Department of Pulmonary Medicine, Helsinki University Central Hospital
and
Department of Medical Genetics, University of Helsinki

IDIOPATHIC PULMONARY FIBROSIS -

FROM EPIDEMIOLOGY
TO GENE MAPPING

Ulla Hodgson

Academic dissertation

To be publicly discussed with the permission of the Faculty of Medicine, University of Helsinki, in Lecture Hall 4, Meilahti Hospital, Helsinki, on the March 3rd 2006, at 12 noon.

Helsinki 2006

Supervised by

Docent Tarja Laitinen, M.D., Ph.D.
Department of Medical Genetics
University of Helsinki

Professor Pentti Tukiainen, M.D., Ph.D.
Department of Pulmonary Diseases
Helsinki University Central Hospital

Reviewed by

Docent Lauri Tammilehto, M.D., Ph.D.
Department of Pulmonary Diseases
Jorvi Hospital
Helsinki University Central Hospital

Docent Markus Perola, M.D., Ph.D.
Department of Molecular Medicine
National Public Health Institute

Official Opponent

Professor Leif Groop, M.D., Ph.D.
Department of Clinical Sciences/Diabetes & Endocrinology
Lund University
University Hospital Malmö

As a well spent day bring happy sleep,
So life, well used, bring happy death.
Time abides long enough,
For those who make use of it.

- Leonardo da Vinci -

CONTENTS

LIST OF ORIGINAL PUBLICATIONS

This thesis is based on the following original publications, which are referred to in the text by their Roman numerals.

I   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.

II   Hodgson U, Pulkkinen V, Dixon M, Peyrard-Janvid M, Rehn M, Lahermo P, Ollikainen V, Salmenkivi K, Kinnula V, Kere J, Tukiainen P, Laitinen T. ELMOD2 is a Candidate Gene for Familial Idiopathic Pulmonary Fibrosis. Submitted.

III Hodgson U, Tukiainen P, Laitinen T. The polymorphism C5507G of Complement Receptor 1 does not explain Idiopathic Pulmonary Fibrosis among the Finns. Respiratory Medicine 2005;99:265-7.

IV   Kinnula V, Hodgson U, Lakari E, Tan R, Sormunen R, Soini Y, Kakko S, Laitinen T, Oury T, Pääkkö P. Extracellular superoxide dismutase has highly specific localization in idiopathic pulmonary fibrosis/ usual interstitial pneumonia. Histopathology. In press.

ABBREVIATIONS

AEC	alveolar epithelial cell
AIP	acute interstitial pneumonia
Arg	arginine
ATS	American Thoracic Society
BAL	bronchoalveolar lavage
COP	cryptogenic organizing pneumonia
CR1	complement receptor 1
DIP	desquamative interstitial pneumonia
EBV	Epstain-Barr virus
ECM	extracellular matrix
ELMOD2	ELMO domain containing 2
ERS	European Respiratory Society
ECSOD	extracellular superoxide dismutase
EST	expressed sequence tag
ET	endothelin
FGF	fibroblast growth factor
FPF	familial pulmonary fibrosis
Gly	glysine
HLA	human leucocyte antigen
HRCT	high resolution computed tomography
IBD	identical-by-descent
IBS	identical-by-state
IIP	idiopathic interstitial pneumonia
IPF	idiopathic pulmonary fibrosis
LD	linkage disequilibrium
LIP	lymphoid interstitial pneumonia
LOD	logarithm of odds
MHC	major histocompatibility complex

MMP	matrix metalloproteinase
NO	nitric oxide
NPL	non-parametric linkage
NSIP	non-specific interstitial pneumonia
OR	odds ratio
PAI	plasminogen activator inhibitor
PCR	polymerase chain reaction
PDGF	platelet derived growth factor
Pro	proline
RA	rheumatoid arthritis
RBILD	respiratory bronchiolitis with interstitial pneumonia
RNS	reactive nitrogen species
ROS	reactive oxygen species
SNP	single nucleotide polymorphism
SOD	superoxide dismutase
SP-C	surfactant protein C
TBB	transbronchial biopsy
TGF-β	transforming growth factor-β
TIMP	tissue inhibitor of matrix metalloproteinase
TNF-α	tumor necrosis factor-α
UIP	usual interstitial pneumonia
VATS	video-assisted thoracoscopy
VEGF	vascular endothelial growth factor

ABSTRACT

Idiopathic pulmonary fibrosis is the most common of the idiopathic interstitial pneumonias, and is distinguished from other interstitial pneumonias by the histological pattern, clinical manifestation, and poor outcome with usual survival of less than three years. There is no curative treatment yet available. The pathogenesis and etiology of idiopathic pulmonary fibrosis are unknown. The reports of multiple affected family members in the same family observed worldwide support the influence of genetic factors in the etiology.

In this study we evaluated the first nationwide prevalence of idiopathic pulmonary fibrosis using the recent diagnostic criteria. In Finland the prevalence of idiopathic pulmonary fibrosis was 16-18/100 000 inhabitants. The prevalence revealed variation in geographical distribution. In eastern and southern Savo, the prevalence was 45/100 000 inhabitants. We identified multiplex families with idiopathic pulmonary fibrosis, and calculated that the prevalence of familial idiopathic pulmonary fibrosis is 5.9/1 million population in Finland, explaining 3.3-3.7% of all idiopathic pulmonary fibrosis. The origins of the patients with familial idiopathic pulmonary fibrosis tended to cluster within the Savo and the nearby Carelia regions. Although no obvious genealogical loops between the families were observed, the clustering can be explained by a founder effect, i.e. the affected family members most likely share a common disease causing allele introduced by a common ancestor.

Parallel to the mapping of novel susceptibility genes for idiopathic pulmonary fibrosis, we tried to identify polymorphisms within candidate genes which could functionally play a part in the pathogenesis of the disease. Increased activity of extracellular superoxide dismutase in the extracellular matrix protects against lung injury caused by free radicals, and a synonymous Arg213Gly polymorphism in the extracellular superoxide dismutase gene is reported to result in higher serum levels. To study the possibility of an association of the Arg213Gly polymorphism with idiopathic pulmonary fibrosis we screened 63 patients and 61 population based controls. One of the patients and three controls carried the Gly213 allele, thus no association was detected. Association with the Pro1827Arg polymorphism of the Complement receptor 1 gene was studied among 96 patients and 164 controls. None of the 520 chromosomes studied carried the Arg1827 allele.

9

We performed a genome-wide scan with six multiplex families. Three regions on chromosomes 3, 4, and 13 obtained NPL scores of 1.7, 1.7, and 1.6, respectively, and on chromosomes 9 and 12 possible shared haplotypes were seen. These five loci were selected for fine mapping in an extended data set with the original six pedigrees, two additional multiplex families, four singletons, and 12 trios originating from the enrichment area. After hierarchial fine mapping with 63 markers on chromosomes 3q13 and 13q31 the NPL scores increased to 2.1 and 2.4, respectively, but no shared haplotype associated with the disease was detected, and the shared haplotypes on chromosomes 9 and 12 broke down. On chromosome 4q31.1 the NPL score increased to 2.1, and a 110 kb shared haplotype was carried by one third of the affected families (8/24), while none of the unaffected family members carried it. The carriership of the susceptibility haplotype was 34% among all the genotyped uniplex and multiplex families (12/35), and 7.7% (11/143) among 143 controls (27 family based, 23 regional, and 93 Finnish unrelated controls). The susceptibility haplotype obtained an odds ratio of 6.3 (p=0.0001, 95%CI=2.3–15.9). The critical region harbors two novel candidate genes, ELMOD2, and LOC152586. An *in vitro* translation assay of LOC152586, however, failed to yield a stable polypeptide. mRNA expression of ELMOD2 was decreased in lung biopsies derived from patients suffering from idiopathic pulmonary fibrosis (N=6) compared to healthy controls (N=7). Based on its expression and functional properties, potentially being involved in apoptosis, phagocytosis, cell engulfment, and cell migration, ELMOD2 is a prime positional candidate susceptibility gene for familial IPF.

INTRODUCTION

Idiopathic pulmonary fibrosis (IPF) is a chronic progressive lung disease. Patients suffer from cough and dyspnoe for two years on average before diagnosis (King et al. 2001). According to the recent international diagnostic criteria, IPF refers to the histological pattern of usual interstitial pneumonia (UIP) (ATS 2000). IPF differs from other idiopathic interstitial pneumonias (IIP) in terms of disease pattern, response to immunosuppressive therapy, and lethal outcome. The average survival after diagnosis is less than three years (Nicholson et al. 2000, Collard et al. 2003). No curative treatment is yet available.

The etiology and pathogenesis of IPF remain unknown. IPF is a multifactorial disease in which genetic factors also interact. The most convincing evidence of the importance of genetic factors is based on reports on familial cases worldwide (Peabody et al.1950, Bonnani et al. 1965, Javaheri et al. 1980, Uchiyama et al. 1997, Marshall et al. 2000, Thomas et al. 2002, Lee et al. 2005, Steele et al. 2005). Familial and sporadic IPF do not seem to differ in either their clinical characteristics or outcome (Marshall et al. 2000, Lee et al. 2005). Thus results from studies on familial IPF can reveal genes and molecular signalling pathways also important in sporadic IPF. Attempts to study the genetics of IPF have been mostly based on association studies with candidate genes (Whyte et al. 2000, Hutyrova et al. 2002, Latsi et al. 2003, Zorzetto et al. 2003, Lawson et al. 2004), but the reported associations have remained unconfirmed. IPF affects patients at older ages, which offers a challenge to gather informative multiplex families to perform linkage studies. Defining a precise phenotype is crucial; several interstitial pneumonias have been misdiagnosed as IPF. The recent diagnostic criteria and better imaging technologies such as high resolution computed tomography (HRCT) are likely to reduce the confusion. Based on the difficulties in phenotyping and in collecting multiplex families genome wide scans of IPF have not previously been published.

Trying to detect linkage in a complex disease using genetically isolated populations has been shown to be in some cases advantageous (Lander and Schork 1994, Risch 2000). The Finns have lived in isolation for centuries, and the population remained small until it expanded only during the last century. The genetic diversity is limited, shaped by founder effects, genetic drift, and isolation (de la Chapelle and Wright 1998, Peltonen et al. 2000, Kere 2001). Founder populations, such as the Finns, exhibit linkage disequilibrium and haplotype sharing over long genetic distances, and patients have most probably inherited the same disease-causing allele from a common ancestor.

Using population isolates has resulted in successful identification of disease-related genes and molecular genetic mechanisms in complex diseases such as asthma and hyperlipidemia (Laitinen et al 2004, Pajukanta et al. 2004). With the advantages of population isolates and precise phenotyping, mapping genes responsible for IPF may be successful.

This study attempted to identify genetic factors that may be relevant in the pathogenesis of IPF. We started with an epidemiological study by evaluating the prevalence of IPF in Finland, which was the first nationwide prevalence study using the novel international classification, and we identified multiplex families with at least two family members with IPF. Polymorphisms Pro1827Arg in the gene for Complement receptor 1 and Arg213Gly for the ECSOD gene had been verified and are likely to result in functional changes (Xiang et al. 1999, Folz and Crapo 1996). Pro1827Arg had already been suggested to be associated with IPF (Zorzetto et al. 2003). We aimed to study the association of these polymorphisms with IPF among Finnish patients. With multiplex families we performed a genome-wide scan to detect possible linkage to IPF. With an additional study population, we went further to perform haplotype association studies aimed at limiting a critical region in order to identify candidate genes for IPF using the methods of positional cloning.

REVIEW OF THE LITERATURE

1. IPF

1.1 From history to novel classification

The earliest description of IPF is considered to be a publication by Hamman and Rich in 1944. Three patients, who were treated at the John Hopkins Hospital from the years 1931 to 1933, suffered from extreme dyspnoea and cyanosis in a way that was new to the physicians. The patients died within three weeks to three months after entering the hospital. At autopsy the lungs showed widespread connective tissue hyperplasia throughout the interstitial structures, and the alveolar walls were replaced by scar tissue. In 1967, Scadding and Hinson reported on sixteen patients with diffuse fibrosing alveolitis. The progressive inflammatory and fibrous interstitial diseases were verified by surgical lung biopsy performed by thoracotomy. Because it became obvious that idiopathic intersitial diseases included various clinical conditions, they have been classified into distinguishable entities according to their histological appearance since 1969. The first pathological classification by Liebow and Carrington (1969) included five subtypes: usual interstitial pneumonia (UIP), desquamative interstitial pneumonia (DIP), bronchiolitis obliterans with interstitial pneumonia (BIP), lymphoid interstitial pneumonia (LIP), and giant cell interstitial pneumonia (GIP). In the 1980`s, large studies described the clinical features and outcome of IPF, or as it was previously called, cryptogenic fibrosing alveolitis (Turner-Warwick et al. 1980, Tukiainen et al. 1983). In 1998, Katzenstein and Myers offered a renewed classification based upon the identification and definition of new histopathological entities. They divided the IIPs into four subtypes: UIP, DIP, acute interstitial pneumonia (AIP), and non-specific interstitial pneumonia (NSIP). Because the overlapping of the clinical entities – their symptoms, radiological and histological findings, and the diseases courses – caused some confusion in the diagnostic criteria and terminology, there was a need for a novel international standard. In 2002 an international consensus statement defining the clinical, radiologic, and pathologic approach to the classification of IIPs was produced as a collaborative effort of the American Thoracic Society (ATS), European Respiratory Society (ERS), and American College of Chest Physicians (ACCP) (ATS 2002). The novel classification, valid today, divides the IIPs into IPF, and IIPs other than IPF comprising DIP, respiratory bronchiolitis with interstitial lung disease (RBILD), AIP, cryptogenic organizing pneumonia (COP), NSIP, and LIP. IPF represents approximately 60% of all IIPs (Bjoraker 1998, Thomeer 2001, Gross and Hunninghake 2001).

1.2 Epidemiology

There are limited data on the epidemiology of IPF. One of the oldest prevalence estimates frequently referred to in the literature, from 3 to 5/100 000, is based on the Lung Program published in 1972 by The National Heart and Lung Institute in the United States (DHEW 1972). In the Moravian and Silesian populations of the Czech Republic, where the proportion of biopsy-verified cases was 38%, the prevalence tended to increase from 7 to 12/100 000 during the years 1981 to 1990 (Kolek 1994). A large epidemiological study in New Mexico reported the prevalence of IPF for males as 20 and for females as 13/100 000, when the diagnosis was confirmed by clinical or autopsy data (Coultas et al. 1994). Some of the studies focus mainly on population cohorts which have been exposed to fibrogenic dusts. According to those studies the prevalence estimates for IPF have ranged from 3 to 6/100 000 (Iwai et al. 1994, Hubbard et al. 1996, Scott et al. 1990). All of these studies were conducted before the novel IIP classification, and it is likely that these figures include a mixture of several subtypes of IIPs counted as IPF. The incidence estimates are not precise either, suggesting incidences of 11 for males and 7/100 000 per year for females (Coultas et al.1994), and increases with advancing age (Scott et al. 1990, Mannino et al. 1996, Coultas et al. 1994).

IPF has been reported worldwide (Kolek 1994, Coultas et al. 1994, ATS 2000, Zorzetto et al. 2003, Miyake et al. 2005). Some predominance in males is observed (Coultas et al. 1994). The mortality for IPF is estimated to be 1-3.3/100 000 in Japan (Iwai et al. 1994). Some evidence that age-adjusted mortality rates are higher among whites than blacks has been shown (Mannino et al. 1996). This might just reflect inadequate reporting rather than differences in the disease course. Variation in the age-adjusted mortality from IPF is also observed among different geographical areas in the United States; lowest in the midwest and northeast and highest in the west and southeast, and in the United Kingdom; highest in industrialized central areas of England and Wales (Mannino et al.1996, Johnston et al. 1990). This variation is, however, thought to reflect merely the poor exclusion of occupational and environmental fibrogenic exposures, and might be related to inadequate diagnostic criteria (ATS 2000, Fellrath and duBois 2003).

1.3 Etiology

Several environmental factors have been studied as triggers of the disease. Viral infections have been associated with the pathogenesis of IPF. Epstain-Barr virus (EBV) has been studied using various approaches. An association between IPF and serological evidence of active EBV infection is observed (Vergnon et al. 1984). Egan and collagues (1995) reported EBV capsid antigen in epithelial cells of IPF patients by immunofluorescent staining. A high incidence of co-existent or preceding influenza, cytomegalovirus, and hepatitis C infections have been reported among IPF patients, and thereby these have been proposed as etiological agents (ATS 2000).

Numerous environmental exposures have been offered as etiological candidates. Metal and wood dust exposure, especial dust containing steel, brass, lead, and pine wood, have been associated with IPF (Iwai et al. 1994, Scott et al.1990, Hubbard et al. 1996, Baumgartner et al. 2001). In these studies the diagnosis of IPF was not verified with HRCT and/or biopsy specification, which diminishes the reliance of the observed associations.

Cigarette smoking has been identified as a potential risk factor with an odds ratio (OR) of 2.9 (Ryu et al. 2001). A surprising observation has been reported that current smokers with IPF survive longer than non-smokers (King et al. 2001). It is possible that ingredients of tobacco smoke may affect fibroblast function. It is also possible that smokers express other respiratory tract-related illnesses, such as chronic bronchitis and emphysema, causing symptoms such as cough and dyspnoe, and their IPF is diagnosed earlier – paradoxically leading to better survival than seen among non-smokers.

The use of antidepressants and chronic aspiration have also been suggested as etiological factors of IPF (Hubbard et al. 1998, Tobin et al. 1998).

Autoimmune activity may be increased in IPF. Patients with IPF often express systemic symptoms such as fever and arthralgia. Some are positive for rheumatoid factor (19%) and/or antinuclear factor (26%), and antibodies against antisynthetases (Holgate et al. 1983, duBois and Wells 2001, Imbert-Masseau et al. 2003) as markers of autoimmune activity. A recent study showed that an interaction between an endogenous antigen, expressed by type II epithelial cells, and circulating auto-antibodies results in increased TGF-β and tenascin production. That experiment suggests that

the biological activity of autoantibodies may play a role in the pathogenesis of IPF (Wallace and Howie 2001).

1.4 Pathological findings

For a long time it was supposed that IPF is an inflammatory disease with chronic alveolitis preceding fibrotic response (Keogh and Crystal 1982). It has become obvious from multiple sources, however, that inflammation is not the trigger for fibrogenesis of IPF type (Selman et al. 2001, Pardo and Selman 2002). Inflammatory cells and intra-alveolar macrophage accumulation do not belong to the major histological features of IPF. Alveolitis, if it is observed, is mild and seen in late and early diseases, indicating that inflammation does not have to precede fibrosis (Katzenstein and Myers 1998). Adamson and his collagues (1998) showed that severe injury and retarded repair of alveolar epithelium disturbs normal epithelial-fibroblast interactions and is sufficient to promote fibrosis without preceding inflammation. Immunosuppressive and anti-inflammatory treatments used for years in the treatment of IPF do not improve disease outcome (ATS 2000).

The most recent hypothesis is based on an abnormal wound healing process. The proper fibrotic cascade is not known, but various steps are speculated. A schematic illustration of hypothesized pathogenesis is shown in Figure 1. Many environmental, physical, and chemical factors generate reactive oxidative molecules that cause nonspecific damage to cells and the extracellular matrix (ECM) when produced in excess (Kinnula and Crapo 2003). Multiple injuries damage and activate alveolar epithelial cells (AEC) (Pardo and Selman 2002). AECs synthesize molecules, such as tissue factors and plasminogen activator inhibitors (PAI-1 and PAI-2), that are profibrotic and also procoagulant (Imokawa et al. 1999). AECs express multiple cytokines and growth factors. AECs seem to be the main sites of synthesis of platelet-derived growth factor, transforming growth factor (TGF-β), tumor necrosis factor alfa (TNF-α), connective tissue growth factor, and endotelin-1 (ET-1), which induce migration and proliferation of fibroblasts (Antoniades et al. 1990, Giaid et al. 1993, Kapanci et al. 1995, Pan et al. 2001, Pardo and Selman 2002, Chambers et al. 2003, Shi-wen et al. 2004).

Intra-alveolar activation of the coagulation cascade has been documented in pulmonary fibrosis. Plasmin can degrade a number of ECM molecules and activate procollagens (Kotani et al. 1995; Fujimoto et al. 2003). Direct thrombin inhibition decreases lung collagen accumulation in bleomycin-induced pulmonary fibrosis in mice (Howell et al. 2001).

