

Landmark papers in respiratory medicine

Updates in using a molecular classifier to identify usual interstitial pneumonia in conventional transbronchial lung biopsy samples

The most common fibrosing interstitial lung disease (ILD) is idiopathic pulmonary fibrosis (IPF), with an incidence of 14–60 cases per 100 000 inhabitants per year in North America [1] and 3–9 cases per 100 000 per year in Europe [2]. IPF is a chronic, progressive fibrosing interstitial lung disease characterised by continued scarring of the lung parenchyma and associated with a steady worsening of respiratory symptoms, quality of life and pulmonary function, ultimately leading to death [1, 3], and a median survival of 3–5 years from the time of diagnosis [4, 5]. A precise diagnosis of the underlying ILD entity is essential for prognostication and choice of therapy as treatments differ between ILD subtypes, including that some drugs may be detrimental to an IPF patient. However, the diagnosis of ILD is sometimes difficult, partly imprecise, and frequently characterised by delay, misdiagnosis, use of costly and invasive procedures, and high use of healthcare resources.

In the absence of an alternative cause of interstitial pneumonia, IPF is defined by the presence of a usual interstitial pneumonia (UIP) pattern on a high-resolution computed tomography (HRCT) of the lungs or on biopsy [1, 6]. In the correct clinical context, an UIP or probable UIP pattern on HRCT may suffice for diagnosis [1, 7]. However, the combination of clinical information and HRCT pattern does not allow for a secure diagnosis in nearly half of the

patients [8]. Current guidelines advocate the use of surgical lung biopsy (SLB) in such cases, if patients are eligible for SLB [6]. However, SLB might be associated with a meaningful mortality and complication rate as well as costs [9]. Further difficulties include interrater disagreement and the possibility of sampling error [10].

Several alternative approaches are being investigated. As the diagnostic yield with transbronchial lung biopsy (TBLB) is low, current research has assessed the role of transbronchial lung cryobiopsy (TBLC). TBLC was shown to be a relatively safe method that provided significantly larger lung samples, where UIP can be observed with high confidence and a good overall agreement between observers [9]. Therefore, TBLC may be an adequate substitute for SLB in centres with expertise performing TBLC in the histopathological diagnosis of ILDs, as it has shown a relevant diagnostic value in the context of an experienced interdisciplinary team and may enable histopathological evaluation even in patients with a more severe disease not suitable for SLB [11].

Another important endoscopic technique is bronchoalveolar lavage (BAL) as it can provide additional information. The main indication for BAL is to obtain cell morphology, differential cytology and microbiology. In IPF, neutrophilia and eosinophilia are found commonly, while a lymphocytosis might be associated with, for example, hypersensitivity

Cite as: Crespo A, Alfaro T, Somogyi V, *et al.* Updates in using a molecular classifier to identify usual interstitial pneumonia in conventional transbronchial lung biopsy samples. *Breathe* 2020; 16: 200067

 @ERSpublications

A molecular classifier using a machine-learning algorithm based on genomic data could provide an objective method to aid clinicians and multidisciplinary teams to establish the diagnosis of IPF in less-invasive transbronchial lung biopsy samples <https://bit.ly/2QLdWim>



CrossMark



© ERS 2020

pneumonitis. However, BAL cytology results are unspecific [12].

Finally, multidisciplinary diagnosis is associated with great improvements in the diagnostic confidence, and evidence-based guidelines recommend an interdisciplinary approach involving clinicians, radiologists and pathologists to achieve a final diagnosis with the greatest confidence [10].

The aim of this paper is to provide an update on the role of molecular analysis in the diagnosis of IPF, using a machine learning algorithm in less-invasive TBLB samples, as a first step into precision medicine in this disease [13].

Biomarkers

Genomic, transcriptomic, metabolomic, proteomic and epigenetic biomarkers, gene co-expression networks, drug-gene interaction testing, lung microbiome and mitochondrial DNA analyses, gene expression in BAL assays, and quantitative CT and machine learning algorithms are helping understanding of pathomechanisms, detecting significant features, and determining predictive value of disease risk, outcome and mortality.

