



REVIEW

Genetics of fibrosing lung diseases

J.C. Grutters* and R.M. du Bois#

ABSTRACT: Genetic studies in familial lung fibrosis have demonstrated an association with surfactant protein C genes: two mutations have been found resulting in protein misfolding and causing type-II epithelial cell injury. Remarkably, different histological patterns were observed in the affected subjects, suggesting the influence of modifier genes and/or environmental factors. Surfactant protein C gene variations have not, however, been associated with sporadic cases, *i.e.* idiopathic pulmonary fibrosis (IPF).

Susceptibility to IPF probably involves a combination of polymorphisms related to epithelial cell injury and abnormal wound healing. To date, the genetic associations with IPF that have been reported in different cohorts include the genes encoding tumour necrosis factor (TNF; -308 adenine), interleukin-1 receptor antagonist (+2018 thymidine) and association with severity and progression (interleukin-6/TNF receptor II and transforming growth factor- β 1 (TGFB1; +869 cytosine)), but none of these associations have been replicated by others.

Unlike in IPF, immunological inflammation seems to be more prominent in the pathogenesis of scleroderma lung fibrosis, being an autoimmune disease with specific autoantibodies, such as antitopoisomerase antibodies, in patients with diffuse lung disease, and anticentromere antibodies, in patients with pulmonary vascular disease. Antitopoisomerase antibody positivity is associated with the carriage of human leukocyte antigen DRB1*11 and DPB1*1301 alleles, suggesting the recognition of a specific amino-acid motif. Extended haplotype analysis also supports the conclusion that TNF may be the primary association with anticentromere positivity. Intriguingly, associations with TGFB1 and genes involved in extracellular matrix homeostasis have been reported in this disease.

In conclusion, significant steps forward have been taken in the understanding of the genetic contribution to fibrosing lung diseases, but major challenges lay ahead. It is the present authors' opinion that only a combined approach studying large numbers of familial and sporadic cases, all clinically well phenotyped, using multiple distinct cohorts, and genotyped according to relevant gene ontologies will be successful. It will be necessary to be particularly vigilant with regard to phenotype; the absence of very strong reproducible associations may be because of the rigidity of phenotype definition, coupled with the possibility that idiopathic pulmonary fibrosis may still be a heterogeneous group of diseases, despite the more rigid definition set out by the European Respiratory Society/American Thoracic Society statement.

KEYWORDS: Familial, genetics, interstitial lung disease, pulmonary fibrosis

Diffuse parenchymal lung diseases (DPLDs) comprise >200 entities and include a wide spectrum of diseases, many uncommon and of unknown aetiology, but together accounting for ~15% of respiratory practice [1]. An important heterogeneous group of DPLDs are the idiopathic interstitial pneumonias (IIPs), which result from damage to the lung parenchyma by varying mechanisms of inflammation and fibrosis. Recently, an international statement on IIPs has been published that, for the first time, provides an integrated clinical, radiological and pathological approach to the

classification of these diseases, including seven clinicoradiopathological entities [2]. Idiopathic pulmonary fibrosis (IPF) is the best-known entity and can now be diagnosed even without the need for surgical lung biopsy. Nonspecific interstitial pneumonia (NSIP), cryptogenic organising pneumonia, acute interstitial pneumonia, respiratory bronchiolitis-associated interstitial lung disease, desquamative interstitial pneumonia and lymphoid interstitial pneumonia are the other disease entities [2]. Importantly, radiological and histological patterns seen in IIPs can also be found in the context of systemic diseases, or exposure to

AFFILIATIONS

*Heart Lung Centre Utrecht, Dept of Pulmonology, St. Antonius Hospital, Nieuwegein, The Netherlands.

#Clinical Genomics Group, Dept of Occupational and Environmental Medicine, Royal Brompton Hospital and National Heart and Lung Institute, Imperial College of Science, Technology and Medicine, London, UK.

CORRESPONDENCE

R.M. du Bois
Clinical Genomics Group
1B Manresa Road
London
SW3 6LR
UK

Fax: 44 2073518336
E-mail: r.dubois@rbh.nthames.nhs.uk

Received:
November 22 2004
Accepted:
January 03 2005



drugs or other environmental factors. For example, NSIP is the most common DPLD in patients with systemic sclerosis.

It is widely believed that genetic factors play an important role in the aetiopathogenesis of many DPLDs, but the identification and understanding of these factors is in its infancy. This review focuses on key issues in the genetic research of fibrosing lung diseases, and, in particular, summarises current knowledge about genetic predisposition to familial pulmonary fibrosis, sporadic IPF and lung fibrosis in systemic sclerosis.

KEY CONSIDERATIONS

Many DPLDs can result in "pulmonary fibrosis". Historically, (idiopathic) interstitial pneumonias were often considered as being synonymous with "pulmonary fibrosis", and inappropriate attempts were made to derive meaningful correlations of treatment response and outcome in this population. Additionally, lung fibrosis can result from many other pathological processes, including infection, malignancy and surgical procedures. Thus, the terms "lung/pulmonary fibrosis" and also "fibrosing alveolitis" should not be used as a diagnostic label. Recent refinement in the classification of IIPs has provided an important handle for diagnostic categorisation of fibrosing lung diseases [2]. Even within these diagnostic categories, considerable heterogeneity in terms of clinical characteristics, disease activity, severity, prognosis and response to therapy can be observed. Therefore, in the context of genetic studies, precise phenotyping of each case is extremely important.

It is thought that the development of most fibrosing lung diseases occurs in susceptible individuals following an exposure to a variety of potential environmental triggers. Evidence to support this hypothesis derives from the observation of varying susceptibility to environmental causes; reports on familial clustering of DPLDs, including IPF, sarcoidosis, hypersensitivity pneumonitis, Langerhans' cell histiocytosis, pulmonary alveolar proteinosis and desquamative interstitial pneumonia; and from human leukocyte antigen (HLA) association studies.

Fibrosing lung diseases are referred to as "complex diseases", *i.e.* it is thought that multiple genetic loci, each exerting variable relatively small effects, are involved. Some alleles might predispose to disease, whereas others might be protective (susceptibility alleles), or might be involved in severity/progression (modifying genes).

GENETIC RESEARCH STRATEGIES

Three main strategies can be applied to investigation of the role of genetic factors in fibrosing lung diseases. These strategies do not differ from those used in other fields.

The first strategy is linkage analysis in families. Linkage analysis requires the use of a large number of DNA markers that are closely spaced throughout the genome in order to attempt to identify chromosomal regions that cosegregate with the occurrence of disease or a specific phenotype in families with lung fibrosis. Once a candidate region is sufficiently small, the disease-associated gene can be located through positional cloning if there are no other candidate genes in this area. There are advantages and disadvantages to such a classic linkage analysis. An advantage is that *a priori* no candidate

gene has to be defined. The definition/selection of the latter is often an educated guess or based on insight. A clear disadvantage is the preferable need for three generations: DNA from grandparents, parents and children are needed, and, when the disease occurs late in life, grandparental information is often lacking. A second problem with these studies is the need for substantial sample sizes.

The second strategy is based on case-control association studies, which seek evidence for a significant association between an allele of a candidate gene or gene haplotype (series of polymorphisms in the same gene that cosegregate) and a disease/phenotype characteristic by comparing the allele/genotype frequencies of affected subjects and carefully selected and, if necessary, age-matched controls. There is no need for information derived from grandparents or children and, hence, this strategy is often the preferred route for late-onset diseases. Case-control studies are relatively easy to perform in a short period of time, which is one of the reasons why they have become increasingly popular. Case-control studies, however, have the problem that a significant association between an allele and a disease does not mean that the candidate gene is a causal one. Population stratification in the (distant) past can lead to allele frequency differences between diseased and healthy subjects without the gene being causative or disease related. A positive finding in a case-control study, therefore, always needs confirmation in another group of subjects whom it might reasonably be assumed are not related to the first group of subjects, *e.g.* a positive finding in Japanese subjects should be confirmed in subjects who are African, Dutch, English or of some other race before that allele can be added to the list of "approved" candidate genes. The obvious need for confirmatory studies may render the case-control study less attractive.

Thirdly, the affected sibling pair (ASP) method might be used. This studies identity by descent sharing between two siblings affected by the disease of interest. Basically, the ASP test compares the frequencies of alleles in diseased and non-diseased subjects and shares many of the characteristics of the case-control study, the difference being that related subjects (siblings) now form the control group. The statistical evaluation of such correlated information is more demanding because the straightforward Chi-squared test cannot be used: specialised statistical software must be used. The major advantage of the ASP test is, however, that population stratification is no longer a problem: it starts from the notion that, if population stratification does exist, the siblings of the diseased subject will be affected by that process to the same degree, and so any difference in allele frequency between diseased and nondiseased subjects must reflect a genuine association.

FAMILIAL PULMONARY FIBROSIS

Familial pulmonary fibrosis or familial IIP is identified by confirming IIP in two or more members of the same family. It is of note that different subtypes of IIP can be found in an affected family, which sometimes causes difficulties in establishing a precise familial disease phenotype. The most frequently reported IIP entity in familial-type lung fibrosis is IPF/usual interstitial pneumonia (UIP; ~80%) [3].