Myofibroblasts within the early fibrotic lesions, as well as intra-alveolar and interstitial activated fibroblasts and myofibroblasts, and type 2 pneumocytes produce ECM proteins (Selman et al. 1986, Pääkkö et al. 2000, Ramos et al. 2001). ECM is composed of fibronectin, elastic fibers, proteoglycans such as tenascin, and especially composed of abundant fibrillar collagens. The activated fibroblasts, myofibroblasts and ECM form fibroblastic foci, pathognomonic of IPF. An imbalance between matrix metalloproteinases (MMPs) that break down matrix proteins and their inhibitors, tissue inhibitors of matrix metalloproteinases (TIMPs), lead to the progressive deposition of ECM (Selman et al. 2000, Ruiz et al. 2003). MMPs may also cause the release of fibrosis-promoting cytokines and growth factors and lead to the initiation and progression of pulmonary fibrosis (Winkler and Fowlkes 2002). Interstitial neovascularization may enhance fibrogenesis, but the role of angiogenic molecules is still unclear (Keane et al. 1997). Factors that have angiogenic activity (interleukin-8 [IL-8], epithelial neutrophil-activating peptide-78) were found at higher levels in tissue specimens of IPF patients (Keane 1997, Keane 2001), and high serum levels of vascular endothelial growth factor (VEGF), IL-8, and ET-1 appear to have a role in the progression of IIP (Simler et al. 2004).

Myofibroblasts produce 1) angiotensinogens such as angiotensin II that provokes AEC apoptosis, and 2) gelatinises A and B, that may increase basement membrane disruption and allow fibroblast migration that in turn might hamper the repair of AECs (Ramos et al. 2001, Ruiz et al. 2003). These mechanisms further provoke unsuccessful re-epithelization.

There is evidence that, besides alveolar epithelium, bronchiolo-alveolar junctions also represent a relevant and specific target of injury in IPF (Chilosi et al. 2002). Epithelial cells of IPF patients expressing ΔN-p63, a member of the p53 tumor suppressor gene family, were observed at sites of abnormal proliferation at the bronchiolo-alveolar junctions, characterized by epithelial hyperplasia, squamous metaplasia, and abnormal p53 nuclear accumulations – features that were not observed in normal lung, or in COP, NSIP, DIP, or AIP lung (Chilosi et al. 2002). The products of the p53 gene family may also reflect the malignant transformation observed in IPF (Turner-Warwick et al. 1980, Chilosi et al. 2002,).

Based on present evidence, it is proposed that the earliest morphologic change associated with progressive fibrosis is the presence of fibroblastic foci (Kuhn and McDonald 1991, Kazenstein and Myers 1998, King et al. 2001). Fibroblasts within these foci keep modifying their phenotypes,

parallel to that seen in skin wounds, from a migratory through a proliferative phase to a profibrotic phenotype, producing ECM components (Selman et al. 2001). Whether fibroblasts from patients with IPF are genetically susceptible to abnormal response after lung injury is not clear.
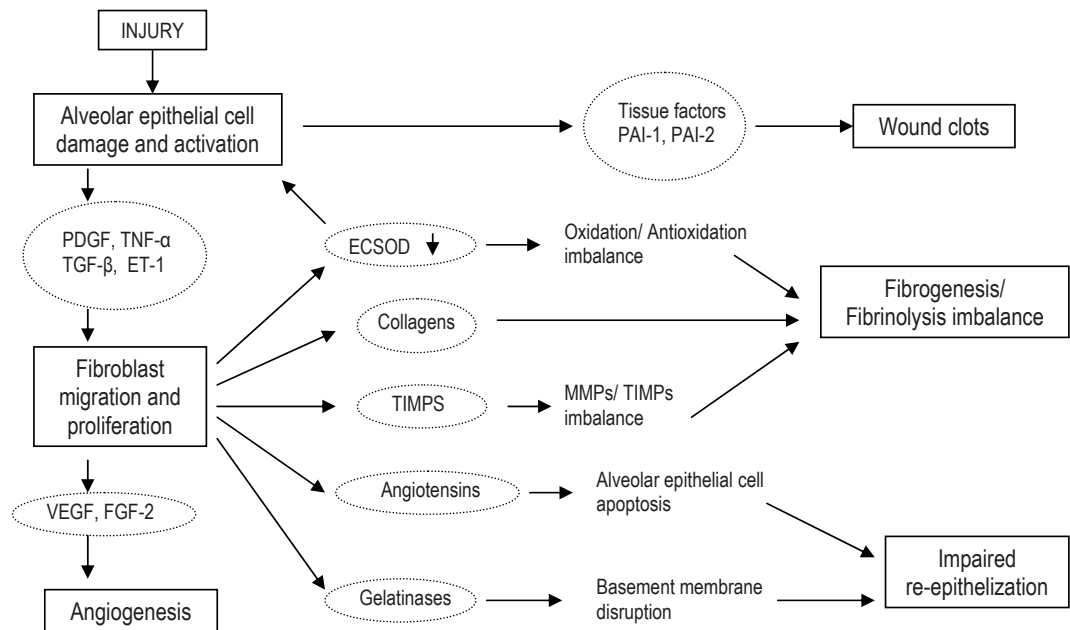
Figure 1. A schematic illustration of possible steps from injury to usual interstitial pneumonia (UIP) (modified from Selman et al. 2001). Multiple injuries damage alveolar epithelial cells (AECs). Activated AECs secrete antifibrinolytic tissue factors plasminogen activator inhibitors (PAI) -1 and -2, which provoke wound clot formation. Activated AECs also secrete several growth factors and cytokines that enhance fibroblast migration and proliferation, and differentiation into myofibroblasts. Proliferated fibroblasts and myofibroblasts secrete extracellular matrix (ECM) proteins, mainly collagens, and tissue inhibitors of metalloproteinases (TIMPs), which in turn favors in imbalance fibrinogenetic deposition of ECM over fibrinolysis. Angiogenic factors, vascular endothelial growth factor (VEGF) and fibroblast growth factor-2 (FGF) enhance neovascularization. Myofibroblasts and AECs secrete gelatinases that damage basement membrane, and angiotensinogens that induce AEC death. Both of these result in impaired re-epithelization.

1.5 Diagnosis of IPF

Patients with IPF typically suffer from shortness of breath and cough for more than six months before diagnosis. General symptoms, such as fever, arthralgia and myalgia, may also occur. Digital clubbing is seen in around 50% of the patients. Auscultation of the lungs reveals fine bibasilar inspiratory cracles (Velcro rales). The average age at the time of diagnosis is 60-66 years (Turner-Warwick et al. 1980, Johnston et al. 1997, Lee et al. 2005). It is important to exclude other causes of interstitial pulmonary diseases. IPF is also seen as pulmonary involvement in various systemic diseases, such as rheumatoid arthritis (RA) and systemic sclerosis (Wells et al. 1993). IPF type fibrosis is considered to be the most common pulmonary manifestation of RA, present in approximately 20% of outpatients (Dawson et al. 2001), although a pathological approach suggests that cases previously classified as IPF are likely to show a pattern of NSIP rather than UIP, particularly in relation to systemic sclerosis (Nicholson et al. 2002). The median survival from diagnosis is 2.5-4.5 years (Schwarz et al. 1994, Nicholson et al. 2000, King et al. 2001, Collard et al. 2003).

There are no specific laboratory findings for IPF. Bronchoalveolar lavage (BAL) fluid cell differential may show neutrophilia and the total cell count may be increased, as a sign of immunoactivation (Kinnula and Tukiainen 2004). Most importantly, however, BAL excludes other conditions, such as exposure to asbestos and infectious agents. Sedimentation rate and CRP might be elevated. Rheumatoid factors and antinuclear antibodies are seen in up to 30% of IPF patients (Holgate et al. 1983, Fellrath and duBois 2003). Pulmonary function tests show restriction with reduced vital capacity, and/or lowered diffusing capacity for carbon monoxide, and/or reduced arterial oxygen pressure (King et al. 2001, Kinnula and Tukiainen 2004).

Conventional chest radiography typically shows bibasilar nodular or reticular infiltrations. HRCT appears to be a valuable tool in diagnosing IPF and distinguishing it from other IIPs (Tung et al. 1993, Johkoh et al. 1999). The HRCT technique with slices of 1-2 mm in thickness and an algorithm that maximizes spatial resolution gives detailed images of lung parenchyma. Typical features are peripheral reticular opacities, most marked at lower zones, honeycombing is common, as well as traction bronchiectasis. Mild ground glass attenuation might also exist. Architectural distortion of the parenchyma is often evident. The characteristic radiological features are located most typically at lower lobes and peripherally (ATS 2000). HRCT, evaluated by an experienced

observer, appeared to distinguish histologically verified UIP in 71% (25/35) of 129 patients with IIP (Johkoh et al. 1999).

Figure 2. Typical high resolution computed tomography (HRCT) image of idiopathic pulmonary fibrosis (IPF) with bibasal honeycombing and traction bronchiectasis.

The gold standard in diagnosing IPF is a surgical lung biopsy showing a pattern of UIP. International guidelines recommend taking the biopsy if there are no contraindications for surgery. It is especially important if a patient presents clinical, physiological, or radiological features that are not typical for IPF (Costabel and King 2001, ATS 2002). Surgical lung biopsy, taken with video-assisted thoracoscopy (VATS) or open thoracotomy, provides tissue samples to distinguish UIP. VATS has been shown to be a safe procedure (Bensard et. al. 1993, Sihvo and Salo 2003). Transbroncial biopsy does not provide enough tissue for IPF diagnosis (Akira et al. 1993). Typical for the histological appearance in UIP is that in areas of normal lung, scattered fibroblastic foci, architectural destruction, and honeycombing fibrosis appear randomly (Katzenstein and Myers 1998, Travis et al. 2000, Flaherty et al. 2001). That is why it is recommended to take multiple tissue samples from macroscopically normal appearing lung areas, as well as from the affected lung (Flaherty et al. 2001). The interstitial fibrotic foci composed of active proliferating fibroblasts and myofibroblasts are thought to be the key features of the active fibrotic process (Selman et al 2001). The fibrotic changes show temporal heterogeneity from scattered fibroblastic foci to dense acellular collagen. Interstitial inflammation is mild or moderate. The histological changes appear most prominently in the peripheral subpleural parenchyma. Areas of normal (or close to normal) lung should be seen in tissue samples in order to exclude other interstitial diseases (ATS 2002).
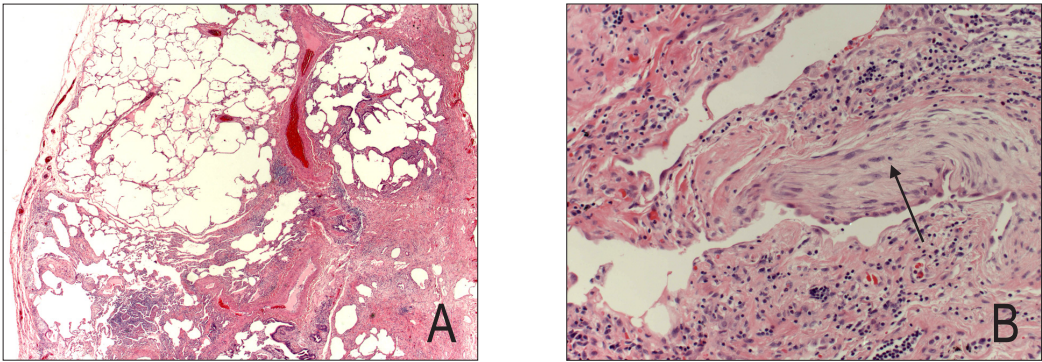
Figure 3. Typical features of usual interstitial pneumonia (UIP) in surgical biopsy. Honeycombing, temporal heterogeneity and architectural destruction (A), and a fibroblastic focus (arrow) (B) can be identified.

IPF is very often difficult to distinguish from the fibrotic form of NSIP. Biopsy specimens from multiple lobes may show patterns of both UIP and NSIP, and these NSIP-like areas could be present in the majority of UIP cases (Flaherty et al. 2001, Kazensten et al. 2002). In such cases the pathological diagnosis remains UIP and the clinical outcome is similar to that of IPF (Flaherty et al. 2001, Kazensten et al. 2002).

1.6 Treatment

For decades it has been the tradition to treat IPF patients with corticosteroids and immunosuppressive agents. There is, however, no supportive evidence that immunosuppressive agents improve survival or the quality of life (Maple et al. 1996, Mason et al. 1999, Selman et al. 2004). Because of the lack of a more effective therapy, according to the present international guidelines of ATS/ERS, combined therapy of corticosteroid and azathioprime, or cyclophosphamide should be used for a minumum of 6 months if no intolerable side effects occur. The therapy should be carried out only in patients who show objective evidence of continued improvement or stabilization of the condition (ATS 2000). During therapy the patients should be monitored carefully because of the possible and probable adverse effects (ATS 2000).

Some patients can get symptomatic relief from antitussive agents and opioids (ATS 2000). For patients who deteriorate despite optimal medical treatment lung transplantation can be considered (ATS 2000).

Despite the use of aggressive treatments, IPF is in the majority of cases a progressive, irreversible and fatal disease (Selman et al. 2001, Gross and Hunninghake 2001, duBois and Wells 2001). It is

unlikely that any of the present treatments will improve the prognosis of IPF, and therapies based on alternative approaches are needed. The possible targets are to inhibit fibroblast proliferation and ECM accumulation, to induce apoptosis of myofibroblasts, and to prevent the epithelial damage or to provoke its repair (Selman et al. 2001, Fellrath and duBois 2003). There are plenty of ongoing clinical trials based on these alternative mechanisms: pirfenidone inhibits TGF-β-stimulated collagen synthesis and decreases the ECM (Raghu et al. 1999), interferon-γ inhibits fibrogenesis (Ziesche et al. 1999), N-acetylcysteine is a glutathione precursor that prevents epithelial cell injury mediated by oxygen radicals (Behr et al. 1997), captopril inhibits the angiotensin-converting enzyme that induces AEC apoptosis (Uhal et al. 1998), and bosentan is a endothelin-1- inhibitor that is one of the cytokines associated with fibroblast proliferation (Fellrath and duBois 2003, Selman et al. 2004).

2. GENE MAPPING OF COMPLEX DISEASES

2.1. General

Using gene mapping approaches in the identification genes causing simple Mendelian diseases has become a straightforward process. When a defect in a gene is detected only among patients and never among healthy controls the connection between the gene and the disease is easy to establish. Complex diseases do not follow classic Mendelian inheritance, but involve multiple genetic and environmental determinants. Susceptibility alleles are found both among patients and controls (low penetrance), but the frequencies are different. The disease-predisposing allele is likely to result in disease together with favourable environmental and/or other genetic factors. With different environmental and/or genetic factors the allele is carried by unaffected as well. The incomplete correspondence between a single allele and the disease makes both the mapping and the identification of the disease-causing genes difficult. A mutation in a gene can result in causing a disease in various ways. Only one amino acid change in the product of the gene can alter the function. There is increasing evidence that identified polymorphisms not only in coding, but also in non-coding sequences, can change expression levels and splicing (Pagani and Baralle 2004). The expression levels can also be contolled by several trans-acting factors, such as transcription factors that can be located far from the gene (Pastinen and Hudson 2004). During the recent years, genes and molecular genetic mechanisms responsible for complex diseases, such as in asthma (Laitinen et al. 2004), hyperlipidemia (Pajukanta et al. 2004), and psoriasis (Asumalahti et al. 2000), have successfully been identified.

The methods of locating susceptibility genes are based mainly on either linkage or association approaches. The goal of a linkage study is to identify chromosomal loci that are present in affected family members more often than is explained by Mendelian segregation; the alleles are used as tools for assessing the linkage properties of these loci. In association studies particular alleles are the subject of study, and might even be the cause of the phenotype. Association is a population property and can be assessed in unrelated individuals (Ewens and Spielman 2001). These two approaches can completement each other: linkage studies can bring up hypotheses that can be verified in larger population- or case-control association studies.

There are several aspects concerning the difficulties in mapping complex diseases. One genotype can lead to different phenotypes or different genotypes can lead to identical clinical outcomes. All

individuals with a predisposing allele may not manifest the disease, in the case of incomplete penetrance. Individuals without the predisposing allele may also manifest the disease, termed as phenocopies (Lander and Schork 1994). Polygenic inheritance makes mapping more difficult, because some diseases require inherited susceptibility alleles in multiple genes, as a single locus is alone incapable of causing the disease.

The number of disease-causing alleles in every gene varies plenty, as do the allele frequencies. If a disease-causing allele is very frequent in a population, it may be difficult to distinguish its` association to the disease. Those genes with only a few harmful alleles can be identified more easily. According to the Common Disease/Common Variant (CD/CV) hypothesis common diseases are due to the alleles with relatively high frequencies (Reich and Lander 2001). Allelic diversity is also thought to be similar for common and rare diseases (Hartl and Campbell 1982). In real populations, however, there tends to be multiple disease-predisposing alleles for rare diseases (Reich and Lander 2001).

The distances between observed loci are measured in genetic units, centimorgans (cM). One cM corresponds to the distance within which one recombination is supposed to occur in every 100 meioses and corresponds approximately to a physical length of one megabase (Mb) of DNA (Thompson et al. 1991). When alleles on the same chromosome tend to occur together more often than expected under random segregation they are said to be in linkage disequilibrium (LD) (Terwilliger and Ott 1994).

2.2 Linkage analysis

In order to find linkage beween IPF and disease-predisposing loci it is crucial to define the proper phenotype of the affected family members and to identify multiplex families with linkage information. In some diseases, such as IPF, with the late onset, and short survival after diagnosis, the recruitment period is extremely short. Family members may have died before the disease was manifested, or their symptoms have been interpreted as other than IPF before the improved diagnostic techniques and the present classification were available – both examples result in ignoring the true familiality, and excluding informative families.

By means of positional cloning, disease-predisposing genes are identified purely due to their chromosomal location, without concern for their function. Botsein et al. (1980) recognized naturally

occurring DNA sequence variations, simple tandem repeats (also called microsatellites). These occur in mammalian genomes at fairly regular intervals, and can be used as genetic markers in gene mapping strategies (Weber and May 1989).

Linkage analysis tries to locate chromosomal regions (containing a disease-predisposing gene) by detecting marker loci and disease phenotype that segregate together. Two-point linkage analysis shows whether a particular allele and the disease have a tendency to be co-inherited from parents to offspring, which could indicate a short distance between the allele and disease-causing gene (Terwilliger and Ott 1994). Multipoint analysis gathers information from multiple markers simultaneously to estimate whether an allele at a given point is shared identical-by-descent (IBD) (Kruglyak and Lander 1995, Ott 1996), and thus tests for linkage to an extended chromosomal region rather than to a single point.

Linkage analysis can be performed by either a parametric or a non-parametric approach. Parametric linkage analyses estimate the recombination fraction (equivalent to genetic distance when small) between a disease and marker loci when the mode of inheritance is known (Ott 1996). In the case the mode of inheritance is known, the statistical power is generally higher than with non-parametric linkage analysis (NPL). For NPL analysis the mode of inheritance does not need to be determined. The simplest NPL analysis is estimating whether affected siblings share an IBD allele more often than expected according to Mendelian transmission (Terwilliger and Ott 1994). The method is nowadays mostly used as an extended version, an extended relative pair analysis, and it takes into account all of the affected family members instead of the sibpairs. NPL method compares the alleles shared by affected individuals, and the unaffected individuals are ignored. NPL uses all the genotype data collected from all available pedigree members, and it computes the probable IBD statuses also for missing genotypes (Kruglyak et al. 1996).

The complex inheritance of diseases such as IPF offer challenges. Some individuals who inherit a predisposing allele do not manifest the disease (incomplete penetrance), while others who have not inherited the allele may anyhow get the disease due to environmental or random causes (phenocopy). Genetic heterogeneity, where a chromosomal region may segregate with the disease in some of the families, but not in others, may hamper gene mapping. A disease-causing allele is hard to map if it occurs at high frequency in the population. One gene may interfere with the expression of another gene located at a different locus (epistasis), and multiple, often seemingly unrelated, physical effects may be caused by a single altered gene or a pair of altered genes (pleiotropy)

(Lander and Schork 1994, Tabor et al. 2002, Risch 2000). It is therefore crucial to gather all the possible available marker and genotype data from the pedigrees, which is computationally demanding. A widely used computer package is GENEHUNTER (Kruglyak et al. 1996). The package can simultaneously analyse the data with parametric and non-parametric methods, conduct non-parametric two-point and multipoint analyses, determine the information content for each marker, re-construct haplotypes, and analyse affected sibpairs (Kryglyak et al. 1996, Ewens and Spielman 2001).

The statistical measurement in parametric analyses is the logarithm ($\log_{10}$) of odds (LOD) score. It is a ratio of the likelihoods of 1) linkage between a locus and a marker at a certain recombination fraction (θ) compared to 2) no linkage ($\theta = 0.5$) (Terwilliger and Ott 1994). The recombination fraction (θ) measures the extent of linkage between two loci, i.e. the probability of a recombination occurring between two loci. The most likely θ gives the highest LOD score (Pawlowitzki et al. 1997). Non-parametric linkage can be analysed by two additional tests: S_{pairs} and S_{all}, which are based on IBD statuses. S_{pairs} consideres all affected sibpairs, while S_{all} consideres all family members, with the assumption that if many affected relatives share the same allele IBD, linkage to the disease-causing allele is more likely than in the case of only affected siblings sharing it. These calculations result in NPL scores NPL_{pairs} and NPL_{all}, very often presented as Z_{pairs} and Z_{all} (Kruglyak 1996).