Improvements in TBLC, new endobronchial procedures and improvement of radiological techniques have all emerged as promising ways to increase diagnostic sensitivity and confidence for IPF in expert referral institutions. New biomarkers may provide a less invasive approach and an objective tool to define the UIP pattern from a TBLB specimen and reach a confident diagnosis of IPF.

Many of the biomarker studies are small, single-centre, retrospective trials with a small sample; however, the first longitudinal, prospective biomarker study (PROFILE) has been recently published [14]. Although, reliable predictive diagnostic, therapeutic and prognostic biomarkers are still missing, genetic testing has resulted in important findings in recent years.

Genetic variants of TOLLIP and MUC5B are associated with a susceptibility for IPF, outcomes, and perhaps, with therapy responsiveness. Shortening of telomere-length is associated with poor prognosis [15].

Lung microbiome studies provided information on IPF subjects with progressive or stable disease, suggesting that the bacterial burden itself may be important for disease progression [15–17].

Previous transcriptomic studies have mainly been performed using mass tissue. With the emergence of novel technologies, the regulatory and transcriptional networks can examine the lung from a variety of viewpoints using available mass data and profiles of disease and single cell microenvironments. This will enable the integration of information about the lung affected by IPF and thus improve diagnostics and therapy [18]. Subsequent transcriptomic studies

on lung and peripheral blood samples revealed the role of genes involved in alveolar epithelial injury and remodelling in IPF pathogenesis. It has been shown that pathways in IPF are aberrantly stimulated and that different phenotypes vary in their gene expression patterns [15]. Mitochondrial aberrations have been proven to have an effect on the extent of the alterations in microRNA expression and a gene signature consisting of 52 differently expressed genes could effectively categorise patients at high *versus* low mortality risk [15]. Furthermore, the possibility of using transcriptomic data to reconstruct the UIP pattern has recently been established [7]. Thus, the transcriptomic signature, which may also predict the biological response to antifibrotic therapy, could be very promising to implement personalised medicine in IPF [15].

Machine learning

Machine learning is a method of constructing a process or set of rules (an algorithm) learned on a data record that can be used to make accurate predictions on a new sample of data. These techniques have been widely used to overcome medical challenges and have improved our understanding of signalling pathways, significant phenotypes and help to estimate disease risk. In machine learning, large samples (usually several thousand) from medical imaging data and serial data are needed, while clinical trials in IPF often have a smaller sample size because of the difficulty recruiting patients with rare diseases such as IPF.

In 2015, KIM *et al.* [19] used machine learning to develop a genomic data algorithm from surgical biopsy samples to detect a molecular signature for UIP that is concordant with histopathological diagnoses derived from the same tissue.

In this context, very recently RAGHU *et al.* [13] studied the utility and validity of a molecular UIP signature, which can be identified by a machine learning algorithm, in less-invasive TBLB samples.

Methods

To achieve this goal, RAGHU *et al.* [13] prospectively recruited 237 patients for a Novel Genomic Test (BRAVE) study in 29 US and European sites.

Patients underwent assessment for new-onset ILD by clinically indicated diagnostic SLB, TBLB or TBLC. Histopathological UIP pattern (classic, difficult, or favouring classic UIP) or non-UIP pattern were assessed in each patient. After exclusions, diagnostic histopathology and RNA sequence data from 90 patients were used to train a machine learning algorithm (Envisia Genomic Classifier; Veracyte, San Francisco, CA, USA) to identify an UIP pattern [13].

The Envisia genomic classifier uses total RNA extracted from TBLB specimen to run next generation RNA sequencing [20]. The genetic count data of 190 genes are then entered into the classifier, the machine learning algorithm, to generate either a UIP or non-UIP classification result. To ensure that the classifier was trained only on pathological data for UIP, radiology diagnoses were not used as a criterion to include or exclude patients from the study or to inform reference labels [13]. A limitation of this RNA molecular classifier is that it doesn't distinguish among causes of UIP, which can be found in many clinical settings (*e.g.* hypersensitivity pneumonitis, rheumatoid arthritis, drug-induced lung toxicity).

