Systems medicine advances in interstitial lung disease

Flavia R. Greiffo[1], Oliver Eickelberg[1,2] and Isis E. Fernandez[1]

Affiliations: [1]Comprehensive Pneumology Center, Ludwig-Maximilians-Universität, University Hospital Grosshadern and Helmholtz Zentrum München and Member of the German Center for Lung Research, Munich, Germany. [2]Division of Respiratory Sciences and Critical Care Medicine, Dept of Medicine, University of Colorado, Denver, CO, USA.

Correspondence: Isis E. Fernandez, Comprehensive Pneumology Center, Ludwig-Maximilians-Universität and Helmholtz Zentrum München, Max-Lebsche-Platz 31, 81377 Munich, Germany.
E-mail: isis.fernandez@helmholtz-muenchen.de

ABSTRACT   Fibrotic lung diseases involve subject–environment interactions, together with dysregulated homeostatic processes, impaired DNA repair and distorted immune functions. Systems medicine-based approaches are used to analyse diseases in a holistic manner, by integrating systems biology platforms along with clinical parameters, for the purpose of understanding disease origin, progression, exacerbation and remission.

Interstitial lung diseases (ILDs) refer to a heterogeneous group of complex fibrotic diseases. The increase of systems medicine-based approaches in the understanding of ILDs provides exceptional advantages by improving diagnostics, unravelling phenotypical differences, and stratifying patient populations by predictable outcomes and personalised treatments. This review discusses the state-of-the-art contributions of systems medicine-based approaches in ILDs over the past 5 years.

Introduction

Interstitial lung diseases (ILDs) are fibrosing diseases, characterised by a reversible or nonreversible limitation in the gas exchange capacity of the lung, induced by known or unknown causes. This occurs as a secondary effect to the excessive accumulation of cells from distinct sources (*e.g.* mesenchymal, epithelial and inflammatory), wound healing products and extracellular matrix (ECM) in the lung interstitium. ILDs refer to a large group of diseases with a high mortality index, overlapping clinical features, unpredictable clinical progression and no available curative therapies, as is the case for idiopathic pulmonary fibrosis (IPF) [1, 2].

Systems biology is a biology-based interdisciplinary area that studies complex interactions within biological systems, using a holistic approach to biological research. Driven by high-throughput "omic" technologies, it enables multiscale and insightful overviews of cells, organisms and populations. Systems medicine integrates systems biology into modelling of pathological mechanisms, along with clinical parameters. Dynamic analysis between clinical and omics-generated data through bioinformatic and computational tools helps to dissect altered pathways, and to understand disease establishment, progression and remission [3].

The application of systems medicine to ILDs seeks to analyse their heterogeneity in a comprehensive manner, with the purpose of identifying biomarkers and genetic factors that improve disease understanding from the physio-pathological and clinical perspectives. While there are many key studies that have increased our knowledge of IPF and other ILDs, this article concisely focuses on omic-related studies, which contributed to the field during the past 5 years. Contributions published prior to this time frame were comprehensively reviewed in 2012 by HERAZO-MAYA and KAMINSKI [4].

Systems medicine and ILDs

With scientific advancement and accessibility at reasonable costs, the widespread use of high-throughput omic technologies has increased greatly over the last decade. This phenomenon has had a major impact on our understanding of multiple diseases, where omic-generated data have led to a personalised approach, decreasing mortality and improving survival, *e.g.* in the case of *HER2* (human epidermal growth factor receptor type 2)-positive breast cancer patients or *EGFR* (epidermal growth factor receptor)-mutated patients. Omic-driven personalised medicine has the potential to allow individualised treatment selection, determined by specific characteristics of the patient and disease.

There are overlapping similarities among ILDs subtypes, complicating the precise diagnosis and representing a daily challenge in clinical practice. The available treatment options are limited, increasing the need for molecular fingerprinting of patient populations. Thus, this complex and heterogeneous group of diseases requires the integrated approach of systems medicine to clearly discriminate and better understand them. We consider that advances in multiplex approaches allow us to glimpse critical players in these biological systems (figure 1). However, confirmatory testing of individual interactions is critical to demonstrate the importance at the molecular, cellular and organism level. A combination of holistic and traditional reductionist approaches will thus be needed for further understanding this multifaceted disease process.

Systems medicine in IPF

IPF has the worst prognosis of all ILDs, with a median survival of 3–5 years after diagnosis and no curative treatment available [2, 5]. Distinct clinical phenotypes with different patterns of survival have being described in IPF [5]. Therefore, systems medicine represents a new era for IPF. Interesting biomarkers have been discovered using omic analysis (table 1) and the challenge now is to establish them in routine clinical practice. Currently, only clinical and physiological changes are used to characterise disease progression.