Epidemiological studies

The precise incidence and prevalence of IPF are not known. Its incidence has been estimated at 7–11 per 100,000 in the general population [4]. The prevalence has been estimated at 13–20 per 100,000 population and increases with the ageing process [4]. MARSHALL *et al.* [5] were the first to report a large cohort of familial IIP. They identified 25 UK families, comprising 67 cases, which were all classified as IPF on the basis of high-resolution computed tomographic findings (93%) and/or the histological pattern (32%). The authors estimated that familial cases account for 0.5–2.2% of all patients with IPF, with a prevalence of 1.34 cases per million in the UK population. In contrast, LOYD [6] recently suggested that the real incidence of familial IPF might be considerably higher. The author reported that, in the Vanderbilt lung transplantation programme, nine of the 47 (19%) individuals undergoing transplantation for this condition showed a positive family history.

In a nationwide epidemiological study in Finland including all pulmonary clinics in the country (n=29), a prevalence of 5.9 per million population was found for familial IPF [7]. Interestingly, this study revealed evidence of geographical clustering of multiplex families, suggesting a recent founder effect in patients with familial IPF [7].

In the largest familial study undertaken to date, coordinated by D. Schwartz in the USA, 75 families with two or more affected individuals have been identified. In 56 of these, detailed phenotyping has shown that 25 families exhibited uniform IPF, whereas, in the remaining 31, phenotypic heterogeneity was observed. The group is undertaking detailed linkage analysis in an endeavour to find a region sufficiently tight for positional cloning and hopefully to identify a key gene or genes (D. Schwartz, Duke University Medical Center, Durham, NC, USA, personal communication).

Mode of inheritance

The familial form of IPF/UIP is probably transmitted as an autosomal dominant trait with reduced penetrance [3, 8, 9]. An interesting observation in this respect has been made by BITTERMAN *et al.* [9]. They evaluated 17 unaffected members of three families with an autosomal dominant form of IPF and found evidence of alveolar inflammation in approximately half of them. In addition, these subjects showed no progression towards lung fibrosis during a follow-up of 2–4 yrs. Unfortunately, to the present authors' knowledge, longer-term follow-up data from this study cohort have not been reported.

Surfactant protein C gene mutations

As the results of genetic studies analysing candidate loci near the HLA region of chromosome 6 and on chromosome 14 in familial lung fibrosis have remained largely negative [6, 10], clearly one of the most intriguing genetic findings to date in this disease has been the identification of causal mutations in the surfactant protein gene (SFTPC). NOGEE *et al.* [11] were the first to describe a mutation in this gene that was associated with NSIP in an infant whose mother had desquamative interstitial pneumonia. Heterozygous guanine (G) to adenine (A) transition substitution (>) of the first base of intron 4 (IVS4+1G>A) was present in both patients, and caused skipping of exon 4 with the deletion of 37 amino acids. The putatively encoded product thus lacked a cysteine residue that

is important for disulphide-mediated protein folding. In the patients, lack of mature surfactant protein (SP)-C in lung tissue and bronchoalveolar lavage fluid was observed, supporting the concept that the precursor protein was not being processed and secreted normally. The SFTPC IVS4+1G>A mutation was identified on only one allele in each patient, consistent with an autosomal dominant pattern [11]. Recently, transfection studies demonstrated that the mutant protein diverts its wild-type counterpart to aggresomes, thus providing a molecular mechanism for the dominant negative effect observed *in vivo* [12].

Consistent with this finding, THOMAS *et al.* [8] reported on another mutation in the gene encoding the hydrophobic lung-specific SP-C. In a large familial IIP kindred, including 14 affected members and spanning six decades, they found a thymidine (T) to A transversion in exon 5 (position +128) in DNA from all available affected family members. This mutation substituted a polar residue (glutamine) for a highly conserved neutral amino acid (leucine 188) that was predicted to hinder processing of SP-C precursor protein. This was confirmed by immunostaining for pro-SP-C, which showed very abnormal distribution of staining in lung from affected patients in this family compared with normal lung [8]. Interestingly, again, two different pathological diagnoses (UIP and NSIP) in affected relatives sharing the same SFTPC mutation were found. In addition, the histological diagnosis varied with age, *i.e.* UIP in adulthood and NSIP in childhood. Although the possibility of NSIP occurring as precursor lesion to UIP could not be excluded in this family, it is more likely that the pleiotropic effects of the genetic defect are caused by modifier genes and/or environmental factors.

SFTPC has been localised on the short arm of chromosome 8 [13]. It is a relatively small gene spanning 3.5 kilobases and composed of six exons [14], with expression restricted to type-II alveolar cells. Interestingly, two other mutations in SFTPC (exons 3 and 5) have recently been reported in infantile pulmonary alveolar proteinosis with or without interstitial lung disease [15, 16]. Figure 1 shows a diagram of the SFTPC gene with the nucleotide positions of the previously described mutations.

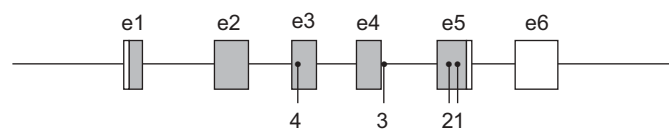


FIGURE 1. Nucleotides 22,041,255–22,043,927 of chromosome 8, showing the current nucleotide positions (according to RIVA and KOHANE [17]) of the surfactant protein C gene mutations currently known to be associated with interstitial lung diseases: 1) 22,043,461T>A (glutamine substituted for leucine 188) [8]; 2) 22,043,398 G>A (glutamine substituted for arginine 167) [16]; 3) 22,042,998 G>A (skipping of exon 4) [11]; and 4) 22,042,547T>C (threonine substituted for isoleucine 73) [15, 16] (□: noncoding region; ■: coding region). Mutations 1 and 3 have been attributed to the occurrence of pulmonary fibrosis in different families [8, 11]. e: exon.

Proposed surfactant protein C gene mutation disease mechanisms

SP-C is perhaps the most hydrophobic protein yet encountered in the mammalian proteome, bearing a domain rich in valine, leucine and isoleucine residues that form a stable α -helical structure approximately spanning the phospholipid bilayer [18]. The α -helical SP-C molecule is membrane associated and palmitoylated, and readily forms insoluble random structures in the aqueous environment [19]. Missense or short deletion mutations, as seen in the studies of NOGEE *et al.* [11] and THOMAS *et al.* [8], result in the production of a stable mRNA that produces an abundance of a misfolded protein, which, by accumulation or complex formation, may cause type-II epithelial cell injury. However, AMIN *et al.* [20] described a family with similar lung pathology in whom expression of pro-SP-C and active SP-C was undetectable. Therefore, the severe lung disease caused by abnormalities in SP-C can result from production and accumulation of an abnormal pro-SP-C as well as its absence. A deficiency of SP-C in surfactant could cause abnormal shear forces in the alveoli, thereby causing mechanical injury of the respiratory epithelium, which, in turn, may contribute to the pathogenesis of (familial) pulmonary fibrosis [6].

Hermansky–Pudlak syndrome gene mutations

Hermansky–Pudlak syndrome (HPS) is an autosomal recessive disorder characterised by the classic triad of oculocutaneous albinism, bleeding and lysosomal ceroid (a chromolipid related to lipofuscin) storage resulting from defects in multiple cytoplasmic organelles: melanosomes, platelet-dense granules, and lysosomes [21]. A certain fraction of HPS patients also suffer from pulmonary fibrosis, which is often fatal before 40 yrs of age. Although the disease occurs worldwide, it is most common in north-western Puerto Rico, where its frequency is 1 in 1,800 owing to a founder effect. In Europe, the disease has also been identified in a village in the Swiss Alps, and in persons of Dutch and Turkish ethnicity. Four disease phenotypes are currently recognised, *i.e.* HPS-1, HPS-2, HPS-3 and HPS-4, of which HPS-1 and HPS-4 are associated with severe pulmonary fibrosis. The HPS phenotypes are recognised to result from mutations in four genes: HPS-1 gene, adaptin β 3A gene (ADTB3A), HPS-3 gene, and HPS-4 gene, respectively [21]. To date, only ADTB3A produces a product of known function. This gene codes for the β 3A subunit of the adapter complex-3 coat protein, a heterotetrameric complex that mediates vesicle formation [22, 23]. Although no direct evidence exists to support the hypothesis, it would be logical that the other HPS-causing genes are also involved in vesicle formation.

The pathological features of HPS lung fibrosis have been studied by NAKATANI *et al.* [24], and are characterised by extensive proliferation of type-II epithelial cells with characteristic foamy swelling/degeneration, patchy fibrosis with lymphocytic and histiocytic infiltration, and honeycomb change. Unlike in UIP/IPF, the honeycombing was not found to show a predilection for the lower lobes or subpleural regions. Type-II epithelial cells are characterised by an accumulation of phospholipid, with weak positivity for surfactant protein immunohistochemically. Ultrastructurally, the presence of numerous giant lamellar bodies that compress

the nucleus with occasional cytoplasmic disruption has been demonstrated [24]. Together these findings suggest a form of cellular degeneration with an accumulation of surfactant (giant lamellar body degeneration), indicating that there might be a basic defect in the vesicle formation/secretion process of surfactant by type-II epithelial cells in HPS.

SPORADIC IDIOPATHIC PULMONARY FIBROSIS

Although SFTPC mutations form a genetic basis for some familial cases of IPF, they are not associated with all of the families with IPF. In addition, the majority of patients with IPF lack a family history. Therefore, the development of IPF is likely to be determined by multiple genetic factors that each contributes to a modest effect on predisposition to this disease. In combination with appropriate environmental or cellular triggers, individuals who possess these predispositional factors may develop IPF. To date, the lack of genome-wide linkage analysis studies and knowledge of chromosomal loci for IPF make the positional cloning of these genes impossible. Therefore, case–control association studies are presently the method of choice for the genetic study of IPF.

