Candidate genes of idiopathic pulmonary fibrosis: current evidence and research

Wei Zhou1,2
Yaping Wang1,2
1Department of Medical Genetics, 2Jiangsu Key Laboratory of Molecular Medicine, Nanjing University School of Medicine, Nanjing, People’s Republic of China

Abstract: Idiopathic pulmonary fibrosis (IPF) is a group of common and lethal forms of idiopathic interstitial pulmonary disease. IPF is characterized by a progressive decline in lung function with a median survival of 2–3 years after diagnosis. Although the pathogenesis of the disease remains unknown, genetic predisposition could play a causal role in IPF. A set of genes have been identified as candidate genes of IPF in the past 20 years. However, the recent technological advances that allow for the analysis of millions of polymorphisms in different subjects have deepened the understanding of the genetic complexity of IPF susceptibility. Genome-wide association studies and whole-genome sequencing continue to reveal the genetic loci associated with IPF risk. In this review, we describe candidate genes on the basis of their functions and aim to gain a better understanding of the genetic basis of IPF. The discovered candidate genes may help to clarify pivotal aspects in the diagnosis, prognosis, and therapies of IPF.

Keywords: idiopathic pulmonary fibrosis, candidate genes, susceptibility

Introduction

Idiopathic pulmonary fibrosis (IPF) is a chronic lung disease characterized by progressively worsening dyspnea and pulmonary function. Although published data on the epidemiology of IPF in the world are limited, the majority of studies have shown an increase in its incidence and prevalence during recent decades. A systematic review assessed all the available population-based studies from 1968 to 2012 and estimated a conservative incidence range of three to nine cases per 100,000 per year in Europe and North America, whereas the incidences were lower in East Asia and South America. Patients with IPF exhibited variable disease courses and worse prognosis and had a median survival of 2–3 years after the initial diagnosis. In 2011, the American Thoracic Society/European Respiratory Society panel of experts issued a new classification scheme for IPF. It was defined as a specific form of chronic, progressive fibrosing interstitial pneumonia of unknown cause, occurring mainly in older adults and associated with the histopathological and/or radiological pattern of usual interstitial pneumonia. During recent decades, although there have been significant advances in the understanding of the pathogenesis of IPF, the prognosis of the disease was barely improved. Lung transplantation is still the only treatment option that successfully increases the survival.

Since Haschek et al proposed the “alveolar epithelial disrepair” hypothesis to challenge the traditional “inflammatory theory”, the pathogenesis of lung fibrosis has remained controversial. The point of conflict was whether the alveolar impairment was caused by an inflammation reaction or by an abnormal repair. Although the exact

Correspondence: Yaping Wang
Department of Medical Genetics, Nanjing University School of Medicine, Nanjing 210093, People’s Republic of China
Tel +86 25 8368 6495
Fax +86 25 8368 6451
Email wangyap@nju.edu.cn
etiology of pulmonary fibrosis is uncertain, the process of fibrosis is likely to involve an interaction between exogenous and endogenous.10 Cigarette smoking is the most important environmental risk factor associated with IPF. In a multicenter case–control study, a higher proportion of IPF patients had a history of smoking compared with controls.11 Exposure to several environment/occupational agents, particularly metal, wood dust and farming, or stone cutting/polishing, has also been observed to increase the risk of IPF.12,13 It is also worth noting that the comorbid conditions, such as gastroesophageal reflux,14 pulmonary hypertension,15 emphysema,16 viral or bacterial lung infections,17 and radiotherapy, may be the risk factors for IPF.

