monotherapy and 20 (31.3%) received immunosuppressants (IS), before antifibrotics became the recommended treatment for IPF. Four patients (6.3%) had no treatment or had it suspended due to drug intolerance. Thirty-seven patients (57.8%) also received proton pump inhibitor (PPI) therapy.

Polymorphisms and IPF susceptibility

The distribution of allele and genotype frequencies of the probed SNPs among patients and controls is shown in Table 3. Genotype distributions all conformed to the Hardy–Weinberg equilibrium (Table S2 in Online Resource file).

Table 1 Clinical and demographic characteristics of IPF patients and controls

	Patients ($N=64$)	Controls ($N=74$)	P-value
Gender			0.559[a]
Female	22 (34.4%)	29 (39.2%)	
Male	42 (65.6%)	45 (60.8%)	
Age (years)	70.3 (±9.6)	73.1 (±7.0)	0.058[b]
Smoking			0.508[a]
Non-smoker	28 (43.8%)	39 (52.7%)	
Ex-smoker	34 (53.1%)	32 (43.2%)	
Smoker	2 (3.1%)	3 (4.1%)	
Follow-up duration (months)	47.1 (±27.2)		
Number of deaths	36 (56.2%)		
FVC (% pred)	83.4 (±16.9)		
TLC (% pred)	77.1 (±15.5)		
DLCO (% pred)	46.1 (±17.5)		
Comorbidities			
COPD	2 (3.1%)		
Pulmonary hypertension	21 (32.8%)		
Lung cancer	6 (9.4%)		
OSA	13 (20.3%)		
GERD	10 (15.6%)		
Metabolic conditions	37 (57.8%)		
Cardiovascular disease	34 (53.1%)		
Depression	10 (15.6%)		

Values are presented as mean (±SD) and counts as n (%)

% pred percentage of predicted normal value, *IPF* idiopathic pulmonary fibrosis, *FVC* forced vital capacity, *TLC* total lung capacity, *DLCO* diffusing capacity of the lung for carbon monoxide, *COPD* chronic obstructive lung disease, *OSA* obstructive sleep apnea, *GERD* gastro-esophageal reflux disease

[a]Chi-squared test

[b]t-test

Table 2 Pharmacological agents administered to IPF patients

	Patients ($N=64$)
Antifibrotics	35 (54.7%)
Immunosuppressants*	20 (31.3%)
NAC monotherapy*	5 (7.8%)
No treatment	4 (6.3%)
PPI	37 (57.8%)

Counts (%) at end of follow-up

Treatments were administered according to IPF management practice guidelines available at the time of prescription

IPF idiopathic pulmonary fibrosis, *NAC* N-acetylcysteine, *PPI* proton pump inhibitor

*Up until antifibrotics became the recommended treatment for IPF

The MUC5B rs35705950 SNP revealed a significant association with IPF: the frequency of minor allele T was significantly higher in IPF patients than in controls (45.3% vs 11.5%, OR = 6.38, 95% CI 3.46–11.79, $P < 1.0 \times 10^{-6}$). Accordingly, the genotypes of heterozygous and homozygous individuals carrying the T allele were significantly more frequent in the IPF group (75% vs 20.3%, OR = 11.80, 95% CI 5.30–26.28, $P < 1.0 \times 10^{-6}$), indicating this variant allele also confers susceptibility to IPF in the Portuguese population. Statistically significant *P*-values on genotypic and allelic distribution analyses remain so, even after Bonferroni's correction for multiple testing.

No differences were found in genotypic and allelic distributions between IPF patients and controls for any of the five TOLLIP SNPs, nor the SNPs in SERPINB1 and PLAU genes.

Multi-locus interactions (MDR analysis)

Average CV testing balanced accuracy (TBA) of all models containing the MUC5B SNP was consistently above 0.7, but they were not statistically significant in the explicit test for interactions. Additionally, in models not containing this SNP, there was a sharp drop in average CV TBA indicating that the MUC5B SNP was the main driver for the association with IPF susceptibility. The explicit test for interactions did not detect any significant multi-locus, gender-locus, or smoker-locus interaction. When exploring data using the information gain metrics described by Moore and Hu [27], a weak evidence of a synergistic interaction between the SERPINB1 SNPs rs386713 and rs408265 was stated. However, this interaction was not significant and was correlated with the MUC5B SNP main effect, and thus did not contribute any information for prediction (not shown).

Polymorphisms, haplotypes, and IPF survival

Kaplan–Meier analyses of MUC5B, TOLLIP, SERPINB1, and PLAU genotypes showed statistically significant associations with survival for TOLLIP SNPs rs111521887 and rs5743894 (log-rank test *P*-values < 0.001), and close to significance for rs5743854 ($P = 0.096$) (Fig. 1). Patients with genotypes G/G, C/C, and G/G, respectively, had worst survival than those without these genotypes.

Haplotypes defined by these three SNPs and the neighboring MUC5B polymorphism [5′-rs35705950(M)-rs111521887(T)-rs5743894(T)-rs5743854(T)-3′] were reconstructed and five major haplotypes were identified (Table S3 in Online Resource file). Among these, haplotypes T-C-T-C and T-G-C-C were significantly more frequent and G-C-T-C was significantly less frequent in patients than in controls. In the IPF group, the TOLLIP SNPs are in strong linkage disequilibrium (LD) with each other, and moderate LD with the MUC5B SNP, as indicated by the pairwise LD measures shown in Table S4, Fig. S1 in Online Resource file.

Kaplan–Meier analysis of the haplotypes revealed a strong statistically significant association of haplotype T-C-T-C with increased survival (log-rank test *P*-value = 0.0004) (Fig. 2). The association remained significant at a 5% level after Bonferroni's correction for multiple comparisons ($P = 0.0128$).

Patients carrying the T-C-T-C haplotype had significantly longer time to death and fewer deaths compared to non-carriers ($P = 0.017$ and $P = 0.006$, respectively); all other parameters did not vary (Table 4).

To evaluate the influence of the T-C-T-C haplotype on the survival of IPF patients on AF treatment, we have analyzed patients treated with these drugs for at least one year. Kaplan–Meier analysis of these patients ($n = 28$) revealed a trend for better survival of T-C-T-C haplotype carriers in response to AF treatment (HR = 0.31, 95% CI 0.07–1.47, log-rank test *P*-value = 0.122) (Fig. 3). No difference in the frequency of patients on AF treatment was observed between carriers and non-carriers of the haplotype (Table S5). Conversely, survival analysis probing the MUC5B SNP alone (GG genotype vs. TT/TG genotypes) in patients treated with antifibrotics did not show any association (HR = 0.76, 95% CI 0.15–3.98, log-rank test *P*-value = 0.745) (Fig. S2 in Online Resource file).

Haplotypes defined by the three SERPINB1 SNPs, which were in high linkage disequilibrium with one another, revealed no significant associations with either IPF susceptibility or survival (not shown).

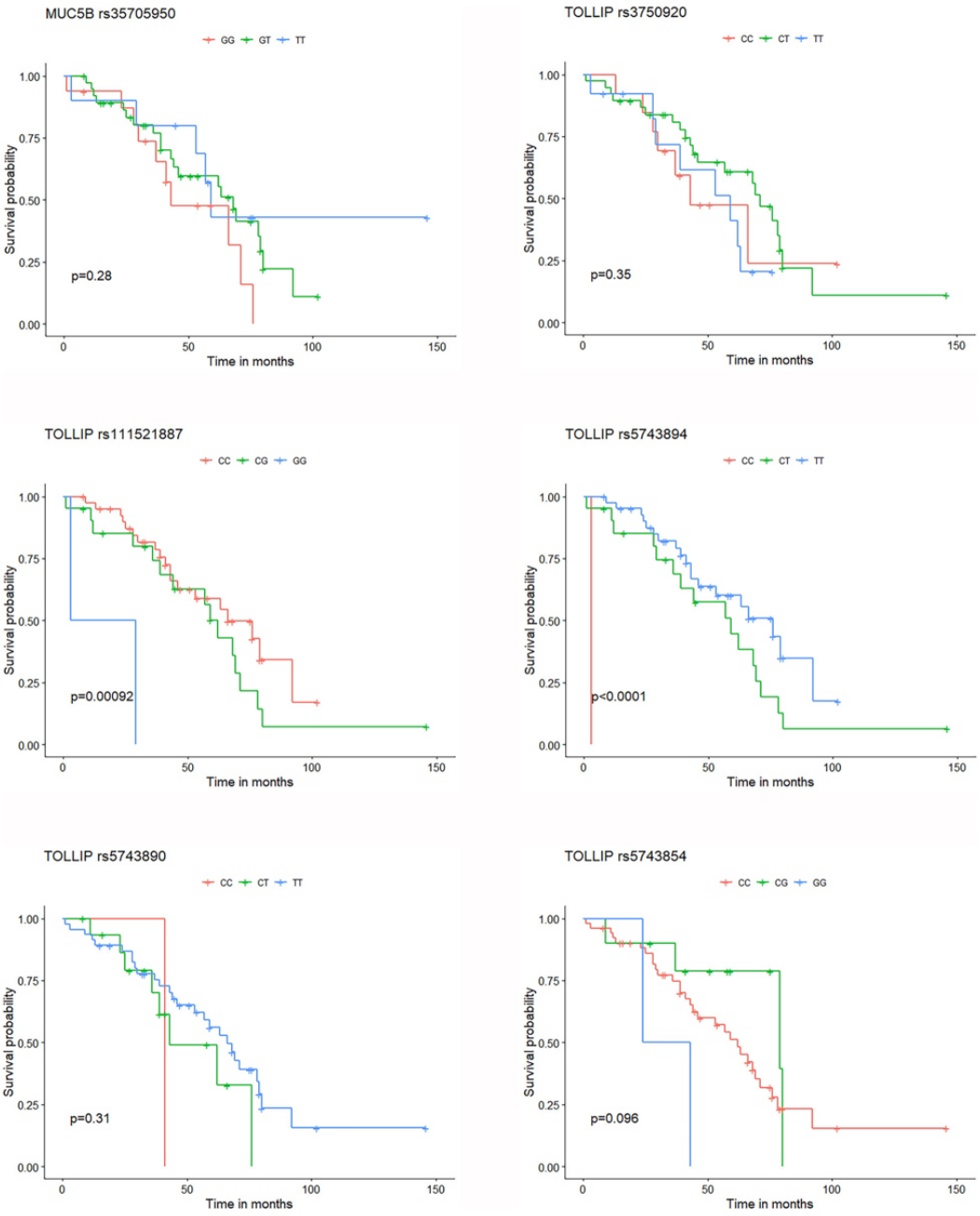


Fig. 1 Kaplan–Meier curves showing IPF patients' survival according to MUC5B rs35705950, TOLLIP rs3750920, TOLLIP rs111521887, TOLLIP rs5743894, TOLLIP rs5743890 and TOLLIP rs5743854 genotypes. Time in months; *P*-values for log-rank test

Uni- and multivariate analysis for predictors of survival

To further explore the role of demographic, clinical, and genetic factors as predictors of survival, univariate and multivariate regression analyses were employed. Univariate analyses using Cox regression did not reveal any association between gender, age, smoking, comorbidities, or AF treatment and survival (Table 5). Conversely, better pulmonary function (FVC, TLC and DLCO) associated

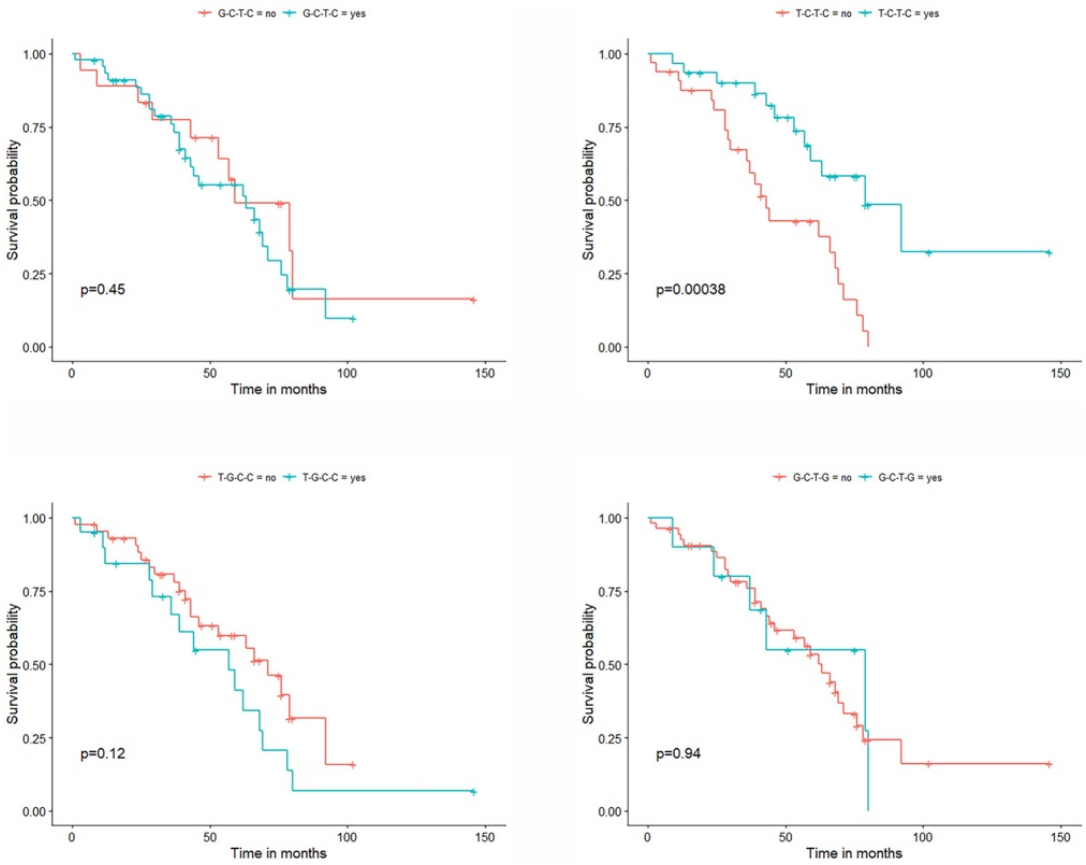


Fig. 2 Kaplan–Meier curves showing IPF patients' survival according to MUC5B rs35705950-TOLLIP rs111521887-TOLLIP rs5743894-TOLLIP rs5743854 haplotypes. Time in months; *P*-values for log-rank test

significantly with longer survival (HR < 1, $P = 4.79 \times 10^{-4}$, $P = 0.001$ and $P = 0.038$, respectively). Three multivariate Cox regression models were built for TOLLIP SNPs rs111521887, rs5743894, and rs5743854, adjusting for FVC (Table 5). Only the rs5743854 SNP attained statistical significance, with the G/G genotype showing an association with worst survival, compared to the C/C genotype (HR = 9.27, $P = 0.006$). The TOLLIP rs111521887 G/G genotype tended to behave as a risk factor for worst survival compared to C/C (HR = 5.99, $P = 0.092$); whereas the rs5743894 T/T genotype tended to show a protective effect (HR = 0.58, $P = 0.113$). Statistically significant p-values on univariate regression analyses remain so, even after Bonferroni correction for multiple testing.

No prognostic association was found for the remaining TOLLIP SNPs or any variants of the PLAU, SERPINB1, and MUC5B genes, when analyzed independently.

Haplotypes defined by the M950-T887-T894-T854 block, on the other hand, were independently associated with decreased mortality under multivariate analyses. The best fitting multivariate Cox regression model built with forward conditional stepwise selection and considering demographic and clinical covariates, showed that the T-C-T-C haplotype was a strong predictor of better survival (HR = 0.37, 95% CI 0.17–0.78, $P = 0.009$, after adjustment for FVC) (Table 5). The T-C-T-C haplotype and FVC associations remained equally significant (HR = 0.96, 95% CI 0.94–0.99, $P = 0.008$ and HR = 0.37, 95% CI 0.17–0.78, $P = 0.009$, respectively) after adjustment for treatment with antifibrotics as a time-dependent covariate (HR = 0.84, 95% CI 0.38–1.88, $P = 0.679$) (Table S6 in Online Resource file).

Discussion

Our study shows that the MUC5B promoter SNP rs35705950 is significantly associated with IPF susceptibility in the Portuguese, as documented in other populations [28]. Several

Table 3 Genotype and allele distribution of MUC5B, TOLLIP, SERPINB1 and PLAU SNPs in IPF patients and controls

	Genotype			Allele	
MUC5B					
rs35705950	**GG** (%)	**GT** (%)	**TT** (%)	**G** (%)	**T** (%)
IPF	16 (25.0)	38 (59.4)	10 (15.6)	70 (54.7)	58 (45.3)
Control	59 (79.7)	13 (17.6)	2 (2.7)	131 (88.5)	17 (11.5)
	–	OR 10.78 (4.66–24.91)	OR 18.44 (3.67–92.77)	–	OR 6.38 (3.46–11.79)
P	**–**	$\mathbf{<1.0\times10^{-6}}$	$\mathbf{1.32\times10^{-5}}$	$\mathbf{<1.0\times10^{-6}}$	
TOLLIP					
rs3750920	**CC** (%)	**CT** (%)	**TT** (%)	**C** (%)	**T** (%)
IPF	13 (20.3)	38 (59.4)	13 (20.3)	64 (50.0)	64 (50.0)
Control	23 (31.1)	41 (55.4)	10 (13.5)	87 (58.8)	61 (41.2)
P	–	0.230	0.124	0.144	
rs111521887	**CC** (%)	**CG** (%)	**GG** (%)	**C** (%)	**G** (%)
IPF	41 (64.1)	21 (32.8)	2 (3.1)	103 (80.5)	25 (19.5)
Control	56 (75.7)	17 (23.0)	1 (1.4)	129 (87.2)	19 (12.8)
P	–	0.173	0.401	0.130	
rs5743894	**TT** (%)	**TC** (%)	**CC** (%)	**T** (%)	**C** (%)
IPF	42 (65.6)	21 (32.8)	1 (1.6)	105 (82.0)	23 (18.0)
Control	57 (77.0)	16 (21.6)	1 (1.4)	130 (87.8)	18 (12.2)
P	–	0.136	0.830	0.176	
rs5743890	**TT** (%)	**TC** (%)	**CC** (%)	**T** (%)	**C** (%)
IPF	46 (71.9)	17 (26.6)	1 (1.6)	109 (85.2)	19 (14.8)
Control	49 (66.2)	23 (31.1)	2 (2.7)	121 (81.8)	27 (18.2)
P	–	0.529	0.607	0.450	
rs5743854	**CC** (%)	**CG** (%)	**GG** (%)	**C** (%)	**G** (%)
IPF	52 (81.3)	10 (15.6)	2 (3.1)	114 (89.1)	14 (10.9)
Control	57 (77.0)	16 (21.6)	1 (1.4)	130 (87.8)	18 (12.2)
P	–	0.395	0.517	0.851	
SERPIN					
rs57721448	**GG** (%)	**GT** (%)	**TT** (%)	**G** (%)	**T** (%)
IPF	53 (82.8)	10 (15.6)	1 (1.6)	116 (90.6)	12 (9.4)
Control	55 (74.3)	19 (25.7)	0 (0.0)	129 (87.2)	19 (12.8)
P	–	0.162	0.311	0.364	
rs386713	**TT** (%)	**TC** (%)	**CC** (%)	**T** (%)	**C** (%)
IPF	25 (39.1)	31 (48.4)	8 (12.5)	81 (63.3)	47 (36.7)
Control	28 (37.8)	40 (54.1)	6 (8.1)	96 (64.9)	52 (35.1)
P	–	0.698	0.507	0.784	
rs408265	**CC** (%)	**CT** (%)	**TT** (%)	**C** (%)	**T** (%)
IPF	24 (37.5)	29 (45.3)	11 (17.2)	77 (60.2)	51 (39.8)
Control	22 (29.7)	38 (51.4)	14 (18.9)	82 (55.4)	66 (44.6)
P	–	0.352	0.511	0.426	
PLAU					
rs2227564	**CC** (%)	**CT** (%)	**TT** (%)	**C** (%)	**T** (%)
IPF	44 (68.8)	16 (25.0)	4 (6.3)	104 (81.3)	24 (18.8)
Control	52 (70.3)	21 (28.4)	1 (1.4)	125 (84.5)	23 (15.5)
P	–	0.788	0.136	0.479	

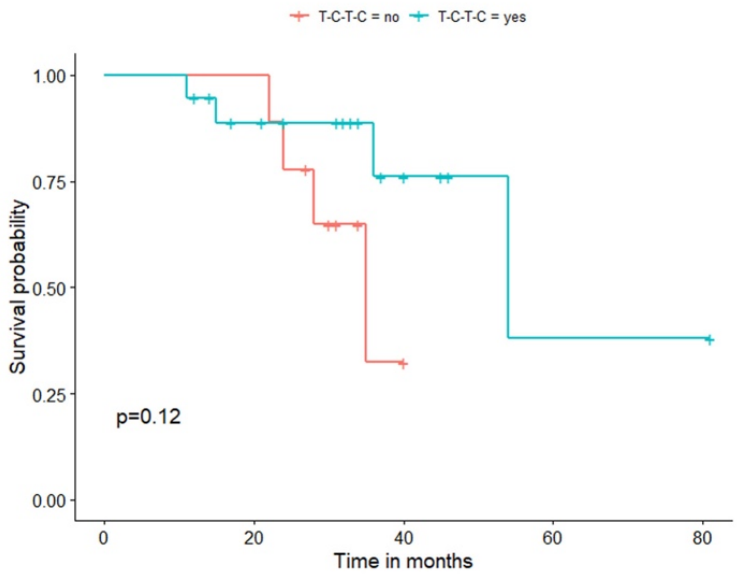


Fig. 3 Kaplan–Meier curves showing IPF patients' survival according to MUC5B rs35705950-TOLLIP rs111521887-TOLLIP rs5743894-TOLLIP rs5743854 haplotype T-C-T-C for patients treated with antifibrotics for at least 1 year. Time in months; *P*-values for log-rank test

Table 4 Clinical and demographic characteristics of patients with and without the T-C-T-C haplotype defined by the rs35705950-rs111521887-rs5743894-rs5743854 MUC5B-TOLLIP block

	T-C-T-C haplotype ($N=31$)	other ($N=33$)	*P*-value
Gender			0.217[a]
Female	13 (41.9%)	9 (27.3%)	
Male	18 (58.1%)	24 (72.7%)	
Age (years)	69.0 (62.0–78.0)	73.0 (65.0–78.0)	0.480[b]
Ever smoked	15 (48.4%)	21 (63.6%)	0.219[a]
Smoking			0.282[a]
Non-smoker	16 (51.6%)	12 (36.4%)	
Ex-smoker	15 (48.4%)	19 (57.6%)	
Smoker	0 (0.0%)	2 (6.1%)	
Follow-up duration (months)	54.0 (39.0–75.0)	36.0 (24.0–59.0)	0.017[b]
Number of deaths	12 (38.7%)	24 (72.7%)	0.006[a]
FVC (% pred)	84.0 (73.7–94.0)	81.0 (72.2–91.2)	0.583[b]
TLC (% pred)	75.8 (65.7–87.0)	76.2 (69.7–86.0)	0.797[b]
DLCO (% pred)	43.0 (39.7–60.0)	43.2 (34.0–54.5)	0.372[b]
COPD	0 (0.0%)	2 (6.1%)	0.493[a]
Pulmonary hypertension	10 (32.3%)	11 (33.3%)	0.927[a]
Lung cancer	4 (12.9%)	2 (6.1%)	0.419[a]
OSA	7 (22.6%)	6 (18.2%)	0.662[a]
GERD	3 (9.7%)	7 (21.2%)	0.305[a]
Metabolic conditions	18 (58.1%)	19 (57.6%)	0.968[a]
Cardiovascular disease	17 (54.8%)	17 (51.5%)	0.790[a]
Depression	3 (9.7%)	7 (21.2%)	0.305[a]
Antifibrotics*	15 (45.5%)	20 (64.5%)	0.126[a]

Values are presented as median (IQR) and counts as *n* (%)

% pred percentage of predicted normal value, *IPF* idiopathic pulmonary fibrosis, *FVC* forced vital capacity, *TLC* total Lung Capacity, *DLCO* diffusing capacity of the Lung for carbon monoxide, *COPD* chronic obstructive lung disease, *OSA* obstructive sleep apnea, *GERD* gastro-esophageal reflux disease

*At the end of follow-up

[a]Chi-squared test

[b]Wilcoxon's rank sum test

Table 5 Univariate and multivariate Cox proportional hazard regressions predicting patients' death

	HR	95% CI	*P*-value
Univariate analysis			
Gender			
Female	–	–	
Male	1.24	0.62, 2.47	0.551
Age	1.03	0.99, 1.06	0.151
Smoking			
Non-smoker	–	–	
Ex-smoker	0.88	0.45, 1.72	0.708
Smoker	1.20	0.15, 9.27	0.864
FVC (% pred)	0.95	0.93, 0.98	4.79×10^{-4}
TLC (% pred)	0.96	0.94, 0.98	0.001
DLCO% (% pred)	0.97	0.94, 1.00	0.038
COPD	1.60	0.381–6.743	0.521
Pulmonary hypertension	1.32	0.667–2.596	0.429
Lung cancer	0.47	0.111–1.969	0.301
OSA	0.59	0.228–1.515	0.271
GERD	1.001	0.414–2.416	0.999
Metabolic conditions	0.69	0.353–1.360	0.286
Cardiovascular disease	0.67	0.344–1.294	0.231
Depression	1.71	0.794–3.669	0.171
Antifibrotics[#]	0.68	0.32–1.46	0.325
Multivariate analysis[*]			
Model 1			
FVC (% pred)	0.95	0.93, 0.98	4.80×10^{-4}
TOLLIP rs111521887			
CC	–	–	
CG	1.65	0.83, 3.26	0.153
GG	5.99	0.74, 48.2	0.092
Model 2			
FVC (% pred)	0.95	0.93, 0.98	4.57×10^{-4}
TOLLIP rs5743894			
CT	–	–	
TT	0.58	0.30, 1.14	0.113
Model 3			
FVC (% pred)	0.95	0.93, 0.98	1.25×10^{-4}
TOLLIP rs5743854			
CC	–	–	
CG	0.64	0.22, 1.86	0.414
GG	9.27	1.92, 44.7	0.006
Model 4			
FVC (% pred)	0.96	0.94, 0.99	0.006
MUC5B-TOLLIP haplotype			
Without	–	–	
With T-C-T-C haplotype	0.37	0.17, 078	0.009

HR hazard ratio, *95% CI* confidence interval, *% pred* percentage of predicted normal value, *IPF* idiopathic pulmonary fibrosis, *FVC* forced vital capacity, *TLC* total lung capacity, *DLCO* diffusing capacity of the lung for carbon monoxide, *COPD* chronic obstructive lung disease, *GERD* gastro-esophageal reflux disease, *OSA* obstructive sleep apnea, MUC5B-TOLLIP haplotype (rs35705950-rs111521887-rs5743894-rs5743854 block)

[#]At the end of follow-up

[*]Multivariate Cox regression models obtained by forward conditional stepwise selection

studies have shown that MUC5B, the principal macromolecule in airway mucus, seems to play a critical role in IPF pathogenesis, but with varying outcomes depending on ethnicity [28–32]. Seibold et al. (2011) first evaluated the MUC5B polymorphism as a risk factor for IPF, showing that the minor allele T was associated with a 14-fold increased expression of MUC5B in the lung, and had the most substantial impact on pulmonary fibrosis risk [14]. Thereafter, numerous case–control or cohort studies were carried out, mostly in Caucasians and some in Asian populations [29–34]. Our findings are in line with those reported for Caucasian populations, with significantly higher frequencies of the minor allele T in IPF patients [14, 35].

Other common variants reportedly associated with IPF susceptibility include SNPs in the genomic region of the TOLLIP gene. The minor alleles of intronic SNPs rs111521887 and rs5743894 were reported to confer susceptibility to IPF, while that of the rs5743890 SNP was found to reduce such a risk [7, 8]. In our study, none of the TOLLIP SNPs, on their own, showed an association with IPF risk, which may indicate a potential distinct feature in Portuguese patients.

In addition to influencing IPF susceptibility, SNPs within MUC5B and TOLLIP seem to have an impact on disease survival as well. Peljto et al. [15] showed that the MUC5B rs35705950 minor allele was paradoxically associated with a twofold decrease in mortality risk, while increasing the IPF susceptibility risk [15]. We did not observe an association between the MUC5B polymorphism and IPF survival, in line with recent studies probing patients from different ethnicities, which also do not corroborate the reported association [33, 34].

Other studies have implicated the TOLLIP rs5743890 SNP in IPF survival, reporting an increased mortality in carriers of the minor allele C [7, 35]. In addition, Oldham et al. (2015) showed that TOLLIP's synonymous substitution rs3750920 was associated with a favorable response to N-acetylcysteine, as carriers of the TT genotype (25% of IPF) treated with NAC had a significantly reduced composite endpoint risk, including death, hospitalization, and FVC decline [36]. We could not test any SNP genotype effect on survival in the context of NAC treatment due to the reduced number of patients subjected to this treatment (7.8%). In our study, no association was seen between the TOLLIP rs5743890 or rs3750920 SNPs and IPF survival. Instead, we found evidence for a tentative association of the minor alleles of TOLLIP SNPs rs111521887 (G), rs5743894 (C), and rs5743854 (G), in homozygosity, with increased mortality in IPF patients. Multivariate analysis confirmed the risk of mortality was higher for patients homozygous for the minor allele G of TOLLIP rs5743854, after adjustment for a number of covariates. The frequency of the GG genotype, however, is very low (0.022; $n=134$), in accordance to

that found in Iberian (0.019; $n = 107$) and European populations (0.020; $n = 503$) of the 1000 Genomes Project [37]. The rs5743854 polymorphism is located within the TOLLIP promoter and has potential transcriptional consequences.

Haplotypes defined by neighboring SNPs can provide better insights into the influence of genetic loci than the information produced by the study of SNPs individually [38, 39]. It has been demonstrated that genetic associations can be fine-mapped using SNP haplotypes within regions of interest previously identified by single-SNP genome-wide association studies [40]. We show that carriers of the haplotype T-C-T-C defined by the MUC5B promoter SNP and the three TOLLIP SNPs that tended to associate individually with survival in our study, had a significant increase in survival compared to those without the haplotype, the difference in median survival time amounting to an impressive 36 months. Multivariate analysis, using as covariates demographics and pulmonary function, confirmed the association, and further demonstrated that the haplotype protective effect was independent of the administration of antifibrotic drugs, pirfenidone or nintedanib. Interestingly, when evaluating the impact of the T-C-T-C haplotype on disease survival of IPF patients treated with antifibrotics for at least a year, we found a trend for longer survival in carriers. A similar HR to that seen for the T-C-T-C haplotype carriers in the multivariate analysis was obtained (HR = 0.31), although it did not reach statistical significance ($P = 0.122$ for log-rank test) conceivably due to the small sample size. Whether this could reflect an influence of the haplotype on response to AF therapy in these patients—in addition to its protective effect regardless of treatment—requires further investigation. Biondini et al. (2021) reported an association of the MUC5B rs35705950 T allele with longer survival in 88 IPF patients on AF therapy (TT/TG vs. GG carriers), but the MUC5B rs35705950 genotype did not seem to affect response to this treatment [41]. Our analysis of IPF patients on AF treatment did not replicate the reported association; however, the number of such patients is considerably smaller in our study.

Recent data highlights a role for TOLLIP on autophagy and apoptosis in the early and late stages of the disease, suggesting a protective effect against apoptosis, oxidative damage, and fibrosis [5, 42]. Furthermore, TOLLIP was shown to protect bronchial epithelial cells from bleomycin-induced apoptosis in vitro [5]. Accordingly, TOLLIP expression in lungs from patients with IPF was decreased relative to that of normal lungs [5]. It had been previously reported that carriers of the minor alleles of TOLLIP SNPs rs111521887, rs5743894, and rs5743890 had reduced expression of TOLLIP in IPF lung tissues, relative to homozygous carriers of the common alleles, further suggesting that variants within this gene may modify the course of disease [7]. Notably, in our study, the minor alleles of the former two SNPs tend to associate with decreased survival in IPF, while the common alleles integrate haplotypes whose carriers show better survival compared to non-carriers. These findings suggest that the region defined by this haplotype block contains or segregates with a genetic factor as-yet unidentified that influences the clinical course of disease.

The variants of the novel candidate genes surveyed, although potentially promising, did not reveal significant associations with IPF susceptibility or prognosis. SERPINB1 and PLAU were chosen as candidates in this study given their involvement in the lung fibrotic process. Both genes are highly expressed in the lung, and in addition to playing a role in tissue homeostasis suggestive of a likely involvement in the lung parenchyma pathophysiology [43, 44], their expression is misregulated in lung tissue of IPF patients, according to a comprehensive collection of public microarray and RNA-Seq study results [45]. IPF pathogenesis involves tissue turnover mediated by proteases upregulated during disease, especially neutrophil elastase [43]. SERPINB1, a member of the serpin family of proteinase inhibitors, is among the most efficient inhibitors of neutrophil serine proteases, protecting tissues from damage at inflammatory sites and contributing to tissue homeostasis [44, 46]. SERPINB1 expression in lung tissue of IPF patients is downregulated by an average 1.5-fold according to several microarray and RNA-Seq studies (AFFY_ST1_0-7, AFFY_U133PLUS_2-2, AGIL_4×44 K-6, mRNASeq_HUMAN_GL3, totalRNASeq_HUMAN_GL4) [45]. Variants of the SERPINB1 gene may be associated with regulatory factors of functional consequence at the level of expression.

The PLAU gene encodes a key serine protease known as urokinase-type plasminogen activator (uPA) that converts plasminogen to plasmin, which lyses fibrin and degrades extracellular matrix directly, or indirectly by activating matrix metalloproteinases and collagenase. The uPA/fibrinolysin system participates in the repair process of lung fibrosis and is implicated in IPF [47, 48]. PLAU expression in lung tissue of IPF patients is upregulated by an average 1.6-fold according to several microarray and RNA-Seq studies (AFFY_ST1_0-7, AGIL_4×44 K-6, mRNASeq_HUMAN_GL3, totalRNASeq_HUMAN_GL4) [45]. The missense variant in PLAU underlies lower affinity to the fibrin clot and a reduction in fibrin degradation and has been associated with several respiratory and lung diseases [49, 50].

We have also used MDR to investigate whether multilocus interactions, rather than individual effects of the probed variants, could underlie IPF susceptibility, as seen in other pathologies [24]. All meaningful interactions correlated with the MUC5B main effect and collapsed without it, therefore not contributing any information for phenotype prediction. Thus, no multi-locus interactions between the probed SNPs associated with the IPF susceptibility trait were identified in our study. Nonetheless, we cannot

exclude the possibility that the modest sample size in our study may have limited MDR ability to detect gene–gene interactions.

In conclusion, we show that the MUC5B rs35705950 minor allele is associated with an increased IPF risk in the Portuguese and that a MUC5B-TOLLIP SNPs-defined haplotype is strongly associated with disease survival. In spite of the sample size, which constitutes the main limitation of our study, we propose this haplotype has the potential to be used as a marker of prognostic value in IPF patient stratification, subject to further multicenter and prospective validation.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00109-022-02242-y.

Acknowledgements We are indebted to Instituto de Saúde Pública da Universidade do Porto (ISPUP) for providing samples from the EPI-Porto cohort to be used as controls; we thank the cohort management and all participants. We acknowledge the contribution of Ana Isabel Loureiro, MD (Centro Hospitalar de Trás-os-Montes e Alto Douro) and Rui Rolo, MD (Hospital de Braga) who referred some IPF patients to our center. We are grateful to Catarina Teixeira Antunes and Joana Filipa Amorim Reis for assistance in the genotyping.

Author contribution Conceptualization: P.C.M., M.L.S., J.A.M., and A.M.; methodology: P.C.M., M.L.S., A.M., and E.M.; data curation: P.C.M., M.L.S., and B.A.L.; formal analysis: P.C.M., M.L.S., B.A.L., and E.M.; investigation: P.C.M., M.L.S., C.D.V., A.C.F., N.M., H.N-B., and A.M.; validation: P.C.M., M.L.S., and J.H.M.; writing—original draft: P.C.M., M.L.S., and B.A.L.; writing—review and editing: all authors; resources: P.C.M., M.L.S., A.C.F., N.M., and H. N-B; funding acquisition: P.C.M., M.L.S., J.A.M., and A.M.; supervision and project administration: P.C.M., M.L.S., and A.M.

Funding This work was supported by the Portuguese Society of Pulmonology (SPP) (Grant - SPP/Bolsa Novartis 2012) and Boehringer-Ingelheim (unrestricted research grant).

Data availability The datasets generated during and/or analyzed during the current study are not publicly available due to privacy or ethical restrictions, but are available from the corresponding author on reasonable request.

Declarations

Ethics approval This study was performed in line with the principles of the Declaration of Helsinki. Approval was granted by the Ethics Committee of Centro Hospitalar Universitário de São João (CHUSJ) (CE-203–2012).

Consent to participate Written informed consent was obtained from all individual participants in the study.

Competing interests Patrícia Caetano Mota has received speaker fees from Boehringer-Ingelheim and research grants from Boehringer-Ingelheim and Novartis; and has participated in research with Boehringer-Ingelheim and Roche, for which her institution has been remunerated.

Hélder Novais-Bastos and António Morais have attended advisory boards for Boehringer-Ingelheim and Roche; have received speaker fees from Boehringer-Ingelheim and Roche; have participated in research with Boehringer-Ingelheim and Roche, for which their institution has been remunerated. Hélder Novais-Bastos has also received a research grant from Boehringer-Ingelheim.
The remaining authors have no relevant financial or non-financial interests to disclose.