Following the traditional paradigm for the pathogenesis of IPF, candidate genes that have been analysed for this disease are primarily involved in inflammatory responses, *e.g.* tumour necrosis factor (TNF)- α and genes in the interleukin (IL)-1 gene cluster. However, increasing evidence since the late 1990s suggests that the inflammatory response may not precede the fibrosis nor play a major role in the development of IPF [25]. Instead, the evolving hypothesis proposes that IPF is a disease of impaired wound healing involving the epithelial/fibroblast pathway [26]. The triggering event in IPF is thought to be multiple and continuous microscopic insults to the alveolar epithelial cells, suggesting that different gene ontologies might be relevant. Specifically, candidate genes might be found in surfactant genes and genes involved in oxidant/antioxidant mechanisms. Epithelial injury is followed by activation of the coagulation cascade, and aberrant wound healing subsequently leads to fibroblast proliferation, transformation into myofibroblasts, formation of fibroblastic foci, deposition of extracellular matrix and, ultimately, end-stage fibrosis. It is obvious from this extensive pathogenetic process that many candidate genes can be identified with reasonable logic. In that the insult to the epithelial cell is probably the first event in this process, genes involved in the epithelial cell response to this insult might be considered as the most attractive targets, but the scientific literature is very light in this regard.

Although many candidate genes for a key role in IPF pathogenesis can be proposed, only limited numbers have been evaluated to date, and even fewer of them have demonstrated confirmed associations. Candidate genes and polymorphisms can best be categorised according to specific biological domains playing a role in IPF pathogenesis. Each major category will be discussed, including a review of the results reported to date, which are also summarised in table 1. It is of note that, because of the intrinsic pitfalls of genetic association studies [27], and the chances of spurious classification of phenotypes and insufficiently powered case–control studies, positive and negative findings should be interpreted with caution. Positive findings, suggesting putative disease gene variants in the development of disease, need further

TABLE 1 Genes and gene polymorphisms that have been evaluated in sporadic idiopathic pulmonary fibrosis

Gene ontology	Gene symbol	Polymorphism	Association with disease susceptibility	Pop. (cases/controls n)	[Refs]
Pro-/anti-inflammation					
IL-1 cluster	IL1A	-889C>T	No	CZ W (54/199)	[29] ⁺
	IL1B	-511C>T	No	CZ W (54/199)	[29]
TNF and related genes	IL1RN	+3953C>T	No	CZ W (54/199)	[29]
		+2018C>T [#]	+2018T (OR 3.8)	UK W (88/88); IT W (61/103)	[30]
		Intron 2 VNTR [#]	No	CZ W (54/199); AU W (22/140)	[29, 31]
	TNF	-238G>A	No	UK W (74/201)	[32]
		-308G>A	-308A (OR 13.9)	UK W-1 (88/88); IT (61/103); AU W (22/140)	[30, 31] [§]
		+488G>A	No	UK W (74/192)	[32]
		+676T>G	No	UK W (74/192)	[32]
	TNFR2	+1663G>A	No	UK W (74/192)	[32]
		+1668T>G	No	UK W (74/192)	[32]
		+1690T>C	No	UK W (74/192)	[32] ^f
+249A>G		No	UK W (74/201)	[32]	
LTA	+365C>G	No	UK W (74/201)	[32]	
	+720C>A	No	UK W (74/201)	[32]	
IL6	-174G>C	No	UK W (74/201); AU W (22/140)	[31, 32]	
	Intron 4A>G	No	UK W (74/201)	[32] ^{##}	
Chemokine-related genes	IL10	+43G>A (G15R)	No	UK W (96/96)	[33]
	IL8	-353A>T	No	UK W (71/194)	[34]
+293G>T		No	UK W (71/194)	[34]	
+678T>C		No	UK W (71/194)	[34]	
+2607G>C		No	UK W (71/194)	[34]	
CXCR1	CXCR2	+785C>T	No	UK W (71/194)	[34]
		+1208T>C	No	UK W (71/194)	[34]
Th1/Th2 response genes	IL12B	+1440G>A	No	UK W (71/194)	[34]
		3'UTR 1188A>C	No	UK W (73/157)	[35]
Complement receptor genes	IFNG	3'UTR 5644A>G	No	UK W (73/157)	[35]
		CR1	+3650A>G	No	IT W (74/166)
		Intron 27 520T>C	No	IT W (74/166)	[36]
		+5507C>G (P1827R)	+5507G (OR 6.2)	IT W (74/166)	[36]
Pulmonary surfactant					
Surfactant protein genes	SFTPA1	6A, 6A ² , 6A ³ , 6A ^{4†}	6A ⁴ in NS only (OR 3.7)	MX (84/194)	[37]
	SFTPA2	1A, 1A ⁰ , 1A ¹ , 1A ² , 1A ^{5†}	No	MX (84/194)	[37]
	SFTPB	1580T>C (I131T)	1580C in Sm only (OR 7.6)	MX (84/194)	[37]
	SFTPC	438C>A (N138T)	No	MX (84/194)	[37]
		582G>A (N186S)	No	MX (84/194)	[37]
	SFTPD	ATG>ACG (M11T)	No	MX (84/194)	[37]
		ACA>GCA (T160A)	No	MX (84/194)	[37]
Coagulation cascade					
PAI genes	PAI1	Promoter region 4G/5G	No	USA W (62/2120)	[38]
Fibroblast-related pathways					
TGF-β genes	TGFB1	+869T>C (L10P)	No	ES W (128/140)	[39] ^{††}
		+915G>C (R25P)	No	ES W (128/140); AU W (22/140)	[31, 39]
RAAS-related genes	ACE	I/D	D	USA W (24/6000)	[40] ^{††}

Pop.: population; IL: interleukin; TNF: tumour necrosis factor; Th: T-helper cell; PAI: plasminogen activator inhibitor; TGF: transforming growth factor; RAAS: renin-angiotensin-aldosterone system; IL1RN: IL-1 receptor antagonist gene; TNFR2: TNF receptor II gene; LTA: lymphotoxin-α gene; CXCR: CXC chemokine receptor gene; IFNG: interferon-γ gene; CR: complement receptor; SFTP: surfactant protein gene; TGFB1: TGF-β1 gene; ACE: angiotensin-converting enzyme; C: cytosine; T (nucleotide); thymidine; G (nucleotide); guanine; A (nucleotide); adenine; >: substitution; VNTR: variable number of tandem repeats; G15R: arginine I substituted for glycine (G (amino acid)) 15; P: proline; I: isoleucine; T (amino acid): threonine; N: asparagines; S: serine; M: methionine; A (amino acid): alanine; L: leucine; UTR: untranslated region; I/D: insertion/deletion; OR: odds ratio (maximum reported in homozygous subjects); NS: nonsmokers; Sm: smokers; AU: Australia; CZ: Czech Republic; ES: Spain; IT: Italy; MX: Mexico; W: White. [#]: in strong linkage disequilibrium; [†]: allelic variants; ⁺: association between -889T and lower carbon monoxide diffusing capacity of the lung (DL_{CO}) [28]; [§]: no association in UK W-2 [32]; ^f: co-carriage of C-allele and IL6 intron 4G; ^{##}: association between IL6 intron 4GG genotype and lower DL_{CO}; ^{††}: association between +869C (10P) and deterioration in gas exchange; ^{††}: small and heterogeneous study cohort.

replication and, ultimately, the demonstration of a biologically functional effect. Negative findings do not rule out a gene when only a single nucleotide polymorphism has been studied. Before a candidate gene can be confidently removed from the list, detailed knowledge of linkage disequilibrium across the gene is regarded as a prerequisite.

Pro-/anti-inflammation

Interleukin-1 gene cluster

IL-1 comprises two structurally distinct forms, IL-1 α and IL-1 β . Both are potent proinflammatory cytokines with fibrogenic properties. The genes encoding IL-1 α (IL1A) and IL-1 β (IL1B), and their naturally occurring inhibitor IL-1 receptor antagonist (IL1RN), are localised in a cluster on chromosome 2q14. WHYTE *et al.* [30] have demonstrated an association between IL1RN and IPF, *i.e.* the rarer allele of the +2018C>T polymorphism conferred an increased risk of IPF in subjects from the UK and Italy. However, an IL1RN association with susceptibility to IPF could not be confirmed by others [29]. The same investigators did not find an association for multiple polymorphisms in IL1A and IL1B either [29]. However, an association between the -889T allele of IL1A and severity of gas transfer deficits in patients with IPF has been reported [28].

Tumour necrosis factor and related genes

Expression analysis of TNF- α in lung tissue from patients with IPF has shown elevated levels in (regenerating) type-II pneumocytes [41–43]. WHYTE *et al.* [30] first reported an association between the TNF gene and IPF. Carriage of the TNF -308A allele was associated with increased risk of IPF in well-defined UK (odds ratio (OR) 1.85) and Italian (OR 2.50) cohorts [30]. RIHA *et al.* [31] recently confirmed this association in 22 Australian IPF patients. However, a study by PANTELIDIS *et al.* [32] did not detect an association between the TNF -308A allele and IPF in UK patients. In addition, no association was found for multiple other polymorphisms in the TNF gene, and polymorphisms in the genes encoding TNF receptor II (TNFR2) and lymphotoxin- α [34]. Interestingly, the G allele of the IL6 intron 4 A>G polymorphism was associated with lower carbon monoxide diffusing capacity of the lung, and co-carriage of the TNFR2 1690C allele and the IL6 intron 4G allele was associated with presence of disease, suggesting a combinatory effect of these two genes in IPF development [32].