Lander and Kruglyak (1995) suggested significance criteria for LOD scores in complex diseases. The criteria are valid only in perfect data sets with complete pedigrees, and with 100% successful genotypes – the criteria that correlate poorly with real data sets. A LOD score of 3.3 or higher is considered as significant evidence for linkage, and a LOD score >1.9 for suggestive linkage. The performance of NPL_{all} is roughly comparable to that of LOD score analysis under the correct inheritance model. NPL_{all} appears to be a non-parametric pedigree-analysis method that loses relatively little power when compared with the best parametric method (Lander and Kruglyak 1995). Altmuller et al. (2001) reviewed 101 genome-wide scans in complex human diseases. Of these, 67 showed no significant linkage according to the categories proposed by Lander and Kruglyak (1995), suggesting that significant linkage is hard to find in genome-wide scans in complex diseases. Some concern has been raised whether the thresholds should be lower among genome scans in complex diseases (Altmuller et al. 2001, Wiltshire et al. 2002, Sawcer et al. 1997, Göring et al. 2001). Therefore, a method of choice to estimate the power to detect linkage in complex diseases is simulation. Randomly generated genotypes, real marker densities, and true

pedigree structures are simulated N times. When the considered NPL score exceeds the NPL score that is observed once per genome scan at random it is proposed to show suggestive linkage, and significant linkage is obtained when the NPL score exceeds the one observed once per every 20 genome scans at random (Lander and Kruglyak 1995, Laitinen et al. 2001). Wiltshire et al. (2002) simulated an experimental data set with a 10 cM marker map and 15% missing genotypes, and found that an independent region showing evidence for linkage (IRL) with a LOD score of 1.51-1.55 is expected to occur only once by chance. They proposed this locus-counting method (IRLs) as an additional method for evaluating the results of genome scans on complex diseases.

2.3 Association analysis

Association studies are based on consideration of whether a genetic polymorphism is overpresented (positive association) or underpresented (negative association) in the studied phenotype compared to a control population (Lander and Schork 1994, Baur and Knapp 1997). Associations can be performed for any DNA polymorphism, but they are more meaningful when applied to functionally significant variations in genes having a clear biological relation to the disease. Multimarker haplotypes are more informative than single alleles. In the case where true association has been detected, it can be due to 1) an allele actually being the disease-causing allele, 2) an allele being in linkage disequilibrium with the disease-causing allele, or 3) an artifact of population admixture: any trait present frequently in an ethnic group shows positive correlation with any high frequency allele (false positive finding) (Lander and Schork 1994). Because of the later the selection of a control group is crucial. While studies are performed in relatively homogenous populations, within an ethnic group, the control population should represent the same ethnic origin (Lander and Schork 1994, Risch 2001). Population stratification can be partly avoided by using non-affected family-based controls (Lander and Schork 1994, Risch 2001). The affected and non-affected family members alike carry background hereditability and experience similar environmental factors. Therefore, an observed association to the disease is most likely to be true positive. Sampling families of varying ethnicity is advantageous to enhance evidence of causality as well as to identify genetic and/or environmental modifying factors (Risch 2000). Consistent replications of the association in assorted studies with different populations strengthen the evidence of causality.

Improved techniques for genotyping polymorphisms and extended computational possibilities have made it possible to look for differences in genetic variants throughout the entire genome between affected individuals and controls. Whole-genome linkage disequilibrium mapping is based on the

fact that some of the polymorphisms and disease-predisposing alleles are located nearby each other (Kruglyak 1999). LD varies depending on the genomic region and population history (Wright et. al. 1999). The average LD distance rarely extends beyond around 3 kb in the general population, however, it can extend further in young founder populations (Kruglyak 1999, Wright et. al, 1999). The markers of choice are biallelic single nucleotide polymorphisms (SNPs) because they are extremely numerous (approximately 3 million SNPs, one in every 1 kb), are expected to exist in the human genome, they have low mutation rates, and analysis is quite easy to automate. Current estimates of the number of SNPs needed for whole-genome association studies vary considerably, from 100 000 to 500 000 (Kruglyak 1999, Palmer and Cardon 2005). Whole-genome LD mapping is thus far financially available for very few researchers, but with improving technology it will hopefully be easier to access. The interest in average LD lengths has been, however, replaced by awareness of special LD patterns (Cardon and Abecasis 2003). It seems that the genome is composed of a series of high LD regions (blocks) flanked by very low LD regions (recombination hotspots) (Cardon and Abecasis 2003). Preliminary results suggest that three to five haplotypes can account for 90% of all haplotypes in the human population, and each haplotype block is on average 22-44 kb long; shorter in populations originated from Africa than in the other examined samples (Daly et al. 2001, Gabriel et al. 2002). These blocks could be detected with few markers and less money. The ongoing haplotype mapping project (HapMap) aims to characterize patterns of LD in the human genome (Gabriel et al. 2002, www.hapmap.org).

Association studies with candidate genes have been widely used in complex diseases, but the results have been conflicting (Risch 2000, Tabor et al. 2002). Considering the location and function of each polymorphism, along with more information about LD patterns and potential haplotypes, and with the knowledge of the sequences and functions of the candidate genes, selecting precise candidate genes and variants to perform association studies seems more promising (Risch 2000, Tabor et al. 2002).

2.4 Characteristics of the Finnish population

The majority of the genes of the Finnish population are thought to originate from a small number of founders who immigrated from the south around 2000 years ago (Lahermo et al. 1999, Peltonen et al. 1999). The population grew very slowly, at the beginning of the 18th century there were approximately 250 000 Finns. After that the population expanded rapidly, today being 5,2 million inhabitants.

Until the reign of Gustavus of Vasa (1523-1560), most of the Finns lived on the southern and western coastlines, in the early settlement area. During that reign people of the early settlement, mostly farmers from southern Savo, moved to eastern, central, and northern parts of Finland, to the late settlement area. The immigrated groups lived for centuries in subisolates. Until the Second World War many of these northeastern settlements consisted of descendants of 40-60 founding families (Varilo et al. 2000). During the immigration in the 16th century the national system of parish and tax records was established, which still works as a magnificent tool to solve pedigree structures and to trace common ancestors (Peltonen et al. 2000).

During immigration the Finns have gone through multiple genetic bottlenecks. Because of the small number of founders the population of Finland has experienced genetic drift that favored retention of some genes and removed others. Genes with allele frequencies of a few per cent showed more than 10-fold differences among seven communities in a study by Nevanlinna (1972). For common alleles, the overall frequencies in Finland are similar to other European countries, but very rare alleles have been almost totally lost from most subisolates (Kere 2001).

Since Norio (1966) brought up the genetic background of congenital nephritic syndrome, at least 36 monogenic disorders have been successfully identified among Finns (Norio 2003). Complex traits are not inherited according to Mendelian transmission, and their phenotypes are more diverse. Many complex diseases tend to manifest at older ages, and allow the disease-causing allele to be transmitted, leading to the existence of even dominant disease alleles in pedigrees. When a disease can be caused by each of several rare alleles, poor genetic diversity in isolated population is advantageous, as has been shown in nonpolypotic colon cancer, combined hyperlipidemia, and long QT syndrome (Peltomäki 1993, Vuorio et al. 1997, Piippo et al. 2000).

After a founder mutation has been introduced into a population, in young populations fewer crossovers have influenced the flanking chromosomal regions, and chromosomes in individuals include longer regions of common ancestral chromosomes. This has been regarded as advantageous in using population isolates for mapping. The levels of LD between common markers close to each other, however, do not show remarkable differences between Finland and other countries. The LD patterns are preserved only over short (<1 cM) distances when comparing alleles of linked loci within randomly selected individuals from the sub-population of Kuusamo, and across Finland (Varilo et al. 2000). On the other hand, in isolated founder populations affected individuals are

obviously related to each other and their predisposing alleles are likely to be IBD instead of identical by state (IBS). Because the affected individuals most likely have inherited a rare (disease-predisposing) allele from a common ancestor, there may be long, conserved haplotypes adjacent to rare disease alleles, and associations with the surrounding markers are expected, as well as higher levels of LD. The extent of the LD around a disease gene reflects the age of the mutation, and the extremes of LD intervals vary in Finland from 11 cM to 13 cM (Höglund et al. 1995, Varilo et al. 1996, Peltonen et al. 1999).

Because of the genetic bottlenecks, random drift, and the high levels of LD over long chromosomal regions, the genetic and haplotype diversity is far lower in population isolates such as Finland, which provides advantageous conditions for the identification of disease alleles.

In addition to genetic homogeneity, cultural homogeneity favors performing genetic studies in Finland. Diagnostic criteria across Finland are extremely uniform. Patients are entitled to compensation for medication through a nationwide institution, The Social Insurance Institution of Finland. The desicion for compensation is based on reports of diagnostic criteria that have to be fulfilled according to the ICD-10 criteria. The diagnoses are registered in hospital discharge records. The standardized quality of health care and health care registers make diagnoses relatively comparable and reliable.

3. GENETICS OF IPF

Rare, late onset diseases are a challege for linkage studies. IPF manifests often after the age of 60 and leads to death within a couple of years. Thus genotyping of the parents of the probands is very often impossible, and even the availability of DNA from all affected siblings is difficult. To find families with enough linkage information to identify meaningful linkage has been an obstacle to performing genome screens. IPF is the most lethal form of IIPs and still, under the new classification era, many IIPs are misdiagnosed as IPF. Most of the studies concerning the genetics of IPF are case-controlled, association-based studies on sporadic cases, using microsatellite or SNP analysis of candidate genes.

The only published linkage study attempting to locate a chromosomal region for IPF was performed on chromosome 14 (Loyd 2003). An interest in chromosome 14 was aroused due to two reports which showed an increase in the frequency of Z and S alleles of the α-1-antitrypsin gene among IPF patients (Geddes 1977), and an association between IPF and immunoglobulin Gm allotypes (Musk et al.1986), both genes encoded on chromosome 14. Therefore, a collaborative study was performed with 70 individuals from 11 families using 14 markers distributed across chromosome 14 to establish or exclude linkage. All two-point and multipoint LOD scores were either negative or near zero suggesting that with this data set linkage to chromosome 14 can be excluded (Loyd 2003).

3.1 IPF in families

The most convincing evidence for genetic predisposition in the pathogenesis of IPF is the existence of multiplex families, i.e. families with more than one affected member. Peabody and Hayes (1950) presented an identical twin pair, whose symptoms started within 16 months of each other and were diagnosed as having IPF. MacMillan used the term familial pulmonary fibrosis already in 1951 to describe a mother and her adult daughter who both died from fibrotising lung disease. In 1965, Bonnani and collagues described a family with eight affected family members, with ages at death ranging from 38-52 years. Since then many anecdotal cases of familial idiopathic pulmonary (FPF) fibrosis have been reported (Javaheri et al. 1980, Bitterman et al. 1986, Uchiyama et al. 1997). Twenty-five families with two or more affected family members were identified in the United Kingdom through a questionnaire mailed to respiratory physicians (Marshall et al. 2000). Altogether 67 individuals were diagnosed with IPF; the diagnoses were confirmed in 32% of the cases with a biopsy, although this occurred prior to the recent diagnostic criteria. They calculated a

prevalence of 1.34/million for FPF. In order to define clinical features of FPF, Lee and collagues (2005) described 15 families with at least two affected family members. They compared the clinical features and survival with nonfamilial cases, and found no differences. Steele and collagues (2005) recently reported 111 families with two or more cases of IIP, IPF among them.

Although the model of inheritance is not clear, the autosomal dominant model with varying penetrance has been proposed (Marshall et al. 1997, Marney et al. 2001). Most of the described pedigrees demonstrate vertical transmission from a parent to an offspring (Bonnani et al. 1965, Bitterman et al. 1986, Uchiyama et al. 1997, Marshall et al. 2000, Thomas et al. 2002, Lee et al. 2005), confirming this model. Sex-linked inheritance was excluded by reports of father-son transmissions (Koch 1965, Adelman 1966). According to recently published reports, patients within the same family have been affected by distinct subtypes of IIP, mostly IPF and NSIP. This may reflect that the same genetic mechanisms may result in diverse interstitial responses (Thomas 2002, Steele 2005)

3.2 Candidate genes

Genes that have been proposed as candidate genes for IPF are categorized based on their possible functional role in the pathogenesis of IPF. Polymorphisms in the genes have often been amino acid changes that could alter the protein function. The genetic studies of IPF have mostly been association studies with case-control designs. The results have been predominantly either conflicting or proposed associations have not been confirmed. On the other hand, candidate genes cannot be ruled out based on negative results when only some of the polymorphisms in a gene have been studied, or when studies have been performed with one distinct population or with a small number of individuals.

3.2.1. Immune response genes

The earliest studies focused on a possible association between human leucocyte antigen (HLA) genes of the major histocompatibility complex (MHC) and IPF. Some association was detected between an increase of the serologically-detected HLA-DR2 antigen and IPF (Libby et al. 1983), but the descriptions of the radiological findings suggest the patients also represented other forms of IIPs. Fulmer and collagues (1978) detected no differences in 35 serologically detected HLA A and B antigens between IPF patients and controls. Only very recently, MHC polymorphisms were

evaluated with modern gene technology. Falfan-Valencia and collagues (2005) studied the HLA-B, -DRB1, and -DQB1 loci in a cohort of 75 IPF patients and 95 controls. Three haplotypes were significantly increased in IPF: 1) HLA-B*15-DRB1*0101-DQB1*0501 (OR=10.72, CI=1.43-459.6); 2) HLA-B*52-DRB1*1402-DQB1*0301 (OR=4.42, CI=1.21-24.1); and 3) HLA-B*35-DRB1*0407-DQB1*0302 (OR=4.73, CI=1.53-19.5). BAL cells collected from patients with these haplotypes showed reduced epithelial growth rates, and caused epithelial cell apoptosis. Their findings suggest that some MHC polymorphisms cause susceptibility to IPF.

TNF-α and Interleukins 1 and 6 (IL-1, IL-6) are proinflammatory and fibrogenic cytokines (Pan et al. 1996, Chambers et al. 2003). Whyte and collagues (2000) compared 88 IPF patients from England and 61 patients from Italy to 191 controls to detect SNP polymorphisms in the IL-1 receptor antagonist (IL1RN), and Tumor Necrosis Factor-α (TNF-α) genes. The authors suggest that carrying the IL1RN allele +2018T and the non-coding TNF-α -308A allele is associated with increased risk of developing IPF. The association of the TNF-α -308A allele with IPF was confirmed by Riha et al. (2004) in their study of 22 IPF patients, while Pantelidis and collagues (2001) were unable to confirm the association. They screened altogether 12 SNPs in the TNF-α, lymphotoxin-α, TNF-receptor 2, and interleukin-6 (IL6) genes in 74 IPF patients from England and 201 population-based controls. However, the authors suggested that the cocarriage of intronic SNPs, allele 4GG in IL6, and allele 1690C in TNF-R2 were increased among IPF patients compared to the controls. Further, the GG genotype in intron 4 of IL6 was found to be independently associated with lower diffusing capacity for carbon monoxide levels. No association was observed between IPF and the studied polymorphisms by Hutyrova and collagues (2002), who explored the polymorphisms in the IL1α, IL1β, and IL1RN genes that are clustered on chromosome 2q13-21. The study population consisted of 54 IPF patients, 95 sarcoidosis patients, and 199 controls from the Czech Republic, all of west Slavonic ancestry.

Cytokine mediators that provoke T cell-response towards Th2 predominance act as functional candidate genes for IPF. Associations between the SNPs C1188 in the IL12 gene, on chromosome 5q31-33, that has been proven to cause a functional change, and G5644 in the IFN-γ gene, on chromosome 12q14, with susceptibility to IPF were studied. Seventy-three patients and 157 population-based controls carried similar allele frquencies (Latsi et al. 2003).

Intra-alveolar activation of the coagulation system due to reduced fibrinolytic fuction has been suggested to play a role in the pathogenesis of IPF (Fujimoto et al. 2003). The frequencies of

polymorphism 4G/5G in the plasminogen activator inhibitor-1 promoter, encoded on chromosome 7q21-22, did not differ between 62 patients with IPF and controls (Kim et al. 2003).

The Complement Receptor 1 (CR1) gene (known also as the complement component 3b/4b receptor or C3-binding protein) is located on chromosome 1q32. CR1 is expressed in different cells, including phagocytes, lymphocytes, and dendritic cells (Fearon and Wong 1983), but mainly on the surface of erythrocytes (Siegel et al. 1981). The primary function of CR1 is mediating the adherence of target molecule coated with the complement components C3b and C4b, and their transport to the reticuloendothelial system, the spleen and liver, to be destroyed (Cornacoff et al. 1983, Davies et al. 1992). Immune complexes that escape the reticuloendothelial system can in turn increase tissue damage associated with many immune diseases.

A total of 11 nonsynonymous polymorphisms in the coding sequence of CR1 have been reported (http://www.ncbi.nlm.nih.gov/SNP). The substitution C5507G in exon 33 leading to an amino acid change from proline to arginine (Pro1827Arg) has been reported (Zorzetto et al. 2002). The polymorphism creates a potential cleavage site for trypsin-like proteases that allow its accelerated proteolysis and, in return, decrease the activity of the complement (Herrera et al. 1998). Pro1827Arg has been associated previously with two parenchymal pulmonary diseases, sarcoidosis and IPF, among Italian patients (Zorzetto et al. 2002, Zorzetto et al. 2003). Arg1827 homozygosity was significantly more common in patients with sarcoidosis (OR=3.1, 95%CI=.5-6.7) and IPF (OR=6.2, 95%CI=2.2-18.4) than in control subjects or COPD patients.

3.2.2 Genes in biochemical defence

Lungs are directly exposed to high levels of environmental oxidants, such as pollutants, ozone, fibers, cigarette smoke, and radiation, as well as hyperoxia (Kinnula and Crapo 2003). In these circumstances several endogenous enzymes and reaction pathways can be activated to generate reactive oxygen species (ROS). Reactive nitrogen species (RNS) can be generated in parallel, when also endogenous nitric oxide (NO) is present. Constitutive and inducible forms of NO are widely expressed in lungs (Kinnula et al. 1995, Kinnula et al. 2005). Although ROS and RNS are essential for normal physiological functions, such as taking a part in phagocytosis (Ottonello et al. 1995), mediating smooth muscle relaxation (Dupuy et al. 1992), and modulating pathways in cell signalling (Thannickal and Fanburg 2000), they are extremely potent in causing nonspecific damage to cells and ECM (Kinnula and Crapo 2003).

Antioxidants defend against ROS and RNS by either enzymatic or nonenzymatic ways (Bowler and Crapo 2002). Extracellular superoxide dismutase (ECSOD) is an antioxidative enzyme belonging to a family of superoxide dismutases (SOD). It is expressed in various tissues, highly in lung, and is abundant in airway epithelial and vascular endothelium, and especially in the extracellular matrix (Marklund 1984, Folz and Crapo 1994, Oury et al. 1996). ECSOD is the major enzyme capable in dismutating superoxide radicals into hydrogen peroxide and oxygen in the ECM (Fattman et al. 2003, Kinnula and Crapo 2003).

Imbalance between oxidants and antioxidants in the lung provokes oxidative stress. The importance of ESCOD against oxidative injury has been confirmed both in transgenic (Bowler et al. 2002, Folz et al. 1999, Ghio et al. 2002) and knockout mice (Carlsson et al. 1995, Fattman et al. 2003), thus suggesting that ESCOD may be one of the major enzymes against oxidant-mediated parenchymal lung injury, such as that seen in IPF.

Surfactant protein C (SP-C) is a membrane protein expressed by epithelial cells in the lung. Mature SP-C maintains a biophysical surface activity and may also protect the surfactant film layer. The SP-C gene is located on chromosome 8p21. Multiple heterozygous alleles in the SP-C gene have been reported in association with interstitial lung diseases (Nogee et al. 2001, Nogee et al. 2002). One of these is a single nucleotide change T128A resulting in an amino acid change from leucine to glutamine (Leu188Glu). There is some evidence that the Leu188Glu mutation may cause misfolding and trapping of proSP-C preventing delivery to distal secretory components, and this could eventually lead to type II cellular damage (Thomas et al. 2002). The Leu188Glu mutation was studied in a large kindred of 97 individuals; 14 of them were affected, 11 adults and 3 children. All the children were diagnosed as NSIP verified with a biopsy, and all the adults were diagnosed as UIP, 6/11 had biopsy verification. The Leu188Glu mutation was detected among all the affected individuals, but in only two non-affected family members (Thomas et al. 2002). The presence of two different pathologic diagnoses in one kindred all sharing the same mutation might reflect different pathogenesis in the children in comparison with the adults, or the phenotypes may potentially overlap. Two affected brothers with probable IPF were sequenced to screen for the reported SP-C gene mutations, but none was observed (Yoshioka et al. 2004). The SP-C gene was also sequenced in 89 sporadic IPF patients and 46 NSIP patients (Lawson et al. 2004). Ten SNPs were revealed among IPF patients, but in none of the controls. Only one SNP in a single patient created an amino acid change, and one intronic SNP suggested a potential splicing site. The authors

concluded that according to their findings mutations in the SP-C gene do not explain the pathogenesis of IPF in the majority of sporadic cases.

