

PCSK6 and Survival in Idiopathic Pulmonary Fibrosis

Justin M. Oldham MD*¹, Richard J. Allen PhD*², Jose M. Lorenzo-Salazar MSc*³, Philip L. Molyneaux MBBS*^{4,5}, Shwu-Fan Ma PhD⁶, Chitra Joseph PhD⁷, John S. Kim MD⁶, Beatriz Guillen-Guio PhD^{2,8}, Tamara Hernández-Beeftink MSc⁸, Jonathan A. Kropski MD⁹, Yong Huang MD⁶, Cathryn T. Lee MD¹⁰, Ayodeji Adegunsoye MD¹⁰, Janelle Vu Pugashetti MD¹¹, Angela L. Linderholm PhD¹¹, Vivian Vo BS¹¹, Mary E. Streck MD¹⁰, Jonathan Jou MD¹², Adrian Muñoz-Barrera³; Luis A. Rubio-Rodriguez³; Richard Hubbard MBBS^{13,14}, Nik Hirani MBBS¹⁵, Moira K.B. Whyte MBBS¹⁵, Simon Hart MBBS¹⁶, Andrew G. Nicholson MBBS^{4,5}, Lisa Lancaster MD⁹, Helen Parfrey MBBS¹⁷, Doris Rassel MBBS¹⁷, William Wallace MBBS¹⁵, Eleanor Valenzi MD¹⁸, Yingze Zhang PhD¹⁸, Josyf Mychaleckyj, DPhil;¹⁹ Amy Stockwell MS²⁰, Naftali Kaminski MD²¹, Paul J. Wolters MD²², Maria Molina-Molina MD^{23,24}, William A. Fahy²⁵, Fernando J. Martinez MD²⁶, Ian P. Hall BMBCh^{7,14}, Martin D. Tobin MBChB^{2,27}, Toby M. Maher BM^{4,5,28}, Timothy S. Blackwell MD⁹, Brian L. Yaspan PhD²⁰, R Gisli Jenkins MBBS#^{4,5}, Carlos Flores PhD#^{3,8,24}, Louise V Wain PhD#^{2,27}, Imre Noth MD#⁶

*These authors contributed equally

#These authors contributed equally

Affiliations

1. Division of Pulmonary and Critical Care Medicine, University of Michigan, Ann Arbor, MI USA
2. Department of Health Sciences, University of Leicester, Leicester, UK

3. Genomics Division, Instituto Tecnológico y de Energías Renovables (ITER), Santa Cruz de Tenerife, Spain
4. National Heart and Lung Institute, Imperial College London, London, UK
5. Royal Brompton and Harefield Hospitals, Guy's and St. Thomas' NHS Foundation Trust, London, UK
6. Division of Pulmonary and Critical Care Medicine, University of Virginia, Charlottesville, VA USA
7. University of Nottingham, Division of Respiratory Medicine, Nottingham, UK
8. Research Unit, Hospital Universitario Nuestra Señora de Candelaria, Santa Cruz de Tenerife, Spain
9. Division of Pulmonary and Critical Care Medicine, Vanderbilt University, Nashville, TN USA
10. Section of Pulmonary and Critical Care Medicine, University of Chicago, Chicago, IL USA
11. Division of Pulmonary, Critical Care and Sleep Medicine, University of California Davis, Davis, CA USA
12. Department of Surgery, University of Illinois College of Medicine, Peoria, IL
13. Division of Epidemiology and Public Health, University of Nottingham, Nottingham, United Kingdom
14. National Institute for Health Research, Nottingham Biomedical Research Centre, Nottingham University Hospitals NHS Trust, Nottingham, United Kingdom
15. Medical Research Council Centre for Inflammation Research, The University of Edinburgh, Edinburgh, United Kingdom

16. Respiratory Research Group, Hull York Medical School, Castle Hill Hospital,
Cottingham, United Kingdom
17. Cambridge Interstitial Lung Disease Service, Royal Papworth Hospital, Cambridge,
United Kingdom
18. Division of Pulmonary, Allergy and Critical Care Medicine, The University of
Pittsburgh, Pittsburgh, PA USA
19. Center for Public Health Genomics; University of Virginia, Charlottesville, VA, USA
20. Genentech Inc., San Francisco, CA USA
21. Section of Pulmonary, Critical Care and Sleep Medicine, Yale School of Medicine,
New Haven, CT USA
22. Division of Pulmonary, Critical Care, Allergy and Sleep Medicine, University of
California at San Francisco, San Francisco, CA USA
23. Servei de Pneumologia, Laboratori de Pneumologia Experimental, IDIBELL, Campus
de Bellvitge, Universitat de Barcelona, Barcelona, Spain,
24. CIBER de Enfermedades Respiratorias, Instituto de Salud Carlos III, Madrid, Spain
25. Discovery Medicine, GlaxoSmithKline, Stevenage, UK
26. Weill Cornell Medical Center, New York, NY USA
27. National Institute for Health Research, Leicester Respiratory Biomedical Research
Centre, Glenfield Hospital, Leicester, UK
28. Division of Pulmonary and Critical Care Medicine, University of Southern California,
Los Angeles, CA USA

Corresponding Author

Justin M. Oldham

6301 MSRB3

1150 W. Medical Center Dr.

Ann Arbor, MI 48109-5642

oldhamj@med.umich.edu

Author Contributions

JMO, RJA, JMLS, SFM, RGJ, CF, LVW and IN designed and supervised the study. JMO, PLM, JAK, AA, JVP, MES, RH, NH, MW, SH, HP, NK, PW, MMM, FJM, TMM, TSB, RGJ and IN recruited patients for the study. JMO, PLM, SFM, JAK, AL, VV, YZ, AS, NK, PW, IH, MDT, WAF, TMM, TSB, BLY, RGJ, CF and IN contributed to DNA extraction and genotyping. CJ, AGN, DR, WW and RGJ secured lung tissue and performed the IHC analysis. JMO, RJA, JMLS, CJ, JSK, BGG, THB, YH, AS, BLY, RGJ, CF, LVW performed the analyses. JMO, PLM, JK, CTL, AA, JVP, MES, EV, YZ, KG, DK, NK, PW, MMM, FJM, TMM, RGJ and IN collected paired clinical data. JMO, JMLS, RJA, RGJ, CF, LW and IN wrote the manuscript with the input of all authors. All authors approved the final version of this manuscript.

Funding

NHLBI - K23HL138190 (JMO), R56HL158935 (JMO), K23HL150301 (JSK), K23HL146942 (AA), T32HL007605 (CTL), RO1HL130796 (IN), UG3HL145266 (IN)

Ministry of Science and Innovation (RTC-2017-6471-1; AEI/FEDER UE (CF)

Instituto de Salud Carlos III co-financed by the ERDF “A way of making Europe” from the EU (PI20/00876, CF), Instituto Tecnológico y de Energías Renovables (OA17/008 (CF) Consejería de Educación-Gobierno de Canarias and Cabildo Insular de Tenerife (BOC n.º 173, 24/08/2017 (JMLS)

UK National Institute for Health Research (NIHR) - RGJ holds an NIHR Research Professorship (RP-2017-08-ST2-014) and has MRC grants (G0901226 and MR/V00235X/1). TMM is supported by an NIHR Clinician Scientist Fellowship (NIHR reference CS-2013-13-017). MDT is supported by an NIHR Senior Investigator Award (NIHR201371). The views expressed are our own and not necessarily those of the NHS, the NIHR, or the UK Department of Health.