Cytokines considered to have an anti-inflammatory effect, for example IL-10, have a regulatory role affecting TNF- α production. Alveolar macrophages from patients with IPF exhibit elevated levels of TNF- α relative to IL-10, suggesting that, in this disease, the normal homeostatic mechanisms fail to regulate the TNF- α inflammatory response [33]. There is evidence that IL-10 gene haplotypes and promoter polymorphisms are associated with differential IL-10 production, and ~75% of interindividual variation in IL-10 production appears to be genetic [44, 45]. Recently, WHITTINGTON *et al.* [33] screened the coding sequence and 3' untranslated region of IL10 for polymorphisms in 96 patients with IPF. Although they identified a novel polymorphism at nucleotide position +43 from the start (G>A), causing an amino acid change and resulting in lower levels of IL-10 secretion, no difference in allele frequency between patients and controls was observed.

Chemokine genes

Neutrophil accumulation in the lower respiratory tract is a typical finding in IPF, and the recruitment and activation of neutrophils is believed to play a fundamental role in the development of lung injury that precedes normal repair. IL-8 has been found at increased levels in the lungs of patients with IPF, and several studies have shown a correlation between IL-8 levels in bronchoalveolar lavage fluid or expression of alveolar macrophage mRNA IL-8 and bronchoalveolar lavage fluid neutrophil count in this disease. Only one study has investigated single-nucleotide polymorphisms in the genes encoding IL-8 and its receptor, but no association was found [34].

Genes involved in T-helper cell type 1/type 2 response

Ample data support the role of T-helper cell type 2 regulatory cytokines in the pathogenesis of lung fibrosis [46]. This type of cytokine response is associated with high levels of IL-4, -5 and -13, and a paucity of interferon (IFN)- γ [47–49]. IFN- γ shows antifibrotic properties, and this imbalance might thus contribute to the excessive fibroblast activation, deposition of collagen and scar formation that occurs in IPF. Functional polymorphisms in genes encoding cytokines influencing T-helper cell type 1/type 2 balance, and especially IFN- γ and its receptors, have, therefore, been studied in IPF.

IL-12 plays a key role in inducing IFN- γ production, and a single nucleotide polymorphism in the 3' untranslated region of the IL12 p40 gene (*i.e.* IL12B) at position 1188 has been shown to correlate with increased IL-12 secretion [50]. LATSI *et al.* [35] found no association of this polymorphism with susceptibility to IPF. Similarly, they found equal distribution of a potentially functional polymorphism in the IFN- γ gene (position 5644 in the 3' untranslated region) in IPF and control subjects. However, absence of an association with just a single nucleotide polymorphism does not exclude a role for this gene in genetic susceptibility to IPF.

Complement receptor genes

Complement receptor 1 (CR1/CD35/C3b/C4b receptor) is important in the clearance of immune complexes. The C>G polymorphism at nucleotide position +5507 in exon 33 has been correlated with the levels of CR1 expressed on erythrocytes, which may directly affect the clearance of immune complexes [36]. Although there is no firm evidence for a role of immune complex clearance in the pathogenesis of IPF, this is an attractive concept. ZORZETTO *et al.* [36] analysed this and two other polymorphic sites in the CR1 gene using an Italian IPF cohort. They demonstrated a significant association between the GG genotype of the C5507G polymorphism and IPF (OR 6.2) [36]. As this GG genotype is thought to result in low expression of CR1 on erythrocytes, the authors speculated that, in a subset of subjects who develop IPF, CR1 polymorphisms related to a low CR1/erythrocyte ratio might contribute to impaired clearance of immune complexes containing viral particles and/or complement-opsinised viruses. In conjunction with other environmental and genetically determined factors, this could result in repeated episodes of lung injury and subsequent aberrant wound healing.

Pulmonary surfactant

Association of mutations in SFTPC with familial IPF suggests that genes involved in maintaining the alveolar structure and function are important for the development of IPF. In addition to SFTPC, SFTPA1, SFTPA2, SFTPB and SFTPD are also members of the surfactant protein gene family [51]. Using a Mexican IPF cohort, SELMAN *et al.* [37] evaluated gene polymorphisms in SP-A1, SP-A2, SP-B, SP-C and SP-D genes. Stratification of the study cohort based on their smoking habit demonstrated associations for the 6A⁴ haplotype of SFTPA1 with smoking IPF patients, and for the 1580C allele of SFTPB with smoking IPF patients. Functional analyses of the SP-A with different SFTPA1 haplotypes demonstrated 6A⁴-specific aggregation of SP-A. It is of note that the SFTPB 1580C allele has also been shown to confer risk of chronic obstructive pulmonary disease, making a specific role in IPF pathogenesis unlikely [52]. Furthermore, associations with sporadic IPF cases were not detected for multiple polymorphisms in SFTPC and SFTPD [37].

More recently, LAWSON *et al.* [53] reported SFTPC sequence results from 89 patients with sporadic IPF and found evidence for a role of genetic mutations in only 1% of cases. These findings may be in line with the results of AMIN *et al.* [20], who demonstrated preserved SP-C expression in bronchoalveolar lavage fluid and lung biopsy specimens from 19 unrelated IPF patients. Therefore, it is unlikely that SFTPC mutations contribute to the pathogenesis of IPF in the majority of sporadic cases.

Oxidative/antioxidative balance

Several studies have shown excessive oxidant production in IPF, as well as deficiencies in glutathione production [54, 55]. Increased production of oxidants has been suggested to contribute to epithelial cell apoptosis, which is an important early feature in the pathogenesis of pulmonary fibrosis. Loss of the alveolar epithelial protective barrier could subsequently lead to exposure of the underlying basement membrane to oxidative injury, resulting in degradation of key constituents of basement membrane. Loss of the integrity of the subepithelial basement membrane is commonly thought to be an essential and unique element in the progressive pulmonary fibrosis of IPF [56]. Therefore, genes involved in oxidative/antioxidative balances can be regarded as important candidates for genetic studies in IPF. However, to date, no such studies have been reported.

Coagulation cascade

Increased local procoagulant and antifibrinolytic activities have been found in patients with IPF, suggesting that fibrin matrix removal is reduced, and, as a consequence, fibroblast migration, extracellular matrix deposition and the magnitude of the fibrotic response are increased [57, 58]. Levels of tissue factor and plasminogen activator inhibitor (PAI)-1 and PAI-2 are significantly elevated in bronchoalveolar lavage fluid obtained from IPF patients, and the compensatory increase in tissue factor pathway inhibitor appears to be insufficient to counterbalance tissue factor, which leads to a hypercoagulable state in IPF lungs [59]. KIM *et al.* [38] have recently evaluated a polymorphism in the PAI-1 gene, but did not find an association with IIP. However, further systematic analysis of

these genes using functional genomic tools is warranted to determine their roles in the development of IPF.

Fibroblast-related pathways

Transforming growth factor- β

Fibroblast migration/proliferation and phenotypic change to myofibroblasts, with subsequent accumulation and remodeling of extracellular matrix are central to the pathogenesis of IPF [25, 46]. Transforming growth factor (TGF)- β 1 has been shown to be a critical mediator of lung fibrosis in animal models, and a number of studies have shown that antagonising TGF- β 1 prevents the development of tissue fibrosis [60]. Furthermore, targeted overexpression of TGF- β 1 has been shown to produce progressive fibrosis [61]. Interestingly, recent insights from rat models show that TGF- β 3 is associated with "regular" wound healing, whereas TGF- β 1 is associated with fibrotic wound healing [62]. TGF- β 3, therefore, seems to act as an endogenous counterbalance of TGF- β 1.

There is evidence that the circulating concentration of TGF- β 1 is under predominantly genetic control [63], implying that genetic variation at the TGF- β 1 gene (TGFB1) locus might influence diseases, including IPF, in which TGF- β 1 is implicated. Against this background, XAUBET *et al.* [39] have assessed polymorphisms in TGFB1 in White IPF patients from Spain. They studied two exon 1 polymorphisms, at positions +869 (T>C) and +915 (G>C), both resulting in amino acid substitutions. The authors found no association of either polymorphism with susceptibility to IPF. Interestingly, however, the +869C allele (codon 10 proline) was associated with increased deterioration of the disease, as measured by gas exchange [39]. Although this allele was not associated with other functional parameters of IPF, and the levels of TGF- β in both case and control subjects were not determined, the study provides the first evidence for TGFB1 as a potential determinant of disease progression in IPF. However, further evaluation using independent IPF cohorts is necessary to determine the true roles of the +869C allele and other polymorphisms in this gene.

Renin-angiotensin-aldosterone system

The local renin-angiotensin-aldosterone system in the lung might play a role in lung fibrogenesis. Angiotensin-converting enzyme (ACE) increases angiotensin II levels, and signalling *via* angiotensin II receptor type 1 (AT₁) has been shown to promote lung fibrosis in rats, possibly through upregulation of TGF- β 1 [64]. The ACE inhibitor captopril has an inhibitory effect on human lung fibroblasts *in vitro* [65], and inhibition of AT₁ has been shown to reduce bleomycin-induced lung fibrosis in rats [64]. However, a recent retrospective study from the Mayo Clinic could not demonstrate a beneficial survival effect of ACE inhibition in IPF [66].