Genomics and transcriptomics

Advances in genomic techniques have allowed high-throughput analysis and discovery of gene deregulation in IPF [6]. In particular, genetic studies have contributed to a better understanding of IPF, *e.g.* the expression of *MUC5B* (mucin 5B) and *TOLLIP* (Toll-interacting protein) [7–9]. Polymorphism in the promoter region of *MUC5B* (rs35705950) is associated with a higher likelihood of IPF development, although patients carrying this allele present a milder disease course and improved survival [7]. *MUC5B* was found expressed in areas of microscopic honeycombing and honeycomb cysts [8]. To date, the precise role of *MUC5B* in the pathophysiology of IPF is unclear.

Furthermore, variants in *TOLLIP* have also been linked to susceptibility and treatment responses in IPF. Single nucleotide polymorphisms within *TOLLIP* (rs5743890/rs3750920) are associated with increased mortality risk (rs5743890) and better response to *N*-acetylcysteine treatment (rs3750920) in IPF patients [9, 10]. Similarly, *DSP* (desmoplakin) variance (rs2076295) is associated with increased risk of IPF. *MUC5B* and *DSP* expression in the lung, especially in the airway epithelium, indicates the contribution of the aberrant epithelium in IPF [11].

Most IPF cases are sporadic. However, genetic variations also include an autosomal dominance pattern of inheritance, leading to familial pulmonary fibrosis. Several studies showed a strong relationship between familial IPF and telomerase mutations and their shortening. Recently, a prospective study performed genetic evaluations of IPF patients and affected relatives, and confirmed a strong relationship between familial IPF and telomerase mutations and shortening [12, 13]. Several studies have reported that familial IPF is associated with variances of the genes *TERT* (telomerase reverse transcriptase), *TERC* (telomerase RNA component), *DKC1* (dyskerin pseudouridine synthase 1), *TINF* (TERF1 interacting nuclear factor 2), *RTEL1* (regulator of telomere elongation helicase 1) and *PARN* (poly(A)-specific RNase) [12, 13]. Familial IPF also presents mutations within surfactant protein-encoding genes *SFTPA2* (surfactant protein A2) and *SFTPC* (surfactant protein C), and *ABCA3* (ATP-binding cassette subfamily A member 3) [12, 14].

MicroRNAs (miRNAs) are small noncoding RNAs involved in the regulation of gene and protein expression, thus altering cellular phenotypes. IPF patients display downregulation of miRNA levels in
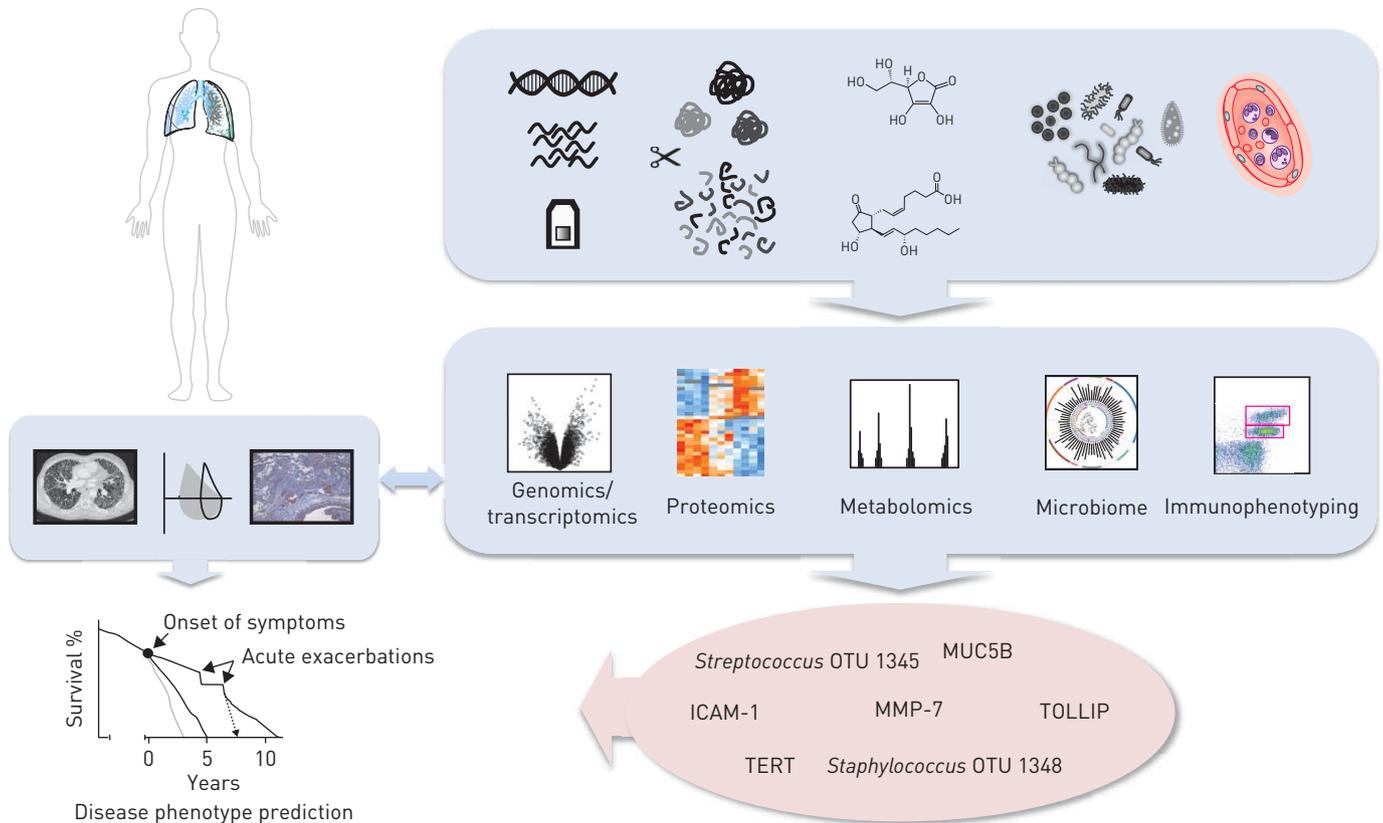
FIGURE 1 Systems medicine-based approaches in interstitial lung diseases seek to analyse biological products (*e.g.* RNA, DNA, proteins, metabolites, microbiome, *etc.*) and, through massive data generation and integration with clinical features, help to identify biomarkers that can predict disease phenotypes. OTU: operational taxonomic unit; MUC5B: mucin 5B; ICAM-1: intracellular adhesion molecule-1; MMP-7: metalloproteinase-7; TOLLIP: Toll-interacting protein; TERT: telomerase reverse transcriptase.