British Lung Foundation Chair in Respiratory Research (C17-3) (TMM), GlaxoSmithKline/British Lung Foundation Chair in Respiratory Research (C17-1) (LWW) Wellcome Trust - WT202849/Z/16/Z (MDT), 221680/Z/20/Z (BGG). For the purpose of open access, the author has applied a CC BY public copyright license to any Author Accepted Manuscript version arising from this submission.

BREATHE — The Health Data Research Hub for Respiratory Health (MC_PC_19004) (LWW) Action for Pulmonary Fibrosis Research Fellows (RJA and PLM).

This research used the ALICE and SPECTRE High Performance Computing Facilities at the University of Leicester and TeideHPC (<https://teidehpc.iter.es/en/home>) at Instituto Tecnológico y de Energías Renovables.

Role of the Funding Source

This study was funded by the agencies listed above, which had no role in the design or conduct of the study. All authors had full access to all data, and the corresponding author had final responsibility for the decision to submit for publication.

Running Head: PCSK6 and Survival in IPF

Manuscript word count: 2727

Abstract word count: 250/250

Abstract

Rationale: Idiopathic pulmonary fibrosis (IPF) is a devastating disease characterized by limited treatment options and high mortality. Novel therapies and prognostic biomarkers are needed.

Objective: To identify and validate molecular determinants of IPF survival.

Methods: A staged genome-wide association study (GWAS) was performed using paired genomic and survival data. Stage I cases were drawn from centers across the US and Europe and stage II cases from Vanderbilt University. Cox proportional hazards regression was used to identify gene variants associated with differential transplant-free survival (TFS). Stage I variants with nominal significance ($p < 5 \times 10^{-5}$) were advanced for stage II testing and meta-analyzed to identify those reaching genome-wide significance ($p < 5 \times 10^{-8}$). Downstream analyses were performed for genes and proteins associated with variants reaching genome-wide significance.

Main Results: After quality controls, 1481 stage I cases and 397 stage II cases were included in the analysis. After filtering, 9,075,629 variants were tested in stage I, with 158 meeting advancement criteria. Four variants associated with TFS with consistent effect direction were identified in stage II, including one in an intron of proprotein convertase subtilisin/kexin type 6 (*PCSK6*) reaching genome-wide significance (HR 4.11; 95%CI 2.54-6.67; $p = 9.45 \times 10^{-9}$). *PCSK6* protein was highly expressed in IPF lung parenchyma and negatively correlated with survival. Peripheral blood *PCSK6* gene expression and plasma concentration were associated with reduced transplant-free survival.

Conclusions: We identified four novel variants associated with IPF survival, including one in *PCSK6* that reached genome-wide significance. Downstream analyses suggested that PCSK6 protein may serve as prognostic biomarker in IPF and potential therapeutic target.

Abstract word count: 250/250

Keywords

Idiopathic Pulmonary Fibrosis, Genome-wide association study, Genomics, Survival, PCSK6, protein

Introduction

Idiopathic Pulmonary Fibrosis (IPF) is a devastating disease characterized by progressive lung scarring and poor survival.^{1,2} Two anti-fibrotic therapies have been approved for the treatment of IPF after randomized controlled trials demonstrated efficacy in slowing lung function decline.^{3,4} Despite this advance, outcomes remain poor and anti-fibrotic therapy appears to provide only modest survival benefit.⁵ To improve IPF outcomes, novel therapeutic targets are needed.

We and others have identified molecular IPF risk factors through unbiased investigation of the genome, transcriptome, and proteome.⁶⁻¹⁵ Among the strongest molecular determinants of IPF is a common variant in the promoter region of *MUC5B*, which increases the odds of developing IPF by nearly 5-fold per risk allele.⁶⁻⁹ Despite this strong association with IPF onset, the *MUC5B* promoter was paradoxically associated with improved survival.¹⁶ Few other susceptibility-associated gene variants have been shown to reliably predict differential IPF survival, suggesting that molecular determinants of IPF susceptibility and progression may have limited overlap.

To better understand molecular drivers of IPF progression and identify new therapeutic targets, we conducted a two-stage, multi-center, international genome-wide association study (GWAS) of IPF survival, followed by downstream analysis of genes and proteins associated with top survival-associated variant to determine whether these circulating biomarkers also predicted differential survival.

Methods

Cohorts and case selection

All patients provided informed consent for research blood draw in accordance with protocols approved by the institutional review board at each participating institution. GWAS stage I cases consisted of unrelated IPF patients of European ancestry from three previously described case-control GWAS datasets from the United States (US),⁹ United Kingdom (UK),⁷ and a combined cohort from the US, UK and Spain (UUS)⁶. Available outcome data was gathered for all cases meeting international consensus criteria for IPF¹⁷ and survival plotted for individual cohorts within each dataset. Patients without available outcome data were excluded, as were clinical trial cohorts due to short follow-up (**Supplemental Methods**). Stage II cases consisted of previously described, unrelated IPF patients of European ancestry from Vanderbilt University.¹⁸

Genotyping and quality control

Genotypes were generated for stage I cases using SNP (single nucleotide polymorphism) genotyping arrays (Thermo Fisher Scientific, Waltham, MA) according to previously described methods.^{6,7,9} Imputation for stage I cases was performed using the Michigan Imputation Server using the Haplotype Reference Consortium panel (v1.1 2016). Genotypes for the stage II cases were determined by whole genome sequencing, as previously described.¹⁸ Stringent quality control measures were applied with a two-tier variant filtering scheme as follows: those ranging $0.5 \leq \text{MAF} < 1\%$ were retained when imputation $R^2 \geq 0.8$; and those with $\text{MAF} \geq 1\%$ when imputation $R^2 \geq 0.5$. In addition, variants deviating from Hardy-Weinberg equilibrium ($p < 1.0 \times 10^{-6}$) were removed.

Genome-wide survival analysis

The primary endpoint assessed was transplant-free survival (TFS), defined as the time in months from site-determined date of IPF diagnosis to event (death or lung transplant) or censoring date. Variants associated with differential TFS were identified using a multivariable Cox proportional hazards regression model adjusted for age, sex, center, and first ten genetic principal components. Variant genotypes were treated as a continuous variable with each patient having an imputed genotype dosage between zero and two risk alleles. Variants present in at least two datasets (US, UK and UUS) were meta-analyzed using METAL (v2011-03-25) to generate stage I results.

Variants nominally associated with TFS in stage I were defined as those with Wald $p < 0.05$ in at least two datasets with the same direction of effect and $p < 5.0 \times 10^{-5}$ in stage I meta-analysis. Conditional analysis of these SNPs to deduce their independence was performed with GCTA-COJO v1.26. The proportional hazards assumption was then assessed for each independent variant meeting advancement criteria by testing whether Schoenfeld residual rank varied by genotype strata. Variants that satisfied the proportional hazards assumption were advanced for stage II testing. Stage I and II cases were then meta-analyzed using METAL with the genome-wide significance threshold set at $p < 5.0 \times 10^{-8}$. *In silico* assessments were used to infer the biological effect of variants associated with TFS after stage II testing.