Studies have also demonstrated the genetic basis for IPF. The most persuasive evidence is a familial clustering of the disease, familial pulmonary fibrosis (FPF), which accounts for <5% of all cases.18 More recent data suggested that in up to 20% of IPF patients, the FPF could be underestimated.19 FPF is clinically and histologically indistinguishable from sporadic IPF except that the age of onset tends to be earlier. Subsequently, several gene variants and dysfunctional proteins were detected to be associated with the susceptibility of FPF. Three alveolar stability-associated genes (surfactant protein C, SFTPC; surfactant protein A2, SFTP A2; and ATP-binding cassette member A3, ABCA3) were detected to concentrate in FPF. As the incidence and prevalence of IPF increased with age, the disease was considered as an age-related condition with the distinctive features of degenerative disorders, such as telomerase dysfunction and shortening.20 Approximately 10% of FPF and 1%–3% of sporadic IPF patients carry variations in two major telomerase components: telomerase reverse transcriptase (TERT) and telomerase RNA component (TERC). Recent research has shown that dyskerin, dyskeratosis congenita 1 (DKC1), and regulator of telomere elongation helicase 1 (RTEL1) are also involved in the telomere shortening and may deteriorates the survival in IPF and supported the hypothesis that abnormal telomeres were involved in the disease pathogenesis and outcome.21 In 2011, a common variant (rs35705950) in the promoter region of MUC5B was detected to be strongly associated with both FPF and sporadic IPF. Sporadic reports showed that variants in multiple genes related to immunity and inflammation (TOLLIP, ELMOD2, major histocompatibility complex [MHC], interleukins [ILs], tumor necrosis factor [TNF], and transforming growth factor β [TGF-β]) can be found in some cases of IPF patients.

In this review, we present the current evidence and research on these genes that may contribute to the disease in order to uncover the underlying pathophysiological changes in FPF and sporadic IPF.

The genes associated with alveolar stability

SFTPC

SFTPC has been widely reported to be associated with various interstitial lung diseases (ILDs), including IPF. The SFTPC gene is located on the short arm of chromosome 8 (gene ID: 6440) with six exons encoding a 197 amino acid precursor protein. In 2001, Nogee et al reported the first case who was an infant with a family history of ILD and carried a heterozygous substitution of A for G in the splice donor site of intron 4.22 This nucleotide substitution resulted in the skipping of exon 4 and a 37 amino acid deletion in the C-terminal domain of the precursor protein. As a result, the mature surfactant protein C (SP-C) was absent in the lung tissue and bronchoalveolar lavage fluid. Later, two other heterozygous mutations, Leu188 to Gln (L188Q) and Ile73 to Thr (I73T), were identified in SFTPC gene.23,24 These two mutations may result in abnormal intracellular trafficking of protein and an accumulation of aberrantly processed prosurfactant protein C (pro-SP-C) within alveoli.

Lawson et al showed ten single-nucleotide polymorphisms (SNPs) in the SFTPC sequence in 89 IPF patients but not in 104 controls.25 These mutations were mainly detected in FPF. Markart et al reported two nonsynonymous variants in the SFTPC gene in a group of sporadic IPF cases, but their allele frequencies were similar to that of controls.26 Subsequently, at least 40 pathogenic mutations in the human SFTPC gene have been discovered through genetic analyses on the DNA samples from infants, children, and adult patients with ILDs. Compared with FPF, the mutations of SFTPC gene are not frequently detected in sporadic cases of IPF. The association observed between the mutations and the ILD led several groups to investigate its molecular mechanism.

The pro-SP-C encoded by the SFTPC would undergo several rounds of proteolytic cleavage to form the mature 35-residue hydrophobic protein (SP-C) and then secrete to the alveolar space. The SP-C, together with surfactant phospholipids and other surfactant proteins, is essential for maintaining alveolar stability.27 The abnormal SP-C proteins resulted from the SFTPC mutants can induce endoplasmic reticulum (ER) stress, inhibit the ubiquitin/proteasome system, and activate the apoptotic pathways.28 The cultured type II alveolar epithelial cells (AECII) transfected with SFTPC<sup>Δ3566Q</sup> caused the development of abnormal lamellar bodies, ER stress, and unfold protein response.29
Some fatal mutations have been found to be located in the BRICHOs domain of pro-SP-C, a special domain detected in 12 protein families with a wide range of functions and disease associations. For the pro-SP-C, the BRICHOs domain acts as a chaperone for SP-C during biosynthesis. It plays a pivotal role in the maturation of pro-SP-C, proper folding process and correct packaging, and eventually enters the lamellar bodies with other surfactant components for exocytosis. This process may explain why SP-C mutations can result in pulmonary fibrosis that occurs as a consequence of unfolded pro-SP-C but not due to deficiency, which is different from infant/adult respiratory distress. Epithelial to mesenchymal transition (EMT) has also been proven to occur in the process of collagen deposition in the alveolar wall and pulmonary fibrosis. Targeted expression of SFTPcL188Q in A549 cells resulted in decreased expression of E-cadherin and zonulaoccluden-I and increased expression of the mesenchymal marker α-smooth muscle actin, showing the relationship between the variants of SFTPC and EMT.

