Rare surfactant-related variants in familial and sporadic pulmonary fibrosis

Rachel M. Sutton1,2 | Humberto Trejo Bittar1,2 | Daniel I. Sullivan1,2 | Agustin Gil Silva1,2 | Harinath Bahudhanapati1,2 | Anishka H. Parikh1,2 | Yingze Zhang1,2,3 | Kevin Gibson1,2 | John F. McDyer2 | Daniel J. Kass1,2 | Jonathan K. Alder1,2

1Dorothy P. and Richard P. Simmons Center for Interstitial Lung Disease, University of Pittsburgh, Pittsburgh, Pennsylvania, USA
2Division of Pulmonary, Allergy and Critical Care Medicine, Pittsburgh, Pennsylvania, USA
3Department of Human Genetics, University of Pittsburgh, Pittsburgh, Pennsylvania, USA

Abstract
The role of constitutional genetic defects in idiopathic pulmonary fibrosis (IPF) is increasingly appreciated. Monogenic disorders associated with IPF affect two pathways: telomere maintenance, accounting for approximately 10% of all patients with IPF, and surfactant biology, responsible for 1%–3% of cases and often co-occurring with lung cancer. We examined the prevalence of rare variants in five surfactant-related genes, SFTPA1, SFTPA2, SFTPC, ABCA3, and NKX2-1, that were previously linked to lung disease in whole genome sequencing data from 431 patients with IPF. We identified functionally deleterious rare variants in SFTPA2 with a prevalence of 1.3% in individuals with and without a family history of IPF. All individuals had no personal history of lung cancer, but substantial bronchiolar metaplasia was noted on lung explants and biopsies. Five patients had novel missense variants in NKX2-1, but the contribution to disease is unclear. In general, patients were younger and had longer telomeres compared with the majority of patients with IPF suggesting that these features may be useful for identifying this subset of patients in the clinic. These data suggest that SFTPA2 variants may be more common in unselected IPF cohorts and may manifest in the absence of personal/family history of lung cancer or IPF.

Keywords
ABCA3, genetics, IPF, lung cancer, NKX2-1, SFTPA1, SFTPA2, SFTPC

1 | INTRODUCTION

Idiopathic pulmonary fibrosis (IPF) is defined as “a specific form of chronic, progressive, fibrosing interstitial pneumonia of unknown cause,” and is diagnosed by exclusion of alternative causes of interstitial lung disease (ILD) and the presence of characteristic imaging or histopathologic findings (Raghu et al., 2018). IPF is more commonly diagnosed in older adults and despite advances in management is associated with a poor prognosis (Lederer & Martinez, 2018). Historically IPF is associated with a median survival of only 2–years, however a significant number of patients will have survival greater than 5 years (Kim et al., 2015). While assigned a
single clinical moniker, IPF likely is composed of several distinct pathologies that have a common clinical presentation.

A number of genetic factors have been described that increase susceptibility to IPF. Genetic variants in telomere-related genes that encode the proteins responsible for maintaining and protecting the terminal sequences of chromosomes have been described in familial pulmonary fibrosis (Armanios et al., 2007), and are the most common identifiable Mendelian cause (Borie et al., 2017; van Moorsel et al., 2021). Ten percent of IPF patients without a family history also carry rare Mendelian variants in telomerase and other telomere-related genes (Dressen et al., 2018; Petrovski et al., 2017). Short telomeres in the absence of an identifiable causal genetic variant are also known to be a risk factor for the development of pulmonary fibrosis in patients with both familial as well as sporadic disease (Alder et al., 2008; Cronkhite et al., 2008).

IPF is also associated with defects in genes related to surfactant biology. Variants in genes encoding for surfactant protein C (SFTPC) and surfactant proteins A1 (SFTPA1) and A2 (SFTPA2) have been reported to cause pulmonary fibrosis in adults (Garcia, 2011; Nogee, 2019; van Moorsel et al., 2015). SFTPA1 and SFTPA2 are nearly identical paralogous genes that encode surfactant proteins SFTPA1 and SFTPA2, respectively. SFTPA1 and SFTPA2 are produced and secreted by type II alveolar epithelial and club cells into the alveolar and bronchial space, respectively. These proteins are classified as collectins and contain an N-terminal collagen-like and a C-type lectin binding domain with roles in innate immunity, surfactant structure, and regulation of surfactant production (Kishore et al., 2005). Variants in SFTPA2 in two large families with an autosomal dominant inheritance pattern of IPF and lung cancer were first reported in 2009 (Wang et al., 2009). Subsequently, additional reports of familial pulmonary fibrosis associated with variants in SFTPA1 and SFTPA2 variants have been reported (Coghlan et al., 2014; Doubková et al., 2019; Legendre et al., 2020; Liu et al., 2020, 2021; van Moorsel et al., 2015; Nathan et al., 2016; Sritharan et al., 2018; Takezaki et al., 2019), though the prevalence in large unselected cohorts and their clinical features are not fully described.

Variants in the lipid transporter, ABCA3, located at the limiting membrane of lamellar bodies, are typically associated with pediatric ILD and are inherited in an autosomal recessive manner (Beers & Mulugeta, 2017). Pediatric disease is typically caused by compound heterozygosity of different types of deleterious variants, but several cases of adult ILD caused by homozygous variants in ABCA3 have been reported (Campo et al., 2014; Coghlan et al., 2014; Tomer et al., 2021) as well as potential interactions between ABCA3 variants and variants in other surfactant-related genes (Bullard & Nogee, 2007; Crosso et al., 2010). At least two adult patients with IPF have been identified that were homozygous for the most common deleterious variant in ABCA3, p.(Glu292Val) (Campo et al., 2014; Coghlan et al., 2014). The full contribution of ABCA3 variants to IPF pathogenesis is not known.