References

1. Coward WR, Saini G, Jenkins G (2010) The pathogenesis of idiopathic pulmonary fibrosis. Ther Adv Respir Dis 4:367–388. https://doi.org/10.1177/1753465810379801
2. Raghu G, Remy-Jardin M, Myers JL, Richeldi L, Ryerson CJ, Lederer DJ et al (2018) Diagnosis of idiopathic pulmonary fibrosis. An official ATS/ERS/JRS/ALAT clinical practice guideline. Am J Respir Crit Care Med 198:e44–68. https://doi.org/10.1164/rccm.201807-1255ST
3. Selman M, King J, Pardo A (2001) Idiopathic pulmonary fibrosis: prevailing and evolving hypotheses about its pathogenesis and implications for therapy. Ann Intern Med 134:136–151
4. Lv M, Liu Y, Ma S, Yu Z (2019) Current advances in idiopathic pulmonary fibrosis: the pathogenesis, therapeutic strategies and candidate molecules. Future Med Chem 11:2595–2620. https://doi.org/10.4155/fmc-2019-0111
5. Li X, Kim SE, Chen TY, Wang J, Yang X, Tabib T et al (2020) Toll Interacting Protein protects bronchial epithelial cells from bleomycin-induced apoptosis. FASEB J 34:9884. https://doi.org/10.1096/fj.201902636RR
6. Spagnolo P, Grunewald J, Du Bois RM (2014) Genetic determinants of pulmonary fibrosis: evolving concepts. Lancet Respir Med 2:416–428
7. Noth I, Zhang Y, Ma SF, Flores C, Barber M, Huang Y et al (2013) Genetic variants associated with idiopathic pulmonary fibrosis susceptibility and mortality: a genome-wide association study. Lancet Respir Med 1:309–317. https://pubmed.ncbi.nlm.nih.gov/24429156/
8. Fingerlin TE, Murphy E, Zhang W, Peljto AL, Brown KK, Steele MP et al (2013) Genome-wide association study identifies multiple susceptibility loci for pulmonary fibrosis. Nat Genet 45:613–620. https://www.nature.com/articles/ng.2609
9. Allen RJ, Porte J, Braybrooke R, Flores C, Fingerlin TE, Oldham JM et al (2017) Genetic variants associated with susceptibility to idiopathic pulmonary fibrosis in people of European ancestry: a genome-wide association study. Lancet Respir Med 5:869–880. https://pubmed.ncbi.nlm.nih.gov/29066090/
10. Newton CA, Molyneaux PL, Oldham JM (2018) Clinical genetics in interstitial lung disease. Front Med 5:116. https://doi.org/10.3389/fmed.2018.00116
11. Coghlan MA, Shifren A, Huang HJ, Russell TD, Mitra RD, Zhang Q et al (2014) Sequencing of idiopathic pulmonary fibrosis-related genes reveals independent single gene associations. BMJ Open Respir Res 1:e000057. https://bmjopenrespres.bmj.com/content/1/1/e000057
12. Petrovski S, Todd JL, Durheim MT, Wang Q, Chien JW, Kelly FL et al (2017) An exome sequencing study to assess the role of rare genetic variation in pulmonary fibrosis. Am J Respir Crit Care Med 196:82–93. https://pubmed.ncbi.nlm.nih.gov/28099038/
13. Seibold MA, Wise AL, Speer MC, Steele MP, Brown KK, Loyd JE et al (2011) A common MUC5B promoter polymorphism and pulmonary fibrosis. N Engl J Med 364:1503–1512. https://pubmed.ncbi.nlm.nih.gov/21506741/
14. Zhang Y, Noth I, Garcia JGN, Kaminski N (2011) A variant in the promoter of MUC5B and idiopathic pulmonary fibrosis. N Engl J Med.364:1576–1577. https://doi.org/10.1056/NEJMc1013504

15. Peljto AL, Zhang Y, Fingerlin TE, Shwu-Fan M, Garcia JGN, Richards TJ et al (2013) Association between the MUC5B promoter polymorphism and survival in patients with idiopathic pulmonary fibrosis. JAMA 309:2232–2239. https://pubmed.ncbi.nlm.nih.gov/23695349/
16. Evans CM, Fingerlin TE, Schwarz MI, Lynch D, Kurche J, Warg L et al (2016) Idiopathic pulmonary fibrosis: a genetic disease that involves mucociliary dysfunction of the peripheral airways. Physiol Rev 96:1567–1591. https://doi.org/10.1152/physrev.00004.2016
17. Ramos E, Lopes C, Barros H (2004) Investigating the effect of nonparticipation using a population-based case–control study on myocardial infarction. Ann Epidemiol 14:437–441
18. Raghu G, Collard HR, Egan JJ, Martinez FJ, Behr J, Brown KK 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 183:788–824. https://doi.org/10.1164/rccm.2009-040GL
19. Robalo Cordeiro C, Campos P, Carvalho L, Campainha S, Clemente S, Figueiredo L et al (2016) Consensus document for the diagnosis and treatment of idiopathic pulmonary fibrosis: joint consensus of Sociedade Portuguesa de Pneumologia, Sociedade Portuguesa de Radiologia e Medicina Nuclear e Sociedade Portuguesa de Anatomia Patológica. Rev Port Pneumol (English Ed) 22:112–122
20. Raghu G, Rochwerg B, Zhang Y et al (2015) An Official ATS/ERS/JRS/ALAT Clinical practice guideline: treatment of idiopathic pulmonary fibrosis. An update of the 2011 clinical practice guideline [published correction appears in Am J Respir Crit Care Med. 2015 Sep 1;192(5):644. Dosage error in article text]. Am J Respir Crit Care Med 192:e3–e19. https://doi.org/10.1164/rccm.201506-1063ST
21. Stephens M, Smith NJ, Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. Am J Hum Genet 68:978–989
22. Stephens M, Scheet P (2005) Accounting for decay of linkage disequilibrium in haplotype inference and missing-data imputation. Am J Hum Genet 76:449–462. https://doi.org/10.1086/428594
23. Barrett JC, Fry B, Maller J, Daly MJ (2005) Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 21:263–265. https://pubmed.ncbi.nlm.nih.gov/15297300/
24. Ritchie MD, Hahn LW, Roodi N, Bailey LR, Dupont WD, Parl FF et al (2001) Multifactor-dimensionality reduction reveals high-order interactions among estrogen-metabolism genes in sporadic breast cancer. Am J Hum Genet 69:138–147. https://pubmed.ncbi.nlm.nih.gov/11404819/
25. Moore JH, Andrews PC (2015) Epistasis analysis using multifactor dimensionality reduction. Methods Mol Biol 1253:301–314. https://doi.org/10.1007/978-1-4939-2155-3_16
26. Greene CS, Himmelstein DS, Nelson HH, Kelsey KT, Williams SM, Andrew AS et al (2010) Enabling personal genomics with an explicit test of epistasis. Pacific Symp Biocomput 2010:327–336. https://doi.org/10.1142/9789814295291_0035
27. Moore JH, Hu T (2015) Epistasis analysis using information theory. Methods Mol Biol 1253:257–268. https://doi.org/10.1007/978-1-4939-2155-3_13
28. Zhu QQ, Zhang XL, Zhang SM, Tang SW, Min HY, Yi L et al (2015) Association between the MUC5B promoter polymorphism rs35705950 and idiopathic pulmonary fibrosis: a meta-analysis and trial sequential analysis in caucasian and asian populations. Medicine (Baltimore) 94:e1901. https://pubmed.ncbi.nlm.nih.gov/26512610/
29. Stock CJ, Sato H, Fonseca C, Banya WAS, Molyneaux PL, Adamali H et al (2013) Mucin 5B promoter polymorphism is associated with idiopathic pulmonary fibrosis but not with development of lung fibrosis in systemic sclerosis or sarcoidosis. Thorax 68:436–441. https://thorax.bmj.com/content/68/5/436
30. Horimasu Y, Ohshimo S, Bonella F, Tanaka S, Ishikawa N, Hattori N et al (2015) MUC5B promoter polymorphism in Japanese patients with idiopathic pulmonary fibrosis. Respirology 20(3):439–444. https://doi.org/10.1111/resp.12466
31. Peljto AL, Selman M, Kim DS, Murphy E, Tucker L, Pardo A, et al (2015) The MUC5B promoter polymorphism is associated with idiopathic pulmonary fibrosis in a Mexican cohort but is rare among Asian ancestries. Chest 147:460–464. https://pubmed.ncbi.nlm.nih.gov/25275363/
32. Wang C, Zhuang Y, Guo W, Cao L, Zhang H, Xu L et al (2014) Mucin 5B promoter polymorphism is associated with susceptibility to interstitial lung diseases in Chinese males. PLoS One 9:e104919. https://doi.org/10.1371/journal.pone.0104919
33. Jiang H, Hu Y, Shang L, Li Y, Yang L, Chen Y (2015) Association between MUC5B polymorphism and susceptibility and severity of idiopathic pulmonary fibrosis. Int J Clin Exp Pathol 8:14953–14958
34. Van Der Vis JJ, Snetselaar R, Kazemier KM, Ten Klooster L, Grutters JC, Van Moorsel CHM (2016) Effect of Muc5b promoter polymorphism on disease predisposition and survival in idiopathic interstitial pneumonias. Respirology 21:712–717. https://doi.org/10.1111/resp.12728
35. Bonella F, Campo I, Zorzetto M, Boerner E, Ohshimo S, Theegarten D et al (2021) Potential clinical utility of MUC5B und TOLLIP single nucleotide polymorphisms (SNPs) in the management of patients with IPF. Orphanet J Rare Dis 16(1):1–9. https://doi.org/10.1186/s13023-021-01750-3
36. Oldham JM, Ma SF, Martinez FJ, Anstrom KJ, Raghu G, Schwartz DA et al (2015) TOLLIP, MUC5B, and the response to N-acetylcysteine among individuals with idiopathic pulmonary fibrosis. Am J Respir Crit Care Med 192(12):1475–1482. https://doi.org/10.1164/rccm.201505-1010OC
37. Auton A, Abecasis GR, Altshuler DM, Durbin RM, Bentley DR, Chakravarti A et al (2015) A global reference for human genetic variation. Nature 526:68–74. https://www.nature.com/articles/nature15393
38. Shirali M, Knott SA, Pong-Wong R, Navarro P, Haley CS (2018) Haplotype heritability mapping method uncovers missing heritability of complex traits. Sci Reports 8:1–9. https://www.nature.com/articles/s41598-018-23307-4
39. Schrodi SJ, Garcia VE, Rowland C, Jones HB (2006) Pairwise linkage disequilibrium under disease models. Eur J Hum Genet 15:212–220. https://www.nature.com/articles/5201731
40. Khankhanian P, Gourraud PA, Lizee A, Goodin DS (2015) Haplotype-based approach to known MS-associated regions increases the amount of explained risk. J Med Genet 52:587–594. https://pubmed.ncbi.nlm.nih.gov/26185143/
41. Biondini D, Cocconcelli E, Bernardinello N, Lorenzoni G, Rigobello C, Lococo S et al (2021) Prognostic role of MUC5B rs35705950 genotype in patients with idiopathic pulmonary fibrosis (IPF) on antifibrotic treatment. Respir Res 22:98. https://doi.org/10.1186/s12931-021-01694-z
42. Li X, Goobie GC, Gregory AD, Kass DJ, Zhang Y (2021) Toll-interacting protein in pulmonary diseases abiding by the goldilocks principle. Am J Respir Cell Mol Biol 64:536–546. https://doi.org/10.1165/rcmb.2020-0470TR
43. Obayashi Y, Yamadori I, Fujita J, Yoshinouchi T, Ueda N, Takahara J (1997) The role of neutrophils in the pathogenesis of idiopathic pulmonary fibrosis. Chest 112:1338–1343
44. Baumann M, Pham CTN, Benarafa C (2013) SerpinB1 is critical for neutrophil survival through cell-autonomous inhibition of cathepsin G. Blood 121:3900–3907. http://ashpublications.org/blood/article-pdf/121/19/3900/1218553/3900.pdf
45. Hruz T, Laule O, Szabo G, Wessendorp F, Bleuler S, Oertle L et al (2008) Genevestigator v3: a reference expression database for the meta-analysis of transcriptomes. Adv Bioinform 2008:420747. https://doi.org/10.1155/2008/420747

46. Gong D, Farley K, White M, Hartshorn KL, Benarafa C, Remold-O'Donnell E (2011) Critical role of serpinB1 in regulating inflammatory responses in pulmonary influenza infection. J Infect Dis 204:592–600. https://doi.org/10.1093/infdis/jir352
47. Gharaee-Kermani M, Hu B, Phan SH, Gyetko MR (2008) The role of urokinase in idiopathic pulmonary fibrosis and implication for therapy. Expert Opin Investig Drugs 17:905–916. https://doi.org/10.1517/13543784.17.6.905
48. Tucker TA, Idell S (2021) The contribution of the urokinase plasminogen activator and the urokinase receptor to pleural and parenchymal lung injury and repair: a narrative review. Int J Mol Sci 22:1437. https://doi.org/10.3390/ijms22031437
49. Idell S, Cohen AB (1985) Bronchoalveolar lavage in patients with the adult respiratory distress syndrome. Clin Chest Med 6:459–471
50. Chang LC, Tseng JC, Hua CC, Liu YC, Shieh W Bin, Wu HP (2006) Gene polymorphisms of fibrinolytic enzymes in coal workers' pneumoconiosis. Arch Environ Occup Health 61:61–66. https://pubmed.ncbi.nlm.nih.gov/17649957/

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Authors and Affiliations

Patrícia Caetano Mota[1,2,4] · Miguel Luz Soares[2,3,4] · Carlos Daniel Vasconcelos[2,3,4] · António Carlos Ferreira[2,3,4] · Bruno A. Lima[5] · Elisabetta Manduchi[6] · Jason H. Moore[7] · Natália Melo[1] · Hélder Novais-Bastos[1,2,4] · José Miguel Pereira[8] · Susana Guimarães[9] · Conceição Souto Moura[9] · José Agostinho Marques[1,2,4] · António Morais[1,2,4]

1 Departamento de Pneumologia, Centro Hospitalar Universitário de São João, EPE, Alameda Professor Hernâni Monteiro, 4200-319 Porto, Portugal

2 Faculdade de Medicina da Universidade do Porto, Porto, Portugal

3 Laboratório de Apoio à Investigação em Medicina Molecular (LAIMM), Departamento de Biomedicina, Faculdade de Medicina da Universidade do Porto, Porto, Portugal

4 Instituto de Investigação e Inovação em Saúde, Universidade do Porto, Porto, Portugal

5 Oficina de Bioestatística, Ermesinde, Portugal

6 Department of Biostatistics, Epidemiology & Informatics and Institute for Biomedical Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

7 Department of Computational Biomedicine, Cedars-Sinai Medical Center, Los Angeles, CA, USA

8 Departamento de Radiologia, Centro Hospitalar Universitário de São João, EPE, Alameda Professor Hernâni Monteiro, 4200-319 Porto, Portugal

9 Departamento de Anatomia Patológica, Centro Hospitalar Universitário de São João, EPE, Alameda Professor Hernâni Monteiro, 4200-319 Porto, Portugal

ONLINE RESOURCE

PREDICTIVE VALUE OF COMMON GENETIC VARIANTS IN IDIOPATHIC PULMONARY FIBROSIS

Authors' full names: Patrícia Caetano Mota[1,2,4], Miguel Luz Soares[2,3,4], Carlos Daniel Vasconcelos[2,3,4,*], António Carlos Ferreira[2,3,4,*], Bruno A Lima[5], Elisabetta Manduchi[6], Jason H. Moore[7], Natália Melo[1], Hélder Novais-Bastos[1,2,4], José Miguel Pereira[8], Susana Guimarães[9], Conceição Souto Moura[9], José Agostinho Marques[1,2,4], António Morais[1,2,4]

Authors' affiliation(s):

[1]Departamento de Pneumologia, Centro Hospitalar Universitário de São João, EPE, Porto, Portugal

[2]Faculdade de Medicina da Universidade do Porto, Porto, Portugal

[3]Laboratório de Apoio à Investigação em Medicina Molecular (LAIMM), Departamento de Biomedicina, Faculdade de Medicina da Universidade do Porto, Porto, Portugal

[4]Instituto de Investigação e Inovação em Saúde, Universidade do Porto, Porto, Portugal

[5]Oficina de Bioestatística, Ermesinde, Portugal

[6]Department of Biostatistics, Epidemiology & Informatics and Institute for Biomedical Informatics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

[7]Department of Computational Biomedicine, Cedars-Sinai Medical Center, Los Angeles, CA, USA

[8]Departamento de Radiologia, Centro Hospitalar Universitário de São João, EPE, Porto, Portugal

[9]Departamento de Anatomia Patológica, Centro Hospitalar Universitário de São João, EPE, Porto, Portugal

[*]These authors contributed equally to this work

Corresponding author full contact details:

Patrícia Caetano Mota

Centro Hospitalar Universitário de São João, EPE, Porto

Alameda Professor Hernâni Monteiro

4200-319 Porto

Portugal

email: patcmota@gmail.com

Table S1. Primers and restriction enzymes for RFLP analyses

Locus	Primer name	Primer sequence (5' to 3')	Annealing (°C)	Amplicon Size (bp)	Restriction Enzyme	RFLP fragments (bp)
MUC5B	rs35705950_Fw	GCCAGAATGAGGGACAGTGA	60	375	HhaI	G: 189+105+81 T: 270+105
	rs35705950_Rv	ACTTTGCCCTCGTCCCTC				
TOLLIP	rs5743854_Fw	CATCCCCGTTCCGTTCAC	60	250	SbfI	G: 250 C: 134+116
	rs5743854_Rv	TTCCTCCTCGGTCCTCTGT				
	rs5743890_Fw	GGCTCCCCTGGAAATGGTAA	62	408	EarI	T: 408 C: 310+98
	rs5743890_Rv	GTCAGGACACAGAGAAAAGGC				
	rs5743894_Fw	GGCCTCTAACACTGTCCCTT	62	440	FspI	T: 440 C: 247+193
	rs5743894_Rv	AAAAGCCAGTCAAGCAGCAG				
	rs111521887_Fw	GAGGGTTCAGTGAGGACAGG	60	460	AflIII	C: 460 G: 333+127
	rs111521887_Rv	ACTCAACATGGATGGCAGGT				
	rs3750920_Fw	GGTCTCATCGATGGGGTGTT	62	421	MspI	C: 226+195 T: 421
	rs3750920_Rv	CCCTGCTGTTTTATGGGCTC				
SERPINB1	rs57721448_Fw	GAGGTCGTCTGTCTCCGTC	60	140	BsaJI	G: 98+42 T: 140
	rs57721448_Rv	CTCCCTCGCTCAGCTTCC				
	rs386713_Fw	GGAAGCTGAGCGAGGGAG	60	266	HgaI	T: 266 C: 154+112
	rs386713_Rv	GCTCACTTCTGCTTGCACTT				
	rs408265_Fw	AAGGTGTGAATGACTGAGACGA	60	111	StyI	C: 80+31 T: 111
	rs408265_Rv	CCAAATTTGCTTCCTTCTCTCcA				
PLAU	rs2227564_Fw	CCTGCTATGAGGGGAATGGT	62	826	AluI	C: 623+134+69 T: 420+203+134+69
	rs2227564_Rv	TACATGAAGCAGTGTGTGGC				

RFLP, Restriction Fragment Length Polymorphism; Fw, Forward; Rv, Reverse; bp, base pairs

Table S2. Chi-squared tests for Hardy-Weinberg equilibrium

Locus	SNP	Population	
		Patients *P*	Controls *P*
MUC5B	rs35705950	0.113	0.242
TOLLIP	rs3750920	0.134	0.217
	rs111521887	0.726	0.820
	rs5743894	0.366	0.918
	rs5743890	0.685	0.718
	rs5743854	0.113	0.918
SERPINB1	rs57721448	0.520	0.205
	rs386713	0.735	0.110
	rs408265	0.661	0.736
PLAU	rs2227564	0.151	0.486

Table S3. Haplotype frequencies of the M950-T887-T894-T854 MUC5B-TOLLIP haplotype block

Haplotype	Overall Freq.	Patients Freq.	Control Freq.	Chi Square	*P*
GCTC	0.553	0.427	0.662	15.325	<0.0001
TCTC	0.174	0.274	0.088	16.582	<0.0001
GCTG	0.092	0.070	0.112	1.450	0.2285
GGCC	0.075	0.042	0.104	3.822	0.0506
TGCC	0.073	0.138	0.017	14.662	0.0001
TCTG*	0.021	0.033	0.010	1.895	0.1686
GGTC*	0.004	0.002	0.007	0.427	0.5136
TGTC*	0.004	0.008	0.000	1.160	0.2814
GGTG*	0.003	0.006	0.000	0.927	0.3357

*Rare haplotypes (≤ 2 occurrences)

Table S4. Measures of pairwise linkage disequilibrium between MUC5B rs35705950 and TOLLIP rs111521887, rs5743894 and rs5743854 SNPs in patients and controls

SNP A	SNP B	D'	CI 95%	R^2	Dist bp	χ^2	*P*
IPF patients							
rs35705950	**rs111521887**	0.663	0.398 - 0.927	0.129	71485	15.085	0.0001
rs35705950	**rs5743894**	0.748	0.499 - 0.997	0.148	83551	19.622	0.0000
rs35705950	**rs5743854**	-0.417	-0.906 - 0.072	0.018	90033	1.778	0.1824
rs111521887	**rs5743894**	1	1.000 - 1.000	0.903	12066	115.517	0.0000
rs111521887	**rs5743854**	-1	-1.000 - -0.998	0.030	18548	3.815	0.0508
rs5743894	**rs5743854**	-1	-1.000 - -0.999	0.027	6482	3.443	0.0635
Controls							
rs35705950	**rs111521887**	-0.121	-1.000 - 1.000	0.000	71485	0.020	0.8882
rs35705950	**rs5743894**	-0.048	1.000 - 1.000	0.000	83551	0.003	0.9575
rs35705950	**rs5743854**	-0.092	1.000 - 1.000	0.000	90033	0.003	0.9575
rs111521887	**rs5743894**	1	1.000 - 1.000	0.940	12066	139.132	0.0000
rs111521887	**rs5743854**	-1	-1.000 - -0.996	0.020	18548	3.018	0.0823
rs5743894	**rs5743854**	-1	-1.000 - -0.996	0.019	6482	2.837	0.0921

D', D prime between two SNPs; R^2, correlation coefficient between two SNPs; Dist bp, distance between SNPs in base pairs

Table S5. Distribution of patients with and without MUC5B/TOLLIP T-C-T-C haplotype according to treatment

	T-C-T-C haplotype (N = 31)	Other (N = 33)	p-value[1]
Antifibrotics	15 (45.5%)	20 (64.5%)	0.126
Immunosuppressants	7 (22.6%)	13 (39.4%)	0.147
NAC monotherapy	1 (3.2%)	4 (12.1%)	0.356
No treatment	3 (9.7)	1 (3.0%)	0.347
PPI	18 (58.1%)	19 (57.6%)	0.968

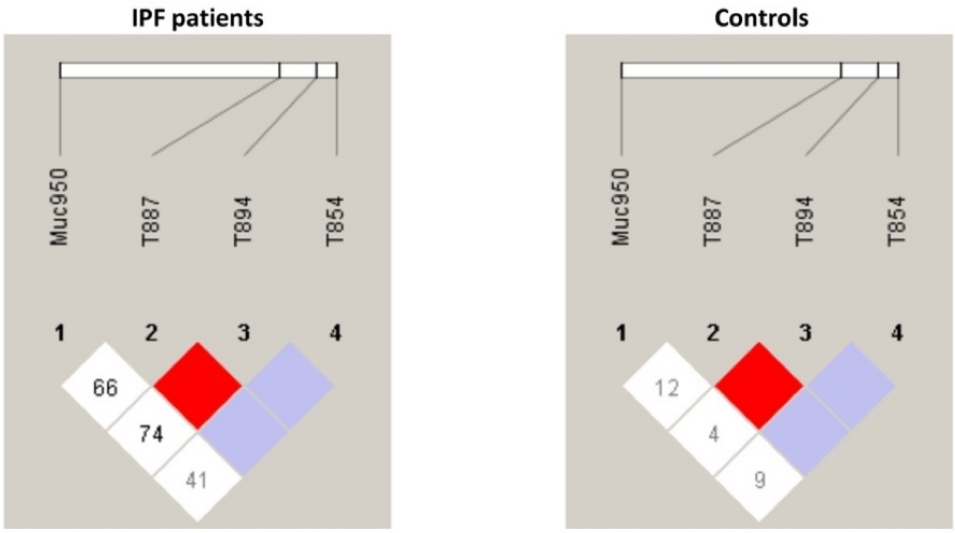
Counts n (%) at end of follow-up
[1] Chi-squared test
NAC, N-acetylcysteine; PPI, proton pump inhibitor

Table S6. Multivariate Cox proportional hazard regression model predicting patients' death. T-C-T-C haplotype and FVC associations were adjusted for antifibrotics as a time-dependent covariate

	HR	**95% CI**	**p-value**
*Multivariate analysis**			
FVC (% pred)	0.96	0.94, 0.99	0.008
MUC5B-TOLLIP haplotype			
without	—	—	
with T-C-T-C haplotype	0.37	0.17, 0.78	0.009
Antifibrotics[#]	0.84	0.38, 1.88	0.679

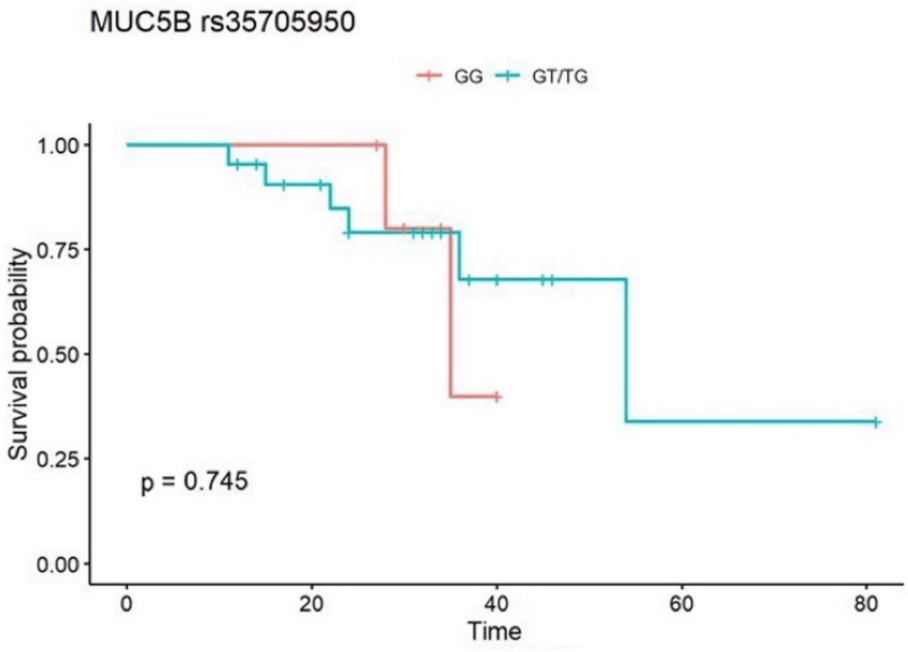
HR Hazard Ratio; 95% CI Confidence Interval
% pred, percentage of predicted normal value
FVC, Forced Vital Capacity
MUC5B-TOLLIP haplotype (rs35705950-rs111521887-rs5743894-rs5743854)
[#] Time-dependent covariate
* Multivariate Cox regression model obtained by forward conditional stepwise selection

Fig. S1 Linkage disequilibrium (LD) plot of the 91-kb M950-T887-T894-T854 block on chromosome 11p15.5



The plot was drawn using Haploview software v4.2.
D' values are shown inside the squares. Empty squares have a D' = 1 (100)
Red and blue boxes represent D' = 1 (LOD ≥ 2 and LOD < 2, respectively). White boxes represent D' < 1 and LOD < 2. LOD, log of likelihood odds ratio.

Fig. S2 Kaplan–Meier curves showing IPF patients' survival according to MUC5B rs35705950 TT/TG and GG genotypes of patients treated with antifibrotics for at least one year.



Time in months; p-value for Log-rank test

SECOND STUDY

POLYMORPHISMS AND HAPLOTYPES OF *TOLLIP* AND *MUC5B* ARE ASSOCIATED WITH SUSCEPTIBILITY AND SURVIVAL IN PATIENTS WITH FIBROTIC HYPERSENSITIVITY PNEUMONITIS

Mota PC, Soares ML, Ferreira AC, Santos RF, Rufo JC, Vasconcelos D, Carvalho A, Guimarães S, Vasques-Nóvoa F, Cardoso C, Melo N, Alexandre AT, Coelho D, Novais-Bastos H, Morais A

Pulmonology. 2024 Feb 2:S2531-0437(24)00007-2. doi: 10.1016/j.pulmoe.2024.01.002. Epub ahead of print. PMID: 38309995

ARTICLE IN PRESS
JID: PULMOE [mSP6P;February 2, 2024;18:04]
Pulmonology 000 (xxxx) 1–11

PULMONOLOGY
www.journalpulmonology.org



ORIGINAL ARTICLE

Polymorphisms and haplotypes of *TOLLIP* and *MUC5B* are associated with susceptibility and survival in patients with fibrotic hypersensitivity pneumonitis

P.C. Mota[a,b,e], M.L. Soares[c,d,e,*], A.C. Ferreira[c,e], R.F. Santos[e,f], J.C. Rufo[g,h,i], D. Vasconcelos[c,e], A. Carvalho[b,j], S. Guimarães[b,k], F. Vasques-Nóvoa[b,l,m], C. Cardoso[a,b,e], N. Melo[a], A.T. Alexandre[a], D. Coelho[a,b,e], H. Novais-Bastos[a,b,e], A. Morais[a,b,e]

[a] *Departamento de Pneumologia, Centro Hospitalar Universitário de São João, EPE, Porto, Portugal*
[b] *Faculdade de Medicina da Universidade do Porto, Porto, Portugal*
[c] *Laboratório de Apoio à Investigação em Medicina Molecular (LAIMM), Departamento de Biomedicina, Faculdade de Medicina da Universidade do Porto, Portugal*
[d] *LAIMM, Núcleo de Recursos Laboratoriais, Unidade de Gestão de Conhecimento, Departamento de Recursos Comuns, Faculdade de Medicina da Universidade do Porto, Portugal*
[e] *Instituto de Investigação e Inovação em Saúde, Universidade do Porto, Porto, Portugal*
[f] *Escola Superior de Saúde - Instituto Politécnico do Porto, Portugal*
[g] *Indoor Air Quality and Respiratory Health Lab, Laboratory for Integrative and Translational Research in Population Health (ITR), Porto, Portugal*
[h] *EPIUnit - Instituto de Saúde Pública, Universidade do Porto, Porto, Portugal*
[i] *Center for Translational Health and Medical Biotechnology Research (T.Bio), Escola Superior de Saúde, Instituto Politécnico do Porto, Porto, Portugal*
[j] *Departamento de Radiologia, Centro Hospitalar Universitário de São João, EPE, Porto, Portugal*
[k] *Departamento de Anatomia Patológica, Centro Hospitalar Universitário de São João, EPE, Porto, Portugal*
[l] *Departamento de Medicina Interna, Centro Hospitalar Universitário de São João, EPE, Porto, Portugal*
[m] *UnIC@RISE, Department of Surgery and Physiology, Portugal*

KEYWORDS
Hypersensitivity pneumonitis;
MUC5B;

Abstract
Introduction and Objectives: Hypersensitivity pneumonitis (HP) is an interstitial lung disease with diverse clinical features that can present a fibrotic phenotype similar to idiopathic pulmonary fibrosis (IPF) in genetically predisposed individuals. While several single nucleotide

Abbreviations: HP, Hypersensitivity pneumonitis; ILDs, Interstitial lung diseases; fHP, Fibrotic hypersensitivity pneumonitis; IPF, Idiopathic pulmonary fibrosis; SNPs, Single nucleotide polymorphisms; MUC5B, Mucin 5B; TOLLIP, Toll-interacting protein; FVC, Forced vital capacity; TLC, Total lung capacity; DLCO, Diffusing capacity of the lung for carbon monoxide; % pred, percentage of predicted; % pred, percentage of predicted; BAL, Bronchoalveolar lavage; RFLP, Restriction fragment length polymorphism; PCR, Polymerase chain reaction; dbSNP, Single nucleotide polymorphisms database; HRCT, High-resolution computed tomography; UIP, Usual interstitial pneumonia; TBLC, Transbronchial lung cryobiopsy; LD, Linkage disequilibrium; MAF, Minor allele frequency; HR, Hazard Ratio.

* Corresponding author at: Laboratório de Apoio à Investigação em Medicina Molecular, Núcleo de Recursos Laboratoriais, Unidade de Gestão de Conhecimento, Departamento de Recursos Comuns, Faculdade de Medicina da Universidade do Porto, Rua Dr. Plácido Costa, s/n - Piso 4, 4200-450 Porto, Portugal.
E-mail address: mlsoares@med.up.pt (M.L. Soares).

https://doi.org/10.1016/j.pulmoe.2024.01.002
2531-0437/© 2024 Sociedade Portuguesa de Pneumologia. Published by Elsevier España, S.L.U. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Please cite this article in press as: P.C. Mota, M.L. Soares, A.C. Ferreira et al., Polymorphisms and haplotypes of *TOLLIP* and *MUC5B* are associated with susceptibility and survival in patients with fibrotic hypersensitivity pneumonitis, Pulmonology (2024), https://doi.org/10.1016/j.pulmoe.2024.01.002

ARTICLE IN PRESS
JID: PULMOE [mSP6P;February 2, 2024;18:04]

TOLLIP;
Genetic variants;
Haplotypes;
Susceptibility;
Survival;
Fibrosis

polymorphisms (SNPs) have been associated with IPF, the genetic factors contributing to fibrotic HP (fHP) remain poorly understood. This study investigated the association of *MUC5B* and *TOLLIP* variants with susceptibility, clinical presentation and survival in Portuguese patients with fHP.

Material and Methods: A case-control study was undertaken with 97 fHP patients and 112 controls. Six SNPs residing in the *MUC5B* and *TOLLIP* genes and their haplotypes were analyzed. Associations with risk, survival, and clinical, radiographic, and pathological features of fHP were probed through comparisons among patients and controls.

Results: *MUC5B* rs35705950 and three neighboring *TOLLIP* variants (rs3750920, rs111521887, and rs5743894) were associated with increased susceptibility to fHP. Minor allele frequencies were greater among fHP patients than in controls (40.7% vs 12.1%, $P<0.0001$; 52.6% vs 40.2%, $P = 0.011$; 22.7% vs 13.4%, $P = 0.013$; and 23.2% vs 12.9%, $P = 0.006$, respectively). Haplotypes formed by these variants were also linked to fHP susceptibility. Moreover, carriers of a specific haplotype (G-T-G-C) had a significant decrease in survival (adjusted hazard ratio 6.92, 95% CI 1.73−27.64, $P = 0.006$). Additional associations were found between *TOLLIP* rs111521887 and rs5743894 variants and decreased lung function at baseline, and the *MUC5B* SNP and radiographic features, further highlighting the influence of genetic factors in fHP.

Conclusion: These findings suggest that *TOLLIP* and MUC5B variants and haplotypes may serve as valuable tools for risk assessment and prognosis in fibrotic hypersensitivity pneumonitis, potentially contributing to its patient stratification, and offer insights into the genetic factors influencing the clinical course of the condition.

© 2024 Sociedade Portuguesa de Pneumologia. Published by Elsevier España, S.L.U. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Introduction

Hypersensitivity pneumonitis (HP) is an interstitial lung disease (ILD), with a heterogeneous clinical presentation and evolution, recently classified as comprising two forms – fibrotic and non-fibrotic.[1-3] HP is triggered by exposure to a variety of inciting antigens that results in inflammation, causes persistent lung injury and, in genetically predisposed individuals, leads to fibrotic disease, similar to idiopathic pulmonary fibrosis (IPF).[4,5] The differential diagnosis of these two fibrosing entities is challenging but imperative given the therapeutic and prognostic implications.[1,4,6,7] In fibrotic HP (fHP), antigen avoidance, if recognized, is the mainstay of treatment, associated with immunosuppression, when necessary.[8,9] Antifibrotics, the standard treatment for IPF, are only prescribed later in cases of HP with progressive pulmonary fibrosis.[10-12]

Several single nucleotide polymorphisms (SNPs) have been associated with IPF susceptibility and survival, such as the mucin 5B (*MUC5B*) rs35705950 promoter polymorphism and several toll-interacting protein (*TOLLIP*) SNPs.[13-17] Whether independently associated or integrating haplotypes potentially predicting altered survival,[18] these variants highlight the contribution of genetic determinants to the development and evolution of the fibrosing phenotype.

In fHP, in contrast, very little is known about predisposing genetic factors.[19] To date, no genome-wide association studies have been performed although a few studies probing candidates chosen on the basis of association with other ILDs have identified associations with disease susceptibility and outcomes in fHP.[20-22] Among those, the *MUC5B* promoter variant and telomere length were implicated as susceptibility and survival determinants in patients with chronic HP,[20,23,24] as extensively reported for IPF.[25] No study has yet focused on the involvement of the various IPF-related *TOLLIP* SNPs in the context of fHP.

The objective of this study was to determine whether the *MUC5B* promoter and neighboring *TOLLIP* polymorphisms, and respective haplotypes, are associated with susceptibility, clinical presentation and survival in patients with fHP.

Materials and methods

Participants and study design

A case-control study was conducted with 97 non-familial fHP patients followed from 2013 to 2022 in a tertiary Portuguese hospital in Porto. This group included cases enrolled in the FIBRALUNG (Fibrosing ILD Biomarkers That Rule Acceleration) project, the first national registry and biobank of ILDs.[26] A control group with 112 subjects was selected from EPIPorto,[27] a population-based cohort of healthy adults ($n = 74$), and NETDiamond,[28,29] a prospective cohort for the study of therapeutic targets in heart failure, selecting those without respiratory disease ($n = 38$).

Diagnosis of fHP was established according to international guidelines.[1,2,6] Demographic and clinical information, including lung function, bronchoalveolar lavage (BAL) as well as radiographic and histopathological data, were collected on enrolment.

The study protocol was reviewed and approved by the Ethics Committee of the Institution and performed in accordance with the Helsinki Declaration. Written informed consent was obtained from all participants.

Genotyping

Individuals were genotyped for six SNPs: *TOLLIP* rs3750920, rs111521887, rs5743894, rs5743890, rs5743854, and *MUC5B* rs35705950, as described,[18] and by RFLP and real-time PCR (details in Supplementary Methods M1, Table S1).

ARTICLE IN PRESS
JID: PULMOE [mSP6P;February 2, 2024;18:04]
Pulmonology 00 (xxxx) 1–11

Radiology

High-resolution computed tomography (HRCT) scans were independently reviewed by a radiologist with experience in ILD CT scan interpretation, blinded to clinical diagnosis. Detailed information on image analysis and classification scores can be found in Supplementary Methods M2.

Histopathology

Lung biopsies were scored and reviewed by an expert lung pathologist in ILDs (surgical lung biopsy or transbronchial lung cryobiopsy) blinded to the original diagnosis, using a guided pathology data collection form. Detailed information on histopathological analyses and classification scores can be found in Supplementary Methods M3.

Statistical analysis

Statistical analyses were performed in R (v4.2.0). Haplotype frequencies were estimated using Haploview v4.2. Pairwise linkage disequilibrium (LD) between SNPs was calculated using SNPAlyze v8.1.1.0 (Dynacom Co., Japan).

Continuous data are presented as means ± standard deviation or medians and interquartile range. Categorical data are presented as frequencies and percentages. Chi-squared tests were used to compare allele and genotype frequencies between patients and controls.

Overall survival was analyzed with time from diagnosis to death (in months) censured by end of follow-up or lung transplantation. Transplant-free survival considered lung transplantation equivalent to death. Kaplan-Meier plots and log-rank tests were used for survival and transplant-free survival.

Univariate and multivariate Cox proportional hazards regressions were employed to identify variables that predict survival status, with results being expressed as hazard ratio (HR) and respective 95% confidence interval (CI).

Two-sided P-values lower than 0.05 were considered statistically significant.

Results

Demographic and clinical features of the study subjects

The baseline characteristics of participants are shown in Table 1. There were no significant differences between the groups regarding gender and smoking history, although fHP patients were younger than control subjects (mean 67.5 vs 71.2 years, P = 0.002). Antigen exposure was identified in 93% of patients, with avian and molds being the most frequently reported.

All patients had radiographic signs of fibrosis. Histological confirmation of fHP diagnosis was necessary in 55.7% of patients. Most patients were initially treated with immunosuppressants with the addition of antifibrotics in cases of progressive pulmonary fibrosis.

Overall median follow-up duration, defined as time from diagnosis to death/transplant or end of study, was 24 months. Twenty-six patients (26.8%) died during follow-up.

MUC5B and *TOLLIP* associations with fHP susceptibility and survival

Genotype frequencies of the probed SNPs are distributed as shown in Table 2, all conforming to the Hardy-Weinberg equilibrium (Supplementary Table S2).

The frequency of the *MUC5B* rs35705950 minor allele T was significantly increased in fHP patients (40.7% vs 12.1%, OR=5.01, 95% CI 3.06–8.21, $P < 0.0001$). Accordingly, the genotypes of T allele carriers were significantly more frequent in the fHP group (69.1% vs 22.3 %, OR=7.77, 95 % CI 4.18-14.43, $P < 0.0001$), indicating this variant allele also confers susceptibility to fHP.

TOLLIP rs3750920, rs111521887 and rs5743894 SNPs (henceforth also referred to as T920, T887 and T894) were significantly associated with fHP. The minor allele frequencies (MAF) were significantly increased in fHP patients relative to those of controls (52.6% vs 40.2%, P = 0.011; 22.7% vs 13.4%, P = 0.013; and 23.2% vs 12.9%, P = 0.006; respectively). The overall genotypic distributions of these SNPs were also associated with fHP (P = 0.029, P = 0.049, P = 0.029), with homozygotes bearing the minor alleles at higher risk of developing the disease (significantly so for T920 TT OR=2.90, P = 0.012; and T894 CC OR=5.04, P = 0.040; and borderline significant for T887 GG, P = 0.052). Allele and genotype distributions of *TOLLIP* rs5743890 and rs5743854 SNPs did not differ significantly between patients and controls.

Kaplan-Meier survival analysis of *MUC5B* and *TOLLIP* SNPs did not show any association between individual genotypes and survival in fHP patients (Supplementary Fig. S1).

MUC5B and *TOLLIP* haplotype associations with fHP susceptibility and survival

Haplotypes defined by the *MUC5B* and three neighboring *TOLLIP* SNPs that had shown an association with susceptibility to fHP (MUC5B-T920-T887-T894 haplotype block) were reconstructed (Supplementary Table S3). Six major haplotypes were defined, of which five showed an association with susceptibility to the disease. Haplotypes T-T-C-T, T-T-G-C, and T-C-C-T were significantly more frequent, and haplotypes G-C-C-T and G-T-C-T were significantly less frequent in patients than in controls.

