Idiopathic pulmonary fibrosis biomarkers: clinical utility and a way of understanding disease pathogenesis

Matthew Flynn
Elisabeth S Baker
Daniel J Kass
Dorothy P and Richard P Simmons
Center for Interstitial Lung Disease, Division of Pulmonary, Allergy, and Critical Care Medicine, University of Pittsburgh, Pittsburgh, PA, USA

Correspondence: Daniel J Kass
Dorothy P and Richard P Simmons
Center for Interstitial Lung Disease, Division of Pulmonary, Allergy, and Critical Care Medicine, University of Pittsburgh, 3459 Fifth Avenue, NW628, Pittsburgh, PA 15213, USA
Tel +1 412 624 7444
Email kassd2@upmc.edu

Abstract: Idiopathic pulmonary fibrosis (IPF) is a typically fatal disease that remains incompletely understood despite intense study and the arrival of drugs that may alter the natural history of the disease. Rendering an accurate diagnosis and predicting prognosis remain challenging problems to clinicians. One potential solution to these clinical problems is the identification of IPF biomarkers, easily measured factors that can be employed to predict clinical behavior. Candidate biomarkers have been identified by research in the laboratory on potential culprit cells or genes that may contribute to the pathogenesis of IPF. In this review, we present the current data on a number of well-studied IPF biomarker candidates and their potential role in the pathogenesis of disease. We also establish a framework for evaluating utility of incorporating these IPF biomarkers into clinical practice.

Keywords: idiopathic pulmonary fibrosis, usual interstitial pneumonia, biomarker, matrix metalloproteinases

Introduction

Idiopathic pulmonary fibrosis (IPF) is a chronic and progressive scarring of the lung parenchyma characterized histologically by the usual interstitial pneumonia (UIP) pattern – a disease limited to the lungs and lacking a well-defined etiology.1 IPF has remained a rare disease, occurring predominantly in older men. Recent epidemiologic studies suggest an increasing incidence, possibly as high as 94 per 100,000, using refined definitions of disease.2 The mechanism of disease remains unclear. Recent theories have moved away from the long-standing paradigm of chronic inflammation to one that favors recurrent alveolar epithelial cell injury and aberrant repair.3 The interplay of pro-fibrotic cell types and the mediators associated with their recruitment, proliferation, and differentiation is presently the subject of intensive research.4 Another elusive feature of IPF is its heterogeneous natural history: many patients demonstrate a stable or slow decline, while for others, the decline is precipitous or fraught with the so-called acute exacerbations, sudden deteriorations of pulmonary function of unclear etiology.5 Regardless, overall prognosis is quite dismal, with a median survival of 3.8 years from diagnosis.2 For decades, the only US Food and Drug Administration-approved treatment for this deadly disease has been lung transplantation, but with the approval of pirfenidone and nintedanib to treat IPF,6,7 the IPF landscape is sure to change.
The clinical problem
When evaluating a patient with pulmonary fibrosis, the clinician is faced with the following questions:

1. Is this disease IPF? There are several diseases that may present clinical, radiographic, and pathologic manifestations that are quite similar to IPF. Making an accurate clinical diagnosis is critical as these diseases may be treated differently. Even multidisciplinary experts demonstrate significant interobserver disagreement in IPF diagnosis based upon high-resolution computed tomography (CT) and histologic review when blinded.8
2. If it is IPF, will this patient benefit from the newly available treatments? At an estimated cost upwards of $100,000 per year for a course of treatment with pirfenidone or nintedanib,9 predicting a priori who is likely or unlikely to benefit is critical.10
3. Who will experience a more progressive course of disease? In practice, IPF patients tend to be referred for transplantation evaluation early on in the course of their disease – an international consensus report from 2006 outlined a guideline for referral in any patient with radiographic or histologic evidence of UIP.11 But identifying who should be prioritized for early transplant from those who will follow a more indolent course is currently one of the most critically important but unanswered questions for IPF clinicians.

Biomarkers appear to be the answer to these important clinical questions.

What is a biomarker?
Before exploring the current literature on IPF-specific biomarkers, it is beneficial first to consider the subject of biomarkers more generally and to create a framework for evaluating their potential utility. As defined by a National Institutes of Health expert-working group, a biomarker is “a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention”.12 As it stands, a biomarker denotes many things, including vital signs, general chemistry panels, specialized biochemical analyses, physiologic testing, and radiographic studies. While all of these measurements may have a role in the care of IPF patients, we focus this review on molecular and cellular biomarkers. Molecular and cellular biomarkers encompass any measurable element from the lung via exhaled condensate or bronchoalveolar lavage fluid (BALF) as well as urine, blood, or saliva. Technological advances in next-generation sequencing techniques have greatly expanded the armamentarium of potential biomarkers, but to date, there are no measured markers that have been implemented clinically except in Japan.13 A biomarker with clinical utility requires a sufficient measure of accuracy and reproducibility – that is, one would like a biomarker to have strong performance characteristics, including high sensitivity and specificity, and strong positive or negative predictive values and likelihood ratios, depending on specific clinical usage. Additional factors that are equally as important include safety and ease of obtaining the sample, acceptable costs to perform the test, and reasonable availability of testing.

