Molecular Biomarkers in Idiopathic Pulmonary Fibrosis

Brett Ley, MD
Kevin K. Brown, MD
Harold R. Collard, MD†

1Division of Pulmonary and Critical Care Medicine, Department of Medicine, University of California, San Francisco, San Francisco, California

2Department of Medicine, National Jewish Health and the University of Colorado, Denver, Colorado

† Corresponding Author: Harold R. Collard, MD, 505 Parnassus Avenue, Box 0111, San Francisco, CA 94143, Email: hal.collard@ucsf.edu, Phone: 415-502-2361, Fax: 415-502-1321

Word Count: 4727

Running Head: Molecular biomarkers in IPF
ABSTRACT

Molecular biomarkers are highly desired in idiopathic pulmonary fibrosis (IPF) where they hold the potential to elucidate underlying disease mechanisms, accelerated drug development, and advance clinical management. Currently, there are no molecular biomarkers in widespread clinical use for IPF, and the search for potential markers remains in its infancy. Proposed core mechanisms in the pathogenesis of IPF for which candidate markers have been offered include alveolar epithelial cell dysfunction, immune dysregulation, and fibrogenesis. Useful markers reflect important pathologic pathways, are practically and accurately measured, have undergone extensive validation, and are an improvement upon the current approach for their intended use. The successful development of useful molecular biomarkers is a central challenge for the future of translational research in IPF and will require collaborative efforts among those parties invested in advancing the care of patients with IPF.
INTRODUCTION

Idiopathic pulmonary fibrosis (IPF) is a chronic fibrotic lung disease of unknown cause that generally affects adults over the age of 50 years. On surgical lung biopsy, it is characterized by the underlying histopathologic pattern of usual interstitial pneumonia (UIP) in the absence of secondary causes or associations (i.e. idiopathic UIP). IPF has a poor prognosis with an average survival of 3-5 years, but there is appreciable heterogeneity among individual patients in disease course. At present, there are no FDA approved medications for IPF, leaving limited therapeutic options for patients with progressive disease.

In categorizing patients with IPF, clinicians and researchers currently rely on clinical data— history and physical exam, radiology, pulmonary function tests, and histopathologic pattern—that do not directly reflect the underlying pathobiologic mechanisms of the disease. We are unable, therefore, to adequately sub-phenotype patients with IPF who may have quantitatively or qualitatively different biologic mechanisms responsible for their underlying disease. This is a major limitation to both clinical care and research in the field.

In this article, we explore the field of molecular biomarkers, which we use to refer to any objectively quantifiable substance at the cellular level or smaller (e.g. protein) that can be measured in biological tissue or fluids from patients with IPF. At their core, molecular biomarkers should reflect and inform the underlying biological mechanisms.
involved in a disease. This article aims to provide a framework for thinking about molecular biomarker development for IPF, using three core mechanistic themes relevant to IPF to review the steps necessary to the development of molecular biomarkers and to describe how the current crop of molecular biomarkers perform.

DEVELOPMENT OF MOLECULAR BIOMARKERS FOR IPF

In theory, molecular biomarkers may be used in IPF in several ways. These include identifying patients at risk for developing IPF (predisposition biomarkers); making the diagnosis of IPF (diagnostic or screening biomarkers); determining a patient's baseline prognosis, staging disease severity, and monitoring for progression (prognostic biomarkers); and identifying target engagement with a specific mechanistic response or predicting response or toxicity to therapy, either by identifying a specific subgroup most likely to respond to a therapy (therapeutic, predictive, or companion biomarker) or by substituting for a clinically meaningful outcome/endpoint (surrogate biomarker/endpoint). (10, 23) Currently, there are no molecular biomarkers in widespread clinical use for IPF for any of these uses.

There are two fundamentally different approaches to biomarker discovery which both have value. In the hypothesis-based approach, a candidate marker is selected a priori based on pre-existing rationale/evidence. This is the approach that has been taken in the majority of biomarker studies in IPF to date. This approach has the advantage of a strong biologic rationale and preliminary data, but suffers from a lack of efficiency in the
discovery process. The unbiased or hypothesis-free approach utilizes systems biology methods (e.g. genomic, transcriptomic, proteomic, and integratomic) to screen a large number of candidate markers for their association with the disease or outcome of interest, greatly increasing the efficiency and scope of the discovery process but increasing the risk of false discovery.

Most fundamentally, molecular biomarkers should reflect the presence and activity of relevant pathobiologic processes or mechanisms. In addition, biomarker measurement should be simple, technically accurate, broadly reproducible, and have acceptable risk. Therefore, measurement from easily obtainable body fluids or tissues is preferred (i.e. blood/serum, urine, exhaled breath condensates > BALF > transbronchial biopsy > surgical lung biopsy). Molecular biomarkers require broad validation across sexes, ages, ethnicities, and disease severity in order to assure generalizability. Practically, validation of the biomarker begins with a derivation (or discovery) cohort of appropriate size and relevance to its proposed role, in which reference ranges, thresholds, inter- and intra-subject variability, and predictive performance for the biomarker’s intended use are measured with the appropriate statistical tests. While techniques to reduce optimism (e.g. cross-validation, bootstrap resampling) can be used for “internal validation”, true validation (external validation) involves replicating results in one or more separate cohorts (ideally prospective and multicentered).
If considered for use in clinical practice, molecular biomarkers should improve upon current clinical practice compared to non-biomarker-driven strategies and their use should ultimately lead to improved patient outcomes. For example, a novel prognostic biomarker should add statistical performance to well-established prognostic variables (e.g. pulmonary function tests), and its measurement should lead to alterations in clinical management (e.g. change of therapy, referral for lung transplantation). If molecular biomarkers do not have all of these clinical characteristics, they still may add research value by providing insights into disease biology. Figure 1 outlines the ideal qualities of molecular biomarkers in IPF.