Pairwise LD measures indicated moderate to strong LD among the *TOLLIP* SNPs and moderate LD between these and the *MUC5B* SNP in the fHP group, while in controls linkage between tightly linked T887 and T894 and the T920 SNP was only moderate, and with the *MUC5B* SNP was considerably low (Supplementary Table S4 and Fig. S2).

Notably, Kaplan-Meier survival analysis of the haplotypes revealed a statistically significant association of haplotype G-T-G-C with reduced survival (censoring for lung transplant) (log rank test P-value = 0.006) and reduced transplant-free survival (log rank test P-value = 0.021) in fHP (Fig. 1). Multivariate Cox regression analysis showed that the presence of the G-T-G-C haplotype was independently associated with shorter survival after adjustment for age and lung function parameters (overall survival, HR 6.92, P = 0.006; transplant-free survival, HR 5.93, P = 0.009)

ARTICLE IN PRESS
JID: PULMOE [mSP6P;February 2, 2024;18:04]

Table 1 Clinical and demographic characteristics of fHP patients and controls.

	fHP (n = 97)	Controls (n = 112)	p value
Gender (n,%)			0.375
Female	44 (45.4 %)	44 (39.3 %)	
Male	53 (54.6 %)	68 (60.7 %)	
Age (years)[a]	67.5 (8.7)	71.2 (8.4)	0.002
Ever smoker (n,%)	47 (48.5 %)	51 (45.9 %)	0.718
Follow up time (months)[b]	24 (16–56)		
Deaths (n,%)	26 (26.8 %)		
Transplants (n,%)	4 (4.1 %)		
Pulmonary function tests			
FVC (% pred)[a]	80.8 (17.3)		
TLC (% pred)[a]	76.9 (17.7)		
DLCO (% pred)[a]	53.9 (16.8)		
Antigen exposure (n,%)			
Avian	70 (72.2 %)		
Molds	37 (38.1 %)		
Other Organics	18 (18.6 %)		
Inorganics	21 (21.6 %)		
Unknow	7 (7.2 %)		
Multiple	46 (47.4 %)		
BAL cell count			
Total cells ($\times 10^4 mL^{-1}$)[b]	1.6 (0.7–2.8)		
Macrophages[b]	71.9 (58.7–81.4)		
Lymphocytes[b]	12 (6.6–20.1)		
Neutrophils[b]	5.4 (3.2–11.2)		
Eosinophils[b]	2.8 (1.2–7.1)		
Lung biopsy (n,%)	54 (55.7 %)		
Treatment (n,%)			
Immunosuppressants	53 (54.6 %)		
Imunosuppressants+Antifibrotics	25 (25.8 %)		
No treatment	19 (19.6 %)		
Comorbidities (n,%)			
COPD	3 (3.1 %)		
Pulmonary hypertension	7 (7.2 %)		
Lung cancer	13 (13.4 %)		
OSA	39 (40.2 %)		
GERD	21 (21.6 %)		
Metabolic conditions	29 (29.9 %)		
Cardiovascular disease	55 (56.7 %)		
Depression	20 (20.6 %)		

Values are presented as [a]mean (±standard deviation), [b]median (Q1-Q3) and number (%).
fHP, fibrotic hypersensitivity pneumonitis; COPD, chronic obstructive lung disease; OSA, obstructive sleep apnea; GERD, gastro-esophageal reflux disease;% pred, percentage of predicted normal value; FVC, forced vital capacity; TLC, total lung capacity; DLCO, diffusing capacity of the lung for carbon monoxide.

(Table 3). Despite being a relatively low frequency haplotype, the difference in median survival time between non-carriers and carriers was 21 months.

Other predictors of survival

Univariate analyses did not reveal associations with gender, age, and smoking. Conversely, better pulmonary function (FVC, TLC and DLCO) associated significantly with longer survival (HR 0.97, P = 0.015; HR 0.98 P = 0.019 and HR 0.96, P = 0.002; respectively) (Table 3).

MUC5B and *TOLLIP* associations with pulmonary function and clinical features

The presence of at least one minor allele of *TOLLIP* SNPs T887 and T894 was associated with lower baseline FVC% pred (genotypes GG/CG vs. CC, 76.3% vs 83.6%, P = 0.045; and genotypes CC/CT versus TT, 75.9% vs 83.9%, P = 0.027; respectively), whereas for T920 it was of borderline statistical significance (79.2% vs 87.5%, P = 0.060) (Table 4). *MUC5B* rs35705950 and the *TOLLIP* T890 and T854 polymorphisms showed no such association (Table 4 and Supplementary Table S5).

ARTICLE IN PRESS
JID: PULMOE [mSP6P;February 2, 2024;18:04]
Pulmonology 00 (xxxx) 1–11

Table 2 Genotype and allele distributions of *MUC5B* and *TOLLIP* SNPs in fHP patients and controls.

	Genotype			*P* value	Allele		*P* value
***MUC5B* rs35705950**				**< 0.0001**			**<0.0001**
	GG (%)	**GT** (%)	**TT** (%)		**G** (%)	**T** (%)	
HP	30 (30.9)	55 (56.7)	12 (12.4)		115 (59.3)	79 (40.7)	
Control	87 (77.7)	23 (20.5)	2 (1.8)		197 (87.9)	27 (12.1)	
	–	OR 6.93 (3.71–13.37)	OR 17.40 (4.42–116.03)		OR 5.01 (3.06–8.21)		
P*	–	**<0.0001**	**<0.0001**				
***TOLLIP* rs3750920**				**0.029**			**0.011**
	CC (%)	**CT** (%)	**TT** (%)		**C** (%)	**T** (%)	
HP	19 (19.6)	54 (55.7)	24 (24.7)		92 (47.4)	102 (52.6)	
Control	39 (34.8)	56 (50.0)	17 (15.2)		134 (59.8)	90 (40.2)	
	–	OR 1.98 (1.03–3.90)	OR 2.90 (1.28–6.75)		OR 1.65 (1.12–2.43)		
P*		**0.044**	**0.012**				
***TOLLIP* rs111521887**				**0.049**			**0.013**
	CC (%)	**CG** (%)	**GG** (%)		**C** (%)	**G** (%)	
HP	60 (61.9)	30 (30.9)	7 (7.2)		150 (77.3)	44 (22.7)	
Control	84 (75.0)	26 (23.2)	2 (1.8)		194 (86.6)	30 (13.4)	
					OR 1.90 (1.14–3.16)		
P*		0.130	0.052				
***TOLLIP* rs5743894**				**0.029**			**0.006**
	TT (%)	**TC** (%)	**CC** (%)		**T** (%)	**C** (%)	
HP	59 (60.8)	31 (32.0)	7 (7.2)		149 (76.8)	45 (23.2)	
Control	85 (75.9)	25 (22.3)	2 (1.8)		195 (87.1)	29 (12.9)	
	–		OR 5.04 (1.01–25.13)		OR 2.03 (1.22–3.39)		
P*		0.066	**0.040**				
***TOLLIP* rs5743890**				0.227			0.245
	CC (%)	**TC** (%)	**TT** (%)		**T** (%)	**C** (%)	
HP	0 (0.0)	25 (25.8 %)	72 (74.2)		169 (87.1)	25 (12.9)	
Control	3 (2.7)	32 (28.6 %)	77 (68.8)		186 (83.0)	38 (17.0)	
P*	–	0.985	0.985				
***TOLLIP* rs5743854**				0.668			0.371
	CC (%)	**CG** (%)	**GG** (%)		**C** (%)	**G** (%)	
HP	77 (79.4)	18 (18.6)	2 (2.1)		172 (88.7)	22 (11.3)	
Control	83 (74.1)	26 (23.2)	3 (2.7)		192 (85.7)	32 (14.3)	
P*	–	0.396	0.721				

OR, Odds Ratio (95 % Confidence Interval).
*P**, Logistic regression *P* value of each minor allele genotype vs. reference genotype (ancestral allele).
Alleles are reported in the Forward strand orientation, except rs5743854 in conformity with the dbSNP database minor allele designation.

Exposure to birds as an inciting antigen was tendentially more frequent in carriers of the G-T-G-C haplotype ($P = 0.067$). Indeed, all carriers of the haplotype were exposed to birds, whereas 30.3% of the non-carriers were not (Table 4).

Radiographic analyses showed that *MUC5B* rs35705950 minor alleles were associated with a lower total score of fibrosis ($P = 0.023$), but not to honeycombing or a pattern indicative of definite or probable UIP. No associations between radiographic features and genotype distributions were seen for the *TOLLIP* SNPs (Table 5).

BAL sample analyses revealed a tendency for a larger amount of neutrophils in patients bearing at least one of the minor alleles of T887 (G) and T894 (C) polymorphisms. Conversely, for *MUC5B* rs35705950 this trend was noted but for the ancestral allele (G). Accordingly, carriers of the G-T-G-C haplotype, which is associated with poorer prognosis, showed a

ARTICLE IN PRESS
JID: PULMOE [mSP6P;February 2, 2024;18:04]

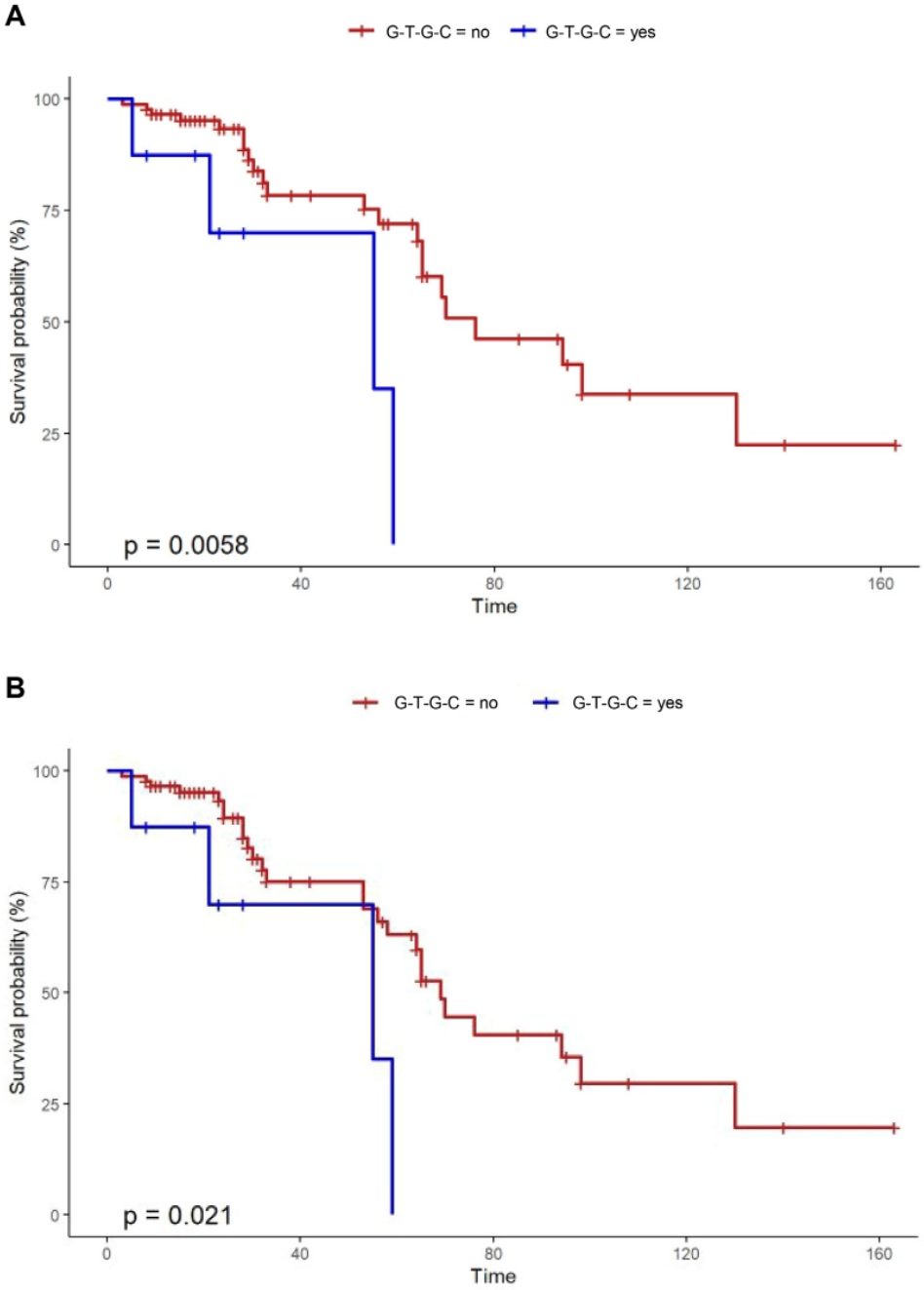


Fig. 1 Kaplan-Meier plots showing fHP patients' overall survival (A) and transplant-free survival (B) according to haplotype G-T-G-C defined by the rs35705950-rs3750920-rs111521887-rs5743894 *MUC5B-TOLLIP* block.
Time in months; P value for log-rank test.

median neutrophil amount that more than doubled that of non-carriers although not reaching statistical significance ($P = 0.085$) (Table 5).

No additional significant associations emerged between the probed SNPs or respective haplotypes and histopathological features of fibrosis (Table 5).

Discussion

This study investigated specific genetic variants as determinants of susceptibility and survival in patients with fHP. The *MUC5B* promoter rs35705950 variant and *TOLLIP* rs3750920, rs111521887 and rs5743894 variants are associated with

ARTICLE IN PRESS
JID: PULMOE [mSP6P;February 2, 2024;18:04]
Pulmonology 00 (xxxx) 1–11

Table 3 Univariate and multivariate Cox proportional hazard regressions predicting patients' death.

	HR	95 % CI	p-value
Univariate analysis			
Gender	1.36	0.63 – 2.92	0.434
Age	1.05	1.00 – 1.11	0.060
Smoking	1.70	0.79 – 3.66	0.173
FVC (% pred)	0.97	0.95 – 0.99	**0.015**
TLC (% pred)	0.98	0.97 – 1.00	**0.019**
DLCO (% pred)	0.96	0.93 – 0.98	**0.002**
G-T-G-C haplotype	4.26	1.38 – 13.13	**0.012**
*Multivariate analysis**			
Overall survival			
G-T-G-C haplotype	6.92	1.73 – 27.64	**0.006**
Age	1.08	1.00 – 1.16	**0.046**
FVC (% pred)	0.99	0.94 – 1.03	0.547
TLC (% pred)	0.99	0.96 – 1.02	0.535
DLCO (% pred)	0.97	0.94 – 1.01	0.161
Transplant-free survival			
G-T-G-C haplotype	5.93	1.55 – 22.68	**0.009**
Age	1.04	0.98 – 1.11	0.202
FVC (% pred)	0.97	0.93 – 1.01	0.206
TLC (% pred)	0.99	0.96 – 1.02	0.486
DLCO (% pred)	0.98	0.95 – 1.01	0.186

HR, Hazard Ratio; 95 % CI, Confidence Interval.
% pred, percentage of predicted normal value.
FVC, Forced Vital Capacity; TLC, Total Lung Capacity; DLCO, Diffusing Capacity of the Lung for Carbon Monoxide.
* Multivariate Cox regression models obtained by forward conditional stepwise selection.

susceptibility to fHP in this Portuguese cohort. Although such an association had been previously reported for *MUC5B* rs35705950,[20-22] this is the first study implicating *TOLLIP* variants as susceptibility determinants in fHP. Notably, we also found that a haplotype defined by these four variants is significantly associated with poorer survival in these patients.

The role of the *MUC5B* polymorphism in the development and progression of lung fibrosis has been extensively investigated in IPF and is considered the most influential genetic risk factor of this disease.[13-15,30] The involvement of *MUC5B* in HP was described in recent studies[20-22] with results corroborated by those presented herein. Furusawa et al identified six common IPF genetic risk variants with a positive direction of risk effect in HP cases, with the *MUC5B* variant showing the strongest association.[21]

Other common variants reportedly associated with IPF susceptibility include SNPs in the genomic region of the *TOLLIP* gene. Notwithstanding some inconsistent reports, it is generally accepted that *TOLLIP* polymorphisms play a role in IPF pathophysiology. The minor alleles of intronic SNPs rs111521887 and rs5743894 were reported to confer susceptibility to IPF, while that of the rs5743890 SNP was found to reduce such risk.[30,31] Moreover, *TOLLIP*'s exonic variant rs3750920 was associated with a favorable response to N-acetylcysteine, with carriers of the TT genotype showing a significantly reduced composite endpoint risk, including death, hospitalization, and FVC decline.[17]

Table 4 *MUC5B* and *TOLLIP* associations with demographic and clinical features.

Genotype/Haplotype	*MUC5B* rs35705950			*TOLLIP* rs3750920			*TOLLIP* rs111521887			*TOLLIP* rs5743894			G-T-G-C haplotype*		
	GG (*n* = 30)	GT/TT (*n* = 67)	P value	CC (*n* = 19)	CT/TT (*n* = 78)	P value	CC (*n* = 60)	CG/GG (*n* = 37)	P value	TT (*n* = 59)	TC/CC (*n* = 38)	P value	No (*n* = 89)	Yes (*n* = 8)	P value
Gender (female/male)	11 (36.7 %)	33 (49.3 %)	0.250	11 (57.9 %)	33 (42.3 %)	0.221	23 (38.3 %)	21 (56.8 %)	0.077	23 (39.0 %)	21 (55.3 %)	0.116	39 (43.8 %)	5 (62.5 %)	0.309
	19 (63.3 %)	34 (50.7 %)		8 (42.1 %)	45 (57.7 %)		37 (61.7 %)	16 (43.2 %)		36 (61 %)	17 (44.7 %)		50 (56.2 %)	3 (37.5 %)	
Age (years)	66.2 (8.7)	68.1 (8.7)	0.324	68.8 (8.7)	67.1 (8.7)	0.447	67.4 (8.7)	67.5 (8.8)	0.953	67.6 (8.6)	67.2 (8.7)	0.812	67.4 (8.9)	68.3 (5.8)	0.794
Ever smoker	17 (56.7 %)	30 (44.8 %)	0.279	9 (47.4 %)	38 (48.7 %)	0.916	31 (51.7 %)	16 (43.2 %)	0.420	30 (50.8 %)	17 (44.7 %)	0.557	44 (49.4 %)	3 (37.5 %)	0.518
Death	9 (30.0 %)	17 (25.4 %)	0.629	6 (31.6 %)	20 (25.6 %)	0.577	17 (28.3 %)	9 (24.3 %)	0.814	17 (28.8 %)	9 (28.7 %)	0.644	22 (24.7 %)	4 (50.0 %)	0.204
Pulmonary function tests															
FVC (% pred)	82.8 (16.3)	79.9 (17.7)	0.449	87.5 (14.2)	79.2 (17.6)	0.060	83.6 (14.8)	76.3 (20.1)	**0.045**	83.9 (14.7)	75.9 (19.9)	**0.027**	81.1 (16.9)	77.7 (23.1)	0.623
TLC (% pred)	76.4 (15.2)	77.2 (18.9)	0.855	80.7 (13.6)	76.0 (18.6)	0.309	77.8 (14.4)	75.4 (22.4)	0.536	77.8 (14.5)	75.4(22.1)	0.397	76.7 (17.3)	79.1 (24.3)	0.741
DLCO (% pred)	50.4 (16.8)	55.6 (16.8)	0.184	51.1 (10.2)	54.6 (18.0)	0.454	53.8 (15.4)	54.1 (19.4)	0.929	53.9 (15.6)	54.0 (19.1)	0.900	54.1 (16.2)	51.8 (24.8)	0.730
Antigen exposure															
Avian	21 (70.0 %)	49 (73.1 %)	0.750	11 (57.9 %)	59 (75.6 %)	0.122	40 (66.7 %)	30 (81.1 %)	0.124	40 (67.8 %)	30 (78.9 %)	0.232	62 (69.7 %)	8 (100 %)	0.067
Molds	10 (33.3 %)	27 (40.3 %)	0.514	10 (52.6 %)	27 (34.6 %)	0.147	21 (35 %)	16 (43.2 %)	0.417	21 (35.6 %)	16 (42.1 %)	0.519	33 (37.1 %)	4 (50.0 %)	0.471
Other organics	4 (13.3 %)	14 (20.9 %)	0.376	2 (10.5 %)	16 (20.5 %)	0.315	9 (15.0 %)	9 (24.3 %)	0.251	9 (15.3 %)	9 (23.7 %)	0.297	17 (19.1 %)	1 (12.5 %)	0.645
Inorganics	7 (23.3 %)	14 (20.9 %)	0.788	3 (15.8 %)	18 (23.1 %)	0.489	13 (21.7 %)	8 (21.6 %)	0.996	12 (20.3 %)	9 (23.7 %)	0.696	19 (21.3 %)	2 (25.0 %)	0.810

Values are presented as mean (SD) and counts as n (%).
* Defined by the rs35705950-rs3750920 -rs111521887- rs5743894 *MUC5B-TOLLIP* block.
FVC, forced vital capacity; TLC, total lung capacity; DLCO, diffusing capacity of the lung for carbon monoxide.

ARTICLE IN PRESS
JID: PULMOE [mSP6P;February 2, 2024;18:04]

Table 5 MUC5B and TOLLIP associations with BAL, radiographic and histopathological features.

Genotype/Haplotype	*MUC5B* rs35705950			*TOLLIP* rs3750920			*TOLLIP* rs11521887			*TOLLIP* rs5743894			G-T-G-C haplotype*		
	GG (*n* = 30)	GT/TT (*n* = 67)	P value	CC (*n* = 19)	CT/TT (*n* = 78)	P value	CC (*n* = 60)	CG/GG (*n* = 37)	P value	TT (*n* = 59)	TC/CC (*n* = 38)	P value	No (*n* = 89)	Yes (*n* = 8)	P value
BAL features															
Lymphocytes	13.8 (9.6–21.1)	11.0 (6–19.2)	0.117	11.1 (5.2–13.7)	12.0 (7.0–20.2)	0.322	12.2(6.6–19.4)	11.0 (6.5–20.9)	0.615	12.2(6.6–19.4)	9.7 (6.0–20.6)	0.384	12(6.6–20.9)	8.1 (6.4–16.0)	0.377
Neutrophils	6.4 (3.4–12.1)	5.4 (3.2–11.0)	0.283	4.0 (3–8)	5.8 (3.2–11.2)	0.329	5.0 (3.2–9.9)	8.0 (3.8–12.2)	0.142	4.9 (3.2–9.4)	8.0 (4.2–13.5)	0.082	5.5 (3.2–10.5)	11.2(7.9–18.2)	0.085
Eosinophils	3.0 (1.2–12.7)	2.8 (1.4–5.8)	0.660	3.0 (0.6–9.2)	2.7 (1.4–2.6)	0.882	2.5 (1.1–5.8)	4.0 (1.6–9.3)	0.356	2.4 (1.0–5.8)	4.4 (1.7–9.9	0.210	2.8 (1.1–6.6)	4.1 (2.3–10.4)	0.332
Radiographic features															
UIP + probable UIP	5 (17.9 %)	23 (34.8 %)	0.238	4 (22.2 %)	24 (31.5 %)	0.743	16 (28.1 %)	12 (32.5 %)	0.490	16 (28.6 %)	12 (31.6 %)	0.616	25 (29.0 %)	3 (37.5 %)	0.755
Indeterminated for UIP	8 (28.6 %)	13 (19.7 %)		5 (27.8 %)	16 (21.1 %)		11 (19.3 %)	10 (27.0 %)		11 (19.6 %)	10 (26.3 %)		20 (23.3 %)	1 (12.5 %)	
Alternative diagnosis	15 (53.6 %)	30 (45.5 %)		9 (50 %)	36 (47.4 %)		30 (52.6 %)	15 (40.5 %)		29 (51.8 %)	16 (42.1 %)		41 (47.7 %)	4 (50.0 %)	
Honeycombing	13 (46.4 %)	24 (36.4 %)	0.361	8 (44.4 %)	29 (38.2 %)	0.624	25 (43.9 %)	12 (32.4 %)	0.268	24 (42.9 %)	13 (34.2 %)	0.400	35 (40.7 %)	2 (25.0 %)	0.385
Fibrosis score	9.5 (7–12)	8 (5.0–10.8)	**0.023**	8.5 (5.8–11)	8 (6–11)	0.979	9 (6.3–11)	8 (6.0–11.0)	0.482	8.5 (6.0–11.0)	8 (6.0–10.8)	0.516	8 (6.0–11.0)	10 (6.0–12.5)	0.455
Histopathological features															
Honeycombing	6 (33.3 %)	14 (38.9 %)	0.690	3 (21.2 %)	17 (42.5 %)	0.160	12 (33.3 %)	8 (44.4 %)	0.425	11 (31.4 %)	9 (47.4 %)	0.247	18 (35.3 %)	2 (66.7 %)	0.274
Area of fibrosis (%)	31.7 (16.2)	31.4 (13.1)	0.941	30.7 (17.2)	31.8 (13.0)	0.806	31.1 (14.7)	32.2 (12.9)	0.775	31.0 (14.9)	32.4 (12.6)	0.735	31.5 (14.2)	30.2 (12.4)	0.866
Fibroblast foci															
0	10 (55.6 %)	12 (33.3 %)	0.270	3(21.4 %)	19 (47.5 %)	0.206	17 (47.2 %)	5 (27.8 %)	0.390	16 (45.7 %)	6 (31.5 %)	0.600	20 (39.2 %)	2 (66.7 %)	0.561
1	3 (16.7 %)	7 (19.5 %)		4 (28.6 %)	6 (15.0 %)		6 (16.7 %)	4 (22.2 %)		6 (17.2 %)	4 (21.1 %)		10 (19.6 %)	0 (0.0 %)	
>1	5 (27.7 %)	17 (47.2 %)		7 (50.0 %)	15 (37.5 %)		13 (36.1 %)	9 (50.0 %)		13 (37.1 %)	9 (47.4 %)		21 (41.2 %)	1 (33.3 %)	
Fibrosis distribution															
Peribronchiolar	6 (33.3 %)	7 (20.0 %)	0.677	4 (28.6 %)	9 (23.1 %)	0.816	10 (28.6 %)	3 (16.7 %)	0.428	10 (29.4 %)	3 (15.8 %)	0.438	12 (24.0 %)	1 (33.3 %)	0.849
Peribronchiolar (+paraseptal or subpleural)	8 (44.4 %)	20 (57.1 %)		7 (50.0 %)	21 (53.8 %)		16 (45.7 %)	12 (66.7 %)		16 (47.1 %)	12 (63.2 %)		27 (54.0 %)	1 (33.3 %)	
Heterogeneous	3 (16.7 %)	7 (20.0 %)		2 (14.3 %)	8 (20.5 %)		7 (20.0 %)	3 (16.7 %)		6 (17.6 %)	4 (21.1 %)		9 (18.0 %)	1 (33.3 %)	
Undetermined	1 (5.6 %)	1 (2.9 %)		1 (7.1 %)	1 (2.6 %)		2 (5.7 %)	0 (0.0 %)		2 (5.9 %)	0 (0.0 %)		2 (4.0 %)	0 (0.0 %)	
Inflammatory infiltrate															
Light	14 (77.8 %)	24 (66.7 %)	0.532-	10 (71.4 %)	28 (70.0 %)	0.920	26 (72.2 %)	12 (66.7 %)	0.756	26 (74.3 %)	12 (63.2 %)	0.534	37 (72.5 %)	1 (33.3 %)	0.148
Moderate-severe	4 (22.2 %)	12 (33.3 %)		4 (28.6 %)	12 (30 %)		10 (27.8 %)	6 (33.3 %)		9 (25.7 %)	7 (36.8 %)		14 (27.5 %)	2 (66.7 %)	
Peribronchiolar metaplasia	2(1–3)	1 (1.9–2.3)	0.749	1 (0.0–2.8)	2(1.0–3.0)	0.390	2(0.8–3.0)	1 (1.0–2.0)	0.829	2(0.5–3.0)	1 (1.0–2.5)	0.623	2 (1.0–3.0)	1 (1–1)	0.353
Giant cells granulomas present	10 (55.5 %)	6 (16.7 %)	**0.003**	6 (42.9 %)	10 (25.0 %)	0.208	10 (27.8 %)	6 (33.3 %)	0.673	10 (28.6 %)	6 (31.6 %)	0.817	15 (29.4 %)	1 (33.3 %)	0.885

Values are presented mean (SD) or as median (Q1-Q3) and counts as n (%).
* Defined by the rs35705950-rs3750920 -rs111521887- rs5743894 *MUC5B-TOLLIP* block.
BAL, bronchoalveolar lavage; UIP, usual interstitial pneumonia.

ARTICLE IN PRESS
JID: PULMOE [mSP6P;February 2, 2024;18:04]
Pulmonology 00 (xxxx) 1–11

Having probed five *TOLLIP* SNPs, including those above, we have uncovered a significant correlation between genotypes of *TOLLIP* rs3750920, rs111521887 and rs5743894 and susceptibility to fHP, with the minor alleles associating with higher risk. Interestingly, none of these SNPs independently associated with IPF risk in a previous study with Portuguese patients,[18] raising the possibility that unlike *MUC5B* rs35705950, *TOLLIP* can be an informative marker in a future strategy for differential diagnosis between the two fibrotic diseases, at least in the Portuguese. Several studies have implicated *TOLLIP* as playing a protective role in autophagy and apoptosis.[32,33] *TOLLIP*'s expression is decreased in IPF lungs when compared to normal lungs,[32] which has been linked to minor allele carriers of variants T887, T894 and T890, further suggesting that these could modify the course of disease.[30] In our study, the minor alleles of the former two variants tended to or associated with susceptibility (respectively) and integrated the haplotype associated with poorer prognosis in fHP. Interestingly, genotypes containing at least one minor allele of *TOLLIP* variants rs111521887 and rs5743894 were significantly associated with lower baseline FVC% pred, whereas for rs3750920 the decrease was of borderline statistical significance. In Japanese fHP patients genotyped for *TOLLIP* rs3750920 and rs5743899 (residing between T887 and T894 and not evaluated here), Katayanagi et al reported that the GG (minor allele) genotype of rs5743899 only, was associated with rapid FVC deterioration over time.[34] The relationship between rs3750920 genotypes and FVC at baseline was not stated.

The lack of association between *TOLLIP* rs5743890 and fHP risk is in line with that reported by Ley et al.[20] We are not aware of studies concerning *TOLLIP* rs5743854 in the context of fHP.

We found no association between the *MUC5B* variant and baseline FVC, in contrast with a recent study in Polish patients.[35] However, the latter reported a *MUC5B* MAF of only 17 %, with a GT/TT prevalence of 30.2 %, which contrasts strongly with those reported here and elsewhere.[20]

Ley et al reported that the *MUC5B* rs35705950 minor allele was associated with radiographic evidence of moderate to severe fibrosis and traction bronchiectasis in HP patients.[20] In our study the minor allele associated with a lower total score of fibrosis; nonetheless, HRCT fibrosis was scored differently and the median scores in both genotypic groups (GT/TT=8 and GG=9.5) already represent at least moderate fibrosis.

The use of SNP haplotypes within regions of interest previously identified by single-SNP genome-wide association studies can prove very useful in uncovering and fine-mapping genetic associations,[36] and provide better insights into the influence of genetic loci than the study of SNPs individually.[37,38] We show that four haplotypes, defined by the *MUC5B* promoter SNP and the three *TOLLIP* SNPs that associated individually with susceptibility to fHP, were also strong predictors of fHP risk. Remarkably, carriers of the haplotype G-T-G-C had a significant decrease in survival compared to non-carriers, with the median survival time differing by a remarkable 21 months. This association was confirmed by multivariate analysis adjusting for age and lung function. We had previously reported an association of *MUC5B* plus *TOLLIP*-defined haplotypes with susceptibility and survival in IPF using a similar approach,[18] suggesting that the region comprising these variants contains or segregates with an as-yet unidentified genetic determinant(s) that influence(s) the clinical course of both disorders. This, in spite of the absence of an independent association of each *TOLLIP* variant and the *MUC5B* promoter SNP with fHP survival. These results are in accordance with the report by Ley et al, who also found no independent association between *MUC5B* rs35705950 or *TOLLIP* rs5743890 and survival in patients with fHP.[20]

The G-T-G-C haplotype is also associated with a feature that can negatively impact in disease prognosis. All carriers of the haplotype had been subjected to bird exposure, an antigen that is frequently described in more severe and fibrotic forms of the disease,[39] despite the lack of an association with compromised lung function in our study.

Overall, our findings suggest that *TOLLIP* variants, as well as haplotypes (incorporating *MUC5B* rs35705950), could be of potential use for patients' risk stratification and as markers of prognostic value in fHP. Multicenter and prospective studies are needed to validate these results and to elucidate the role of both genes in the pathophysiology of hypersensitivity pneumonitis.

Funding

This work was supported by the Portuguese Society of Pulmonology (SPP) (SPP/Bolsa Novartis 2012 grant); Boehringer-Ingelheim (Unrestricted Research Grant); Fundação para a Ciência e a Tecnologia, I.P. (PTDC/MEC-RES/0158/2020 grant); and Fundação Amélia de Mello and D. José de Mello.

JCR is funded by Fundação para a Ciência e a Tecnologia through the Stimulus for Scientific Employment Individual Support (2020.01350.CEECIND).

Conflicts of interest

Patrícia Caetano Mota has received speaker fees from Boehringer-Ingelheim and research grants from Boehringer-Ingelheim and Novartis; and has participated in research with Boehringer-Ingelheim and Roche, for which her institution has been remunerated.

Hélder Novais-Bastos and António Morais have attended advisory boards for Boehringer-Ingelheim and Roche; have received speaker fees from Boehringer-Ingelheim and Roche; have participated in research with Boehringer-Ingelheim and Roche, for which their institution has been remunerated. Hélder Novais-Bastos has also received a research grant from Boehringer-Ingelheim.

The remaining authors have no relevant financial or non-financial interests to disclose.

Acknowledgements

We are indebted to Instituto de Saúde Pública da Universidade do Porto (ISPUP) for providing samples from the EPI-Porto cohort to be used as controls; we thank the cohort management and all participants. We are grateful to Catarina Teixeira Antunes and Joana Filipa Amorim Reis for assistance in the genotyping.

ARTICLE IN PRESS
JID: PULMOE [mSP6P;February 2, 2024;18:04]

Supplementary materials

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.pulmoe.2024.01.002.

References

1. Raghu G, Remy-Jardin M, Ryerson CJ, et al. Diagnosis of hypersensitivity pneumonitis in adults. An official ATS/JRS/ALAT clinical practice guideline. Am J Respir Crit Care Med. 2020;202(3):e36–69.
2. Fernandez Perez ER, Travis WD, Lynch DA, et al. Diagnosis and evaluation of hypersensitivity pneumonitis: CHEST guideline and expert panel report. Chest. 2021;160(2):e97–e156.
3. Santos V, Martins N, Sousa C, et al. Hypersensitivity pneumonitis: main features characterization in a Portuguese cohort. Pulmonology. 2020;26(3):130–7.
4. Vasakova M, Selman M, Morell F, Sterclova M, Molina-Molina M, Raghu G. Hypersensitivity pneumonitis: current concepts of pathogenesis and potential targets for treatment. Am J Respir Crit Care Med. 2019;200(3):301–8.
5. Nogueira R, Melo N, Novais EBH, et al. Hypersensitivity pneumonitis: antigen diversity and disease implications. Pulmonology. 2019;25(2):97–108.
6. Marinescu DC, Raghu G, Remy-Jardin M, et al. Integration and application of clinical practice guidelines for the diagnosis of idiopathic pulmonary fibrosis and fibrotic hypersensitivity pneumonitis. Chest. 2022;162(3):614–29.
7. Raghu G, Remy-Jardin M, Richeldi L, et al. Idiopathic pulmonary fibrosis (an update) and progressive pulmonary fibrosis in adults: an official ATS/ERS/JRS/ALAT clinical practice guideline. Am J Respir Crit Care Med. 2022;205(9):e18–47.
8. Terras Alexandre A, Martins N, Raimundo S, et al. Impact of Azathioprine use in chronic hypersensitivity pneumonitis patients. Pulm Pharmacol Ther. 2020;60:101878.
9. Raimundo S, Pimenta AC, Cruz-Martins N, et al. Insights on chronic hypersensitivity pneumonitis' treatment: factors associated with a favourable response to azathioprine. Life Sci. 2021;272:119274.
10. Richeldi L, du Bois RM, Raghu G, et al. Efficacy and safety of nintedanib in idiopathic pulmonary fibrosis. N Engl J Med. 2014;370(22):2071–82.
11. Noble PW, Albera C, Bradford WZ, et al. Pirfenidone in patients with idiopathic pulmonary fibrosis (CAPACITY): two randomised trials. Lancet. 2011;377(9779):1760–9.
12. Flaherty KR, Wells AU, Cottin V, et al. Nintedanib in progressive fibrosing interstitial lung diseases. N Engl J Med. 2019;381(18):1718–27.
13. Seibold MA, Wise AL, Speer MC, et al. A common MUC5B promoter polymorphism and pulmonary fibrosis. N Engl J Med. 2011;364(16):1503–12.
14. Zhang GY, Liao T, Gao WY. MUC5B promoter polymorphism and pulmonary fibrosis. N Engl J Med. 2011;365(2):178.. author reply -9.
15. Peljto AL, Selman M, Kim DS, et al. The MUC5B promoter polymorphism is associated with idiopathic pulmonary fibrosis in a Mexican cohort but is rare among Asian ancestries. Chest. 2015;147(2):460–4.
16. Bonella F, Campo I, Zorzetto M, et al. Potential clinical utility of MUC5B und TOLLIP single nucleotide polymorphisms (SNPs) in the management of patients with IPF. Orphanet J Rare Dis. 2021;16(1):111.
17. Oldham JM, Ma SF, Martinez FJ, et al. TOLLIP, MUC5B, and the response to N-acetylcysteine among individuals with idiopathic pulmonary fibrosis. Am J Respir Crit Care Med. 2015;192(12):1475–82.
18. Mota PC, Soares ML, Vasconcelos CD, et al. Predictive value of common genetic variants in idiopathic pulmonary fibrosis survival. J Mol Med. 2022;100(9):1341–53.
19. Blach J, Mackiewicz B. How much do we know about genetic predisposition of hypersensitivity pneumonitis? Ann Agric Environ Med. 2022;29(2):306–8.
20. Ley B, Newton CA, Arnould I, et al. The MUC5B promoter polymorphism and telomere length in patients with chronic hypersensitivity pneumonitis: an observational cohort-control study. Lancet Respir Med. 2017;5(8):639–47.
21. Furusawa H, Peljto AL, Walts AD, et al. Common idiopathic pulmonary fibrosis risk variants are associated with hypersensitivity pneumonitis. Thorax. 2022;77(5):508–10.
22. Furusawa H, Cardwell JH, Okamoto T, et al. Chronic hypersensitivity pneumonitis, an interstitial lung disease with distinct molecular signatures. Am J Respir Crit Care Med. 2020;202(10):1430–44.
23. Ley B, Torgerson DG, Oldham JM, et al. Rare protein-altering telomere-related gene variants in patients with chronic hypersensitivity pneumonitis. Am J Respir Crit Care Med. 2019;200(9):1154–63.
24. Bouros D, Tzouvelekis A. Telomeropathy in chronic hypersensitivity pneumonitis. Am J Respir Crit Care Med. 2019;200(9):1086–7.
25. Peljto AL, Blumhagen RZ, Walts AD, et al. Idiopathic pulmonary fibrosis is associated with common genetic variants and limited rare variants. Am J Respir Crit Care Med. 2023;207(9):1194–202.
26. https://clinicaltrials.gov/study/NCT05635032?cond=ILD&term=fibralung&rank=1.
27. Ramos E, Lopes C, Barros H. Investigating the effect of nonparticipation using a population-based case-control study on myocardial infarction. Ann Epidemiol. 2004;14(6):437–41.
28. Pintalhao M, Vasques-Novoa F, Couto-Viana B, et al. Relaxin-2, pathophysiological insights and outcomes in heart failure with preserved ejection fraction: findings from the NETDiamond cohort. Int J Cardiol. 2022;365:87–90.
29. Leite AR, Neves JS, Angelico-Goncalves A, et al. Clinical and pathophysiologic insights of free triiodothyronine/free thyroxine ratio in patients with heart failure with preserved ejection fraction: data from the NETDiamond cohort. Cardiology. 2023;148(3):239–45.
30. Noth I, Zhang Y, Ma SF, et al. Genetic variants associated with idiopathic pulmonary fibrosis susceptibility and mortality: a genome-wide association study. Lancet Respir Med. 2013;1(4):309–17.
31. Fingerlin TE, Murphy E, Zhang W, et al. Genome-wide association study identifies multiple susceptibility loci for pulmonary fibrosis. Nat Genet. 2013;45(6):613–20.
32. Li X, Goobie GC, Gregory AD, Kass DJ, Zhang Y. Toll-interacting protein in pulmonary diseases. abiding by the goldilocks principle. Am J Respir Cell Mol Biol. 2021;64(5):536–46.
33. Li X, Goobie GC, Zhang Y. Toll-interacting protein impacts on inflammation, autophagy, and vacuole trafficking in human disease. J Mol Med. 2021;99(1):21–31.
34. Katayanagi S, Setoguchi Y, Kitagawa S, Okamoto T, Miyazaki Y. Alternative gene expression by TOLLIP variant is associated with lung function in chronic hypersensitivity pneumonitis. Chest. 2022;161(2):458–69.
35. Lewandowska KB, Szturmowicz M, Lechowicz U, et al. The presence of T Allele (rs35705950) of the MUC5B gene predicts lower baseline forced vital capacity and its subsequent decline in patients with hypersensitivity pneumonitis. Int J Mol Sci. 2023;24(13).