PCSK6 tissue expression

Formalin-fixed paraffin-embedded human lung tissue sections obtained from patients with IPF undergoing surgical lung biopsy were compared to control subjects undergoing lung resection for malignancy, with sections distal to areas of malignancy utilized.

Immunohistochemistry was performed using standard methods (**Supplemental Methods**) and mean staining intensity of PCSK6 protein was compared between IPF cases and non-IPF controls using a Mann-Whitney U-test.

PCSK6 clinical outcome association

The association between circulating *PCSK6* gene expression and TFS was assessed using three previously published microarray datasets from the COMET trial, Imperial College and the University of Chicago, which were analyzed separately with results meta-analyzed and presented as a forest plot.¹⁹ Circulating plasma PCSK6 protein concentration was then determined in patients with IPF from UC-Davis and UChicago (**Supplemental Methods**), log₂ transformed, and tested for TFS association using Cox proportional hazards regression.²⁰ The proportional hazards assumption was satisfied for all downstream survival analyses.

Results

Case selection for stage I

Patients comprising the US cohort included those from the University of Chicago (n=118) and University of Pittsburgh (n=200) (**Figure E1, Figure E2**). Those comprising the UK cohort included patients from the University of Edinburgh (n=119), Trent Lung Fibrosis

Study (n=210), a subset of those participating in the prospective, multi-center PROFILE study (NCT 01134822) (n=175), and aggregated patients from smaller UK centers (Hull and Papworth) (n=61) (**Figure E1, Figure E3**). Patients comprising the UUS cohort included those from the University of Chicago (independent of those from the US cohort, n=187), PROFILE study (independent of those in the UK dataset, n=299), University of California (Davis and San Francisco) (n=84) and aggregated patients from centers in Spain (n=28) (**Figure E1, Figure E4**).

Baseline characteristics and outcomes

Following phenotypic exclusions, 1481 patients comprising stage I were included in the analysis. These included 318 patients from the US dataset, 565 from the UK dataset, and 598 from the UUS dataset. Baseline characteristics for each dataset are shown in **Table 1**. The mean age ranged from 67 to 72 years and males comprised 71-75% of each dataset. The mean percent predicted FVC and diffusion capacity of the lung for carbon monoxide (DLCO) was lowest in patients comprising the US dataset and highest in the UK dataset. A majority of patients in each dataset were classified as gender, age, physiology²¹ (GAP) stage I or II. Death events were highest in the UK cohort and lowest in the UUS cohort (**Figure E5**). Median survival was 39.3 months in the US dataset, 53.2 months in the UK dataset and 40.6 months in the UUS dataset. Median survival was 48 months in the Vanderbilt University validation cohort (**Figure E6**).

Genome-wide survival analysis

After filtering, 7,873,835 variants in the US dataset, 8,591,398 variants in the UK dataset, and 8,620,496 variants in the UUS dataset were tested for TFS association. Quintile-quintile plots for each stage I cohort suggested acceptable inflation (**Figure E7**). After stratifying stage I cohorts by minor allele frequency (MAF), inflation was higher for rare variants compared to low and high frequency variants, but within an acceptable range for each group ($\lambda < 1.1$) (**Figure E8**). For meta-analysis, 9,075,629 variants were tested for TFS association in the aggregated stage I cohorts. One hundred and sixty-one independent SNPs had Wald $p < 0.05$ in at least two datasets with the same direction of effect and $p < 5.0 \times 10^{-5}$ in stage I meta-analysis (**Figure 1**). Of those, 158 satisfied the proportional hazards assumption and advanced for stage II testing (**Table E1**).

Genotype data was available in the Vanderbilt University cohort for 154 of the 158 variants advanced from stage I. Six variants were associated with TFS in the Vanderbilt University cohort at $p < 0.05$, including four with consistent effect direction that strengthened in TFS association after meta-analysis (**Table 2; Table E2**). These four were rs184498750 near Succinate-CoA Ligase GDP/ADP-Forming Subunit Alpha (*SUCLG1*), rs60514164 near ubiquitin-conjugating enzyme E2Q family member 2 (*UBE2Q2*), rs35647788 in an intron of Proprotein Convertase Subtilisin/Kexin Type 6 (*PCSK6*), and rs3893252 in an intron of Deleted In Azoospermia-Associated Protein 1 (*DAZAP1*) (**Table 2**). Of these, rs35647788 (*PCSK6*) showed strong TFS association in stage I (HR 4.76; 95% CI 2.62-8.64; $p = 2.96 \times 10^{-7}$) and stage II (HR 3.12; 95% CI 1.37-7.11; $p = 6.70 \times 10^{-3}$) cohorts and crossed the genome-wide significance threshold in meta-analysis (HR 4.11; 95% CI 2.54-6.67; $p = 9.45 \times 10^{-9}$).

With the exception of rs60514164 (*UBE2Q2*) (MAF=8%), these SNPs were low frequency, with MAF of ~1% in the study population. Regional association plots for each of the four variants are shown in **Figure E9**.

In multivariable analysis, each variant except rs3893252 (*DAZAP1*) maintained survival association after adjustment for relevant confounders of IPF survival (**Table E3**). Among patients with the rs35647788 (*PCSK6*) variant, all were heterozygotes (**Table E4**) and were evenly distributed across centers comprising the UK and UUS cohorts. No rs35647788 (*PCSK6*) variants were observed in the US cohort despite good imputation quality ($r^2=0.74$). In sensitivity analysis of the *PCSK6* variant, results were consistent when censoring transplants (**Table E5**). *In silico* testing revealed functional effects for each of the four variants associated with TFS after stage II meta-analysis (**Table E6**). None of the variants identified had known association with fibrotic lung disease.

PCSK6 tissue expression

Morphologic assessment of histological sections from lung tissue in patients with IPF were compared with control subjects without fibrotic lung disease. In IPF lung, cytoplasmic *PCSK6* expression localized to ciliated epithelial cells and alveolar epithelial cells and was markedly higher than *PCSK6* expression in non-IPF control sections (**Figure 2**). Western Blot confirmed the presence of only a single *PCSK6* band (**Figure E10**). Relative staining intensity was two-fold higher in IPF lung samples (n=86) compared with non-IPF controls (n=9) ($p<0.001$) (**Figure 2**).

PCSK6 clinical outcome association

Increased PCSK6 protein staining score was associated with reduced TFS in those with available survival data (n=71), with staining scores above the median associated with greater than 2-fold increased risk of death or lung transplant (HR 2.41; 95% CI 1.12-5.16; p=0.024) (**Figure 3a**). When assessing *PCSK6* gene expression in the COMET (n=75), Imperial College (n=55) and University of Chicago (n=45) cohorts, increasing PCSK6 expression was associated with increased mortality risk in each cohort, with each one-unit increased associated with greater than three-fold increased risk of death or lung transplant in meta-analysis (HR 3.43; 95% CI 1.62-7.25; p=0.0012) (**Figure 3b**). When assessing PCSK6 plasma concentration in patients with IPF from UC-Davis (n=138) and UChicago (n=181), increasing plasma PCSK6 concentration was associated with reduced TFS, with each one-unit change in log-transformed plasma concentration associated with a nearly 50% increase in outcome risk (HR 1.47; 95% CI 1.14, 1.89; p=0.0031). These results were consistent across UC-Davis (HR 1.47, 95% CI 1.14-1.89; p=0.08) and UChicago (HR 1.34, 95% CI 0.92-1.94; p=0.12) cohorts. After stratification of the combined cohort by tertiles, those with PCSK6 concentration in the highest tertile displayed significantly worse survival than those in the second and third tertiles (p=0.0018) (**Figure 3c**).