NKK2-1 is a core transcription factor responsible for the regulation of surfactant genes and expressed during development of the thyroid, lung, and forebrain. NKK2-1 has been reported to encode two isoforms that differ by 30 amino acids, but the smaller 371 amino acid protein is far more abundant (Kolla et al., 2007; Li et al., 2000). Variants in NKK2-1 are associated with autosomal dominant brain-lung-thyroid syndrome and are most frequently null mutation (nonsense, frameshifting, splice-site, or large deletions) suggesting that haploinsufficiency is the mechanism responsible for the inheritance pattern. The most common clinical manifestations are benign chorea associated with hypothyroidism, however, there is considerable heterogeneity in clinical presentation, and some individuals present with only one body system affected. Pediatric ILD is the most prevalent lung finding (Hamvas et al., 2013; Thorwarth et al., 2014) though cases of adult-onset ILD have been reported in combination with lung cancer (Borie et al., 2021).

We recently conducted a large sequencing study of patients with predominantly sporadic IPF (Alder et al., 2022) and analyzed the prevalence of pathogenic variants in genes associated with telomere maintenance. In the current study we examined the prevalence of rare variants in surfactant-related genes, tested their functional consequences, and review their clinical characteristics.

2 | METHODS

2.1 | Subjects and study approval

The cohort of patients evaluated for this study was described previously (Alder et al., 2022). This study was approved by the institutional review board of the University of Pittsburgh. All subjects provided written, informed consent before enrollment in the study. Research subjects were recruited from the ILD specialty clinic or lung transplantation clinics at the University of Pittsburgh Medical Center (UPMC). All patients were diagnosed with idiopathic pulmonary fibrosis according to consensus guidelines of the American Thoracic Society and European Respiratory Societies at the time of their enrollment (American Thoracic Society & European Respiratory Society, 2000) (Raghu et al., 2011). Collected patient characteristics including sex, age, age at diagnosis, family history, pulmonary function, imaging, and survival were collected via retrospective chart review.

2.2 | Genetic analysis

Whole genome sequencing (WGS) was performed on all samples included in this study. Three hundred and ninety-one samples were sequenced at the UPMC genome center and 40 samples were sequenced at Knome (now PierianDX). Control samples were provided by the UPMC Genome Center and patients with ILD were excluded. All samples where average coverage of target genes were less than 10x were excluded from this analysis. We focused our analysis on five genes associated with surfactant biology that have been previously reported in adult patients with IPF; SFTPC, SFTPA1, SFTPA2, ABCA3, and NKK2-1. Variants analyzed included those with a
minor allele frequency (MAF) <0.0001 in the gnomAD database (v2.1.1) that were nonsynonymous exonic variants or splice-site variants. The term rare variant includes any variant meeting the above criteria. These rare variants were then further classified using the American College of Medical Genetics (ACMG) guidelines as variants of uncertain significance (VUS), likely pathogenic or pathogenic. The ACMG guidelines also include the classification of likely benign and benign variants, however, no variants met that classification due to the filtering of rare exonic or splice-site variants (Richards et al., 2010). All variants met the criteria PM2, given the frequency was less than 0.01% and it is unknown if reported variants in gnomAD have lung disease. Variants met the pathogenic strong (PS3) criteria if the protein product of each variant genes failed to be secreted in our in vitro assay. PP3 was included only if SIFT (Ng & Henikoff, 2003) and PolyPhen-2 (Adzhubei et al., 2010) agreed that the variant was predicted to be damaging. Conversely, BP4 was applied when SIFT and polyphen agreed that the variant was predicted to be tolerated. All of the SFTPA2 variants were located in mutational hotspot so PM1 was applied. Patients harboring the SFTPA2 p.(Val178Met) variant were classified with PP5 as this variant in mutational hotspot so PM1 was applied. Patients harboring the variant was predicted to be tolerated. All of the variants that the protein product of each variant genes failed to be secreted in our in vitro assay. PP3 was included only if SIFT (Ng & Henikoff, 2003) and PolyPhen-2 (Adzhubei et al., 2010) agreed that the variant was predicted to be damaging. Conversely, BP4 was applied when SIFT and polyphen agreed that the variant was predicted to be tolerated. All of the SFTPA2 variants were located in mutational hotspot so PM1 was applied. Patients harboring the SFTPA2 p.(Val178Met) variant were classified with PP5 as this variant has been reported by two independent groups. We tested for relatedness among individuals that shared the same variant using KING (Manichaikul et al., 2010). Conservations studies were performed by retrieving the proteins sequence for SFTPA1, SFTPA2, and NKX2-1 from each of the indicated species from RefSeq and aligning the peptide sequences using Clustal Omega.

2.3 | Telomere length analysis
Telomere length was estimated using WGS data and the Telseq software package (Ding et al., 2014). A nomogram of telomere lengths was established as described previously (Alder et al., 2022).

2.4 | Histologic analysis
Histologic samples were available for review from three of the patients. One of the samples with a variant in SFTPA2 p.(Tyr186Cys) was from an explanted lung while the other two were from biopsies taken during clinical care. Multiple slides from each individual were reviewed by a board-certified cardiothoracic pathologist.

2.5 | Expression constructs, tissue culture, and transfection
The expression constructs for epitope-tagged SFTPA2 were described previously (Wang et al., 2009). A gBlock with each of the variants we identified was synthesized at IDT and then incorporated into the expression plasmid using a Gibson Assembly (Gibson et al., 2009). Each plasmid was sequence verified. A549 cells were acquired from the American Type Culture Collection (ATCC) and propagated in Dulbecco's modified Eagle medium (DMEM) supplemented with 10% fetal bovine serum (FBS) and penicillin, streptomycin, and l-glutamine. Cells were transfected with Fugene HD Transfection reagent (Roche). Forty-eight hours after transfection, cells and supernatant were collected for Western blot analysis. Fifteen microliters of supernatant was mixed directly with loading buffer and 20 μg of cell lysates, extracted in radioimmunoprecipitation assay buffer, were separated on Any kD gels (Bio-Rad) and blotted using standard procedures. Antibodies for V5 (Thermo Fisher Scientific), LAMC1 (Sigma-Aldrich), and Tubulin (Bio-Rad) were used at 1:1000 overnight before imaging (Chemidoc MP; Bio-Rad).