A schematic workflow to translate a candidate biomarker into a clinical test is presented in Figure 1. The potential for molecular biomarkers in IPF is significant, and research suggests their possible value in confirming an accurate diagnosis of IPF, determining genetic predisposition for the development of IPF, providing prognostic information regarding disease course and mortality, measuring disease progression and activity, and gauging response to potential therapies.

What do candidate biomarkers tell us about IPF pathogenesis?
In their seminal review of IPF, published in 2001, Selman et al very elegantly refuted the commonly held notion of inflammation as the prominent causative factor of IPF.1 The studies of the last 14 years since this publication were some of the most critically important studies of IPF. Many genes and pathways of the deranged epithelial phenotype in IPF...
were identified that led to the study of many of the epithelial cell-derived candidate biomarkers described below. Pathologic imbalance of proteases and antiprotease studies also lead to candidate biomarkers, particularly the matrix metalloproteinases (MMPs), described in the Fibrogenesis/matrix remodeling section. What perhaps has been most unexpected was the recent identification of genetic factors such as Mucin 5B (MUC5B) polymorphism or the resurgence of inflammation as a critical component of the fibrotic process. In their updated review of IPF in 2011, King et al note that pulmonary fibrosis remains a disease of aberrant healing but that inflammation may play a role in the progression of disease in a subset of patients. This shift in thinking, we suggest, is reflected in part by the relative increase in the number of inflammatory candidate biomarkers including semaphorin 7a (Sema7a), BLyS, and chemokine (C-X-C motif) ligand 13 (CXCL13) compared to epithelial cell-derived biomarkers. This new focus of IPF biomarker research may also reflect the shift away from lung tissue-derived biomarkers to the less invasive peripheral blood biomarkers. Therefore, we suggest that candidate biomarkers speak to prevailing theories of the pathogenesis of IPF, and as these theories have evolved over time, so have the biomarkers. We have grouped the biomarkers below into categories based on their purported site of action. A similar categorization of IPF biomarkers has recently been published. The list below is roughly chronological. We review the clinical data, and we also discuss these biomarkers within the context of the larger shifts in our understanding of IPF. We first focus on biomarkers derived from epithelial cell injury, followed by a discussion of extracellular signaling molecules, matrix remodeling, cellular markers, and finally inflammatory biomarkers. Table 1 summarizes the data.

### Alveolar epithelial cell damage/dysfunction

Krebs von den Lungen 6

Krebs von den Lungen 6 (KL-6) antigen is a transmembrane mucin, or glycoprotein, encoded by MUC1 (Mucin 1, cell surface associated) that is expressed on the apical surface of type II alveolar epithelial cells. The precise function is unknown, but studies suggest that KL-6 has chemotactic activity on fibroblasts as well as anti-apoptotic effects on stimulated fibroblasts via interaction through an unknown receptor. KL-6 is in clinical practice in Japan, and several studies have been completed in this population. While KL-6 level elevation may be sensitive for interstitial lung disease (ILD), with differentiation demonstrated between ILD and controls,

### Table 1 Summary of candidate biomarker data

<table>
<thead>
<tr>
<th>Mechanistic category</th>
<th>Candidate biomarker</th>
<th>Predisposition</th>
<th>Diagnosis</th>
<th>Prognosis</th>
<th>Therapeutic monitoring</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alveolar epithelial cell damage/dysfunction</td>
<td>KL-6</td>
<td>–</td>
<td>–</td>
<td>++</td>
<td>+</td>
</tr>
<tr>
<td>Extracellular signals</td>
<td>SP-A + SP-D</td>
<td>–</td>
<td>–</td>
<td>+</td>
<td>–</td>
</tr>
<tr>
<td>Fibrogenesis/matrix remodeling</td>
<td>OPN</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Immune dysregulation/inflammation</td>
<td>MMP7</td>
<td>–</td>
<td>–</td>
<td>++</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>LOXL2</td>
<td>–</td>
<td>–</td>
<td>+</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>Periostin</td>
<td>–</td>
<td>–</td>
<td>++</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>Fibrocytes</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>T-cells</td>
<td>CCL18</td>
<td>–</td>
<td>–</td>
<td>+</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>YKL-40</td>
<td>–</td>
<td>–</td>
<td>+</td>
<td>±</td>
</tr>
<tr>
<td></td>
<td>IL-8</td>
<td>–</td>
<td>–</td>
<td>+</td>
<td>++</td>
</tr>
<tr>
<td></td>
<td>ICAM-1</td>
<td>–</td>
<td>–</td>
<td>++</td>
<td>–</td>
</tr>
<tr>
<td>B-cells and autoimmunity</td>
<td>MUC5B</td>
<td>–</td>
<td>–</td>
<td>+</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>Telomere shortening</td>
<td>+</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td></td>
<td>TOLLIP</td>
<td>+</td>
<td>–</td>
<td>+</td>
<td>–</td>
</tr>
</tbody>
</table>

**Notes:** The pathways reviewed in this manuscript are listed based on their proposed mechanistic category. We have rated the studies as follows: ++ denotes relatively strong evidence for the biomarker based on multiple studies or on large single studies; + denotes small single-study evidence for biomarker utility; ± denotes candidate biomarkers with equivocal evidence; – denotes candidate biomarkers without data.