CANDIDATE MOLECULAR BIOMARKERS IN IPF

Over the past 15 years, the pathogenic paradigm for IPF has shifted from one of uncontrolled inflammation towards one of alveolar epithelial cell dysfunction, immune dysregulation, and fibroproliferation/fibrogenesis/matrix remodeling. In this paradigm, alveolar epithelial stress and dysfunction results in activation of pro-fibrotic signaling pathways, fibroblast proliferation, and exuberant extracellular matrix deposition. The resulting tissue destruction and architectural distortion eventually leads to organ dysfunction (e.g., restricted ventilation, hypoxemia), disability (shortness of breath and exercise limitation), and death from respiratory failure. In the remainder of this article, we use these proposed core mechanistic pathways to discuss and contextualize candidate biomarkers (Figure 2). Table 1 demonstrates the level of evidence supporting a clinical role for each candidate biomarker.
Core Pathway 1: Alveolar Epithelial Cell Dysfunction

Surfactant proteins

Surfactant proteins are synthesized and secreted by type II alveolar epithelial cells (AECs). Their functions include facilitating surfactant function and transport, and innate host defense. Quantitative or qualitative abnormalities in surfactant proteins are hypothesized to increase alveolar epithelial endoplasmic reticulum (ER) stress and trigger the unfolded protein response (UPR). Both mutations in genes encoding surfactant proteins and the level of surfactant proteins in blood and the extracellular lining fluid of the lung in patients have been proposed as potential molecular biomarkers of IPF.

Mutations in genes encoding surfactant proteins

Mutations in surfactant protein genes, especially A2 and C, have provided important insights into the pathogenesis of IPF. For example, some mutations have been shown to be associated with the development of pulmonary fibrosis, and the induction of chronic ER stress and the UPR in AECs. In animal models, a second pro-fibrotic stimulus (e.g. viral infection, bleomycin) may be necessary for the development of fibrosis, suggesting IPF may involve a “two-hit” pathobiology, similar to that proposed for some malignancies. Because IPF is an uncommon condition in the general population, and surfactant protein mutations are rarely associated with sporadic IPF, the most likely
role for the use of surfactant protein mutations as molecular biomarkers in clinical practice would be identifying at-risk individuals within families with pulmonary fibrosis known to have specific surfactant protein mutations.

Surfactant protein levels

Surfactant protein levels measurable in BALF and blood, specifically SPA and SPD, have been evaluated as diagnostic and prognostic molecular biomarkers in IPF. Because surfactant proteins are synthesized and secreted by type II AECs, their levels in extracellular fluids may reflect alterations in production and secretion by abnormal type II AECs, increased permeability of the epithelial-endothelial barrier, and/or type II AEC injury leading to non-secretory release. Gene expression profiling in IPF lungs demonstrates overexpression of the SPA1 gene in some patients with progressive disease compared to those with relatively stable disease lending support to altered surfactant production in patients with progressive IPF.(11) Both SPA and SPD levels are decreased in BALF of patients with IPF compared to healthy controls but are similar to other ILDs including NSIP.(8, 34) Conversely, serum levels of both SPA and SPD are elevated in IPF compared to healthy controls.(28, 34, 62) In small studies, serum levels of SPA and/or SPD were significantly higher in subjects with IPF compared to other ILDs, but with insufficient evidence to establish them as diagnostic markers.(28, 34, 62) Serum SPD levels have been reported to be significantly elevated in acute exacerbations of IPF (AE-IPF) compared to patients with stable IPF suggesting that type II AECs may play a direct or indirect role in the pathobiology of AE-IPF.(18) Prognostically, higher
serum levels of both SPA and SPD have been associated with reduced survival in IPF independent of clinical variables. However, in a recent prediction model study including 118 IPF patients, neither SPA nor SPD significantly improved discrimination of survival times when added to a clinical model (age, FVC, DLCO, and change in FVC). In two prospective treatment trials of pirfenidone, longitudinal changes in SPA and SPD levels were not significantly different between treatment and placebo groups. In sum, serum levels of SPA and SPD may be markers of abnormal function, injury, apoptosis or proliferation of type II AECs, but their measurement has yet to be proven clinically useful in making the diagnosis or predicting prognosis in patients with IPF. Whether serial measurements can reflect disease activity over time is unknown.