Zuo and collagues (2002) used cDNA microarrays to identify candidate genes. They collected a pool of RNA obtained from five patients with pulmonary fibrosis; three had a histologic verification of IPF, one also had rheumatoid arthritis and one had Sjögren`s syndrome. Controls consisted of eight individuals who did not suffer from parenchymal disease. An array with 7129 genes revealed 164 genes with significantly different expression profiles between the two groups. The expression was increased for genes encoding smooth muscle proteins, immunoglobulins, and some complement factors, and genes encoding proteins that are involved in ECM formation, degradation, and signaling. Among all, matrilysin (matrix metalloproteinase 7, MMP-7), encoded on chromosome11q21-22, was the most up-regulated gene. They continued to study MMP-7 in an experimental model and administered bleomycin to the lungs of MMP-7 knockout mice, and their wild-type controls. Two to three weeks later, the histologic changes were milder among the MMP-7-knockout mice compared to the changes in wild-type controls. They suggested matrilysin as a mediator of pulmonary fibrosis.

3.3 Mouse models

Plenty of mouse models have been used in order to clarify the steps of the fibrotic cascade. Fibrosis has been introduced to the mice by several ways. Intratracheal instillation of bleomycin has been widely used. Other ways to induce fibrosis to mice are irradiation and administration of various fibrogenic agents. These models do not, however, represent the model of the pathogenesis of IPF, as it is understood today. The models include known etiological factors, inflammation precedes fibrosis, and, in mice as in humans, bleomycin- induced fibrosis responds quite well to steroid treatment, unlike in IPF.

A strain-dependent variability in the response to fibrosis-inducing agents has been recognized. 129J mice exhibit a reduced fibroproliferative response compared to C57BL/6 mice, which develop clear fibroblastic foci at the sites of injury. 129J mice are regarded as fibrosis-resistant and C57BL/6 mice as fibrosis-prone strains (Marshall et al. 1997, Brody et al. 2002). The 129J strain was crossed to the C57BL/6 strain, and the offspring were subsequently backcrossed to the inbred 129J founders. The backcross produced a generation of which approximately 25% carried a phenotype that, similarly to 129J, was protected from fibrosis. In both original strains the TGF-β gene was

overexpressed, which caused a significantly lower fibroproliferative response in 129J mice compared with C57BL/6 mice. Both experiments suggest that 129J mice carry a gene or a cluster of genes that protect them from the development of fibrosis, but what the gene/s is/are, is unknown (Warshamana et al. 2002).

Several transgenic or knockout mice have been developed in order to identify genes that could play a role in the pathogenesis of lung fibrosis. TGF-α, 5-lipoxygenase-derived leucotriene, PAI-1, or nuclear factor erythroid 2-related factor knockout mice developed significantly less lung fibrosis after being exposed to bleomycin than did the wild-type mice (Madtes et al. 1999, Cho et al. 2002, Peters-Golden et al. 2002, Chuang-Tsai et al. 2003), whereas the knockout for cyclo-oxygenase-2 led to increased lung fibrosis (Bonner et al. 2002). Overexpression of Smad7 protein in mice led to suppression of type I procollagen and reduced hydroxyproline content in lung tissue, and the mice did not develop fibrosis after bleomycin treatment (Nakao et al. 1999). Mice with overexpressed ET-1 developed progressive pulmonary fibrosis with accumulation of ECM proteins, and recruitment of inflammatory cells (predominantly CD4-positive cells) (Hocher et al. 2000).

Th1-biased C57BL/6 mice and Th2-biased DBA/2 mice were exposed to bleomycin either once or three consecutive times (Chuang-Tsai et al. 2003). Bleomycin induced a Th2-like environment in both strains. Fibrosis was detectable only after multiple exposures in Th2-biased DBA/2 mice. When Th1-biased C57BL/6 mice were knocked out from endothelial nitric oxide synthetase (which protects endothelial cells from oxidative injury), the similar prolonged fibrosis after multiple exposures was present. These experiments indicate that the fibrotic cascade is complex and can be determined by a number of genetic, and also environmental, factors (Chuang-Tsai et al. 2003).

Mice models have also been used to study gene expression profiles using cDNA microarray techniques. Munger and collagues (1999) described a mouse strain (129) that develop inflammation but not fibrosis in response to bleomycin. The strain had a null mutation of the epithelial-restricted integrin β6-subunit gene (β6$^{-/-}$). The mice were exposed either to bleomycin or saline, and gene expression of these β6$^{-/-}$ mice was compared to that in wild-type mice. The expression pattern of approximately 6 000 genes and ESTs were analysed simultaneously (Kaminski et al. 2000). A dramatic increase of expression among wild-type mice compared to the β6$^{-/-}$ mice was found in a cluster of genes that are involved in the formation of the ECM, regulate responses to the ECM, and are induced by DNA damage, such as p21. Quantitative comparison among these genes revealed even higher expression response of osteopontin, tenascin-C, tropoelastin, and heme-oxygenase.

Macrophage-restricted metalloproteinase expression was increased more than 20-fold in β6$^{-/-}$ mice compared to wild-type mice. The identified group of genes might play a role in the fibrotic process, as may the epithelial-restricted integrin β6-subunit itself.

A genome-wide scan has been performed in the C57BL/6J mice strain susceptible to fibrosis (Haston et al. 2002). The mice were given whole lung irradiation and the mice which developed histologically verified fibrosis were genotyped. The screen was performed with 154 markers with a mean intermarker distance of 10 cM. The data set revealed regions with LOD scores > 2.8 on chromosomes 1, 6, 17, 18, and 19. The study group had previously identified an association of the fibrotic phenotype with a marker on chromosome 17 (Haston and Travis 1997). Combining the two studies they concluded that on chromosome 17 there is a fibrosis susceptibility locus that includes a universal ´fibrotic` gene. Candidate genes mapping to that region in the mouse include TNF, manganese superoxide dismutase, plasminogen, and p21. In genomic comparisions, this linkage region matches with regions on human chromosomes 4p16 and the HLA-system on chromosome 6p.

AIMS OF THIS STUDY

1. To evaluate the prevalence of idiopathic pulmonary fibrosis in Finland according to the novel ATS/ERS 2000 international diagnostic criteria.

2. To evaluate the prevalence of familial idiopathic pulmonary fibrosis in Finland.

3. To identify novel candidate genes for idiopathic pulmonary fibrosis by a genome-wide scan and positional cloning.

4. To verify whether polymorphisms Pro1827Arg of the CR1 gene and Arg213Gly of the ECSOD gene, that functionally may be significant in the pathogenesis of the disease, are associated with idiopathic pulmonary fibrosis among Finnish patients.

MATERIALS AND METHODS

1. Patient selection

1.1 Identifying IPF patients

To identify all the IPF patients in Finland, we contacted all pulmonary clinics (N=29) in Finland during the years 1997-1998. Hospital databases were screened with the diagnosis J84.1, i.e. alveolitis fibroticans idiopathica in accordance with the ICD-10 classification. To evaluate as many patients as possible representing different parts of Finland, we selected all five university hospitals, the two largest central hospitals, and the largest regional hospital. To reconfirm the diagnosis we reviewed 50% of the medical records. In four central hospitals and in one regional hospital a local specialist in pulmonary medicine confirmed the diagnosis. For the rest of the hospitals (N=16) the number of IPF patients was extrapolated using statistics from the evaluated hospitals.

We mailed a questionnaire to all the still living patients identified by the primary screen (N=1212). In the questionnaire we asked whether they have or have had other affected family members and for the names and birthplaces of their parents and grandparents. We received 675 (56%) replies. From those patients (N=88) who reported an affected family member, we asked for more detailed pedigree information and with their permission examined their medical records. A familial case of IPF was confirmed when the medical data showed that both the proband and his or her affected family member fulfilled the diagnostic criteria. By using the Finnish church records we traced back 3-5 generations of all identified families to locate the earliest available birth places for family members to the confirm the geographic origin of the family.

1.2 Family selection for the genome-wide scan and for haplotype association analysis

In the genome-wide scan we included six pedigrees (Figure 1 of the original communication II). A total of 17 affected individuals were recognized, of which three were deceased at the time of the study. A total of 33 non-affected family members donated their blood samples for the genome scan analysis. For the haplotype association analysis the data set was extended by 12 nuclear families (proband, her or his spouse, and one offspring), two multiplex pedigrees without linkage information, and four singletons with a positive family history, but no DNA samples available from other family members. All nuclear families and singletons originated from the province of Savo and

the multiplex families from Central Southern Finland. A total of 33 patients and their 60 non-affected family members were genotyped. In addition to family based controls (N=27), we genotyped two markers (2,902,739 and 2,921,606; equalling their physical position in contig NT_106606.15) in 23 unrelated individuals from the Savo region (Koskenmies et al. 2001), and in 93 healthy blood donors from across Finland to estimate the frequency of the 2266-susceptibility haplotype in Finland.

1.3 Replication data set

During the study we continued to collect a replication data set (familial cases across the country, in collaboration with Finnish pulmologists and all cases in the Helsinki University district when first degree family members were available for genotyping). To date, the replication data set consists of one sib pair originating from the central part of Finland, and 10 sporadic cases from the Helsinki University district area. We have evaluated the susceptibility haplotype by genotyping the 2,902,739 and 2,921,606 markers for these samples (similarly as described in paragraph 1.2).

1.4 Patients and controls for association studies on the CR1 and ESCOD genes

For the association study on CR1 (study III) the population consisted of 96 patients. The control population consisted of 96 voluntary blood donors across Finland and 68 healthy controls from the Savo region.

Table 1. Clinical characteristics of the patients genotyped for the Arg213Gly variant in the ECSOD gene, the Pro1827Arg variant in the CR1 gene, and included in the fine mapping study. VC = vital capacity, DLCO = diffusing capacity for carbon monoxide.

Characteristics of the patients		**ECSOD**		**CR1**		**Fine mapping**	
Age, average (years);	(range)	62	(26-81)	62	(26-83)	62	(26-81)
VC (% of predicted);	(range)	71	(35-96)	74	(35-102)	69	(30-102)
DLCO (% of predicted);	(range)	57	(29-91)	58	(28-91)	57	(28-91)
Biopsy (N)		11	17 %	17	18 %	6	18 %
Male/ Female (N)		23/ 40		42/ 54		14/ 19	

For the association study on ECSOD (study IV) the patient population consisted of 63 patients (40 females and 23 males). The control population consisted of 61 unrelated population based controls; 28 of these were the spouses of the IPF patients.

2. Diagnostic criteria

The IPF diagnoses were made in accordance with the ATS/ERS international consensus statement (ATS 2000). The diagnostic criteria are shown in Table 2. For non-affected family-members, an interview revealed no clinical IPF in their medical history.

Table 2. The diagnostic criteria for IPF. When a biopsy was unavailable, all the major criteria and at least 3 of 4 minor criteria had to be fulfilled. For patients with a surgical biopsy showing UIP, only the major criteria were considered to be relevant.

	Major Criteria		**Minor Criteria**
1)	Exclusion of other known causes of interstitial lung disease	1)	Age >50 years
2)	Abnormal pulmonary function with restriction and/or decreased diffusing capacity	2)	Insidious onset of otherwise unexplained dyspnoe
3)	Bibasilar reticular abnormalities on HRCT scans	3)	Duration of illness >3 months
4)	BAL or TBB not pointing to another disease	4)	Bibasilar, inspiratory crackles on auscultation

All patients and controls signed a consent form when they donated their blood sample to the study. The Ethics committee of the Department of Medicine in Helsinki University Hospital, and the Ministry of Social Affairs and Health of Finland have approved the study.

3. Genome scan and fine mapping markers and genotyping methods

DNA was extracted from peripheral blood leukocytes by a standard non-enzymatic method (Lahiri et al. 1991). For the genome-wide scan, we used the Applied Biosystems Linkage Mapping Set MD-10 of 337 microsatellite markers. Genomic DNA (20 ng) was dried on microtiter plates for each PCR assay. The PCRs were performed in 5 µl volumes using reagent concentrations and temperature profiles as recommended by the reagent manufacturer (Applied Biosystems). The average interval between the markers was 10.6 cM. The genotyping success rate was 83%. The fluorescently-labeled PCR products were pooled (10-20 markers in each pool) and electrophoresed

in a Megabase 1000 capillary electrophoresis instrument (Molecular Dynamics). The alleles were called using the Genetic Profiler 1.1 software (Molecular Dynamics).

Fine mapping of chromosomes 3, 4, 9, 12, and 13 was performed using a total of 63 additional microsatellite markers and one SNP. The order of the primers and their sequences used in genotyping are shown in Table E1 in the original communication (II). Fine mapping markers were genotyped either using capillary electrophoresis as described above or gel electrophoresis on an ABI377 sequencer. Alleles were called using Genotyper 2.0 (Applied Biosystems). PCR amplifications with fluorescently-labeled primers were done in 10 µl reaction volumes containing 20 ng of genomic DNA, 0.5 µM of each primer, 0.2 µM of each dNTPs and 0.05 µM DNA polymeraze enzyme (AmpliTaqGold, Applied Biosystems) in the buffer with 2.5 mM MgCl$_2$. The samples were denatured at 94°C for 10 minutes, and then subjected to 30 seconds at 94°C, 30 seconds at 57°C and 30 seconds at 72°C for 35 cycles and were elongated for 10 minutes at 72°C. One SNP was genotyped from the short coiled-coil protein (SCOC) intron 1 (2739448A>G). Amplified PCR fragments were digested with the restriction enzyme EcoRI (New England BioLabs). The lengths of the PCR products were 175 and 117 basepairs (bp) for the minor alleles (A) and 292 bp for the major alleles (G). The reaction products were electrophoresed through 3% agarose gels stained with ethidium bromide for 2.5 hours and photographed under UV-illumination. Errors in Mendelian inheritance for all markers were checked using the PEDcheck program (O'Connell and Weeks 1998).

4. Genotyping of the Pro1827Arg variant in the CR1 gene

Genomic DNA comprising the Pro1827Arg polymorphism in CR1 was amplified using two primer pairs: 1) 5`CTTTTGTCCAAATCCTCCAG and 3`AAAGTTAAGCTCACAAACAAATACCA; and 2) 5`TTCAACCTCATTGGGGAGAG and 3`GGCAGGGCTGCTCCAAA. The PCR conditions were as follows: amplifications were performed in 10 ul reaction volumes containing 20 ng genomic DNA, 0.25 mM of each primer, 0.2 mM of each dNTP, and 0.25 U thermostable DNA polymerase (AmpliTaqGold; Applied Biosystems) in the buffer recommended by the manufacturer with 2.5 to 3.5 mM MgCl$_2$. The samples were denatured for 10 minutes at 94˚C, subjected to 31 to 35 cycles each of 30 seconds at 94˚C, 30 seconds at 59˚C, 30 seconds at 72˚C, and elongated for 10 minutes at 72˚C. The polymorphism was studied using two restriction enzymes HpyCH4III (PCR amplicon 1) and MnlI (PCR amplicon 2) (New England BioLabs). Different PCR amplicons were used to confirm unambiguous allele calling. The restricted PCR fragments were electrophoresed

through 2% agarose and 5% high resolution agarose gel (MetaPhor), respectively. PCR fragments were stained with ethidium bromide and photographed under UV-illumination. The length of the PCR product 1 for the major allele (C, referring to Pro1827) was 328 bp, and for the minor allele (G, referring to Arg1827) 164 bp + 164 bp (HpyCH4III specific restriction site). The lengths of PCR product 2 for the major allele (Pro1827) were 37 + 29 + 9 bp, and in the presence of the minor allele (Arg1827) 66 + 9 bp (MnlI specific restriction site).

5. Genotyping of the Arg213Gly variant in the ECSOD gene

The Arg213Gly variant in ECSOD was amplified using the primers 5´CGCCAGGCGCGGGAACACTCAG3´ and 5´GGCGGACTTGCACTCGCTCTCG3´. One mismatch was induced in both primers: one to delete the second digestion site for MwoI and the other to reduce the formation of secondary structures by the primer pair. PCR amplifications were done in 10 µl reaction volumes containing 100 ng of genomic DNA with AmpliTaqGold Polymerase. After amplification, the PCR fragments were digested with MwoI (New England BioLabs) in the buffer recommended by the manufacturer. The lengths of the PCR products were 63 bp for the minor allele (the Gly213 allele) and 28 bp and 35 bp for the major alleles (the Arg213 allele). The reaction products were electrophoresed through 4 % agarose gels stained by Gelstar (FMC Bioproducts) and photographed under UV-illumination.

6. Sequencing

A critical 110 kB region identified in the genome-scan and exons 4 to 9 of the ELMOD2 gene were sequenced in order to identify all the sequence polymorphisms. The reference sequence was generated by assembling the sequence from the Human Genome Project (in contig NT_016606.15 from position 2866166 to 2975765, a total of 109599 bp). Repeat regions such as SINE, LINE, LTR, MER1, and MER2 elements covered 48% of the sequence. A total of 157 PCR amplicons were designed to be ~650 bp in length with ~100 bp overlap with adjacent assays. Repeat regions were re-sequenced when possible to design primers in a unique sequence within the range mentioned above. SNP discovery was carried out in two selected patients (an affected father and his daughter), both heterozygous carriers of the identified susceptibility haplotype.

To reveal mutations in coding regions of the ELMOD2 gene, one patient representing each genome scan family and the haplotype-carrying father and his daughter were sequenced for all nine exons, and exon-intron boundaries of the ELMOD2 gene.

PCRs were performed in 10 µl reactions containing 5 ng DNA, 3 mM MgCl$_2$, 0.8 µM of each primer, 0.2 mM of each dNTP, and 0.03 U/µl of Hot Star Taq^{TM} polymerase (Qiagen). Reactions were heated at 95 ^{o}C for 15 minutes, subjected to 40 cycles of amplification (30 seconds at 95°C, 30 seconds at 50-62°C, and 30 seconds at 72°C) and a final extension of 10 minutes at 72°C. Unincorporated primers and dNTPs were removed by incubation with 0.4 U/µl of Exonuclease I (New England BioLabs) and 0.072 U/µl shrimp alkaline phosphatase (Amersham Biosciences) for 30 minutes at 37°C. All exons were sequenced on both strands using the DYEnamic ET Dye terminator kit (Amersham Biosciences), following the manufacturer's instructions. Cleaned sequencing products were injected for 80 seconds at 3 kW and electrophoresed for 100 minutes at 9 kW or 180 minutes at 6 kW using MegaBACE long read matrix on a MegaBACE 1000 instrument (Amersham Biosciences). The sequences were analysed with the Sequence Analyser v3.0 software (Amersham Biosciences).

To detect sequence variations, the sequence chromatograms were inspected visually by two independent readers and the sequences aligned using the Pregap and Gap4 softwares from the Staden package (http://staden.sourceforge.net).

Genomic DNA comprising the Pro1827Arg polymorphism (PCR done with amplicon pair 1) was sequenced in an ABI3730 Automatic DNA sequencer (Applied Biosystems) (original communication III).

7. mRNA expression

Human Multiple Tissue cDNA panels I & II (BD Biosciences) were used for ELMOD2 and LOC152586 mRNA expression studies. ELMOD2 expression was studied also with the commercial Human Blood Fractions MTC cDNA panel (BD Biosciences), commercial fibroblast cell lines CCL-151 (healthy) and CCL-134 (derived from IPF patient), and lung biopsies. IPF biopsies were derived during transplantation operations from surgically verified UIP patients (N=6), and controls consisted of healthy lung areas derived from patients who went through thoracotomy because of solitary lung infiltrates (N=7). A pulmonary pathologist verified that the controls represented

healthy lung areas, and the diagnoses of all patients were based on light microscopic evaluations using the histologic criteria presented by Katzenstein and Myers (1998). The biopsies were retrieved from the Departments of Pathology and Pulmonary Medicine, Helsinki University Central Hospital.