**Abbreviations:** KL-6, Krebs von den Lungen 6; SP, surfactant protein; OPN, osteopontin; VEGF, vascular endothelial growth factor; MMP7, matrix metalloproteinase-7; LOXL2, lysyl oxidase-like 2; CCL18 chemokine (C-C motif) ligand 18; IL-8, interleukin-8; ICAM-1, intercellular adhesion molecule-1; Sema7a, semaphorin 7a; HSP70, heat shock protein 70; CXCL13, chemokine (C-X-C motif) ligand 13; MUC5B, Mucin 5B; TOLLIP, toll-interacting protein.
it is not specific for IPF. It is elevated in numerous ILD conditions, without significant differences between the UIP and nonspecific interstitial pneumonia (NSIP) histologies; therefore, it appears to have limited utility in diagnosis.\textsuperscript{17,18} This biomarker has shown promise in prognostication: in a smaller study with 27 patients, the authors found that lower baseline levels of KL-6 at time of diagnosis, <1,000 U/mL, have a more favorable 3-year survival rate.\textsuperscript{19} In a larger scale study of 219 patients with ILD, of which 183 were determined to have UIP pattern on high-resolution CT, it was found that patients with lower baseline KL-6 levels had significantly improved mortality over a 6-year period. The cut-off used in this study was again 1,000 U/mL, and, notably, those subjects with markedly elevated KL-6 levels had particularly poor survival.\textsuperscript{20} An additional study subsequently determined that baseline KL-6 levels above 1,300 U/mL were sensitive predictors for higher incidence of experiencing an acute exacerbation of IPF over a 3-year follow-up period, making it a potentially valuable tool for prognostication.\textsuperscript{21}

While the approved therapies for IPF are new, with limited data available to date, there is some evidence that KL-6 may be a marker for response to drug therapy. A small study of 14 patients with acute exacerbations of IPF examined KL-6 levels pre-pulse dose steroid administration as well as 1 week and 3 weeks post-administration: patients who survived 3 months had decreasing levels of KL-6 at each follow-up interval, which could suggest that KL-6 levels fall either in response to therapy or that elevated KL-6 levels reflect poorer prognosis.\textsuperscript{22} Finally, although not a primary end point, a recent study investigating the use of pirfenidone found that the medication not only attenuated a decline in pulmonary function but also reduced serum levels of KL-6, again suggesting the biomarker as an indicator of therapeutic response.\textsuperscript{23} Probably, the most significant limitation to a wider application of the results of most KL-6 studies is the limited population of the studies, although a study including German subjects looking at MUC1 polymorphisms suggests that extrapolation may be possible.\textsuperscript{13} As a molecular biomarker in IPF, KL-6 seems to demonstrate promise primarily in prognostication, with some data hinting at possibilities for therapeutic response monitoring.

**Surfactant proteins A and D**

Surfactant proteins (SPs) are lipoprotein complexes synthesized and secreted by type II alveolar epithelial cells with identified roles in reducing the surface tension at the lung air–liquid interface and promoting lung host defense. SFTPA (SP-A) and SFTPD (SP-D), specifically, have demonstrated roles in the innate immune system through direct binding to microorganisms and regulation of macrophages, and also in the adaptive immune system through interacting with both dendritic cells and T-cells. Beyond protecting the host, evidence also points to a role in modulating the inflammatory response to protect the lung from injury.\textsuperscript{24,25} One suggested mechanism is that chronic type II cell injury leads to cell turnover, possible somatic DNA mutations, resulting in apoptosis via endoplasmic reticulum stress secondary to protein misfolding or telomerase shortening. Subsequently, through debated pathways, epithelial cell injury results in fibrosis.\textsuperscript{25} Strengthening the case, limited studies have linked mutations in SP-A with familial IPF; although given sparse available data to date, its potential use in screening at-risk populations cannot be determined.\textsuperscript{26} Mutations in SP-C (SFTPC) have clearly been associated with familial pulmonary fibrosis.\textsuperscript{27–30}

SPs have been measured in both serum and BALF samples, with results suggesting possible roles in diagnosis and prognostication. In a 2002 study referenced above, 33 patients with ILD, including 21 diagnosed with IPF, were found to have serum levels of SP-A and SP-D that were significantly elevated compared to controls; significant differences between IPF and other ILD subjects were observed only with SP-D.\textsuperscript{17} A small study evaluating levels of SP-A in BALF samples of IPF and sarcoidosis patients suggests that lower levels of SP-A are present in IPF patients.\textsuperscript{31} This is similar to findings in a prior study from 1995 that measured SP-A in the BALF of IPF patients versus healthy subjects, with results significant for lower SP-A levels in IPF, as well as lower levels equating to diminished survival.\textsuperscript{32} However, when compared to other ILD patients in a separate study, a reduction in BALF SP-A content was not found to be specific for IPF.\textsuperscript{33} Further contradicting those prior studies, a more recent evaluation of BALF in patients with IPF and hypersensitivity pneumonitis showed increased levels of SP-A in hypersensitivity pneumonitis (HP) compared to control samples.\textsuperscript{34}