Overlap between IPF risk and transplant free survival

Variants previously associated with IPF risk⁶⁻¹⁰ were investigated for outcome association (**Supplementary Methods**). None of the 15 genetic variants with previously associated with IPF risk⁶⁻¹⁰ were associated with TFS after Bonferroni correction (p=0.0033) (**Table E7**). As previously reported,^{16,22} individuals with the *MUC5B* promoter polymorphism

(rs3570590) displayed better overall survival, though this did not reach significance after adjustment for multiple testing. None of the four validated survival variants were associated with differential IPF risk (**Table E8**). When combining the effect of thousands of IPF risk variants in a polygenic risk score, this risk score was not significantly associated with TFS for any significance threshold used (**Figure E11**), again suggesting that variants that affect disease risk may have little impact on survival times after diagnosis.

Discussion

In this investigation, we conducted the first GWAS of IPF survival, identifying a variant intronic to *PCSK6* that associated with differential TFS at genome-wide significance in two independent IPF cohorts totaling nearly two thousand patients. We subsequently found that PCSK6 protein was highly expressed in IPF lung tissue, localizing to alveolar epithelial cells, which play a key role in IPF onset and progression.²³ Finally, we found that PCSK6 lung staining, peripheral blood gene expression and circulating plasma concentration negatively correlated with TFS across independent IPF cohorts. To our knowledge, this study is the first to systematically identify gene variants associated with IPF survival and the first to identify PCSK6 as a potential therapeutic target in IPF.

PCSK6, also called *PACE4*, encodes a widely expressed calcium-dependent serine endoprotease with known expression in alveoli and alveolar macrophages.²⁴ PCSK6 is a critical mediator of TGF- β processing and is crucial for reproduction, embryological development and blood pressure regulation.²⁵⁻²⁹ A *PCSK6* gene variant has been implicated in the development of hypertension²⁹ and dysregulated *PCSK6* gene expression has been

linked to vascular disease,³⁰⁻³² and cardiac remodeling following myocardial ischemia.³³ These cardiovascular remodeling effects make PCSK6 of particular relevance to IPF, as *PCSK6* overexpression can lead to increased collagen I and III deposition, TGF- β activation and extracellular matrix formation,^{33,34} which are cardinal features of IPF pathogenesis.² Additionally, PCSK6 may bind tissue inhibitors of metalloproteinases (TIMPs),³⁵ potentially counteracting pro-fibrotic metalloproteinases.³⁶

PCSK6 dysregulation has also been implicated in the development of cancer of the lung,³⁷ breast,³⁸ ovary³⁹ thyroid⁴⁰ and prostate.⁴¹ PCSK6 has been shown to regulate apoptosis in prostate cancer⁴² and pancreatic cancer⁴³ and also linked to increased cancer cell invasiveness by enhancing bioactivity of matrix metalloproteinases and cytokines.⁴⁴ Accordingly, PCSK6 has been proposed as an anti-tumor therapeutic target^{45,46} and a bioavailable formulation of an anti-PCSK6 molecule is currently under investigation.⁴⁷ *In vitro* PCSK6 inhibition has already been shown to reduce fibroblast proliferation, migration and invasion in rheumatoid arthritis-associated synovitis,⁴⁸ suggesting it as a candidate target for the treatment of pulmonary fibrosis.

The *PCSK6* variant identified in this study is rare and unlikely to singularly explain subsequent gene expression and protein findings. *In silico* analyses identified several nearby regulatory elements which may include common functional variants, with smaller effects, in linkage disequilibrium with this *PCSK6* variant. While beyond the scope of this investigation, further research is needed to explore other rare variants in *PCSK6* that may have an additive

effect with the one identified, along with regulatory elements, expression quantitative trait loci and more complex structural variants that may be contributing to our findings.

This study also highlights important differences between genomic determinants of IPF susceptibility and survival, suggesting that genes involved in disease onset may be independent of those driving disease progression. No survival-associated variant identified in this analysis was associated with IPF risk. While replicating prior IPF risk association for the *MUC5B* promoter polymorphism, we found only weak association with improved survival, an observation that may be influenced by index event bias.²² As none of the survival-associated variants showed an association with disease risk, it is unlikely these survival results are affected by index event bias. These findings have potential implications for drug development, as genes associated with IPF survival may represent more effective therapeutic targets than those associated with IPF onset.

This study has several limitations. First, we acknowledge the relatively small sample size used to conduct this analysis. Despite being the largest genomic IPF outcome analysis reported to date, the modest size of this cohort limited our ability to identify higher frequency SNPs with modest effect sizes. Given sample size constraints for this rare disease, we pursued a two-stage approach with meta-analysis of candidate variants rather than a discovery/replication approach, which would have required substantially higher sample sizes in each cohort. The consistent effect association across cohorts and genome-wide significance for the *PCSK6* variant after meta-analysis increases confidence that this represents a true association, as does the downstream clinical outcome analysis showing

PCSK6 gene expression and protein concentration to be associated with differential TFS. Next, there were likely differences between cohorts used to conduct this study. Survival was best among patients in the UUS cohort, which may reflect selection bias, as these patients had less follow-up compared with those in the US and UK datasets. Additionally, a large proportion of patients comprising the UUS dataset were recruited after the US approval of pirfenidone and nintedanib, which might explain the better overall survival and patients recruited before 2012 may have been exposed to potentially harmful immunosuppression.⁴⁹ The timing of study recruitment relative to IPF diagnosis was also unknown, which could have influenced results. Finally, we only assessed individuals of European ancestry. Validation of results is required in patients of other ancestral backgrounds.

Conclusion

Here we present results from the first GWAS of IPF survival conducted to date. This study sheds important light on the genetics of IPF progression and identified novel variants which may contribute to this process, including rs35647788 in an intron of *PCSK6*. Downstream analysis demonstrated PCSK6 protein lung staining, peripheral blood gene expression and circulating plasma concentration to be associated with reduced IPF survival, suggesting PCSK6 may serve as a potential therapeutic target in patients with IPF.

Figures

Figure 1. Manhattan plot of stage I gene variants associated with IPF survival. SNPs reaching nominal significance ($p < 5 \times 10^{-5}$) fall above the blue line and those reaching genome-wide significance ($p < 5.0 \times 10^{-8}$) fall above the red line.