members of the let-7, mir-29 and mir-30 families, and upregulation in members of the mir-155 and mir-21 families, which modulate biological pathways and modify the IPF phenotype [15]. Altered expression of let-7 family members leads to changes in epithelial–mesenchymal transition in lung epithelial cells and inhibition of mir-21 modulates fibrosis [15]. mir-29 is also being explored as a potential target for IPF therapy as it decreases collagen expression in fibrotic lungs [16].

Moreover, IPF shares features with other chronic lung diseases, such as chronic obstructive pulmonary disease (COPD), where the molecular mechanism of lung injury leads to airway fibrosis. KUSKO *et al.* [17] revealed a shared mRNA–miRNA transcriptional network between IPF and COPD. Specifically, upregulation of mir-96 in both diseases may control part of the shared disease gene expression network. KUSKO *et al.* [17] describe overexpression of mir-96 as a key novel regulator of p53 expression in both epithelial cells and fibroblasts, which modulates the expression of genes related to the "emphysema–IPF" gene network [17]. Currently, miRNA therapy is being explored as a therapeutic option in diverse fibrotic conditions.

Cell-based RNA genomics can also provide an insight into disease features and targetable molecules. In line with the study by KUSKO *et al.* [17], single-cell RNA sequencing analysis of IPF and control lungs identified the cross-talk of four distinct lung epithelial cell subtypes (alveolar type 2, goblet, basal and indeterminate) [18]. XU *et al.* [18] showed that cells isolated from IPF patients express genes associated with activation of canonical transforming growth factor (TGF)-β, HIPPO/YAP, PI3K/AKT, p53 and WNT signalling cascades, which are activated in an integrated network [18]. Myofibroblast differentiation and massive ECM deposition are both ideal targetable phenomena in fibrotic diseases [19]. PARKER *et al.* [19] used fibroblasts to explore at the RNA level how proteins present in the ECM of acellular lungs can provoke a feedback loop to normal living cells exposed to an aberrant environment. The results suggest that characterisation of lung proteins, specifically the lung fibrotic ECM, will help to not only determine its composition, but also to define targetable molecules for advanced stages of fibrosis [19].

Advances in genomic technologies show an exciting time ahead, where analysis of the genetic makeup of IPF patients will dictate therapeutic approaches and predict disease outcome. Hence, implementing routine

TABLE 1 Potential interstitial lung disease (ILD) targets identified through systems medicine-based approaches