Serum testing has yielded mixed results: a large study of IPF patients in 2002 found limited utility in SP serum levels in distinguishing IPF from other types of ILD, but it did find elevated levels of both SP-A and SP-D to correlate with poor survival in IPF patients.\textsuperscript{35} Subsequently, a study of 82 IPF patients determined that elevated SP-A levels at initial diagnosis are significantly associated with 1-year mortality or transplant, while SP-D was not (although it did have a similar trend).\textsuperscript{36} Meanwhile, results looking only at the value of SP-D as a marker of prognosis did find significantly
increased mortality in IPF patients with elevated levels.\textsuperscript{37} As a biomarker, SPs appear to have limited clinical utility in the differentiation of IPF from other ILDs, with more encouraging results in the field of prognosis.\textsuperscript{38}

**Extracellular signals**

**Osteopontin**

Osteopontin (OPN) is a glycosylated phosphoprotein that is expressed in many tissues throughout the body, and within IPF lungs, it is localized to alveolar epithelial cells and macrophages.\textsuperscript{39} Significant advances in the understanding of this marker were made by Pardo et al, who discovered OPN to be one of the most highly upregulated genes in IPF via microarray gene expression studies on IPF lung tissue.\textsuperscript{39} In addition to the aforementioned immunohistochemical results, this group also found significantly elevated levels of OPN in BALF samples of 18 IPF patients compared to controls. They also demonstrated upregulation of, and co-localization with, another biomarker MMP7, and demonstrated in vitro associations of OPN with human fibroblast and epithelial cell growth and migration.\textsuperscript{39} A small study of 17 ILD patients and nine patients with sarcoidosis found significantly elevated levels of serum OPN in both conditions compared to controls; there was also a significant difference found between ILD and sarcoidosis patients, although IPF patients were statistically indistinguishable from other ILD conditions.\textsuperscript{40} This study noted an inverse correlation with arterial partial pressure of oxygen (PaO\textsubscript{2}) levels but not with pulmonary function test (PFT) values. There was no significant correlation between OPN and other biomarkers, including KL-6 and SP-A and SP-D.\textsuperscript{40} Despite the above discoveries strongly supporting a role for OPN in the pathogenesis of IPF, the data are insufficient to support its use as a biomarker for IPF.

**Vascular endothelial growth factor**

Vascular endothelial growth factor (VEGF) is a signaling protein that functions via tyrosine kinase receptors to modulate various angiogenic processes, including inhibition of endothelial cell apoptosis, regulation of vascular permeability, and proliferation, migration, and differentiation of endothelial cells.\textsuperscript{41} This protein is highly expressed in alveolar epithelial cells, as well as lung macrophages, smooth muscle, and myofibroblasts.\textsuperscript{42} A number of VEGF isoforms exist, the first discovered and most well known being VEGF-A; however, most IPF literature does not differentiate, and discussions refer to pan-VEGF isoforms unless specifically noted.\textsuperscript{43} A study by Meyer et al measuring VEGF levels in BALF and serum found significantly depressed values in BALF of IPF patients compared to controls and sarcoidosis and cystic fibrosis patients; serum levels did not demonstrate any significant differences.\textsuperscript{44} A few years later, another investigation of BALF by Koyama et al determined that VEGF concentrations were significantly reduced in all chronic pulmonary disease states evaluated (IPF, ILD related to connective tissue disease, and sarcoidosis), as well as in healthy smokers.\textsuperscript{45} Serum VEGF levels were reevaluated in a 2010 study that included 41 IPF patients, with results identifying significantly elevated VEGF levels in IPF patients with high Alveolar–arterial (A–a) oxygen gradients compared to IPF patients with lower A–a gradients or healthy controls. There was an additional finding of baseline serum VEGF values inversely correlating with subsequent forced vital capacity (FVC) declines.\textsuperscript{46} A very recent study of serum VEGF-A plasma levels found significantly elevated levels in IPF patients compared with controls, without any correlation with PFTs.\textsuperscript{47} As a biomarker, VEGF BALF levels appear significantly associated with IPF and could possibly be a component of evaluation in patients with an uncertain diagnosis, but it is probably of limited utility. Serum levels are more practical as a biomarker, and may have some utility in a diagnostic algorithm, as well as prognosis of clinical course. Nintedanib, one of the two newly approved therapies for IPF, is a tyrosine kinase inhibitor that targets VEGF signaling along with other pathways.\textsuperscript{7} Phase III trials have shown that treatment slows the decline of pulmonary function. Given the possibility that VEGF signaling is a part of pulmonary fibrosis, more study is necessary to determine how VEGF will perform as a biomarker of treatment success.