SP-C, together with SP-A, SP-B, and SP-D, constitutes 10% of the protein components of pulmonary surfactant in humans, and the remaining 90% is lipids. The main function of the surfactant is to reduce the surface tension at the air–water interface and prevent alveolar collapse. Because the surfactant alterations and alveolar type II cell apoptosis had been implicated in the pathogenesis of IPF, the hereditary and acquired disorders associated with the surfactants attracted more attention in the etiology study of IPF.

SFTPA2
The identification of the SFTPc mutation led to a search for a susceptible gene in encoding the other surfactant proteins in IPF. Wang et al. used linkage analysis to map a susceptible locus in the 15.7 Mb region of chromosome 1Q22 in a four-generation family with IPF and identified a transversion mutation (GGG to GTG) in codon 231 of one SFTPα2 allele that encoded SP-A (SP-A2). This mutation is predicted to change a highly conserved glycine residue to a valine (G231V). Subsequently, a heterozygous transversion mutation (TTC to TCC) in codon 198 that substituted a serine residue for a phenylalanine (F198S) in the same gene was found in a 45-year-old man with IPF and lung cancer in another family. Both of the mutations, G213V and F183S, occurred in the highly conserved carbohydrate-recognition domain of the protein and may disrupt cellular trafficking of the protein, probably by interfering with its proper folding.

Unlike SP-C, SP-A is a large multimeric protein that belongs to a family of collectin proteins characterized by an NH2-terminal collagen-like region and a C-terminal lectin domain. Because it is homologous to collectin, SP-A has always been thought to play a role in the innate defense system and pulmonary immune response. It may interact with pathogen-associated molecular patterns anchored on membranes via the calcium-dependent carbohydrate-binding domains. SP-A has been associated with SP-D in immune cells, in which it activates various cellular functions and acts as an opsonin in the modulation of macrophage phagocytosis and the production of reactive oxygen species. Experimental data showed that the expression of mutant proteins increased the transcription of a BiP-reporter construct, expression of BiP protein, and production of an ER stress-induced XBP-1-spliced product. The patients with a mutant protein of SP-A2 showed a higher level of TGF-β1 than that in the control subjects. This finding suggested that the function lesion of SP-A2 may cause the disease through a change in the secretion of TGF-β, a cytokine known as a fibrogenic mediator in epithelial cells.

Although the pathogenic mechanism of the gene mutation associated with pulmonary fibrosis remains to be clarified, the linkage analysis and the candidate gene approaches indicated that the SFTPα2 may be a susceptible gene in IPF.

ABCA3
ABCA3 has been identified as a neonatal respiratory distress syndrome or ILD-associated protein in childhood. It is a transporter protein predominantly expressed in the AECII and is involved in the transport of lipids across plasma membranes. The ABCA3 gene encodes a 1,704 amino acid protein located on the limiting membrane of lamellar bodies, the secretory vesicle for surfactant, suggesting that it may play a role in surfactant metabolism and transport. There is a wide spectrum of variants in the ABCA3 gene, and >150 distinct mutations have been found to cause genetic abnormalities in surfactant metabolism, which would lead to neonatal respiratory distress and pediatric ILD. The lung disorder caused by ABCA3 mutations occurs in an autosomal recessive pattern, and the compound heterozygote or homozygous mutations could be detected in both alleles of the patient. ABCA3 may also act as a modifier of the phenotype associated with SFTPc mutation, as the heterozygous ABCA3 mutation was found in severely affected infants with SFTPc, a commonly identified substitution of isoleucine by threonine in codon 73. A sequence analysis in a full-term female baby who developed severe chronic lung disease revealed that she was a homozygous carrier of a mutation p.Trp308Arg (c.922T>C) at exon 9 of the ABCA3 gene but without any
mutation in SFTPC. A novel homozygous G>T transition at nucleotide 2891 in exon 21 of the ABCA3 gene, resulting in G964D, was found in a large kindred of pulmonary fibrosis. Kumar et al reported an IPF patient who carried a compound heterozygote for ABCA3, E292V, and G964D, further supporting the possibility that adult-onset fibrotic lung disease may be due to the mutations in ABCA3. In our unreported study, a novel variant in exon 24 of ABCA3 may be a risk factor for IPF in the Chinese population.