A functional insertion/deletion polymorphism in intron 16 of the ACE gene has been found to be responsible for almost half of the variance in serum ACE levels [67, 68]. The deletion allele, correlating with high serum ACE levels, has been shown to associate with systemic sclerosis, a disease showing a high propensity for lung fibrosis [69]. In addition, MORRISON *et al.* [40] have detected an increased deletion allele frequency in a small cohort of mixed IIP cases (69 *versus* 54% in controls), which definitively requires further confirmation.

Extracellular matrix

An essential hallmark of IPF is the exorbitant production of extracellular matrix molecules, including collagen, tenascin and proteoglycans. There is clearly an imbalance between the production and degradation of extracellular matrix. In this context, matrix-degrading enzymes, such as matrix metalloproteinases (MMPs) and their inhibitors (tissue inhibitor of metalloproteinases (TIMPs)), are of interest. ZUO [70] has immunohistochemically demonstrated increased expression of MMP-7 (matrilysin) in IPF lungs, which has been confirmed by others [71]. Therefore, the MMP-7 gene might be an important candidate gene for further genetic analysis, although the exact mechanism by which this molecule plays a role remains largely unknown. In general, however, MMPs are thought not only to be involved in the breakdown and remodelling that occurs during injury, but may also cause the release of growth factors and cytokines known to influence growth and differentiation of target cells within the lung [72]. In addition, RAMOS *et al.* [73] have suggested increased production of TIMPs in IPF, accounting for the inability to degrade matrix. Promoter polymorphisms, which can influence the production levels of these molecules, are therefore of special interest in this respect.

Apoptosis

In normal wound healing, the number of fibroblasts gradually declines as the healing process or active fibrosis is successfully completed or terminated, but seems to persist in lung tissue from patients with IPF [74]. An important mechanism herein appears to be apoptosis. In the presence of TGF- β , fibroblasts are protected against apoptosis, but some studies have suggested increased rates of apoptosis [75]. In addition, the pattern of expression of pro-apoptotic proteins and inhibitors of apoptosis needs further study before candidate genes can be designated.

In addition to concepts of apoptosis, studies have investigated variations in mechanisms of wound repair, including p53. The p53 gene encodes a nuclear protein that binds to and modulates the expression of genes that play an important role in DNA repair, cell division and cell death by apoptosis. HOJO *et al.* [76] have demonstrated a close relationship between IPF and the presence of heterogeneous point mutations in fibrotic lung tissue. The occurrence of somatic p53 mutations in IPF is primarily interpreted in the context of the increased tumourigenesis seen in this disease. However, it is also tempting to speculate on a role for somatic mutations in this and other apoptosis-regulating genes in the pathogenesis of IPF itself.

Angiogenic/angiostatic balance

Angiogenesis is the formation of new vessels from pre-existing vasculature. The activation and resolution of angiogenesis is fundamental to wound healing. A microenvironment that encourages angiogenesis has been thought to be a key process in IPF. Although several studies have reported increased angiogenic activity in IPF [77, 78], recent evaluations of interstitial capillary density in IPF lung biopsy specimens could not demonstrate excessive neovascularisation in characteristic fibrotic lesions. However, one of these studies showed increased vascular density in areas of minimal fibrosis [78, 79].

Interestingly, COSGROVE *et al.* [80] have just described the presence of a potent angiostatic mediator, pigment epithelium-derived factor (PEDF), in IPF lungs. They demonstrated that vascular density is regionally decreased in IPF within the fibroblastic foci, and that, within these areas, PEDF levels were increased, whereas vascular endothelial growth factor levels were decreased. Furthermore, PEDF co-localised with TGF- β 1, suggesting that this fibrogenic cytokine might regulate PEDF expression. Taken together, these results suggest a role for dysregulated angiogenesis in the pathogenesis of IPF, and, more specifically, highlight the regional nature of a varied angiogenic response.

PULMONARY FIBROSIS IN SYSTEMIC SCLEROSIS

Systemic sclerosis (scleroderma) is a disease of unknown origin, characterised by excessive deposition of collagen and other connective tissue macromolecules in skin and multiple internal organs, prominent and often severe alternations in the microvasculature, and humoral and cellular immunological abnormalities. One of the major organ manifestations of this clinically heterogeneous disease is pulmonary fibrosis. Scleroderma lung fibrosis appears to be quite distinct from IPF, despite some superficial similarities of clinical, physiological and radiological features. Specifically, response to therapy and survival from this type of diffuse lung disease is better than in IPF even when matched for a variety of indices, including severity of disease at presentation and duration of disease. The histological pattern of scleroderma lung disease is most commonly NSIP [81] by comparison with the UIP pattern of disease that defines IPF. One of the key histological features of NSIP is a homogeneous distribution pattern of combinations of interstitial chronic inflammation and fibrosis. Areas of dense fibrosis and fibroblastic foci, as seen in UIP, are inconspicuous or absent.

The pathogenesis of systemic sclerosis is extremely complex [82], and, unlike in UIP/IPF, immunological inflammation seems to be a crucial component in the pathogenesis. Besides fundamental abnormalities in cells of the immune system, particularly T- and B-cells, alterations in fibroblasts and endothelial cells are intimately involved in the development of the clinical and pathological manifestations of the disease. These abnormalities result in the characteristic triad of pathological changes in systemic sclerosis: 1) humoral and cellular abnormalities, which include the production of numerous autoantibodies, chronic mononuclear cell infiltration of affected tissues, and dysregulation of lymphokine and growth factor production; 2) severe and often progressive cutaneous and visceral fibrosis; and 3) obliteration of the lumen of small arteries and arterioles. At present, it is not clear which of these alterations is of primary importance or how they interrelate to cause the progressive fibrotic process in systemic sclerosis.

Remarkably, >90% of patients with systemic sclerosis harbour antinuclear antibodies in their serum. There are three predominant antibodies: anticentromere antibodies, antitopoisomerase antibodies (ATAs), and anti-RNA polymerase antibodies. These autoantibodies are almost totally mutually exclusive and define different clinical subsets of the disease with reasonable accuracy [83, 84]. In particular, the presence of

ATAs (also known as anti-Scl-70 antibodies) in a scleroderma patient is very strongly associated with the risk of development of pulmonary fibrosis (relative risk 17) [83, 84]. In addition, ATA positivity was strongly associated with carriage of HLA-DRB1*11 alleles (previously included as part of HLA-DR5), *i.e.* 39 of 54 (72%) patients who tested positive for ATA carried these alleles *versus* ~18% carriership in controls ($p=0.00006$) [84]. Within HLA-DRB1*11, the allelic subtype *1104 seems to be associated with ATA positivity and the presence of lung fibrosis [85]. GILCHRIST *et al.* [84] also showed that HLA DPB1*1301 was tightly associated with ATA presence.

As the early phases of pathogenesis of systemic sclerosis are thought to involve a T-cell-mediated response to an antigenic trigger, possibly epitopes of DNA topoisomerase I, resulting in the production of antibodies directed against this enzyme (ATA), the described HLA associations focus on the major histocompatibility complex (MHC) region on chromosome 6 for the identification of predisposing genetic factors for this disease. The contribution of genetic factors is further supported by the observation of familial clustering of the disease, the high frequency of autoimmune disorders and autoantibodies in family members of patients with systemic sclerosis, and differences in prevalence and clinical manifestations among different ethnic groups [82, 86, 87]. Besides MHC-based genes (*i.e.* HLA genes and pro-inflammatory genes, such as TNF), non-MHC based genes encoding pro-/anti-inflammatory cytokines and chemokines, and genes involved in fibroblast and endothelial cell functioning are also important candidates for a role in the genetics underlying systemic sclerosis. The genes and gene polymorphisms that have been evaluated to date in diffuse systemic sclerosis, *i.e.* systemic sclerosis associated with pulmonary fibrosis, are discussed below and summarised in table 2.

Major histocompatibility complex

Strong associations have been found between HLA-DRB1*11, and also HLA-DPB1*1301, and diffuse systemic sclerosis, and there is some evidence to suggest that it is an amino-acid motif, shared by the different class II susceptibility alleles, that may be pivotal in predisposing to autoantibody formation [84, 85].

However, the MHC region contains multiple extended haplotypes, and, therefore, HLA associations might also point to another nearby gene on chromosome 6p21 that is causally involved in systemic sclerosis. The known linkage disequilibrium between the TNF locus and HLA class II genes has led to studies to define complex haplotypes in that region. SATO *et al.* [88] have fine mapped across the TNF locus in scleroderma subsets and found an association between the potentially functional TNF -857C>T polymorphism and lung fibrosis. Intergenic haplotype construction revealed, however, that this was fully explained by linkage disequilibrium with HLA-DRB1*11. Remarkably, they found that another TNF allele, TNF -863A, was very strongly associated with positivity for anticentromere antibodies and therefore “protective” against pulmonary fibrosis [88]. This polymorphism has proven functionality, *i.e.* the TNF -863A allele is associated with high TNF- α production *in vitro* [94]. This would suggest a differential pathogenetic role for TNF- α across different scleroderma subsets. TNF-related associations might have

immense practical implications now that anti-TNF strategies (*e.g.* infliximab or etanercept) have been proven to be highly effective in rheumatoid arthritis and Crohn’s disease, and are increasingly being applied to other chronic inflammatory processes. Therefore, potential intervention with specific strategies, such as TNF- α blockade in systemic sclerosis, might be possible, provided it can be directed to those clinical phenotypes most affected by the effects of this pro-inflammatory cytokine (conceivably anticentromere antibody-positive subjects according to the study of SATO *et al.* [88]).