Total RNA was extracted from snap-frozen lung biopsies by mechanical homogenization and Phase Lock Gel Heavy kit according to the manufacturer´s (Eppendorf) instructions. cDNA was synthesized with TaqMan reverse transcription reagents (Applied Biosystems) and oligo dT primers according to the manufacturer´s instructions. PCRs were performed in 25 µl volumes containing 1.0 µl cDNA as template, 2.5 µl 10X PCR Gold buffer, 1.5 mM MgCl$_2$, 0.2 mM dNTPs (Finnzymes), 200 nM primer, and 0.25 U AmpliTaqGold (Applied Biosystems) under the following conditions: 94°C for 10 minutes; 36 cycles of 94°C for 30 seconds, 60°C for 30 seconds, 72°C for 1 minutes followed by a final extension of 72°C for 10 minutes. The primer sequences for ELMOD2 were 5'-TTCTTTGTGGGAGTTCTTCTA-3' (spanning exons 2-9) and 5'-TGAAAAGATTAAAGGACTTTTACTGGA-3', respectively, and for LOC152586 5'-CCGTCCCTGGCATTATCC-3' and 5'-CTCAGTCGCTGCAATTTCC-3' (covering 1143 bp of the cDNA), respectively. The accuracy of the PCR fragments was verified by sequencing. Primers for GAPDH were supplied with the kit and used according to the manufacturer´s instructions.

The relative mRNA expression of ELMOD2 in the patient group was compared with controls after equalizing with the relative expression of mRNA GAPDH used as an endogenous reference gene. Syngene GeneTools software (Synoptics Ltd.) was used for detecting differences between intensities of the corresponding lanes.

8. *In vitro* translation

Capped RNA molecules of the LOC152586 gene were transcribed from an IMAGEclone 5267198 DNA construct with T7 RNA polymerase (mMESSAGE mMASCHINE system, Ambion). Translation was performed with rabbit reticulocyte lysate translation machinery (Riboprobe in vitro translation system, Promega) in the presence of Redivue-L-[S^{35}]-methionine (Amersham Biosciences) in the reaction mixture. The Xenopus elongation factor α (pTRI-Xef) DNA template was used as a positive control for transcription and translation, while in negative control, water replaced DNA. The translated polypeptides were detected by autoradiography after Tris-Tricine SDS-PAGE.

9. *In situ* hybridization

The cDNA sequence of ELMOD2 was amplified from the IMAGEclone 3897166 with a primer pair containing the promoter sequence for SP6 and T7 RNA polymerase: forward (sense <u>T7</u>) 5'-<u>TAATACGACTCACTATAGG</u>GTTCCGTCGTTTCCGTT-3' and reverse (antisense <u>SP6</u>) 5'-<u>ATTTAGGTGACACTATAGA</u>ATTACAATCCAGTAAAAGTCCTTT-3'. Antisense and sense probes for ELMOD2 were transcribed by SP6 and T7 RNA polymerases (MAXIscript *in vitro* transcription kit, Ambion) in the presence of digoxigenin-11-uridine-5´-triphosphate (Dig-11-UTP, Roche).

Non-radioactive *in situ* hybridization on tissue sections was performed with a Ventana Discovery™ device. In brief, the samples were deparaffinized with heat treatment followed by post-fixation and RiboClear pre-treatment. The samples were protease-treated for 18 minutes and hybridized for 6 hours at 65°C with both antisense and sense probes. The slides were then washed three times with 0.1X SSC (15 mM NaCl, 150 nM Sodium citrate, pH 7.0) at 75°C followed by the detection step, which included 20 minutes incubation with biotinylated anti-DIG antibody (Jackson ImmunoResearch Laboratories) and 2 hours incubation with the BCIP/NBT substrate. After the labelling, the slides were washed, dehydrated, and mounted with Mountex (HistoLab). All the reagents for Discovery™ were provided by Ventana Medical Systems except for protease K (Roche), which was used at a concentration of 350 ng/µl.

10. Statistical analyses

10.1 Linkage analysis

Genome-wide nonparametric multipoint linkage analysis and haplotype re-construction were done with GENEHUNTER 2.0 for six multiplex families (Kruglyak 1996).

10.2 Power estimations for linkage analysis (simulations)

To estimate the genome-wide power to detect linkage in our data set, we performed power simulations using the following model of inheritance: autosomal dominant with reduced penetrance of 0.9 and a frequency of the susceptibility allele in the population of 0.005. For linkage we used a two-point analysis with the 10 cM marker map. All markers had four equally informative alleles and no missing data was allowed. Iterations were made using the true pedigree structures, and the number of iterations was 2000.

10.3 Association analyses

Two-tailed Student's tests were used to compare the differences between familial and sporadic IPF patients in the age of onset of the disease (study I), and to compare mRNA expression between cases and controls (study II). Comparisons between the patient and control groups were done using the chi square test (studies III and IV), and comparisons of susceptibility haplotype carriership between affected individuals and controls were analysed using Fisher's two-sided exact tests. Odds ratios were calculated using the interphase at http://home.clara.net/sisa/fisher.html (study II).

RESULTS

1. Prevalence

From the hospital data bases we were able to identify a total of 1445 in- or out-patients who were given the diagnosis J84.1 (Alveolitis fibroticans idiopathica) during the years 1997-1998 in Finnish pulmonary clinics. To verify the diagnoses we evaluated 50% of the medical records from eight centers; all five university hospitals, the two largest central hospitals and the largest regional hospital. These eight centers offered medical care for 66% of all identified patients. In each of these hospitals we used the proportion of confirmed diagnoses to extrapolate the total number of IPF patients per center. The percentages of confirmed diagnoses in different centres varied between 49% and 77%. This range of variation was then used in the extrapolation of the number of IPF patients in the rest of the centres. Using this approach we were able to identify a total of 833-943 IPF patients, equivalent to a prevalence of 16-18/100 000 among the Finnish population of 5.17 million. Gender predominance was not perceived among the patients.

IPF was more frequent in eastern Finland. In an area covering two hospital districts, the southern and eastern Savo Hospital districts, the prevalence was 45/100 000, which was more than twice the average (Figure 4). To identify the right patients, the medical histories in these centers were first reviewed by the local pulmologists and later a specialist from our group (U.H.) also went them through to see that there were no false positives to affect the high prevalence.

The international guidelines (ATS/ERS 2000) encourage taking a lung biopsy to verify the diagnosis. According to the evaluated medical records, in 28.2% of cases (N =75) the IPF diagnosis was verified with a biopsy featuring a UIP pattern and the major criteria. In the remaining 71.8% of cases the diagnosis was based on the major and minor criteria, without biopsy verification.

2. Familial IPF

Eighty-eight of the patients identified in the primary screen reported to us that they have or have had an affected family member. Medical records were then studied in detail. In 17 families we were able to verify 2 to 5 affected first-degree family members. The clinical course of the disease, HRCT scan findings, and histological appearance were indistinguishable between the familial and sporadic IPF patients. The age of onset, however, was slightly younger among the familial cases compared to that among the sporadic cases (mean 61.9 yrs vs. 65.3 yrs, p=0.11). Again, no gender predominance was perceived. Based on this study approach, the prevalence of familial IPF was 5.9 per million (31 living patients at the time of the study), explaining 3.3-3.7% of all the IPF cases.

To study the geographic distribution of the multiplex families, we plotted the birthplaces of the parents of the probands on to a map of Finland (Figure 4). The majority of the parents originated from the same region where we also observed the highest prevalence of IPF. In this enrichment area we observed 50-fold risk of FPF compared to the rest of the country. The birthplaces of the parents clustered in certain neighbouring municipalities. The most significant clustering of 10 parents was found in 3 rural municipalities (Tuusniemi, Heinävesi and Kerimäki) in the province of Savo. The second cluster of five parents originated from three rural municipalities (Impilahti, Sortavala and Ruskeala) in Carelia, a province that formerly belonged to Finland. These two clusters are situated within 200 kilometers of each other. The Finnish origin of the families from Carelia was confirmed by their family names and their membership in the Lutheran church. By using parish records, we have been able to trace each family back 3-5 generations and found that these families were already settled in these regions generations ago, but no obvious loops between the pedigrees were observed.
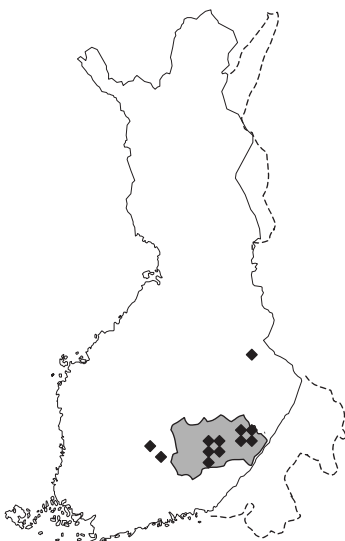

Figure 4. Birthplaces of the parents of the multiplex families in the genome-wide scan (♦). The southern and eastern Savo Hospital districts, where the prevalence was more than twice the average, are shown in gray.

3. Genome wide linkage analysis

Genotyping was performed in six multiplex families, where 14 living patients and their 33 first-degree relatives provided linkage information (GS in Figure 5). All of the formerly identified 17 multiplex families were not suitable for the linkage analysis because 11 of these families did not provide information for linkage analysis. Five of the six pedigrees (Pedigrees 3, 4, 5, 13, and 17 in Figure 5) originated from the enrichment area in Savo (Figure 4). The sixth pedigree (Pedigree 9 in Figure 5) originated 150 km west of the other families.

For the genome-wide scan we used a commercially available microsatellite marker set (Applied Biosystems Linkage Mapping Set MD-10). The success rate of genotypes was 83%. The average interval between the markers was 10.6 cM. None of the markers showed an exceptional tendency for Mendelian errors. In the linkage analysis, the three best loci were located on chromosome 3, marker *D3S1278* (NPL-score 1.7, p=0.06, information content 57%), on chromosome 4, marker *D4S424* (NPL 1.7, p=0.05, information 60%), and on chromosome 13, marker *D13S265* (NPL 1.6, p=0.06, information 71%). Using nonparametric linkage analysis (Kruglyak et al. 1996) none of the loci reached genome-wide significance.

To better understand the power of our data set to detect true linkage in a genome-wide scan, we performed simulations. In the simulations we used the true pedigree structures and made the assumption that FPF is inherited in an autosomal dominant fashion with reduced penetrance. In the case of true linkage, the locus was identified in 60% of the permutations with a LOD score >1, in 22% of the permutations with a LOD score >2, and 4% of the permutations with a LOD score >3. In the case of no linkage, LOD scores >1 at any given locus were observed only in 1.6% of the simulations. The results suggested that because of the small number of pedigrees, we could still miss true linkage with a likelihood that should be considered. Therefore, we visually inspected the haplotypes reconstructed by GENEHUNTER in all the chromosomal regions that showed positive NPL-scores and were potentially shared by the affected individuals within a family and across the families. Two chromosomal regions appeared to be of interest. On chromosome 9, all three patients in one family shared a 33 cM haplotype (markers *D9S175*, *D9S167*, *D9S283*, *D9S287*), while in five other families, 10 of 16 patients shared an allele as part of the haplotype either at *D9S175* or *D9S167*. The best NPL-score of 1.3 was obtained for *D9S167* (p=0.09, information 68%). On chromosome 12, a 25 cM haplotype (*D12S364*, *D12S310*, *D12S1617*, *D12S345*) was shared by six

of 16 patients and all 16 patients shared the same allele at least at marker *D12S310* (NPL 0.48, p=0.3, information 56%). These five loci on chromosomes 3, 4, 9, 12, and 13 were all selected for fine mapping. The genome-wide scan results are shown in the original communication (II, Figure 2).
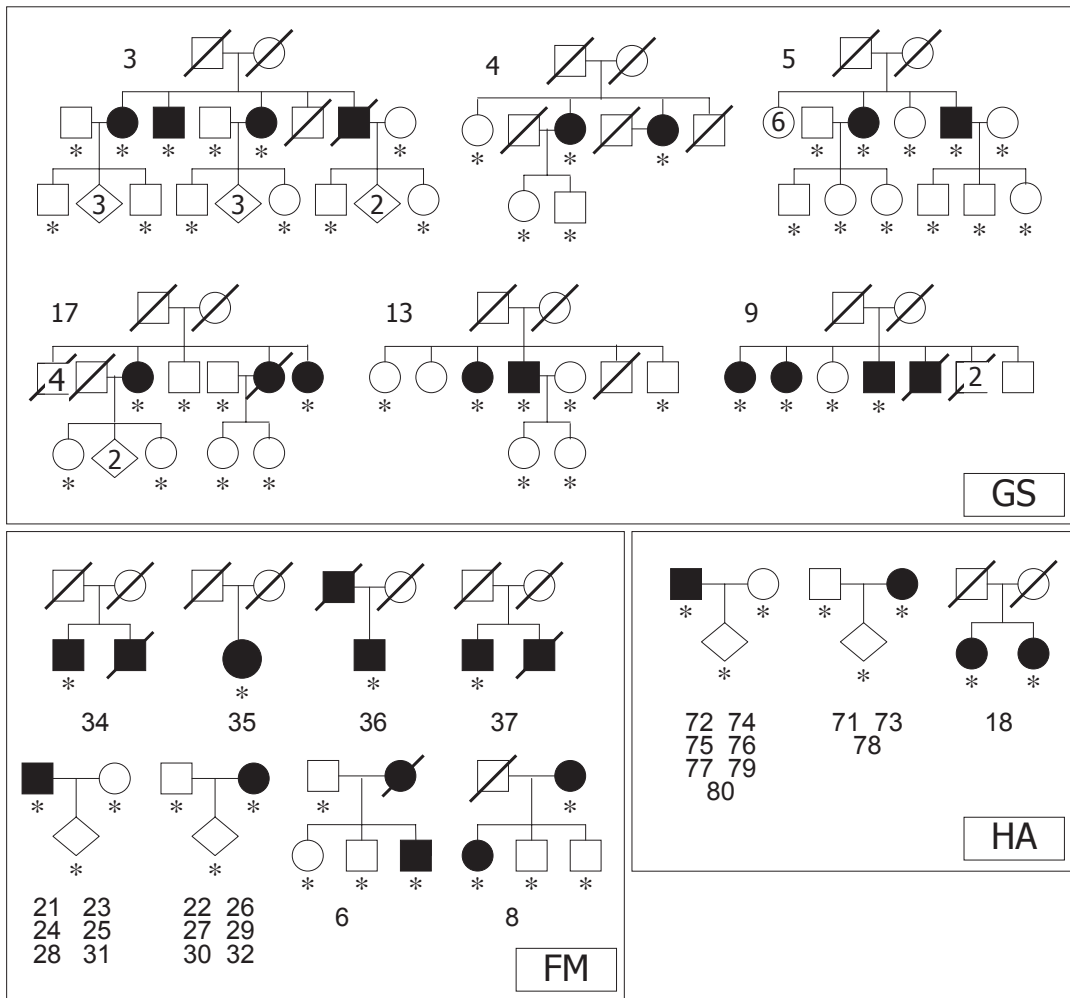

Figure 5. Pedigree structures in the genome-wide scan for familial IPF (GS), fine mapping (GS+ FM), and the susceptibility haplotype association study (GS+ FM+ HA). Affected family members are shown in black. Individuals who donated blood for genotyping are marked with an asterisk. The pedigree identification numbers are also used in Table 1, and Figures 1 and 3, of the original communication (II).

4. Haplotype association analysis on chromosomes 3, 4, 9, 12, and 13

To maximize the information content for linkage and to detect or exclude a potential haplotype association, we used the principals of hierarchical genotyping in the extended family data set (GS + FM in Figure 5). The probands of the 12 nuclear families originating from the province of Savo, as well as the four singletons who had a positive family history, were also genotyped with fine mapping markers. We added a total of nine markers to 3q13, 28 markers to 4q31, six markers to 9q21, four markers to 12p12-q12, and 17 markers to 13q31. With the additional markers linkage became stronger on chromosomes 3 (at *D3S1303*, NPL 2.07, p=0.05, information 83%), chromosome 4 (at *D4S1586*, NPL 2.09, p=0.02, information 74%), and chromosome 13 (at *D13S281*, NPL 2.4, p=0.01, information 84%), but weaker on chromosome 12 (at *D12S310*, NPL 0.1, p=0.4, information 67%), and chromosome 9 (at *D9S167*, NPL 1.0, p=0.2, information 79%). In additional to weakened linkage results on chromosomes 9 and 12, potentially shared haplotypes were broken down (marker densities within the linkage peak after fine mapping were on average 4.1 cM and 3.7 cM, respectively) and the chromosomes were excluded. On chromosome 3 with additional 9 markers (marker density on average 4.4 cM), no evidence of a shared haplotype between patients was found and it was excluded. On chromosome 13, a conserved haplotype among patients was observed, but the same haplotype was as common among the non-affected family members, showing no evidence of a disease association.

Contrary to other loci, on chromosome 4 the patients shared a haplotype in eight of the 24 families (Figure 3, original communication II). The shared haplotype was at the shortest 110 kb and at the longest 13 cM, defined by 4-16 consecutive highly informative markers. At the shortest, the families shared four consecutive alleles 2266. The susceptibility 2266-haplotype was delimited upstream and downstream by two recombinations.

In the Finnish population the frequency of the shared 2266-haplotype (2873878*2, 2894859*2, 2902739*6, and 2921606*6) is rather low. Among the unrelated family-based controls, nobody (N=27 tested) was a carrier. To get a better estimate of the 2266-haplotype frequency, we genotyped markers 2902739 and 2921606. These markers were always inherited in LD and informative enough to tag the 2266-haplotype. The estimation was performed in a study group which consisted of 1) the patients (one family was represented by one affected family member) and their available spouses who were already genotyped in the genome scan and fine mapping studies (N=24 patients, N=27 controls) (GS+FM in Figure 5), 2) the replication data set with 10 patients

and one sibpair (N=11 patients) (HA in Figure 5), and 3) additional regional controls (N=23) and unrelated Finnish individuals (N=93). The total number of genotyped families was 35; 13 of these were multiplex families and 22 seemingly uniplex families. The total number of controls was 143; 50 regional and 93 across Finland. None of the patients or controls was a homozygous carrier of the haplotype. In 38% (5/13) multiplex families, the affected family members shared the susceptibility haplotype (families 6, 17, 18, 34, 36 in Figure 5). Correspondingly, in 27% (7/22) uniplex families (families 21, 23, 24, 31, 73, 74, and 77 in Figure 5) the proband was a 2266-haplotype carrier. Among the control individuals from the Province of Savo, the heterozygous carriership of the haplotype was 4% (2/50), and 9.6% (9/93) among the controls across Finland, respectively. The carriership of the susceptibility haplotype was significantly higher among the patients (12/35) when compared to the regional controls (2/50 individuals, OR=12.5, 95%CI=2.6–60.6, p=0.0004), and to the combined pool of controls (11/143 individuals, OR=6.3, 95%CI=2.5-15.9, p=0.0001).

5. Positional candidate genes

Based on the human gene annotations publicly available, at least two genes are located within the susceptibility haplotype (http://www.sanger.ac.uk, http://www.genome.ucsc.edu, Stausberg et al. 2002). One of the genes is ELMOD2 (also known as MGC10084, HGNC:28111, 9830169G11Rik) (GenBank Source: BC015168, GeneID: 255520), and the other is LOC152586 (GeneID: 152586, similar to RIKEN cDNA 4933434I20). The genes are encoded in opposite directions, but are not overlapping. Three exons, 1-3, of the ELMOD2 gene are located within the 110 kb critical region, whereas the entire coding sequence of LOC152586 occurs totally within that region (Figure 4, original communication II).

ELMOD2 consists of nine exons encoding a 293 amino acid protein. The experiments with the commercially available human tissue cDNA panels showed that this gene is expressed in multiple tissues and cell types, including the lung, and both in healthy and IPF-derived fibroblasts (Figure 5, original communication II). mRNA expression of ELMOD2 was further studied with lung biopsies derived from patients with verified UIP (N=6). The biopsies were taken during lung transplantation, which was performed because of end-stage IPF. The control biopsies were from healthy lung areas, confirmed by a pulmonary pathologist, from patients who underwent oparations because of solitary lung nodules (N=7). The mean intensity of mRNA expression was significantly decreased among

the patients compared to the controls (p=0.05) (Figure 7, original communication II). Using *in situ* hybridization we detected the expression of ELMOD2 in epithelial cells and alveolar macrophages of a healthy adult lung (Figure 6, original communication II). The function of ELMOD2 is poorly known, but it belongs to a protein family that expresses a highly conserved domain found in a number of eukaryotic proteins including CED-12 and ELMO1-3. These molecules are known to interact in signaling pathways involved with apoptosis, phagocytosis, cell engulfment, and cell migration.

LOC152586 is even less well characterized. Several overlapping expressed sequence tags (ESTs) that are most probably encoded by a single gene can be found in the Human Genome Assembly Data Base. Extensive altered splicing is suggested from different transcripts and all potential open reading frames are short. All the identified exons are located within the critical region. Based on RT-PCR results this gene was only expressed in the testis (Figure 5, original communication II). We carried out an *in vitro* translation assay using an IMAGEclone 5267198 that contained one of the longest potential open reading frames (147 amino acid), and we failed to produce any peptide (Figure 5, original communication II).