	Target	Reference(s)
IPF		
Genomics	MUC5B	[7, 8]
	TOLLIP	[9]
	let-7, mir-30, mir-155, mir-21	[15]
	mir-29	[15, 16]
Proteomics	CCL24, surfactant protein A2, NF-κB, peroxisome proliferator-activated receptor-γ	[20–22]
	Platelet-derived growth factor receptor-α	[23]
	Ficolin-2, cathepsin-S, legumain, inducible T-cell costimulator, trypsin-3	[25]
Metabolomics	Ornithine	[27]
	Lactate dehydrogenase-5	[28]
	6-Phosphofructo-2-kinase/fructose-2,6-biphosphatase 3	[51]
Microbiome	Staphylococcus OTU 1348	[30–32]
	Streptococcus OTU 1345	[30]
Peripheral blood phenotyping	Metalloproteinase-7	[33]
	CD28, inducible T-cell costimulator, lymphocyte-specific protein tyrosine kinase, interleukin-2 inducible T-cell kinase	[34, 35]
	Neoepitopes BGM, C1M, C3M, C5M, C6M, CRPM	[36]
	Plasma B-lymphocyte stimulating factor, B-cells	[37]
	Myeloid-derived suppressor cells	[38]
Familial IPF		
Genomics	TERT, TERC, DKC1, TINF2, RTEL1, PARN	[12, 13]
	SFTPA2, SFTPC, ABCA3	[12, 14]
Non-IPF ILDs		
Genomics	Telomere length	[43]
	RTEL1	[44]
	DPP9, DSP, FAM13A, IVD, DISP2, OBFC1, ATP11A, MUC2	[45]
	ADAMTS4, ADAMTS9, AGER, HIF1A, SERPINE2, SELE, RTKN2, PI15	[48]

IPF: idiopathic pulmonary fibrosis; OTU: operational taxonomic unit. See main text for further details of gene symbols.

clinical genotyping and biomarker testing is essential for future patient stratification and personalised treatment.

Proteomics

As a substantial part of systems medicine, proteomics covers information on protein abundance, variation and modification, including their partners and networks, in order to explain cellular processes. Given the impact of aberrant biological processes in IPF, proteomics analysis provides global protein quantification as a critical tool to identify disease-driving molecules and potential targets. The most universal and powerful method for global protein measurement is mass spectrometry, which uses liquid chromatography together with high-resolution tandem mass spectrometry to identify and quantify peptides at a large scale. These techniques, together with bioinformatic tools, contribute to a better understanding of protein biochemistry, nature and interaction.

It is possible that many of the proteins dysregulated in IPF may not cross the lung endothelial barrier or become diluted out by more abundant constituents, thus being undetectable in plasma [20]. Based on this concept, initial studies with proteomics in IPF have been primarily performed in bronchoalveolar lavage fluid (BALF). Osteopontin has been a more strongly validated marker increased in the BALF of IPF patients. Other targets have also been associated with disease, *e.g.* the CC chemokine ligand CCL24, surfactant protein A2, and transcriptional factors NF-κB, peroxisome proliferator-activated receptor-γ and c-Myc [20–22]. Recently, it was reported from proteomics analysis of fibroblast surface fractions that platelet-derived growth factor receptor (PDGFR)-α expression was altered by pro-fibrotic TGF-β in lung

fibroblasts from IPF patients [23]. The authors suggest a potential cross-talk between two critical signalling pathways in fibrosis, *i.e.* TGF-β/PDGFR-α, which affects myofibroblast differentiation in the context of IPF [23], currently insufficiently targeted by approved therapies.

In an iTRAQ (isobaric tag for relative and absolute quantitation)-based proteomics study, NIU *et al.* [24] identified a set of proteins in the serum of IPF patients. The identified targets were related to the protein activation cascade, regulation of response to wounding and extracellular components (see table 1). Unfortunately, no correlation with clinical parameters was performed. Recently, using a 1129-analyte SOMAmer (slow off-rate modified aptamer) array in IPF plasma, a six-analyte index predicted better progression-free survival in IPF. Specifically, high levels of ficolin-2, cathepsin-S, legumain and soluble vascular endothelial growth factor receptor-2, and low levels of inducible T-cell costimulator (ICOS) or trypsin-3 were associated with IPF progression [25]. The use of this index should be further validated in independent cohorts for general applicability.

Metabolomics

Metabolites are the second product of metabolic reactions catalysed by enzymes that naturally occur in the cells, complementing gene and protein expression [26]. Metabolic changes of the lung are involved in IPF pathogenesis, as reported by KANG *et al.* [27]. IPF presented 25 metabolite signatures, which indicated alteration in metabolic pathways of ATP degradation, glycolysis, glutathione biosynthesis and ornithine aminotransferase pathways [27]. Interestingly, ornithine has a negative correlation with forced vital capacity (FVC), suggesting the potential role of ornithine in IPF pathophysiology [27]. Moreover, lactic acid levels and lactate dehydrogenase (LDH)-5 levels were elevated in IPF lungs when compared with controls [28]. Increased levels of LDH-5 activate the TGF-β pathway, inducing myofibroblast differentiation [28].