ARTICLE IN PRESS
JID: PULMOE [mSP6P;February 2, 2024;18:04]
Pulmonology 00 (xxxx) 1–11

36. Khankhanian P, Gourraud PA, Lizee A, Goodin DS. Haplotype-based approach to known MS-associated regions increases the amount of explained risk. J Med Genet. 2015;52(9):587–94.
37. Schrodi SJ, Garcia VE, Rowland C, Jones HB. Pairwise linkage disequilibrium under disease models. Eur J Hum Genet. 2007;15 (2):212–20.
38. Shirali M, Knott SA, Pong-Wong R, Navarro P, Haley CS. Haplotype heritability mapping method uncovers missing heritability of complex traits. Sci Rep. 2018;8(1):4982.
39. Morell F, Roger A, Reyes L, Cruz MJ, Murio C, Munoz X. Bird fancier's lung: a series of 86 patients. Medicine. 2008;87 (2):110–30.

Supplementary Material

Supplementary methods

M1. Genotyping

Genomic DNA was extracted from peripheral blood using the Wizard Genomic DNA Purification Kit (Promega Corp., Madison, USA) or the QIAamp DNA Blood kit (Qiagen, Hilden, Germany).

For restriction fragment length polymorphism (RFLP) analysis, DNA sequences containing the polymorphic sites were amplified by PCR in 20 µl reactions containing 20 ng of genomic DNA, 1.5 mM of each primer, and Kapa Taq EXtra HotStart ReadyMix (Merck, Darmstadt, Germany). Cycling conditions were 95ºC denaturing for 5 min, 35 cycles of 30 sec at 95ºC, 30 sec at annealing temperature, and 30 sec at 72ºC, with a final extension for 10 min at 72ºC. The amplicons were digested overnight in 15µl reactions with 8µl of PCR product and 3 units of restriction endonuclease. TaqMan assays were carried out according to manufacturer's instructions.

Table S1. Genotyping approaches and conditions

Primers, conditions and restriction enzymes for RFLP analyses							
Locus	Primer name	SNP loc (GRCh38.p14)	Primer sequence (5' to 3')	Annealing (°C)	Amplicon Size (bp)	Restriction Enzyme	RFLP fragments (bp)
MUC5B	rs35705950_Fw	5'-regulatory 1219991	GCCAGAATGAGGGACAGTGA	60	375	HhaI	G: 189+105+81 T: 270+105
	rs35705950_Rv		ACTTTGCCCTCGTCCCTC				
TOLLIP	rs5743854_Fw	5'-regulatory 1310024	CATCCCCGTTCCGTTCAC	60	250	SbfI	G: 250 C: 134+116
	rs5743854_Rv		TTCCTCCTCGGTCCTCTGT				
	rs5743890_Fw	Intronic 1304599	GGCTCCCCTGGAAATGGTAA	62	408	EarI	T: 408 C: 310+98
	rs5743890_Rv		GTCAGGACACAGAGAAAAGGC				
	rs5743894_Fw	Intronic 1303542	GGCCTCTAACACTGTCCCTT	62	440	FspI	T: 440 C: 247+193
	rs5743894_Rv		AAAAGCCAGTCAAGCAGCAG				
	rs111521887_Fw	Intronic 1291476	GAGGGTTCAGTGAGGACAGG	60	460	AflIII	C: 460 G: 333+127
	rs111521887_Rv		ACTCAACATGGATGGCAGGT				
	rs3750920_Fw	Exonic 1288726	GGTCTCATCGATGGGGTGTT	62	421	MspI	C: 226+195 T: 421
	rs3750920_Rv		CCCTGCTGTTTTATGGGCTC				

Taqman SNP Genotyping Assays for Real-time PCR			
Locus	**SNP ID**	**Assay ID***	**Context Sequence [VIC/FAM]**
MUC5B	rs35705950	C_1582254_20	CCTTCCTTTATCTTCTGTTTTCAGC[G/T]CCTTCAACTGTGAAGAGGTGAACTC
TOLLIP	rs3750920	C_1582189_1	CCACCTTGCCCTGCCTCAGGGACTC[C/T]GGGATGGTGATGTGGGTCCAGGCAA
	rs111521887	C_152324875_10	GACCCCCTCTGAGGGAACGGCCAAC[C/G]TGTCCACACCAACCCCCCTCTGAGA
	rs5743854	C_11281491_20	GTGACAGGCCCGACATCCCCTGCAG[C/G]GCGATCTTTCCCTCAGCTGCCTGCC

SNP loc, SNP location; RFLP, Restriction Fragment Length Polymorphism; Fw, Forward; Rv, Reverse; bp, base pairs
*TaqMan® SNP Genotyping Assays (Thermo Fisher Scientific, Massachusetts, USA)

M2. Radiology

All high-resolution computed tomography (HRCT) scans were reviewed by a radiologist with experience in ILD chest CT scan interpretation who was blinded to clinical diagnosis. Only thin-section CT images were analyzed (< 1.5 mm thickness; 1-10 mm interval; lung kernels and lung windows). Coronal and sagittal reconstructions were freely accessible. For each scan, a fibrotic pattern was assigned according to current UIP classification, as indicated by the ATS/JRS/ALAT guidelines: 1-usual interstitial pneumonia (UIP) or probable UIP, 2-indeterminate for UIP or 3-alternative diagnosis.[1] The presence of honeycombing, as defined by the Fleischner Society,[2] was recorded, in addition to an assessment of the global extent of interstitial abnormalities. For each pulmonary lobe (with the lingula as part of the left upper lobe) an estimate of the total percentage of lung affected by fibrosis (i.e., reticular abnormality, traction bronchiectasis and honeycombing) was scored: 0- 0%, 1- 0 to 25%, 2- 25 to 50%, 3- 50 to 75%, 4- >75%, yielding a total ILD extent score of 0-20, which can be qualitatively translated to 0: none, 1-5: mild; 6-10 moderate, > 10: severe.

M3. Histopathology

Pathology variables were assessed blindly for the clinical data based on review of slides from one or two biopsied lobes. The new guidelines for diagnosis of HP and fHP were applied to all cases[3]. The tissue area on each slide was measured using QuPath[4] and the quantification area of fibrosis was manually evaluated using the same program.

For dichotomous variables the presence of a feature in one lobe was considered to be present in that patient; the data from all lobes were averaged for continuous variables. The following variables were examined: presence of honeycombing, presence of peribronchiolar metaplasia, presence of giant cells or granulomas (either interstitial or within airspaces), percentage of fibrosis, presence of fibroblastic foci (none, one or more than one), fibrosis distribution (heterogeneous, peribronchiolar only, peribronchiolar associated with subpleural or paraseptal fibrosis, undetermined) and chronic inflammatory infiltrate (none/light if up to 25%; moderate/severe if more than 25%).[3,5]

Supplementary tables and figures

Table S2. Chi-squared tests for Hardy-Weinberg equilibrium

Locus	SNP	Population	
		Patients *P*	**Controls** *P*
MUC5B	rs35705950	0.139	0.657
TOLLIP	rs3750920	0.311	0.844
	rs111521887	0.249	1.000
	rs5743894	0.389	1.000
	rs5743890	0.354	1.000
	rs5743854	0.342	0.695

Table S3. Haplotype frequencies of the M950-T920 -T887-T894 MUC5B-TOLLIP haplotype block

Haplotype	Overall Freq.	Patients Freq.	Control Freq.	Chi Square	*P*
GCCT	0.451	0.361	0.529	11.821	**0.0006**
GTCT	0.214	0.171	0.251	3.934	**0.0473**
TTCT	0.104	0.162	0.055	12.733	**0.0004**
TTGC	0.081	0.143	0.028	18.288	**<0.0001**
GTGC	0.055	0.045	0.064	0.686	0.4075
TCCT	0.051	0.074	0.031	3.893	**0.0485**
GCGC*	0.024	0.015	0.031	1.165	0.2804
TCGC*	0.015	0.024	0.007	2.177	0.1401
GTGT*	0.002	0	0.004	0.864	0.3525
TTCC*	0.002	0.005	0	1.158	0.282

*Rare haplotypes (≤ 7 occurrences)

Table S4. Measures of pairwise linkage disequilibrium between MUC5B rs35705950 and TOLLIP rs3750920, rs111521887 and rs5743894 SNPs in patients and controls

SNP A	SNP B	D'	CI 95%	R^2	Dist bp	χ^2	*P*
fHP patients							
rs35705950	**rs3750920**	0.503	0.334 - 0.672	0.157	68735	29.197	<0.0001
rs35705950	**rs111521887**	0.573	0.372 - 0.774	0.140	71485	27.700	<0.0001
rs35705950	**rs5743894**	0.583	0.386 - 0.780	0.150	83551	29.452	<0.0001
rs3750920	**rs111521887**	0.631	0.409 - 0.854	0.106	2750	19.515	<0.0001
rs3750920	**rs5743894**	0.636	0.417 - 0.855	0.110	14816	20.651	<0.0001
rs111521887	**rs5743894**	1	1.000 - 1.000	0.971	12066	188.424	<0.0001
Controls							
rs35705950	**rs3750920**	0.428	0.143 - 0.712	0.037	68735	8.962	0.0028
rs35705950	**rs111521887**	0.131	-0.042 - 0.304	0.015	71485	4.158	0.0415
rs35705950	**rs5743894**	0.141	-0.032 - 0.314	0.018	83551	4.589	0.0322
rs3750920	**rs111521887**	0.471	0.207 - 0.735	0.051	2750	12.817	0.0003
rs3750920	**rs5743894**	0.457	0.186 - 0.727	0.046	14816	11.486	0.0007
rs111521887	**rs5743894**	1	1.000 - 1.000	0.962	12066	215.423	<0.0001

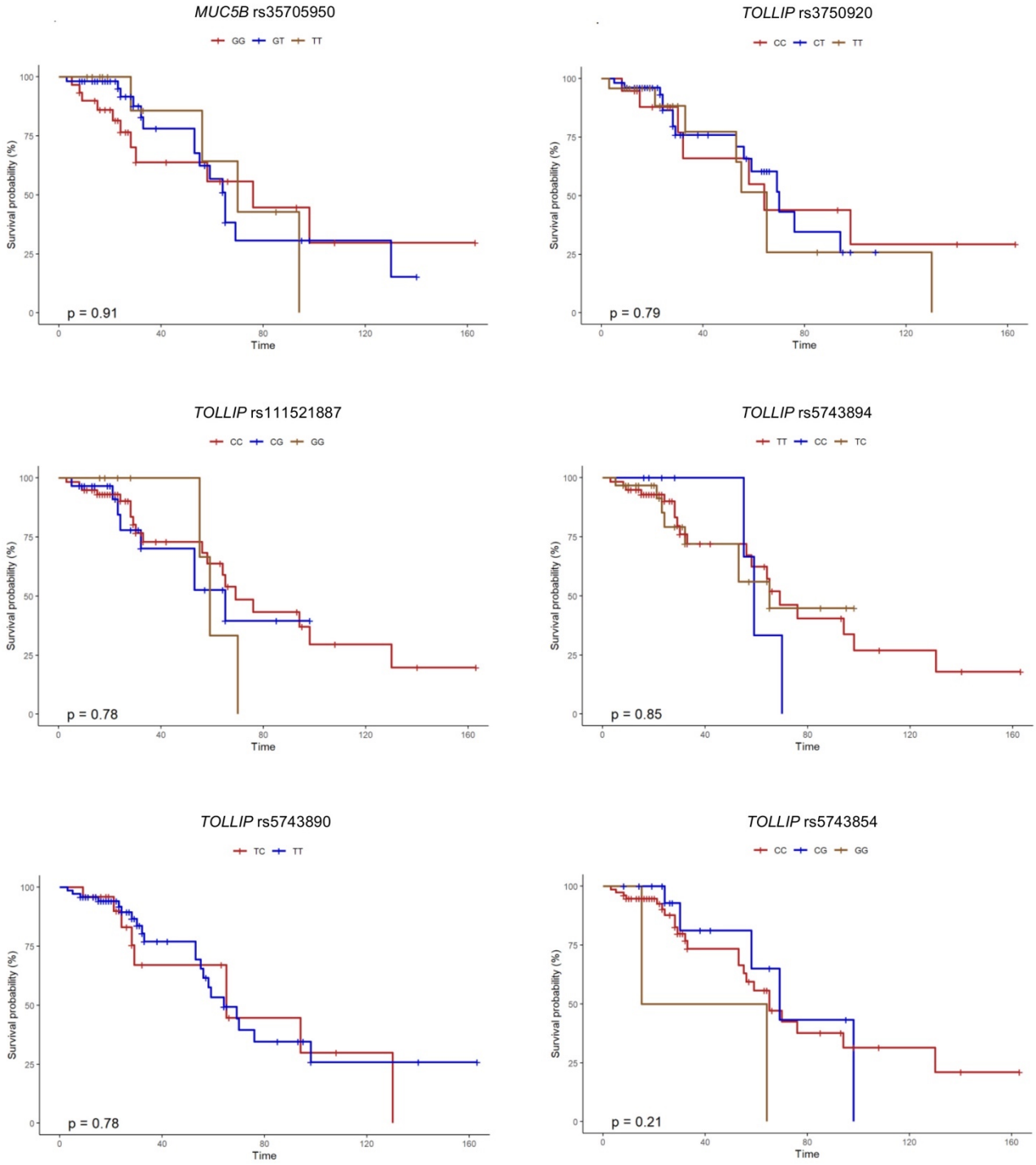
D', D prime between two SNPs; R^2, correlation coefficient between two SNPs; Dist bp, distance between SNPs in base pairs

Table S5. TOLLIP rs5743890 and rs5743854 associations with demographic and clinical features

Genotype	TOLLIP rs5743890			TOLLIP rs5743854		
	TT (n=72)	TC (n=25)	*P* value	CC (n=77)	CG/GG (n=20)	*P* value
Gender (female/male)	36 (50%) 36 (50%)	8 (32%) 17 (68%)	0.119	38 (49.4%) 39 (50.6%)	6 (30%) 14 (70%)	0.121
Age (years)	68.4 (8.3)	64.9 (9.4)	0.083	67.8 (9.0)	66.2 (7.6)	0.447
Ever smoker	33 (45.8%)	14 (56%)	0.381	35 (45.5%)	12 (60.0%)	0.246
Death	19 (26.4%)	7 (28.0%)	1.000	20 (26%)	6 (30%)	0.779
Pulmonary function tests						
FVC (% pred)	81.2 (17.8)	79.7 (15.8)	0.710	80.0 (17.2)	84.0 (17.4)	0.355
TLC (% pred)	77.6 (17.0)	74.8 (19.9)	0.510	76.6 (18.5)	78.0 (14.6)	0.758
DLCO (% pred)	53.4 (17.5)	55.3 (15.3)	0.641	54.8 (17.5)	50.6 (14.1)	0.349
Antigen exposure						
Avian	51 (70.8%)	19 (76.0%)	0.619	57 (74.0%)	13 (65.0%)	0.422
Molds	30 (41.7%)	7 (28.0%)	0.226	28 (36.4%)	9 (45.0%)	0.479
Other organics	14 (19.4%)	4 (16.0%)	0.703	12 (15.6%)	6 (30.0%)	0.140
Inorganics	16 (22.2%)	5 (20.0%)	0.816	16 (20.8%)	5 (25.0%)	0.683
BAL features						
Lymphocytes (%)	11.0 (6.0-18.9)	13.4 (10.3-23.2)	0.093	12.0 (6.6-20.1)	11.7 (6.5-15.2)	0.651
Neutrophils (%)	5.9 (3.2-11.3)	4.7 (2.8-9.3)	0.272	5.2 (3.1-10.1)	5.8 (3.3-11.8)	0.704
Eosinophils (%)	3.2 (1.5-7.8)	2.2 (0.7-5.0)	0.255	2.6 (1.4-5.8)	3.5 (0.7-13.0)	0.494
Radiographic features						
UIP + probable UIP	19 (26.8%)	9 (39.1%)	0.355	22 (29.3%)	6 (31.6%)	0.166
Indeterminate for UIP	18 (25.4%)	3 (13.0%)		14 (18.7%)	7 (36.8%)	
Alternative diagnosis	34 (47.8%)	11 (47.8%)		39 (52.0%)	6 (31.6%)	
Honeycombing present	27 (38.0%)	10 (43.5%)	0.642	30 (40.0%)	7 (36.8%)	0.801
Fibrosis score	9.0 (6.0-11.5)	8.0 (5.0-10.0)	0.470	9.0 (6.0-11.5)	7.0 (5.8-9.0)	0.213
Histopathological features						
Honeycombing present	15 (36.6%)	5 (38.5%)	0.903	16 (44.4%)	4 (22.2%)	0.111
Area of fibrosis (%)	31.0 (14.0)	33.1 (14.5)	0.633	33.0 (14.8)	28.4 (12.2)	0.260
Fibroblast foci						
0	14 (34.1%)	8 (61.5%)	0.200	14 (38.9%)	8 (44.4%)	0.721
1	8 (19.5%)	2 (15.4%)		6 (16.7%)	4 (22.2%)	
>1	19 (46.3%)	3 (23.1%)		16 (44.4%)	6 (33.3%)	
Fibrosis distribution						
Peribronchiolar	9 (22.5%)	4 (30.8%)	0.613	8 (22.9%)	5 (27.8%)	0.740
Peribronchiolar (+paraseptal or subpleural)	23 (57.5%)	5 (38.5%)		18 (51.4%)	10 (55.6%)	
Heterogeneous	7 (17.5%)	3 (23.1%)		7 (20.0%)	3 (16.7%)	
Undetermined	1 (2.5%)	1 (7.7%)		2 (5.7%)	0 (0%)	
Inflammatory infiltrate						
Light	27 (65.9%)	11 (84.6%)	0.197	26 (72.2%)	12 (66.7%)	0.673
Moderate-severe	14 (34.1%)	2 (15.4%)		10 (27.8%)	6 (33.3%)	
Peribronchiolar metaplasia	1.0 (1.0-3.0)	2.0 (1.0-3.0)	0.548	1.5 (1.0-2.3)	1.5 (1.0-3.0)	0.866
Giant cells granulomas	13 (31.7%)	3 (23.1%)	0.553	10 (27.8%)	6 (33.3%)	0.673

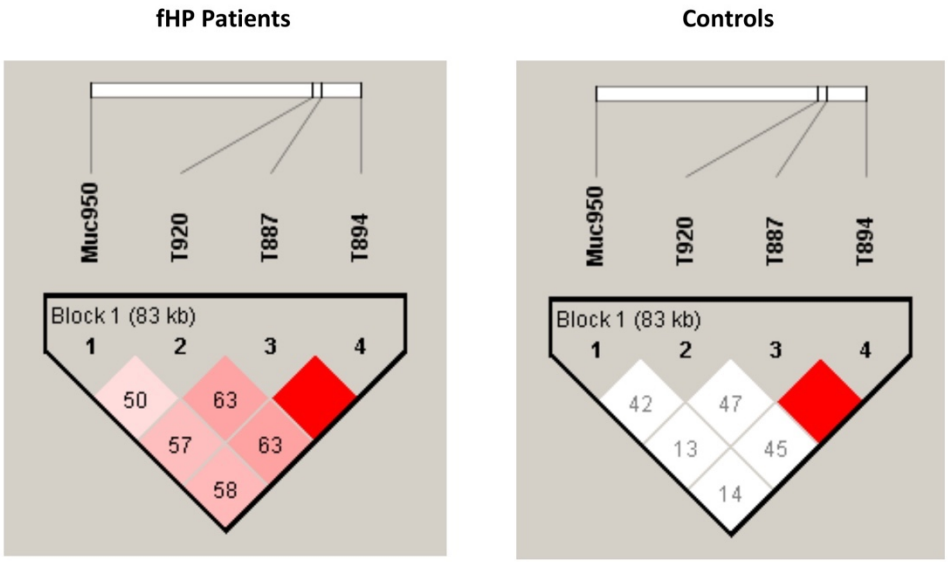
Values are presented as median (Q1-Q3) or mean (SD) and counts as n (%).
FVC, forced vital capacity; TLC, total lung capacity; DLCO, diffusing capacity of the lung for carbon monoxide.

Fig. S1 Kaplan–Meier plots showing fHP patients' transplant free survival according to MUC5B rs35705950 and TOLLIP rs3750920, rs111521887, rs5743894, rs5743890 and rs5743854 genotypes



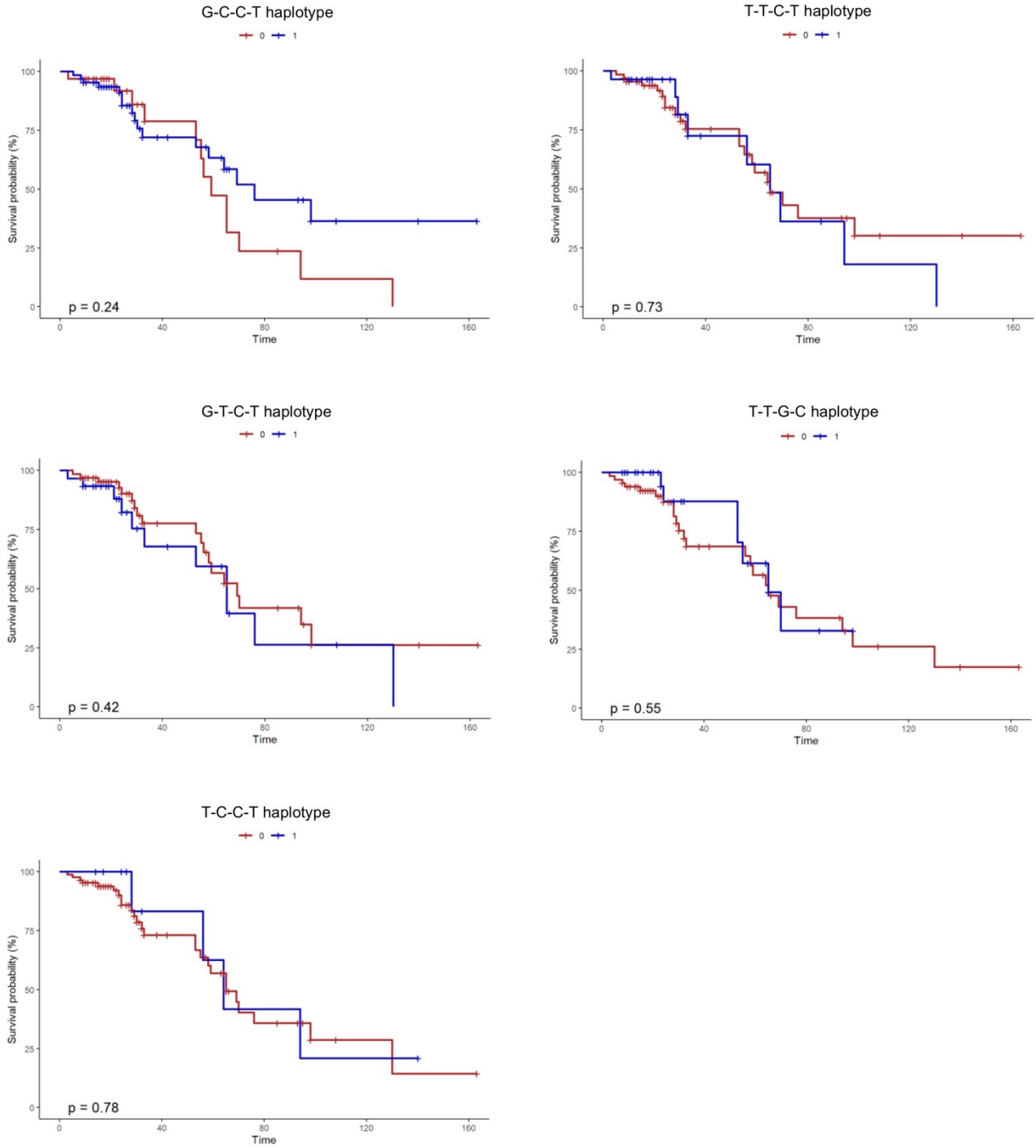
Time in months; P-values for log-rank test

Fig. S2 Linkage disequilibrium (LD) plot of the 83-kb M950-T920-T887-T894 block on chromosome 11p15.5



The plot was drawn using Haploview software v4.2.
D' values are shown inside the squares. Empty squares have a D' =1 (100)
Red boxes represent D' =1 (LOD ≥2). Shades of pink/red represent D' <1 and LOD ≥2.
White boxes represent D' <1 and LOD <2.
LOD, log of likelihood odds ratio.

Fig. S3 Kaplan-Meier plots showing fHP patients' survival according to haplotype defined by the rs35705950-rs3750920-rs111521887-rs5743894 MUC5B-TOLLIP block



Time in months; P value for log-rank test

Supplementary references

1. Raghu G, Remy-Jardin M, Richeldi L, et al. Idiopathic Pulmonary Fibrosis (an Update) and Progressive Pulmonary Fibrosis in Adults: An Official ATS/ERS/JRS/ALAT Clinical Practice Guideline. *Am J Respir Crit Care Med* 2022; **205**(9): e18-e47.
2. Hansell DM, Bankier AA, MacMahon H, McLoud TC, Muller NL, Remy J. Fleischner Society: glossary of terms for thoracic imaging. *Radiology* 2008; **246**(3): 697-722.
3. Raghu G, Remy-Jardin M, Ryerson CJ, et al. Diagnosis of Hypersensitivity Pneumonitis in Adults. An Official ATS/JRS/ALAT Clinical Practice Guideline. *Am J Respir Crit Care Med* 2020; **202**(3): e36-e69.
4. Bankhead P, Loughrey MB, Fernandez JA, et al. QuPath: Open source software for digital pathology image analysis. *Sci Rep* 2017; **7**(1): 16878.
5. Wright JL, Churg A, Hague CJ, Wong A, Ryerson CJ. Pathologic separation of idiopathic pulmonary fibrosis from fibrotic hypersensitivity pneumonitis. *Mod Pathol* 2020; **33**(4): 616-25.

THIRD STUDY

MUC5B AND *TOLLIP* POLYMORPHISMS IN THE DIFFERENTIAL DIAGNOSIS BETWEEN IDIOPATHIC PULMONARY FIBROSIS AND FIBROTIC HYPERSENSITIVITY PNEUMONITIS: A CASE CONTROL STUDY

Mota PC, Soares ML, Ferreira AC, Santos RF, Dias CC, Rufo JC, Vasconcelos D, Carvalho A, Guimarães S, Vasques-Nóvoa F, Cardoso C, Melo N, Alexandre AT, Coelho D, Sokhatska O, Beltrão M, Novais-Bastos H, Morais A

Submitted to Respiratory Research

Special collection: "Pulmonary fibrosis: new insights"

MUC5B and *TOLLIP* polymorphisms in the differential diagnosis between idiopathic pulmonary fibrosis and fibrotic hypersensitivity pneumonitis: a case control study

Patrícia Caetano Mota[1,2], Miguel Luz Soares[3,4], António Carlos Ferreira[3,4], Rita F. Santos[4,5], Claudia Camila Dias[6,7], João Cavaleiro Rufo[8,9], Daniel Vasconcelos[3,4], André Carvalho[2,10], Susana Guimarães[2,11], Francisco Vasques-Nóvoa[12,13], Catarina Cardoso[1,2], Natália Melo[1], André Terras Alexandre[1], David Coelho[1,2], Oksana Sokhatska[14], Marília Beltrão[14], Hélder Novais-Bastos[1,2], António Morais[1,2]

[1]Departamento de Pneumologia, Unidade Local de Saúde de São João, Porto, Portugal

[2]Faculdade de Medicina da Universidade do Porto (FMUP), Instituto de Investigação e Inovação em Saúde, Universidade do Porto, Porto, Portugal

[3]Laboratório de Apoio à Investigação em Medicina Molecular (LAIMM), Departamento de Biomedicina, FMUP, Porto, Portugal

[4]Instituto de Investigação e Inovação em Saúde, Universidade do Porto, Porto, Portugal

[5]Escola Superior de Saúde - Instituto Politécnico do Porto, Porto, Portugal

[6]Unidade de Gestão de Conhecimento, FMUP, Porto, Portugal

[7]Centro de Investigação em Tecnologias e Serviços de Saúde (CINTESIS@RISE), Departamento de Medicina da Comunidade, Informação e Decisão em Saúde (MEDCIDS), FMUP, Porto, Portugal

[8]EPIUnit - Instituto de Saúde Pública, Universidade do Porto, Porto, Portugal

[9]Center for Translational Health and Medical Biotechnology Research (T.Bio), Escola Superior de Saúde, Instituto Politécnico do Porto, Porto, Portugal

[10]Departamento de Radiologia, Unidade Local de Saúde de São João, Porto, Portugal

[11]Departamento de Anatomia Patológica, Unidade Local de Saúde de São João, Porto, Portugal

[12]Departamento de Medicina Interna, Unidade Local de Saúde de São João, Porto, Portugal

[13]UnIC@RISE, Department of Surgery and Physiology, FMUP, Porto, Portugal

[14]Serviço de Imunologia Básica e Clínica, FMUP, Portugal

Corresponding author:

Miguel Luz Soares

Unidade de Biologia Experimental, Departamento de Biomedicina

Faculdade de Medicina da Universidade do Porto.

Rua Dr. Plácido Costa, s/n - Piso 4, 4200-450 Porto, Portugal

mlsoares@med.up.pt

Abstract

Background: Idiopathic pulmonary fibrosis (IPF) and fibrotic hypersensitivity pneumonitis (fHP) are interstitial lung disorders with different pathogenesis and risk factors but sharing clinical, radiographic, and histopathologic characteristics. Genetic features have been increasingly recognized as playing a role in the pathophysiology of both conditions, namely single nucleotide polymorphisms (SNPs) of the *MUC5B* gene promoter and *TOLLIP*. This study aimed at analyzing *MUC5B* and *TOLLIP* SNPs in order to identify signatures that may help to genetically differentiate the two fibrotic phenotypes, with focus on disease risk and survival.

Methods: A case-control study involving 86 IPF patients, 97 fHP patients, and 112 healthy controls was conducted. The *MUC5B* promoter and 5 *TOLLIP* SNPs, as well as haplotypes, were analyzed. Associations with risk and survival were investigated through comparisons between both conditions and between each condition and controls.

Results: Analyses of each SNP individually showed that the *MUC5B* rs35705950 was associated with susceptibility to both IPF and fHP, as previously reported. The minor allele of rs35705950 was significantly more frequent among IPF (42.4% vs 12.1%, $P<0.001$) and fHP patients (40.7% vs 12.1%, $P<0.001$) compared to controls. *TOLLIP* SNPs rs3750920, rs111521887, and rs5743894, on the other hand, were associated with susceptibility to fHP only, with a significantly higher frequency in fHP patients than in controls (52.6% vs 40.2%, $P=0.011$; 22.7% vs 13.4%, $P=0.013$; and 23.2% vs 12.9%, $P=0.006$; respectively). Direct comparisons between IPF and fHP patients did not reveal significant differences in allele and genotype distributions for any SNP. Haplotypes defined by the rs35705950-rs3750920-rs111521887-rs5743894 block showed associations with susceptibility to disease when comparing all patients to controls. Importantly, a haplotype was identified that conferred

significantly better survival to fHP carriers than to IPF carriers (median survival: IPF=59 months vs fHP =98 months; log-rank test *P*=0.006).

Conclusions: While revealing genetic similarities between IPF and fHP, our study also unveiled distinctive genetic features with impact on risk and survival, indicating that *TOLLIP* and *MUC5B* variants and haplotypes may serve as valuable tools for risk assessment and prognosis and for patient stratification in both diseases.

Keywords: Idiopathic pulmonary fibrosis; Hypersensitivity pneumonitis; *MUC5B*; *TOLLIP*; genetic variants; haplotypes; survival.

Background

Fibrosing interstitial lung diseases (fILDs) encompass both idiopathic pulmonary fibrosis (IPF), the prototypical condition in this group, and non-IPF disorders like fibrotic hypersensitivity pneumonitis (fHP). These disorders share overlapping clinical, radiographic, and histopathologic features despite differences in risk factors and pathogenesis [1-3].

IPF is a progressive and ultimately fatal disease characterized by the presence of usual interstitial pneumonia (UIP) on histopathologic and/or radiologic examination [1]. Hallmark features include epithelial cell injury and activation, leading to a fibroproliferative response and subsequent scarring, with minimal to moderate signs of inflammation. This pathology primarily reflects an epithelial fibroblastic disorder [4]. Fibrotic HP arises from significant antigen exposure, leading to persistent lung injury and subsequent inflammatory responses. This inflammation can progress to a late fibrotic phase, akin to IPF, notably in those patients with a UIP-like pattern, a subgroup with worst prognosis [2, 5, 6]. Indeed, evidence suggests the existence of at least two distinct pathways leading to lung fibrosis, in some instances ultimately resulting in a very similar fibrotic phenotype [2]. As a result, distinguishing between these conditions poses a significant challenge, but this differentiation is crucial given the distinct therapeutic requirements at baseline and prognostic implications of each disorder [1-3, 5].

Genetic factors play a role in both diseases, with mucin 5B (*MUC5B*) and toll-interacting protein (*TOLLIP*) single nucleotide polymorphisms (SNPs) having been implicated in susceptibility and survival. Associations of variants of these genes with specific outcomes highlight the contribution of genetic determinants to the development and evolution of the fibrosing phenotype [7-17]. While previous studies have investigated the role of these SNPs in each disease independently, a direct comparison between IPF and fHP in the context of these genetic features has not been conducted.

This study aimed at analyzing the behavior of *MUC5B* and *TOLLIP* SNPs and their haplotypes in a direct comparison between IPF and fHP patients of the same population (i.e., with the same genetic background), in order to identify differentiating elements as well as common features between these fibrotic diseases, with focus on disease risk and survival.

Methods

Participants and study design

A case-control study was undertaken, involving 86 IPF and 97 fHP patients. These patients were followed from 2013 to 2022 at a tertiary hospital in Porto, Portugal. The study cohort comprised individuals enrolled in the FIBRALUNG project, which is the nation's inaugural registry and biobank for ILDs [18]. For the control group, 112 subjects were selected from the EPIPorto cohort [19], a population-based study of healthy adults (n=74), and the NETDiamond cohort [20, 21], a prospective study focused on therapeutic targets in heart failure, excluding those with respiratory conditions (n=38). Diagnoses of IPF and fHP were established through multidisciplinary team assessments in accordance with international guidelines [1-3, 22-25]. Patient data, including demographic information, clinical details, lung function tests, bronchoalveolar lavage (BAL) results, and radiographic and histopathological features, were collected at enrollment. The study protocol received approval from the Ethics Committee of the institution and was conducted in compliance with the principles outlined in the Helsinki Declaration. Written informed consent was obtained from all participants.

Genotyping

Individuals underwent genotyping for six SNPs: *MUC5B* rs35705950 and *TOLLIP* rs3750920, rs111521887, rs5743894, rs5743890, and rs5743854 (5' to 3', on Chr.11) using a combination

of restriction fragment length polymorphism (RFLP) analysis and real-time polymerase chain reaction (PCR). Further details are provided in Supplementary Methods M1 and Table S1.

Radiology

High-resolution computed tomography (HRCT) scans were independently reviewed by a radiologists group experienced in interpreting CT scans for interstitial lung diseases (ILDs), who were blinded to the clinical diagnosis. Further details regarding image analysis and classification scores are provided in Supplementary Methods M2.

Histopathology

Lung biopsies, including transbronchial cryobiopsies (TBLC), were reviewed independently by expert lung pathologists using a guided pathology data collection form, who were blinded to the clinical diagnosis. Further details regarding histopathological analyses and classification scores can be found in Supplementary Methods M3.

Statistical analysis

Continuous data are presented as means ± standard deviation or medians and interquartile range. Categorical data are presented as frequencies and percentages. Chi-squared tests were used to compare allele and genotype frequencies between patients and controls. To test hypotheses about the equality of group means, the T-test for independent groups and one-factor analysis of variance (ANOVA) were used for variables with a normal distribution. Hypotheses about the distribution of continuous variables with a non-normal distribution were tested using the Mann-Whitney and Kruskal-Wallis non-parametric tests, depending on the nature of the hypotheses being tested. Logistic regression was applied to determine the association of *MUC5B* and *TOLLIP* genotypes distributed according to minor allele (dominant model) with

IPF and fHP, in comparison with controls and each other, with results being expressed as odds ratio (OR) and respective 95% confidence interval (CI). Overall survival was analyzed with time from diagnosis to death (in months) censured by end of follow-up. Kaplan-Meier plots and log-rank tests were used for analysis. Univariate and multivariate Cox proportional hazards regressions were employed to identify variables that predict survival status, with results being expressed as hazard ratio (HR) and respective 95% CI. Statistical analyses were performed in R (v4.2.0) and the software Statistical Package for the Social Sciences (SPSS) v. 29.0. Haplotype frequencies were estimated using Haploview v4.2. [26-28]. Pairwise linkage disequilibrium (LD) between SNPs was calculated using SNPAlyze v8.1.1.0 (Dynacom Co., Japan). Two-sided P-values lower than 0.05 were considered statistically significant.

Results

Demographic and clinical features of the study subjects

The characteristics of the study participants at baseline are shown in Tables 1 and 2. Notably, IPF patients were predominantly male, older, and exhibited a higher percentage of smokers than patients diagnosed with fHP. Key pulmonary function measurements showed a significantly lower diffusion capacity of the lung for carbon monoxide (DLCO) in IPF patients, whereas forced vital capacity (FVC) and total lung capacity (TLC) appeared similar between both diseases (Table 1).

Table 1. Clinical and demographic characteristics of study participants

	IPF (n=86)	fHP (n=97)	Controls (n=112)	p value
Gender (n, %)				0.049
Female	24 (27.9%)	44 (45.4%)	44 (39.3%)	
Male	62 (72.1%)	53 (54.6%)	68 (60.7%)	
Age (years)[a]	70.8 (±8.5)	67.5 (±8.7)	71.2 (±8.4)	0.004
Ever smoker (n, %)	55 (64.0%)	47 (48.5%)	51 (45.9%)	0.030
Follow up time (months)[b]	43 (19-69)	24 (16-55)		0.022
Deaths (n, %)	56 (65.1%)	26 (26.8%)		< 0.001
Pulmonary function tests				
FVC (% pred)[a]	84.7 (±17.5)	80.8 (±17.3)		0.071
TLC (% pred)[a]	78.2 (±15.5)	76.9 (±17.7)		0.618
DLCO (% pred)[a]	47.4 (±17.5)	53.9 (±16.8)		0.018
Treatment (n, %)				< 0.001
N-acetylcysteine	5 (5.8%)	0 (0.0%)		
Immunosuppressants	20 (23.3%)	53 (54.6%)		
Antifibrotics	52 (60.5%)	25 (25.8%)		
No treatment	9 (10.5%)	19 (19.6%)		
Comorbidities (n, %)				
COPD	2 (2.3%)	3 (3.1%)		> 0.999
Pulmonary hypertension	22 (25.6%)	7 (7.2%)		< 0.001
Lung cancer	8 (9.3%)	13 (13.4%)		0.385
GERD	12 (14.0%)	21 (21.6%)		0.177
Metabolic conditions	53 (61.6%)	29 (29.9%)		< 0.001
Cardiovascular disease	48 (55.8%)	55 (56.7%)		0.904
Depression	13 (15.1%)	20 (20.6%)		0.334

Values are presented as [a]mean (±standard deviation), [b]median (Q1-Q3) and number (%). IPF, idiopathic pulmonary fibrosis; fHP, fibrotic hypersensitivity pneumonitis; COPD, chronic obstructive lung disease; OSA, obstructive sleep apnea; GERD, gastro-esophageal reflux disease; % pred, percentage of predicted normal value; FVC, forced vital capacity; TLC, total lung capacity; DLCO, diffusing capacity of the lung for carbon monoxide

As expected, IPF patients exhibited a higher occurrence of radiologic UIP pattern and honeycombing. Histological confirmation was necessary in more fHP than IPF cases, with the former displaying fewer instances of honeycombing and fibroblast foci but showing a more predominant peribronchiolar fibrosis and granulomas, which were absent in the latter (Table 2).