2.6 | Luciferase reporter assays
We generated a human SFTP8 reporter construct by synthesizing 838 bps of the proximal promoter (starting at the +1 position) and assembling into the pGL4.10[luc2] vector. Five NKX2-1 expression constructs were made by synthesizing the wild-type and four variant versions of the 371 amino acid isoform of human NKX2-1 and assembling into the pCDNA3.1 expression vector. All synthesis was performed at Twist Biosciences. The reporter and expression constructs were cotransfected, along with a Renilla transfection control plasmid, into HEK293T cells (ATCC; cultured in DMEM plus 10% FBS) using lipofectamine 2000 (Thermo Fisher Scientific) according to the manufacturer's recommendations. Cell lysates were collected 48 h later and the activity of firefly and Renilla luciferase was measured using the Promega Dual-Glo luciferase assays system according to the manufacturer's protocol.

3 | RESULTS
3.1 | Identification of surfactant protein variants
To test the overall prevalence of SFTPC, SFTPA1, and SFTPA2 variants we examined their coding sequences in whole genome data from a cohort of 431 patients with a diagnosis of IPF (Alder et al., 2022). No rare variants were identified in SFTPC, but four rare (MAF <0.0001) nonsense variants in SFTPA2 and one in SFTPA1 were found across seven patients (Table 1). One of the variants, SFTPA2, p.(Val178Met), was found in two patients and was previously reported in at least two families with IPF (Coughlan et al., 2014; Legendre et al., 2020). The other four variants were novel, and their functional effects were unknown. No relationship was identified between individuals that shared the same variant by family history or genetic analysis (Section 2). Due to the highly homologous sequences in SFTPA1 and SFTPA2, we confirmed each of the SFTPA2 variants with long-range polymerase chain reaction (PCR) and Sanger sequencing (data not shown) (Horton et al., 2013). Each of the variants fell within the carbohydrate-binding domain where all previously reported variants are localized and showed variable levels of conservation (Figure 1a,b). In each of the patients with variants in SFTPA1/2, we evaluated if
TABLE 1  Surfactant-associated rare variants identified in this SFTPA1 and SFTPA2

| ID | cDNA            | Protein          | gnomAD MAF     | ACMG classification | Previously reported                                      |
|----|-----------------|------------------|---------------|---------------------|--------------------------------------------------------|
| 421| c.460G>C        | p.(Ala154Pro)    | 0.00004633    | Likely pathogenic   | Not reported                                           |
| 579| c.530T>A        | p.(Phe177Tyr)    | 0.00002391    | VUS                 | Not reported                                           |
| 417| c.532G>A        | p.(Val178Met)    | 0.00001195    | Pathogenic          | Coghlan et al. (2014), Legendre et al. (2020)          |
| 563| c.532G>A        | p.(Val178Met)    | 0.00001195    | Pathogenic          | Coghlan et al. (2014), Legendre et al. (2020)          |
| 111| c.557A>G        | p.(Tyr186Cys)    | 0.00000398    | Likely pathogenic   | Not reported                                           |
| 014| c.557A>G        | p.(Tyr186Cys)    | 0.00000398    | Likely pathogenic   | Not reported                                           |
| 024| c.656G>T        | p.(Arg219Leu)   | –             | VUS                 | Not reported                                           |

Abbreviations: ACMG, American College of Medical Genetics; MAF, minor allele frequency; SFTPA1/2, surfactant proteins A1/2; VUS, variant of unknown significance.

Variants are annotated on the NM_001098668.4 isoform of SFTPA2.

Variant located in SFTPA1.

FIGURE 1  Novel variants in SFTPA2 cluster with previously reported variants and disrupt secretion of the protein. (a) Schematic of SFTPA2 showing annotated domains and the location of previously reported (black, below) and novel (red, above) variant in individuals with pulmonary fibrosis. SS, signal sequence. (b) Alignment of protein sequences from the carbohydrate recognition domain showing the location of the novel variants reported here and their respective conservation. (c) Western blot of media and whole cell lysates from A549 cells that had been transfected with the indicated SFTPA2 expression constructs. Each of the constructs carried a V5-epitope tag to facilitate Western blot analysis and tubulin was used as a load control for cell lysates and LAMC1 was used as a loading control for supernatants. (d) Graph of age at diagnosis from IPF patients with no identifiable genetic variants (IPF-unknown), pathogenic or likely-pathogenic variants in telomere maintenance genes (Telomere), and patients with variants in SFTPA2. Groups were compared with the Kruskal–Wallis test. *p < 0.05 and **p < 0.01. (e) Telomere lengths of patients with variants in SFTPA2 plotted on a nomogram with percentile indicated on the right of the graph. Patient with TERC VUS and TERT promoter variant is indicated with *. IPF, idiopathic pulmonary fibrosis; SFTPA2, surfactant proteins A2; VUS, variant of unknown significance.
they had any variants in telomere-related genes as these are the most common cause of IPF. One of the patients with SFTPA2 p.(Tyr186Cys) variant also harbored a VUS in TERC (r.67g>u), a component of telomerase. The same patient also had a TERT promoter variant that has been reported in IPF patients with telomere-mediated IPF (Maryoung et al., 2017).

3.2 Novel SFTPA2 variants disrupt protein secretion

We next sought to functionally characterize the variants we had identified. Previously reported variants in SFTPA2 have been shown to block secretion of the mature protein (Wang et al., 2009). To examine if the five variants we had identified were functional, we cloned each of the variants into an expression vector and transfected them into A549 cells, a lung adenocarcinoma-derived cell line that has been previously shown to support secretion of this protein (Wang et al., 2009). The wild-type version of SFTPA2 was robustly expressed and secreted into the media. However, cells expressing the p.(Ala154Pro), p.(Val178Met), and p.(Tyr186Cys) variants showed abundant protein in cell lysates but had significantly less or completely lacked secreted protein in the media collected from the same cells (Figure 1c). The SFTPA2 p.(Phe177Tyr) and SFTPA1 p.(Arg219Leu) variants were secreted into the media and were not investigated further given their unknown functional significance.