Krebs von den Lungen-6 (KL6)/Mucin 1 (MUC1)

KL6/MUC1 is a large membrane-bound glycoprotein in the mucin family expressed on the extracellular surface of type II AECs and bronchiolar epithelial cells in the lung, as well as glandular epithelial cells in other tissues including pancreas and breast. Originally investigated as a potential tumor marker for adenocarcinomas, it has since been studied extensively as a diagnostic and prognostic biomarker in ILDs. Expression of KL6 is increased in affected lung and regenerating type II AECs in a variety of ILDs including IPF. In vitro, KL6 has also demonstrated chemotactic, pro-proliferative, and anti-apoptotic effects on lung fibroblasts suggesting a potential pathogenic role in pulmonary fibrosis.
Serum levels of KL6 are significantly elevated in IPF and other ILDs compared to controls without ILD. Elevated serum levels, especially greater than 1000 U/ml, are significantly associated with reduced survival in many ILDs, including IPF. However, in the largest study of IPF patients (n=118), KL6 level at baseline demonstrated poor discrimination for survival time (C-statistic 0.58) and did not improve the prediction of survival beyond clinical predictors (age, FVC, DLCO, and change in FVC). In patients with AE-IPF, serum KL6 was elevated compared to patients with stable IPF, supporting a role for type II AEC injury in AE-IPF. However, when evaluated longitudinally in prospective treatment trials, changes in serum KL6 levels did not track with treatment response. In sum, KL6 may be a marker of type II AEC injury and may indicate worse survival among patients with ILDs, but may not be specific to the pathogenesis of IPF, or useful for diagnosing or prognosticating in IPF beyond readily available clinical information. Whether serial changes in KL6 prove useful for monitoring disease progression and/or diagnosing acute exacerbation in IPF deserves further study.

Mucin 5B (MUC5B)

Genome-wide linkage scanning involving 82 families with familial pulmonary fibrosis (FPF) followed by fine mapping of a region of interest on chromosome 11 in 575 subjects with pulmonary fibrosis (83 FPF, 492 sporadic IPF) and 322 controls, identified a common single nucleotide polymorphism (SNP) in the putative promoter region of the
MUC5B gene (rs35705950) that was highly associated with disease. The association has been confirmed in independent cohorts from the US and Europe. In total, the minor allele occurred in 34% of FPF, 34-42% of sporadic IPF, and 9-11% of control subjects. In the European study, no association was found with systemic sclerosis-associated ILD suggesting potential specificity for IPF. In a population-based study, there was an allele-dose-dependent association with an increased prevalence of interstitial lung abnormalities on chest imaging by computerized tomography (CT) in individuals without clinically recognized interstitial lung disease (i.e. subclinical ILD), lending further epidemiologic support for these genetic polymorphisms in MUC5B as a risk factor for the development of pulmonary fibrosis.

In addition to its association with an increased risk of developing IPF, the MUC5B promoter polymorphism appears to have prognostic value. In two independent IPF cohorts, one drawn from clinical practice (n=148) and one drawn from a large clinical trial (n=438), the presence of the MUC5B mutation (measured by two different assays) was associated with decreased mortality when compared to the wild-type SNP, independent of clinical factors (sex, FVC, and DLCO), and when included in a clinical model, significantly improved the prediction of mortality (C-statistic increased from 0.68-0.69 to 0.71-0.73). It is unclear how to reconcile MUC5B’s association with increased risk of developing IPF, and yet a better prognosis among those with established IPF. The mechanism by which the MUC5B promoter polymorphism could cause or contribute to IPF is unknown. MUC5B is overexpressed in the lung of...
subjects with IPF regardless of allele status, as well as in unaffected carriers compared to unaffected non-carriers, potentially implicating a direct role of the MUC5B protein in the pathogenesis of IPF. Elucidating a mechanistic role for MUC5B in IPF could lead to important advancements in our understanding of disease risk and prognosis in a large proportion of familial and sporadic cases.

Telomeres

Telomeres, the noncoding repetitive nucleotide sequence at the ends of DNA that protect against degradation, shorten with age. After telomeres reach a critical length, activation of a p53-dependent checkpoint leads to apoptosis or cellular senescence. Therefore, short telomeres, potentially leading to AEC senescence and apoptosis, provide an attractive pathogenic link between aging and the development of IPF. In 2007, heterozygous mutations in telomerase genes (TERT and TERC) were reported in 8-15% of families with IPF, and rarely in sporadic cases of IPF. Although telomerase mutations are rare in sporadic cases of IPF, shorter telomeres are common in sporadic IPF compared to age-matched controls. In fact, nearly one-quarter of patients with IPF have peripheral blood leukocyte telomere lengths in the <10th percentile. Short telomeres have also been associated with risk of developing chronic obstructive pulmonary disease, another aging-related lung disease, and thus may not be specific to the pathogenesis of IPF. Recently, peripheral blood leukocyte telomere length has been shown to be predictive of transplant-free survival in three separate cohorts of patients with IPF (total n=342), independent of other prognostic
factors (age, sex, FVC, and DLCO).(90) Pending further validation and delineation of
normal and threshold values, peripheral blood telomere length holds promise as a
mechanistic biomarker that may help predict predisposition to disease and disease
prognosis.