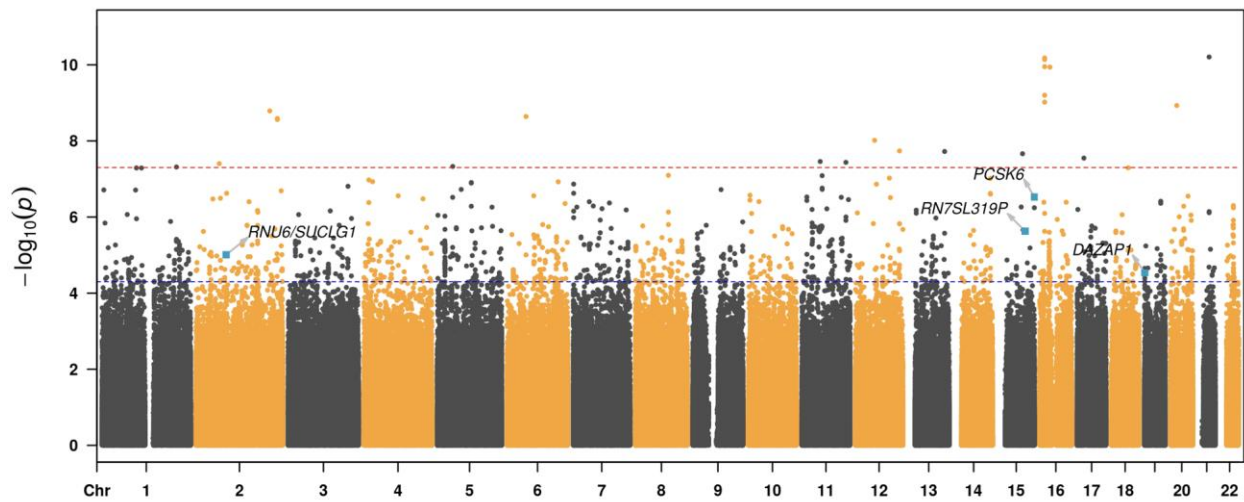


Figure 2. PCSK6 immunohistochemistry showed increased cytoplasmic PCSK6 expression in ciliated epithelial cells (A) and alveolar epithelial cells (B-C) compared with normal lung control sections (D-E). Parallel IPF section sections (F-G) confirm increased PCSK6 staining (F) when compared to control section (G). Human kidney positive control (H) is provided for reference. Comparison of relative PCSK6 staining intensity (I) demonstrated significantly higher mean intensity in IPF lungs when compared to non-IPF control lungs ($p < 0.001$). Magnification is 20 μ m magnification for panels A-E and 50 μ m for panels F-G.

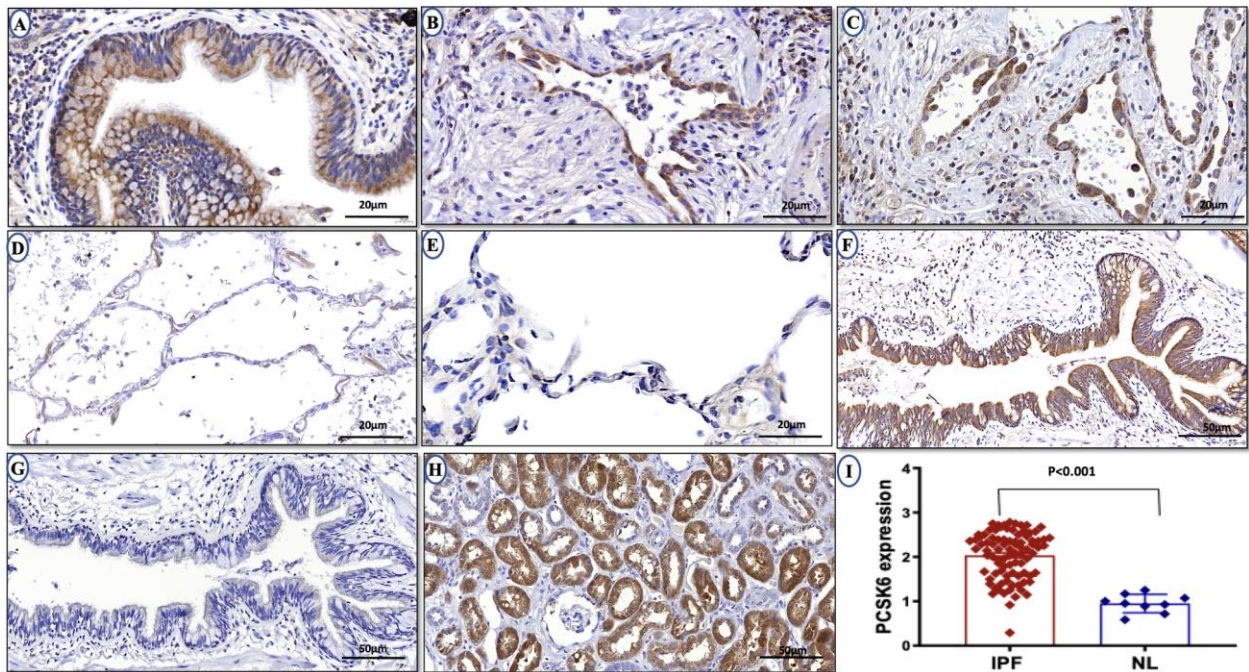
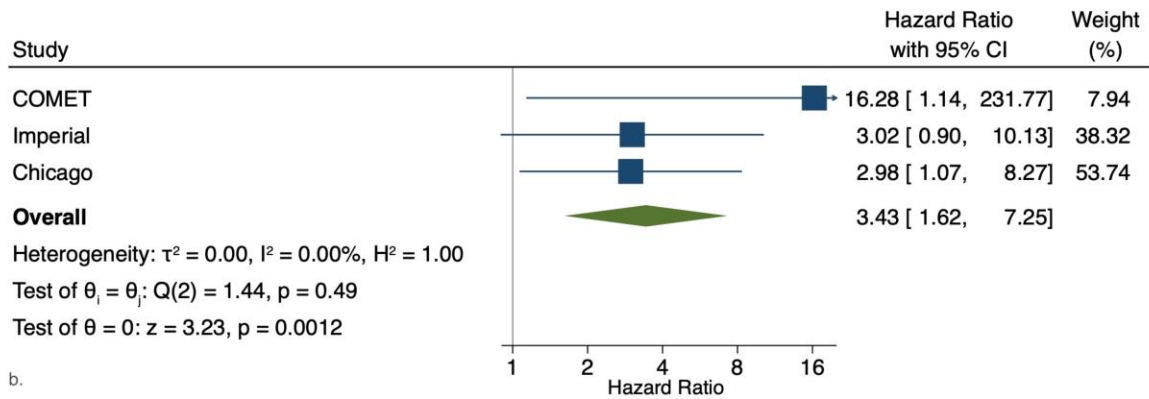
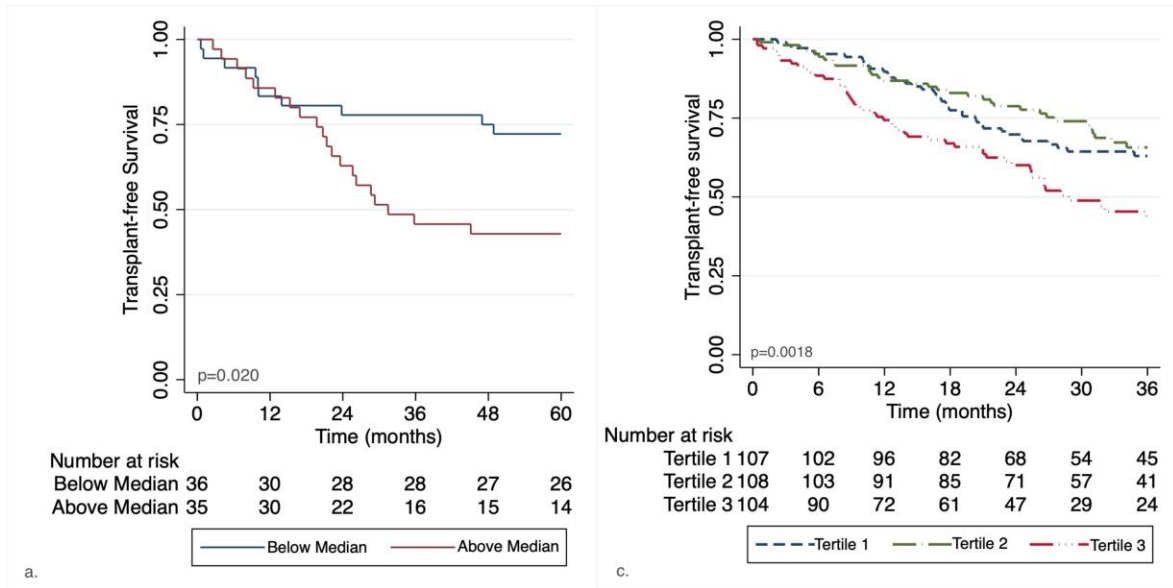