3.3 Clinical characteristics of patients with SFTPA2 variants

We next evaluated the clinical characteristics of each of the individuals found to carry a functional variant (Table 2). The patient with the SFTPA2 p.(Ala154Pro) variant was diagnosed as an elderly male. He had no family history of pulmonary fibrosis and presented with CT findings consistent with usual interstitial pneumonia (UIP) (Figure 2a). The remaining four patients, two each with the SFTPA2 p.(Val178Met) and p.(Tyr186Cys) variants, presented between 37 and 56 years of age (Figure 1d). The two patients with SFTPA2 p.(Val178Met) variants and one patient with the p.(Tyr186Cys) variant reported a family history of ILD, while the other patient with SFTPA2 p.(Tyr186Cys) reported no family history of ILD at the time of diagnosis. The latter patient also had a VUS in TERC and one of the putative TERT promoter variants. All four patients with either of the p.(Val178Met) or p.(Tyr186Cys) variants had computed tomography (CT) images that were inconsistent with a UIP pattern (Raghu et al., 2018), with each demonstrating predominantly findings of ground glass opacities without honeycombing or a clear subpleural and basilar predominance (Figure 2a). Because imaging studies were not consistent with UIP, all four of these patients underwent lung biopsy with histopathology demonstrating a UIP pattern in all four cases. Both samples from the patients with p.(Tyr186Cys) variant showed extensive areas of parenchymal scarring and fibroblastic activity in subpleural (in keeping with the UIP pattern) and peribroncholar (Figure 2b) regions. In addition, one each with SFTPA2 p.(Tyr186Cys) and SFTPA2 p.(Val178Met) demonstrated significant bronchiolar epithelial metaplasia (Figure 2c,d). No granulomas, vasculitis, or mineral dust were observed. Qualitatively, no significant increases in chronic inflammatory cell infiltrates were noted. We examined the telomere length of patients with SFTPA2 variants to test if this test could be used to distinguish these patients and found that all five patients had telomere lengths above the 10th percentile including one patient that carried a variant of unknown significance in TERC (r.67g>u) and a TERT promoter variant. Ultimately two of the three patients had an acute exacerbation while awaiting lung transplantation and died, while the third was successfully transplanted.

3.4 Identification of variants in surfactant-related genes

Variants in ABCA3 and NKX2-1 are typically associated with pediatric lung disease (Nathan et al., 2018), but have been reported to manifest rarely in adults (Borie et al., 2021; Campo et al., 2014). We identified several rare and novel variants in ABCA3 including previously reported pathogenic variants (Supporting Information: Table S1). None of the patients in our cohort were homozygous or compound

| ID | Variant | Sex | Age at onset | Age at diagnosis | Exposures | Family history |
|----|---------|-----|--------------|-----------------|-----------|----------------|
| 421 | p.(Ala154Pro) | M | 73 | 74 | None | None |
| 563 | p.(Val178Met) | F | Early 20s | 37 | 14-pack year former smoker | Father—unknown ILD age 41 |
| 417 | p.(Val178Met) | F | 34 | 42A | 7-pack year former smoker | Multiple family members with ILD |
| 111 | p.(Tyr186Cys) | M | Late 30s | 42 | 20-pack year former smoker | Father—Lung cancer age 72, Mother—Pulmonary fibrosis age 83 |
| 014 | p.(Tyr186Cys) | M | 56 | 56 | None | None |

Abbreviations: F, female; M, male; ILD, interstitial lung disease; SFTPA2, surfactant proteins A2.
heterozygous for rare variants in ABCA3. We found a similar burden of ABCA3 rare variants in our cohort of IPF patients compared to a geographically similar cohort of non-ILD controls (14 out of 431 IPF vs. 16 out of 444 controls; \( p = 0.85 \), Fisher’s exact test). Given the unclear contribution of these variants to disease, we did not investigate them further.

Disease-associated alleles of NKX2-1 are most often large deletions, frame-shifting, or nonsense, but missense variants have also been identified. We identified five novel missense changes in our cohort of IPF patients in distinct regions of the coding sequence (Table 3 and Figure 3a). One of the variants fell in a nonconserved portion of the long isoform, but the other four variants were in conserved amino acids across the coding sequence including one in the homeodomain near where other missense mutations have been reported (Figure 3a,b). No rare variants were identified in a non-ILD control cohort of 444 patients (5 out of 431 IPF vs. 0 out of 444 controls; \( p = 0.03 \), Fisher’s exact test). None of the patients had variants in telomere-related or other known IPF-associated genes.
Functional analysis of NKX2-1 rare variants

NKX2-1 is a transcription factor that regulates the expression of several lung-associated genes including SFTPB and SFTPC. To test the functional consequences of the variants we identified, we cloned ~800 nucleotides of the SFTPB proximal promoter into a luciferase reporter construct and tested if the variants we identified disrupted the capacity of NKX2-1 to activate transcription of SFTPB. Since the 371 amino acid isoform is the dominant form of NKX2-1, we tested each of the corresponding variants on the short isoform and did not test the p.(Gly23Arg) variant as it is located in the long isoform. Despite their strong conservation, none of the variants we identified interfered with the capacity of NKX2-1 to activate transcription of SFTPB (Supporting Information: Figure S1A).

### Table 3

| Patient ID | cDNA | Protein | gnomAD MAF | ACMG classification | Previously reported |
|------------|------|---------|------------|--------------------|--------------------|
| 213        | c.67G>C | p.(Gly23Arg) | - | VUS | Not reported |
| 104        | c.396C>A | p.(Asp132Glu) | - | VUS | Not reported |
| 228        | c.532G>A | p.(Asp178Asn) | - | VUS | Not reported |
| 523        | c.631A>G | p.(Lys211Glu) | - | VUS | Not reported |
| 057        | c.781C>A | p.(Gln261Lys) | - | VUS | Not reported |

Abbreviations: ACMG, American College of Medical Genetics; cDNA, complementary DNA; MAF, minor allele frequency; VUS, variant of unknown significance.