cleaved Cytokeratin 18 (cCK18)
In exploring the putative role of ER stress-UPR-induced AEC apoptosis in the
pathogenesis of IPF, a marker of the activity of this process would be highly desirable.
Cytokeratin 18 is a cytoskeletal protein found in AECs (and other epithelial cells). During
epithelial cell apoptosis, cytokeratin 18 is cleaved by caspases yielding a fragment,
cCK18, which is detectable by a specific antibody. In a recent study, cCK18 was detected
by immunohistochemistry in AECs of IPF lungs but not in normal lung tissue.(16) In vitro,
cCK18 and mediators of the UPR increased following induction of ER stress in type II
AECs. cCK18 was significantly elevated in the serum of IPF patients compared to normal
controls and patients with HP and NSIP, distinguishing IPF from HP/NSIP with moderate
accuracy (AUROC 0.76). However, cCK18 level at baseline was not associated with
disease severity or outcomes in IPF. Therefore, cCK18 may be mechanistically-
informative for the ER stress-UPR-AEC apoptosis in IPF, with elevated levels suggesting
higher degrees of AEC apoptosis in IPF compared to other ILDs. Although serum cCK18
levels at baseline do not appear to be useful for determining prognosis, serial levels
have not been evaluated for measuring ongoing disease activity and progression.
Further study will be necessary to determine whether cCK18 serum levels could serve as an efficacy marker for drugs intended to target ER stress-UPR-AEC apoptotic pathways.

**Core pathway 2: Immune dysregulation**

Innate and adaptive immunity contribute to tissue injury and repair including fibrogenesis in many tissues. The contribution of the immune system to tissue injury and fibrosis in IPF is controversial. Clinically, IPF patients do not respond to traditional immunosuppressive therapies. In fact, treatment of IPF with prednisone and azathioprine was associated with increased mortality and hospitalizations in a recent randomized-controlled trial. In addition, on histopathology, cellular infiltration by inflammatory cells is not conspicuous in IPF, consisting of minimal patchy lymphocytic and plasma cell infiltrates with occasional lymphoid aggregates in areas of dense fibrosis. A more nuanced look at immunity, however, suggests that there are important aspects of the IPF disease pathobiology that may be influenced by components of the immune system.

**Innate Immunity**

**TLR3**

Toll-like receptors (TLRs) are components of the innate immune system that are critical for response to infection and tissue injury. The pathogenic role of a functional SNP in the TLR3 gene (L412F) was recently investigated in IPF. Through a series of in vitro studies, activation of TLR3 in IPF lung fibroblasts with the L412F mutation demonstrated
reduced cytokine responses and dysregulated fibroproliferation compared to wild-type. (61) Also, TLR3\textsuperscript{null} mice had increased collagen and profibrotic cytokine production compared to wild-type. In two independent cohorts of IPF totaling 308 subjects, this mutation was associated with increased mortality independent of clinical parameters (age, FVC, & DLCO), and in a clinical trial cohort it was associated with accelerated physiologic progression. (61) The TLR3 SNP was common (32-43% were heterozygous, 6.5-11% were homozygous) and may be a factor leading to increased progression in established IPF.

**ELMOD2**

Genome-wide linkage scanning of six multiplex FPF families from southeastern Finland, an area with clustering of FPF due to common ancestry, identified mutations in ELMOD2 (ELMO domain containing 2) that was associated with decreased ELMOD2 mRNA expression in the lung. (31) In the lung, ELMOD2 is expressed in macrophages and type II AECs. In vitro, AEC cell lines overexpressing ELMOD2 activated gene expression pathways involved in response to viral infections, and ELMOD2 inhibition blunted antiviral cytokine expression after TLR3 activation, thus supporting an antiviral role for ELMOD2. (74)

**TOLLIP**

In a large genome-wide association study of IPF and control subjects, SNPs in the TOLLIP gene (Toll-interacting protein) were associated with development of IPF and decreased
expression of TOLLIP protein.(60) Counter-intuitively, one of the minor alleles was protective against the development of IPF but associated with increased mortality among established IPF subjects. TOLLIP is an adapter protein with ubiquitin-binding domains that modulates TLR, Interleukin-1, and TGFβ signaling through receptor trafficking and degradation.(110)

**α-defensins**

α-defensins are small antimicrobial proteins secreted by neutrophils and epithelial cells that are involved in multiple functions of innate immunity.(41) In a small study, plasma protein levels were elevated in IPF and correlated with measures of disease severity.(56) More recently, global gene expression profiling demonstrated α-defensins 3 & 4 to be significantly upregulated in the blood and lung tissue of AE-IPF compared to stable IPF.(41) In lung tissue from AE-IPF, α-defensin proteins were localized to type II AECs in the setting of widespread evidence of AEC apoptosis and proliferation. Peripheral blood gene expression profiling of 130 patients with IPF demonstrated that α-defensins 3 & 4 were among thirteen genes whose expression levels distinguished severe from mild IPF as defined by impairment in DLCO.(106) Interestingly, α-defensins are activated by matrix metalloproteinase 7 (MMP7), which is also highly expressed in type II AECs in IPF lungs and may be important in disease pathogenesis (see section on matrix metalloproteinases below).