The genes associated with maintaining of telomere length TERT and TERC

The TERT gene located at chromosome 5p15.33 (gene ID: 7015) encodes the catalytic component of telomerase, and the TERC at chromosome 3q26.2 (gene ID: 7012) is an RNA gene and acts as a component of telomerase to maintain telomere. In germ cells and other stem cells, the integrity of the telomere is maintained by telomerase complex, containing the following two key components: the TERT and the TERC. The telomerase complex uses the template encoded by TERC to add telomere repeats (TTAGGG) to the end of the chromosomes, which works as a “cap” to offset the telomere shortening with cell divisions. Evidence has indicated a relationship between chronological age and the incidence of IPF. Similarly, telomere length could limit tissue renewal capacity, which was implicated in age-related disease. Indeed, IPF is one of the most prevalent clinical manifestations of human telomere syndromes, which have defined pathological ranges of telomere shortening.

Recent research showed that short telomere length was detected in both the peripheral blood and the lung tissue cells of IPF patients. Moreover, in bleomycin-induced lung fibrosis mice, telomerase mutation or telomere shortening caused emphysematous lesions. Overall, the germ-line mutations in telomerase components, TERT and TERC, underlie 8%-15% of FPF and 1%-3% of sporadic pulmonary fibrosis patients. In 2007, Armanios et al identified five heterozygous mutations in TERT and one mutation in TERC in six families, and none of these mutations were found in controls. Later, a genome-wide association study (GWAS) with 159 IPF patients and 934 controls identified a susceptibility SNP (rs2736100) in intron 2 of the TERT gene in a Japanese population. This variant (rs2736100) is located within a linkage disequilibrium block in TERT, suggesting that it could be associated with some other unidentified variations in IPF.

Polymorphisms in TERT and TERC might contribute to telomere attrition, which was reported as a hallmark of aging. Increased mutations affecting telomere function have been reported in IPF patients, but the underlying mechanism for its pathogenesis still remains unclear. Alveolar stem cell senescence may be a driver of the pathogenesis. Evidence has shown that damaged telomere maintenance was related to epithelial cell senescence in lung injury response, which results in impaired lung epithelial homeostasis. Telomere dysfunction in AECs preferentially upregulated the DNA damage response, activated the p53 pathway, and induced cell senescence. Another intriguing target of TERT regulation is the Wnt-β-catenin pathway. In the lung tissues of IPF patients, the Wnt-β-catenin pathway is activated, which correlated with increased expression of (TGF-β) to promote EMT. Further investigations are needed to confirm the role of telomere maintenance and cellular senescence in lung fibrosis and elucidate the association between the aforementioned pathways in lung injury response and fibrotic remodeling.

DKC1 and RTEL1

Up to 25% of IPF patients and 40% of FPF patients exhibited the shortened telomeres, but less than half of these patients had the germline mutation in the TERT or TERC gene. The low frequency of these events suggested that there may be other genes responsible for the telomere length. The DKC1 gene that encodes a component of the telomerase complex was reported to cause autosomal dominant dyskeratosis congenita (DC), a disorder in which 20% of affected individuals develop pulmonary fibrosis. A transversion mutation (c.145A>T) in DKC1 exon 3, which predicted a missense substitution (p.Thr49Ser) near the nuclear localization signal of dyskerin, was identified in a kindred of FPF patients but not in 2,816 controls. Recently, Kropski et al screened a kindred of FPF and found a 69-year-old male patient carrying a novel variant, c.1213A>G, in the DKC1 gene. No mutation was identified in those aforementioned FPF-associated genes (TERT and TERC) in this patient. These results suggested that the DKC1 may be the third telomerase–related gene in the etiology of FPF.