Non-major histocompatibility complex

Chemokine-related genes

RENZONI *et al.* [34] studied variations in the gene encoding IL-8 and its receptors (CXCR1 and CXCR2). An increase in the frequency of the CXCR2 +785C and +1208T allele was found compared with controls in the systemic sclerosis patients. However, this was not found to be specific to those with lung fibrosis, being found in systemic sclerosis patients both with and without diffuse lung disease.

Cytokine genes

TGF- β 1 is a potent profibrotic cytokine secreted from activated macrophages and T-lymphocytes that may play a central role in the pathogenesis of tissue fibrosis in systemic sclerosis [82]. Two studies in systemic sclerosis patients have shown an association between a single nucleotide polymorphism at codon 10 of the gene encoding TGF- β 1 and pulmonary fibrosis (notably also associated with IPF progression), which might suggest that a higher production phenotype of this cytokine is particularly related to this clinical phenotype [89, 90].

Extracellular matrix

Variations in genes ontologically related to extracellular matrix homeostasis are of great interest in fibrosing lung diseases as a whole and, especially, lung fibrosis in systemic sclerosis. To date, only weak associations have been reported in the genes encoding fibronectin 1 and secreted protein acidic and rich in cysteine/osteonectin (table 2) [91, 92]. A genetically thorough study on the role of the fibrillin 1 gene (FBN1) in systemic sclerosis has been performed by TAN *et al.* [93]. Fibrillin 1 is the major constituent of extracellular microfibrils and has widespread distribution in both elastic and nonelastic connective tissue throughout the body. Mutations in FBN1 are the major cause of Marfan syndrome. In the study of TAN *et al.* [93], all 69 known exons of FBN1 were sequenced. Five single-nucleotide polymorphisms were identified and tested in Choctaw and Japanese systemic sclerosis patients and controls. On the basis of the genotype results, haplotypes were inferred, and two haplotypes (Hap-5 and Hap-6) were associated with systemic sclerosis in both populations [93]. Interestingly, one of these haplotypes (Hap-5) was specifically associated with lung fibrosis. These results are consistent with the hypothesis that FBN1 or a nearby gene on chromosome 15q is involved in systemic sclerosis (lung fibrosis) susceptibility.

CONCLUSION

It is disappointing that stronger associations between genetic polymorphisms and idiopathic fibrosing lung diseases have not yet been found, but the threshold of an exciting era awaits. Better definitions will improve phenotyping and the quality of

TABLE 2 Genes and gene polymorphisms that have been evaluated in systemic sclerosis (SSc)-associated pulmonary fibrosis

Gene ontology	Gene symbol	Polymorphism	Association with lung fibrosis in SSc	Pop. (cases/controls n)	Comments/other associations	[Refs]
MHC						
HLA genes	DRB1	HLA typing	DR11; DRB1*1104	UK (202/307); GR (98/130)		[84, 85]
	DPB1	HLA typing	DPB1*1301	UK (202/307)		[84]
TNF	TNF	-237G>A	No	UK W (214/354)		[88]
		-307G>A	No	UK W (214/354)		[88]
		-857C>T	-857T	UK W (214/354)	-857T in LD with DR11	[88]
		-863C>A	No	UK W (214/354)	-863A strongly associates with ACA+	[88]
		-1031T>C	No	UK W (214/354)	-1031C associates with ACA+	[88]
Non-MHC						
Chemokine-related genes	IL8	-353A>T	No	UK W (128/194)		[34]
		+293G>T	No	UK W (128/194)		[34]
		+678T>C	No	UK W (128/194)		[34]
	CXCR1	+2607G>C	No	UK W (128/194)		[34]
	CXCR2	+785C>T	+785C	UK W (128/194)	Also associated with SSc without lung fibrosis	[34]
+1208T>C		+1208T	UK W (128/194)	Also associated with SSc without lung fibrosis	[34]	
Cytokine genes	TGFB1	+1440G>A	No	UK W (128/194)		[34]
		-509C>T	-509T	JP (87/110)	Weak association	[89]
		+869T>C (L10P)	+869C (OR 2.0)	JP (87/110); UK (152/147)	Also associated with SSc without lung fibrosis in UK patients	[89, 90]
ECM	FN1	+915G>C (R25P)	No	UK (152/147)		[90]
		Haelll RFLP	Genotype AB	UK (161/253)		[91]
		MspI RFLP	No	UK (161/253)	Co-carriage of genotypes MspI CD and Haelll AB showed two-fold increased risk of developing lung fibrosis	[91]
	SPARC	HindIII RFLP	No	UK (161/253)		[91]
		TaqI RFLP	No	UK (161/253)		[91]
		+998C>G	No	USA W (94/70); AA (33/28); MA (51/45)		[92]
		+1551C>G	No	USA W (94/70); AA (33/28); MA (51/45)		[92]
+1922T>G	+1922T	USA W (94/70); AA (33/28); MA (51/45)	Weak association	[92]		
FBN1	Hap-1 -12	Hap-5	JP (106/100)	Hap-5/6 associated with SSc in JP and Choctaw pops	[93]	

Pop.: population; MHC: major histocompatibility complex; HLA: human leukocyte antigen; TNF: tumour necrosis factor; ECM: extracellular matrix; IL: interleukin; CXCR: CXC chemokine receptor gene; TGFB1: transforming growth factor-β1 gene; FN: fibronectin; FBN: fibrillin; SPARC: secreted protein acidic and rich in cysteine; G: guanine; A: adenine; C: cytosine; T: thymidine; >: substitution; L10P: proline (P) substituted for leucine (L) 10; R: arginine; RFLP: restriction fragment length polymorphism; Haelll: *Haemophilus aegyptius* III; MspI: *Moraxella* spp. I; HindIII: *Haemophilus influenzae* Rd III; TaqI: *Thermus aquaticus* YTI I; Hap: haplotype; OR: odds ratio (maximum reported in homozygous subjects); GR: Greece; JP: Japan; W: White; AA: African American; MA: Mexican American; LD: linkage disequilibrium; ACA+: anticentromere antibody positivity.

associations. The large USA familial study will also provide novel candidates, and whether familial disease shows similar predispositions to sporadic disease will be awaited with great interest. The dissection of key associations will undoubtedly require collaboration of the highest quality. Too often, small studies that cannot be reproduced are published. Working together will limit these spurious publications, save time and money and permit true identification of the key issues. The European Respiratory Society is encouraged to continue to

facilitate such collaborations both within its organisation and internationally.

REFERENCES

1 Cushley MJ, Davison AG, du Bois RM, *et al.* The diagnosis, assessment and treatment of diffuse parenchymal lung disease in adults. British Thoracic Society recommendations. *Thorax* 1999; 54: Suppl. 1, S1–S30.