*Variants are annotated on the NM_001079668.3 isoform of NKX2-1.
Clinical characteristics of patients with NKX2-1 variants

Patients with variants in NKX2.1 presented at a wide age range (43–73 years) and with diverse clinical presentations (Table 4). None of the patients with NKX2-1 variants reported a family history of ILD although clinical histories were likely focused on pulmonary phenotypes. Two of the patients, p.(Gly23Arg) and p.(Asp178Asn), had CT scans that were consistent with UIP while the others had features that were either possibly suggestive of or inconsistent with UIP. The patient with the p.(Asp132Glu) variant had a significant smoking history and presented with combined emphysema and fibrosis and underwent lung transplantation. Three of the patients had biopsy or explant pathology available and all reported to show UIP (p.(Gly23Arg) and p.(Lys211Glu) or UIP and emphysema (p.(Asp132Glu)). Only one of the patients was taking medications for hypothyroidism (p.(Gly23Arg)) and underwent lung transplantation. Three of the patients had biopsy or explant pathology available and all reported to show UIP (p.(Gly23Arg) and p.(Lys211Glu) or UIP and emphysema (p.(Asp132Glu)). Only one of the patients was taking medications for hypothyroidism (p.(Gly23Arg)) to our knowledge. We examined the telomere length based on WGS data to determine if this could be used to distinguish these patients and found that all five had telomere lengths above the 10th percentile (Supporting Information: Figure S1C).

4 | DISCUSSION

Here we report missense variants in SFTPA2 and NKX2-1 in 11 patients (2.5% of cohort) from a cohort of predominantly sporadic IPF. Functional assessment confirmed that three variants, identified in five patients, resulted in disrupted SFTPA2 secretion, which supports their pathogenicity. Two of these variants are novel and reported here for the first time, while the third variant was previously reported (Coghlan et al., 2014; Legendre et al., 2020). All three variants are located in C-lectin binding domain located in exon 6 of SFTPA2, the same domain affected by all prior known pathogenic variants. Although one individual had variants in multiple genes (SFTPA2 and TERC), none of the other patients had variants in additional genes associated with IPF. The TERC variant was classified as a VUS and the telomeres of this patient were normal length, thus the contribution of telomere dysfunction to the development of lung disease in this patient is unknown. Three of the patients with SFTPA2 variants met the definition of familial pulmonary fibrosis at the time of diagnosis, and two were classified as sporadic IPF. 1.3% of all IPF patients sequenced, including sporadic and familial disease, carried rare variants in SFTPA2 and our data further support that surfactant defects are responsible for a portion of patients with IPF.

The contribution of NKX2-1 variants we identified in our cohort to disease pathogenesis is unclear. Disease-causing NKX2-1 alleles are typically loss-of-function and prevent expression of the full-length protein. Disease-causing missense mutations typically cluster in the DNA-binding domain (homeodomain). Only one of the five missense variants we identified was located in the DNA-binding domain in a patient that developed disease at a young age (43 years). Despite considerable conservation of four of the variants, we could detect no functional differences in the expressed proteins. There were no family histories of lung disease reported in any of the individuals with NKX2-1 variants, but it is possible that our histories missed clinical features outside the lung that are associated with NKX2-1 deficiency. While it is tempting to speculate that the missense variants we identified are hypomorphic alleles, additional studies will be required to determine if the rare variants we identified contribute to lung pathology.

To date, only one patient with sporadic IPF due to a SFTPA2 variant has previously been reported in a study that performed sequencing on 118 sporadic cases (van Moorsel et al., 2015). In our retrospective analysis we identified two additional patients with sporadic IPF and pathogenic SFTPA2 variants. DNA from parents or siblings of these probands was unavailable to determine if these variants were de novo or if the variants show incomplete penetrance and were inherited from an unaffected parent. This finding highlights the possibility of a genetic etiology even in cases of apparently sporadic disease.

The patients with SFTPA2 variants described here demonstrate significant clinical heterogeneity. Age at diagnosis ranged from 37 to 74 years old, and progression to death or transplantation varied from 2 to 10 years. Interestingly, three patients that had a history of smoking presented at a younger age than the those that did not. Only one patient had a CT scan consistent with UIP, while the other four had CT scans inconsistent with UIP. In contrast to previous studies with larger families (Wang et al., 2009), none of the patients had a personal history of lung cancer and only one reported a first-degree relative with lung cancer. However, two of the five patients had

| ID  | Variant       | Sex | Age at onset | Age at diagnosis | Exposures                                   | Family history |
|-----|---------------|-----|--------------|------------------|---------------------------------------------|----------------|
| 213 | p.(Gly23Arg)  | F   | 46           | 54               | 30-pack year former smoker                  | None           |
| 104 | p.(Asp132Glu) | M   | 64           | 64               | 105-pack year former smoker                 | None           |
| 228 | p.(Asp178Asn) | M   | 70           | 72               | None                                        | None           |
| 523 | p.(Lys211Glu) | F   | 43           | 47               | None                                        | None           |
| 057 | p.(Gln211Lys) | M   | 73           | 73               | 27-pack year former smoker                  | None           |

Note: Family history limited in scope to lung disease.
Abbreviations: F, female; M, male.

TABLE 4 Clinical characteristics of NKX2-1 variants
substantial bronchiolar epithelial metaplasia, a preneoplastic change that appears consistent with previous reports of cancer predisposition (Jensen-Taubman et al., 1998). Given the heterogeneous presentation, it is difficult to definitively identify these patients based on clinical findings, however, young age, personal or family history of lung cancer, and normal telomeres likely should increase the index of suspicion for surfactant-mediated disease.

Besides important implications for the family members of affected patients, knowledge of a SFTPA2 variant can have important implications for disease management. One patient reported here with a SFTPA2 p.Y186C variant underwent unilateral lung transplantation. If this had been known in advance, he likely would have been prioritized for bilateral lung transplant given the association with lung cancer (Wang et al., 2009; Yoon et al., 2018). There has been one previously reported patient with an SFTPA2 p.N210T variant who underwent single lung transplant and then passed away less than 2 years later with metastatic lung adenocarcinoma (van Moorsel et al., 2015). The findings, together with our previous report (Alder et al., 2022), support genetic screening for patients with a possible underlying genetic etiology for IPF in the context of lung transplantation.

The pathologic mechanism responsible for disease in patients with SFTPA2 and potentially NKX2-1 is unclear. Our functional studies of SFTPA1 and SFTPA2 were limited to testing for the ability of proteins to be secreted from cells. It is possible that some mutations disrupt other aspects of surfactant biology that are not captured in our functional assays. Similarly, our analysis of NKX2-1 only tested the capacity of the protein to activate a SFTPβ-luciferase reporter. Given that disease requires decades to develop, it is possible that the pathogenic mechanism is subtle and may not be captured in our functional studies. The mechanism responsible for the delayed onset remains elusive and additional investigation is warranted to understand how surfactant dysfunction leads to adult-onset disease.