Microbiome

The human microbiota consists of 10–100 trillion symbiotic microbial cells in a single individual; however, little is known about the contribution of the microbiome to health and disease. Recently, it was reported that IPF patients have higher bacterial load in BALF when compared with controls and the increased bacterial load identifies patients with more rapidly progressive disease [29]. Interestingly, MOLYNEAUX *et al.* [29] found bacterial burden to be independent of *MUC5B* genotype, suggesting a direct connection between host immunity and bacterial load [29]. In addition, it has been described in the COMET cohort that microbial signatures of *Staphylococcus* and *Streptococcus* help predict IPF progression [30]. Disease progression was significantly associated with increased abundance of two specific strains of *Staphylococcus* and *Streptococcus*. The study determined two operational taxonomic units (OTUs) associated with disease progression: *Staphylococcus* OTU 1348 and *Streptococcus* OTU 1345 [30]. This altered microbiome persists over time, which may implicate bacterial communities localised in the lower airways as persistent stimuli for repetitive alveolar injury in IPF [31]. IPF patients from the COMET cohort showed interactions between the host microbiome and progression-free survival [32]. Downregulation of the immune response was associated with changes in the abundance of *Staphylococcus* OTU 1348, which correlated with alterations in circulating leukocyte phenotypes, expression of several Toll-like receptors and fibroblast responsiveness [32].

Peripheral blood phenotyping

In the past, the role of the immune system in IPF has been controversial. However, increasing contributions indicate multiple alterations in the immune compartment of IPF patients, raising interest in the field. The peripheral blood compartment provides an easily accessible liquid biopsy, highly practical to study molecules and cell types as potential biomarkers. IPF patients have an increase in circulating biomarkers, *e.g.* CCL18, metalloproteinase (MMP)-7 and soluble intracellular adhesion molecule-1. To date, MMP-7 is the most validated biomarker to adopt clinically for diagnosis and prognostic evaluation of IPF [33]. Recently, several studies have validated the immunosuppressive environment present in the peripheral blood of IPF patients, showing that the gene expression of CD28, ICOS, lymphocyte-specific protein tyrosine kinase and interleukin-2 inducible T-cell kinase can predict outcome in IPF [34, 35].

In PROFILE, the largest IPF biomarker cohort studied globally, JENKINS *et al.* [36] showed changes in serum concentrations of proteolytically cleaved protein fragments/neopitopes. Levels of fragmented proteins generated by MMP activity and collagen synthesis in the serum of IPF patients were associated with IPF progression, as well as survival rate [36]. Immune cell type dysfunction has also been implicated in IPF. Abnormal B-cells and B-cell stimulator factor are often present in IPF patients, and are highly associated with disease manifestations and patient outcome [37]. Furthermore, myeloid-derived suppressor cells were described as a potent biomarker for IPF, where increased numbers of myeloid-derived suppressor cells measured in the blood of IPF patients correlated with lung function in cross-sectional and longitudinal analysis [38].

Omics in non-IPF ILDs

Non-IPF ILDs are defined by a large group of diseases classified by know causes or associations, *e.g.* idiopathic interstitial pneumonias (IIPs), granulomatous and other forms of ILDs [39]. The majority of non-IPF ILDs are idiopathic nonspecific interstitial pneumonia (NSIP), hypersensitivity pneumonitis, connective tissue disease-associated ILD and sarcoidosis. These diffuse parenchymal lung diseases affect the interstitium of the lung, distort pulmonary architecture and alter the gas exchange ability of the lung. ILDs can have associated causes, or not, but once scarring occurs, it is irreversible. The heterogeneity of non-IPF ILDs is extremely complex, with multiple common features and a high overlap in clinical, radiological and pathological patterns. Therefore, diagnosis requires a multidisciplinary approach and, in some cases, surgical lung biopsy, thus increasing mortality risk [40]. This highlights the need for integrated systems biology platforms to determine the molecular fingerprinting that improves diagnostic accuracy and provides novel targetable molecules.

KIM *et al.* [41] used integrative clustering analysis to integrate multi-omics data and propose an integrative phenotyping framework for identification of disease subtypes. Well-characterised clinical data, mRNA and miRNA profiles were integrated and visualised. Applying this method, clusters of homogeneous disease presentations and intermediate disease characteristics were identified. Validation of this open-access tool with other datasets is needed to support wide clinical usage. The integration of multi-omics data (*e.g.* proteomic, metabolomic, genomic and clinical data) provides a holistic integration of disease-related pathways at multiple levels, giving a futuristic perspective of understanding disease [41]. Using machine learning approaches with high-dimensional transcriptional data, a genomic signature could cluster ILD patients (IIP, NSIP, hypersensitivity pneumonitis and sarcoidosis) with a specificity of 92% (95% CI 81–100%) and a sensitivity of 82% (95% CI 64–95%) for microarray analysis, and a specificity of 95% (95% CI 84–100) and a sensitivity of 59% (95% CI 35–82%) for RNA sequencing. This study continues to support the need for better and less invasive diagnostic methods for ILDs patients [42]. Additionally, telomere length has been shown to be relevant in different ILDs. Telomeres are shorter in ILDs patients when compared with healthy controls, and even shorter in IPF when compared with other IIPs and sarcoidosis [43]. Moreover, a rare loss-of-function variant in *RTEL1* was found in peripheral blood mononuclear cells in one out of 25 families, representing a genetic predisposition for familiar interstitial pneumonia [44]. FINGERLIN *et al.* [45] identified 10 risk loci for fibrotic IIP (fIIP), *e.g.* *DPP9* (dipeptidyl peptidase 9), *DSP*, *FAM13A* (family with sequence similarity 13 member A), *IVD* (isovaleryl-CoA dehydrogenase), *DISP2* (dispatched RND transporter family member 2), *OBFC1* (oligonucleotide/ oligosaccharide-binding fold-containing protein 1), *ATP11A* (ATPase phospholipid transporting 11A) and *MUC2* (mucin 2), by genome-wide association studies. Furthermore, by imputing the data of genome-wide genotypes and conducting RNA sequencing studies, they identified new fIIP risk loci in lung tissue. Fibrotic lung tissue showed a deregulation in the human leukocyte antigen region of chromosome 6 (rs7887), involving and reaffirming the role of an immune deregulation in fibrotic ILDs [46].