**Alveolar Macrophage Activation**
CC chemokine ligand 18 (CCL18)

CCL18, also known as PARC for pulmonary and activation-regulated chemokine, is a chemokine protein produced by alveolar macrophages that stimulates collagen production in pulmonary fibroblasts. In fibrotic ILDs including IPF, there are increased numbers of CCL18-positive macrophages. CCL18 gene expression is increased in alveolar macrophages derived from BAL fluid, and this expression correlates with impairment in pulmonary function. In vitro, upregulation of CCL18 in alveolar macrophages occurs with Th2 cytokine stimulation (e.g. IL-13) and co-culture with fibroblasts via a β2-integrin/scavenger receptor pathway. In turn, CCL18 derived from alveolar macrophages induces fibroblast production of collagen, which appears independent of TGFβ signaling pathways. These studies suggest a positive feedback loop for collagen production in which Th2-activated alveolar macrophages (so called alternatively activated [M2] macrophages) produce CCL18, CCL18 stimulates collagen production by pulmonary fibroblasts, and fibroblast-derived collagen stimulates further CCL18 production by alveolar macrophages. In 72 patients with IPF, serum CCL18 concentration at baseline predicted subsequent physiologic progression and survival independent of other clinical parameters. Therefore, CCL18 may be a mediator and marker of a fibrogenic pathway common to IPF and other fibrotic ILDs, and its serum level may prove to be a useful prognostic biomarker.

YKL40
YKL40 is a chitinase-like protein expressed in many cell types that is involved in tissue response to injury and is dysregulated in many acute and chronic inflammatory conditions including COPD and asthma. Mechanisms by which it may be involved in tissue fibrosis include mediating Th2 cell-IL13 signaling pathways, inflammatory cell and fibroblast survival and proliferation, and alternative alveolar macrophage activation. YKL40 expression localizes to bronchiolar epithelial cells and alveolar macrophages adjacent to fibrotic areas in IPF, and protein levels are elevated in the BALF and serum of patients with IPF compared to healthy controls. Elevated serum concentrations correlate with worse gas exchange, and in small single-center study, both BALF and serum levels were associated with survival in IPF.

**Adaptive Immunity**

*Anti-HSP70 antibodies (Heat Shock Protein)*

In a study investigating the role of antigen-specific autoimmunity in IPF, IgG autoantibodies to HSP70 were identified in the serum of 25% of patients with IPF compared to 3% of healthy controls. HSP70 protein, antibody-antigen complexes, and complement were common in IPF lung tissue. HSP70 protein has been shown to induce CD4 T cells from IPF patients to proliferate and produce profibrotic cytokines (IL-4), and anti-HSP70 antibodies isolated from IPF patients have been shown to activate monocytes and induce IL-8 production. In a small, single-center study, anti-HSP70 IgG positivity in IPF patients was associated with physiologic progression and worse 1-year survival.
C-X-C motif chemokine 13 (CXCL13)

The importance of B cells in the pathogenesis of IPF is unknown, but a potential role is suggested by the identification of lymphoid aggregates in IPF lungs along with evidence for antibody dysregulation in some IPF patients. CXCL13 is a chemokine important for homing CXCR5-expressing B lymphocytes into lymphoid aggregates. In a recent study of 95 subjects with IPF, CXCL13 was significantly elevated in blood and lung tissue compared to controls (healthy subjects and subjects with COPD), and CXCL13 expression localized to lymphoid aggregates in the lung. Both baseline levels of CXCL13 and increases in CXCL13 over time were associated with reduced survival, independent of baseline demographic and physiologic parameters. These findings require external validation but suggest a potential role for CXCL13 as prognostic biomarker in IPF, and the study authors theorize that CXCL13 measurement could identify IPF patients who may be responsive to therapies that target B cells.

T-cell subsets

T cells are the predominant mononuclear cell type in IPF lungs, and their density in IPF lung tissue has been associated with disease severity and poor survival. CD4+ T cells express CD28, a co-stimulatory protein whose expression may be lost after repetitive cycles of antigen-driven stimulation and T cell proliferation. CD4+ CD28null cells are therefore considered a marker of chronic adaptive immune activation. In a study involving 89 patients with IPF, the proportion of CD28+ CD4 T cells in the peripheral
blood (CD28%) was low compared to healthy controls. (27) Also, low CD28% was correlated with low DLCO and reduced survival. In a subgroup of patients with serial measurements, longitudinal changes in CD28% correlated with longitudinal changes in FVC. (27) Clustering of peripheral blood mononuclear cell gene expression profiles in IPF patients identified a subgroup with down-regulated genes involved in the “co-stimulatory signal during T cell activation” pathway including CD28, ICOS, LCK, and ITK. (29) Quantitative expression levels of these four genes correlated with the percentage of peripheral blood CD4+CD28+ T cells, and in two small cohorts, improved prediction of transplant-free survival when added to clinical predictors. (29) These findings were confirmed in a validation cohort and were not explained by immunosuppression use.