Using next-generation sequencing technologies, Stuart et al recently revealed two genes that were also associated with telomere length. One was the PARN gene, which encoded an exoribonuclease. Five damaging mutations of PARN in heterozygous were identified in 78 European FPF patients but not in 2,816 controls (P=1.3×10^-6). The other was RTEL1 gene, an established locus for DC, which harbored more novel damaging and missense variants at conserved residues in patients than in controls (P=1.6×10^-4). Four highly
detrimental mutations were identified in a group of French FPF patients using whole-exome sequencing analysis in this gene. Three of the four mutations were missense mutations (p.Thr49Met, p.Arg213Trp, and p.Phe964Leu). The fourth was one base pair insertion in exon 33 (c.3493dupC, p.Gln1165Profs*22), leading to a frameshift and a premature stop codon. A function prediction showed that these variants may seriously modify the functional domains of the protein.

RTEL1 is a member of the subclass of iron–sulfur cluster-containing DNA helicases that participate in telomere maintenance, and shortened telomeres were detected in the cells from the patients with the RTEL1 mutation. Other researchers found that telomere length in peripheral blood mononuclear cells from patients with RTEL1 mutations was extremely short and similar to the peripheral blood mononuclear cell telomere length in other patients with known telomerase pathway mutations (TERT, TETC, and DKK1), suggesting that the rare loss-of-function variants in RTEL1 represented a genetic risk of FPF. These findings indicated that the RTEL1, as a novel candidate gene, was associated with IPF and confirmed the important role of telomerase dysfunction and shortened telomeres in the pathogenesis of pulmonary fibrosis.

### Common variation and IPF risk MUC5B

A genome-wide linkage screen in 82 multiplex families with familial interstitial pneumonia (FIP) demonstrated an association between idiopathic interstitial pneumonia and the 3.4 Mb region of chromosome 11p15 where the maximum multipoint LOD score was 3.3 (P=0.00004). Consequently, fine mapping was performed by genotyping 306 tagging SNPs in the mucin gene cluster region on the p-terminus of chromosome 11 in the unrelated FIP subjects, IPF subjects, and healthy controls. Nineteen of them were detected to be associated with FIP and/or IPF.

Among these SNPs, the variant (rs35705950) at 3 kb upstream of the MUC5B transcription start site, in the putative promoter, had the most substantial effect on pulmonary fibrosis risk with a frequency of 38% among subjects with IPF, 34% among those with FIP, and 9% in controls. MUC5B (gene ID: 727897) is an evolutionarily conserved gene that encodes mucin glycoprotein, a principal macromolecule in airway mucus. Interestingly, the expression of the MUC5B protein in the lung tissues of IPF patients was ∼14-fold higher than those who did not. This variant of rs35705950 was associated with a 37.4-fold increase in MUC5B protein expression in the lung tissues of unaffected subjects who were heterozygous/homozygous compared with the wild-type allele. The minor allele of the promoter region exhibited a higher frequency in both familial and sporadic IPF patients. Therefore, this variant is currently the most common genetic predisposing humans to IPF. However, our previous study showed that the promoter variant rs35705950 was rare in the Chinese population, although an association was observed between this variant and a risk of ILD. Peljto et al reported that the MUC5B promoter variant rs35705950 was associated with IPF in a Mexican population but not in Koreans.

Recent evidence showed that patients with pulmonary fibrosis had positive MUC5B staining in the cells that were morphologically consistent with AECs, which normally do not produce mucus. The mechanism by which the expression of MUC5B increases and it favors the development of IPF remains uncertain. Based on the relationship between this variant and excess production of MUC5B, the increased MUC5B expression may impair mucus clearance and alveolar repair and further interfere with host defense. As a risk factor for this complex disease, this typically common variation may have potential predictive value for the development of this disease in a population. Although the previous data provide strong evidence for an association between the polymorphism of MUC5B gene and IPF, the molecular consequences of the variation and the contribution of the increased MUC5B protein expression contributes to the development of lung fibrosis require further explanation.