Using the two-dimensional DIGE (difference in gel electrophoresis) technique and MALDI-TOF-MS (matrix assisted laser desorption ionisation-time of flight mass spectrometry), KORFEI *et al.* [47] sought differences between IPF and fibrotic NSIP, reporting that differences in expression of only a few proteins exist between these two entities. Although it is currently a limited and low-sensitivity proteome measuring technique, the authors interestingly reported that intracellular clearance of reactive oxygen species and carbonyl proteins seems to be enhanced in NSIP, due to enhanced expression of antioxidant acting proteins, which may explain the better outcome and survival in patients with NSIP [47].

Assessing the relationship between changes in lung function parameters and gene expression is a helpful method to identify clinically relevant molecules implicated in disease. STEELE *et al.* [48] used tissue gene expression profiling of IIPs to analyse differences and similarities among several IIPs, estimating the relationship between gene regulation and disease progression (using percentage predicted values for FVC and diffusing capacity of the lung for carbon monoxide). Some previously reported fibrotic targets were confirmed, *e.g.* *ADAMTS4* (ADAM metallopeptidase with thrombospondin type 1 motif 4), *ADAMTS9* (ADAM metallopeptidase with thrombospondin type 1 motif 9), *AGER* (advanced glycosylation end-product specific receptor), *HIF1A* (hypoxia inducible factor 1 α subunit), *SERPINE2* (serpin family E member 2) and *SELE* (selectin E), and novel candidates were identified, *e.g.* *RTKN2* (rhotekin 2) and *PI15* (peptidase inhibitor 15). Gene expression changes in those targets were significantly associated with lung function decline in moderate and severe IIP patients [48].

Conclusions

The pathobiology of pulmonary fibrosis involves adaptive and maladaptive pathways that work on multiple biological levels to disturb organ function. A comprehensive understanding of this complex interaction requires the integration of multiple data types: from DNA sequence variations to transcriptomics to

proteomics and ultimately phenomics, notwithstanding data obtained by classical physiology and pathology across different tissue and cell types. The speed of progress in science and technology applied to medicine in recent years has allowed several endotypes of diseases to be defined: from systems medicine-based approaches, leading to detection of specific targets that subclassify disease, to the possibility of offering personalised treatment, as is the case for EGFR in lung cancer.

Genome editing technologies, such as the CRISPR (clustered regularly interspaced palindromic repeat)-associated protein Cas9 (CRISPR-Cas9) system, provide hope that the discovery of critical gene regulators of disease can be applied to gene correction, such that CRISPR-Cas9-mediated gene correction could theoretically be offered for personalised treatments [49, 50]. Recently, XIE *et al.* [51] showed that glycolytic reprogramming is critical to lung myofibroblast differentiation and pulmonary fibrosis. Using CRISPR-Cas9 tools, inhibition of glycolysis was achieved by gene disruption of *PFKFB3* (6-phosphofructo-2-kinase/fructose-2,6-biphosphatase 3, a critical glycolytic enzyme). This example raises our hope that gene editing can be used for individualised therapeutics.

A combination of holistic and traditional reductionist experimental approaches will thus be required for the understanding of this multifaceted disease process. Notably, there are far more omics-related studies investigating IPF than other ILDs. Thus, community efforts must be invested in comprehensively characterising all ILDs, to learn more precisely about their common and distinctive features, and translate this gain of knowledge into more refined treatment options and better patient care than currently practised.

Acknowledgements

References

1   Travis WD, Costabel U, Hansell DM, *et al.* An official American Thoracic Society/European Respiratory Society statement: update of the international multidisciplinary classification of the idiopathic interstitial pneumonias. *Am J Respir Crit Care Med* 2013; 188: 733–748.

2   Fernandez IE, Eickelberg O. New cellular and molecular mechanisms of lung injury and fibrosis in idiopathic pulmonary fibrosis. *Lancet* 2012; 380: 680–688.

3   Capobianco E. Ten challenges for systems medicine. *Front Genet* 2012; 3: 193.

4   Herazo-Maya JD, Kaminski N. Personalized medicine: applying 'omics' to lung fibrosis. *Biomark Med* 2012; 6: 529–540.