Table 2. BAL, radiological and histopathological features of IPF and fHP patients

	IPF (n=86)	fHP (n=97)	p value
BAL features			
Total cells ($\times 10^4 mL^{-1}$)[b]	1.4 (1.0-2.4)	1.6 (0.7-2.7)	0.788
Macrophages[b]	73.0 (66.0-77.8)	71.9 (58.7-81.3)	0.811
Lymphocytes[b]	10.9 (7.0-18.0)	12.0 (6.6-20.0)	0.773
Neutrophils[b]	7.1 (3.0-12.2)	5.4 (3.2-11.2)	0.581
Eosinophils[b]	4.4 (1.8-8.8)	2.8 (1.2-6.6)	0.238
Radiological features			
UIP + probable UIP	63 (75.9%)	28 (29.8%)	<0.001
Indeterminate for UIP	15 (18.1%)	21 (22.3%)	
Alternative diagnosis	5 (6.0%)	45 (47.9%)	
Honeycombing	53 (63.9%)	37 (39.4%)	0.001
Fibrosis score[a]	8 (±3.0)	9 (±3.0)	0.352
Lung biopsy (n, %)	15 (17.4%)	54 (55.7%)	<0.001
Histopathological features			
Honeycombing	11 (73.3%)	20 (37.0%)	0.012
Area of fibrosis[b]	32.5 (28.3-61.5)	30.9 (21.4-40.0)	0.167
Fibroblast foci			0.007
0	1 (6.7%)	22 (40.7%)	
1	1 (6.7%)	19 (18.5%)	
>1	13 (86.7%)	22 (40.7%)	
Fibrosis distribution			<0.001
Subpleural and paraseptal	11 (78.6%)	0 (0.0%)	
Peribronchiolar	0 (0.0%)	13 (24.5%)	
Peribronchiolar (+paraseptal/ subpleural)	2 (14.3%)	28 (52.8%)	
Heterogeneous	0 (0.0%)	10 (18.9%)	
Undetermined	1 (7.1%)	2 (3.8%)	
Inflammatory infiltrate			0.534
Light	12 (80.0%)	38 (70.4%)	
Moderate-severe	3 (20.0%)	16 (29.6%)	
Peribronchiolar metaplasia[b]	2 (0-5)	2 (1-3)	0.678
Granulomas / giant cells	0 (0.0%)	16 (29.6%)	0.015

214 Values are presented as [a]mean (±standard deviation), [b]median (Q1-Q3) and number (%).
215 IPF, idiopathic pulmonary fibrosis; fHP, fibrotic hypersensitivity pneumonitis; BAL, bronchoalveolar
216 lavage; UIP, usual interstitial pneumonia
217

218 Treatment strategies obviously differed between the two conditions: the majority of IPF
219 patients primarily received antifibrotic therapy, while most fHP patients were initially treated
220 with immunosuppressants, later supplemented with antifibrotics in those cases revealing
221 progressive fibrosis (Table 1).

222 The median follow-up duration from diagnosis to death, transplant or the end of the study was
223 notably shorter in fHP cases (24 months) compared to IPF cases (43 months) (P=0.022).
224 However, the mortality rate was markedly higher in IPF patients (65.1%) than in fHP patients
225 (26.8%) (P<0.001).
226 Regarding comorbidities, pulmonary hypertension and metabolic disorders were significantly
227 more prevalent in the IPF group (P<0.001).
228
229 ***MUC5B and TOLLIP associations with IPF and fHP susceptibility and survival***
230 Genotype and allele frequencies of the probed SNPs in IPF and fHP patients, and in control
231 individuals are shown in Table 3. All distributions conformed to the Hardy-Weinberg
232 equilibrium (Supplementary Table S2).
233 Independent comparisons of each disorder versus controls revealed that while the frequency of
234 the *MUC5B* rs35705950 minor allele was significantly higher in patients than in controls
235 (42.4% vs 12.1%, and 40.7% vs 12.1%, P=<0.001 for IPF and fHP, respectively), the minor
236 allele frequencies of *TOLLIP* rs3750920, rs111521887, and rs5743894 SNPs were significantly
237 increased in fHP when compared to controls (52.6% vs 40.2%, P=0.011; 22.7% vs 13.4%,
238 P=0.013; and 23.2% vs 12.9%, P=0.006; respectively), but not in IPF patients (Table 3).

Table 3. Genotype and allele distributions of *MUC5B* and *TOLLIP* SNPs in patients and controls

		IPF n (%)	fHP n (%)	CT n (%)	p-value 3 groups	p-value IPF vs fHP	p-value IPF vs CT	p-value fHP vs CT
MUC5B	GG	24 (27.9%)	30 (30.9%)	87 (77.7%)	<0.001	0.905	<0.001	<0.001
rs35705950	GT	51 (59.3%)	55 (56.7%)	23 (20.5%)				
	TT	11 (12.8%)	12 (12.4%)	2 (1.8%)				
	G	99 (57.6%)	115 (59.3%)	197 (87.9%)	<0.001	0.739	<0.001	<0.001
	T	73 (42.4%)	79 (40.7%)	27 (12.1%)				
TOLLIP	CC	20 (23.3%)	19 (19.6%)	39 (34.8%)	0.090	0.831	0.136	0.029
rs3750920	CT	46 (53.5%)	54 (55.7%)	56 (50.0%)				
	TT	20 (23.3%)	24 (24.7%)	17 (15.2%)				
	C	86 (50.0%)	92 (47.4%)	134 (59.8%)	0.027	0.622	0.051	0.011
	T	86 (50.0%)	102 (52.6%)	90 (40.2%)				
TOLLIP	CC	56 (65.1%)	60 (61.9%)	84 (75.0%)	0.143	0.538	0.294	0.049
rs111521887	CG	27 (31.4%)	30 (30.9%)	26 (23.2%)				
	GG	3 (3.5%)	7 (7.2%)	2 (1.8%)				
	C	139 (80.8%)	150 (77.3%)	194 (86.6%)	0.045	0.413	0.118	0.013
	G	33 (19.2%)	44 (22.7%)	30 (13.4%)				
TOLLIP	TT	57 (66.3%)	59 (60.8%)	85 (75.9%)	0.068	0.296	0.329	0.029
rs5743894	TC	27 (31.4%)	31 (32.0%)	25 (22.3%)				
	CC	2 (2.3%)	7 (7.2%)	2 (1.8%)				
	T	141 (82.0%)	149 (76.8%)	195 (87.1%)	0.024	0.223	0.163	0.006
	C	31 (18.0%)	45 (23.2%)	29 (12.9%)				
TOLLIP	CC	1 (1.2%)	0 (0.0%)	3 (2.7%)	0.490	0.560	0.581	0.227
rs5743890	TC	21 (24.4%)	25 (25.8%)	32 (28.6%)				
	TT	64 (74.4%)	72 (74.2%)	77 (68.8%)				
	C	23 (13.4%)	25 (12.9%)	38 (17.0%)	0.434	0.891	0.336	0.245
	T	149 (86.6%)	169 (87.1%)	186 (83.0%)				
TOLLIP	CC	68 (79.1%)	77 (79.4%)	83 (74.1%)	0.901	0.992	0.716	0.668
rs5743854	CG	16 (18.6%)	18 (18.6%)	26 (23.2%)				
	GG	2 (2.3%)	2 (2.1%)	3 (2.7%)				
	C	152 (88.4%)	172 (88.7%)	192 (85.7%)	0.604	0.931	0.438	0.371
	G	20 (11.6%)	22 (11.3%)	32 (14.3%)				

239 Values are expressed in number (%).
240 IPF, idiopathic pulmonary fibrosis; fHP, fibrotic hypersensitivity pneumonitis; CT, controls
241

Accordingly, the *MUC5B* rs35705950 genotype distributions were significantly different in both conditions, when independently compared to controls (dominant model: IPF, 72.1% vs 22.3%, OR=8.99 95% CI 4.70-17.19, *P*<0.001; and fHP, 69.1% vs 22.3%, OR=7.77 95% CI 4.19-14.43, *P*<0.001) (Fig.1A).

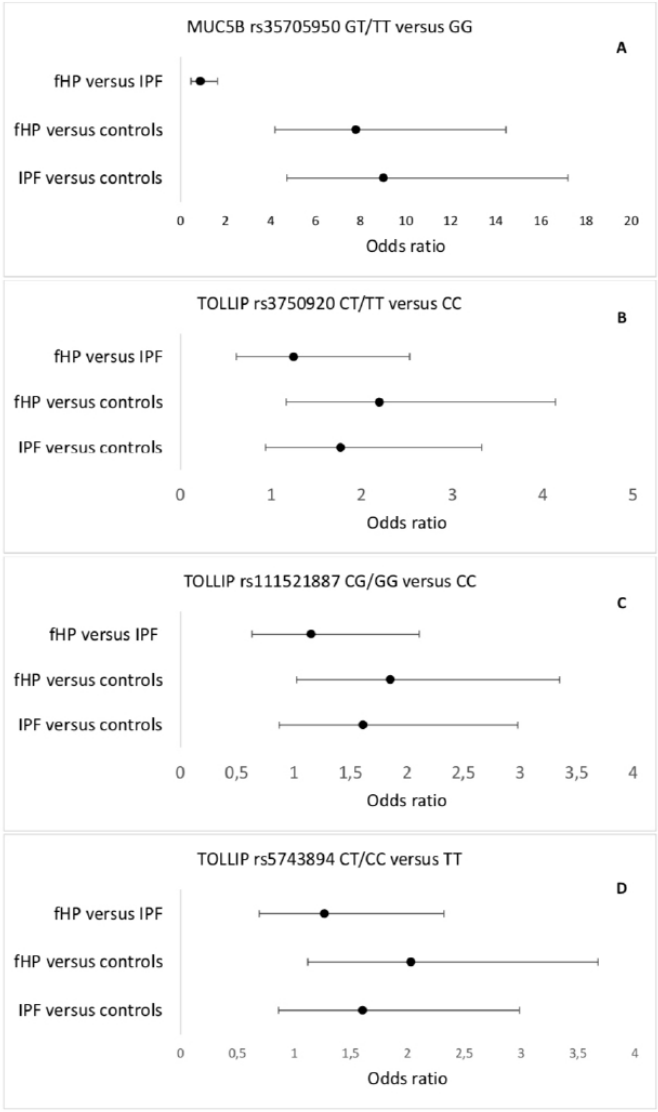


Figure 1. Associations of *MUC5B* and *TOLLIP* genotypes with risk among IPF, fHP and controls.
A) *MUC5B* rs35705950 GT/TT vs GG: IPF vs controls 72.1% vs 22.3%, P<0.001; fHP vs controls: 69.1% vs 22.3%, *P*<0.001; fHP vs IPF P=0.865.

B) *TOLLIP* rs3750920 CT/TT vs CC: IPF vs controls 76.7% vs 65.2%, P=0.079; fHP vs controls 80.4% vs 65.2%, P=0.015; fHP vs IPF, P=0.546.
C) *TOLLIP* rs111521887 CG/GG vs CC: IPF vs controls 30% vs 25%, P=0.131; fHP vs controls 37% vs 25%, P=0.042; fHP vs IPF P=0.648.
D) *TOLLIP* rs5743894 CT/CC vs TT: IPF vs controls 29% vs 24.1%, P=0.138; fHP vs controls: 39.2% vs 24.1%, P= 0.020; fHP vs IPF, P=0.445.
The forest plots depict odds ratios and their associated 95% confidence intervals.
IPF, idiopathic pulmonary fibrosis; fHP, fibrotic hypersensitivity pneumonitis

The genotypes of the *TOLLIP* SNPs, however, showed significant associations with fHP only (*P*=0.015, *P*=0.042, and *P*=0.020, respectively) (Fig.1B-D). As with the alleles, no differences in genotype frequencies were seen in IPF patients versus controls for these SNPs.

Of note, the direct comparison between IPF and fHP patients did not show significant differences in allele or genotype distributions of either *MUC5B* or any of the *TOLLIP* SNPs. These results support the notion that *MUC5B* rs35705950 is as a common genetic risk factor for both diseases, and that the *TOLLIP* rs3750920, rs111521887, and rs5743894 polymorphisms are genetic risk factors for fHP alone.

The lack of differentiation between IPF and fHP in a direct comparison of allelic and genotypic distributions of the probed SNPs individually was further seen when the same comparison was conducted in the context of the patients' radiological UIP pattern or presence of honeycombing on HRCT (Supplementary Table S3). Additionally, no differences were found between IPF patients with a UIP pattern and fHP patients without a UIP pattern (not shown). Kaplan-Meier survival analyses of the *MUC5B* and *TOLLIP* SNP genotypes revealed statistically significant associations for *TOLLIP* rs111521887 and rs5743894 (log rank test *P*=0.007 and *P*=0.002, respectively) in IPF patients. Patients with genotypes G/G and C/C, respectively, had worst survival than those without these genotypes (Supplementary Fig. S1). No associations were seen between individual genotypes and survival in fHP patients (not shown).

283 ***MUC5B and TOLLIP haplotype associations with IPF and fHP susceptibility***

284 To assess the effects of the SNPs in the context of haplotypes, we focused on the haplotype
285 block defined by the *MUC5B* and the three adjacent *TOLLIP* SNPs, which had independently
286 shown an association with susceptibility to fHP (rs35705950-rs3750920-rs111521887-
287 rs5743894 block, henceforth referred to as MT1T2T3). Of the 6 major haplotypes
288 reconstructed from the whole population (IPF, fHP and controls individuals), 5 showed an
289 association with susceptibility to disease. Haplotypes T-T-G-C, T-T-C-T, and T-C-C-T were
290 significantly more frequent, and haplotypes G-C-C-T and G-T-C-T were significantly less
291 frequent in patients (IPF and fHP together) than in controls (Supplementary Table S4). In
292 patients, the *TOLLIP* SNPs were in robust linkage disequilibrium (LD) among themselves and
293 moderately linked with the *MUC5B* SNP, as determined by pairwise LD measures. Conversely,
294 in the control group, while *TOLLIP* rs111521887 and rs5743894 were fully linked, these were
295 only moderately linked with *TOLLIP* rs3750920, and the *TOLLIP* SNPs were in low LD with
296 the *MUC5B* SNP (Supplementary Table S5 and Fig. S2).
297 Haplotypes defined by the block incorporating the three *TOLLIP* SNPs only, did not show any
298 association with either condition (not shown).
299 A direct comparison between IPF and fHP patients did not reveal significant associations of
300 any haplotype with either disease (Supplementary Table S6). Nonetheless, pairwise LD
301 measures for the two disorders indicated that while the three *TOLLIP* SNPs were in full LD in
302 IPF patients, only rs111521887 and rs5743894 were fully linked and linkage of these with
303 rs3750920 was just moderate in fHP patients. In addition, LD between *MUC5B* rs35705950
304 and the *TOLLIP* SNPs was markedly stronger in IPF than in fHP (Fig. S2).
305
306
307

MUC5B and TOLLIP haplotype associations with IPF and fHP survival

Kaplan-Meier survival analyses of the haplotypes defined by the same block showed a significant association between the T-C-C-T haplotype (composed of the *MUC5B* minor allele and the *TOLLIP* ancestral alleles) and improved survival in patients with IPF (log rank test *P*= 0.027). The difference in median survival time between non-carriers and carriers was of 24 months (Fig.2).

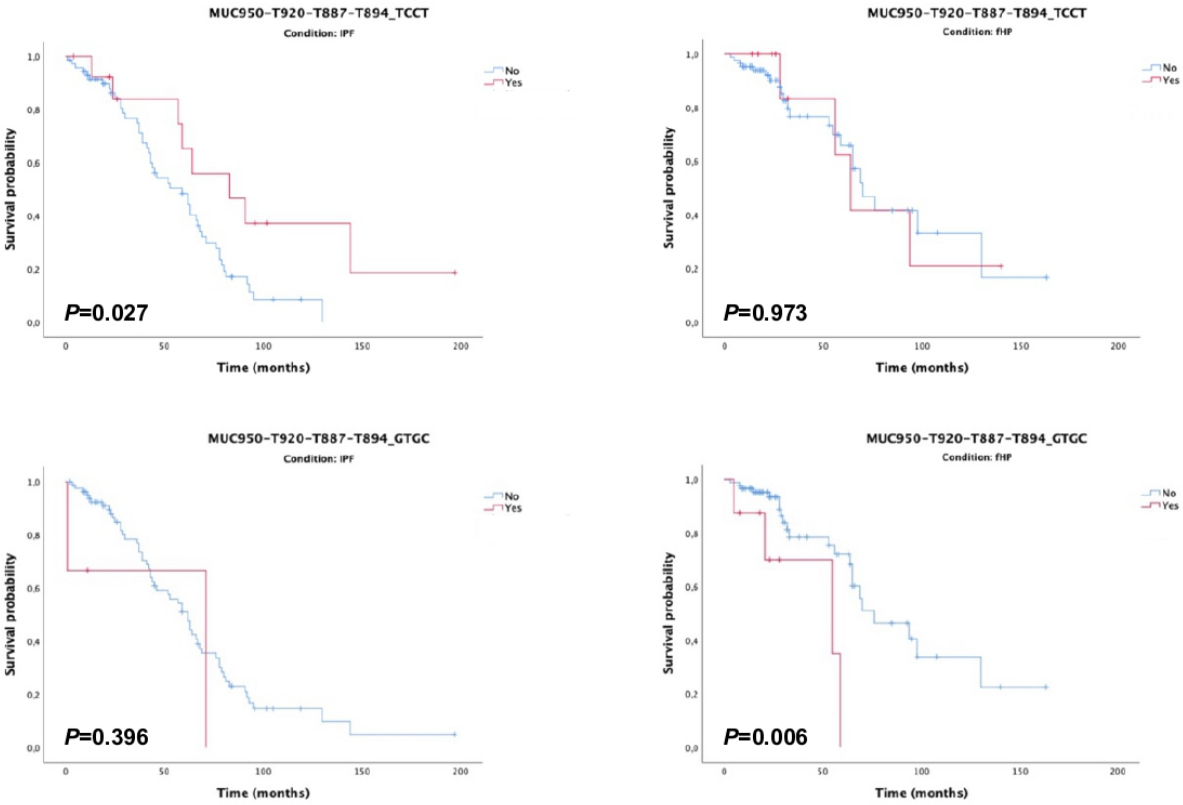


Figure 2. Kaplan-Meier plots representing IPF and fHP patients' survival according to the presence of haplotypes T-C-C-T and G-T-G-C defined by the rs35705950-rs3750920-rs111521887-rs5743894 block. Time in months; *P* value for log-rank test.

Conversely, the haplotype G-T-G-C (integrating the *MUC5B* ancestral allele and the *TOLLIP* minor alleles) was associated with reduced survival in fHP patients (log rank test *P*=0.006)

(Fig.2). Despite its relatively low frequency, carriers of this haplotype showed a median survival time 21 months shorter than non-carriers.

Survival of IPF and HP patients was directly compared in carriers of the 6 major haplotypes defined by the MT1T2T3 block. Notably, Kaplan-Meier analyses revealed that fHP carriers of the G-C-C-T haplotype had a significantly higher survival than IPF carriers of the same haplotype (log rank test *P*=0.006). This haplotype, which is composed of the ancestral alleles of all SNPs and thus the most frequent in both cohorts, conferred a median survival increase of 39 months in fHP patients relative to IPF patients (median survival: IPF=59 months vs fHP=98 months) (Fig.3).

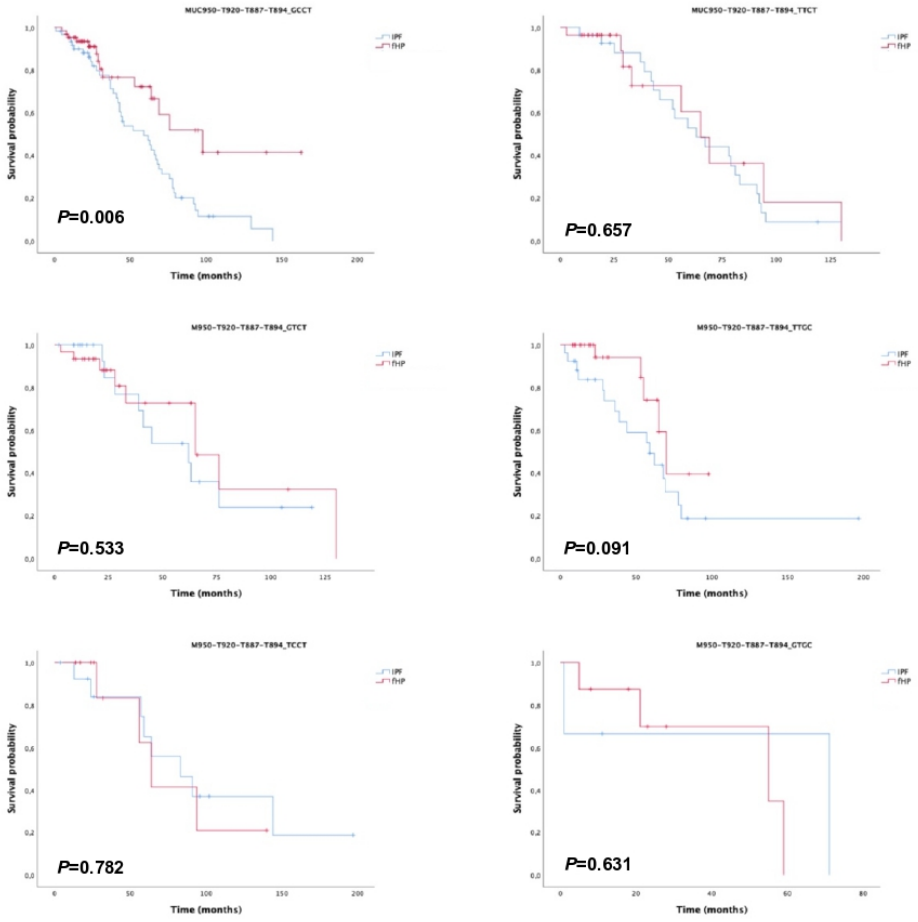


Figure 3. Kaplan-Meier plots comparing IPF and fHP patients' survival according to the 6 major haplotypes defined by the rs35705950-rs3750920-rs111521887-rs5743894 *MUC5B-TOLLIP* block. Time in months; *P* value for log-rank test.

Multivariate Cox regression analysis confirmed that the G-C-C-T haplotype was a strong predictor of better survival for fHP patients relative to IPF patients after adjustment for gender, age and UIP pattern (HR=2.05, 95% CI 1.01-4.18, *P*=0.048, Model 1), as well as when adjusted for gender, age and honeycombing (HR=2.06, 95% CI 1.07-3.97, *P*=0.031, Model 2) (Table 4).

Table 4. Multivariate Cox proportional hazard regressions predicting patients' death in IPF vs. fHP carriers of haplotype G-C-C-T

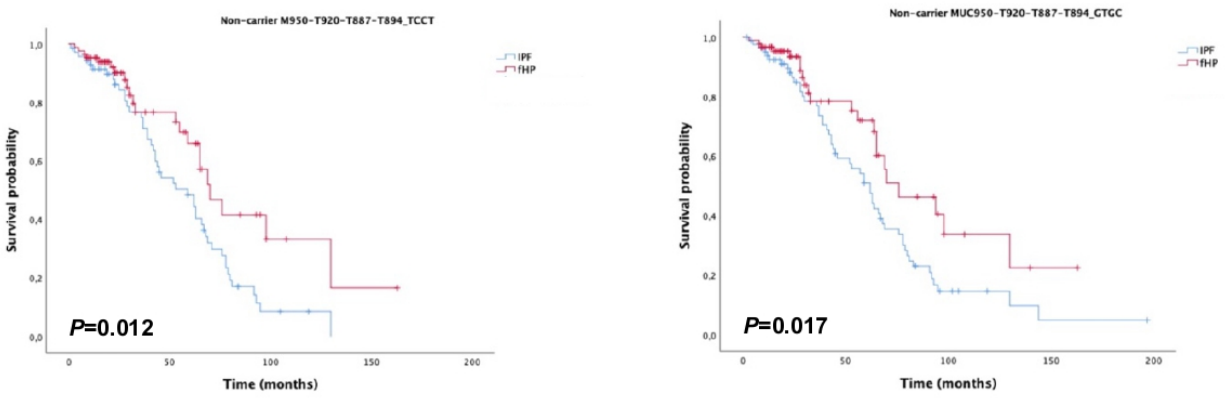
		HR	**95% CI**	**p-value**
Univariate analysis				
Condition				
	fHP	—	—	
	IPF	2.28	1.24-4.18	0.008
*Multivariate analysis**				
Model 1				
Disease	fHP	—	—	
	IPF	2.05	1.01-4.18	0.048
Age		1.02	0.99-1.06	0.233
Gender	female	—	—	
	male	1.51	0.80-2.86	2.862
UIP pattern	no	—	—	
	yes	1.03	0.54-1.94	0.940
Model 2				
Disease	fHP	—	—	
	IPF	2.06	1.07-3.97	0.031
Age		1.02	0.99-1.06	0.226
Gender	female	—	—	
	male	1.51	0.80-2.84	0.204
Honeycombing	no	—	—	
	yes	1.05	0.60-1.84	0.860

HR Hazard Ratio; 95% CI Confidence Interval
IPF, Idiopathic pulmonary fibrosis; fHP, fibrotic hypersensitivity pneumonitis
rs35705950- rs3750920-rs111521887-rs5743894 *MUC5B-TOLLIP* haplotype block
* Multivariate Cox regression models obtained by enter-method.

353

354 Interestingly, while there were no differences in survival between IPF and fHP patients
355 carrying haplotypes T-C-C-T (associated with increased survival in IPF) and G-T-G-C
356 (associated with decreased survival in fHP), KM analyses of IPF vs. fHP in non-carriers of
357 these haplotypes revealed that fHP non-carriers had significantly higher survival than IPF non-
358 carriers (log rank test *P*=0.012 and *P*=0.017, respectively) (Fig.4). The median survival
359 difference was of 11 months (IPF=59 vs fHP=70) for haplotype T-C-C-T and 14 months
360 (IPF=62 vs fHP=76) for haplotype G-T-G-C.

361



369 **Figure 4.** Kaplan-Meier plots comparing IPF and fHP patients' survival in non-carriers of haplotypes
370 T-C-C-T and G-T-G-C defined by the rs35705950-rs3750920-rs111521887-rs5743894 *MUC5B-*
371 *TOLLIP* block. Time in months; *P* value for log-rank test.

372

373 KM analysis of the haplotypes defined by the block integrating the *TOLLIP* SNPs only
374 (T1T2T3), revealed that haplotype C-C-T was also associated with significantly higher survival
375 in fHP comparing to IPF carriers (log rank test *P*=0.028. Median survival: IPF=63 months vs
376 fHP=94 months) (not shown).

377

Discussion

Through an examination of specific genetic markers formerly associated with IPF and fHP, our study sought to unveil genetic patterns that could help differentiate between the two fibrotic disorders, with influence on susceptibility and/or survival, on one hand, or suggestive of shared mechanisms and common pathways underlying both diseases, on the other. The analyses confirmed previous studies showing that the *MUC5B* promoter rs35705950 variant and *TOLLIP* SNPs rs3750920, rs111521887, and rs5743894 differ between patients and controls [12, 17]. Individually, the minor allele of *MUC5B* rs35705950 showed a higher frequency in both IPF and fHP, while the minor alleles of *TOLLIP* rs3750920, rs111521887, and rs5743894 were more frequent only in fHP when compared to controls [10, 12, 14, 17, 29]. However, when directly comparing IPF and fHP, no significant differences were found, even among those with a radiological UIP pattern, illustrating a degree of genetic similarity suggestive of shared fibrosing mechanisms among both fILDs in Portuguese patients, but also revealing an inherent complexity in what concerns the influence of genetics in susceptibility for the two disorders. Interestingly, our analyses focusing on haplotypes formed by the SNPs with significant differences among phenotypes, rather than on each variant individually, revealed associations with survival that may help discriminate between IPF and fHP in carriers (and non-carriers) of specific haplotypes.

Our results are consistent with recent research suggesting some genetic overlap between IPF and fHP, particularly regarding the minor allele of *MUC5B* rs35705950, and highlighting the potential of *TOLLIP* SNPs in predicting risk and survival of fHP patients when compared to controls; but not enabling a clear differentiation between IPF and fHP [16, 17, 30]. The robust association of *MUC5B* rs35705950 with risk for both IPF and fHP, consistent with the literature and our own previous studies, reaffirms its role as a shared genetic risk factor also in Portuguese patients [12, 17]. Additionally, the association of this variant with other fILDs, such as

403 rheumatoid arthritis-associated interstitial lung disease (RA-ILD), was already described,
404 revealing common features among patients with these conditions [31]. For instance, in a
405 multicenter European study, van der Vis et al. (2023) showed that the *MUC5B* minor allele is
406 linked to older age at diagnosis in IPF, as previously observed in fHP and RA-ILD [16, 32, 33].
407 The intricate interactions involving *TOLLIP* SNPs present challenges for interpretation in the
408 context of both IPF and fHP. *TOLLIP* has been implicated in susceptibility to and in
409 progression of IPF, particularly due to its effects on autophagy and apoptosis during the
410 disease's progression stages [34]. Previous research suggests that minor alleles of certain
411 *TOLLIP* SNPs, such as rs111521887 and rs5743894, may contribute to IPF susceptibility,
412 while others like rs5743890 could potentially mitigate this risk, despite an associated increase
413 in mortality [34-36]. However, uncertainty persists regarding whether these associations are
414 independent of the neighboring *MUC5B* rs35705950 variant [37]. In a recent study involving
415 Portuguese patients, no significant associations were observed between five *TOLLIP* SNPs
416 (including those mentioned above) and IPF susceptibility [12]. These observations are further
417 reinforced by the present study, which incorporates a larger sample of IPF patients and controls.
418 Whether they could represent a distinctive feature of Portuguese or even European patients, as
419 suggested by research highlighting the ethnic heterogeneity in genetic associations related to
420 fILDs, remains to be further explored [10, 38].
421 Furusawa et al. (2020) demonstrated significant similarities in gene expression profiles
422 between IPF and fHP by identifying a considerable proportion of genes associated with
423 collagen catabolic processes and epithelial development that were shared between the two
424 conditions [16]. Like Selman et al. (2006), they also reported genes specifically associated with
425 fHP, i.e. showing differential expression in fHP compared to controls but not in IPF [30]. Some
426 of these genes were linked to chemokine-mediated signaling and immune response, such as
427 Biglycan (*BGN*). *BGN* encodes a proteoglycan that binds to toll-like receptors 2 and 4,

428 triggering inflammation and establishing a potential link between immune response and lung
429 fibrosis in fHP [16]. Similarly, *TOLLIP* is a biologically plausible candidate for fHP risk,
430 considering its involvement in innate immunity, protein trafficking, and inflammatory
431 responses [34, 36, 39]. Evidence connecting the *TOLLIP* gene to fHP remains limited, but we
432 have recently shown that the minor alleles of *TOLLIP* rs3750920, rs111521887, and rs5743894
433 were significantly more frequent in fHP patients than in controls [17]. Notwithstanding, and
434 considering the lack of associations between *TOLLIP* SNPs and IPF, the present study did not
435 identify a clear genetic distinction when directly comparing allelic and genotypic frequencies
436 of *TOLLIP* between IPF and fHP patients. Indeed, genotypic and allelic distributions of the
437 *MUC5B* rs35705950 and *TOLLIP* rs3750920, rs111521887, and rs5743894 SNPs did not differ
438 significantly between IPF and fHP patients, even in the context of the presence of an UIP
439 pattern or honeycombing on HRCT. In recent years, it has become evident that an UIP pattern,
440 typically associated with IPF, can also be present in other fILDs and may carry a similarly poor
441 prognosis [31, 40]. However, unlike IPF, the presence of a probable UIP pattern on HRCT is
442 not reliably indicative of underlying UIP when the disease is not idiopathic or when
443 demographic and clinical factors differ from those typically seen in IPF [41]. A potential
444 confounder in our analysis arises from the pooling of patients with UIP and probable UIP
445 patterns within the same category, regardless of the underlying disease. However, this
446 limitation was mitigated by introducing an additional analysis: the presence of honeycombing
447 on HRCT, which is also recognized as a negative prognostic indicator [40, 42]. This too, did
448 not reveal significant genetic differences between the two fibrotic disorders. These
449 observations suggest that the genetic influence on phenotype may arise from interactions with
450 other loci, that the data requires a broader analytical approach based on haplotypes, or both.
451 Abbasi et al. (2022) conducted a pioneering study examining the surfactant protein genes
452 (*SFTPs*) in both IPF and fHP, focusing on SNPs. They identified variants, haplotypes and

variant interactions among the several *SFTPs* with varying effects on both fILDs [43]. Indeed, the use of haplotypes defined by neighboring SNPs can reliably provide better information than that produced by the study of SNPs individually, as variants may be in linkage disequilibrium and thus inherited as a block [44-46].

Haplotypes defined by the SNPs which were associated with either IPF or fHP revealed significant differences in haplotype frequencies between the overall group of patients (IPF and fHP) and controls, reinforcing the notion that the genomic landscape encompassing the *MUC5B* and *TOLLIP* loci is determinant for the phenotype. The three haplotypes significantly more common in the group of patients all incorporated the *MUC5B* rs35705950 minor allele T, whereas the two more frequent in controls bore the ancestral allele G. This is consistent with the significantly higher frequency of the *MUC5B* T allele observed in either IPF and fHP when compared to controls. Notwithstanding, pairwise LD measures indicated that in patients the *TOLLIP* SNPs are only moderately linked with the *MUC5B* SNP, suggesting that the genetic impact on susceptibility to these fibrosing disorders is not solely determined by the latter. Although none of the haplotypes significantly associated with either disease upon directly comparing IPF and fHP haplotype frequencies, pairwise LD measures for the two disorders revealed different patterns of linkage. In IPF patients the *TOLLIP* SNPs were in full LD with each other and in robust linkage with the *MUC5B* SNP, suggesting a tighter and more homogeneous block. Conversely, in fHP patients, only *TOLLIP* rs111521887 and rs5743894 (farther downstream from *MUC5B*) were in full LD, pointing towards the possibility of recombination events taking place between these and rs3750920, and more likely with *MUC5B* rs35705950, thus forming a looser block. This configuration stands somewhat midway between the IPF and the controls LD patterns, with the latter displaying low linkage between fully linked *TOLLIP* rs111521887-rs5743894 and rs3750920 and virtually no linkage to *MUC5B* rs35705950. These observations show that despite the fact that susceptibility to IPF and fHP

478 cannot be distinguished on the basis of the frequency of studied SNPs and haplotypes, the latter
479 can be useful to uncover informative LD patterns.
480 Notably, our analyses using haplotypes unveiled significant associations with prognosis in both
481 IPF and fHP patients. In IPF, carriers of the T-C-C-T haplotype, which is composed of the
482 *MUC5B* minor allele and the *TOLLIP* ancestral alleles, showed increased survival when
483 compared to non-carriers. These results are in accordance with our previous work showing that
484 the same alleles of *MUC5B* rs35705950 and *TOLLIP* rs111521887 and rs5743894 integrated a
485 broader haplotype associated with improved survival in IPF [12]. Interestingly, while both
486 TOLLIP SNPs individually associated with survival in IPF patients this was not the case of
487 *MUC5B* rs35705950. The lack of association between the *MUC5B* rs35705950 minor allele
488 and survival is supported by several studies probing patients from different ethnicities [10, 12,
489 47, 48], which in turn do not corroborate such an association as previously proposed [49]. In
490 fHP, the haplotype G-T-G-C (integrating the *MUC5B* ancestral allele and the *TOLLIP* minor
491 alleles) was associated with reduced survival in carriers. Although the latter alleles were not
492 individually associated with survival in fHP, they were significantly more frequent in fHP
493 patients than in controls, pointing towards a relevant role of *TOLLIP* in this disorder.
494 Importantly, survival analysis in the context of a direct comparison between IPF and fHP
495 patients unveiled a haplotype that promoted better survival in fHP carriers compared to IPF
496 carriers. The effect of the G-C-C-T haplotype (the commonest and composed of the ancestral
497 alleles of all SNPs) was independent of age, gender or the presence of radiological UIP pattern
498 or honeycombing, as determined by multivariate analyses. Furthermore, the same effect of
499 better survival in fHP relative to IPF patients was observed when comparing non-carriers of
500 haplotypes T-C-C-T and G-T-G-C between the two conditions, the former having shown a
501 significant association with increased survival in IPF patients, and the latter with decreased
502 survival in fHP. These observations seem to suggest that, in the context of survival among the

two conditions, in fHP patients, a haplotype integrating the block of *TOLLIP* ancestral alleles (C-C-T) is protective only when these are inherited along with the *MUC5B* ancestral allele, whereas in IPF patients, survival is improved when these are inherited with the *MUC5B* minor allele (this study and ref. 12). This effect cannot be solely attributed to *MUC5B* rs35705950, not only because this SNP is not associated with survival on its own, as mentioned above, but also because the C-C-T haplotype, defined by the T1T2T3 block alone, is itself associated with significantly higher survival in fHP comparing to IPF carriers. Thus, it could be speculated that either the variants themselves or a linked as-yet unidentified genetic factor within the region may influence varying clinical outcomes in the context of specific genomic landscape conformations.

Of note, although DLCO was significantly lower in IPF patients, as opposed to FVC and TLC that were similar between both diseases, it was not included in the multivariate model, as its interpretation warrants caution due to inherent measurement variability within patients, differences in techniques across pulmonary function laboratories, and its lack of specificity for tracking pulmonary fibrosis progression. The severity of DLCO reduction may not always align closely with the disease state, as extensive areas of fibrosis do not invariably result in severely reduced DLCO [50]. Conversely, some ILD patients may exhibit significant reductions in lung volumes or severe hypoxemia despite having normal or only slightly reduced DLCO values [51, 52]. Furthermore, the presence of moderate-to-severe reductions in DLCO alongside normal lung volumes may indicate ILD coexisting with emphysema or pulmonary vascular disease [50, 52]. In fact, in IPF patients, pulmonary hypertension was more frequent.

Limitations of this work include the single center nature of the study, sample size, and some heterogeneity in disease severity at baseline and pharmacological treatment.

Conclusions

Altogether, our results unveil genetic similarities between IPF and fHP alongside with distinctive features with respect to the analyzed loci, emphasizing the need for a deeper understanding of the complexity behind the genetic mechanisms underlying the pathological processes of both fibrotic disorders, aimed at improved patient management and treatment, and paving the way for personalized approaches to diagnosis and therapy in future.

List of abbreviations

fILDs Fibrotic interstitial lung diseases
fHP Fibrotic hypersensitivity pneumonitis
IPF Idiopathic pulmonary fibrosis
SNPs Single nucleotide polymorphisms
MUC5B mucin 5B
TOLLIP Toll-interacting protein
FVC Forced vital capacity
TLC Total lung capacity
DLCO Diffusing capacity of the lung for carbon monoxide
BAL Bronchoalveolar lavage
RFLP Restriction fragment length polymorphism
PCR Protein chain reaction
HRCT High-resolution computed tomography
UIP Usual interstitial pneumonia
TBLC Transbronchial lung *cryobiopsy*
LD Linkage disequilibrium
HR Hazard Ratio
CI Confidence interval
RA-ILD Rheumatoid arthritis-associated interstitial lung disease
GWAS Genome-wide association studies
BGN Biglycan
SP Surfactant proteins

559 **Declarations**

560 **Ethics approval**

561 This study was performed in line with the principles of the Declaration of Helsinki. Approval
562 was granted by the Ethics Committee of Unidade Local de Saúde de São João (CE-203–2012).