**Fibrogenesis/matrix remodeling**

**Matrix metalloproteinases**

MMPs are a collection of zinc-dependent endopeptidases that are collectively involved in the remodeling of the extracellular matrix and are thought to have roles in both normal organogenesis and inflammation. In fibrotic lung disease, they are purported to have a role in the observed dysregulated collagen turnover.\textsuperscript{48} MMP7 has been particularly well studied, with gene and protein expression levels found to be significantly increased in fibrotic lung, localizing to alveolar and bronchiolar epithelial cells. Furthermore, in vivo, MMP7 knockout mice were protected from bleomycin-induced fibrosis.\textsuperscript{49} An additional study found active MMP7 to immunolocalize to alveolar macrophages and hyperplastic epithelial cells in IPF lungs.\textsuperscript{50}

A study aiming to evaluate MMP levels in 20 IPF patients determined that MMP3, MMP7, MMP8, and MMP9 were all
elevated in BALF compared to healthy controls. The authors also found that all four of these MMPs were more elevated in those with poorer prognoses, and in particular, MMP8 and MMP9 were also significantly associated with rapidly declining pulmonary function as defined by persistent FVC reduction >10% over a 1-year period.\textsuperscript{51} A recent evaluation of MMP8 in IPF determined that there were increased BALF and serum levels compared to controls, although these increased levels were not found to correlate with declining lung function or mortality.\textsuperscript{52} Further studies of MMP7 in IPF patients have demonstrated increased BALF levels above normal controls but nonsignificant differences in serum levels and lack of differentiation between IPF and other ILDs.\textsuperscript{50,53,54} An analysis of multiple plasma proteins in IPF versus a control population that included chronic obstructive pulmonary disease (COPD), sarcoidosis, and HP identified MMP1 and MMP7 as primarily overexpressed proteins in IPF. In combination, MMP1 and MMP7 significantly differentiated IPF patients from the other chronic pulmonary conditions; interestingly, MMP7 levels correlated with subclinical ILD and were negatively associated with PFT results, suggesting possible roles in early diagnosis and gauging disease severity.\textsuperscript{55} Finally, a recent large study incorporating 241 IPF patients, with both derivation and validation cohorts, analyzed the concentrations of 92 serum proteins and found five proteins, including MMP7, to be associated with prognosis and mortality.\textsuperscript{56} What is perhaps most notable is that the robust nature of the results led the investigators to derive a severity prediction rule based on MMP7 levels. Overall, MMP7 appears to have the greatest individual clinical utility as a biomarker, primarily in the realm of prognosis, especially when combined with other biomarkers. MMP7 may have limited utility, again in combination with other biomarkers, in differentiation of IPF from other ILDs. The suggestion that MMP7 may have a role in diagnosis of subclinical disease is particularly intriguing, although further investigation is necessary.

**Lysyl oxidase-like 2 (LOXL2)**

Inhibition of LOXL2 in IPF is currently being studied in a clinical trial by Gilead Biosciences. Researchers from Gilead in collaboration with university-based coinvestigators found that higher serum levels of LOXL2 correlated with IPF disease progression.\textsuperscript{57} The authors note that implementation of LOXL2 as a measured biomarker will require validation in larger cohorts.

**Periostin**

Periostin is an extracellular matrix protein, part of a class of “matricellular” proteins that modulate cell–matrix interactions via signaling mechanisms and have identified roles in cardiac development and remodeling, tumorigenesis, and wound repair.\textsuperscript{58–60} The mechanism of action is not well defined in IPF, although a murine model has suggested a role in induction of chemokines and recruitment of neutrophils and macrophages.\textsuperscript{61} A study of 91 ILD patients, 51 with IPF, used immunohistochemistry to demonstrate strong expression of periostin in fibroblastic foci of lung tissue in UIP and NSIP. Further work by this group noted significantly elevated serum levels in IPF and NSIP compared to cryptogenic organizing pneumonia (COP) and normal controls, as well as inverse correlation with 6-month measurements of pulmonary function, suggesting a role for the biomarker in prognostication.\textsuperscript{62} The findings were largely verified by a second study of 54 IPF patients, with immunohistochemistry demonstrating increased staining for periostin in fibrotic lung tissue, as well as increased production of periostin messenger RNA by IPF fibroblasts. Clinically, the group was able to demonstrate correlation with serum periostin levels and disease progression, as defined by death, acute exacerbation of IPF, transplant, reduction of FVC by >10%, or decline of diffusing capacity of the lung for carbon monoxide (DLCO) by >15% at 48-week follow-up.\textsuperscript{63}

**Circulating fibrocytes**

In 2004, two studies published nearly simultaneously\textsuperscript{64,65} identified a bone marrow-derived collagen-producing cell, termed “fibrocyte”, as a critical player in matrix production in experimental lung fibrosis. Since then, multiple studies have examined these cells, defined as expressing both the common leukocyte antigen CD45 and collagen I, when isolated from the peripheral blood\textsuperscript{66} as biomarkers for disease severity in IPF and in other non-IPF diseases.\textsuperscript{67–70} The fibrocyte bridges our categories of matrix and remodeling and immune dysregulation as it exhibits properties of both. “Bursts” of fibrocytes did not correlate with death in a study of Hermansky–Pudlak syndrome,\textsuperscript{67} suggesting that deterioration or death does not correlate with absolute fibrocyte numbers and potentially limits enthusiasm for the fibrocyte as a biomarker.\textsuperscript{71} Further study will be necessary to assess absolute numbers of fibrocytes, their behavior over time, and their relationship with pulmonary function.