5   Raghu G, Collard HR, Egan JJ, *et al.* An official ATS/ERS/JRS/ALAT statement: idiopathic pulmonary fibrosis: evidence-based guidelines for diagnosis and management. *Am J Respir Crit Care Med* 2011; 183: 788–824.

6   Kass DJ, Kaminski N. Evolving genomic approaches to idiopathic pulmonary fibrosis: moving beyond genes. *Clin Transl Sci* 2011; 4: 372–379.

7   Peljto AL, Zhang Y, Fingerlin TE, *et al.* Association between the *MUC5B* promoter polymorphism and survival in patients with idiopathic pulmonary fibrosis. *JAMA* 2013; 309: 2232–2239.

8   Seibold MA, Wise AL, Speer MC, *et al.* A common *MUC5B* promoter polymorphism and pulmonary fibrosis. *N Engl J Med* 2011; 364: 1503–1512.

9   Noth I, Zhang Y, Ma SF, *et al.* Genetic variants associated with idiopathic pulmonary fibrosis susceptibility and mortality: a genome-wide association study. *Lancet Respir Med* 2013; 1: 309–317.

10   Oldham JM, Ma SF, Martinez FJ, *et al.* TOLLIP, MUC5B, and the response to *N*-acetylcysteine among individuals with idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2015; 192: 1475–1482.

11   Mathai SK, Pedersen BS, Smith K, *et al.* Desmoplakin variants are associated with idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2016; 193: 1151–1160.

12   Kropski JA, Young LR, Cogan JD, *et al.* Genetic evaluation and testing of patients and families with idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2017; 195: 1423–1428.

13   Stuart BD, Choi J, Zaidi S, *et al.* Exome sequencing links mutations in *PARN* and *RTEL1* with familial pulmonary fibrosis and telomere shortening. *Nat Genet* 2015; 47: 512–517.

14   Epaud R, Delestrain C, Louha M, *et al.* Combined pulmonary fibrosis and emphysema syndrome associated with *ABCA3* mutations. *Eur Respir J* 2014; 43: 638–641.

15   Pandit KV, Milosevic J, Kaminski N. MicroRNAs in idiopathic pulmonary fibrosis. *Transl Res* 2011; 157: 191–199.

16   Montgomery RL, Yu G, Latimer PA, *et al.* MicroRNA mimicry blocks pulmonary fibrosis. *EMBO Mol Med* 2014; 6: 1347–1356.

17   Kusko RL, Brothers JF II, Tedrow J, *et al.* Integrated genomics reveals convergent transcriptomic networks underlying chronic obstructive pulmonary disease and idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2016; 194: 948–960.

18   Xu Y, Mizuno T, Sridharan A, *et al.* Single-cell RNA sequencing identifies diverse roles of epithelial cells in idiopathic pulmonary fibrosis. *JCI Insight* 2016; 1: e90558.

19   Parker MW, Rossi D, Peterson M, *et al.* Fibrotic extracellular matrix activates a profibrotic positive feedback loop. *J Clin Invest* 2014; 124: 1622–1635.

20   Foster MW, Morrison LD, Todd JL, *et al.* Quantitative proteomics of bronchoalveolar lavage fluid in idiopathic pulmonary fibrosis. *J Proteome Res* 2015; 14: 1238–1249.

21   Carleo A, Bargagli E, Landi C, *et al.* Comparative proteomic analysis of bronchoalveolar lavage of familial and sporadic cases of idiopathic pulmonary fibrosis. *J Breath Res* 2016; 10: 026007.

22   Landi C, Bargagli E, Carleo A, *et al.* A system biology study of BALF from patients affected by idiopathic pulmonary fibrosis (IPF) and healthy controls. *Proteomics Clin Appl* 2014; 8: 932–950.

23   Heinzelmann K, Noskovicova N, Merl-Pham J, *et al.* Surface proteome analysis identifies platelet derived growth factor receptor-alpha as a critical mediator of transforming growth factor-beta-induced collagen secretion. *Int J Biochem Cell Biol* 2016; 74: 44–59.

24   Niu R, Liu Y, Zhang Y, *et al.* iTRAQ-based proteomics reveals novel biomarkers for idiopathic pulmonary fibrosis. *PLoS One* 2017; 12: e0170741.

25   Ashley SL, Xia M, Murray S, *et al.* Six-SOMAmer index relating to immune, protease and angiogenic functions predicts progression in IPF. *PLoS One* 2016; 11: e0159878.

26   Maplestone RA, Stone MJ, Williams DH. The evolutionary role of secondary metabolites – a review. *Gene* 1992; 115: 151–157.

27   Kang YP, Lee SB, Lee JM, *et al.* Metabolic profiling regarding pathogenesis of idiopathic pulmonary fibrosis. *J Proteome Res* 2016; 15: 1717–1724.