Many of the patients identified in this study presented with features that might suggest a diagnosis other than IPF. Most patients had CT scans that were inconsistent with UIP. Review of their medical records showed that the diagnosing physicians recognized that these were difficult to classify cases and hence many underwent surgical lung biopsy (four out of five SFTPA2 and one out of five NKX2-1). Taken together, our findings suggest that younger patients that present with ILD may be enriched for individuals with an underlying genetic etiology and consultation with a genetics provider may be warranted.

In conclusion, we report the identification of five patients with functional missense variants in SFTPA2 identified by WGS of 431 patients with IPF at a single tertiary referral center. We report two out of five patients with a history consistent with sporadic IPF at the time of their diagnosis further supporting that variants in surfactant protein A genes might be a minor, but underrecognized contributor to sporadic IPF. Our study is limited in that family histories were extracted by medical record review rather than a formal family history taken by a trained geneticist or genetic counselor. Due to significant clinical heterogeneity it appears impossible to identify surfactant-mediated IPF based on clinical, radiographic, or histopathologic characteristics alone. A CT scan with ground-glass opacities that is inconsistent with UIP appears to be more common based on our data and previous reports, but a typical UIP pattern is also possible. Young age and normal telomere length may also support further investigation of surfactant genes. Furthermore, histopathology with prominent fibroblast activity outside of areas of subpleural scarring, as well as abundant bronchial epithelial metaplasia might suggest SFTPA2 variant, but were not seen in all specimens. Given important clinical implications and improving availability and cost, genetic testing of patients with IPF might be a useful tool that should be considered to identify more patients with disease due to variants in SFTPA1/2.

ACKNOWLEDGMENTS
We are extraordinarily grateful to all the patients who agreed to participate in this study. Without their participation, none of this could have happened. We are also grateful to the contributions made by Dr. Kevin Misner during his fellowship training. We also acknowledge the help and support from the staff of the UPMC Genome Center. This project was supported by National Heart, Lung, and Blood Institute Grant R01HL135062 (JA), R01 HL133184 (JM), Kathleen Ford Fund for Lung Transplant Research, Richard P. Simmons Chair for Interstitial Lung Disease, and the Simmons Genetics Research Fund.

CONFLICT OF INTEREST
D. J. Kass reports collaborative research funding from Regeneron Pharmaceuticals in pulmonary hypertension, which is unrelated to this article. K. F. Gibson reports membership on the advisory board of Bayer Pharmaceuticals, outside the scope of the submitted work.

ORCID
Jonathan K. Alder (http://orcid.org/0000-0003-3741-6512)

REFERENCES
Adzhubei, I. A., Schmidt, S., Peshkin, L., Ramensky, V. E., Gerasimova, A., Bork, P., Kondrashov, A. S., & Sunyaev, S. R. (2010). A method and server for predicting damaging missense mutations. Nature Methods, 7(4), 248–249. https://doi.org/10.1038/nmeth0410-248
Alder, J. K., Chen, J. J.-L., Lancaster, L., Danoff, S., Su, S., Cogan, J. D., Vulto, I., Xie, M., Qi, X., Tuder, R. M., Phillips, J. A., Lansidorp, P. M., Loyd, J. E., & Armanios, M. Y. (2008). Short telomeres are a risk factor for idiopathic pulmonary fibrosis. Proceedings of the National Academy of Sciences USA, 105(35), 13051–13056. https://doi.org/10.1073/pnas.0804280105
Alder, J. K., Sutton, R. M., Lasella, C. J., Nouraie, M., Koshy, R., Hannan, S. J., Chan, E. G., Chen, X., Zhang, Y., Brown, M., Popescu, I., Veatch, M., Saul, M., Berndt, A., Methé, B. A., Morris, A., Pilewski, J. M., Sanchez, P. G., Morrell, M. R., ... McDyer, J. F. (2022). Lung transplantation for idiopathic pulmonary fibrosis enriches for individuals with telomere-mediated disease. The Journal of Heart and Lung Transplantation, 41(5), 654–663. https://doi.org/10.1016/j.healun.2021.11.008
American Thoracic Society, & European Respiratory Society. (2000). Idiopathic pulmonary fibrosis: Diagnosis and treatment: International Consensus Statement. American Journal of Respiratory and Critical Care Medicine, 161(2 Pt 1), 646–664. https://doi.org/10.1164/ajrccm.161.2.ats3-00
Armanios, M. Y., Chen, J. J.-L., Cogan, J. D., Alder, J. K., Ingersoll, R. G., Markin, C., Lawson, W. E., Xie, M., Vulto, I., Phillips, J. A., Lansdorp, P. M., Greider, C. W., & Loyd, J. E. (2007). Telomerase mutations in families with idiopathic pulmonary fibrosis. New England Journal of Medicine, 356(13), 1317–1326. https://doi.org/10.1056/NEJMoa066157

Beers, M. F., & Mulugeta, S. (2017). The biology of the ABCA3 lipid transporter in lung health and disease. Cell and Tissue Research, 367(3), 481–493. https://doi.org/10.1007/s00441-016-2554-z

Borie, R., Funalot, B., Epaud, R., Delestrain, C., Cazes, A., Gounant, V., Frija, J., Debray, M. P., Zalcman, G., & Crestani, B. (2021). NFKX2.1 (TTF1) germline mutation associated with pulmonary fibrosis and lung cancer. ERJ Open Research, 7(4), 00356-2021. https://doi.org/10.1183/23122541.00356-2021

Borie, R., Kannegiesser, C., Sicre de Fontbrune, F., Gouya, L., Nathan, N., & Crestani, B. (2017). Management of suspected monogenic lung fibrosis in a specialised centre. European Respiratory Review, 26(144), 160122. https://doi.org/10.1183/16000617.0122-2016

Bullard, J. E., & Nogee, L. M. (2007). Heterozygosity for ABCA3 mutations modifies the severity of lung disease associated with a surfactant protein C gene (SFTPC) mutation. Pediatric Research, 62(2), 176–179. https://doi.org/10.1203/PDR.0b013e3180a72588