Regulatory T cells (T regs) play an important role in modulation of the adaptive immune response. In a small study, T regs (CD4+CD25+FOXP3+) were reduced in the BALF and blood of patients with IPF compared to healthy controls and other ILDs, and T reg suppressor function, not number, was correlated with impairment in pulmonary function. (43) Semaphorin 7a (Sema 7a) is a membrane protein that modulates lymphocyte function. Its expression on T regs is increased in the lung and peripheral blood of patients with IPF, and peripheral blood Sema 7a+ T regs are elevated in patients with rapidly progressive disease. (78) Sema 7a+ T regs have reduced expression of regulatory cytokines (e.g. IL-10) and were able to induce fibrosis in a TGFβ-overexpression mouse model.
Core pathway 3: Fibroproliferation, Fibrogenesis, and Extracellular Matrix Remodeling

Matrix Metalloproteinases (MMPs)

MMPs are a large family of zinc-containing metallopeptidases that degrade multiple components of ECM, activate and degrade biologic mediators (e.g. growth factors, chemokines, and cell surface receptors), and facilitate cell migration. There are 23 known MMPs expressed from 24 gene loci, and most are secreted. MMPs are critically important in the homeostasis of the ECM, expressed at low levels in healthy tissue and upregulated in wound healing. MMPs are some of the most highly expressed genes in IPF lungs. In IPF lungs, MMPs are primarily produced by activated type II AECs, but a few are produced by fibroblasts within fibroblast foci.

MMP1 degrades type 1 and 3 collagen and is expressed by type II AECs. MMP1 is increased in lung tissue and BALF from patients with IPF, and serum levels of MMP1 are higher in IPF compared to HP, sarcoidosis, and COPD. MMP7 is also produced by activated type II AECs and upregulated in IPF. It has broad substrate activity including multiple basement membrane and ECM components, signaling molecules, and receptors suggesting a potential role in mediating an array of biologic processes.

MMP7 co-localizes in IPF AECs with osteopontin (OPN), a proinflammatory and profibrotic cytokine (see below), and MMP7 expression/activation is induced by OPN, which in turn is cleaved and activated by MMP7. Similar to MMP1, serum protein levels of MMP7 are higher in IPF compared to HP, sarcoidosis, and COPD. In one study, the combination of serum MMP1 & MMP7 levels was able to distinguish IPF from...
HP with 96% sensitivity and 87% specificity. Serum MMP7 levels also correlated with impairment in pulmonary function (FVC and DLCO). In a candidate serum protein screening approach in IPF patients, MMP7 concentration was identified as significantly associated with survival and progression-free and transplant-free survival in the discovery cohort (n=140) and transplant-free survival in the validation cohort (n=101). MMP7 was further included in a prognostic model along with clinical parameters (sex, FVC, and DLCO) derived from the discovery cohort, demonstrating a good to excellent discriminative performance for survival in the validation cohort (C-statistic 0.73-0.84). However, model calibration was not presented. In a large clinical trial cohort of IPF subjects (n=438), MMP7 was an independent predictor of survival in a model including clinical parameters (sex, FVC, DLCO) and MUC5B genotype (p=0.04). Change in serum MMP7 level is currently being used as a primary efficacy outcome in a phase 2 trial of inhaled carbon monoxide therapy for IPF (clinicaltrials.gov NCT01214187).

Several other MMPs are of interest in the pathobiology of IPF. MMP3 expression is increased in IPF lungs, and MMP3null mice are protected from bleomycin-induced lung injury suggesting a potential pathogenic role. In vitro, MMP3 also cleaves and activates OPN, and exposure of lung epithelial cells to MMP3 results in β-catenin signaling activation via cleavage of E-cadherin and subsequent activation of EMT. MMPs may play a role in fibrocyte migration into lung tissue. Human fibrocytes strongly express MMP 2, 7, 8, and 9, where MMP8 specifically mediates migration through
collagen 1, and MMP2 and 9 may facilitate migration through basement membrane components.(26) MMP9 is expressed in phenotypically altered (Thy-1 negative) fibroblasts within fibroblast foci in response to TGFβ1 stimulation(77) and its activated form is increased in the BALF of patients with rapidly progressive compared to slowly progressive disease.(84) MMPs may also play a role in fibrocyte migration into lung tissue.(26)

Osteopontin (OPN)

OPN is a secreted phosphorylated glycoprotein (also called secreted phosphoprotein 1 or SPP1) that functions as a mediator of inflammation and wound healing in many organs, and may play a role in multiple TGFβ-mediated processes.(22, 102) Its expression is increased in IPF lungs and localizes to AECs and macrophages.(67, 91, 105, 111) In vitro, OPN promotes migration and proliferation of fibroblasts, and its expression is increased by bleomycin-induced lung injury in mice.(67, 91) OPN induces upregulation of MMP7 in AECs and colocalizes with MMP7 in AECS of IPF lungs.(67) OPN is elevated in the BALF and plasma of IPF patients compared to healthy controls, but plasma levels are not significantly different from other ILDs.(36, 67) In a mix of ILD including IPF, plasma OPN levels correlated with oxygenation but not pulmonary function parameters.(36) OPN gene expression is increased in the lung tissue of IPF patients with progressive disease compared to those with stable disease.(11)

Periostin
Periostin is an ECM and intracellular protein (so called matricellular) that promotes ECM deposition, mesenchymal cell proliferation, and wound closure, and contributes to fibrosis in multiple organs. In mice, periostin is upregulated after bleomycin-induced lung injury, and blocking periostin improves survival and limits collagen deposition. Periostin null mice are protected from bleomycin-induced fibrosis. Periostin is increased in lung tissue of IPF patients and localizes to fibroblasts, especially to fibroblasts in fibroblast foci. Serum periostin levels are elevated in IPF and correlate with physiologic progression. In asthma, periostin is secreted by bronchial epithelial cells in response to IL-13 and can serve as a marker of IL-13 expression and treatment response to IL-13 antagonists.