**Immune dysregulation/inflammation**

**Chemokine (C-C motif) ligand 18**

Chemokine (C-C motif) ligand 18 (CCL18), also known as pulmonary and activation-regulated chemokine, is an extracellular signal primarily expressed by alveolar macrophages...
that recruit lymphocytes and dendritic cells to foci of inflammation. Experimental studies indicate the existence of a positive feedback loop whereby alveolar macrophages are stimulated by native collagen to produce CCL18 and fibroblasts respond to CCL18 by synthesizing collagen. In a study of 55 patients with ILDs by Prasse et al, CCL18 levels were found to be elevated in both BALF and serum, most impressively in IPF, NSIP, and scleroderma. Moreover, elevated levels of CCL18 correlate with worse PFT. A further prospective study of 72 IPF patients, using a cut-off value for serum CCL18 concentrations of 150 ng/mL, found that patients above this level have significantly decreased measures of pulmonary function at 6-month follow-up, as well as increased mortality. A more recent investigation of CCL18 levels in the serum, BALF, and alveolar macrophage culture supernatants of patients with a variety of ILDs found elevated levels in various pulmonary conditions, most significantly in HP. This study posited that CCL18 may therefore reflect inflammatory response. Presently, collective results suggest limited utility for diagnostic purposes of this biomarker; however, the correlation with PFT trends, a known marker of disease progression, suggests utility as a prognostic tool in IPF patients.

**YKL-40**

A chitinase-like glycoprotein (chitinase 3-like 1, CHI3L1 [cartilage glycoprotein-39]), studied in the literature as YKL-40, has previously been associated with fibrosis in liver disease and to correlate with the degree of airway remodeling in severe asthma. While the mechanism of YKL-40 is not well understood, research has suggested roles in regulating connective tissue cell proliferation and angiogenesis. Within the lung, immunohistochemistry has demonstrated enhanced expression of YKL-40 in airway epithelial cells and alveolar macrophages. In a study of 63 IPF patients, Furuhashi et al found elevated levels of YKL-40 in serum and BALF samples, as well as a correlation with DLCO and KL-6 levels (serum YKL-40 only), suggesting a possible role for the biomarker in gauging disease activity and prognosis. Further work by Korthagen et al again determined that serum and BALF samples in 85 IPF patients are significantly elevated compared to controls; however, they found no correlation between serum levels and BALF levels or with PFT results. They followed the patients for >4 years and found that combining serum and BALF levels of YKL-40 allowed for stratification by mortality, with those patients having higher levels in both serum and BALF demonstrating increased mortality. Finally, a recent study of 315 patients with various ILDs, including 185 with IPF, determined significant elevations of YKL-40 in all ILDs compared to the control population, with the highest levels noted in NSIP, IPF, and COP. While the biomarker does not appear to have any role in diagnosis, a subset of the study with available serial follow-up did find levels of YKL-40 decreased over time in COP, a disease which tends to respond to available therapies, while remaining consistently elevated in IPF patients, suggesting a possible role in evaluating disease activity and therapeutic monitoring.

**Interleukin-8**

Interleukin-8 (IL-8; also known as CXCL8) is a cytokine produced by phagocytes exposed to inflammatory stimuli and serves as a chemoattractant for neutrophils. A study of the BALF in patients with IPF and fibrosis associated with known rheumatologic disease found significantly greater concentrations of IL-8 and neutrophils in both patient populations compared to healthy controls or those with non-pulmonary rheumatologic disease, suggesting that IL-8-mediated neutrophil chemotaxis is mechanistically involved in pulmonary fibrosis. When compared directly, the IPF population had significantly greater levels of IL-8 than those with systemic rheumatologic disease. A similar study of BALF samples from IPF patients attempted to correlate IL-8 serum levels with pulmonary function testing and PaO2 levels. This study then applied known clinical data and noted a negative correlation in lung function testing and PaO2 levels in patients with elevated serum IL-8 levels, suggesting a possible role for serum IL-8 as a tool for prognostication or monitoring of disease status. These data are bolstered by a more recent investigation of the serum levels of several cytokines in IPF patients that found cytokines IL-2, IL-8, IL-10, and IL-12 to be significantly elevated, with IL-8 specifically negatively correlating with PFTs. While studying a broader ILD population, and including only a small sample of IPF patients, a study by Vasakova et al similarly discovered a negative correlation between IL-8 BALF levels and lung function testing. Finally, in a study by Richards et al referred to elsewhere in this review,
high concentrations of plasma IL-8 reflected poor overall survival, transplant-free survival, and disease progression-free survival in the derivation cohort. Elevated IL-8 was also predictive of overall survival and poor transplant-free survival in an independent validation cohort. Several investigations certainly appear to support a role for IL-8 as a biomarker for prognostication given replicated relations with PFT declines, as well as the validated associations with survival.