Campo, I., Zorzetto, M., Mariani, F., Kadija, Z., Morbini, P., Dore, R., Lantsorp, P. M., Greider, C. W., & Loyd, J. E. (2007). Telomerase mutations in families with idiopathic pulmonary fibrosis. New England Journal of Medicine, 356(13), 1317–1326. https://doi.org/10.1056/NEJMoa066157

Frontiers in Cellular and Molecular Biology, 2021, 1–17. https://doi.org/10.3389/fcimb.2021.683105

Gibson, D. G., Young, L., Chuang, R.-Y., Venter, J. C., Hutchison, C. A., & Smith, H. O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. Nature Methods, 6(5), 343–345. https://doi.org/10.1038/nmeth.1318

Hamvas, A., Deterding, R. R., Wert, S. E., White, F. V., Dishop, M. K., Alfano, D. N., Halfower, A. C., Planer, B., Stephan, M. J., Uchida, D. A., Williams, L. D., Rosenfeld, J. A., Lebel, R. R., Young, L. R., Cole, F. S., & Nogue, L. M. (2013). Heterogeneous pulmonary phenotypes associated with mutations in the thyroid transcription factor gene NFKX2.1. Chest, 144(3), 794–804. https://doi.org/10.1378/chest.12-2502

Horton, C. J., Mitui, M., Leos, N. K., Garcia, C. K., & Park, J. Y. (2013). Long-range PCR based sequencing of the highly homologous genes, SFTPA1 and SFTPA2. Molecular and Cellular Probes, 27(3–4), 115–117. https://doi.org/10.1016/j.mcp.2013.01.001

Jensen-Taubman, S. M., Steinberg, S. M., & Linnola, R. I. (1998). Bronchiolization of the alveoli in lung cancer: Pathology, patterns of differentiation and oncogene expression. International Journal of Cancer, 75(4), 489–496. https://doi.org/10.1002/1097-0215(19980209)75:4%3C489::AID-IJC1%3E3.0.CO;2-P

Kim, H. J., Perlman, D., & Tomic, R. (2015). Natural history of idiopathic pulmonary fibrosis. Respiratory Medicine, 109(6), 661–670. https://doi.org/10.1016/j.rmed.2015.02.002

Kishore, U., Bernal, A. L., Kamran, M. F., Saxena, S., Singh, M., Sarma, P. U., Madan, T., & Chakraborty, T. (2005). Surfactant proteins SP-A and SP-D in human health and disease. Archivum Immunologiae et Therapiae Experimentalis, 53(5), 399–417.

Kolla, V., Gonzales, L. W., Gonzales, J., Wang, P., Angampalli, S., Feinstein, S. I., & Ballard, P. L. (2007). Thyroid transcription factor in differentiating type II cells: Regulation, isoforms, and target genes. American Journal of Respiratory and Cellular Biology, 36(2), 213–225. https://doi.org/10.1016/j.rccm.2006-02070C

Lederer, D. J., & Martínez, F. J. (2018). Idiopathic pulmonary fibrosis. New England Journal of Medicine, 378(19), 1811–1823. https://doi.org/10.1056/NEJMra1705751

Legendre, M., Butt, A., Borie, R., Debray, M.-P., Browdy, D., Filhol-Blin, E., Desroziers, T., Nau, V., Copin, B., Dastot-Le Moal, F., Héry, M., Duquesnoy, P., Allou, N., Bergeron, A., Bermudez, J., Cazes, A., Chene, A. L., Cottin, V., Crestani, B., … Nathan, N. (2020). Functional assessment and phenotypic heterogeneity of SFTPA1 and SFTPA2 mutations in interstitial lung diseases and lung cancer. European Respiratory Journal, 56(6), 2002806. https://doi.org/10.1183/13993003.02806-2020

Li, C., Cai, J., Pan, Q., & Mino, P. (2000). Two functionally distinct forms of NFKX2.1 protein are expressed in the pulmonary epithelium. Biochemical and Biophysical Research Communications, 270(2), 462–468. https://doi.org/10.1006/bbrc.2000.2443

Liu, L., Liu, Y.-J., Guo, T., & Luo, H. (2021). Identification of a missense mutation in the surfactant protein A2 gene in a Chinese family with interstitial lung disease. DNA and Cell Biology, 40(1), 126–131. https://doi.org/10.1089/dna.2020.6045

Liu, L., Qin, J., Guo, T., Chen, P., Ouyang, R., Peng, H., & Luo, H. (2020). Identification and functional characterization of a novel surfactant protein A2 mutation (p.N207Y) in a Chinese family with idiopathic pulmonary fibrosis. Molecular Genetics & Genomic Medicine, 8(9), 1393. https://doi.org/10.1002/mgge.1393

Manichaikul, A., Mychaleckyj, J. C., Rich, S. S., Daly, K., Sale, M., … Chen, W.-M. (2010). Robust relationship inference in genome-wide association studies. Bioinformatics, 26(22), 2867–2873. https://doi.org/10.1093/bioinformatics/bts559

Maryoung, L., Yue, Y., Young, A., Newton, C. A., Barba, C., van Oers, N. S. C., Wang, R. C., & Garcia, C. K. (2017). Somatic mutations in telomerase promoter counterbalance germline loss-of-function mutations. Journal of Clinical Investigation, 127(3), 982–986. https://doi.org/10.1172/JCI91161
van Moorsel, C. H. M., Ten Klooster, L., van Oosterhout, M. F. M., de Jong, P. A., Adams, H., Wouter van Es, H., Ruven, H. J. T., van der Vis, J. J., & Grutters, J. C. (2015). SFTPA2 mutations in familial and sporadic idiopathic interstitial pneumonia. *American Journal of Respiratory and Critical Care Medicine*, 192(10), 1249–1252. https://doi.org/10.1164/ajrccm.201504-0675LE

van Moorsel, C. H. M., Ten Klooster, L., van Oosterhout, M. F. M., Nathan, N., Borensztajn, K., & Clement, A. (2018). Genetic causes and mechanisms of idiopathic pulmonary fibrosis: A joint evidence-based approach. *Respiratory and Critical Care Medicine*, 192(2), 1457–1467. https://doi.org/10.1164/rccm.201802-2784OC

Ng, P. C. (2003). SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Research*, 31(13), 3812–3814. https://doi.org/10.1093/nar/gkg509