Circulating Fibrocytes

Circulating fibrocytes are a subpopulation of leukocytes that express hematopoietic (CD45, CD34) and mesenchymal (Coll 1, fibronectin) markers. Fibrocytes are mesenchymal progenitor cells believed to be capable of producing ECM and differentiating into fibroblasts and myofibroblasts. In murine pulmonary fibrosis models, circulating fibrocytes contributed to the lung fibroblast population and impaired recruitment of fibrocytes protected against fibrosis. Fibrocytes are detected in IPF lungs but not healthy control lungs, and are recruited by CXCL12 expression by AECs. In one study, circulating fibrocytes were higher in IPF patients compared to normal controls and were elevated in AE-IPF compared to stable IPF. In 3 patients who recovered from AE-IPF, circulating fibrocyte percentage returned to pre-
exacerbation levels in recovery. Circulating fibrocyte percent did not correlate with pulmonary function, exercise capacity, or HRCT fibrosis; however, >5% circulating fibrocytes was associated with worse survival.(54)

Conclusions and Future directions

Major advances in the understanding the pathobiology of IPF have occurred over the past decade leading to the recognition of numerous potential molecular biomarkers, but the field of molecular biomarkers for IPF remains in its infancy. Only two prognostic biomarkers (MUC5B and MMP7) appear promising enough at this time to consider for translation into clinical practice. Currently, there is insufficient evidence to support any biomarkers for other clinical roles. Mechanistically-informative molecular biomarkers that are practical, accurate, validated, and clinically useful are needed to move the field forward. Characterization of potential molecular biomarkers for IPF should continue to take advantage of the wealth of data generated by systems biology research, which provides an unbiased approach to identifying and validating candidate biomarkers and mechanistic pathways.

IPF, like many other clinically-defined diseases, is a biologically heterogenous disease, involving multiple complex and interactive mechanisms, the relative importance of which may vary among individual patients. Molecular biomarkers could someday soon lead to diagnostic reclassification and/or sub-phenotyping of IPF patients on the basis of an individual patient’s relevant biology, directly informing treatment options. We
believe the development of clinically useful molecular biomarkers for IPF is the central challenge to translational research in the field over the next decade, and that meeting this challenge will require collaborative efforts among those parties invested in advancing the care of patients with IPF.
Acknowledgements

The authors wish to thank Dean Sheppard for his thoughtful review of the manuscript.
References


85. Selman M, King TE, Pardo A, American Thoracic S, European Respiratory S, and American College of Chest P. Idiopathic pulmonary fibrosis:


Table 1. Candidate molecular biomarkers for idiopathic pulmonary fibrosis and the strength of evidence supporting their clinical role

<table>
<thead>
<tr>
<th>Biomarker</th>
<th>Predisposition</th>
<th>Diagnosis</th>
<th>Prognosis</th>
<th>Therapeutic</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Genomic</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SPC</td>
<td>+/-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SPA2</td>
<td>+/-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MUC5B</td>
<td>+/-</td>
<td></td>
<td>+/</td>
<td></td>
</tr>
<tr>
<td>TERT/TERC</td>
<td>+/-</td>
<td></td>
<td>+/-</td>
<td>+/-</td>
</tr>
<tr>
<td>Telomere length</td>
<td>+/-</td>
<td></td>
<td>+/-</td>
<td>+/-</td>
</tr>
<tr>
<td>ELMO2</td>
<td>+/-</td>
<td></td>
<td>+/-</td>
<td>+/-</td>
</tr>
<tr>
<td>TOLLIP</td>
<td>+/-</td>
<td></td>
<td>+/-</td>
<td>+/-</td>
</tr>
<tr>
<td>TLR3</td>
<td>-</td>
<td></td>
<td>+/-</td>
<td>+/-</td>
</tr>
<tr>
<td><strong>Blood Proteins</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SPA</td>
<td>+/-</td>
<td>+/-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>SPD</td>
<td>+/-</td>
<td>+/-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>KL6/MUC1</td>
<td>+/-</td>
<td>+/-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>α-defensins</td>
<td>+/-</td>
<td></td>
<td>+/-</td>
<td></td>
</tr>
<tr>
<td>cCK18</td>
<td>+/-</td>
<td></td>
<td>+/-</td>
<td></td>
</tr>
<tr>
<td>CCL18</td>
<td>+/-</td>
<td></td>
<td>+/-</td>
<td></td>
</tr>
<tr>
<td>YKL40</td>
<td>+/-</td>
<td></td>
<td>+/-</td>
<td></td>
</tr>
<tr>
<td>Anti-HSP70 IgG</td>
<td>+/-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CXCL13</td>
<td>+/-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MMP7</td>
<td>+/-</td>
<td></td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>MMP1</td>
<td>+/-</td>
<td></td>
<td>+/-</td>
<td></td>
</tr>
<tr>
<td>OPN</td>
<td>+/</td>
<td></td>
<td>+/-</td>
<td></td>
</tr>
<tr>
<td>Periostin</td>
<td>+/-</td>
<td></td>
<td>+/-</td>
<td></td>
</tr>
<tr>
<td><strong>Blood Cells</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CD28% CD4+ T cells</td>
<td>+/</td>
<td></td>
<td>+/-</td>
<td></td>
</tr>
<tr>
<td>Tregs</td>
<td>+/-</td>
<td></td>
<td>+/-</td>
<td></td>
</tr>
<tr>
<td>Semaphorin 7a+</td>
<td>+/-</td>
<td></td>
<td>+/-</td>
<td></td>
</tr>
<tr>
<td>Fibrocytes</td>
<td>+/-</td>
<td></td>
<td>+/-</td>
<td></td>
</tr>
</tbody>
</table>