6. Sequencing

In order to identify possible genetic variation in the susceptibility haplotype, we sequenced this region (from position 2,866,166 to 2,975,765 in the public sequence NT_016606.15) in two individuals (an affected father and his daughter), both heterozygous carriers of the haplotype. When the sequences were compared to the public sequence (NT_016606.15), we observed 39 polymorphisms: 19 were heterozygous and 18 homozygous single nucleotide polymorphisms, one single nucleotide insertion, and one two-nucleotide deletion (Table E2, original communication II). Most of the polymorphisms had not been previously reported. None of them appeared to locate on the coding areas of either ELMOD2 or LOC152586.

All nine exons of the ELMOD2 gene (including exons 4-9 outside of the critical region) were sequenced in the affected father and his daughter, and in six patients representing each genome scan family. No differences between the exons and the public sequence were detected.

7. Association study on the Pro1827Arg variant in the CR1 gene

The Pro1827Arg polymorphism in CR1 was studied among 96 IPF patients and 164 controls. For genotyping we used two different restriction enzymes recognising different restriction sites to detect the studied polymorphism. HpyCH4III did not digest any of the PCR fragments, suggesting that all patients and controls were Pro1827 homozygous. In the absence of positive controls, we confirmed the genotyping results with another restriction enzyme, MnlI. Consistent with the previous results, again, only the major allele (Pro1827) was recognized. Since the results did not match our primary hypothesis, we then verified by sequencing that all patients were Pro1827 homozygous, and that the site is not polymorphic among Finns.

8. Association study on the Arg213Gly variant in the ESCOD gene

The Arg213Gly polymorphism in ECSOD was studied in 63 IPF patients and 61 population-based control subjects. The carriership of the Gly213 allele was 2.5%. One of the patients and three control subjects were heterozygous carriers of the Gly213 allele. There was no association between the Gly213 allele and IPF. The study showed that the polymorphism is very rare among Finns and our study design had limited power to show an association or exclude it.

DISCUSSION

1. Prevalence of IPF

Since the first description on fibrosing alveolitis by Hamman and Rich in 1944 it has become obvious that patients with interstitial pneumonias show differences in the clinical presentation as well as in histological appearance. Gradually it was noticed that IPF differs from other IIPs not only in its clinical and histological manifestations, but also in the pathogenesis and prognosis. Because of the variable and confusing diagnostic criteria and terminology of IIPs, the international multidisciplinary panel, nominated by the American Thoracic Society and the European Respiratory Society, released international consensus statements concerning first, IPF: diagnosis and treatment (ATS 2000), and second, classification of the IIPs (ATS 2002).

Our study on the nationwide prevalence of IPF was the first retrospective survey done according to the novel criteria to diagnose IPF. Evaluation of the case records showed that the diagnosis J84.1 according to the ICD-10 classification, which should refer to IPF, was used quite liberally as a primary diagnosis for diffuse paranchymal lung diseases. In clinical practice, the diagnosis J84.1 was obviously used as a primary diagnosis and then specified later in the diagnostic process when other causes for symptoms and lung involvements were determined. Depending on the center, 23% to 51% of the patients were excluded. The excluded patients suffered from a variety of interstitial lung diseases, such as connective tissue disease-related pulmonary manifestations, cryptogenic organizing pneumonia, eosinophilic pneumonia, asbestosis, radiation therapy-related or nitrofurantoin-induced fibrosis, allergic alveolitis, and eosinophilic pneumonia.

The prevalence of IPF in Finland was 16-18/100 000, which is in concordance with reports from other populations (Scott et al. 1990, Coultas et al. 1994, Iwai et al. 1994, Hubbard et al. 1996). In study I, southern and eastern Savo showed a prevalence of 45/100 000, a more than two-fold increase compared to the nationwide prevalence. The diagnostic accuracy was confirmed by re-evaluating the case records. The re-evaluation further confirmed the higher prevalence and the enriched geographical distribution.

2. Familial IPF

Familial and sporadic forms of IPF are clinically and histologically indistinguishable, suggesting common pathogenic pathways (Marshall et al. 1997, Marshall et al. 2000, Lee et al. 2005). Our clinical findings in familial cases in study I revealed no other differences except a slightly younger age at the time of diagnosis among FPF patients (61.9 yrs vs. 65.3 yrs, p=0.11). Additionally, we observed no differences in the clinical characteristics between the familial patients studied in the genome-scan, fine mapping and haplotype association analyses and the sporadic patients in the haplotype association study (Table 1, original communication II). Therefore, we can assume that by discovering genetic defects in familial IPF, the results can pinpoint genes and signaling pathways of interest also for the sporadic forms of the disease.

During our study we were able to identify 17 families with 2-5 affected family members. Most of the affecteds were siblings, only three parent-offspring pairs were detected (Table 2, original communication I). The low number of parent-offspring pairs may be due to several reasons. Some of the parents had died before the age of 60 from other causes. In some cases the probands most likely have insufficient knowledge of their family disease history, and the diagnostic procedures to identify the disease were limited among the older generations. We had excluded IPF among the non-affected family members through an interview by asking whether they were diagnosed with any IIP. It is known that especially first degree relatives may express some signs of the disease (Bitterman et al. 1996, Steele et al. 2005), and it is therefore possible that we have ruled out some appropriate families. The number of familial cases identified implied a prevalence of 5.9 per million for familial IPF, explaining 3.3–3.7% of IPF in Finland. That is four times higher than reported in Great Britain (Marshall et al. 2000). Preliminary reports from the United States suggest that the familial form may explain even 20% of IPF (Loyd 2003).

The birthplaces of the parents of our IPF patients were plotted on to the map of Finland. Surprisingly, the majority were clustered in the southern and eastern Savo region; the very same region where we observed that the prevalence of IPF was two times higher than in the rest of Finland. Each family has been traced back 3-5 generations and according to parish records these families were settled in these same regions already in the beginning of the 19th century. We were not, however, able to detect any direct genealogical loops between the families. Clustering of IPF families has been partially explained by common environmental factors (Steele et al. 2005). In none

of the familial cases was exposure to any known environmental risk factors for IPF. Therefore, the strong clustering supports the importance of genetic factors in the pathogenesis of IPF.

Among the Finns, for many inherited diseases, one major mutation is present in $\geq$ 70% of disease chromosomes (Peltonen et al. 1999). Although some of the diseases are spread throughout the country, they are mainly caused by the same mutation event carried to Finland. When a disease is caused by one major mutation, LD is likely to be detected. The longer the observed genetic interval showing LD, the more obvious the clustering of the grandparents of patients and the younger the disease-predisposing mutation (Peltonen et al. 1999). Within Finland the longest observed LD intervals are 13 cM for congenital chloride diarrhea (Höglund et al. 1995), in which the families originated from the subpopulation of Kainuu, and 11 cM for vLINCL, in which the origin of families is in southern Botnia (Varilo et al. 1996). The pattern of the distribution of the grandparents of patients with vLINCL (Varilo et al. 1996) remarkably resembles that of the parents of FPF patients. Combining the geographical distribution of the parents of the FPF patients and the knowledge of Finnish population history, we assumed that patients with FPF might share a common, ancestral disease-causing allele, and due to the founder effect, with a even smaller study population this might be possible to identify.

3. Identifying novel candidate genomic region for IPF by genome-wide scan combined with hierarchial association study

Collecting families with $\geq$ 2 affected family members suitable for studying linkage in IPF is challenging; IPF is a quite rare disease with late onset and short survival after diagnosis, and the phenotypes also have to be accurate. In our nationwide study we identified 17 multiplex families. All the families were willing to participate in the study. In six of the families there were sufficient family members delivering linkage information, and they were included in the genome-wide scan without further selection. In three of the originally identified families, all the affected family members had died before we were able to contact them. In six families there was one living patient and in two families an affected parent-offspring pair who participated along in the study with their first degree family members. Five of the six genome scan families originated from the above mentioned Savo cluster (Heinävesi-Tuusniemi-Kerimäki). Based on power estimations assuming that at least some of the probands are offspring of the same ancestor, we presumed that most likely these six families would not provide enough information to find a significant linkage signal (Lander and Kruglyak 1995). Therefore, already before the genome scan, we made simulations to estimate

the power of our data set to obtain significant linkage. According to the simulations, it was possible to miss true linkage, and therefore we included in the fine mapping also the regions that showed potential haplotype sharing.

The genome-wide scan identified five interesting regions; on three chromosomes (3, 4, and 13) loci obtained NPL-scores of 1.6-1.7, and on chromosomes 9 and 12 a possible shared haplotype was detected. These regions were characterized in more detailed with fine mapping, which was performed with a hierarchial strategy. For fine mapping we recruited two other families with FPF without linkage information, 12 trios, and four singletons from the region of IPF enrichment. After fine mapping, the possible haplotypes seen on chromosomes 9 and 12 were broken and the NPL-scores decreased, to 1.0 and 0.1, respectively, and these chromosomes were not studied further. On chromosomes 3 and 13 the NPL-scores increased, to 2.1 and 2.4, respectively. After the fine mapping the average distances between markers were 4.4 cM and 2.2 cM on chromosomes 3 and 13, respectively. On chromosome 3 we were unable to see any shared haplotypes. On chromosome 13, on the other hand, a shared 7 cM haplotype was obvious, but as it was seen as frequently among patients and non-affected family members, we considered it to merely represent a common haplotype.

After adding 31 markers on chromosome 4q31.1 the marker *D4S1586* obtained suggestive linkage with an NPL-score of 2.1 (p=0.02, information 74%). The flanking region included the 110 kb haplotype that was shared by one third of the patients, and was not detected among any of the non-affected family members. Because our data set was small, we screened the susceptibility-haplotype in our replication data set in an affected sister pair and 10 single IPF patients. The sister pair and three singletons carried the susceptibility-haplotype consistent with our previous findings. The susceptibility haplotype was detected in 7.7% of regional (2/50) and non-regional (9/93) Finnish controls, significantly less frequent than among the IPF patients (34%), and resulting in an odds ratio of 6.3 (95%CI=2.5-15.9, p=0.0001). To rule out the possibility that the identified susceptibility haplotype is only a regional enrichment of a rare haplotype we compared the patients and regional controls, and the difference was still highly significant (OR 12.5, 95%CI=2.6-60.6, p=0.0004). If we have misinterpreted true IPF cases as unaffected we may have weaken, but not overestimated, the association.

Although the frequency of the susceptibility haplotype was slightly higher among the multiplex families than among the uniplex families (38% vs 27%), carriership could not separate them

phenotypically. The carriership of the susceptibility haplotype did not seem to modify the clinical disease: the clinical characteristics of the haplotype carriers and non-carriers patients were indistinguishable. Five of the multiplex families and four of the uniplex families originated from the enrichment area. It is presumed that some of the seemingly uniplex families are multiplex families, but due to the lack of pedigree information, and renewed diagnostic methods, the familiality is not recognized. One of the multiplex and one of the uniplex families shared a common 8 Mb haplotype (families 17 and 21 in Figure 3, original communication II), which further suggests that they are related within 10-15 generations.

In order to identify possible mutations the susceptibility haplotype-comprising region was sequenced in a father carrying the haplotype, and his daughter. When compared with the public sequence, the sequencing revealed 37 SNPs (Table E2, original communication II). None of the polymorphisms were located in the coding areas of the LOC152586 gene, or in exons 1-3 of the ELMOD2 gene, nor were they predicted to alter reading frames or altered splicing sites. The identified susceptibility haplotype was a rather SNP- and gene-poor region of the genome. 48% of the region was excluded from the sequencing based on repetitive sequence. Whether the excluded region includes polymorphisms that may result in altered function of these genes is unknown and deserves to be clarified further. Thus far, when no functionally important mutations within the haplotype-comprising region have been identified, we cannot claim any correlation between the susceptibility haplotype and the risk of clinical manifestation of the disease.

4. Characterization of candidate genes

According to the recent public databases, two genes are located within the critical region (http://www.sanger.ac.uk, www.ncbi.nih.gov/IEB/Research/Acembly), ELMOD2 and LOC152586. The sequence of ELMOD2 is defined by 140 cDNA clones. The gene consists at its longest of nine exons and is alternatively spliced into 4 different transcripts, together encoding 4 different protein isoforms. The longest protein isoform is 293 amino acid long, with a molecular weight of 34.9 kDa. There are 3 probable alternative promoters and 2 non-overlapping alternative last exons. The gene covers 29 kb on the direct strand, and exons 1-3 are encoded within the critical region (www.ncbi.nih.gov/IEB/Research/Acembly). According to the public databases and our own mRNA expression studies, ELMOD2 is expressed at very high levels in various tissues and cell types, among them lung and fibroblasts, which are both relevant in the pathogenesis of IPF. Using *in situ* hybridization we recognized expression of ELMOD2 in alveolar macrophages and alveolar

walls in healthy adult lung. mRNA expression studies showed that the expression of ELMOD2 is decreased in IPF lung compared to healthy lung. That might reflect that the lack of ELMOD2 expression is associated with IPF. Because we do not understand the function of ELMOD2, we cannot state whether the lack of expression leads to the fibrotic process, or if it`s just an end-stage phenomenon in mature fibrosis.

ELMOD2 belongs to a family of protein with a highly conserved domain which is found in a number of eukaryotic proteins including CED-12 and ELMO1-3. It is similar to the Caenorhabditis elegans gene, ced-12, which is required for the engulfment of dying cells and cell migration. Its function is poorly understood, but due to the conserved motif it is thought that ELMOD2 plays a part in regulating phagocytosis, and participates in apoptosis and cell migration (MIM 606420-606422, http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM). All of those are functions that may play a role in the pathogenesis of IPF, and thus suggesting ELMOD2 not just a positional, but also a functionally promising candidate gene for IPF.

The other gene, LOC152586, is coded in the reverse direction with the whole 0.5kb coding area within the critical region. The function has not been characterized, but overlapping ESTs most probably are encoded by a single gene, and have been recognized in broad spectrum of human tissues showing extensive splicing. EST sequences include open reading frames; the longest one, 147 bp, we identified from the sequence supported by human cDNA IMAGEclone 5267198. According to our *in vitro* translation experiment LOC152586 may not be a protein-coding gene. The sequence, however, is defined by 3 cDNA clones, and appears to contain a domain that has homology to a mouse genome (ref:NP_036346.1 vs. ref:NP_080509.1). The conserved domain suggests that it may possess a regulative function on translational or transcriptional phases of protein production.

Because the true disease-causing variant may occur only in a subset of patients with the susceptibility haplotype, we sequenced one patient from each of the genome scan families, and a haplotype-carrying father and his daughter for all nine exons and exon-intron boundaries of ELMOD2, but we found no mutations. Therefore, a mutation within the coding region does not explain the altered function or expression of ELMOD2 in our data set.

Intronic SNPs obviously do not change the structure of the encoded protein, but there is increasing evidence that not only coding, but also non-coding, sequences can have deleterious effects on splicing and change expression levels (Pagani et al. 2003, Pagani and Baralle 2004). This has been reported in several disease gene-mapping studies, even though the diseases-causing mechanisms remain poorly understood (Karlin et al. 2002, Laitinen et al. 2004, Pastinen and Hudson 2004). Expression levels are controlled both by cis-acting factors, such as DNA polymorphisms and methylation in the flanking DNA sequence, and trans-acting factors, such as transcription factors, that are in turn influenced by other genetic and environmental modulators (Pastinen and Hudson 2004). Identifying the epigenetic mechanisms and factors that together with the genomic variant(s) change the phenotype will be challenging.

5. Association studies on functional candidate genes CR1 and ECSOD

The CR1 gene is an important part of innate immunity in lung (Cornakoff et al. 1983, Davies et al. 1992). The polymorphism Pro1827Arg in exon 33 creates a potential cleavage site for trypsin-like proteases that can increase the shedding of receptors expressed on the cell surface (Herrera et al. 1998). An association with Pro1827Arg had been reported previously among Italian IPF patients (Zorzetto et al. 2003). According to our results the site was not polymorphic among the Finns.

The CR1 protein is composed of repeated motifs of 65-70 amino acids, and almost identical repeats of sequences occur within the coding region of the CR1 gene. For example, the lengths of both exons 33 and 41 are 228 bp, and they differ from each other by only 7 nucleotides. However, the intronic regions between exons are unique, which makes it possible to design specific amplicons for different exons. Later, in our correspondence to Zorzetto et al. (2005), it became obvious that the polymorphic site they had reported was not in exon 33 but in exon 41 (Hodgson and Laitinen 2005).

Based on unfortunately misleading information of the location of the SNP of interest, our study, in fact, did not confirm or exclude the association reported by Zorzetto et al. (2003). It became also evident that the corresponding C>G substitution in exon 33 is a monomorphic site based on our own and the Italian study (original communication II, Zorzetto et al. 2005). The polymorphism should preferably be called Pro2277Arg rather than Pro1827Arg in the future.

The ESCOD gene maps to chromosome 4 (4pter-q21) (Hendrickson et al. 1990). The ECSOD protein has high affinity for collagens and glycosaminoglycans, and this affinity is regulated by the

six amino acid motifs (Arg-Lys-Lys-Arg-Arg-Arg) within the last 14 amino acids (Sandström et al. 1992, Sandström et al. 1994, Folz and Crapo 1994). A single nucleotide change from C to G that leads to an amino acid change from arginine to glycine, Arg213Gly, (Arg-Lys-Lys-*Gly*-Arg-Arg) has been reported (Folz and Crapo 1994). The polymorphism is associated with higher levels of serum ECSOD (Sandström et al. 1994), assumed to be due to the decreased affinity to ECM (Adachi et al. 1996). The increased activity of ECSOD in the ECM is known to protect from lung injury caused by free radicals (Oury et al. 1996).

In our study the Gly213 allele was shown to be a rare variant in the population (2.5%), which is in acconcordance with observations among Swedish populations (2.2-3.8%) (Marklund et al. 1997). The relationship of this polymorphism with lung diseases has been studied among asthma patients, but no association was revealed (Kinnula et al. 2004). We found no association of the Arg213Gly polymorphism with IPF. Therefore, other factors regulating the role of ECSOD in IPF do matter.

Future challenges

Understanding the molecular genetic mechanisms in the pathogenesis of IPF gives us the opportunity to reach for effective therapies for IPF patients. Because the phenotypes of familial and sporadic IPF cannot be distinguished, familial IPF can give us unique insight into the pathogenesis of IPF. IPF is a complex disease, and it is obvious that identifying one disease-predisposing gene would give us one new tool to discover parts of the etiology of IPF. Although IPF does not differ in its clinical presentation and outcome between different populations, it is crucial that the results we have obtained among Finnish IPF patients be verified in other, and especially larger, populations. Our genome scan among Finnish families pointed to two novel positional candidates. Based on functional and expressional properties, especially ELMOD2 is proposed as a prime candidate susceptibility gene for familial IPF. Characterization of the genes, and the function and expression of both ELMOD2 and LOC152586 is our next challenge on the way to better evaluate the possible role of these genes in the etiology and pathogenesis of IPF.

CONCLUSIONS

We estimated the first nationwide prevalence of IPF according to the present diagnostic criteria. The prevalence of IPF in Finland was 16-18/100 000 inhabitants. A geographical clustering with a prevalence of 45/100 000 was observed in southern and eastern Savo.

The familial form explains 3.3-3.7% of all IPF cases in Finland. The distribution of the origins of identified FPF patients clustered remarkably to southern and eastern Savo. Ten of the families originated from two clusters close to each other in Savo and Carelia. FPF patients may share a common ancestor who introduced the disease-predisposing allele into the population.

The genome-wide scan combined with haplotype analysis with a dense marker map identified a 110 kb region on chromosome 4q31.1. The locus obtained an NPL score of 2.1, exceeding the suggestive linkage threshold. The susceptibility haplotype was shared by one third of the patients.

The susceptibility haplotype harbors two functionally poorly characterized genes, ELMOD2 and LOC152586. ELMOD2 is potentially involved in apoptosis, phagocytosis, cell engulfment, and cell migration. It is expressed in functionally relevant tissue, in lung and in fibroblasts, and its expression is significantly decreased in IPF lung compared to healthy lung. Therefore ELMOD2 becomes a prime candidate gene for familial IPF.

The Pro1827Arg of CR1 gene in exon 33 was formerly reported to be associated with IPF. The site in exon 33 appeared to be monomorphic. The polymorphism is located in exon 41, and should preferably be called Pro2277Arg. The functionally promising polymorphism Arg213Gly in ECSOD was not associated with IPF among Finnish IPF patients.

ACKNOWLEDGEMENTS

I am deeply grateful to all IPF patients and their families for participating this study.

I acknowledge Jodie Painter for revising the language of my thesis.

During these years I have also worked in the Pulmonary Clinics. When the studies did not seem to go further, my collagues have helped to remind me of the value of clinical work and to keep my self-esteem, thank you all for that. With some of them I have been able to ventilate the dualism of balancing work and preparing the thesis at the same time, with some I have made good friends, and some have given really valuable support and help to me. Thank you Aija, Annamari, Eeva-Maija, Heikki, Juho, Laura, Marjukka, and Maija, and the others.