Figure 3. Relationship between PCSK6 and clinically relevant IPF endpoints. Higher PCSK6 staining intensity (a), peripheral blood gene expression (b) and circulating plasma protein concentration (c) are associated with reduced transplant-free survival.



Acknowledgements

We thank all the patients that contributed biologic samples and clinical data for this research.

Declaration of interests

JMO reports grants from the National Heart, Lung, and Blood Institute (NHLBI), American Lung Association and American Thoracic Society related to the submitted work and personal fees from Genentech, Boehringer Ingelheim, United Therapeutics, Lupin pharmaceuticals and AmMax Bio unrelated to the submitted work. PLM has received industry-academic funding from AstraZeneca and GSK and has received speaker and consultancy fees from Boehringer Ingelheim and Hoffman-La Roche outside the submitted work. JSK reports grants from the NHLBI and Pulmonary Fibrosis Foundation. AGN reports consultancy fees for Galapagos, Medical Quantitative Image Analysis and Boehringer Ingelheim, and fees for educational activities from Boehringer Ingelheim and UptoDate. AA reports grants from the National Heart, Lung, and Blood Institute (NHLBI), American College of Chest Physicians, and the Pulmonary Fibrosis Foundation, and speaking and advisory board fees from Genentech and Boehringer Ingelheim. MES has received grants from Boehringer Ingelheim and Galapagos and personal fees from Boehringer Ingelheim and Fibrogen. HP reports grants from Roche; personal fees from Boehringer Ingelheim, Roche and Pliant Therapeutics and is a Trustee for the charity Action for Pulmonary Fibrosis. IPH holds an NIHR Senior Investigator Award. AS and BLY are employees of Roche/Genentech with stock and stock options in Roche. TMM has received industry-academic funding from GlaxoSmithKline (GSK) R&D and UCB; and has received consultancy

or speakers fees from Apellis, AstraZeneca, Bayer, Biogen Idec, Boehringer Ingelheim, Cipla, GSK R&D, InterMune, ProMetic, Roche, Sanofi-Aventis, Sanumed, and UCB. IN reports personal fees and consultancy, advisory board, speaker, and study contracts for Boehringer Ingelheim and Genentech from Boehringer Ingelheim and Genentech; personal fees from Immuneworks, Sanofi Aventis, Global blood therapeutics; and has a patent for Toll-interacting protein (TOLLIP) and pharmacogenetics pending. NK served as a consultant to Biogen Idec, Boehringer Ingelheim, Third Rock, Pliant, Samumed, NuMedii, Theravance, LifeMax, Three Lake Partners, Optikira, Astra Zeneca, Veracyte, Augmanity and CSL Behring, over the last 3 years, reports Equity in Pliant and a grant from Veracyte, Boehringer Ingelheim, BMS and non-financial support from MiRagen and Astra Zeneca. NK has IP on novel biomarkers and therapeutics in IPF licensed to Biotech. RGJ reports grants from GSK, UK MRC, Biogen, AstraZeneca, Galecto and NIHR; personal fees from Boehringer Ingleheim, Bristol Myers Squibb, Chiesi, Galapagos, GSK, Heptares, MedImmune, Pharmakea, Roche/Promedior, Pliant Therapeutics, and Veracyte; and is a Trustee for the charity Action for Pulmonary Fibrosis. LVW has received funding from GlaxoSmithKline. All other authors declare no competing interests.

Data Sharing Statement

GWAS Summary statistics for this study can be accessed from

<https://github.com/genomicsITER/PFgenetics> and will be made available at GWAS Catalog

References

1. Raghu G, Remy-Jardin M, Myers JL, et al. Diagnosis of Idiopathic Pulmonary Fibrosis. An Official ATS/ERS/JRS/ALAT Clinical Practice Guideline. *Am J Respir Crit Care Med*. Sep 1 2018;198(5):e44-e68. doi:10.1164/rccm.201807-1255ST
2. Lederer DJ, Martinez FJ. Idiopathic Pulmonary Fibrosis. *N Engl J Med*. May 10 2018;378(19):1811-1823. doi:10.1056/NEJMra1705751
3. Richeldi L, du Bois RM, Raghu G, et al. Efficacy and safety of nintedanib in idiopathic pulmonary fibrosis. *N Engl J Med*. May 29 2014;370(22):2071-82. doi:10.1056/NEJMoa1402584
4. King TE, Jr., Bradford WZ, Castro-Bernardini S, et al. A phase 3 trial of pirfenidone in patients with idiopathic pulmonary fibrosis. *N Engl J Med*. May 29 2014;370(22):2083-92. doi:10.1056/NEJMoa1402582
5. Dempsey TM, Sangaralingham LR, Yao X, Sanghavi D, Shah ND, Limper AH. Clinical Effectiveness of Antifibrotic Medications for Idiopathic Pulmonary Fibrosis. *Am J Respir Crit Care Med*. Jul 15 2019;200(2):168-174. doi:10.1164/rccm.201902-0456OC
6. Allen RJ, Guillen-Guio B, Oldham JM, et al. Genome-Wide Association Study of Susceptibility to Idiopathic Pulmonary Fibrosis. *Am J Respir Crit Care Med*. Mar 1 2020;201(5):564-574. doi:10.1164/rccm.201905-1017OC
7. Allen RJ, Porte J, Braybrooke R, et al. Genetic variants associated with susceptibility to idiopathic pulmonary fibrosis in people of European ancestry: a genome-wide association study. *Lancet Respir Med*. Nov 2017;5(11):869-880. doi:10.1016/S2213-2600(17)30387-9
8. Fingerlin TE, Murphy E, Zhang W, et al. Genome-wide association study identifies multiple susceptibility loci for pulmonary fibrosis. *Nat Genet*. Jun 2013;45(6):613-20. doi:10.1038/ng.2609
9. 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*. Jun 2013;1(4):309-317. doi:10.1016/S2213-2600(13)70045-6
10. Dhindsa RS, Mattsson J, Nag A, et al. Identification of a missense variant in SPDL1 associated with idiopathic pulmonary fibrosis. *Commun Biol*. Mar 23 2021;4(1):392. doi:10.1038/s42003-021-01910-y
11. Adams TS, Schupp JC, Poli S, et al. Single-cell RNA-seq reveals ectopic and aberrant lung-resident cell populations in idiopathic pulmonary fibrosis. *Sci Adv*. Jul 2020;6(28):eaba1983. doi:10.1126/sciadv.aba1983
12. Todd JL, Neely ML, Overton R, et al. Peripheral blood proteomic profiling of idiopathic pulmonary fibrosis biomarkers in the multicentre IPF-PRO Registry. *Respir Res*. Oct 22 2019;20(1):227. doi:10.1186/s12931-019-1190-z
13. DePianto DJ, Chandriani S, Abbas AR, et al. Heterogeneous gene expression signatures correspond to distinct lung pathologies and biomarkers of disease severity in idiopathic pulmonary fibrosis. *Thorax*. Jan 2015;70(1):48-56. doi:10.1136/thoraxjnl-2013-204596
14. Sivakumar P, Ammar R, Thompson JR, et al. Integrated plasma proteomics and lung transcriptomics reveal novel biomarkers in idiopathic pulmonary fibrosis. *Respir Res*. Oct 24 2021;22(1):273. doi:10.1186/s12931-021-01860-3