Key: No molecular or cellular biomarkers have a proven clinical role in IPF, but the strength of evidence supporting a potential role is denoted as follows:

- **+** Consistent or strong evidence supporting potential role
- **+/-** Conflicting or inadequate evidence to support potential role
- **-** Consistent or strong evidence against a potential role
- **blank** No evidence

Abbreviations: SPC = surfactant protein C gene mutations; SPA2 = surfactant protein A2 gene mutations; MUC5B = mucin 5B promoter polymorphism; TERT/TERC = telomerase gene mutations; ELMO2 = ELMO containing domain 2 gene mutations; TOLLIP = Toll-interacting protein gene mutations; TLR3 = toll-like receptor 3 gene mutation; SPA = surfactant protein A; SPD = surfactant protein D; KL6/MUC1 = Krebs von den Lungen 6/mucin 1; cCK18 = cleaved cytokeratin 18; HSP = heat shock protein; CXCL = C-X-C motif chemokine; MMP = matrix metalloproteinase; OPN = osteopontin; Treg = regulatory T cell
**FIGURE LEGENDS**

**Figure 1.** Ideal qualities of molecular biomarkers for idiopathic pulmonary fibrosis

**Figure 2.** Core mechanisms and candidate molecular biomarkers for idiopathic pulmonary fibrosis

Abbreviations: SPC = surfactant protein C gene mutations; SPA2 = surfactant protein A2 gene mutations; MUC5B = mucin 5B promoter polymorphism; TERT/TERC = telomerase gene mutations; ELMOD2 = ELMO containing domain 2 gene mutations; TOLLIP = Toll-interacting protein gene mutations; TLR3 = toll-like receptor 3 gene mutation; SPA = surfactant protein A; SPD = surfactant protein D; KL6/MUC1 = Krebs von den Lungen 6/mucin 1; cCK18 = cleaved cytokeratin 18; HSP = heat shock protein; CXCL = C-X-C chemokine; MMP = matrix metalloproteinase; OPN = osteopontin; Treg = regulatory T cell; EMT = epithelial to mesenchymal transition; TGF = transforming growth factor
<table>
<thead>
<tr>
<th>Core quality</th>
<th>Additional required qualities for research use</th>
<th>Additional required quality for clinical use</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mechanistic</strong></td>
<td><strong>Practical</strong></td>
<td><strong>Accurate</strong></td>
</tr>
<tr>
<td>Predisposition</td>
<td>• Should reflect mechanisms/biological pathways linked to disease pre-disposition.</td>
<td>• Should be quick, non-invasive and inexpensive since measurement is in asymptomatic people.</td>
</tr>
<tr>
<td>Diagnostic</td>
<td>• Should reflect mechanisms/biological pathways distinct from other ILDs that mimic IPF.</td>
<td>• Measurement should be technically feasible/reproducible.</td>
</tr>
<tr>
<td>Prognostic</td>
<td>• Should provide quantitative assessment of mechanisms/biological pathways relevant to disease progression.</td>
<td>• Should be measureable repeatedly over time with acceptable risk.</td>
</tr>
<tr>
<td>Therapeutic</td>
<td>• Should provide quantitative assessment or indicate presence/absence of mechanisms/biological pathways targeted by the therapeutic intervention</td>
<td>• For surrogates, should be measureable repeatedly over time with acceptable risk.</td>
</tr>
</tbody>
</table>
Immune dysregulation

Infection (viral), microaspiration, inhalational (tobacco)

- SPC
- SPA2
- TERT/TERC
- Short telomeres
- ER stress
- Apoptosis

Epithelial cell dysfunction

- TGFβ

Fibrocytes

- TLR3?
- Periostin
- OPN

Epithelial to mesenchymal transition

- Collagen
- ECM

Fibroblasts and myofibroblasts

- TLR3?
- Periostin
- OPN

- CD4
- FOXP3
- Sema7a
- TGFβ

Fibrogenesis

- CCL18
- CXCL13
- OPN

- YKL40

Immune dysregulation

- CD28
- CD4
- CD25
- CD4+
- CD4+ T cells
- Tregs
- Semaphorin 7a+ T regs
- Fibrocytes

Blood proteins

- SPA, SPD, KL6/MUC1, α-defensins, cCK18, YKL40, Anti-HSP70 IgG

Genomic

- SPC
- TOLLIP
- SPA2
- TLR3
- MUC5B
- TERT/TERC
- Telomere length
- ELMOD2

Blood cells

- CD28+ CD4+ T cells
- Tregs

SPA, SPD, KL6/MUC1, α-defensins, cCK18, YKL40, Anti-HSP70 IgG, CXCL13, MMP7, MMP1, OPN, Periostin

CCL18

CD4

CD4

CD25

CD28

Anti-HSP70 IgG