Nogee, L. M. (2019). Genetic causes of surfactant protein abnormalities. *Current Opinion in Pediatrics*, 31(3), 330–339. https://doi.org/10.1097/MOP.0000000000001751

Petrovski, S., Todd, J. L., Durheim, M. T., Wang, Q., Chien, J. W., Kelly, F. L., Frankel, C., Mebane, C. M., Ren, Z., Bridgers, J., Urban, T. J., Malone, C. D., Finlen Copeland, A., Brinkley, C., Allen, A. S., O’Riordan, T., McHutchison, J. G., Palmer, S. M., & Goldstein, D. B. (2017). An exome sequencing study to assess the role of rare genetic variation in pulmonary fibrosis. *American Journal of Respiratory and Critical Care Medicine*, 196(1), 82–93. https://doi.org/10.1164/ajrccm.201610-2088OC

Raghu, G., Collard, H. R., Egan, J. J., Martinez, F. J., Behr, J., Brown, K. K., Cordier, J. F., Flaherty, K. R., Lasky, J. A., Lynch, D. A., Ryu, J. H., Swigris, J. J., Wells, A. U., Ancochea, J., Bouros, D., Carvalho, C., Costabel, U., Ebina, M., Hansel, D. M., ... ATS/ERS/JRS/ALAT Committee on Idiopathic Pulmonary Fibrosis. (2011). An official ATS/ERS/JRS/ALAT statement: idiopathic pulmonary fibrosis: Evidence-based guidelines for diagnosis and management. *American Journal of Respiratory and Critical Care Medicine*, 183(6), 788–824. https://doi.org/10.1164/ajrccm.2009-040GL

Raghu, G., Remy-Jardin, M., Myers, J. L., Richeldi, L., Ryerson, C. J., Lederer, D. J., Cottin, V., Danoff, S. K., Morell, F., Flaherty, K. R., Wells, A., Martinez, F. J., Azuma, A., Bice, T. J., Bouros, D., Brown, K. K., Collard, H. R., Duggal, A., Galvin, L., Wilson, K. C., ... American Thoracic Society, European Respiratory Society, Japanese Respiratory Society, and Latin American Thoracic Society. (2018). Diagnosis of idiopathic pulmonary fibrosis. An official ATS/ERS/JRS/ALAT clinical practice guideline. *American Journal of Respiratory and Critical Care Medicine*, 198(5), e44–e68. https://doi.org/10.1164/ajrccm.201807-1255ST

Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Hegde, M., Lyon, E., Spector, E., Voelkerding, K., Rehm, H. L., & ACMG Laboratory Quality Assurance Committee. (2015). 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. *Genetics in Medicine*, 17(5), 405–424. https://doi.org/10.1038/gim.2015.30

Srirathan, S. S., Gajewska, M. E., Skyyte, A.-B. S., Madsen, L. B., & Bendsrud, E. (2018). Familial idiopathic pulmonary fibrosis in a young female. *Respiratory Medicine Case Reports*, 24, 1–4. https://doi.org/10.1016/j.rmcr.2018.03.005

Takezaki, A., Tsukumo, S., Setoguchi, Y., Ledford, J. G., Goto, H., Hosomiichi, K., Uehara, H., Nishioka, Y., & Yasutomo, K. (2019). A homozygous SFTPA1 mutation drives necroptosis of type II alveolar epithelial cells in patients with idiopathic pulmonary fibrosis. *Journal of Experimental Medicine*, 216(12), 2724–2735. https://doi.org/10.1084/jem.20182351

Thorwarth, A., Schnittert-Hübener, S., Schrumpf, P., Müller, I., Jyrch, S., Dame, C., Biebermman, H., Kleinau, G., Katchanov, J., Schuelke, M., Ebert, G., Steininger, A., Bönnemann, C., Brockmann, K., Christen, H. J., Crock, P., deZegher, F., Giès, M., Hewitt, J., ... Krude, H. (2014). Comprehensive genotyping and clinical characterisation reveal 27 novel NKX2-1 mutations and expand the phenotypic spectrum. *Journal of Medical Genetics*, 51(6), 375–387. https://doi.org/10.1136/jmedgenet-2013-102248

Tomer, Y., Wambach, J., Knudsen, L., Zhao, M., Rodrigue, L. R., Murthy, A., White, F. V., Venosa, A., Katzen, J., Ochs, M., Hamvas, A., Beers, M. F., & Mulugeta, S. (2021). The common ABCA3E292V variant disrupts AT2 cell quality control and increases susceptibility to lung injury and aberrant remodeling. *American Journal of Physiology—Lung Cellular and Molecular Physiology*, 321(2), L291–L307. https://doi.org/10.1152/ajplung.00400.2020

Wang, Y., Kuan, P. J., Xing, C., Cronkhite, J. T., Torres, F., Rosenblatt, R. L., DiMaio, J. M., Kinch, L. N., Grishin, N. V., & Garcia, C. K. (2009). Genetic defects in surfactant protein A2 are associated with pulmonary fibrosis and lung cancer. *The American Journal of Human Genetics*, 84(1), 52–59. https://doi.org/10.1016/j.ajhg.2008.11.010

Yoon, J. H., Nouriaie, M., Chen, X., Zou, R. H., Sallaeres, J., Veraldi, K. L., Chiarchiaro, J., Lindell, K., Wilson, D. O., Kaminski, N., Burns, T., Trejo Bittar, H., Yousem, S., Gibson, K., & Kass, D. J. (2018). Characteristics of lung cancer among patients with idiopathic pulmonary fibrosis and interstitial lung disease—analysis of institutional and population data. *Respiratory Research*, 19(1), 195. https://doi.org/10.1186/s12931-018-0899-4

**SUPPORTING INFORMATION**

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Sutton, R. M., Bittar, H. T., Sullivan, D. I., Silva, A. G., Bahudhanapati, H., Parikh, A. H., Zhang, Y., Gibson, K., McDyer, J. F., Kass, D. J., & Alder, J. K. (2022). Rare surfactant-related variants in familial and sporadic pulmonary fibrosis. *Human Mutation*, 43, 2091–2101. https://doi.org/10.1002/humu.24476