15. O'Dwyer DN, Norman KC, Xia M, et al. The peripheral blood proteome signature of idiopathic pulmonary fibrosis is distinct from normal and is associated with novel immunological processes. *Sci Rep*. Apr 25 2017;7:46560. doi:10.1038/srep46560
16. Peljto AL, Zhang Y, Fingerlin TE, et al. Association between the MUC5B promoter polymorphism and survival in patients with idiopathic pulmonary fibrosis. *JAMA*. Jun 5 2013;309(21):2232-9. doi:10.1001/jama.2013.5827
17. Raghu G, Remy-Jardin M, Myers J, Richeldi L, Wilson KC. The 2018 Diagnosis of Idiopathic Pulmonary Fibrosis Guidelines: Surgical Lung Biopsy for Radiological Pattern of Probable Usual Interstitial Pneumonia Is Not Mandatory. *Am J Respir Crit Care Med*. Nov 1 2019;200(9):1089-1092. doi:10.1164/rccm.201907-1324ED
18. Dressen A, Abbas AR, Cabanski C, 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*. Aug 2018;6(8):603-614. doi:10.1016/S2213-2600(18)30135-8
19. Huang Y, Oldham JM, Ma SF, et al. Blood Transcriptomic Predicts Progression of Pulmonary Fibrosis and Associates Natural Killer Cells. *Am J Respir Crit Care Med*. Mar 9 2021;doi:10.1164/rccm.202008-3093OC
20. Huang Y, Ma SF, Vij R, et al. A functional genomic model for predicting prognosis in idiopathic pulmonary fibrosis. *BMC Pulm Med*. Nov 21 2015;15:147. doi:10.1186/s12890-015-0142-8
21. Ley B, Ryerson CJ, Vittinghoff E, et al. A multidimensional index and staging system for idiopathic pulmonary fibrosis. *Ann Intern Med*. May 15 2012;156(10):684-91. doi:10.7326/0003-4819-156-10-201205150-00004
22. Dudbridge F, Allen RJ, Sheehan NA, et al. Adjustment for index event bias in genome-wide association studies of subsequent events. *Nat Commun*. Apr 5 2019;10(1):1561. doi:10.1038/s41467-019-09381-w
23. Lederer DJ, Martinez FJ. Idiopathic Pulmonary Fibrosis. *N Engl J Med*. Aug 23 2018;379(8):797-798. doi:10.1056/NEJMc1807508
24. Uhlen M, Fagerberg L, Hallstrom BM, et al. Proteomics. Tissue-based map of the human proteome. *Science*. Jan 23 2015;347(6220):1260419. doi:10.1126/science.1260419
25. Diaz FJ, Sugiura K, Eppig JJ. Regulation of Pcsk6 expression during the preantral to antral follicle transition in mice: opposing roles of FSH and oocytes. *Biol Reprod*. Jan 2008;78(1):176-83. doi:10.1095/biolreprod.107.063537
26. Mujoomdar ML, Hogan LM, Parlow AF, Nachtigal MW. Pcsk6 mutant mice exhibit progressive loss of ovarian function, altered gene expression, and formation of ovarian pathology. *Reproduction*. Mar 2011;141(3):343-55. doi:10.1530/REP-10-0451
27. Constam DB, Robertson EJ. SPC4/PACE4 regulates a TGFbeta signaling network during axis formation. *Genes Dev*. May 1 2000;14(9):1146-55.
28. Bessonard S, Mesnard D, Constam DB. PC7 and the related proteases Furin and Pace4 regulate E-cadherin function during blastocyst formation. *J Cell Biol*. Sep 28 2015;210(7):1185-97. doi:10.1083/jcb.201503042
29. Chen S, Cao P, Dong N, et al. PCSK6-mediated corin activation is essential for normal blood pressure. *Nat Med*. Sep 2015;21(9):1048-53. doi:10.1038/nm.3920
30. Perisic L, Hedin E, Razuvaev A, et al. Profiling of atherosclerotic lesions by gene and tissue microarrays reveals PCSK6 as a novel protease in unstable carotid atherosclerosis.

Arterioscler Thromb Vasc Biol. Oct 2013;33(10):2432-43.

doi:10.1161/ATVBAHA.113.301743

31. Rohl S, Suur BE, Lengquist M, et al. Lack of PCSK6 Increases Flow-Mediated Outward Arterial Remodeling in Mice. *Cells.* Apr 18 2020;9(4)doi:10.3390/cells9041009

32. Rykaczewska U, Suur BE, Rohl S, et al. PCSK6 Is a Key Protease in the Control of Smooth Muscle Cell Function in Vascular Remodeling. *Circ Res.* Feb 28 2020;126(5):571-585. doi:10.1161/CIRCRESAHA.119.316063

33. Kuhn TC, Knobel J, Burkert-Rettenmaier S, et al. Secretome Analysis of Cardiomyocytes Identifies PCSK6 (Proprotein Convertase Subtilisin/Kexin Type 6) as a Novel Player in Cardiac Remodeling After Myocardial Infarction. *Circulation.* May 19 2020;141(20):1628-1644. doi:10.1161/CIRCULATIONAHA.119.044914

34. Tsuji A, Sakurai K, Kiyokage E, et al. Secretory proprotein convertases PACE4 and PC6A are heparin-binding proteins which are localized in the extracellular matrix. Potential role of PACE4 in the activation of proproteins in the extracellular matrix. *Biochim Biophys Acta.* Jan 31 2003;1645(1):95-104. doi:10.1016/s1570-9639(02)00532-0

35. Nour N, Mayer G, Mort JS, et al. The cysteine-rich domain of the secreted proprotein convertases PC5A and PACE4 functions as a cell surface anchor and interacts with tissue inhibitors of metalloproteinases. *Mol Biol Cell.* Nov 2005;16(11):5215-26. doi:10.1091/mbc.e05-06-0504

36. Craig VJ, Zhang L, Hagood JS, Owen CA. Matrix metalloproteinases as therapeutic targets for idiopathic pulmonary fibrosis. *Am J Respir Cell Mol Biol.* Nov 2015;53(5):585-600. doi:10.1165/rcmb.2015-0020TR