563

564 **Consent to participate**

565 Written informed consent was obtained from all individual participants in the study.

566

567 **Availability of data and materials**

568 The datasets used and/or analysed during the current study are available from the corresponding
569 author on reasonable request.

570

571 **Competing interests**

572 The authors have no competing interests.

573

574 **Funding**

575 This research was made possible through the generous support of the Portuguese Society of
576 Pulmonology (SPP) (Grant – SPP/Bolsa Novartis 2012), Boehringer-Ingelheim (Unrestricted
577 Research Grant), Fundação para a Ciência e a Tecnologia, I.P. (within the scope of project
578 PTDC/MEC-RES/0158/2020), and Fundação Amélia de Mello and D. José de Mello. JCR is
579 funded by Fundação para a Ciência e a Tecnologia through the Stimulus for Scientific
580 Employment Individual Support (2020.01350.CEECIND).

581

582 **Authors´contributions**

583 Conceptualization: P.C.M., M.L.S.,and A.M.; methodology: P.C.M., M.L.S. ,and A.M.; data
584 curation: P.C.M., M.L.S., J.R., and C.C.D.; formal analysis: P.C.M., M.L.S., J.R., and C.C.D.;
585 investigation: P.C.M., M.L.S., C.D.V., A.C.F., R.F.S.,O.O., M.B., C.G.C., N.M., A.T.A, D.C.,
586 H.N-B., and A.M.; validation: P.C.M., M.L.S., and A.M.; writing—original draft: P.C.M.,
587 M.L.S., J.R., and C.C.D; writing—review and editing: all authors; resources: P.C.M., M.L.S.,
588 R.F.S., N.M., and H. N-B.; funding acquisition: P.C.M., M.L.S.,H. N-B., and A.M.;
589 supervision and project administration: P.C.M., M.L.S., and A.M..

590

591 **Acknowledgements**

592 We would like to express our gratitude to the Instituto de Saúde Pública da Universidade do
593 Porto (ISPUP) for generously providing samples from the EPI-Porto cohort to serve as controls.
594 Our sincere appreciation goes to the cohort management team and all participants for their
595 valuable contributions. Special thanks to Catarina Teixeira Antunes and Joana Filipa Amorim
596 Reis for their assistance with the genotyping process.

597

598

599

600

601

602

603

REFERENCES

1. Raghu G, Remy-Jardin M, Richeldi L, Thomson CC, Inoue Y, Johkoh T, Kreuter M, Lynch DA, Maher TM, Martinez FJ, et al: **Idiopathic Pulmonary Fibrosis (an Update) and Progressive Pulmonary Fibrosis in Adults: An Official ATS/ERS/JRS/ALAT Clinical Practice Guideline.** *Am J Respir Crit Care Med* 2022, **205:**e18-e47.
2. Marinescu DC, Raghu G, Remy-Jardin M, Travis WD, Adegunsoye A, Beasley MB, Chung JH, Churg A, Cottin V, Egashira R, et al: **Integration and Application of Clinical Practice Guidelines for the Diagnosis of Idiopathic Pulmonary Fibrosis and Fibrotic Hypersensitivity Pneumonitis.** *Chest* 2022, **162:**614-629.
3. Raghu G, Remy-Jardin M, Ryerson CJ, Myers JL, Kreuter M, Vasakova M, Bargagli E, Chung JH, Collins BF, Bendstrup E, et al: **Diagnosis of Hypersensitivity Pneumonitis in Adults. An Official ATS/JRS/ALAT Clinical Practice Guideline.** *Am J Respir Crit Care Med* 2020, **202:**e36-e69.
4. Selman M, King TE, Pardo A, American Thoracic S, European Respiratory S, American College of Chest P: **Idiopathic pulmonary fibrosis: prevailing and evolving hypotheses about its pathogenesis and implications for therapy.** *Ann Intern Med* 2001, **134:**136-151.
5. Vasakova M, Selman M, Morell F, Sterclova M, Molina-Molina M, Raghu G: **Hypersensitivity Pneumonitis: Current Concepts of Pathogenesis and Potential Targets for Treatment.** *Am J Respir Crit Care Med* 2019, **200:**301-308.
6. Churg A, Sin DD, Everett D, Brown K, Cool C: **Pathologic patterns and survival in chronic hypersensitivity pneumonitis.** *Am J Surg Pathol* 2009, **33:**1765-1770.
7. Seibold MA, Wise AL, Speer MC, Steele MP, Brown KK, Loyd JE, Fingerlin TE, Zhang W, Gudmundsson G, Groshong SD, et al: **A common MUC5B promoter polymorphism and pulmonary fibrosis.** *N Engl J Med* 2011, **364:**1503-1512.
8. Zhang GY, Liao T, Gao WY: **MUC5B promoter polymorphism and pulmonary fibrosis.** *N Engl J Med* 2011, **365:**178; author reply 178-179.
9. Peljto AL, Selman M, Kim DS, Murphy E, Tucker L, Pardo A, Lee JS, Ji W, Schwarz MI, Yang IV, et al: **The MUC5B promoter polymorphism is associated with idiopathic pulmonary fibrosis in a Mexican cohort but is rare among Asian ancestries.** *Chest* 2015, **147:**460-464.
10. Bonella F, Campo I, Zorzetto M, Boerner E, Ohshimo S, Theegarten D, Taube C, Costabel U: **Potential clinical utility of MUC5B und TOLLIP single nucleotide polymorphisms (SNPs) in the management of patients with IPF.** *Orphanet J Rare Dis* 2021, **16:**111.
11. Oldham JM, Ma SF, Martinez FJ, Anstrom KJ, Raghu G, Schwartz DA, Valenzi E, Witt L, Lee C, Vij R, et al: **TOLLIP, MUC5B, and the Response to N-Acetylcysteine among Individuals with Idiopathic Pulmonary Fibrosis.** *Am J Respir Crit Care Med* 2015, **192:**1475-1482.
12. Mota PC, Soares ML, Vasconcelos CD, Ferreira AC, Lima BA, Manduchi E, Moore JH, Melo N, Novais-Bastos H, Pereira JM, et al: **Predictive value of common genetic variants in idiopathic pulmonary fibrosis survival.** *J Mol Med (Berl)* 2022, **100:**1341-1353.
13. Blach J, Mackiewicz B: **How much do we know about genetic predisposition of hypersensitivity pneumonitis?** *Ann Agric Environ Med* 2022, **29:**306-308.
14. Ley B, Newton CA, Arnould I, Elicker BM, Henry TS, Vittinghoff E, Golden JA, Jones KD, Batra K, Torrealba J, et al: **The MUC5B promoter polymorphism and**

telomere length in patients with chronic hypersensitivity pneumonitis: an observational cohort-control study. *Lancet Respir Med* 2017, **5:**639-647.

15. Furusawa H, Peljto AL, Walts AD, Cardwell J, Molyneaux PL, Lee JS, Fernandez Perez ER, Wolters PJ, Yang IV, Schwartz DA: **Common idiopathic pulmonary fibrosis risk variants are associated with hypersensitivity pneumonitis.** *Thorax* 2022, **77:**508-510.

16. Furusawa H, Cardwell JH, Okamoto T, Walts AD, Konigsberg IR, Kurche JS, Bang TJ, Schwarz MI, Brown KK, Kropski JA, et al: **Chronic Hypersensitivity Pneumonitis, an Interstitial Lung Disease with Distinct Molecular Signatures.** *Am J Respir Crit Care Med* 2020, **202:**1430-1444.

17. Mota PC, Soares ML, Ferreira AC, Santos RF, Rufo JC, Vasconcelos D, Carvalho A, Guimaraes S, Vasques-Novoa F, Cardoso C, et al: **Polymorphisms and haplotypes of TOLLIP and MUC5B are associated with susceptibility and survival in patients with fibrotic hypersensitivity pneumonitis.** *Pulmonology* 2024.

18. **http://doi.org/10.54499/PTDC/MEC-RES/0158/2020.**

19. Ramos E, Lopes C, Barros H: **Investigating the effect of nonparticipation using a population-based case-control study on myocardial infarction.** *Ann Epidemiol* 2004, **14:**437-441.

20. Pintalhao M, Vasques-Novoa F, Couto-Viana B, Pimentel MJ, Neves JS, Mendonca L, Saraiva F, Lourenco AP, Ferreira G, Macedo F, et al: **Relaxin-2, pathophysiological insights and outcomes in heart failure with preserved ejection fraction: Findings from the NETDiamond cohort.** *Int J Cardiol* 2022, **365:**87-90.

21. Leite AR, Neves JS, Angelico-Goncalves A, Vasques-Novoa F, Saraiva FA, Pinho IB, Oliveira AC, Borges-Canha M, von Hafe M, Vale C, et al: **Clinical and Pathophysiologic Insights of Free triiodothyronine/Free thyroxine Ratio in Patients with Heart Failure with Preserved Ejection Fraction: Data from the NETDiamond Cohort.** *Cardiology* 2023, **148:**239-245.

22. Fernandez Perez ER, Travis WD, Lynch DA, Brown KK, Johannson KA, Selman M, Ryu JH, Wells AU, Tony Huang YC, Pereira CAC, et al: **Diagnosis and Evaluation of Hypersensitivity Pneumonitis: CHEST Guideline and Expert Panel Report.** *Chest* 2021, **160:**e97-e156.

23. Raghu G, Collard HR, Egan JJ, Martinez FJ, Behr J, Brown KK, Colby TV, Cordier JF, Flaherty KR, Lasky JA, 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:**788-824.

24. Raghu G, Remy-Jardin M, Myers JL, Richeldi L, Ryerson CJ, Lederer DJ, Behr J, Cottin V, Danoff SK, Morell F, et al: **Diagnosis of Idiopathic Pulmonary Fibrosis. An Official ATS/ERS/JRS/ALAT Clinical Practice Guideline.** *Am J Respir Crit Care Med* 2018, **198:**e44-e68.

25. Travis WD, Costabel U, Hansell DM, King TE, Jr., Lynch DA, Nicholson AG, Ryerson CJ, Ryu JH, Selman M, Wells AU, et al: **An official American Thoracic Society/European Respiratory Society statement: Update of the international multidisciplinary classification of the idiopathic interstitial pneumonias.** *Am J Respir Crit Care Med* 2013, **188:**733-748.

26. Barrett JC, Fry B, Maller J, Daly MJ: **Haploview: analysis and visualization of LD and haplotype maps.** *Bioinformatics* 2005, **21:**263-265.

27. Stephens M, Scheet P: **Accounting for decay of linkage disequilibrium in haplotype inference and missing-data imputation.** *Am J Hum Genet* 2005, **76:**449-462.

28. Stephens M, Smith NJ, Donnelly P: **A new statistical method for haplotype reconstruction from population data.** *Am J Hum Genet* 2001, **68:**978-989.
29. Peljto AL, Blumhagen RZ, Walts AD, Cardwell J, Powers J, Corte TJ, Dickinson JL, Glaspole I, Moodley YP, Vasakova MK, et al: **Idiopathic Pulmonary Fibrosis Is Associated with Common Genetic Variants and Limited Rare Variants.** *Am J Respir Crit Care Med* 2023, **207:**1194-1202.
30. Selman M, Pardo A, Barrera L, Estrada A, Watson SR, Wilson K, Aziz N, Kaminski N, Zlotnik A: **Gene expression profiles distinguish idiopathic pulmonary fibrosis from hypersensitivity pneumonitis.** *Am J Respir Crit Care Med* 2006, **173:**188-198.
31. Juge PA, Solomon JJ, van Moorsel CHM, Garofoli R, Lee JS, Louis-Sydney F, Rojas-Serrano J, Gonzalez-Perez MI, Mejia M, Buendia-Roldan I, et al: **MUC5B promoter variant rs35705950 and rheumatoid arthritis associated interstitial lung disease survival and progression.** *Semin Arthritis Rheum* 2021, **51:**996-1004.
32. van der Vis JJ, Prasse A, Renzoni EA, Stock CJW, Caliskan C, Maher TM, Bonella F, Borie R, Crestani B, Petrek M, et al: **MUC5B rs35705950 minor allele associates with older age and better survival in idiopathic pulmonary fibrosis.** *Respirology* 2023, **28:**455-464.
33. Palomaki A, FinnGen Rheumatology Clinical Expert G, Palotie A, Koskela J, Eklund KK, Pirinen M, FinnGen, Ripatti S, Laitinen T, Mars N: **Lifetime risk of rheumatoid arthritis-associated interstitial lung disease in MUC5B mutation carriers.** *Ann Rheum Dis* 2021, **80:**1530-1536.
34. Li X, Goobie GC, Gregory AD, Kass DJ, Zhang Y: **Toll-Interacting Protein in Pulmonary Diseases. Abiding by the Goldilocks Principle.** *Am J Respir Cell Mol Biol* 2021, **64:**536-546.
35. Noth I, Zhang Y, Ma SF, Flores C, Barber M, Huang Y, Broderick SM, Wade MS, Hysi P, Scuirba J, et al: **Genetic variants associated with idiopathic pulmonary fibrosis susceptibility and mortality: a genome-wide association study.** *Lancet Respir Med* 2013, **1:**309-317.
36. Fingerlin TE, Murphy E, Zhang W, Peljto AL, Brown KK, Steele MP, Loyd JE, Cosgrove GP, Lynch D, Groshong S, et al: **Genome-wide association study identifies multiple susceptibility loci for pulmonary fibrosis.** *Nat Genet* 2013, **45:**613-620.
37. Allen RJ, Guillen-Guio B, Oldham JM, Ma SF, Dressen A, Paynton ML, Kraven LM, Obeidat M, Li X, Ng M, et al: **Genome-Wide Association Study of Susceptibility to Idiopathic Pulmonary Fibrosis.** *Am J Respir Crit Care Med* 2020, **201:**564-574.
38. Adegunsoye A, Kropski JA, Behr J, Blackwell TS, Corte TJ, Cottin V, Glanville A, Glassberg MK, Griese M, Hunninghake GM, et al: **Genetics and Genomics of Pulmonary Fibrosis: Charting the Molecular Landscape and Shaping Precision Medicine.** *Am J Respir Crit Care Med* 2024.
39. Li X, Goobie GC, Zhang Y: **Toll-interacting protein impacts on inflammation, autophagy, and vacuole trafficking in human disease.** *J Mol Med (Berl)* 2021, **99:**21-31.
40. Dasgupta S, Bhattacharya A, Abhijit RD, Roy Chowdhury S, Chaudhury K: **Risk factors associated with mortality in hypersensitivity pneumonitis: a meta-analysis.** *Expert Rev Respir Med* 2022, **16:**801-811.
41. Selman M, Pardo A, Wells AU: **Usual interstitial pneumonia as a stand-alone diagnostic entity: the case for a paradigm shift?** *Lancet Respir Med* 2023, **11:**188-196.
42. Salisbury ML, Gu T, Murray S, Gross BH, Chughtai A, Sayyouh M, Kazerooni EA, Myers JL, Lagstein A, Konopka KE, et al: **Hypersensitivity Pneumonitis:**

751 **Radiologic Phenotypes Are Associated With Distinct Survival Time and**
752 **Pulmonary Function Trajectory.** *Chest* 2019, **155:**699-711.
753 43. Abbasi A, Chen C, Gandhi CK, Wu R, Pardo A, Selman M, Floros J: **Single**
754 **Nucleotide Polymorphisms (SNP) and SNP-SNP Interactions of the Surfactant**
755 **Protein Genes Are Associated With Idiopathic Pulmonary Fibrosis in a Mexican**
756 **Study Group; Comparison With Hypersensitivity Pneumonitis.** *Front Immunol*
757 2022, **13:**842745.
758 44. Shirali M, Knott SA, Pong-Wong R, Navarro P, Haley CS: **Haplotype Heritability**
759 **Mapping Method Uncovers Missing Heritability of Complex Traits.** *Sci Rep*
760 2018, **8:**4982.
761 45. Schrodi SJ, Garcia VE, Rowland C, Jones HB: **Pairwise linkage disequilibrium**
762 **under disease models.** *Eur J Hum Genet* 2007, **15:**212-220.
763 46. Khankhanian P, Gourraud PA, Lizee A, Goodin DS: **Haplotype-based approach to**
764 **known MS-associated regions increases the amount of explained risk.** *J Med*
765 *Genet* 2015, **52:**587-594.
766 47. Jiang H, Hu Y, Shang L, Li Y, Yang L, Chen Y: **Association between MUC5B**
767 **polymorphism and susceptibility and severity of idiopathic pulmonary fibrosis.**
768 *Int J Clin Exp Pathol* 2015, **8:**14953-14958.
769 48. van der Vis JJ, Snetselaar R, Kazemier KM, ten Klooster L, Grutters JC, van Moorsel
770 CH: **Effect of Muc5b promoter polymorphism on disease predisposition and**
771 **survival in idiopathic interstitial pneumonias.** *Respirology* 2016, **21:**712-717.
772 49. Peljto AL, Zhang Y, Fingerlin TE, Ma SF, Garcia JG, Richards TJ, Silveira LJ,
773 Lindell KO, Steele MP, Loyd JE, et al: **Association between the MUC5B promoter**
774 **polymorphism and survival in patients with idiopathic pulmonary fibrosis.** *JAMA*
775 2013, **309:**2232-2239.
776 50. Miguel-Reyes JL, Gochicoa-Rangel L, Perez-Padilla R, Torre-Bouscoulet L:
777 **Functional respiratory assessment in interstitial lung disease.** *Rev Invest Clin*
778 2015, **67:**5-14.
779 51. Pellegrino R, Viegi G, Brusasco V, Crapo RO, Burgos F, Casaburi R, Coates A, van
780 der Grinten CP, Gustafsson P, Hankinson J, et al: **Interpretative strategies for lung**
781 **function tests.** *Eur Respir J* 2005, **26:**948-968.
782 52. Collard HR, King TE, Jr., Bartelson BB, Vourlekis JS, Schwarz MI, Brown KK:
783 **Changes in clinical and physiologic variables predict survival in idiopathic**
784 **pulmonary fibrosis.** *Am J Respir Crit Care Med* 2003, **168:**538-542.
785
786

Supplementary Material

Supplementary methods

M1. Genotyping

Genomic DNA was extracted from peripheral blood using the Wizard Genomic DNA Purification Kit (Promega Corp., Madison, USA) or the QIAamp DNA Blood kit (Qiagen, Hilden, Germany).

For restriction fragment length polymorphism (RFLP) analysis, DNA sequences containing the polymorphic sites were amplified by PCR in 20 µl reactions containing 20 ng of genomic DNA, 1.5 mM of each primer, and Kapa Taq EXtra HotStart ReadyMix (Merck, Darmstadt, Germany). Cycling conditions were 95ºC denaturing for 5 min, 35 cycles of 30 sec at 95ºC, 30 sec at annealing temperature, and 30 sec at 72ºC, with a final extension for 10 min at 72ºC. The amplicons were digested overnight in 15µl reactions with 8µl of PCR product and 3 units of restriction endonuclease. TaqMan assays were carried out according to manufacturer's instructions.

Table S1. Genotyping methods and conditions

Primers, conditions and restriction enzymes for RFLP analyses

Locus	Primer name	SNP loc (GRCh38.p14)	Primer sequence (5' to 3')	Annealing (°C)	Amplicon Size (bp)	Restriction Enzyme	RFLP fragments (bp)
MUC5B	rs35705950_Fw	5'-regulatory 1219991	GCCAGAATGAGGGACAGTGA	60	375	HhaI	G: 189+105+81 T: 270+105
	rs35705950_Rv		ACTTTGCCCTCGTCCCTC				
TOLLIP	rs5743854_Fw	5'-regulatory 1310024	CATCCCCGTTCCGTTCAC	60	250	SbfI	G: 250 C: 134+116
	rs5743854_Rv		TTCCTCCTCGGTCCTCTGT				
	rs5743890_Fw	Intronic 1304599	GGCTCCCCTGGAAATGGTAA	62	408	EarI	T: 408 C: 310+98
	rs5743890_Rv		GTCAGGACACAGAGAAAAGGC				
	rs5743894_Fw	Intronic 1303542	GGCCTCTAACACTGTCCCTT	62	440	FspI	T: 440 C: 247+193
	rs5743894_Rv		AAAAGCCAGTCAAGCAGCAG				
	rs111521887_Fw	Intronic 1291476	GAGGGTTCAGTGAGGACAGG	60	460	AflIII	C: 460 G: 333+127
	rs111521887_Rv		ACTCAACATGGATGGCAGGT				
	rs3750920_Fw	Exonic 1288726	GGTCTCATCGATGGGGTGTT	62	421	MspI	C: 226+195 T: 421
	rs3750920_Rv		CCCTGCTGTTTTATGGGCTC				

Taqman SNP Genotyping Assays for Real-time PCR

Locus	SNP ID	Assay ID*	Context Sequence [VIC/FAM]
MUC5B	rs35705950	C_1582254_20	CCTTCCTTTATCTTCTGTTTTCAGC[G/T]CCTTCAACTGTGAAGAGGTGAACTC
TOLLIP	rs3750920	C_1582189_1	CCACCTTGCCCTGCCTCAGGGACTC [C/T] GGGATGGTGATGTGGGTCCAGGCAA
	rs111521887	C_152324875_10	GACCCCCTCTGAGGGAACGGCCAAC [C/G] TGTCCACACCAACCCCCCTCTGAGA
	rs5743854	C_11281491_20	GTGACAGGCCCGACATCCCCTGCAG [C/G] GCGATCTTTCCCTCAGCTGCCTGCC

SNP loc, SNP location; RFLP, Restriction Fragment Length Polymorphism; Fw, Forward; Rv, Reverse; bp, base pairs
*TaqMan® SNP Genotyping Assays (Thermo Fisher Scientific, Massachusetts, USA)

M2. Radiology

All high-resolution computed tomography (HRCT) scans underwent assessment by a group of experienced radiologists specializing in ILD interpretation, blinded to clinical diagnosis. Only thin-section CT images with a thickness of less than 1.5 mm and intervals ranging from 1 to 10 mm were included, utilizing lung kernels and lung windows. Additionally, coronal and sagittal reconstructions were available for review. The fibrotic pattern on each scan was classified according to the current usual interstitial pneumonia (UIP) classification outlined by the ATS/JRS/ALAT guidelines: 1) definite UIP or probable UIP, 2) indeterminate for UIP, or 3) alternative diagnosis.[1] The presence of honeycombing, defined by the Fleischner Society, was noted, along with an evaluation of the overall extent of interstitial abnormalities.[2] For each pulmonary lobe (including the lingula as part of the left upper lobe), an estimate of the total percentage of lung affected by fibrosis (including reticular abnormality, traction bronchiectasis, and honeycombing) was assigned a score ranging from 0 to 4, corresponding to 0%, 0-25%, 25-50%, 50-75%, and >75%, respectively. This yielded a total ILD extent score ranging from 0 to 20, which was qualitatively translated to none (0), mild (1-5), moderate (>5-10) or severe (>10) based on predefined thresholds.

M3. Histopathology

Pathology variables were evaluated blindly with respect to clinical data, based on the review of slides obtained from one or two biopsied lobes. All cases were assessed according to the new guidelines for the diagnosis of hypersensitivity pneumonitis (HP) and fibrotic hypersensitivity pneumonitis (fHP).[3] The tissue area on each slide was quantified using QuPath[4,5], and fibrosis quantification was manually performed using the same software. For dichotomous variables, the identification of a feature in one lobe was considered to be present in the patient, while data from all lobes were averaged for continuous variables. The following variables were examined: presence of honeycombing, presence of peribronchiolar metaplasia, presence of giant cells or granulomas (either interstitial or within airspaces), percentage of fibrosis, presence of fibroblastic foci (none, one, or more than one), fibrosis distribution (heterogeneous, peribronchiolar only, peribronchiolar associated with subpleural or paraseptal fibrosis, undetermined), and chronic inflammatory infiltrate (none/light if up to 25%; moderate/severe if more than 25%).

Supplementary tables and figures

Table S2. Chi-squared tests for Hardy-Weinberg equilibrium

Locus	SNP	Population		
		IPF *P*	fHP *P*	Controls *P*
MUC5B	rs35705950	0.076	0.139	0.657
TOLLIP	rs3750920	0.666	0.311	0.844
	rs111521887	>0.999	0.249	>0.999
	rs5743894	0.728	0.389	>0.999
	rs5743890	>0.999	0.354	>0.999
	rs5743854	0.307	0.342	0.695

Table S3. Genotype and allele distributions of *MUC5B* and *TOLLIP* SNPs in IPF and fHP patients according to radiological UIP pattern and presence of honeycombing on HRCT

		UIP or probable UIP (n=91)			**Indeterminate for UIP (n=36)**			**Honeycombing (n=90)**			**No honeycombing (n=87)**		
Genotype / Allele		IPF (n=63) n (%)	fHP (n=28) n (%)	*P* value	IPF (n=15) n (%)	fHP (n=21) n (%)	*P* value	IPF (n=53) n (%)	fHP (n=37) n (%)	*P* value	IPF (n=30) n (%)	fHP (n=57) n (%)	*P* value
MUC5B	GG	19 (30.2%)	5 (17.9%)	0.219	4 (26.7%)	8 (38.1%)	0.473	18 (34.0%)	13 (35.1%)	0.908	5 (16.7%)	15 (26.3%)	0.309
rs35705950	GT_TT	44 (69.8%)	23 (82.1%)		11 (73.3%)	13 (61.9%)		35 (66.0%)	24 (64.9%)		25 (83.3%)	42 (73.7%)	
	G	75 (59.5%)	31 (55.4%)	0.599	16 (53.3%)	26 (61.9%)	0.467	65 (62.5%)	46 (62.2%)	0.963	30 (48.4%)	64 (56.1%)	0.325
	T	51 (40.5%)	25 (44.6%)		14 (46.7%)	16 (38.1%)		39 (37.5%)	28 (37.8%)		32 (51.6%)	50 (43.9%)	
TOLLIP	CC	16 (25.4%)	4 (14.3%)	0.237	4 (26.7%)	5 (23.8%)	>0.999	13 (24.5%)	8 (21.6%)	0.748	7 (23.3%)	10 (17.5%)	0.517
rs3750920	CT_TT	47 (74.6%)	24 (85.7%)		11 (73.3%)	16 (76.2%)		40 (75.5%)	29 (78.4%)		23 (76.7%)	47 (82.5%)	
	C	64 (50.8%)	23 (41.1%)	0.226	17 (56.7%)	19 (45.2%)	0.339	53 (61.0%)	36 (48.6%)	0.761	31 (50.0%)	53 (46.5%)	0.656
	T	62 (49.2%)	33 (58.9%)		13 (43.3%)	23 (54.8%)		51 (49.0%)	38 (51.4%)		31 (50.0%)	61 (53.5%)	
TOLLIP	CC	43 (68.3%)	16 (57.1%)	0.306	9 (60.0%)	11 (52.4%)	0.650	37 (69.8%)	25 (67.6%)	0.821	19 (63.3%)	32 (56.1%)	0.517
rs111521887	CG_GG	20 (31.7%)	12 (42.9%)		6 (40.0%)	10 (47.6%)		16 (30.2%)	12 (32.4%)		11 (36.7%)	25 (43.9%)	
	C	105 (83.3%)	42 (75.0%)	0.188	23 (76.7%)	31 (73.8%)	0.783	88 (84.6%)	60 (81.1%)	0.535	48 (77.4%)	84 (73.7%)	0.585
	G	21 (16.7%)	14 (25.0%)		7 (23.3%)	11 (26.2%)		16 (15.4%)	14 (18.9%)		14 (22.6%)	30 (26.3%)	
TOLLIP	TT	44 (69.8%)	16 (57.1%)	0.238	9 (60.0%)	11 (52.4%)	0.650	37 (69.8%)	24 (64.9%)	0.621	20 (66.7%)	32 (56.1%)	0.341
rs5743894	CT_CC	19 (30.2%)	12 (42.9%)		6 (40.0%)	10 (47.6%)		16 (30.2%)	13 (35.1%)		10 (33.3%)	25 (43.9%)	
	C	20 (15.9%)	14 (25.0%)	0.145	6 (20.0%)	11 (26.2%)	0.542	16 (15.4%)	15 (20.3%)	0.397	12 (19.4%)	30 (26.3%)	0.301
	T	106 (84.1%)	42 (75.0%)		24 (80.0%)	31 (73.8%)		88 (84.6%)	59 (79.7%)		50 (80.6%)	84 (73.7%)	

Values are expressed in number (%).
HRCT, high-resolution computed tomography; UIP, usual interstitial pneumonia

Table S4. Haplotype frequencies of the M950-T920 -T887-T894 *MUC5B-TOLLIP* haplotype block (both diseases versus controls)

Haplotype	Overall Freq.	Patients Freq.	Control Freq.	Chi Square	*P*
GCCT	0.433	0.374	0.530	13.909	$2x10^{-4}$
GTCT	0.194	0.161	0.250	7.084	$7.8x10^{-4}$
TTCT	0.122	0.163	0.054	15.534	$8x10^{-5}$
TTGC	0.101	0.145	0.030	20.291	$6x10^{-6}$
TCCT	0.068	0.090	0.032	7.317	0.007
GTGC	0.047	0.037	0.064	2.263	0.137
GCGC*	0.019	0.011	0.031	3.015	0.083
TCGC*	0.010	0.012	0.005	2.751	0.386
GTGT*	0.003	0.003	0.004	0.122	0.727
TTCC*	0.002	0.003	0.000	0.614	0.433
TTGT*	0.002	0.003	0.000	0.613	0.434

*Rare haplotypes (≤ 7 occurrences)

Table S5. Measures of pairwise linkage disequilibrium between *MUC5B* rs35705950 and *TOLLIP* rs3750920, rs111521887 and rs5743894 SNPs in patients and controls

SNP A	SNP B	D'	CI 95%	R^2	Dist bp	χ^2	*P*
IPF patients							
rs35705950	**rs3750920**	0.512	0.344 - 0.679	0.193	68735	32.582	<0.0001
rs35705950	**rs111521887**	0.695	0.479 - 0.910	0.155	71485	25.918	<0.0001
rs35705950	**rs5743894**	0.751	0.544 - 0.958	0.168	83551	30.868	<0.0001
rs3750920	**rs111521887**	1	1.000 - 1.000	0.237	2750	40.835	<0.0001
rs3750920	**rs5743894**	1	1.000 - 1.000	0.220	14816	37.816	<0.0001
rs111521887	**rs5743894**	1	1.000 - 1.000	0.926	12066	159.284	<0.0001
fHP patients							
rs35705950	**rs3750920**	0.503	0.334 - 0.672	0.157	68735	29.197	<0.0001
rs35705950	**rs111521887**	0.573	0.372 - 0.774	0.140	71485	27.700	<0.0001
rs35705950	**rs5743894**	0.583	0.386 - 0.780	0.150	83551	29.452	<0.0001
rs3750920	**rs111521887**	0.631	0.409 - 0.854	0.106	2750	19.515	<0.0001
rs3750920	**rs5743894**	0.636	0.417 - 0.855	0.110	14816	20.651	<0.0001
rs111521887	**rs5743894**	1	1.000 - 1.000	0.971	12066	188.424	<0.0001
Controls							
rs35705950	**rs3750920**	0.428	0.143 - 0.712	0.037	68735	8.962	0.0028
rs35705950	**rs111521887**	0.131	-0.042 - 0.304	0.015	71485	4.158	0.0415
rs35705950	**rs5743894**	0.141	-0.032 - 0.314	0.018	83551	4.589	0.0322
rs3750920	**rs111521887**	0.471	0.207 - 0.735	0.051	2750	12.817	0.0003
rs3750920	**rs5743894**	0.457	0.186 - 0.727	0.046	14816	11.486	0.0007
rs111521887	**rs5743894**	1	1.000 - 1.000	0.962	12066	215.423	<0.0001

D', D prime between two SNPs; R^2, correlation coefficient between two SNPs; Dist bp, distance between SNPs in base pairs

Table S6. Haplotype frequencies of the M950-T920 -T887-T894 *MUC5B-TOLLIP* haplotype block (IPF versus fHP patients)

Haplotype	Overall Freq.	fHP Freq.	IPF Freq.	Chi Square	*P*
GCCT	0.380	0.371	0.391	0.160	0.689
TTCT	0.162	0.157	0.168	0.076	0.783
GTCT	0.158	0.168	0.148	0.293	0.588
TTGC	0.147	0.150	0.144	0.031	0.861
TCCT	0.086	0.072	0.102	1.028	0.311
GTGC	0.038	0.045	0.029	0.616	0.433
TCGC*	0.014	0.022	0.005	1.970	0.160
GCGC*	0.006	0.009	0.002	0.793	0.373
TTCC*	0.003	0.005	<0.001	0.884	0.347
GTGT*	0.003	<0.001	0.006	1.129	0.288
TTGT*	0.003	<0.001	0.006	1.129	0.288

*Rare haplotypes (≤ 3 occurrences)

Fig. S1 Kaplan-Meier plots showing IPF patients' survival according to *TOLLIP* SNPs

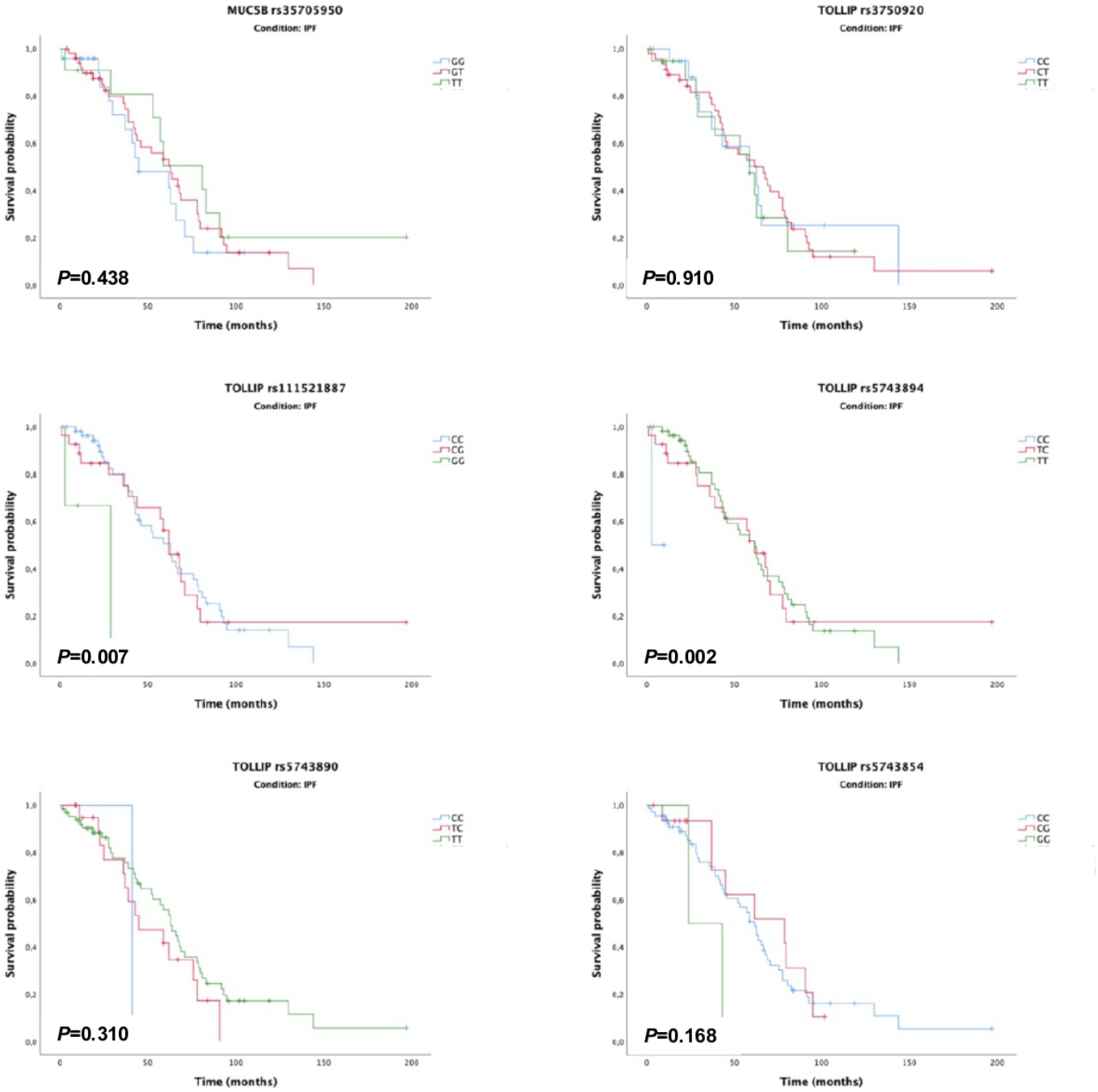
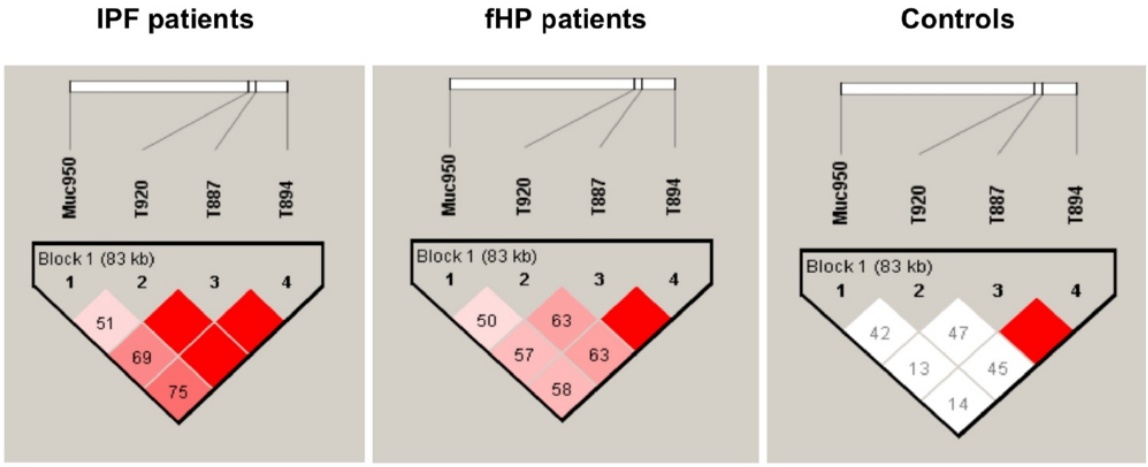


Fig. S2 Linkage disequilibrium (LD) plot of the 83-kb M950-T920-T887-T894 block on chromosome 11p15.5



The plot was drawn using Haploview software v4.2.
D' values are shown inside the squares. Empty squares have a D' =1 (100)
Red boxes represent D' =1 (LOD ≥2). Shades of pink/red represent D' <1 and LOD ≥2.
White boxes represent D' <1 and LOD <2.
LOD, log of likelihood odds ratio.

Supplementary references

1. Raghu G, Remy-Jardin M, Richeldi L, et al. Idiopathic Pulmonary Fibrosis (an Update) and Progressive Pulmonary Fibrosis in Adults: An Official ATS/ERS/JRS/ALAT Clinical Practice Guideline. *Am J Respir Crit Care Med* 2022; **205**(9): e18-e47.
2. Hansell DM, Bankier AA, MacMahon H, McLoud TC, Muller NL, Remy J. Fleischner Society: glossary of terms for thoracic imaging. *Radiology* 2008; **246**(3): 697-722.
3. Bankhead P, Loughrey MB, Fernandez JA, et al. QuPath: Open source software for digital pathology image analysis. *Sci Rep* 2017; **7**(1): 16878.
4. Wright JL, Churg A, Hague CJ, Wong A, Ryerson CJ. Pathologic separation of idiopathic pulmonary fibrosis from fibrotic hypersensitivity pneumonitis. *Mod Pathol* 2020; **33**(4): 616-25.
5. Raghu G, Remy-Jardin M, Ryerson CJ, et al. Diagnosis of Hypersensitivity Pneumonitis in Adults. An Official ATS/JRS/ALAT Clinical Practice Guideline. *Am J Respir Crit Care Med* 2020; **202**(3): e36-e69.