To all my friends I want to express my thanks for sharing the sunny and rainy days, no matter whether that happened at the dinnertable or on the golf course. Thank you Paula and Elina, the three musketeers conquering the world grew into pure friendship.

I am deeply indepted to my parents, Raija and Teuvo Pakkala, who have patiently always been there for me, no matter what I have decided to do. Their love has guided my way. My brother Pekka, and his wife Kaisa, and their adodable daughters Reetta and Essi have been my joy and reminiscence of real life. I want to thank my father-in-law Stanley for refreshing conversations, and his help in reviewing the English in some of our manuscripts.

Finally, I thank my beloved husband Thomas, for love and joy and support. For all these years you have nourished us and kept our home worth living. With you I have been blessed to share the precious time.

Espoo, January 2006

REFERENCES

Adachi T, Yamada H, Yamada Y, Morihara N, Yamazaki N, Murakami T, Futenma A, Kato K, Hirano K. Substitution of glycine to arginine-213 in human extracellular superoxide dismutase impairs affinity for heparin and endothelial cell surface. Biochem J. 1996; 313: 235-239

Adamson IYR, Young L, Bowden DH. Relationships of alveolar epithelial injury and repair to the induction of pulmonary fibrosis. Am J Pathol. 1988,130: 377–383

Adelman AG, Chertkow G, Hayton RC. Familial fibrocystic pulmonary dysplasia. A detailed family study. Canada Med Ass. 1966; 95: 603-610

Akira M, Sakatani M, Ueda E. Idiopathic pulmonary fibrosis: progression of honeycombing at thin-section CT. Radiology. 1993;189: 687-91

Altmuller J, Palmer LJ, Fischer LJ, Fischer G, Scherb H, Wjst M. Genomwide scans of complex human diseases: True linkage is hard to find. Am J Hum Genet. 2001; 69: 936-950

American Thoracic Society (ATS) and European Respiratory Society (ERS). 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

American Thoracic Society. Idiopathic Pulmonary Fibrosis: Diagnosis and Treatment. International Consensus Statement. American Thoracic Society (ATS) and the European Respiratory Society (ERS) 2000; 161: 646–664

Antoniades HN, Bravo MA, Avila RE, Galanopoulos T, Neville-Golden J, Maxwell M, Selman M. Platelet-derived growth factor in idiopathic pulmonary fibrosis. J Clin Invest. 1990; 86: 1055-1064

Asumalahti K, Veal C, Laitinen T, et al. Coding haplotype analysis supports HCR as the putative susceptibility gene for psoriasis at the MHC PSORS1 locus. Hum Mol Genet. 2002; 11: 589-597

Atzori L, Chua F, Dunsmore SE, Willis D, Barbarisi M, McAnulty RJ, Laurent GJ. Attenuation of bleomysin induced pulmonary fibrosis in mice using heme oxygenase inhibitor Zn-deuteroporphyrin IX-2,4-bisethylene glycol. Thorax 2004; 59: 217-223

Baumgartner KB, Coultas DB, Stidley CA, Hunt WC, Colby TV, Waldron JA, the Collaborating Centers. Occupational and environmental risk factors for idiopathic pulmonary fibrosis: a multicenter case control study. Am J Epidemiol. 2000; 152: 307-15

Baur MP, Knapp M. Association studies in genetic epidemiology. In: Pawlowitzki I, Edwards JH, Thompson EA (eds) Genetig Mapping of Disease Genes. Academic Press 1997: 159-172

Behr J, Maier K, Degenkolb B, Krombach F, Vogelmeier C. Antioxidative and clinical effects of high-dose N-acetylcysteine in fibrosing alveolitis. Adjunctive therapy to maintainance immunosuppression. Am J Respir Crit Care Med.1997;156:1897-1901

Bensard DD, McIntyre RC, Waring BJ, Simon JS. Comparion of video thoracoscopic lung biopsy to open lung biopsy in the diagnosis of interstitial lung disease. Chest 1993; 103: 765-770.

Bitterman PB, Rennard SI, Keogh BA, et al. Familial Idiopathic Pulmonary Fibrosis. Evidence of lung inflammation in unaffected family members. N Engl J Med. 1986; 314: 1343-1347

Bjoraker JA, Ryu JH, Edwin MK, Myers JL, Tazelaar HD, Schroeder DR, Offord KP.Prognostic significance of histopathologic subsets in idiopathic pulmonary fibrosis. Am J Respir Crit Care Med. 1998;157:199-203.

Bonnani PP, Frymoyer JW, Jacox RF. A family study of idiopathic pulmonary fibrosis: a possible dysproteinemic and genetically determined disease. Am J Med. 1965; 39:411-421

Bonner JC, Rice AB, Ingram JL, Moomaw CR, Nyska A, Bradbury A, Sessoms AR, Chulada PC, Morgan DL, Zeldin DC, Langenbach R. Susceptibility of cyclooxygenase-2-deficient mice to pulmonary fibrogenesis. Am J Pathol. 2002;161:459-470

Botstein D, White RL, Skolnik MH, Davies RW. Construction of a genetic linkage map in man using restriction fraction length polymorphisms. Am J Hum Genet.1980;32:314-331

Bowler RP, Crapo JD. Oxidative stress in airways: Is there a role for extracellular superoxide dismutase? Am J Repir Crit Care Med. 2002;166:38-43

Bowler RP, Nicks M, Warnick K, Crapo JD. Role of extracellular superoxide dismutase in bleomycin-induced pulmonary fibrosis. Am J Physiol Lung Cell Mol Physiol. 2002;282: 719-726

Brody AR, Warshamana S, Liu J, Tsai S, Pociask DA, Brass DM, Schwartz D. Identifying fibrosis susceptibility genes in two strain of inbred mice. Chest. 2002;121:31

Cardon LR, Abecasis GR. Uning haplotype blocs to map human complex trait loci. Trends Genet. 2003;19:135-140

Carlsson LM, Jonsson J, Edlund T, Marklund SL. Mice lacking extracellular superoxide dismutase are more sensitive to hyperoxia. Proc Natl Acad Sci. 1995;92:6264-6268

Chambers RC, Leoni P, Kaminski N, Laurent GJ, Heller RA. Global expression of profiling of fibroblast responses to transforming growth factor-b1 reveals the induction of inhibitor of differentiation-1 and provides evidence of smooth muscle cell phenotypic switching. Am J Pathol. 2003;162:533-546

Chilosi M, Poletti V, Murer B, Lestani M, Cancellieri A, Montagna L, Piccoli P, Cangi G, Semenzato G, Doglioni C. Abnormal re-epithelization and lung remodelling in idioptahic pulmonary fibrosis: the role of deltaN-p63. Lab Invest. 2002;82:1335-1345

Cho HY, Jedlicka AE, Reddy SPM, Kensler TW, Yamamoto M, Zhang LY, Kleeberger SR. Role of NRF2 protection against hyperoxic lung injury in mice. Am J Cell Mol Biol. 2002;26:175-182

Chuang-Tsai S, Sisson TH, Hattori N, Tsai CG, Subbotina NM, Hansen KE, Simon RH. Reduction in fibrotic tissue formation in mice genetically deficient in plasminogen activator inhibitor-1. Am J Pathol. 2003;163:445-452

Collard HR, King TE Jr, Bartelson BB, Vourlekis JS, Schwarz MI, Brown KK. Changes in clinical and physiologic variables predict survival in idiopathic pulmonary fibrosis. Am J Respir Crit Care Med. 2003;168:538 - 542

Cornacoff J B, Herbert L A, Smead M E, VanAman D J, Birmingham D J, Waxman F J. Primate erythrocyte immune complexes clearing mechanism. J Clin Invest. 1983;71:236

Costabel U, King TE. International consensus statement on idiopathic pulmonary fibrosis. Eur Respir J. 2001;17:163–7

Coultas DB, Zumwalt RE, Black WC, et al. The Epidemiology of Interstitial Lung Disease. Am J Respir Crit Care Med. 1994; 150: 967-972

Daly MJ, Rioux JD, Schaffner SF, Hudson TJ, Lander ES. High-resolution haplotype structure in the human genome. Nature Genet. 2001;29:229-232

Davies KA, Peters AM, Benyon HLC, Walport MJ. Immune complex processing in patients with systemic lupus erythemetous: in vivo imaging and clearance studies. J Clin Invest. 1992;90:2075

Dawson JK, Fewins HE, Desmond J, Lynch MP, Graham DR. Fibrosing alveolitis in patients with rheumatoid arthritis as assessed by high resolution computed tomography, chest radiography, and pulmonary function tests. Thorax. 2001; 56:622-627

de la Chapelle A, Wright FA. Linkage disequilibrium mapping in isolated populations: The example of Finland revisited. Proc Natl Acad Sci.1998;95:12416-12523

DHEW publication [NIH]. Respiratory Diseases Task Force: Report on Problems, Research, Approaches, and Needs, Oct 1972: 73-432

duBois RM, Wells AU. Cryptogenic fibrosing alveolitis/idiopathic pulmonary fibrosis. Eur Respir J. 2001;32:43-55

Dupuy PM, Shore SA, Drazen JM, Frostell C, Hill WA, Zapol WM. Bronchodilator action of inhaled nitric oxide in guinea pigs. J Clin Invest.1992;90:421-428

Ewens WJ, Spielman RS. Locating genes by linkage and association. Theor Popup Biol. 2001;60:135-139

Falfan-Valencia R, Camarena A, Juarez A, Becerril C, Montano M, Cisneros J, Mendoza F, Granados J, Pardo A, Selman M.Major histocompatibility complex and alveolar epithelial apoptosis in idiopathic pulmonary fibrosis. Hum Genet. 2005;30:1-10

Fattman CL, Schaefer LM, Oury TD. Extracellular superoxide dismutase in biology and medicine. Free Radic Biol Med. 2003;35:236-256

Fearon DT, Wong WW. Complement ligand-receptor interactions that mediate biological responses. Annu Rev Immunol. 1983;1:243-71

Fellrath JM, duBois R. Idiopathic pulmonary fibrosis/cryptogenic fibrosing alveolitis. Clin Exp Med. 2003;3:65-83

Flaherty KR, Travis WD, Colby TV, Toews GB, Kazerooni EA, Gross BH, Jain A, Strawderman RL, Flint A, Lynch JP, Martinez FJ. Histopathologic variability in usual and nonspecific interstitial pneumonias. Am J Respir Crit Care Med. 2001;164:1722-7

Folz RJ, Abushamaa AM, Suliman HB. Extracellular superoxide dismutase in the airways of transgenic mice reduces inflammation and attenuates lung toxicity following hyperoxia. J Clin Invest. 1999;103:1055-1066

Folz RJ, Crapo JD. Extracellular superoxide dismutase (SOD3): tissue-specific expression, genomic characterization, and computer-assisted sequence analysis of the human ECSOD gene. Genomics.1994;22:162-171

Fujimoto H, Gabazza E, Hataki O, Yuda H, et al. Thrombin-activatable fibrinolysis iInhibitor and protein C inhibitor in interstitial lung disease. Am J Respir Crit Care Med. 2003;167:1687-1694

Fulmer JD, Sposovska MS, von Gal ER, Crystal RG, Mittal KK. Distribution of HLA antigens in idiopatthic pulmonary fibrosis. Am Rev Respir Dis. 1978;118:141-146

Gabriel SB, Schaffner SF, Nguyen H, et al. The structure of haplotype diversity revealed by high-resolution scanning of human genome. Science. 2002;296:2225-2229

Geddes DM, Webley M, Brewerton DA, Turton CW, Turner-Warwick M, Murphy AH, Ward AM. Alfa-1-antitrypsin phenotypes in fibrosing alveolitis and rheumatoid arthritis. Lancet.1977;2:1049-1051

Ghio AJ, Suliman HB, Carter JD, Abushamaa AM, Folz RJ. Overexpression of extracellular superoxide dismutase decreases lung injury after exposure to oil fly ash. Am J Physiol Lung Cell Mol Physiol. 2002;283:L211-L218

Giaid A, Michel RP, Stewart DJ, Sheppard M, Corrin B, Hamid Q. Expression of endothelin-1 in lung of patients with cryptogenic fibrosing alveolitis. Lancet.1993;341:1550–1554

Gross TJ, Hunninghake GW. Idiopathic pulmonary fibrosis. N Engl J Med. 2001;345:517-525

Göring HHH, Terwilliger JD, Blangero J. Large upward bias in estimation of locus-spesific effects from genomewide scans. Am J Hum Genet. 2001;69:1357-1369

Hamman L, Rich AR. Acute diffuse interstitial fibrosis of the lungs. Bull John Hopkins Hospital.1944;74:177-204

Hartl DL, Campbell RB. Allelic multiplicity in simple Mendelian disorders. Am J Hum Genet. 1982;34:866-873

Haston CK, Travis EL. Murine susceptibility to radiation-induced pulmonary fibrosis influenced by a genetic factor implicated in susceptibility to bleomycin -induced pulmonary fibrosis. Cancer Res.1997;57:5286-5291

Haston CK, Zhou X, Gumbiner-Russo L, Irani R, dejournett R, Gu X, Weil M, Amos CI,Travis EL. Universal and Radiation-induced loci influence murine susceptibility to radiation-induced pulmonary fibrosis. Cancer Res.2002;62:3782-3788

Hendrickson DJ, Fisher JH, Jones C, Ho YS. Regional localization of human extracellular superoxide dismutase gene to 4pter-q21.Genomics.1990;8:736-738

Herrera A H, Xiang L, Martin S G, Lewis J, Wilson J G. Analysis of Complement Receptor Type 1 (CR1) Expression on erythrocytes and of CR1 Allelic Markers in Caucasian and African American Populations. Clin Immunol and Immunopathol. 1998; 87:176-183

Hocher B, Schwarz A, Fagan KA, Thöne-Reineke C, El-Hag K, Kusserow H, Elitok S, Bauer C, Neumayer HH, Rodman DM, Theuring F. Pulmonary fibrosis and chronic lung inflammation in ET-1 transgenic mice. Am J Respir Cell Mol Biol. 2000;23:19-26

Hodgson U, Laitinen T. Complement receptor 1 gene C5507G polymorphism. Respir Med. 2005, in press

Holgate ST, Haslam P, Turner-Warwick M.The significance of antinuclear and DNA antibodies in cryptogenic fibrosing alveolitis.Thorax.1983;38:67-70

Howell DC, Goldsack NR, Marshall RP, McAnulty RJ, Starke R, Purdy G, Laurent GJ, Chambers RC. Direct thrombin inhibition reduces lung collagen, accumulation, and connective tissue growth factor mRNA levels in bleomycin-induced pulmonary fibrosis. Am J Pathol. 2001;159:1383-1395

Hubbard R, Lewis S, Richards K, Johnston I, Britton J. Occupational exposure to metal or wood dust and aetiology of cryptogenic fibrosing alveolitis. Lancet. 1996; 347: 284-289

Hubbard R, Venn A, Smith C, Cooper M, Johnston I, Britton J. Exposure to commonly prescribed drugs and the etiology of cryptogenic fibrosing alveolitis: a case-control study. Am J Respir Crit Care Med.1998; 157: 743-747

Hutyrova B, Pantelidis P, Drabek JI, Zurkova M, Kolek V, Lenhart K, Welsh KI, duBois RM, Petek M. Interleukin-1 gene cluster polymorphisms in sarcoidosis and idiopathic pulmonary fibrosis. Am J Respir Crit Care Med. 2002;165:148-151

Höglund P, Sistonen P, Norio R, Holmberg C, Dimberg A, Gustavson KH, de la Chapelle A, Kere J. Fine mapping of the congenital chloride diarrhea gene by linkage disequilibrium. Am J Hum Genet. 1995;57:95-102

Ikawa M, Nakanishi T, Yamada S, Wada I, Kominami K, Tanaka H, Nozaki M, Nishimune Y, Okabe M. Calmegin is required for fertilin alpha/beta heterodimerization and sperm fertility. Dev Biol.2001; 240: 254-261,

Ikawa M, Wada I, Kominami K, Watanabe D, Toshimori K, Nishimune Y, Okabe M. The putative chaperone calmegin is required for sperm fertility. Nature.1997;387: 607-611,

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

Iwai K, Mori T, Yamada N, Yamaguchi M, Hosoda Y. Idiopathic pulmonary fibrosis: epidemiologic approaches to occupational exposure. Am J Respir Crit Care Med. 1994;150: 670-675

Javaheri S, Lederer DH, Pella JA, et al. Idiopathic Pulmonary Fibrosis in Monozygotic Twins. Chest 1980; 78: 591-594

Johkoh T, Ikezoe J, Kohno N, Takeuchi H, Yamagami N, Tomiyama H, Kondoh S, Kido J, Arisawa J, Kozuka T. High-resolution CT and pulmonary function tests in collagen vascular disease: comparison with idiopathic pulmonary fibrosis. Eur J Radiol. 1994;18: 113-121

Johnston I, Britton J, Kinnear W, Logan R. Rising mortality from cryptogenic fibrosing alveolitis. Br Med J.1990; 301: 1017-1021

Johnston ID, Prescott RJ, Chalmers JC, Rudd RM. British Thoracic Society study of cryptogenic fibrosing alveolitis: current presentation and initial management. Fibrosing Alveolitis Subcommittee of the Research Committee of the British Thoracic Society. Thorax. 1997;52:38-44

Kaminski N, Allard JD, Pittet JF, Zuo F, Griffiths MJD, Morris D, Huang X, Sheppard D, Heller RA. Global analysis of gene expression in pulmonary fibrosis reveals distinct programs regulating lung inflammation and fibrosis. Proc Nat Am Sci. 2000; 97: 1778-83

Kapanci Y, Desmouliere A, Pache JC, Redard M, Gabbiani G. Cytoskeletal protein modulation in pulmonary alveolar myofibroblasts during idiopathic pulmonary fibrosis. Possible role of transforming growth factor beta and tumor necrosis factor alpha. Am J Respir Crit Care Med. 1995;152:2163–2169

Katzenstein AL, Myers JL.Idiopathic pulmonary fibrosis:clinical relevance of pathologic classification. Am J Respir Crit Care Med. 1998;157:1301-1305

Katzenstein AL, Myers JL. Nonspecific interstitial pneumonia and the other idiopathic interstitial pneumonias: Classification and diagnostic criteria. Am J Surg Pathol. 2000;224:19-33.