**Intercellular adhesion molecule-1**

Intercellular adhesion molecule-1 (ICAM-1) is a transmembrane glycoprotein expressed on leukocytes and endothelial cells, with roles in T-cell-mediated cell defense and inflammation, participating in activated leukocyte migration. In 1992, Shijubo et al evaluated the serum levels of ICAM-1 in a variety of patient populations, including IPF, non-IPF ILD, pneumonia, and non-pulmonary inflammatory disorders, with a finding that levels are significantly elevated in IPF compared to other conditions. Additional immunohistochemistry analysis on lung tissue from the IPF patients demonstrated expression of ICAM-1 on alveolar epithelial cells. A follow-up study investigated the serum and BALF levels of ICAM-1 in IPF patients and again found significant elevations in the serum of IPF patients compared to both sarcoidosis patients and healthy volunteers, although BALF levels in IPF patients were not significantly elevated. In another study of the serum levels of several ICAM proteins in IPF, the authors found both ICAM-1 and ICAM-2 to be significantly elevated in IPF patients and secondary ILD patients. The group then reviewed PFT results and found that while ICAM-2 was significantly inversely associated with DLCO levels, ICAM-1 levels were not significantly associated with any particular PFT. Recently, this peripheral blood protein was also implicated in disease outcome predictions, including overall mortality, transplant-free survival, and even progression-free survival in the previously referred to study by Richards et al. Overall, early studies of this biomarker provide support for ICAM-1 having a plausible role in IPF pathogenesis, although its uses as even a diagnostic tool appear limited given similarly elevated levels in various ILD conditions. The recent larger study by Richards et al, however, strongly suggests at least a role in prognostication.

**T-cells: Sema7a and CD28**

Sema7a is a membrane-bound protein that is part of a family of proteins initially identified as having a role in axonal formation in embryonic development. Individually, Sema7a, which is expressed on activated T-cells, has identified functions in T-cell-mediated inflammation and immune regulation, and has been demonstrated to be a necessary component of transforming-growth factor beta 1-mediated pulmonary fibrosis in animal models. A study by Reilkoff et al demonstrated that Sema7a expression localizes to macrophages and CD4+ lymphocytes in IPF lungs and is expressed on peripheral CD4+ cells in IPF serum, specifically T-regulatory cells (Tregs). Furthermore, the study demonstrates significantly increased expression of Sema7a+ Tregs in IPF subjects compared to a control population, and in follow-up, noted the highest increases in patients with progressive IPF, suggesting a role in prognostication. Treg participation in IPF has a somewhat conflicting history, however, with studies demonstrating both protective and causative roles in fibrosis, suggesting that further research on this promising subset of Tregs is warranted.

Markedly decreased expression of CD28, a necessary component of T-cell activation, on the surface of CD4+ T-cells was correlated with a more severe 1-year outcome in IPF. These findings were supported further by genomic analysis of peripheral blood mononuclear cells from two separate IPF cohorts that showed that low expression of four genes, LCK, ITK, ICOS, and CD28, was associated with increased IPF mortality. These four genes all map to the T-cell co-stimulatory signal pathway and suggest that repeated cycles of T-cell stimulation are associated with the development of a pathologic T-cell phenotype and a significantly worsened prognosis in IPF.

**B-cell pathways: autoantibodies, BlyS, and CXCL13**

In what appears to be a return to some of the earliest work on the pathogenesis of IPF, several recent studies have focused directly on autoantibodies and B-cells, using BlyS and CXCL13 as markers of B-cell activation. Recent data implicate B-cell pathways in the pathogenesis of IPF. Isolation of autoantibodies to HSP70 in patients with IPF was associated with a poorer prognosis. BlyS, also known as B-cell-activating factor (BAFF, official gene name TNFSF13b, tumor necrosis factor (ligand) superfamily, member 13b), is a cytokine in the tumor necrosis factor family that is secreted by lymphocytes and is critical to B-cell maturation and antibody production. Xue et al found that plasma BlyS was significantly elevated in patients with IPF when compared to controls without lung disease and with COPD. Furthermore, there was a significant correlation between elevated pulmonary artery pressures and BlyS levels in patients with IPF, as well as lung function, as measured by...
FVC. Post hoc analyses from this study demonstrated that IPF patients with the most elevated BLyS levels (highest 25%) had increased 1-year mortality.

Another biomarker of interest is CXCL13, an agent that promotes B-cell trafficking to immunologic foci and has also been implicated in autoimmunity. DePlanto et al looked at lung tissue from patients with IPF and controls and correlated distinct histological features (lymphoid aggregates) of the IPF lung with elevated CXCL13 expression, by immunohistochemistry, plasma levels, and transcriptomics. Similar to BLyS described above, elevated CXCL13 levels correlated negatively with both lung function, measured by DLCO, and survival in patients with IPF. Vuga et al measured CXCL13 expression in both lung tissue and plasma in patients with IPF and COPD controls and found significantly increased CXCL13 levels in the IPF population, which correlated negatively with survival and pulmonary function as measured by DLCO.