### The genes related to immunity and inflammation

**Toll-like receptor-associated genes**

Despite the controversy on corticosteroids and immune suppressants, the innate and adaptive immune responses were thought to contribute to the pathogenesis of lung fibrosis. Toll-like receptors (TLRs), as a family of pattern recognition receptors, mainly express by immune-related cells. TLR3 specifically recognizes double-stranded RNA and induces multiple intracellular events responsible for innate antiviral immunity. The TLR3 polymorphism (L412F) that would defect the function of TLR3 was found in patients with IPF. The further study showed that the lung fibroblasts from IPF patient with the variant L412F in TLR3 could reduce cytokine responses and deregulate fibroproliferation when compared with the wild genotype.

TOLLIP is a ubiquitin-binding protein that interacts with several components of the TLR signaling cascade. It is an important regulator of innate immune responses mediated...
The pathogenesis of IPF is complex and remains to be clarified. The proinflammatory cytokines may play an important role in the initiation and development of pulmonary fibrosis. Vasakova et al investigated the cytokines polymorphisms of the IL-1α, IL-1β, IL-1R, IL-1RA, IL-2, IL-4, IL-4RA, IL-6, IL-10, IL-12, TNF-α, IFN-γ, and TGF-β genes among 30 IPF patients and 103 healthy volunteers and found the CT genotypes at the promoter regions, the positions −590 and −33, of IL-4 gene was significantly associated with IPF susceptibility.70 Ahn et al demonstrated that IL-8 rs4073 T allele is risk factor for IPF in the Korean population.71 However, Alhamad et al72 examined the allele distributions of polymorphisms in TNF-α, IFN-γ, IL-6, IL-10, and TGF-β1 and the serum levels of cytokines in 60 patients with IPF and 150 healthy controls. Their results indicated that the genotypes of the polymorphisms investigated in the cytokine genes did not predispose to the susceptibility of IPF but played an important role in disease severity.

The other candidate genes related to IPF DSP and DPP9

The GWAS by Fingerlin et al identified multiple susceptibility loci for pulmonary fibrosis, including two cell–cell adhesion molecules, DSP and DPP9, which suggested that defects in cell adhesion or the cellular cytoskeleton could predispose to injury in response to stimuli.73 DSP encodes desmoplakin, a component of the desmosome that tightly links adjacent cells and forms a dynamic structure with other proteins that can bind the cytoskeleton to the cell membrane. The DPP9 is a member of the same protein family as fibroblast activation protein, which was shown to express in fibroblastic foci and in fibrotic interstitium but not in adjacent healthy lung of IPF patients. These results suggested that pulmonary fibrosis might be caused by defects in cell–cell adhesion or the cytoskeleton and DSP and DPP9 could be considered as the candidate genes for IPF.

PINK1

Some age-related neurodegenerative disease has an etiology that is related to mitochondrial dysfunction. Bueno et al recently showed that AECII in the lungs of IPF patients exhibits marked accumulation of dysmorphic and dysfunctional mitochondria.74 Their results also indicated that impaired mitochondria in IPF and aging lungs were associated with low expression of PTEN-induced putative kinase 1 (PINK1). Patel et al75 found that TGF-β could induce lung epithelial cell mitochondrial ROS and depolarization and stabilizes PINK1 protein. PINK1 can ameliorate epithelial cell death and may be necessary to limit fibrogenesis. This evidence indicated that mitochondrial dysfunction may have a role in the pathogenesis of pulmonary fibrosis, and
mitochondrial-associated genes could be recognized as a novel genetic susceptibility loci.

**Conclusion**

In recent decades, tremendous progress has been made in understanding the genetic predisposition of IPF. Genetic association studies play a pivotal role in demonstrating the genetic cause of IPF and revealing the underlying mechanisms of pulmonary fibrosis. The gene variants mentioned in our review only identified in part of IPF patients, and the various pathological phenotypes presented in patients indicated a complex genetic basis of IPF. A increased list of genes involved in IPF pathogenesis would be obtained with high-throughput analysis and next-generation sequencing in the future. Moreover, more effort should be concentrated on investigating the roles of genetic variants and interaction between the genetic background and environmental factors. Elucidation of the genetic basis and pathogenesis of this disease may produce radical changes in the diagnosis and management of IPF.

**Acknowledgment**

This study on IPF susceptibility gene has been supported by the National Natural Science Foundation of China (grant number 81270152).

**Disclosure**

The authors report no conflicts of interest in this work.

**References**