- 2 American Thoracic Society/European Respiratory Society International Multidisciplinary Consensus Classification of the Idiopathic Interstitial Pneumonias. *Am J Respir Crit Care Med* 2002; 165: 277–304.
- 3 Steele MP, Speer MC, Loyd JE, *et al.* Familial pulmonary fibrosis (FPF) in the United States: clinical characteristics and patterns of inheritance. *Am J Respir Crit Care Med* 2004; 169: A550.
- 4 Coultas DB, Zumwalt RE, Black WC, Sobonya RE. The epidemiology of interstitial lung diseases. *Am J Respir Crit Care Med* 1994; 150: 967–972.
- 5 Marshall RP, Puddicombe A, Cookson WO, Laurent GJ. Adult familial cryptogenic fibrosing alveolitis in the United Kingdom. *Thorax* 2000; 55: 143–146.
- 6 Loyd JE. Pulmonary fibrosis in families. *Am J Respir Cell Mol Biol* 2003; 29: Suppl. 3, S47–S50.
- 7 Hodgson U, Laitinen T, Tukiainen P. Nationwide prevalence of sporadic and familial idiopathic pulmonary fibrosis: evidence of founder effect among multiplex families in Finland. *Thorax* 2002; 57: 338–342.
- 8 Thomas AQ, Lane K, Phillips J III, *et al.* Heterozygosity for a surfactant protein C gene mutation associated with usual interstitial pneumonitis and cellular nonspecific interstitial pneumonitis in one kindred. *Am J Respir Crit Care Med* 2002; 165: 1322–1328.
- 9 Bitterman PB, Rennard SI, Keogh BA, Wewers MD, Adelberg S, Crystal RG. Familial idiopathic pulmonary fibrosis. Evidence of lung inflammation in unaffected family members. *N Engl J Med* 1986; 314: 1343–1347.
- 10 Musk AW, Zilko PJ, Manners P, Kay PH, Kambh MI. Genetic studies in familial fibrosing alveolitis. Possible linkage with immunoglobulin allotypes (Gm). *Chest* 1986; 89: 206–210.
- 11 Nogee LM, Dunbar AE III, Wert SE, Askin F, Hamvas A, Whitsett JA. A mutation in the surfactant protein C gene associated with familial interstitial lung disease. *N Engl J Med* 2001; 344: 573–579.
- 12 Wang WJ, Mulugeta S, Russo SJ, Beers MF. Deletion of exon 4 from human surfactant protein C results in aggresome formation and generation of a dominant negative. *J Cell Sci* 2003; 116: 683–692.
- 13 Fisher JH, Emrie PA, Drabkin HA, *et al.* The gene encoding the hydrophobic surfactant protein SP-C is located on 8p and identifies an EcoRI RFLP. *Am J Hum Genet* 1988; 43: 436–441.
- 14 Glasser SW, Korfhagen TR, Perme CM, Pilot-Matias TJ, Kister SE, Whitsett JA. Two SP-C genes encoding human pulmonary surfactant proteolipid. *J Biol Chem* 1988; 263: 10326–10331.
- 15 Brasch F, Griese M, Tredano M, *et al.* Interstitial lung disease in a baby with a *de novo* mutation in the SFTPC gene. *Eur Respir J* 2004; 24: 30–39.
- 16 Tredano M, Griese M, Brasch F, *et al.* Mutation of SFTPC in infantile pulmonary alveolar proteinosis with or without fibrosing lung disease. *Am J Med Genet* 2004; 126A: 18–26.
- 17 Riva A, Kohane IS. SNPper: retrieval and analysis of human SNPs. *Bioinformatics* 2002; 18: 1681–1685. Version 118. <http://snpper.chip.org/bio/show-sequence/?TYPE=U&GENE=18686>. Date last accessed: October 6 2004.
- 18 Whitsett JA. Genetic basis of familial interstitial lung disease: misfolding or function of surfactant protein C? *Am J Respir Crit Care Med* 2002; 165: 1201–1202.
- 19 Baatz JE, Smyth KL, Whitsett JA, Baxter C, Absolom DR. Structure and functions of a dimeric form of surfactant protein SP-C: a Fourier transform infrared and surfactometry study. *Chem Phys Lipids* 1992; 63: 91–104.
- 20 Amin RS, Wert SE, Baughman RP, *et al.* Surfactant protein deficiency in familial interstitial lung disease. *J Pediatr* 2001; 139: 85–92.
- 21 Anderson PD, Huizing M, Claassen DA, White J, Gahl WA. Hermansky-Pudlak syndrome type 4 (HPS-4): clinical and molecular characteristics. *Hum Genet* 2003; 113: 10–17.
- 22 Dell'Angelica EC, Ooi CE, Bonifacino JS. β 3A-adaptin, a subunit of the adaptor-like complex AP-3. *J Biol Chem* 1997; 272: 15078–15084.
- 23 Huizing M, Sarangarajan R, Strovel E, Zhao Y, Gahl WA, Boissy RE. AP-3 mediates tyrosinase but not TRP-1 trafficking in human melanocytes. *Mol Biol Cell* 2001; 12: 2075–2085.
- 24 Nakatani Y, Nakamura N, Sano J, *et al.* Interstitial pneumonia in Hermansky-Pudlak syndrome: significance of florid foamy swelling/degeneration (giant lamellar body degeneration) of type-2 pneumocytes. *Virchows Arch* 2000; 437: 304–313.
- 25 Selman M, King TE, Pardo A. Idiopathic pulmonary fibrosis: prevailing and evolving hypotheses about its pathogenesis and implications for therapy. *Ann Intern Med* 2001; 134: 136–151.
- 26 Selman M, Pardo A. The epithelial/fibroblastic pathway in the pathogenesis of idiopathic pulmonary fibrosis. *Am J Respir Cell Mol Biol* 2003; 29: Suppl. 3, S93–S97.
- 27 Colhoun HM, McKeigue PM, Davey SG. Problems of reporting genetic associations with complex outcomes. *Lancet* 2003; 361: 865–872.
- 28 du Bois RM. The genetic predisposition to interstitial lung disease: functional relevance. *Chest* 2002; 121: Suppl. 3, 14S–20S.
- 29 Hutvyrova B, Pantelidis P, Drabek J, *et al.* Interleukin-1 gene cluster polymorphisms in sarcoidosis and idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2002; 165: 148–151.
- 30 Whyte M, Hubbard R, Meliconi R, *et al.* Increased risk of fibrosing alveolitis associated with interleukin-1 receptor antagonist and tumor necrosis factor- α gene polymorphisms. *Am J Respir Crit Care Med* 2000; 162: 755–758.
- 31 Riha RL, Yang IA, Rabnott GC, Tunnicliffe AM, Fong KM, Zimmerman PV. Cytokine gene polymorphisms in idiopathic pulmonary fibrosis. *Intern Med J* 2004; 34: 126–129.
- 32 Pantelidis P, Fanning GC, Wells AU, Welsh KI, du Bois RM. Analysis of tumor necrosis factor- α , lymphotoxin- α , tumor necrosis factor receptor II, and interleukin-6 polymorphisms in patients with idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2001; 163: 1432–1436.
- 33 Whittington HA, Freeburn RW, Godinho SI, Egan J, Haider Y, Millar AB. Analysis of an IL-10 polymorphism in idiopathic pulmonary fibrosis. *Genes Immun* 2003; 4: 258–264.
- 34 Renzoni E, Lympany P, Sestini P, *et al.* Distribution of novel polymorphisms of the interleukin-8 and CXCR1 and 2 genes in systemic sclerosis and cryptogenic fibrosing alveolitis. *Arthritis Rheum* 2000; 43: 1633–1640.

- 35 Latsi P, Pantelidis P, Vassilakis D, Sato H, Welsh KL, du Bois RM. Analysis of IL-12 p40 subunit gene and IFN- γ G5644A polymorphisms in idiopathic pulmonary fibrosis. *Respir Res* 2003; 4: 6. <http://respiratory-research.com/content/4/1/6>.
- 36 Zorzetto M, Bombieri C, Ferrarotti I, et al. Complement receptor 1 gene polymorphisms in sarcoidosis. *Am J Respir Cell Mol Biol* 2002; 27: 17–23.
- 37 Selman M, Lin HM, Montano M, et al. Surfactant protein A and B genetic variants predispose to idiopathic pulmonary fibrosis. *Hum Genet* 2003; 113: 542–550.
- 38 Kim KK, Flaherty KR, Long Q, et al. A plasminogen activator inhibitor-1 promoter polymorphism and idiopathic interstitial pneumonia. *Mol Med* 2003; 9: 52–56.
- 39 Xaubet A, Marin-Arguedas A, Lario S, et al. Transforming growth factor- β 1 gene polymorphisms are associated with disease progression in idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2003; 168: 431–435.
- 40 Morrison CD, Papp AC, Hejmanowski AQ, Addis VM, Prior TW. Increased D allele frequency of the angiotensin-converting enzyme gene in pulmonary fibrosis. *Hum Pathol* 2001; 32: 521–528.
- 41 Pan LH, Ohtani H, Yamauchi K, Nagura H. Co-expression of TNF alpha and IL-1 beta in human acute pulmonary fibrotic diseases: an immunohistochemical analysis. *Pathol Int* 1996; 46: 91–99.
- 42 Nash JR, McLaughlin PJ, Butcher D, Corrin B. Expression of tumour necrosis factor- α in cryptogenic fibrosing alveolitis. *Histopathology* 1993; 22: 343–347.
- 43 Piguat PF, Ribaux C, Karpuz V, Grau GE, Kapanci Y. Expression and localization of tumor necrosis factor- α and its mRNA in idiopathic pulmonary fibrosis. *Am J Pathol* 1993; 143: 651–655.
- 44 Eskdale J, Gallagher G, Verweij CL, Keijsers V, Westendorp RG, Huizinga TW. Interleukin 10 secretion in relation to human IL-10 locus haplotypes. *Proc Natl Acad Sci U S A* 1998; 95: 9465–9470.
- 45 Turner DM, Williams DM, Sankaran D, Lazarus M, Sinnott PJ, Hutchinson IV. An investigation of polymorphism in the interleukin-10 gene promoter. *Eur J Immunogenet* 1997; 24: 1–8.
- 46 Sime PJ, O'Reilly KM. Fibrosis of the lung and other tissues: new concepts in pathogenesis and treatment. *Clin Immunol* 2001; 99: 308–319.
- 47 Ando M, Miyazaki E, Fukami T, Kumamoto T, Tsuda T. Interleukin-4-producing cells in idiopathic pulmonary fibrosis: an immunohistochemical study. *Respirology* 1999; 4: 383–391.
- 48 Wallace WA, Ramage EA, Lamb D, Howie SE. A type 2 (Th2-like) pattern of immune response predominates in the pulmonary interstitium of patients with cryptogenic fibrosing alveolitis (CFA). *Clin Exp Immunol* 1995; 101: 436–441.
- 49 Prior C, Haslam PL. *In vivo* levels and *in vitro* production of interferon-gamma in fibrosing interstitial lung diseases. *Clin Exp Immunol* 1992; 88: 280–287.
- 50 Seegers D, Zwiers A, Strober W, Pena AS, Bouma G. A TaqI polymorphism in the 3'UTR of the IL-12 p40 gene correlates with increased IL-12 secretion. *Genes Immun* 2002; 3: 419–423.
- 51 Pantelidis P, Veeraraghavan S, du Bois RM. Surfactant gene polymorphisms and interstitial lung diseases. *Respir Res* 2002; 3: 14. <http://respiratory-research.com/content/3/1/14>.
- 52 Guo X, Lin HM, Lin Z, et al. Surfactant protein gene A, B, and D marker alleles in chronic obstructive pulmonary disease of a Mexican population. *Eur Respir J* 2001; 18: 482–490.
- 53 Lawson WE, Grant SW, Ambrosini V, et al. Genetic mutations in surfactant protein C are a rare cause of sporadic cases of IPF. *Thorax* 2004; 59: 977–980.
- 54 Cantin AM, Hubbard RC, Crystal RG. Glutathione deficiency in the epithelial lining fluid of the lower respiratory tract in idiopathic pulmonary fibrosis. *Am Rev Respir Dis* 1989; 139: 370–372.
- 55 Crystal RG. Oxidants and respiratory tract epithelial injury: pathogenesis and strategies for therapeutic intervention. *Am J Med* 1991; 91: Suppl. 3C, 39S–44S.
- 56 Katzenstein AL. Pathogenesis of "fibrosis" in interstitial pneumonia: an electron microscopic study. *Hum Pathol* 1985; 16: 1015–1024.
- 57 Kotani I, Sato A, Hayakawa H, Urano T, Takada Y, Takada A. Increased procoagulant and antifibrinolytic activities in the lungs with idiopathic pulmonary fibrosis. *Thromb Res* 1995; 77: 493–504.
- 58 Imokawa S, Sato A, Hayakawa H, Kotani M, Urano T, Takada A. Tissue factor expression and fibrin deposition in the lungs of patients with idiopathic pulmonary fibrosis and systemic sclerosis. *Am J Respir Crit Care Med* 1997; 156: 631–636.
- 59 Fujii M, Hayakawa H, Urano T, et al. Relevance of tissue factor and tissue factor pathway inhibitor for hypercoagulable state in the lungs of patients with idiopathic pulmonary fibrosis. *Thromb Res* 2000; 99: 111–117.
- 60 Noble PW. Idiopathic pulmonary fibrosis. New insights into classification and pathogenesis usher in a new era therapeutic approaches. *Am J Respir Cell Mol Biol* 2003; 29, Suppl. 3, S27–S31.
- 61 Sime PJ, Xing Z, Graham FL, Csaky KG, Gauldie J. Adenovector-mediated gene transfer of active transforming growth factor- β 1 induces prolonged severe fibrosis in rat lung. *J Clin Invest* 1997; 100: 768–776.
- 62 Kolb M, Bonniaud P, Ask K, et al. TGF β 3 has differential effects in the lung dependent on the underlying pathogenetic condition. *Eur Respir J* 2003; 24: Suppl. 48, 481s.
- 63 Grainger DJ, Heathcote K, Chiano M, et al. Genetic control of the circulating concentration of transforming growth factor type β 1. *Hum Mol Genet* 1999; 8: 93–97.
- 64 Otsuka M, Takahashi H, Shiratori M, Chiba H, Abe S. Reduction of bleomycin induced lung fibrosis by candesartan cilexetil, an angiotensin II type 1 receptor antagonist. *Thorax* 2004; 59: 31–38.
- 65 Nguyen L, Ward WF, Ts'ao CH, Molteni A. Captopril inhibits proliferation of human lung fibroblasts in culture: a potential antifibrotic mechanism. *Proc Soc Exp Biol Med* 1994; 205: 80–84.
- 66 Nadrous HF, Ryu JH, Douglas WW, Decker PA, Olson EJ. Impact of angiotensin-converting enzyme inhibitors and statins on survival in idiopathic pulmonary fibrosis. *Chest* 2004; 126: 438–446.