Genetic mutations
As discussed, biomarkers are not limited to the measure of protein levels in peripheral blood or BALF, and recent genetic investigations of IPF patients have yielded a number of intriguing variations that may hold clinical utility. In 2011, Seibold et al identified a common polymorphism in the promoter region of MUC5B, a gel-forming mucin gene expressed by bronchial epithelial cells, while performing a genome-wide scan of 82 families with familial interstitial pneumonia (FIP). When applied to a population including FIP, IPF, and control patients, the mutation was identified in >30% of FIP and IPF patients but only 9% of controls. This group also demonstrated that the mutation is associated with increased gene expression, although the exact mechanism of pathogenesis remains unclear.

In support of the data from Seibold, Zhang et al confirmed this association in a larger case–control study including 341 IPF patients and 802 controls. While the MUC5B polymorphism is strongly associated with IPF, a retrospective analysis found improved survival in IPF patients with the polymorphism compared to those without, a finding that leads to speculation of differing phenotypes in those with the mutation. Mutations in genes encoding telomerase components (TERT/TERC) that lead to telomere shortening also appear associated with IPF. Even in the absence of known telomere-shortening mutations that were initially discovered in evaluations of familial IPF patients, significantly more patients with sporadic IPF have been noted to have circulating leukocytes with short telomeres compared to controls. Similar findings by Alder et al found that nonfamilial IPF patients, even in the absence of familial mutations, have shorter peripheral blood leukocyte telomeres than age-matched controls. Finally, a more recent genome-wide association study identified a common single-nucleotide polymorphism in TOLLIP, a toll-interacting protein felt to play a role in innate immune system regulation, that correlates with increased mortality. Clearly, these biomarkers have tantalizing utility in early prognostication and perhaps even identifying patients at risk for development of IPF. Similar studies have been identifying additional genetic variants that, with further investigation, will likely broaden the genetic understanding of IPF and possibly serve similar functions as biomarkers.

Future directions
We hope that we have made a convincing argument that the study of biomarkers is an essential part of IPF research. We suggest that biomarkers have the potential to impact all aspects of the clinical care of IPF patients including diagnosis and prognosis. There is a need for biomarkers obtained by less invasive techniques. With literature suggesting that there is increased short-term mortality following lung biopsy as well as possible increased risk of acute exacerbation, it is likely that there will be fewer and fewer lung biopsies in the future. The most recent consensus guidelines even dissuade from routine bronchoscopy for cellular analysis in the majority of patients. We are primarily left with peripheral blood sampling, with most studies to date focusing on measurable levels of peripheral blood proteins. Newer studies will likely focus on the data that are obtainable from peripheral blood mononuclear cells. Another potential source of information may be the isolation of circulating microRNA species.

How should biomarkers be studied? Hypothesis-generating omics experiments will be critical in this regard. The first step already exists. The Lung Genomics Research Consortium (LGRC) was funded to be the largest research database of clinical and genomics information in lung disease. This public database represents transcriptomic analysis of lung tissues from patients with IPF combined with cross-sectional clinical data. There are several shortcomings, however, with the LGRC data. First, for the study of biomarkers, as we have noted above, we must be able to derive data from more accessible sources such as blood. Second, because these data are cross-sectional, we do not see how these biomarkers behave over time. Therefore, we suggest that biomarker studies must accompany every clinical trial in IPF. Clinical trial participants represent a large and “captive” audience. For example, the ASCEND trial sponsored by Intermune
studied 555 patients worldwide. Every clinical trial visit represents an opportunity to collect clinical data and blood for biomarker discovery. Multiple omics technologies may be employed here including transcriptomics, proteomics, or metabolomics. We will need to see how these biomarkers perform as prospective markers of disease severity or of response to therapy.\textsuperscript{124}

**Conclusion**

Despite recent advances that have expanded upon our understanding of IPF pathogenesis, as well as clinical trials demonstrating the first effective therapies aside from lung transplantation, we still have an incomplete understanding of this devastating disease. For this particular reason, further investigations with biomarkers will not only impact clinical care but also answer some of the most basic questions of IPF pathogenesis. This review, which has attempted to analyze the data surrounding some of the best-studied biomarkers, universally suggests that no one biomarker is yet ready for regular clinical practice. For clinicians, biomarkers that could aid in the accurate diagnosis of IPF, as well as those that could assist in early prognostication, would be of immense benefit. The review suggests that we are without strong evidence to support use of biomarkers in sorting through a differential diagnosis. On the contrary, the evidence does suggest that we are nearing development of viable clinical prediction tools, which could aid in selection of patients for treatment initiation and prioritization in transplantation. It may be possible to create a strong model using pooled results of several biomarkers that independently appear to provide clinical prognostic information.\textsuperscript{56}

Finally, with the recent development and approval of IPF therapies, we have entered a new and very exciting era for IPF care where biomarker studies can begin to focus on response to therapy.

**Disclosure**

Dr Kass is the site investigator for clinical trials in IPF sponsored by Gilead Sciences. The other authors report no conflicts of interest in this work.

**References**