Katzenstein AL, Zisman DA, Litzky LA, Nguyen BT, Kotloff RM. Usual interstitial pneumonia. Histological study of biopsy and explant specimens. Am J Surg Pathol. 2002; 26:1567-1577

Keane MP, Arenberg DA, Lynch JP, et al. The CXC chemokines, IL-8 and ILP10, regulate angiogenic activity in idiopathic pulmonary fibrosis. J Immunol.1997:1437-1443

Keane MP, Belperio JA, Burdick MD, Lynch JP, Fishbein MC, Srieter RM. ENA-78 is an important angiogenic factor in idiopathic pulmonary fibrosis. Am J Respir Crit Care Med. 2001;164:2239-2242

Keogh BA, Crystal RG.Alveolitis: the key to the interstitial lung disorders.Thorax.1982;37:1-10

Kere J. Human population genetics: lessons from Finland. Annu Rev Genomics Hum Genet. 2001;2:103-28

Kim KK, Flaherty KR, Long Q, Hattori N, Sisson TH, Travis WD, Martinez FJ, Murray S, Simon RH. A plasminogen activator inhibitor-1 promoter polymorphism and idiopathic interstitial pneumonia. Mol Med. 2003;9:52-56

King TE, Schwarz MI, Brown K, Tooze JA, Colby TV, Waldron JA, Flint A, Thurbeck W, Cherniack RM. Idiopathic pulmonary fibrosis. Relationship between histopathologic features and mortality. Am J Respir Crit Care Med. 2001;164:1025-1032

Kinnula V, Crapo JD, Raivio K. Generation and disposal of reactive oxygen metabolites in the lung. Lab Invest.1995;73:3-19

Kinnula V, Crapo JD. Superoxide dismutases in the lung and human lung disease. Am J Respir Crit Care Med.2003;167:1600-1619

Kinnula VL, Fattman CL, Tan RJ, Oury TD. Oxidative stress in pulmonary fibrosis: a possible role for redox modulatory therapy. Am J Respir Crit Care Med. 2005;172:417-22

Kinnula V, Tukiainen P. Uusia tuulia keuhkofibroosien luokittelussa, diagnostiikassa ja hoitovaihtoehdoissa. Duodecim. 2004;120:1228-1235

Kinnula VL, Lehtonen S, Koistinen P, Kakko S, Savolainen M, Kere J, Ollikainen V, Laitinen T. Two functional variants of the superoxide dismutase genes in Finnish families with asthma. Thorax. 2004;59:116-119

Kittles RA, Perola M, Peltonen L, Bergen AW, Aragon RA, Virkkunen M, Linnoila M, Goldman D, Long JC. Dual origins of Finns revealed by Y chromosome haplotype variation: Am J Hum Genet.1998;62:1171-1179

Koch B. Familial fibrocystic pulmonary dysplasia: observations in one family. Canada Med Ass. 1965;92:801-808

Kolek V. Epidemiology of Cryptogenic Fibrosing Alveolitis in Moravia and Silesia. Acta Univ Palacki Olomuc Fac Med. 1994; 137: 49-50

Koskenmies S, Lahermo P, Julkunen H, Ollikainen V, Kere J, Widen E. Linkage mapping of systemic lupus erythematotus (SLE) in Finnish families multiply affected by SLE. J Med Genet. 2004;41:e2-5

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

Kruglyak L. Prospects for whole-genome linkage disequilibrium mapping of common disease genes. Nature Genet. 1999;22:139-144

Kruglyak L, Daly MJ, Reeve-Daly MP, Lander ES. Parametric and nonparametric linkage analysis: a unified multipoint approach. Am J Hum Genet. 1996;58:1347-1363

Kruglyak L, Lander ES. A nonparametric approach for mapping quantative trait loci. Genetics. 1995;139:1421-1428

Lahermo P, Savontaus ML, Sistonen P, et al. Y chromosomal polymorphisms reveal founding lineages in the Finns and the Saami. Eur J Hum Genet. 1999;7:447-458

Lahiri DK, Nurnberg JI. A rapid non-enzymatic method for the preparation of HMW DNA from blood for RFLP studies. Nucleid Acids Res. 1991;19:5444

Laitinen T, Daly MJ, Rioux JD, et al. A susceptibility locus for asthma- related traits on chromosome 7 revealed by genome-wide scan in a founder population. Nat Genet. 2001;28:87-91

Laitinen T, Polvi A, Rydman P, et al. Characterization of a common susceptibility locus for astma-related traits. Science. 2004; 304:300-304

Lander E, Kruglyak L. Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. Nat Genet. 1995;11:241-247

Lander E, Schork NJ. Genetic dissection of complex traits. Science. 1994;265:2037-2048

Latsi P, Pantelidis P, Vassilakis D, Sato H, Welsh KI, duBois R. Analysis of IL-12 p40 subunit gene and IFN-g G5644A polymorphisms in idiopathic pulmonary fibrosis. Respir Research. 2003;4:1-5

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

Lee H-L, Ryu JH, Wittmer MH, Hartman TE, Lymp JF, Tazelaar HD, Limper AH. Familial idiopathic pulmonary fibrosis- Clinical features and outcome. Chest. 2005;127:2034-2041

Libby DM, Gibofsky A, Fotino M, Waters SJ, Smith JP. Immunogenetic and clinical findings in idiopathic pulmonary fibrosis. Am Rev Respir Dis. 1983;127:618-622

Liebow AA, Carrington CB. The interstitial pneumonias. In:Simon M, Potchen EJ, LeMay M.Frontiers of pulmonary radiology. 1st ed. Grune and Stratton, New York. 1969:102-141

Loyd J. Pulmonary fibrosis in families. Am J Respir Cell Mol Biol.2003;29:S47-S50

MacMillan JM. Familial pulmonary fibrosis. Dis Chest.1951;20:426-436

Madtes DK, Elston AL, Hackman RC, Dunn AR, Clark JG. Transforming growth factor-a deficiency reduces pulmonary fibrosis in transgenic mice. Am J Respir Cell Mol Biol. 1999;20:924-934

Mannino DM, Etzel RA, Parrish RG. Pulmonary fibrosis deaths in the United States, 1979-1991. An analysis of multiple-cause mortality data. Am J Respir Crit Care Med. 1996;153:1548-52

Mapel DW, Samet JM, Coultas DB. Corticosteroids and the treatment of idiopathic pulmonary fibrosis. Past, present, and future. Chest. 1996;110:1058-1067

Marklund SL. Extracellular superoxide dismutase in human tissues and human cell lines. J Clin Invest.1984;74:1398-1403

Marklund SL, Nilsson P, Israelsson K, Schampi I, Peltonen M, Asplund K. Two variants of extracellular superoxide dismutase: relationship to cardiovascular risk factors in an unselected middle-aged population. J Intern Med. 1997;242:5-14

Marney A, Lane KB, Phillips John III, Riley DJ, Loyd J. Idiopathic pulmonary fibrosis can be autosomal dominant trait in some families. Chest. 2001;120:56S

Marshall RP, McNaulty RJ, Laurent GJ. The Pathogenesis of Pulmonary Fibrosis: Is There a Fibrosis Gene? Int J Biochem Cell Biol. 1997; 29: 107-120

Marshall RP, Puddicombe A, Cookson WOC, et al. Adult Familial Cryptogenic Fibrosing alveolitis in the United Kingdom. Thorax. 2000; 55: 143-146

Mason RJ, Schwarz MI, Hunninghake GW, Musson RA. NHLBI Workshop Summary. Pharmacological therapy for idiopathic pulmonary fibrosis. Past, present, and future. Am J Respir Crit Care Med. 1999;160:1771-1777

Miyake Y, Sasaki S, Yokoyama T, Chida K, Azuma A, Suda T, Kudoh S, Sakamoto N, Okamoto K, Kobashi G, Washio M, Inaba Y, Tanaka H. Occupational and environmental factors and idiopathic pulmonary fibrosis in Japan. Ann Occup Hyg. 2005;49:259-65

Munger JS, Huang X, Kawakatsu H, Griffits MJ, Dalton SL, Wu J, Pittet JF, Kaminki N, Garat C, Matthay MA, Rifkin DB, Sheppard D. The integrin alfavbeta6 binds and activates latent TGF beta 1: a mechanism for regulating pulmonary inflammation and fibrosis. Cell. 1999;96:319-328

Musk AW, Zilko PJ, Manners P, Kay PH, Kamboh MI. Genetic studies in familial fibrosing alveolitis. Possible linkage with immunoglobulin allotypes (Gm). Chest.1986;89:206-210

Nakao A, Fujii M, Matsumura R, Kumano K, Saito Y, Miyazono K, Iwamoto I. Transient gene transfer and expression of Smad7 prevents bleomycin-induced lung fibrosis in mice. J Invest. 1999;104:5-11

Nevanlinna HR. Finnish population structure and hereditary diseases. Duodecim. 1972;881:4-14

Nicholson AG, Colby TV, duBois RM, Hansell DM, Wells AU. the prognostic significance of the histologic pattern of interstitial pneumonia in patients presenting with the clinical entity of cryptogenic fibrosing alveolitis. Am J Respir Crit Care Med. 2000;162:2213-2217

Nicholson AG, Colby TV, Wells AU. Histopathological approach to patterns of interstitial pneumonia patient with connective tissue disease. Sarcoidosis Vasc Diffuse Lung Dis. 2002;19:10-17

Nogee L, Dunbar A, Wert S, Askin F, Hamvas A, Whitsett JA. Mutations in the surfactant protein C gene associated with interstitial lung disease. Chest. 2002;121:20–21S

Nogee LM, Dunbar AE, 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

Norio R. Heredity in the congenital nephrotic syndrome; a genetic study of 57 Finnish families with a review of reported cases. Ann Paediatr Fenn.1966;12:Suppl 27

Norio R. Finnish Disease Heritage I: Characteristics, causes, background. Hum Genet. 2003;112:441-456

O'Connell JR, Weeks DE. PedCheck: a program for identification of genotype incompatibilities in linkage analysis. Am J Hum Genet. 1998 ;63:259-66

Ott J. Human genetics: Complex traits on the map. Nature. 1996;379:772-773

Ottonello L, Dapino P, Pastorino G, Dallegri F, Scchetti C. Neutrophil dysfunction and increased susceptibility to infection. Eur J Clin Invest. 1995;25:687-692

Oury TD, Crapo JD, Valnickova Z, Enghild JJ. Human extracellular superoxide dismutase is a tetra composed of two disulphide-linked dimers: a simplified,high-yield purificationof extracellular superoxide dismutase.Biochem J.1996;317:51-57

Oury TD, Day BJ, Crapo JD. Extracellular superoxide dismutase in vessels and airways in humans and baboons. Free Radic Biol Med. 1996;20:957-965

Pajukanta P, Lilja H, Sinsheimer JS, et al. Familial combined hyperlipidemia is associated with upstream transcriptional factor 1 (USF1). Nature Genet. 2004;36.371-376

Palmer LJ, Cardon LR. Shaking the tree: mapping complex disease genes with linkage disequilibrium. Lancet. 2005;366:1223-1234

Pan LH, Yamauchi K, Uzuki M, Nakanishi T, Takigawa M, Inoue H, Sawai T. Type II alveolar epithelial cells and interstitial fibroblasts express connective tissue growth factor in IPF. Eur Respir J. 2001;17:1220–1227

Pantelidis P, Fanning GC, Wells AU, Welsh KI, duBois RM. Analysis of tumor necrosis factor alfa, lymphotoxin alfa, tumor necrosis receptor II, and interleukin-6 polymorphisms in patients with idiopathic pulmonary fibrosis. Am J Respir Crit Care Med. 2001;163:1432-1426

Pardo A, Selman M. Idiopathic pulmonary fibrosis: new insights in its pathogenesis. Int J Biochem Cell Biol. 2002;34:1534-1538

Peabody JW, Peabody JW Jr,Hayes EW, Hayes EW Jr. idiopathic pulmonary fibrosis: Its occurence in identical twin sisters. Dis Chest. 1950;18:330-343

Peltomäki P. et al. Genetic mapping of a locus predisposing to human colorectal cancer. Science.1993;260:810-812

Peltonen L, Jalanko A, Varilo T. Molecular genetics of the Finnish disease heritage. Hum Mol Gen.1999;8:1913-1923

Peltonen L, Palotie A, Lange K. Use of population isolates for mapping complex traits. Nature Rev. 2000;1:182-190

Peters-Golden M, Bailie M, Marshall T, Wilke C, Phan SH, Toews GB, Moore BB. Protection from pulmonary fibrosis in leucotriene-deficiet mice. Am J Respir Crit Care Med. 2002;165:229-235

Piippo K, Laitinen P, Swan H, et al. Homozygosity for a HERG potassium channel mutation causes a severe form of long QT syndrome: identification of an apparent founder mutation in the Finns.
J Am Coll Cardiol. 2000;35:1919-25

Pääkkö P, Kaarteenaho-Wiik R, Pollänen R, Soini Y. Tenscin mRNA expression at the foci of recent injury in usual interstitial pneumonia. Am J Crit Care Med. 2000;161:967-972

Raghu G, Johnson WC, Lockhart D, Mageto Y. Treatment of pulmonary fibrosis with a new antifibrotic agent, pirfenidone:results of a prospective, open-label Phase II study. Am J Respir Crit Care Med. 1999;159:1061-1069

Ramos C, Montano M, Garcia-Alvarez J, Ruiz V, Uhal BD, Selman M, Pardo A. 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

Reich DE, Lander ES. On the allelic spectrum of human disease. Trends Genet. 2001;17:502-510

Riha R, Yang I, Rabnott GC, Tunnicliffe A, Fong K, Zimmerman P. Cytokine gene polymorphisms in idiopathic pulmonary fibrosis. Int Med J. 2004;34:126-129

Rioux JD, Daly MJ, Silverberg MS, et al. Hierarchical linkage disequilibrium mapping of a susceptibility gene for Crohn`s disease on the cytokine clusted on chromosome 5. Nature Genet. 2001;29:223-228

Risch N. Searching for genetic determinants in the new millennium. Nature. 2000;405:847-856

Risch N. Implications of multilocus inheritance for gene-disease association studies. Theor Popul Biol. 2001;60:215-220

Ruiz V, Ordonez RM, Berumen J, Ramirez R, Uhal B, Becerril C, Pardo A, Selman M. Unbalanced collagenases/TIMP-1 expression and epithelial apoptosis in experimental lung fibrosis. Am J Physiol Lung Mol Physiol. 2003;285:L1026-1036

Ryu JH, Colby TV, Hartman TE, Vassallo R. Smoking-related interstitial lung diseases: a concise review. Eur Respir J.2001; 17: 122-13

Sandström J, Carlsson L, Marklund SL, Edlund T. The heparin-binding domain of extracellular superoxide C and formation of variants with reduced heparin affinity. J Biol Chem. 1992;267:18205-18209

Sandström J, Nilsson P, Karlsson K, Marklund SL. 10-fold increase in human plasma extracellular superoxide dismutase content caused by a heparin-binding domain. J Biol Chem. 1994;269:19163-19166

Sawcer S, Jones HB, Judge D, Visser F, Compston A, Goodfellow PN, Clayton D. Empirical genome-wide significance levels established by whole genome simulations. Genet Epidemiol. 1997;14:223-229

Scadding JC, Hinson KFW. Diffuse fibrosing alveolitis (diffuse interstitial fibrosis of the lungs):correlation of histology at biopsy with prognosis.Thorax.1967;22:291-304

Schwartz DA, Helmers RA, Galvin JR, Van Fossen DS, Frees KL, Dayton CS, Burmeister LF, Hunninghake GW. Determinants of survival in idiopathic pulmonary fibrosis. Am J Respir Crit Care Med. 1994;149: 450-454

Scott J, Johnston I, Britton J. What causes cryptogenic fibrosing alveolitis? A case-control study of environmental exposure to dust. Br Med J.1990; 301:1015-1017

Selman M, King TE, Pardo A. Idiopathic pulmonary fibrosis: Prevaling and evolving hypotheses about its pathogenesis and implications for therapy. Ann Int Med. 2001;134:136–151

Selman M, Montano M, Ramos C, Chapela R. Concentration, biosynthesis and degradation of collagen in idiopathic pulmonary fibrosis. Thorax. 1986;41:355-359

Selman M, Thannickal VJ, Pardo A, Zisman DA, Martinez FJ, Lynch JP 3rd. Idiopathic pulmonary fibrosis: pathogenesis and therapeutic approaches. Drugs. 2004;64:405-30

Shi-wen X, Howat SL, Renzoni EA, et al. J Biol Chem. 2004,279:23098-23103

Siegel I, Liu TL, Gleicher N. The red-cell immune system. Lancet. 1981;2:556-559

Sihvo E, Salo J. Torakoskopia. Duodecim. 2001;117:1630–1635

Simler NR, Brenchley PE, Horrocks AW, Greaves SM, Hasleton PS, Egan JJ. Angiogenic cytokines in patients with idiopathic interstitial pneumonia. Thorax. 2004;59:581-585

Spielman RS, Mcinnis RE, Ewens WJ. Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). Am J Hum Genet. 1993;52:506-516

Steele M, Speer MC, Loyd JE, Brown KK, Herron A, Slifer SH, Burch LH, Wahidi MM, Phillips JA III, Sporn TA, McAdams HP, Schwarz MI, Schwartz DA. the clinical and pathological features of familial interstitial pneumonia (FIP). Am J Respir Crit Care Med. 2005;172:1146-52.

Stralin P, Marklund SL. Multiple cytokines regulate the expression of extracellular superoxide dismutase in human vascular smooth muscle cells. Atherosclerosis. 2000;151:433-441

Tabor HK, Rissch NJ, Myers RM. Candidate-gene approaches for studying complex genetic traits:practical considerations. Nature Rev. 2002;3:1-7

Terwilliger JD, Ott J. In: Handbook of human genetic linkage. John Hopkins University Press, Baltimore, 1994

Thannickal VJ, Fanburg BL.Reactive oxygen species in cell signalling. Am J Physiol Lung Cell Mol Physiol. 2000;279:1005-1028

Thomas AQ, lane K, Phillips John III, et al. Heterozygosity for a surfactant protein C gene mutation associated with usual interstitial pneumonitis and cellular nonspesific interstitial pneumonitis in one kindred. Am J Respir Crit Care Med. 2002; 165:1322-1328

Thomeer MJ, Costabe U, Rizzato G, Poletti V, Demedts M. Comparison of registries of interstitial lung diseases in three European countries. Eur Respir J. 2001;.32:114-118

Tobin RW, Pope CE II, Pellegrini CA, Emond MJ, Sillery JIM, Raghu G. Increased prevalence of gastroesophageal reflux in patients with idiopathic pulmonary fibrosis. Am J Respir Crit Care Med. 1998;158:1804-1806

Tukiainen P, Taskinen E, Holsti P, Korhola O, Valle M. Prognosis of cryptogenic fibrosing alveolitis. Thorax. 1983;38:349-355.

Tung KT, Wells AU, Rubens MB, Kirk JM, du Bois RM, Hansell DM. 1993. Accuracy of the typical computed tomographic appearances of fibrosing alveolitis. Thorax. 48: 334-338

Turner-Warwick M, Burrows B, Johnson A.Cryptogenic fibrosing alveolitis: response to corticosteroid treatment and its effect on survival. Thorax. 1980;35:593-599

Uchiyama B, Matsubara N, Nukiwa T, et al. Intensive Interview and Examination of Family Members Revealed More than 30% Incidence of Familial Clustering of Pulmonary Fibrosis. Eur Respir J. 1997; 10 (Suppl 25): 141-142

Uhal BD, Gidea C, Bargout R, et al. Captopril inhibits apoptosis in human lung epithelial cells. A potential antifibrotic mechanism. Am J Physiol. 1998; 275:1013-1017

Wallace WAH, Howie SEM. Upregulation of tenascin and TGF production in a type II alveolar epithelial cell line by antibody against a pulmonary auto-antigen. J Pathol.2001;195:251-256

Van Valkenburgh H, Shern JF, Sharer JD, Zhu X, Kahn RA. ADP-ribosylation factors (ARFs) and ARF-like 1 (ARL1) have both specific and shared effectors: characterizing ARL1-binding proteins. J Biol Chem. 2001;276:22826-22837

Wang DG et al. Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. Science. 1998;280:1077-1082

Varilo T, Laan M, Hovatta I, Wiebe V, Terwilliger JD, Peltonen L. Linkage disequilibrium in isolated populations: Finland and a young sub-population of Kuusamo. Eur J Hum Gen. 2000;8:604-612

Varilo T, Savukoski M, Norio R, Santavuori P, Peltonen L, Järvelä I. The age of human mutation:geneological and linkage disequilibrium analysis of the CLN5 mutation in the Finnish population. Am J Hum Genet.1996;58:506-512

Warshamana GS, Pociask DA, Sime P, Schwartz DA, Brody AR. Susceptibility to asbestos-induced and transforming growth factor-b1-induced fibroproliferative lung disease in two strains in mice. Am J Respir Cell Mol Biol. 2002;27:705-713

Weber JL, May PE. Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. Am J Hum Genet.1989;44:388-396

Wells AU, Hansell DM, duBois RM. Interstitial lung diseases in the collagen vasular diseases. Semin Repir Med. 1993;14.333-343

Vergnon JM, Vincent M, de The G, Mornex JF, Weynants P, Brune J.Cryptogenic fibrosing alveolitis and Epstein-Barr virus: an association? Lancet. 1984;2:768-771

Whyte M, Hubbard R, Meliconi R, et al. Increased risk of fibrosing alveolitis associated with interleukin-1 receptor antagonist and tumor necrosis factor-a gene polymorphisms. Am J Respir Crit Care Med. 2000;162:755-758

Wiltshire S, Cardon LR, McCArthy MI. Evaluating results od genomewide linkage scans of complex traits by locus counting. Am J Hum Genet. 2002;71:1175-1182

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;283:1-11

Wright AF, Carothers AD, Pirastu M. Population choice in mapping genes for complex diseases. Nat Genet. 1999;23:397-404

Vuorio AF, Turtola H, Piilahti KM, Repo P, Kontula K. Familial hypercholesterolemia in the Finnish North Karelia. Arterioscler Thromb. 1997;17:3127-3138

Vyse TJ, Todd JA. Genetic analysis of autoimmune disease. Cell. 1996;85:311-318

Ziesche R, Hofbauer E, Wittman K, Petkov V, Block LH. A preliminary study on long-term treatment with interferon g-1b and low-dose prednisolone in patients with idiopathic pulmonary fibrosis. N Engl J Med. 1999;341:1264-1269

Zorzetto M, Bombieri C, Ferrarotti I, Medaglia S, Agostini C, Tinelli C, Malerba G, Carrabino N, Beretta A, Casali L, Pozzi E, Pignatti P F, Semenzato G, Cuccia M C, Luisetti M. Complement Receptor 1 Gene Polymorphisms in Sarcoidosis. Am J Respir Cell Mol Biol. 2002;27:17-23

Zorzetto M, Ferrarotti I, Trisolini R, Agli L L, Scabini R, Novo M, De Silvestri A, Patelli M, Martinetti M, Cuccia M C, Poletti V, Pozzi E, Luisetti M. Complement Receptor 1 Gene Polymorphisms Are Associated with Idiopathic Pulmonary Fibrosis. Am J Respir Crit Care Med. 2003;168: 330-334

Zorzetto M, Ferrarotti I, Campo I, Luisetti M.Complement receptor 1 gene polymorphism in Finland.Respir Med. 2005, in press