- 67 Cambien F, Alhenc-Gelas F, Herbeth B, *et al.* Familial resemblance of plasma angiotensin-converting enzyme level: the Nancy Study. *Am J Hum Genet* 1988; 43: 774–780.
- 68 Rigat B, Hubert C, Alhenc-Gelas F, Cambien F, Corvol P, Soubrier F. An insertion/deletion polymorphism in the angiotensin I-converting enzyme gene accounting for half the variance of serum enzyme levels. *J Clin Invest* 1990; 86: 1343–1346.
- 69 Fatini C, Gensini F, Sticchi E, *et al.* High prevalence of polymorphisms of angiotensin-converting enzyme (I/D) and endothelial nitric oxide synthase (Glu298Asp) in patients with systemic sclerosis. *Am J Med* 2002; 112: 540–544.
- 70 Zuo F, Kaminski N, Eugui E, *et al.* Gene expression analysis reveals matrilysin as a key regulator of pulmonary fibrosis in mice and humans. *Proc Natl Acad Sci U S A* 2002; 99: 6292–6297.
- 71 Kaminski N. Microarray analysis of idiopathic pulmonary fibrosis. *Am J Respir Cell Mol Biol* 2003; 29: Suppl. 3, S32–S36.
- 72 Winkler MK, Fowlkes JL. Metalloproteinase and growth factor interactions: do they play a role in pulmonary fibrosis? *Am J Physiol Lung Cell Mol Physiol* 2002; 283: L1–L11.
- 73 Ramos C, Montano M, Garcia-Alvarez J, *et al.* Fibroblasts from idiopathic pulmonary fibrosis and normal lungs differ in growth rate, apoptosis, and tissue inhibitor of metalloproteinases expression. *Am J Respir Cell Mol Biol* 2001; 24: 591–598.
- 74 Phan SH. Fibroblast phenotypes in pulmonary fibrosis. *Am J Respir Cell Mol Biol* 2003; 29: Suppl. 3, S87–S92.
- 75 du Bois RM. Sarcoidosis: do our genes really matter? *Am J Respir Crit Care Med* 2001; 164: 725–726.
- 76 Hojo S, Fujita J, Yamadori I, *et al.* Heterogeneous point mutations of the p53 gene in pulmonary fibrosis. *Eur Respir J* 1998; 12: 1404–1408.
- 77 Keane MP, Arenberg DA, Lynch JP III, *et al.* The CXC chemokines, IL-8 and IP-10, regulate angiogenic activity in idiopathic pulmonary fibrosis. *J Immunol* 1997; 159: 1437–1443.
- 78 Renzoni EA. Neovascularization in idiopathic pulmonary fibrosis: too much or too little? *Am J Respir Crit Care Med* 2004; 169: 1179–1180.
- 79 Ebina M, Shimizukawa M, Shibata N, *et al.* Heterogeneous increase in CD34-positive alveolar capillaries in idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2004; 169: 1203–1208.
- 80 Cosgrove GP, Brown KK, Schiemann WP, *et al.* Pigment epithelium-derived factor in idiopathic pulmonary fibrosis: a role in aberrant angiogenesis. *Am J Respir Crit Care Med* 2004; 170: 242–251.
- 81 Bouros D, Wells AU, Nicholson AG, *et al.* Histopathologic subsets of fibrosing alveolitis in patients with systemic sclerosis and their relationship to outcome. *Am J Respir Crit Care Med* 2002; 165: 1581–1586.
- 82 Jimenez SA, Derk CT. Following the molecular pathways toward an understanding of the pathogenesis of systemic sclerosis. *Ann Intern Med* 2004; 140: 37–50.
- 83 Briggs DC, Vaughan RW, Welsh KI, Myers A, Dubois RM, Black CM. Immunogenetic prediction of pulmonary fibrosis in systemic sclerosis. *Lancet* 1991; 338: 661–662.
- 84 Gilchrist FC, Bunn C, Foley PJ, *et al.* Class II HLA associations with autoantibodies in scleroderma: a highly significant role for HLA-DP. *Genes Immun* 2001; 2: 76–81.
- 85 Vlachoyiannopoulos PG, Dafni UG, Pakas I, Spyropoulou-Vlachou M, Stavropoulos-Giokas C, Moutsopoulos HM. Systemic scleroderma in Greece: low mortality and strong linkage with HLA-DRB1*1104 allele. *Ann Rheum Dis* 2000; 59: 359–367.
- 86 Johnson RW, Tew MB, Arnett FC. The genetics of systemic sclerosis. *Curr Rheumatol Rep* 2002; 4: 99–107.
- 87 Mayes MD. Scleroderma epidemiology. *Rheum Dis Clin North Am* 2003; 29: 239–254.
- 88 Sato H, Lagan AL, Alexopoulou C, *et al.* The TNF-863A allele strongly associates with anticentromere antibody positivity in scleroderma. *Arthritis Rheum* 2004; 50: 558–564.
- 89 Sugiura Y, Banno S, Matsumoto Y, *et al.* Transforming growth factor β 1 gene polymorphism in patients with systemic sclerosis. *J Rheumatol* 2003; 30: 1520–1523.
- 90 Crilly A, Hamilton J, Clark CJ, Jardine A, Madhok R. Analysis of transforming growth factor β 1 gene polymorphisms in patients with systemic sclerosis. *Ann Rheum Dis* 2002; 61: 678–681.
- 91 Avila JJ, Lympany PA, Pantelidis P, Welsh KI, Black CM, du Bois RM. Fibronectin gene polymorphisms associated with fibrosing alveolitis in systemic sclerosis. *Am J Respir Cell Mol Biol* 1999; 20: 106–112.
- 92 Zhou X, Tan FK, Reveille JD, *et al.* Association of novel polymorphisms with the expression of SPARC in normal fibroblasts and with susceptibility to scleroderma. *Arthritis Rheum* 2002; 46: 2990–2999.
- 93 Tan FK, Wang N, Kuwana M, *et al.* Association of fibrillin 1 single-nucleotide polymorphism haplotypes with systemic sclerosis in Choctaw and Japanese populations. *Arthritis Rheum* 2001; 44: 893–901.
- 94 Udalova IA, Richardson A, Denys A, *et al.* Functional consequences of a polymorphism affecting NF- κ B p50–p50 binding to the TNF promoter region. *Mol Cell Biol* 2000; 20: 9113–9119.