DISCUSSION

IPF and fHP are the two most prominent fILDs of our national routine ILD practice. Despite distinct etiologies, these diseases often share clinical, radiographic, and histopathologic features, posing challenges in differential diagnosis despite their varied risk factors and pathogenesis (39). Genetic factors have been implicated in the pathogenesis and progression of both diseases, as evidenced by familial and sporadic occurrences (63, 71, 115). The parallel phenotypic outcome of diffuse lung fibrosis through different pathways requires a precise differential diagnosis, particularly in cases demonstrating a radiological UIP pattern, with distinct therapeutic and prognostic implications.

Existing data on the genetics of sporadic forms of both diseases predominantly focuses on the individual study of common variants, particularly SNPs in the *MUC5B* and *TOLLIP* loci (63, 115). However, a comprehensive exploration of the impact of wider genomic regions defined by these variants, as haplotypes, has not been commonly undertaken in the current literature. The use of SNP-defined haplotypes within specific regions of interest previously identified through single-SNP GWS holds immense potential. This approach can be instrumental in unveiling and refining genetic associations, and in offering deeper insights into the influence of genetic loci than the study of SNPs individually (173-175).

The objective of this thesis was to conduct a comprehensive analysis using a three-step strategy to examine the independent impact of *MUC5B* and neighboring *TOLLIP* SNPs, as well as that of their respective haplotypes, in the context of comparisons among IPF and fHP patients versus healthy controls, respectively, and with direct comparisons between patients with both conditions. The primary goal was to identify potential associations pinpointing genetic features that would enable the discrimination between IPF and fHP, or the identification of similarities in genetic mechanisms promoting both diseases.

Case-control studies were designed and included patients followed in the Pulmonology Department of Local Health Unit of São João in Porto. Diagnosis of IPF and fHP was established according to current guidelines (2, 8, 34, 38, 39, 42, 176). Control individuals

were obtained from the EPIPorto and NETDiamond cohorts (177-179), as described in the second and third studies.

The main findings can be summarized as follows:

1. In patients with IPF, the *MUC5B* rs35705950 SNP displayed a significant association with the disease. T allele carriers were notably more prevalent among IPF patients than among controls (75.0% vs 20.3%, $P<1.0x10^{-6}$). This result was largely confirmatory of numerous studies conducted previously in ethnically diverse populations (80, 150, 151, 153, 154). While the genotypic and allelic distributions of *TOLLIP*, plasminogen activator urokinase (*PLAU)*, and serpin family B member 1 (*SERPINB1)* SNPs did not exhibit significant differences between the IPF and controls, an important finding emerged regarding a *MUC5B*-*TOLLIP* haplotype block. Defined by *MUC5B* rs35705950 and *TOLLIP* rs111521887, rs5743894, and rs5743854 SNPs, the T-C-T-C haplotype was identified as an independent protective factor in IPF survival (HR = 0.37, 95% CI 0.17–0.78, *P*=0.009).
2. In patients with fHP, *MUC5B* rs35705950 and three neighboring *TOLLIP* variants rs3750920, rs111521887, and rs5743894, were associated with susceptibility to the disease. MAFs were markedly higher among fHP patients when compared to controls (40.7% vs 12.1%, *P*<0.0001; 52.6% vs 40.2%, *P*=0.011; 22.7% vs 13.4%, *P*=0.013; and 23.2% vs 12.9%, *P*=0.006, respectively). Likewise, haplotypes formed by these variants were linked to increased susceptibility to fHP. This constituted the first report associating *TOLLIP* SNPs to susceptibly in fHP. Notably, carriers of a specific haplotype, G-T-G-C, displayed a significant decrease in survival (HR 6.92, 95% CI 1.73–27.64, *P*=0.006), emerging as an independent detrimental factor in fHP prognosis. Associations were also observed between *TOLLIP* rs111521887 and rs5743894 variants and reduced lung function at baseline, while the *MUC5B* SNP showed significant correlations with specific radiographic features.
3. When comparing IPF and fHP, *MUC5B*, as already expected, but also *TOLLIP* SNPs were not discriminatory between both diseases, including for the UIP pattern or

presence of honeycombing in HRCT. Haplotypes defined by the same *MUC5B-TOLLIP* block, rs35705950-rs3750920-rs111521887-rs5743894, showed associations with susceptibility to disease when comparing all patients to controls. Importantly, a haplotype was identified, G-C-C-T, that conferred significantly better survival to fHP carriers than to IPF carriers (median survival: IPF = 59 months vs fHP = 98 months; *P*=0.006).

These findings underscore a substantial influence of genetic factors in the onset and progression of IPF and fHP.

In the subsequent sections, a comprehensive and detailed discussion of the results obtained from each phase of the investigation is presented. This discussion aims to provide in-depth insights and analysis of the outcomes derived from the three distinct studies integrating the thesis. A summary of the main results is presented in **Table 4**.

FIRST STUDY

The primary aim of this initial investigation was to ascertain the potential association between specific SNPs in *MUC5B* and *TOLLIP* loci, never before studied in the context of a Portuguese population, and IPF susceptibility and survival, as well as to probe an involvement of two new candidate genes - *PLAU* and *SERPINB1* - in this disease.

In total, ten SNPs were subjected to analysis: *MUC5B* rs35705950, *TOLLIP* rs5743854, rs5743890, rs5743894, rs111521887 and rs3750920, *PLAU* rs2227564, and *SERPIN* rs408265, rs386713 and rs57721448. A case-control study was conducted with 64 patients and 74 controls.

The analysis revealed an evident correlation between the *MUC5B* rs35705950 SNP and IPF in this Portuguese population. The prevalence of the minor allele T was markedly higher in IPF patients than in control subjects (45.3% vs 11.5%, OR=6.38, 95% CI 3.46–11.79, $P<1.0\text{x}10^{-6}$). Correspondingly, heterozygous and homozygous individuals carrying the T allele were significantly more prevalent in the IPF group (75% vs 20.3%, OR=11.80, 95% CI 5.30–26.28, $P<1.0\text{x}10^{-6}$). These findings are consistent with observations in other Caucasian populations, emphasizing notably higher frequencies of the minor allele T

among IPF patients (68, 78, 147, 154, 156, 165). A recent meta-analysis, incorporating data from 24 case-control studies involving 6,749 IPF patients and 13,898 healthy controls, corroborated the notion that the presence of the T allele significantly increases the risk of developing IPF, with a stronger association observed among Caucasians compared to Asians (158). The increased risk was evident among homozygotes for the T allele, as illustrated by a substantial difference between the OR values of homozygotes versus heterozygotes (GT genotype), which was similar in magnitude to that reported in our study (TT genotype OR=18.44 versus GT genotype OR=10.78 (158, 180).

The *MUC5B* rs35705950 T allele predisposes individuals to IPF but has also been associated with improved survival in some groups (70, 79, 145, 163, 181). The lower bacterial burden and slower functional decline reported in carriers of this allele could serve as potential explanations (80, 167). This prognostic benefit was further corroborated by a recent study by Biondini et al. (2021), which focused specifically on IPF patients undergoing antifibrotic therapy. This study highlighted a difference in overall survival of 42 months for the GG genotype versus 74 months for TT/GT genotypes (182). Similarly, Van der Vis et al. (2023) reported that in IPF patients aged over 56 years, transplant-free survival was significantly better for *MUC5B* minor allele carriers (45 months) compared to non-carriers (29 months) (163). However, these observations are not supported by our study, nor by others that have investigated IPF patients, including across different ethnicities (153, 154, 165). The *MUC5B* rs35705950 T allele is associated with the overproduction and accumulation of mucin in distal airspaces, which can lead to dysfunction. This dysfunction promotes the retention of inhaled substances such as air pollutants, cigarette smoke, microorganisms, and endogenous inflammatory debris. Over time, this retention may result in temporally and spatially distinct areas of microscopic scarring and progressive fibroproliferation in the lung (77). Cui et al. (2023) showed that patients with high genetic susceptibility to IPF (including the *MUC5B* promoter polymorphism) living in areas of heavy air pollution from nitrogen oxide and particulate matter <2.5 μm in aerodynamic diameter ($PM_{2.5}$) had a threefold risk of developing IPF compared with those with low genetic susceptibility and low air pollution exposure (86). Despite this synergistic effect described between $PM_{2.5}$ and *MUC5B* genotype on IPF incidence, Sese et al. (2024) did

not observe any interaction between *MUC5B* genotype and $PM_{2.5}$ exposure on IPF survival (183). Thus, this association remains controversial, and the influence of other genetic, epigenetic, and environmental factors could regulate *MUC5B* expression, contributing to different prognostic scenarios.

Conflicting findings regarding risk and survival have also emerged in what concerns the *TOLLIP* SNPs. Specifically, the minor alleles of intronic SNPs rs111521887 and rs5743894 were purported to increase susceptibility to IPF, and the rs5743890 SNP to reduce such risk (64, 65). In our study, none of the five *TOLLIP* SNPs, on their own, showed any association with IPF risk, which could indicate a potential distinct feature in Portuguese patients. However, this absence of association had already been noted in other studies (68, 165). Peljto et al. (2023) reported an initial significant association between a specific region within the *TOLLIP* gene (chr11:1284193–1287389) and IPF (P<0.0001), which became non-significant after adjusting for the *MUC5B* polymorphism (P=0.79), suggesting a potential interplay or confounding effect between the two genetic loci in relation to IPF (68).

Regarding survival, we observed a potential association between the minor alleles of *TOLLIP* rs111521887 (G), rs5743894 (C), and rs5743854 (G) in homozygosity and increased mortality among IPF patients. Multivariate analyses showed that only the rs5743854 association remained statistically significant, with the GG genotype associating with worst survival, compared to the CC genotype (HR=9.27, P=0.006). Importantly, this association remained independent of FVC, as better pulmonary function was linked with improved survival in univariate analysis. Our findings did not corroborate associations reported in the literature, which were solely related to *TOLLIP* rs5743890 (associated with increased mortality in carriers of the minor allele C) and rs3750920 (associated with decreased mortality, reduced hospitalization, and FVC decline in carriers of the TT genotype treated with NAC) (64, 165, 166). More recently, Whalen et al. (2024) also reported a positive interaction between *TOLLIP* rs3750920 (genotype TT) and doxycycline in IPF patients, highlighting a survival benefit (184). This finding supports the potential of *TOLLIP* genotyping for targeted treatment in IPF and underscores the need for pharmacogenomics studies in patients with fILDs that involve a broader spectrum of genetic variants.

The collective impact of SNPs of neighboring *MUC5B* and *TOLLIP* was investigated by analyzing haplotypes formed by the block encompassing M950-T887-T894-T854. Five frequent haplotypes were identified, of which two were significantly more frequent in patients (T-C-T-C and T-G-C-C) and one in control individuals (G-C-T-C), revealing haplotypic associations with IPF risk. Notably, carriers of the T-C-T-C haplotype, which incorporated the *MUC5B* promoter minor allele (T) and the wild-type (ancestral) alleles of the three *TOLLIP* SNPs, was significantly associated longer survival when compared to non-carriers. Accordingly, the minor alleles of those three *TOLLIP* SNPs, which were absent from this protective haplotype, were individually associated with a worse prognosis. The protective effect of this haplotype persisted significantly even after adjusting for FVC and treatment with AF as a time-dependent covariate (HR=0.37, 95% CI 0.17–0.78, *P*=0.009). These findings suggest that the region defined by this haplotype block contains or segregates with a genetic factor as-yet unidentified that influences the clinical course of disease. Additionally, our data suggests that it may also influence the response to antifibrotic therapy in these patients, but a larger sample is required for validation and such potential effect warrants further investigation.

Our first study also explored novel candidate genes - *SERPINB1* and *PLAU* - that have a significant presence and play roles in lung fibrosis (185, 186). While these genes exhibited altered expression in lung tissues of IPF patients, the variants surveyed in this study did not associate with IPF susceptibility or prognosis.

Complex traits are believed to result from variation within multiple genes and their interaction, as well as from interactions with environmental factors, therefore not following readily predictable patterns of inheritance. In an attempt to investigate whether multi-locus interactions, rather than individual (or local) effects of the probed variants could underlie IPF susceptibility, we have also employed a multifactor dimensionality reduction (MDR) analysis (187, 188), a strategy that has been successfully employed to detect epistatic interactions among unlinked loci in several other pathologies classified as complex traits (187). We found no predictive multi-locus interactions of IPF susceptibility among the 4 loci. It is noteworthy, however, that the modest sample size of our cohorts may have limited MDR's ability to detect gene-gene interactions.

In summary, this first study corroborated the association of the *MUC5B* rs35705950 minor allele with an increased risk of IPF in the Portuguese population and revealed a robust correlation between a *MUC5B*-*TOLLIP* SNP-defined haplotype and disease survival. Despite the small sample size, we suggest that this haplotype holds promise as a prognostic marker in stratifying IPF patients. However, its definitive utility would benefit from subsequent validation through larger, multicenter, and prospective studies.

SECOND STUDY

The focus of the second study was to explore potential associations of *MUC5B* and *TOLLIP* SNPS (and respective haplotypes) with susceptibility, clinical presentation, and survival in Portuguese patients diagnosed with fHP. Six SNPs were analyzed: *MUC5B* rs35705950 and *TOLLIP* rs3750920, rs111521887, rs5743894, rs5743890, and rs5743854. A case-control study was conducted with 97 fHP patients and 112 control individuals.

In this cohort of Portuguese patients, the *MUC5B* promoter variant rs35705950 and *TOLLIP* SNPs rs3750920, rs111521887, and rs5743894 were associated with susceptibility to fHP. While several published reports had previously shown an association of the *MUC5B* promoter SNP with susceptibility to fHP (81, 189, 190), our study was the first to implicate *TOLLIP* variants as risk determinants of the disorder, namely the minor allele of the rs111521887 and rs5743894 SNPs, which were previously associated with IPF risk in the literature (64), but not confirmed in Portuguese patients as reported in the first study (180).

The frequency of the minor allele T of the *MUC5B* rs35705950 SNP was significantly increased among fHP patients compared to controls, reinforcing the significance of the *MUC5B* promoter SNP as a predictive risk factor across various forms of fILD, including in Portuguese patients with fHP and IPF (141-143, 180, 191). Furusawa et al. (2022) conducted a case–control study involving 226 HP patients and 1355 controls to examine the distribution of 10 IPF-associated polymorphisms. Six common variants were significantly associated with HP, with *MUC5B* rs35705950 showing the strongest

association with both HP overall (OR=2.11, P=1.74×10^{-6}) and fHP specifically (OR=2.32, P=2.61×10^{-6}) (189). These observations have led to the conclusion that excess *MUC5B* is in itself fibrogenic, predisposing individuals to the development of progressive fibrosing lung disease, although the mechanisms behind it are unclear.

As mentioned, *TOLLIP* rs3750920, rs111521887, and rs5743894 SNPs showed significant associations with fHP. In contrast, none of the *TOLLIP* polymorphisms revealed an independent association with IPF risk in our previous study, suggesting a difference between *MUC5B* and *TOLLIP* in their significance as markers for distinguishing between the two fibrotic diseases, particularly in the Portuguese population. While the proximity of *MUC5B* could theoretically explain the association with *TOLLIP*, this was not the case in our initial IPF study, where a similar risk exists for *MUC5B* but not for *TOLLIP* (180).

On the other hand, it's interesting to note that the survival analyses conducted in fHP patients, did not reveal any association between genotypes of *MUC5B* or *TOLLIP* and survival outcomes. These results are in accordance with the report by Ley et al. (2017), who also found no independent association between *MUC5B* rs35705950 or *TOLLIP* rs5743890 and survival in patients with fHP (81). This instance contrasts with the results obtained in the context of IPF (first study) where *TOLLIP* rs5743854 had shown an independent association with poorer survival in IPF patients (180). Taken together, these results point to the prospect of using *TOLLIP* as an informative marker in future strategies for differential diagnosis between the two conditions.

Analyses of haplotypes defined by a block encompassing the *MUC5B* promoter SNP and the three adjacent *TOLLIP* SNPs, all of which had shown an association with fHP susceptibility when analyzed independently, revealed an association between the G-T-G-C haplotype and survival in fHP. Carriers of this haplotype exhibited a shorter overall survival, as well as a shorter transplant-free survival. This effect was independent, as shown by multivariate adjustment for age and lung function parameters (overall survival, HR=6.92, P=0.006; transplant-free survival, HR=5.93, P=0.009).

The methodology used for analyzing haplotypes in this study mirrors the approach employed in the first study that enabled the identification of an association with survival in IPF patients, although in opposite directions. The identification of haplotypes associated with survival across different studies reinforce the potential functional

significance of the studied variants in diverse fILDs. As previously stated, the region encompassing these variants seems to contain or be linked to (an) unidentified genetic factor(s) that can influence the clinical risk and progression of both disorders.

The evaluation of potential interactions between genetic (alleles, genotypes and haplotypes) and clinical, radiological, and pathological features was conducted alongside susceptibility and survival analyses. This approach aimed to explore how genetic variation might influence the initial presentation of the disease in this cohort of patients. Associations were found between *TOLLIP* rs111521887 and rs5743894 variants and decreased lung function at baseline, and the *MUC5B* SNP and radiographic features, further highlighting the influence of genetic factors in fHP. BAL samples analysis also revealed a tendency towards a higher count of neutrophils in patients carrying at least one minor allele of the *TOLLIP* rs111521887 (G) and rs5743894 (C) SNPs. Conversely, for the *MUC5B* rs35705950, this trend was observed but with the ancestral allele (G). Notably, carriers of the G-T-G-C haplotype, linked with a poorer prognosis, tended to exhibit a median neutrophil count that more than doubled that of non-carriers. De Sadeleer et al. (2020) suggested that low BAL lymphocytosis (< 20%) predicts poor fHP outcomes and indicates a lack of response to corticosteroid treatment (192). In our sample, the median lymphocyte count was below 20%, but values displayed a wide range (from 1.6% to 51.0%) and it was not significantly different among alleles, genotypes, or the G-T-G-C haplotype. The observed disparity in neutrophil counts, particularly when lymphocyte counts remain within the normal range, might represent a subgroup of patients potentially resembling IPF, wherein immunosuppression might not be advisable.

In summary, the findings of the second study implicate *TOLLIP* and *MUC5B* variants and respective haplotypes in risk and survival in fHP, suggesting they may be invaluable tools in the context of a wider strategy for risk evaluation and prognosis in fHP. Moreover, these variants, mainly in *TOLLIP* genes, might contribute considerably to patient stratification, and provide useful insights into the genetic mechanisms influencing the clinical trajectory of the condition. Nonetheless, once more, validation through

multicenter and prospective studies is necessary to confirm these results and gain a deeper understanding of the roles played by both genes in the pathophysiology of fHP.

THIRD STUDY

In the two previous studies, the influence of *MUC5B* and *TOLLIP* variants in IPF and fHP risk and survival were individually addressed. However, there has been a lack of direct comparisons between the two conditions regarding the genetic characterization of these and other variants. Some studies have looked at fILDs in general, grouping various diseases, but few have directly compared IPF and fHP, the two most common fILDs in our routine clinical practice (193-195). Abbasi et al. (2022) took a step in this direction by evaluating genetic variants related to surfactant proteins. They discovered that IPF and fHP share some SRGs variants, suggesting common underlying mechanisms in surfactant production, but there are also differences (196). These highlight distinct pathogenic mechanisms between an inflammation-driven fibrosis (fHP) and an epithelial-driven fibrosis (IPF) (196). Additionally, it's worth noting that these fILDs can present a similar radiological UIP pattern, which has prognostic implications. Exploring the interaction between genetic markers and radiological phenotype can aid or not in distinguishing between these two diseases (197, 198).

The primary objective of a third study was to directly compare the behavior of *MUC5B* and neighboring *TOLLIP* SNPs, along with their respective haplotypes, in IPF and fHP to identify associations that could either distinguish the conditions or unveil common pathways. The same set of SNPs was analyzed: *MUC5B* rs35705950 and *TOLLIP* rs3750920, rs111521887, rs5743894, rs5743890, and rs5743854. A case-control study was conducted with 86 IPF and 97 fHP patients, alongside 112 controls.

In general, as expected, notable differences were observed between the clinical, radiographic, histopathological features, treatment approaches, and outcomes of patients with IPF and fHP. IPF patients were predominantly male, older, had a higher prevalence of smoking history and presented more comorbidities (such as pulmonary hypertension and metabolic conditions) compared to those diagnosed with fHP. Key pulmonary function metrics revealed significantly lower DLCO in IPF patients, while FVC

and TLC were similar between the two groups. As expected, IPF patients displayed a higher incidence of UIP pattern (definite and probable) and honeycombing. Histological confirmation was more frequent in fHP cases, showing fewer instances of honeycombing and fibroblast foci but indicating a higher presence of granulomas. Treatment strategies varied, with a majority of IPF patients initially receiving antifibrotic therapy, while most fHP patients were initially treated with immunosuppressants, later supplemented with antifibrotics in cases of progressive fibrosis. The mortality rate for IPF was three times higher than that for fHP. However, the follow-up time differed significantly between the two diseases, with a longer follow-up period for IPF. This discrepancy in follow-up time can be considered a limitation of the study.

This work provides further evidence supporting the association of the minor allele of *MUC5B* rs35705950 with IPF and fHP when comparing each condition with the control population. Among them, no significant differences were observed in allelic or genotypic distributions of this variant, as anticipated, constituting a shared genetic risk factor in Portuguese patients, rather than a discriminatory one. The older age of carriers of the rs35705950 minor allele, regardless of the fILD, had already been described in literature as a common feature of IPF and fHP (163, 190, 199). The absence of an association of *MUC5B* with survival in both diseases was confirmed.

TOLLIP SNPs did not exhibit a significant association with the risk of IPF, as demonstrated in the first study, even with an increased number of patients and controls (180). Similarly, the associations identified between *TOLLIP* rs3750920, rs111521887, and rs5743894 SNPs and the risk of fHP persisted (191). Notwithstanding, when comparing the two diseases, no significant differences were seen in allelic or genotypic distributions of *TOLLIP* SNPs. These observations suggest that the *TOLLIP* SNPs are useful to provide insights into the risk of fHP when compared to controls, but not when compared with IPF. This lack of differentiation between IPF and fHP persisted when the patients' radiological UIP pattern or the presence of honeycombing on HRCT were taken into consideration, suggesting that IPF and fHP patients with a UIP pattern (definite or probable) may be genetically indistinguishable. From a pathogenic perspective, one might speculate that there is a singular disease entity that should be approached in a unified manner. This aligns with the viewpoint of certain authors advocating for the

consolidation of UIP as a single diagnostic entity, highlighting its undeniable advantages (46). Moreover, Furusawa et al. (2020), despite some differences, also reported significant similarities in gene expression profiles between IPF and fHP, mainly in those genes associated with collagen catabolic processes and epithelial development that were shared between the two conditions (190).

Our results emphasize the significance of examining neighboring genetic variants in a wider genomic context, namely the *TOLLIP* SNPs, through haplotype analysis to gain a better understanding of their potential in differentiating between IPF and fHP. Notably, analysis of the haplotypes concerning these genes has not been extensively explored in previous literature. One exception is the study by Bonella et al. (2021), which effectively showed that the haplotype formed by the *MUC5B*-*TOLLIP* block, rs35705950-rs3750920-rs5743890 T-T-T, was more frequent in IPF patients than in healthy controls, although no association with survival was addressed (165).

Consequently, we focused our analysis on the haplotypes examined in the second study, which include the *MUC5B* promoter variant, associated with both diseases, and the three *TOLLIP* SNPs associated with fHP risk - M*UC5B-TOLLIP* block, rs35705950-rs3750920-rs111521887-rs5743894. Significant differences in haplotype frequencies between the overall group of patients (IPF and fHP) and controls were observed, but not between the diseases, reinforcing the notion that the genomic landscape encompassing the *MUC5B* and *TOLLIP* loci is at least influential for the fibrosing phenotype. Our haplotype analyses revealed significant associations with prognosis in both IPF and fHP patients. Notably, the haplotype including the wild-type (ancestral) alleles of *MUC5B* and *TOLLIP* SNPs (G-C-C-T) showed significant differences in survival between IPF and fHP carriers, effectively differentiating the two conditions. This effect was independent of age, gender, or the presence of a radiological UIP pattern or honeycombing, as determined by multivariate analyses. A similar effect was noted for the haplotype formed solely by the ancestral alleles of the same *TOLLIP* SNPs (C-C-T, included in the block above), further reinforcing the association. Furthermore, IPF carriers of the T-C-C-T haplotype, composed of the *MUC5B* minor allele and the *TOLLIP* ancestral alleles, demonstrated increased survival compared to non-carriers. These findings align with our first study showing that the same alleles of *MUC5B* rs35705950 and *TOLLIP*

rs111521887 and rs5743894 formed a haplotype associated with improved survival in IPF. Conversely, in fHP, the G-T-G-C haplotype (comprising the *MUC5B* ancestral allele and the *TOLLIP* minor alleles) was linked to reduced survival in carriers. Although the *TOLLIP* alleles were not individually associated with survival in fHP, they were significantly more frequent in fHP patients than in controls, indicating a relevant role of *TOLLIP* in this disorder.

Altogether, and considering survival only, our observations suggest that in fHP patients, a haplotype integrating the *TOLLIP* ancestral alleles, C-C-T, is protective only when inherited with the *MUC5B* ancestral allele. In contrast, in IPF patients, survival is improved when these *TOLLIP* alleles are inherited with the *MUC5B* minor allele. This effect cannot be solely attributed to *MUC5B* rs35705950, as this SNP is not associated with survival on its own. Moreover, the C-C-T haplotype, defined by the *TOLLIP* SNPs block alone, is associated with significantly higher survival in fHP compared to IPF carriers. Therefore, it can be speculated that either the variants themselves or a linked, unidentified genetic factor within the region may influence varying clinical outcomes in specific genomic landscape conformations.

In summary, this third study highlights genetic similarities between IPF and fHP, particularly regarding the role of *MUC5B* and *TOLLIP* SNPs in disease susceptibility, even in patients presenting with a radiological UIP pattern or honeycombing. However, survival analysis underscores the potential importance of their haplotypes for prognostic stratification of these diseases. Notably, certain haplotypes showed significant differences in survival outcomes between IPF and fHP patients, suggesting their utility in prognostic assessment. Despite these promising findings, further validation through larger, multicenter, and prospective studies are crucial to confirm these results and elucidate the practical utility of *TOLLIP* and *MUC5B* haplotypes as markers for distinguishing between IPF and fHP in terms of survival. These insights could profoundly impact diagnostic and therapeutic decision-making for fILDs.

Gene	SNP (minor allele)	IPF vs CT	fHP vs CT	IPF vs fHP	Literature
MUC5B	rs35705950 (T)	↑ risk ∅ survival	↑ risk ∅ survival	∅ risk and survival	↑ risk for both diseases (78, 81, 158, 165, 189, 190) ↑, ↓ or ∅ IPF survival (79, 153, 165) ↓ fHP survival (81)*
TOLLIP	rs3750920 (T)	∅ risk	↑ risk ∅ survival ↓ FVC*	∅ risk and survival	Genotype TT favorable response to NAC in IPF (166)
	rs111521887 (G)	∅ risk ↓ survival (GG)*	↑ risk ∅ survival ↓ FVC BAL neutrophilia*	∅ risk and survival	↑ IPF risk (64)
	rs5743894 (C)	∅ risk ↓ survival (CC)*	↑ risk ∅ survival ↓ FVC BAL neutrophilia*	∅ risk and survival	↑ IPF risk (64)
	rs5743890 (C)	∅ risk and survival	∅ risk and survival	∅ risk and survival	↓ IPF risk (64) ↓ IPF survival (64, 165) ∅ fHP risk and survival (81)
	rs5743854 (G)	∅ risk ↓ survival (GG)	∅ risk and survival	∅ risk and survival	
Haplotypes					
*M950-T887-T894-T854*_TCTC		↑ survival			
*M950-T920-T887-T894*_TCCT		↑ survival		∅ survival	
*M950-T920-T887-T894*_GTGC			↓ survival Birds* BAL neutrophilia*	∅ survival	
*M950-T920-T887-T894*_GCCT				↑ fHP survival	

Table 4. Summary of the results highlighting the influence of *MUC5B* and *TOLLIP* SNPs on IPF and fHP in Portuguese patients. IPF, idiopathic pulmonary fibrosis; fHP, fibrotic hypersensitivity pneumonitis; CT, controls; SNP, single nucleotide polymorphism. M950-T887-T894-T854, rs35705950-rs111521887-rs5743894-rs5743854 *MUC5B-TOLLIP* haplotype block; M950-T920-T887-T894, rs35705950-rs3750920-rs111521887-rs5743894 *MUC5B-TOLLIP* haplotype block. *not reaching statistical significance but trending towards it.

CONCLUSIONS AND FUTURE PERSPECTIVES

The impact of genetics on susceptibility and progression of sporadic forms of fILDs is a crucial area of research. Common genetic variants, particularly in genes such as *MUC5B* and *TOLLIP*, have been linked to increased susceptibility in conditions like IPF and fHP. Understanding how these genetic markers behave in different disease phenotypes not only aids in accurate diagnosis but also provides valuable insights into prognosis and facilitates a personalized approach to treatment.

The initial individual analyses of *MUC5B* and *TOLLIP* SNPs did not provide a direct differentiation between IPF and fHP. However, a specific haplotype containing the ancestral alleles of *MUC5B* and *TOLLIP* SNPs, for instance, was associated with prolonged survival in fHP when compared to IPF. This emphasizes the importance of analyzing haplotypes instead of individual SNPs, providing a deeper understanding of genetic variation and its impact on diseases prognosis.

Overall, our findings highlight the significant role of *MUC5B* rs35705950 as a shared genetic risk factor in Portuguese patients with both IPF and fHP. The *TOLLIP* SNPs provide information pertaining the individual risk of fHP when compared to controls but do not demonstrate significant differences when compared to IPF. The impact of *MUC5B* and *TOLLIP* haplotypes is substantial, influencing both the risk and survival of patients with IPF and fHP, particularly in survival comparative analyses. Our studies underscore the potential utility of these genetic markers in facilitating the prognostic stratification of the two diseases. The absence of distinct genetic features between patients with IPF and fHP exhibiting a UIP pattern supports the notion that they represent a unified entity, and suggests that additional factors, such as TRGs variants and TL, for example, should be integrated into a wider strategy to ascertain the best pharmacological approach.

It is important to acknowledge that while this research provides valuable insights, further validation and replication of these findings are imperative to establish the robustness and reliability of the identified genetic associations. Additionally, a deeper understanding of the complex polygenic nature of these diseases is crucial, as they may involve yet-to-be-identified variants.

The integration of genetic tests into routine clinical practice should be promoted, not only for suspected cases of FPF but also for sporadic forms of fILDs. The information derived from these tests can guide diagnosis, potentially obviating the need for a biopsy in specific situations, but always after integration with other diagnostic data and discussion in multidisciplinary meetings. The utility, safety, and cost-effectiveness of such an approach must be thoroughly considered.

While associations with susceptibility and survival are evident, future research should explore the impact of genetic factors on functional and radiological progression, as well as responses to treatments, as potential findings can significantly influence the initial management of patients with fILDs.

The findings described in this thesis highlight the importance of continued investigation into the genetic mechanisms of IPF and fHP for improved patient management and treatment outcomes, paving the way for personalized approaches to diagnosis and therapy in the future.

REFERENCES

1. Travis WD, Costabel U, Hansell DM, King TE, Jr., Lynch DA, Nicholson AG, et al. An official American Thoracic Society/European Respiratory Society statement: Update of the international multidisciplinary classification of the idiopathic interstitial pneumonias. Am J Respir Crit Care Med. 2013;188(6):733-48.
2. Raghu G, Remy-Jardin M, Richeldi L, Thomson CC, Inoue Y, Johkoh T, et al. Idiopathic Pulmonary Fibrosis (an Update) and Progressive Pulmonary Fibrosis in Adults: An Official ATS/ERS/JRS/ALAT Clinical Practice Guideline. Am J Respir Crit Care Med. 2022;205(9):e18-e47.
3. Wijsenbeek M, Suzuki A, Maher TM. Interstitial lung diseases. Lancet. 2022;400(10354):769-86.
4. Kaul B, Cottin V, Collard HR, Valenzuela C. Variability in Global Prevalence of Interstitial Lung Disease. Front Med (Lausanne). 2021;8:751181.
5. Hilberg O, Hoffmann-Vold AM, Smith V, Bouros D, Kilpelainen M, Guiot J, et al. Epidemiology of interstitial lung diseases and their progressive-fibrosing behaviour in six European countries. ERJ Open Res. 2022;8(1).
6. Glenn LM, Troy LK, Corte TJ. Diagnosing interstitial lung disease by multidisciplinary discussion: A review. Front Med (Lausanne). 2022;9:1017501.
7. Nogueira R, Melo N, Novais EBH, Martins N, Delgado L, Morais A, et al. Hypersensitivity pneumonitis: Antigen diversity and disease implications. Pulmonology. 2019;25(2):97-108.
8. 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.
9. Borie R, Kannengiesser C, Antoniou K, Bonella F, Crestani B, Fabre A, et al. European Respiratory Society statement on familial pulmonary fibrosis. Eur Respir J. 2023;61(3).
10. Wijsenbeek M, Cottin V. Spectrum of Fibrotic Lung Diseases. N Engl J Med. 2020;383(10):958-68.
11. Adegunsoye A, Oldham JM, Bellam SK, Montner S, Churpek MM, Noth I, et al. Computed Tomography Honeycombing Identifies a Progressive Fibrotic

Phenotype with Increased Mortality across Diverse Interstitial Lung Diseases. Ann Am Thorac Soc. 2019;16(5):580-8.

12. Patolia S, Tamae Kakazu M, Chami HA, Chua A, Diaz-Mendoza J, Duggal A, et al. Bronchoalveolar Lavage Lymphocytes in the Diagnosis of Hypersensitivity Pneumonitis among Patients with Interstitial Lung Disease. Ann Am Thorac Soc. 2020;17(11):1455-67.
13. Meyer KC, Raghu G. Bronchoalveolar lavage for the evaluation of interstitial lung disease: is it clinically useful? Eur Respir J. 2011;38(4):761-9.
14. Hutchinson JP, McKeever TM, Fogarty AW, Navaratnam V, Hubbard RB. Surgical lung biopsy for the diagnosis of interstitial lung disease in England: 1997-2008. Eur Respir J. 2016;48(5):1453-61.
15. Durheim MT, Kim S, Gulack BC, Burfeind WR, Gaissert HA, Kosinski AS, et al. Mortality and Respiratory Failure After Thoracoscopic Lung Biopsy for Interstitial Lung Disease. Ann Thorac Surg. 2017;104(2):465-70.
16. Troy LK, Grainge C, Corte TJ, Williamson JP, Vallely MP, Cooper WA, et al. Diagnostic accuracy of transbronchial lung cryobiopsy for interstitial lung disease diagnosis (COLDICE): a prospective, comparative study. Lancet Respir Med. 2020;8(2):171-81.
17. Maldonado F, Danoff SK, Wells AU, Colby TV, Ryu JH, Liberman M, et al. Transbronchial Cryobiopsy for the Diagnosis of Interstitial Lung Diseases: CHEST Guideline and Expert Panel Report. Chest. 2020;157(4):1030-42.
18. Korevaar DA, Colella S, Fally M, Camuset J, Colby TV, Hagmeyer L, et al. European Respiratory Society guidelines on transbronchial lung cryobiopsy in the diagnosis of interstitial lung diseases. Eur Respir J. 2022;60(5).
19. Cottin V, Wollin L, Fischer A, Quaresma M, Stowasser S, Harari S. Fibrosing interstitial lung diseases: knowns and unknowns. Eur Respir Rev. 2019;28(151).
20. Noble PW, Albera C, Bradford WZ, Costabel U, Glassberg MK, Kardatzke D, et al. Pirfenidone in patients with idiopathic pulmonary fibrosis (CAPACITY): two randomised trials. Lancet. 2011;377(9779):1760-9.

21. Richeldi L, du Bois RM, Raghu G, Azuma A, Brown KK, Costabel U, et al. Efficacy and safety of nintedanib in idiopathic pulmonary fibrosis. N Engl J Med. 2014;370(22):2071-82.
22. Idiopathic Pulmonary Fibrosis Clinical Research N, Raghu G, Anstrom KJ, King TE, Jr., Lasky JA, Martinez FJ. Prednisone, azathioprine, and N-acetylcysteine for pulmonary fibrosis. N Engl J Med. 2012;366(21):1968-77.
23. Johannson KA, Chaudhuri N, Adegunsoye A, Wolters PJ. Treatment of fibrotic interstitial lung disease: current approaches and future directions. Lancet. 2021;398(10309):1450-60.
24. Wells AU, Brown KK, Flaherty KR, Kolb M, Thannickal VJ, Group IPFCW. What's in a name? That which we call IPF, by any other name would act the same. Eur Respir J. 2018;51(5).
25. Flaherty KR, Wells AU, Cottin V, Devaraj A, Walsh SLF, Inoue Y, et al. Nintedanib in Progressive Fibrosing Interstitial Lung Diseases. N Engl J Med. 2019;381(18):1718-27.
26. Lederer DJ, Martinez FJ. Idiopathic Pulmonary Fibrosis. N Engl J Med. 2018;378(19):1811-23.
27. Hoffmann-Vold AM, Allanore Y, Alves M, Brunborg C, Airo P, Ananieva LP, et al. Progressive interstitial lung disease in patients with systemic sclerosis-associated interstitial lung disease in the EUSTAR database. Ann Rheum Dis. 2021;80(2):219-27.
28. Behr J, Prasse A, Kreuter M, Johow J, Rabe KF, Bonella F, et al. Pirfenidone in patients with progressive fibrotic interstitial lung diseases other than idiopathic pulmonary fibrosis (RELIEF): a double-blind, randomised, placebo-controlled, phase 2b trial. Lancet Respir Med. 2021;9(5):476-86.
29. George PM, Spagnolo P, Kreuter M, Altinisik G, Bonifazi M, Martinez FJ, et al. Progressive fibrosing interstitial lung disease: clinical uncertainties, consensus recommendations, and research priorities. Lancet Respir Med. 2020;8(9):925-34.
30. Wijsenbeek M, Kreuter M, Olson A, Fischer A, Bendstrup E, Wells CD, et al. Progressive fibrosing interstitial lung diseases: current practice in diagnosis and management. Curr Med Res Opin. 2019;35(11):2015-24.

31. Zamora-Legoff JA, Krause ML, Crowson CS, Ryu JH, Matteson EL. Progressive Decline of Lung Function in Rheumatoid Arthritis-Associated Interstitial Lung Disease. Arthritis Rheumatol. 2017;69(3):542-9.
32. Maher TM, Corte TJ, Fischer A, Kreuter M, Lederer DJ, Molina-Molina M, et al. Pirfenidone in patients with unclassifiable progressive fibrosing interstitial lung disease: a double-blind, randomised, placebo-controlled, phase 2 trial. Lancet Respir Med. 2020;8(2):147-57.
33. Rajan SK, Cottin V, Dhar R, Danoff S, Flaherty KR, Brown KK, et al. Progressive pulmonary fibrosis: an expert group consensus statement. Eur Respir J. 2023;61(3).
34. Raghu G, Remy-Jardin M, Ryerson CJ, Myers JL, Kreuter M, Vasakova M, et al. Diagnosis of Hypersensitivity Pneumonitis in Adults. An Official ATS/JRS/ALAT Clinical Practice Guideline. Am J Respir Crit Care Med. 2020;202(3):e36-e69.
35. Olson AL, Gifford AH, Inase N, Fernandez Perez ER, Suda T. The epidemiology of idiopathic pulmonary fibrosis and interstitial lung diseases at risk of a progressive-fibrosing phenotype. Eur Respir Rev. 2018;27(150).
36. Olson A, Hartmann N, Patnaik P, Wallace L, Schlenker-Herceg R, Nasser M, et al. Estimation of the Prevalence of Progressive Fibrosing Interstitial Lung Diseases: Systematic Literature Review and Data from a Physician Survey. Adv Ther. 2021;38(2):854-67.
37. Faverio P, Piluso M, De Giacomi F, Della Zoppa M, Cassandro R, Harari S, et al. Progressive Fibrosing Interstitial Lung Diseases: Prevalence and Characterization in Two Italian Referral Centers. Respiration. 2020;99(10):838-45.
38. Fernandez Perez ER, Travis WD, Lynch DA, Brown KK, Johannson KA, Selman M, et al. Diagnosis and Evaluation of Hypersensitivity Pneumonitis: CHEST Guideline and Expert Panel Report. Chest. 2021;160(2):e97-e156.
39. Marinescu DC, Raghu G, Remy-Jardin M, Travis WD, Adegunsoye A, Beasley MB, et al. Integration and Application of Clinical Practice Guidelines for the Diagnosis of Idiopathic Pulmonary Fibrosis and Fibrotic Hypersensitivity Pneumonitis. Chest. 2022;162(3):614-29.