28   Kottmann RM, Kulkarni AA, Smolnycki KA, *et al.* Lactic acid is elevated in idiopathic pulmonary fibrosis and induces myofibroblast differentiation via pH-dependent activation of transforming growth factor-beta. *Am J Respir Crit Care Med* 2012; 186: 740–751.

29   Molyneaux PL, Cox MJ, Willis-Owen SA, *et al.* The role of bacteria in the pathogenesis and progression of idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2014; 190: 906–913.

30   Han MK, Zhou Y, Murray S, *et al.* Lung microbiome and disease progression in idiopathic pulmonary fibrosis: an analysis of the COMET study. *Lancet Respir Med* 2014; 2: 548–556.

31   Molyneaux PL, Willis Owen SA, Cox MJ, *et al.* Host–microbial interactions in idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2017; 195: 1640–1650.

32   Huang Y, Ma SF, Espindola MS, *et al.* Microbes associate with host innate immune response in idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2017; 196: 208–219.

33   Ley B, Collard HR, King TE Jr. Clinical course and prediction of survival in idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2011; 183: 431–440.

34   Herazo-Maya JD, Noth I, Duncan SR, *et al.* Peripheral blood mononuclear cell gene expression profiles predict poor outcome in idiopathic pulmonary fibrosis. *Sci Transl Med* 2013; 5: 205ra136.

35   Huang Y, Ma SF, Vij R, *et al.* A functional genomic model for predicting prognosis in idiopathic pulmonary fibrosis. *BMC Pulm Med* 2015; 15: 147.

36   Jenkins RG, Simpson JK, Saini G, *et al.* Longitudinal change in collagen degradation biomarkers in idiopathic pulmonary fibrosis: an analysis from the prospective, multicentre PROFILE study. *Lancet Respir Med* 2015; 3: 462–472.

37   Xue J, Kass DJ, Bon J, *et al.* Plasma B lymphocyte stimulator and B cell differentiation in idiopathic pulmonary fibrosis patients. *J Immunol* 2013; 191: 2089–2095.

38   Fernandez IE, Greiffo FR, Frankenberger M, *et al.* Peripheral blood myeloid-derived suppressor cells reflect disease status in idiopathic pulmonary fibrosis. *Eur Respir J* 2016; 48: 1171–1183.

39   du Bois RM. Strategies for treating idiopathic pulmonary fibrosis. *Nat Rev Drug Discov* 2010; 9: 129–140.

40   Hutchinson JP, Fogarty AW, McKeever TM, *et al.* In-hospital mortality following surgical lung biopsy for interstitial lung disease in the USA: 2000–2011. *Am J Respir Crit Care Med* 2016; 193: 1161–1167.

41   Kim S, Herazo-Maya JD, Kang DD, *et al.* Integrative phenotyping framework (iPF): integrative clustering of multiple omics data identifies novel lung disease subphenotypes. *BMC Genomics* 2015; 16: 924.

42   Kim SY, Diggans J, Pankratz D, *et al.* Classification of usual interstitial pneumonia in patients with interstitial lung disease: assessment of a machine learning approach using high-dimensional transcriptional data. *Lancet Respir Med* 2015; 3: 473–482.

43   Snetselaar R, van Moorsel CH, Kazemier KM, *et al.* Telomere length in interstitial lung diseases. *Chest* 2015; 148: 1011–1018.

44   Cogan JD, Kropski JA, Zhao M, *et al.* Rare variants in *RTEL1* are associated with familial interstitial pneumonia. *Am J Respir Crit Care Med* 2015; 191: 646–655.

45   Fingerlin TE, Murphy E, Zhang W, *et al.* Genome-wide association study identifies multiple susceptibility loci for pulmonary fibrosis. *Nat Genet* 2013; 45: 613–620.

46   Fingerlin TE, Zhang W, Yang IV, *et al.* Genome-wide imputation study identifies novel HLA locus for pulmonary fibrosis and potential role for auto-immunity in fibrotic idiopathic interstitial pneumonia. *BMC Genet* 2016; 17: 74.

47   Korfei M, von der Beck D, Henneke I, *et al.* Comparative proteome analysis of lung tissue from patients with idiopathic pulmonary fibrosis (IPF), non-specific interstitial pneumonia (NSIP) and organ donors. *J Proteomics* 2013; 85: 109–128.

48   Steele MP, Luna LG, Coldren CD, *et al.* Relationship between gene expression and lung function in idiopathic interstitial pneumonias. *BMC Genomics* 2015; 16: 869.

49   Doudna JA, Charpentier E. Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science* 2014; 346: 1258096.

50   Sander JD, Joung JK. CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat Biotechnol* 2014; 32: 347–355.

51   Xie N, Tan Z, Banerjee S, *et al.* Glycolytic reprogramming in myofibroblast differentiation and lung fibrosis. *Am J Respir Crit Care Med* 2015; 192: 1462–1474.