37. Lin YE, Wu QN, Lin XD, Li GQ, Zhang YJ. Expression of paired basic amino acid-cleaving enzyme 4 (PACE4) correlated with prognosis in non-small cell lung cancer (NSCLC) patients. *J Thorac Dis.* May 2015;7(5):850-60. doi:10.3978/j.issn.2072-1439.2015.05.09

38. Lapierre M, Siegfried G, Scamuffa N, et al. Opposing function of the proprotein convertases furin and PACE4 on breast cancer cells' malignant phenotypes: role of tissue inhibitors of metalloproteinase-1. *Cancer Res.* Oct 1 2007;67(19):9030-4. doi:10.1158/0008-5472.CAN-07-0807

39. Longuespee R, Couture F, Levesque C, et al. Implications of Proprotein Convertases in Ovarian Cancer Cell Proliferation and Tumor Progression: Insights for PACE4 as a Therapeutic Target. *Transl Oncol.* May 9 2014;doi:10.1016/j.tranon.2014.04.008

40. Fradet L, Temmar R, Couture F, Belzile M, Fortier PH, Day R. Evaluation of PACE4 isoforms as biomarkers in thyroid cancer. *J Otolaryngol Head Neck Surg.* Oct 19 2018;47(1):63. doi:10.1186/s40463-018-0311-x

41. Delic S, Lottmann N, Jetschke K, Reifenberger G, Riemenschneider MJ. Identification and functional validation of CDH11, PCSK6 and SH3GL3 as novel glioma invasion-associated candidate genes. *Neuropathol Appl Neurobiol.* Apr 2012;38(2):201-12. doi:10.1111/j.1365-2990.2011.01207.x

42. Yao Z, Sun B, Hong Q, et al. PACE4 regulates apoptosis in human prostate cancer cells via endoplasmic reticulum stress and mitochondrial signaling pathways. *Drug Des Devel Ther.* 2015;9:5911-23. doi:10.2147/DDDT.S86881

43. Tian XF, Huang GM, Zang HL, Cao H. PACE4 regulates apoptosis in human pancreatic cancer Panc1 cells via the mitochondrial signaling pathway. *Mol Med Rep.* Dec 2016;14(6):5205-5210. doi:10.3892/mmr.2016.5885

44. Bassi DE, Fu J, Lopez de Cicco R, Klein-Szanto AJ. Proprotein convertases: "master switches" in the regulation of tumor growth and progression. *Mol Carcinog*. Nov 2005;44(3):151-61. doi:10.1002/mc.20134
45. Levesque C, Fugere M, Kwiatkowska A, et al. The Multi-Leu peptide inhibitor discriminates between PACE4 and furin and exhibits antiproliferative effects on prostate cancer cells. *J Med Chem*. Dec 13 2012;55(23):10501-11. doi:10.1021/jm3011178
46. Levesque C, Couture F, Kwiatkowska A, et al. PACE4 inhibitors and their peptidomimetic analogs block prostate cancer tumor progression through quiescence induction, increased apoptosis and impaired neovascularisation. *Oncotarget*. Feb 28 2015;6(6):3680-93. doi:10.18632/oncotarget.2918
47. Kwiatkowska A, Couture F, Ait-Mohand S, et al. Enhanced anti-tumor activity of the Multi-Leu peptide PACE4 inhibitor transformed into an albumin-bound tumor-targeting prodrug. *Sci Rep*. Feb 14 2019;9(1):2118. doi:10.1038/s41598-018-37568-6
48. Wang F, Wang L, Jiang H, Chang X, Pan J. Inhibition of PCSK6 may play a protective role in the development of rheumatoid arthritis. *J Rheumatol*. Feb 2015;42(2):161-9. doi:10.3899/jrheum.140435
49. 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*. May 24 2012;366(21):1968-77. doi:10.1056/NEJMoa1113354

Table 1. Baseline characteristics and outcomes of stage I and II datasets

Characteristic	Stage I (n=1481)			Stage II (n=397)
	US* (n=318)	UK** (n=565)	UUS*** (n=598)	Vanderbilt****
Age	67.3 (8.9)	72.1 (8.4)	69.8 (8.3)	65.7 (9.0)
Male sex	234 (73.6)	403 (71.3)	449 (75.1)	295 (74.3)
FVC, % predicted	62.7 (17.9)	77.7 (18.5)	70.2 (18.6)	65.0 (16.1)
DLCO, % predicted	36.7 (14.6)	43.3 (14.3)	39.9 (14.6)	39.1 (13.5)
GAP Stage				
I	72 (23.1)	106 (35.2)	135 (24.2)	115 (29.8)
II	149 (47.8)	144 (47.8)	286 (51.4)	201 (52.1)
III	91 (29.2)	51 (16.9)	136 (24.4)	70 (18.1)
Death	189 (59.4)	366 (64.8)	257 (43.0)	202 (50.9)
Transplant	52 (16.4)	2 (0.4)	26 (4.4)	32 (8.1)
Death or transplant	241 (75.8)	366 (64.8)	283 (47.3)	234 (58.9)
Median survival months (IQR)	39.3 (13.4-70.6)	53.2 (24.8-92.5)	40.6 (19.9-75.5)	48 (15-105)

* n for missing data: FVC (n=6); DLCO (n=31); GAP Stage (n=6)

** n for missing data: FVC (n=241); DLCO (n=264); GAP Stage (n=241)

*** n for missing data: FVC (n=30); DLCO (n=41); GAP Stage (n=30)

**** n for missing data: FVC (n=4); DLCO (n=9); GAP stage (n=4)

Table 2. Results for the SNPs nominally associated across stages and with same direction of effects on IPF survival														
Variant location (hg19) and nearest gene						Stage I results				Stage II results			Meta-analysis	
Chr	Position	SNP rsID	Gene	REF	EA	R²	EAF	HR [95% CI]	P-value	EAF	HR [95% CI]	P-value	HR [95% CI]	P-value
2	84291167	rs184498750	<i>SUCLG1</i>	G	T	0.86	1.1%	3.11 [1.88-5.15]	9.83x10 ⁻⁶	1.2%	2.05 [1.79-4.18]	0.049	2.71 [1.79-4.08]	2.07x10 ⁻⁶
15	76081200	rs60514164	<i>UBE2Q2</i>	C	T	0.77	7.0%	1.60 [1.32-1.95]	2.35x10 ⁻⁶	7.7%	1.42 [1.04-1.93]	0.026	1.55 [1.31-1.83]	2.23x10 ⁻⁷
15	101914234	rs35647788	<i>PCSK6</i>	C	T	0.93	0.8%	4.76 [2.62-8.64]	2.96x10 ⁻⁷	0.8%	3.12 [1.37-7.11]	6.72x10 ⁻³	4.11 [2.54-6.67]	9.45x10 ⁻⁹
19	1412985	rs3893252	<i>DAZAP1</i>	C	T	0.82	0.6%	3.57 [1.97-6.49]	2.91x10 ⁻⁵	1.5%	2.09 [1.05-4.15]	0.036	2.84 [1.81-4.45]	5.81x10 ⁻⁶

Abbreviations: Ref = reference allele; EA = effect allele; R²= lowest imputation quality value across studies; EAF = effect allele frequency; HR = hazard ratio; CI = confidence interval