40. Nasser M, Larrieu S, Si-Mohamed S, Ahmad K, Boussel L, Brevet M, et al. Progressive fibrosing interstitial lung disease: a clinical cohort (the PROGRESS study). Eur Respir J. 2021;57(2).
41. Santos V, Martins N, Sousa C, Jacob M, Padrao E, Melo N, et al. Hypersensitivity pneumonitis: Main features characterization in a Portuguese cohort. Pulmonology. 2020;26(3):130-7.
42. Robalo Cordeiro C, Campos P, Carvalho L, Campainha S, Clemente S, Figueiredo L, et al. Consensus document for the diagnosis and treatment of idiopathic pulmonary fibrosis: Joint Consensus of Sociedade Portuguesa de Pneumologia, Sociedade Portuguesa de Radiologia e Medicina Nuclear e Sociedade Portuguesa de Anatomia Patologica. Rev Port Pneumol (2006). 2016;22(2):112-22.
43. Coelho DB, Santos V, Alexandre AT, Bastos HNE, Mota PC, Morais A, et al. Risk Factors for Acute Exacerbations of Fibrotic Hypersensitivity Pneumonitis-Is Exposure a Trigger That We're Missing? Arch Bronconeumol. 2023;59(1):59-60.
44. Fernandez Perez ER, Swigris JJ, Forssen AV, Tourin O, Solomon JJ, Huie TJ, et al. Identifying an inciting antigen is associated with improved survival in patients with chronic hypersensitivity pneumonitis. Chest. 2013;144(5):1644-51.
45. Walsh SL, Sverzellati N, Devaraj A, Wells AU, Hansell DM. Chronic hypersensitivity pneumonitis: high resolution computed tomography patterns and pulmonary function indices as prognostic determinants. Eur Radiol. 2012;22(8):1672-9.
46. Selman M, Pardo A, Wells AU. Usual interstitial pneumonia as a stand-alone diagnostic entity: the case for a paradigm shift? Lancet Respir Med. 2023;11(2):188-96.
47. Wright JL, Churg A, Hague CJ, Wong A, Ryerson CJ. Pathologic separation of idiopathic pulmonary fibrosis from fibrotic hypersensitivity pneumonitis. Mod Pathol. 2020;33(4):616-25.
48. Santos G, Fabiano A, Mota PC, Rodrigues I, Carvalho D, Melo N, et al. The impact of nintedanib and pirfenidone on lung function and survival in patients with idiopathic pulmonary fibrosis in real-life setting. Pulm Pharmacol Ther. 2023;83:102261.

49. Morisset J, Johannson KA, Vittinghoff E, Aravena C, Elicker BM, Jones KD, et al. Use of Mycophenolate Mofetil or Azathioprine for the Management of Chronic Hypersensitivity Pneumonitis. Chest. 2017;151(3):619-25.
50. Terras Alexandre A, Martins N, Raimundo S, Melo N, Catetano Mota P, Novais EBH, et al. Impact of Azathioprine use in chronic hypersensitivity pneumonitis patients. Pulm Pharmacol Ther. 2020;60:101878.
51. Raimundo S, Pimenta AC, Cruz-Martins N, Rodrigues MC, Melo N, Mota PC, et al. Insights on chronic hypersensitivity pneumonitis' treatment: Factors associated with a favourable response to azathioprine. Life Sci. 2021;272:119274.
52. Cottin V, Richeldi L, Rosas I, Otaola M, Song JW, Tomassetti S, et al. Nintedanib and immunomodulatory therapies in progressive fibrosing interstitial lung diseases. Respir Res. 2021;22(1):84.
53. Vasakova M, Morell F, Walsh S, Leslie K, Raghu G. Hypersensitivity Pneumonitis: Perspectives in Diagnosis and Management. Am J Respir Crit Care Med. 2017;196(6):680-9.
54. Kouranos V, Jacob J, Nicholson A, Renzoni E. Fibrotic Hypersensitivity Pneumonitis: Key Issues in Diagnosis and Management. J Clin Med. 2017;6(6).
55. Selman M, King TE, Pardo A, American Thoracic S, European Respiratory S, American College of Chest P. Idiopathic pulmonary fibrosis: prevailing and evolving hypotheses about its pathogenesis and implications for therapy. Ann Intern Med. 2001;134(2):136-51.
56. King TE, Jr., Pardo A, Selman M. Idiopathic pulmonary fibrosis. Lancet. 2011;378(9807):1949-61.
57. Han MK, Zhou Y, Murray S, Tayob N, Noth I, Lama VN, et al. Lung microbiome and disease progression in idiopathic pulmonary fibrosis: an analysis of the COMET study. Lancet Respir Med. 2014;2(7):548-56.
58. Distler JHW, Gyorfi AH, Ramanujam M, Whitfield ML, Konigshoff M, Lafyatis R. Shared and distinct mechanisms of fibrosis. Nat Rev Rheumatol. 2019;15(12):705-30.

59. Vasakova M, Selman M, Morell F, Sterclova M, Molina-Molina M, Raghu G. Hypersensitivity Pneumonitis: Current Concepts of Pathogenesis and Potential Targets for Treatment. Am J Respir Crit Care Med. 2019;200(3):301-8.
60. Buendia-Roldan I, Santiago-Ruiz L, Perez-Rubio G, Mejia M, Rojas-Serrano J, Ambrocio-Ortiz E, et al. A major genetic determinant of autoimmune diseases is associated with the presence of autoantibodies in hypersensitivity pneumonitis. Eur Respir J. 2020;56(2).
61. Selman M, Pardo A, King TE, Jr. Hypersensitivity pneumonitis: insights in diagnosis and pathobiology. Am J Respir Crit Care Med. 2012;186(4):314-24.
62. Kolb M, Vasakova M. The natural history of progressive fibrosing interstitial lung diseases. Respir Res. 2019;20(1):57.
63. Spagnolo P, Grunewald J, du Bois RM. Genetic determinants of pulmonary fibrosis: evolving concepts. Lancet Respir Med. 2014;2(5):416-28.
64. Noth I, Zhang Y, Ma SF, Flores C, Barber M, Huang Y, et al. Genetic variants associated with idiopathic pulmonary fibrosis susceptibility and mortality: a genome-wide association study. Lancet Respir Med. 2013;1(4):309-17.
65. Fingerlin TE, Murphy E, Zhang W, Peljto AL, Brown KK, Steele MP, et al. Genome-wide association study identifies multiple susceptibility loci for pulmonary fibrosis. Nat Genet. 2013;45(6):613-20.
66. Allen RJ, Guillen-Guio B, Oldham JM, Ma SF, Dressen A, Paynton ML, et al. Genome-Wide Association Study of Susceptibility to Idiopathic Pulmonary Fibrosis. Am J Respir Crit Care Med. 2020;201(5):564-74.
67. Koudstaal T, Funke-Chambour M, Kreuter M, Molyneaux PL, Wijsenbeek MS. Pulmonary fibrosis: from pathogenesis to clinical decision-making. Trends Mol Med. 2023;29(12):1076-87.
68. Peljto AL, Blumhagen RZ, Walts AD, Cardwell J, Powers J, Corte TJ, et al. Idiopathic Pulmonary Fibrosis Is Associated with Common Genetic Variants and Limited Rare Variants. Am J Respir Crit Care Med. 2023;207(9):1194-202.
69. Kaur A, Mathai SK, Schwartz DA. Genetics in Idiopathic Pulmonary Fibrosis Pathogenesis, Prognosis, and Treatment. Front Med (Lausanne). 2017;4:154.

70. Zhang D, Newton CA, Wang B, Povysil G, Noth I, Martinez FJ, et al. Utility of whole genome sequencing in assessing risk and clinically relevant outcomes for pulmonary fibrosis. Eur Respir J. 2022;60(6).
71. Newton CA, Molyneaux PL, Oldham JM. Clinical Genetics in Interstitial Lung Disease. Front Med (Lausanne). 2018;5:116.
72. Coghlan MA, Shifren A, Huang HJ, Russell TD, Mitra RD, Zhang Q, et al. Sequencing of idiopathic pulmonary fibrosis-related genes reveals independent single gene associations. BMJ Open Respir Res. 2014;1(1):e000057.
73. Petrovski S, Todd JL, Durheim MT, Wang Q, Chien JW, Kelly FL, et al. An Exome Sequencing Study to Assess the Role of Rare Genetic Variation in Pulmonary Fibrosis. Am J Respir Crit Care Med. 2017;196(1):82-93.
74. Moss BJ, Ryter SW, Rosas IO. Pathogenic Mechanisms Underlying Idiopathic Pulmonary Fibrosis. Annu Rev Pathol. 2022;17:515-46.
75. Khera AV, Chaffin M, Aragam KG, Haas ME, Roselli C, Choi SH, et al. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. Nat Genet. 2018;50(9):1219-24.
76. Moll M, Peljto AL, Kim JS, Xu H, Debban CL, Chen X, et al. A Polygenic Risk Score for Idiopathic Pulmonary Fibrosis and Interstitial Lung Abnormalities. Am J Respir Crit Care Med. 2023;208(7):791-801.
77. Seibold MA, Wise AL, Speer MC, Steele MP, Brown KK, Loyd JE, et al. A common MUC5B promoter polymorphism and pulmonary fibrosis. N Engl J Med. 2011;364(16):1503-12.
78. Zhang Y, Noth I, Garcia JG, Kaminski N. A variant in the promoter of MUC5B and idiopathic pulmonary fibrosis. N Engl J Med. 2011;364(16):1576-7.
79. Peljto AL, Zhang Y, Fingerlin TE, Ma SF, Garcia JG, Richards TJ, et al. Association between the MUC5B promoter polymorphism and survival in patients with idiopathic pulmonary fibrosis. JAMA. 2013;309(21):2232-9.
80. Stock CJ, Sato H, Fonseca C, Banya WA, Molyneaux PL, Adamali H, et al. Mucin 5B promoter polymorphism is associated with idiopathic pulmonary fibrosis but not with development of lung fibrosis in systemic sclerosis or sarcoidosis. Thorax. 2013;68(5):436-41.

81. Ley B, Newton CA, Arnould I, Elicker BM, Henry TS, Vittinghoff E, et al. The MUC5B promoter polymorphism and telomere length in patients with chronic hypersensitivity pneumonitis: an observational cohort-control study. Lancet Respir Med. 2017;5(8):639-47.
82. Kropski JA, Pritchett JM, Zoz DF, Crossno PF, Markin C, Garnett ET, et al. 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. 2015;191(4):417-26.
83. Zhang D, Newton CA. Familial Pulmonary Fibrosis: Genetic Features and Clinical Implications. Chest. 2021;160(5):1764-73.
84. Newton CA, Oldham JM, Ley B, Anand V, Adegunsoye A, Liu G, et al. Telomere length and genetic variant associations with interstitial lung disease progression and survival. Eur Respir J. 2019;53(4).
85. Cutting CC, Bowman WS, Dao N, Pugashetti JV, Garcia CK, Oldham JM, et al. Family History of Pulmonary Fibrosis Predicts Worse Survival in Patients With Interstitial Lung Disease. Chest. 2021;159(5):1913-21.
86. Cui F, Sun Y, Xie J, Li D, Wu M, Song L, et al. Air pollutants, genetic susceptibility and risk of incident idiopathic pulmonary fibrosis. Eur Respir J. 2023;61(2).
87. Bennett D, Mazzei MA, Squitieri NC, Bargagli E, Refini RM, Fossi A, et al. Familial pulmonary fibrosis: Clinical and radiological characteristics and progression analysis in different high resolution-CT patterns. Respir Med. 2017;126:75-83.
88. Lee HL, Ryu JH, Wittmer MH, Hartman TE, Lymp JF, Tazelaar HD, et al. Familial idiopathic pulmonary fibrosis: clinical features and outcome. Chest. 2005;127(6):2034-41.
89. Krauss E, Gehrken G, Drakopanagiotakis F, Tello S, Dartsch RC, Maurer O, et al. Clinical characteristics of patients with familial idiopathic pulmonary fibrosis (f-IPF). BMC Pulm Med. 2019;19(1):130.
90. Borie R, Kannengiesser C, Sicre de Fontbrune F, Gouya L, Nathan N, Crestani B. Management of suspected monogenic lung fibrosis in a specialised centre. Eur Respir Rev. 2017;26(144).

91. van Moorsel CHM, van der Vis JJ, Grutters JC. Genetic disorders of the surfactant system: focus on adult disease. Eur Respir Rev. 2021;30(159).
92. Stanley SE, Gable DL, Wagner CL, Carlile TM, Hanumanthu VS, Podlevsky JD, et al. Loss-of-function mutations in the RNA biogenesis factor NAF1 predispose to pulmonary fibrosis-emphysema. Sci Transl Med. 2016;8(351):351ra107.
93. Kannengiesser C, Manali ED, Revy P, Callebaut I, Ba I, Borgel A, et al. First heterozygous NOP10 mutation in familial pulmonary fibrosis. Eur Respir J. 2020;55(6).
94. Gable DL, Gaysinskaya V, Atik CC, Talbot CC, Jr., Kang B, Stanley SE, et al. ZCCHC8, the nuclear exosome targeting component, is mutated in familial pulmonary fibrosis and is required for telomerase RNA maturation. Genes Dev. 2019;33(19-20):1381-96.
95. Alonso-Gonzalez A, Tosco-Herrera E, Molina-Molina M, Flores C. Idiopathic pulmonary fibrosis and the role of genetics in the era of precision medicine. Front Med (Lausanne). 2023;10:1152211.
96. Liu Q, Zhou Y, Cogan JD, Mitchell DB, Sheng Q, Zhao S, et al. The Genetic Landscape of Familial Pulmonary Fibrosis. Am J Respir Crit Care Med. 2023;207(10):1345-57.
97. Newton CA, Batra K, Torrealba J, Kozlitina J, Glazer CS, Aravena C, et al. Telomere-related lung fibrosis is diagnostically heterogeneous but uniformly progressive. Eur Respir J. 2016;48(6):1710-20.
98. Newton CA, Oldham JM, Applegate C, Carmichael N, Powell K, Dilling D, et al. The Role of Genetic Testing in Pulmonary Fibrosis: A Perspective From the Pulmonary Fibrosis Foundation Genetic Testing Work Group. Chest. 2022;162(2):394-405.
99. Savage SA, Alter BP. Dyskeratosis congenita. Hematol Oncol Clin North Am. 2009;23(2):215-31.
100. Diaz de Leon A, Cronkhite JT, Yilmaz C, Brewington C, Wang R, Xing C, et al. Subclinical lung disease, macrocytosis, and premature graying in kindreds with telomerase (TERT) mutations. Chest. 2011;140(3):753-63.
101. Borie R, Tabeze L, Thabut G, Nunes H, Cottin V, Marchand-Adam S, et al. Prevalence and characteristics of TERT and TERC mutations in suspected genetic pulmonary fibrosis. Eur Respir J. 2016;48(6):1721-31.

102. Nunes H, Jeny F, Bouvry D, Picard C, Bernaudin JF, Menard C, et al. Pleuroparenchymal fibroelastosis associated with telomerase reverse transcriptase mutations. Eur Respir J. 2017;49(5).
103. Ley B, Torgerson DG, Oldham JM, Adegunsoye A, Liu S, Li J, et al. Rare Protein-Altering Telomere-related Gene Variants in Patients with Chronic Hypersensitivity Pneumonitis. Am J Respir Crit Care Med. 2019;200(9):1154-63.
104. Juge PA, Borie R, Kannengiesser C, Gazal S, Revy P, Wemeau-Stervinou L, et al. Shared genetic predisposition in rheumatoid arthritis-interstitial lung disease and familial pulmonary fibrosis. Eur Respir J. 2017;49(5).
105. Newton CA, Zhang D, Oldham JM, Kozlitina J, Ma SF, Martinez FJ, et al. Telomere Length and Use of Immunosuppressive Medications in Idiopathic Pulmonary Fibrosis. Am J Respir Crit Care Med. 2019;200(3):336-47.
106. Stuart BD, Choi J, Zaidi S, Xing C, Holohan B, Chen R, et al. Exome sequencing links mutations in PARN and RTEL1 with familial pulmonary fibrosis and telomere shortening. Nat Genet. 2015;47(5):512-7.
107. Diaz de Leon A, Cronkhite JT, Katzenstein AL, Godwin JD, Raghu G, Glazer CS, et al. Telomere lengths, pulmonary fibrosis and telomerase (TERT) mutations. PLoS One. 2010;5(5):e10680.
108. Zhang D, Eckhardt CM, McGroder C, Benesh S, Porcelli J, Depender C, et al. Clinical Impact of Telomere Length Testing for Interstitial Lung Disease. Chest. 2024.
109. Dodson LM, Baldan A, Nissbeck M, Gunja SMR, Bonnen PE, Aubert G, et al. From incomplete penetrance with normal telomere length to severe disease and telomere shortening in a family with monoallelic and biallelic PARN pathogenic variants. Hum Mutat. 2019;40(12):2414-29.
110. van der Vis JJ, van der Smagt JJ, van Batenburg AA, Goldschmeding R, van Es HW, Grutters JC, et al. Pulmonary fibrosis in non-mutation carriers of families with short telomere syndrome gene mutations. Respirology. 2021;26(12):1160-70.
111. Phillips-Houlbracq M, Mal H, Cottin V, Gauvain C, Beier F, Sicre de Fontbrune F, et al. Determinants of survival after lung transplantation in telomerase-related gene mutation carriers: A retrospective cohort. Am J Transplant. 2022;22(4):1236-44.

112. Zhang D, Adegunsoye A, Oldham JM, Kozlitina J, Garcia N, Poonawalla M, et al. Telomere length and immunosuppression in non-idiopathic pulmonary fibrosis interstitial lung disease. Eur Respir J. 2023;62(5).
113. Adegunsoye A, Morisset J, Newton CA, Oldham JM, Vittinghoff E, Linderholm AL, et al. Leukocyte telomere length and mycophenolate therapy in chronic hypersensitivity pneumonitis. Eur Respir J. 2021;57(3).
114. Michalski JE, Schwartz DA. Genetic Risk Factors for Idiopathic Pulmonary Fibrosis: Insights into Immunopathogenesis. J Inflamm Res. 2020;13:1305-18.
115. Adegunsoye A, Vij R, Noth I. Integrating Genomics Into Management of Fibrotic Interstitial Lung Disease. Chest. 2019;155(5):1026-40.
116. Zhang D, Povysil G, Newton CA, Maher TM, Molyneaux PL, Noth I, et al. Genome-wide Enrichment of TERT Rare Variants in Idiopathic Pulmonary Fibrosis Patients of Latino Ancestry. Am J Respir Crit Care Med. 2022;206(7):903-5.
117. Alder JK, Chen JJ, Lancaster L, Danoff S, Su SC, Cogan JD, et al. Short telomeres are a risk factor for idiopathic pulmonary fibrosis. Proc Natl Acad Sci U S A. 2008;105(35):13051-6.
118. Kirwan M, Dokal I. Dyskeratosis congenita: a genetic disorder of many faces. Clin Genet. 2008;73(2):103-12.
119. Cogan JD, Kropski JA, Zhao M, Mitchell DB, Rives L, Markin C, et al. Rare variants in RTEL1 are associated with familial interstitial pneumonia. Am J Respir Crit Care Med. 2015;191(6):646-55.
120. Borie R, Bouvry D, Cottin V, Gauvain C, Cazes A, Debray MP, et al. Regulator of telomere length 1 (RTEL1) mutations are associated with heterogeneous pulmonary and extra-pulmonary phenotypes. Eur Respir J. 2019;53(2).
121. Philippot Q, Kannengiesser C, Debray MP, Gauvain C, Ba I, Vieri M, et al. Interstitial lung diseases associated with mutations of poly(A)-specific ribonuclease: A multicentre retrospective study. Respirology. 2022;27(3):226-35.
122. Alder JK, Stanley SE, Wagner CL, Hamilton M, Hanumanthu VS, Armanios M. Exome sequencing identifies mutant TINF2 in a family with pulmonary fibrosis. Chest. 2015;147(5):1361-8.

123. Nogee LM, Dunbar AE, 3rd, Wert S, Askin F, Hamvas A, Whitsett JA. Mutations in the surfactant protein C gene associated with interstitial lung disease. Chest. 2002;121(3 Suppl):20S-1S.
124. Takezaki A, Tsukumo SI, Setoguchi Y, Ledford JG, Goto H, Hosomichi K, et al. A homozygous SFTPA1 mutation drives necroptosis of type II alveolar epithelial cells in patients with idiopathic pulmonary fibrosis. J Exp Med. 2019;216(12):2724-35.
125. Legendre M, Butt A, Borie R, Debray MP, Bouvry D, Filhol-Blin E, et al. Functional assessment and phenotypic heterogeneity of SFTPA1 and SFTPA2 mutations in interstitial lung diseases and lung cancer. Eur Respir J. 2020;56(6).
126. Wang Y, Kuan PJ, Xing C, Cronkhite JT, Torres F, Rosenblatt RL, et al. Genetic defects in surfactant protein A2 are associated with pulmonary fibrosis and lung cancer. Am J Hum Genet. 2009;84(1):52-9.
127. Young LR, Nogee LM, Barnett B, Panos RJ, Colby TV, Deutsch GH. Usual interstitial pneumonia in an adolescent with ABCA3 mutations. Chest. 2008;134(1):192-5.
128. Epaud R, Delestrain C, Louha M, Simon S, Fanen P, Tazi A. Combined pulmonary fibrosis and emphysema syndrome associated with ABCA3 mutations. Eur Respir J. 2014;43(2):638-41.
129. Klay D, Platenburg M, van Rijswijk R, Grutters JC, van Moorsel CHM. ABCA3 mutations in adult pulmonary fibrosis patients: a case series and review of literature. Curr Opin Pulm Med. 2020;26(3):293-301.
130. Justet A, Klay D, Porcher R, Cottin V, Ahmad K, Molina Molina M, et al. Safety and efficacy of pirfenidone and nintedanib in patients with idiopathic pulmonary fibrosis and carrying a telomere-related gene mutation. Eur Respir J. 2021;57(2).
131. Silhan LL, Shah PD, Chambers DC, Snyder LD, Riise GC, Wagner CL, et al. Lung transplantation in telomerase mutation carriers with pulmonary fibrosis. Eur Respir J. 2014;44(1):178-87.
132. Borie R, Kannengiesser C, Hirschi S, Le Pavec J, Mal H, Bergot E, et al. Severe hematologic complications after lung transplantation in patients with telomerase complex mutations. J Heart Lung Transplant. 2015;34(4):538-46.

133. Nathan N, Giraud V, Picard C, Nunes H, Dastot-Le Moal F, Copin B, et al. Germline SFTPA1 mutation in familial idiopathic interstitial pneumonia and lung cancer. Hum Mol Genet. 2016;25(8):1457-67.
134. Brudon A, Legendre M, Mageau A, Bermudez J, Bonniaud P, Bouvry D, et al. High risk of lung cancer in surfactant-related gene variant carriers. Eur Respir J. 2024;63(5).
135. Diesler R, Legendre M, Si-Mohamed S, Brillet PY, Wemeau L, Manali ED, et al. Similarities and differences of interstitial lung disease associated with pathogenic variants in SFTPC and ABCA3 in adults. Respirology. 2024;29(4):312-23.
136. Thouvenin G, Nathan N, Epaud R, Clement A. Diffuse parenchymal lung disease caused by surfactant deficiency: dramatic improvement by azithromycin. BMJ Case Rep. 2013;2013.
137. Klay D, Grutters JC, van der Vis JJ, Platenburg M, Kelder JC, Tromp E, et al. Progressive Disease With Low Survival in Adult Patients With Pulmonary Fibrosis Carrying Surfactant-Related Gene Mutations: An Observational Study. Chest. 2023;163(4):870-80.
138. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med. 2015;17(5):405-24.
139. Alder JK, Hanumanthu VS, Strong MA, DeZern AE, Stanley SE, Takemoto CM, et al. Diagnostic utility of telomere length testing in a hospital-based setting. Proc Natl Acad Sci U S A. 2018;115(10):E2358-E65.
140. Putman RK, Gudmundsson G, Araki T, Nishino M, Sigurdsson S, Gudmundsson EF, et al. The MUC5B promoter polymorphism is associated with specific interstitial lung abnormality subtypes. Eur Respir J. 2017;50(3).
141. Juge PA, Lee JS, Ebstein E, Furukawa H, Dobrinskikh E, Gazal S, et al. MUC5B Promoter Variant and Rheumatoid Arthritis with Interstitial Lung Disease. N Engl J Med. 2018;379(23):2209-19.
142. Zhang Y. MUC5B and short telomere length in hypersensitivity pneumonitis. Lancet Respir Med. 2017;5(8):603-4.

143. Platenburg M, Wiertz IA, van der Vis JJ, Crestani B, Borie R, Dieude P, et al. The MUC5B promoter risk allele for idiopathic pulmonary fibrosis predisposes to asbestosis. Eur Respir J. 2020;55(4).
144. Johnson C, Rosen P, Lloyd T, Horton M, Christopher-Stine L, Oddis CV, et al. Exploration of the MUC5B promoter variant and ILD risk in patients with autoimmune myositis. Respir Med. 2017;130:52-4.
145. Dressen A, Abbas AR, Cabanski C, Reeder J, Ramalingam TR, Neighbors M, et al. Analysis of protein-altering variants in telomerase genes and their association with MUC5B common variant status in patients with idiopathic pulmonary fibrosis: a candidate gene sequencing study. Lancet Respir Med. 2018;6(8):603-14.
146. Li X, Goobie GC, Gregory AD, Kass DJ, Zhang Y. Toll-Interacting Protein in Pulmonary Diseases. Abiding by the Goldilocks Principle. Am J Respir Cell Mol Biol. 2021;64(5):536-46.
147. Allen RJ, Porte J, Braybrooke R, Flores C, Fingerlin TE, Oldham JM, et al. Genetic variants associated with susceptibility to idiopathic pulmonary fibrosis in people of European ancestry: a genome-wide association study. Lancet Respir Med. 2017;5(11):869-80.
148. Partanen JJ, Happola P, Zhou W, Lehisto AA, Ainola M, Sutinen E, et al. Leveraging global multi-ancestry meta-analysis in the study of idiopathic pulmonary fibrosis genetics. Cell Genom. 2022;2(10):100181.
149. Zhu QQ, Zhang XL, Zhang SM, Tang SW, Min HY, Yi L, et al. Association Between the MUC5B Promoter Polymorphism rs35705950 and Idiopathic Pulmonary Fibrosis: A Meta-analysis and Trial Sequential Analysis in Caucasian and Asian Populations. Medicine (Baltimore). 2015;94(43):e1901.
150. Horimasu Y, Ohshimo S, Bonella F, Tanaka S, Ishikawa N, Hattori N, et al. MUC5B promoter polymorphism in Japanese patients with idiopathic pulmonary fibrosis. Respirology. 2015;20(3):439-44.
151. Peljto AL, Selman M, Kim DS, Murphy E, Tucker L, Pardo A, et al. The MUC5B promoter polymorphism is associated with idiopathic pulmonary fibrosis in a Mexican cohort but is rare among Asian ancestries. Chest. 2015;147(2):460-4.

152. Wang C, Zhuang Y, Guo W, Cao L, Zhang H, Xu L, et al. Mucin 5B promoter polymorphism is associated with susceptibility to interstitial lung diseases in Chinese males. PLoS One. 2014;9(8):e104919.
153. Jiang H, Hu Y, Shang L, Li Y, Yang L, Chen Y. Association between MUC5B polymorphism and susceptibility and severity of idiopathic pulmonary fibrosis. Int J Clin Exp Pathol. 2015;8(11):14953-8.
154. van der Vis JJ, Snetselaar R, Kazemier KM, ten Klooster L, Grutters JC, van Moorsel CH. Effect of Muc5b promoter polymorphism on disease predisposition and survival in idiopathic interstitial pneumonias. Respirology. 2016;21(4):712-7.
155. Li X, Kim SE, Chen TY, Wang J, Yang X, Tabib T, et al. Toll interacting protein protects bronchial epithelial cells from bleomycin-induced apoptosis. FASEB J. 2020;34(8):9884-98.
156. Borie R, Crestani B, Dieude P, Nunes H, Allanore Y, Kannengiesser C, et al. The MUC5B variant is associated with idiopathic pulmonary fibrosis but not with systemic sclerosis interstitial lung disease in the European Caucasian population. PLoS One. 2013;8(8):e70621.
157. Hunninghake GM, Hatabu H, Okajima Y, Gao W, Dupuis J, Latourelle JC, et al. MUC5B promoter polymorphism and interstitial lung abnormalities. N Engl J Med. 2013;368(23):2192-200.
158. Wu X, Li W, Luo Z, Chen Y. The minor T allele of the MUC5B promoter rs35705950 associated with susceptibility to idiopathic pulmonary fibrosis: a meta-analysis. Sci Rep. 2021;11(1):24007.
159. Kishore A, Zizkova V, Kocourkova L, Petrkova J, Bouros E, Nunes H, et al. Association Study for 26 Candidate Loci in Idiopathic Pulmonary Fibrosis Patients from Four European Populations. Front Immunol. 2016;7:274.
160. O'Dwyer DN, Armstrong ME, Trujillo G, Cooke G, Keane MP, Fallon PG, et al. The Toll-like receptor 3 L412F polymorphism and disease progression in idiopathic pulmonary fibrosis. Am J Respir Crit Care Med. 2013;188(12):1442-50.
161. Falfan-Valencia R, Camarena A, Pineda CL, Montano M, Juarez A, Buendia-Roldan I, et al. Genetic susceptibility to multicase hypersensitivity pneumonitis is

associated with the TNF-238 GG genotype of the promoter region and HLA-DRB1*04 bearing HLA haplotypes. Respir Med. 2014;108(1):211-7.

162. Fingerlin TE, Zhang W, Yang IV, Ainsworth HC, Russell PH, Blumhagen RZ, et al. Genome-wide imputation study identifies novel HLA locus for pulmonary fibrosis and potential role for auto-immunity in fibrotic idiopathic interstitial pneumonia. BMC Genet. 2016;17(1):74.
163. van der Vis JJ, Prasse A, Renzoni EA, Stock CJW, Caliskan C, Maher TM, et al. MUC5B rs35705950 minor allele associates with older age and better survival in idiopathic pulmonary fibrosis. Respirology. 2023;28(5):455-64.
164. Sese L, Borie R, Kannengiesser C, Cottin V, Israel-Biet D, Crestani B, et al. Impact of Air Pollution and MUC5B Genotype on Survival in Idiopathic Pulmonary Fibrosis. Ann Am Thorac Soc. 2023.
165. Bonella F, Campo I, Zorzetto M, Boerner E, Ohshimo S, Theegarten D, et al. Potential clinical utility of MUC5B und TOLLIP single nucleotide polymorphisms (SNPs) in the management of patients with IPF. Orphanet J Rare Dis. 2021;16(1):111.
166. Oldham JM, Ma SF, Martinez FJ, Anstrom KJ, Raghu G, Schwartz DA, et al. TOLLIP, MUC5B, and the Response to N-Acetylcysteine among Individuals with Idiopathic Pulmonary Fibrosis. Am J Respir Crit Care Med. 2015;192(12):1475-82.
167. Allen RJ, Oldham JM, Jenkins DA, Leavy OC, Guillen-Guio B, Melbourne CA, et al. Longitudinal lung function and gas transfer in individuals with idiopathic pulmonary fibrosis: a genome-wide association study. Lancet Respir Med. 2023;11(1):65-73.
168. Blach J, Mackiewicz B. How much do we know about genetic predisposition of hypersensitivity pneumonitis? Ann Agric Environ Med. 2022;29(2):306-8.
169. Camarena A, Juarez A, Mejia M, Estrada A, Carrillo G, Falfan R, et al. Major histocompatibility complex and tumor necrosis factor-alpha polymorphisms in pigeon breeder's disease. Am J Respir Crit Care Med. 2001;163(7):1528-33.
170. Freitas C, Lima B, Melo N, Mota P, Novais-Bastos H, Alves H, et al. Distinct TNF-alpha and HLA polymorphisms associate with fibrotic and non-fibrotic subtypes of hypersensitivity pneumonitis. Pulmonology. 2023;29 Suppl 4:S63-S9.

171. Bouros D, Tzouvelekis A. Telomeropathy in Chronic Hypersensitivity Pneumonitis. Am J Respir Crit Care Med. 2019;200(9):1086-7.
172. Katayanagi S, Setoguchi Y, Kitagawa S, Okamoto T, Miyazaki Y. Alternative Gene Expression by TOLLIP Variant Is Associated With Lung Function in Chronic Hypersensitivity Pneumonitis. Chest. 2022;161(2):458-69.
173. Shirali M, Knott SA, Pong-Wong R, Navarro P, Haley CS. Haplotype Heritability Mapping Method Uncovers Missing Heritability of Complex Traits. Sci Rep. 2018;8(1):4982.
174. Schrodi SJ, Garcia VE, Rowland C, Jones HB. Pairwise linkage disequilibrium under disease models. Eur J Hum Genet. 2007;15(2):212-20.
175. Khankhanian P, Gourraud PA, Lizee A, Goodin DS. Haplotype-based approach to known MS-associated regions increases the amount of explained risk. J Med Genet. 2015;52(9):587-94.
176. 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.
177. Ramos E, Lopes C, Barros H. Investigating the effect of nonparticipation using a population-based case-control study on myocardial infarction. Ann Epidemiol. 2004;14(6):437-41.
178. Pintalhao M, Vasques-Novoa F, Couto-Viana B, Pimentel MJ, Neves JS, Mendonca L, et al. Relaxin-2, pathophysiological insights and outcomes in heart failure with preserved ejection fraction: Findings from the NETDiamond cohort. Int J Cardiol. 2022;365:87-90.
179. Leite AR, Neves JS, Angelico-Goncalves A, Vasques-Novoa F, Saraiva FA, Pinho IB, et al. Clinical and Pathophysiologic Insights of Free triiodothyronine/Free thyroxine Ratio in Patients with Heart Failure with Preserved Ejection Fraction: Data from the NETDiamond Cohort. Cardiology. 2023;148(3):239-45.
180. Mota PC, Soares ML, Vasconcelos CD, Ferreira AC, Lima BA, Manduchi E, et al. Predictive value of common genetic variants in idiopathic pulmonary fibrosis survival. J Mol Med (Berl). 2022;100(9):1341-53.

181. Dhooria S, Bal A, Sehgal IS, Prasad KT, Kashyap D, Sharma R, et al. MUC5B Promoter Polymorphism and Survival in Indian Patients With Idiopathic Pulmonary Fibrosis. Chest. 2022;162(4):824-7.
182. Biondini D, Cocconcelli E, Bernardinello N, Lorenzoni G, Rigobello C, Lococo S, et al. Prognostic role of MUC5B rs35705950 genotype in patients with idiopathic pulmonary fibrosis (IPF) on antifibrotic treatment. Respir Res. 2021;22(1):98.
183. Sese L, Borie R, Kannengiesser C, Cottin V, Israel-Biet D, Crestani B, et al. Impact of Air Pollution and MUC5B Genotype on Survival in Idiopathic Pulmonary Fibrosis. Ann Am Thorac Soc. 2024;21(3):519-23.
184. Whalen W BK, Kim JS, Simmons W, Ma SF, Kaner RJ, Martinez FJ, Anstrom KJ, Parfrey H, Maher T, Hammond M, Clark AB, Thickett D, Jenkins RG, Wilson AM, Noth I. TOLLIP Single Nucleotide Polymorphism and Antimicrobial Treatment Effect in Idiopathic Pulmonary Fibrosis. Am J Respir Crit Care Med. 2024.
185. Obayashi Y, Yamadori I, Fujita J, Yoshinouchi T, Ueda N, Takahara J. The role of neutrophils in the pathogenesis of idiopathic pulmonary fibrosis. Chest. 1997;112(5):1338-43.
186. Gharaee-Kermani M, Hu B, Phan SH, Gyetko MR. The role of urokinase in idiopathic pulmonary fibrosis and implication for therapy. Expert Opin Investig Drugs. 2008;17(6):905-16.
187. Ritchie MD, Hahn LW, Roodi N, Bailey LR, Dupont WD, Parl FF, et al. Multifactor-dimensionality reduction reveals high-order interactions among estrogen-metabolism genes in sporadic breast cancer. Am J Hum Genet. 2001;69(1):138-47.
188. Moore JH, Andrews PC. Epistasis analysis using multifactor dimensionality reduction. Methods Mol Biol. 2015;1253:301-14.
189. Furusawa H, Peljto AL, Walts AD, Cardwell J, Molyneaux PL, Lee JS, et al. Common idiopathic pulmonary fibrosis risk variants are associated with hypersensitivity pneumonitis. Thorax. 2022;77(5):508-10.
190. Furusawa H, Cardwell JH, Okamoto T, Walts AD, Konigsberg IR, Kurche JS, et al. Chronic Hypersensitivity Pneumonitis, an Interstitial Lung Disease with Distinct Molecular Signatures. Am J Respir Crit Care Med. 2020;202(10):1430-44.

191. Mota PC, Soares ML, Ferreira AC, Santos RF, Rufo JC, Vasconcelos D, et al. Polymorphisms and haplotypes of TOLLIP and MUC5B are associated with susceptibility and survival in patients with fibrotic hypersensitivity pneumonitis. Pulmonology. 2024.
192. De Sadeleer LJ, Hermans F, De Dycker E, Yserbyt J, Verschakelen JA, Verbeken EK, et al. Impact of BAL lymphocytosis and presence of honeycombing on corticosteroid treatment effect in fibrotic hypersensitivity pneumonitis: a retrospective cohort study. Eur Respir J. 2020;55(4).
193. Wei R, Li C, Zhang M, Jones-Hall YL, Myers JL, Noth I, et al. Association between MUC5B and TERT polymorphisms and different interstitial lung disease phenotypes. Transl Res. 2014;163(5):494-502.
194. Selman M, Pardo A, Barrera L, Estrada A, Watson SR, Wilson K, et al. Gene expression profiles distinguish idiopathic pulmonary fibrosis from hypersensitivity pneumonitis. Am J Respir Crit Care Med. 2006;173(2):188-98.
195. Isshiki T, Koyama K, Homma S, Sakamoto S, Yamasaki A, Shimizu H, et al. Association of rs3750920 polymorphism in TOLLIP with clinical characteristics of fibrosing interstitial lung diseases in Japanese. Sci Rep. 2021;11(1):16250.
196. Abbasi A, Chen C, Gandhi CK, Wu R, Pardo A, Selman M, et al. Single Nucleotide Polymorphisms (SNP) and SNP-SNP Interactions of the Surfactant Protein Genes Are Associated With Idiopathic Pulmonary Fibrosis in a Mexican Study Group; Comparison With Hypersensitivity Pneumonitis. Front Immunol. 2022;13:842745.
197. Perez-Padilla R, Salas J, Chapela R, Sanchez M, Carrillo G, Perez R, et al. Mortality in Mexican patients with chronic pigeon breeder's lung compared with those with usual interstitial pneumonia. Am Rev Respir Dis. 1993;148(1):49-53.
198. Salisbury ML, Gu T, Murray S, Gross BH, Chughtai A, Sayyouh M, et al. Hypersensitivity Pneumonitis: Radiologic Phenotypes Are Associated With Distinct Survival Time and Pulmonary Function Trajectory. Chest. 2019;155(4):699-711.
199. Palomaki A, FinnGen Rheumatology Clinical Expert G, Palotie A, Koskela J, Eklund KK, Pirinen M, et al. Lifetime risk of rheumatoid arthritis-associated interstitial lung disease in MUC5B mutation carriers. Ann Rheum Dis. 2021;80(12):1530-6.