The primary study end-point was the validation of the classifier in 49 patients by comparison with diagnostic histopathology [13]. To assess clinical utility, the authors compared the agreement and confidence level of diagnosis made by central multidisciplinary teams (MDTs) based on anonymised clinical information and radiology results plus either molecular classifier or histopathology results.

A recent study reported on the reproducibility and robustness of the Envisia test that can differentiate in a minimally invasive way between UIP and non-UIP pathologies in TBLB [20]. Genome analysis and machine learning have been shown to increase the usefulness of TBLB for the diagnosis of UIP with higher sensitivity and specificity than the pathology of TBLB itself, and so may be able to inform about the diagnosis of patients with suspected IPF [21].

Results

The results of the molecular classifier interpreted together with longitudinal clinical evaluation and radiological features in a multidisciplinary discussion can help to differentiate and individually treat patients with an ILD [13].

The genome classifier showed a specificity of 88% (95% CI 70–98%) and sensitivity of 70% (95% CI 47–87%) for UIP in 49 patients with a diagnostic biopsy (area under the curve 0.87; 95% CI 0.76–0.98), which was similar between those with SLB and TBLC. For this population with a 47% prevalence of histological UIP, the positive predictive value (PPV) was 84% and the negative predictive value (NPV) was 77%.

When the analysis was repeated on the 42 patients for which SLB was recommended by guidelines, *i.e.* excluding those with a definite HRCT UIP pattern, a PPV of 80% for UIP was found.

When comparing the outcomes of two separate MDTs for 94 patients, one team having access to the molecular classifier and the other to histology results, there was an overall agreement of 86% (95% CI 78–92%), with no differences between patients with or without diagnostic histology. The overall proportion

of cases with a high confidence or confident diagnosis was similar between both MDTs, but was higher for the molecular classifier MDT in the 18 patients with a final diagnosis of IPF. For the 46 patients with a diagnostic histology, there was a higher proportion of confident diagnoses on the histology MDT, while the reverse was observed for those with a non-diagnostic or non-classifiable fibrosis biopsy result.

An analysis of the 190 genes used by the molecular classifier, showed that 124 are among the 1000 genes that are differently expressed in TBLB samples from patients with and without UIP, which supports the biological plausibility of the classifier results.

Discussion

Together, these results suggest that an RNA analysis of TBLB coupled with a machine learning algorithm can identify the UIP histological pattern with a high specificity and sensitivity. This technique can be particularly useful for improving MDT diagnosis of IPF in patients without a definitive UIP pattern on HRCT, thus preventing the need for SLB.

Some of the limitations of this study are that only patients that underwent a biopsy were included, and those patients declining or not eligible for biopsy are not represented. In addition, the authors acknowledge that some patients had a diagnosis based on cryobiopsy or TBLB, which is currently not recommended by IPF guidelines. However, cryobiopsy has been shown to be useful for the diagnosis of ILD on the recent COLDICE study [22] and it is unclear whether a TBLB-based molecular classifier would overrule results of cryobiopsy and have less potential adverse effects.

The results of this study are also difficult to apply to those patients with an HRCT pattern suggestive of an alternative diagnosis. Specifically, hypersensitivity pneumonitis (HP), which was suspected on the HRCT for 15 patients out of whom the classifier resulted in a UIP result for seven, is a subject for further study. Chronic HP patients may show a biopsy pattern of UIP [23], which is associated with a worse prognosis. The disease behaviour is also considered on MDTs and can help identify phenotypes, such as progressive fibrosing ILD [24]. The authors are planning to perform further prospective studies on the molecular classifier that include progression and outcomes as end-points.

Summary

A molecular classifier using a machine-learning algorithm based on genomic data could provide an objective method to help clinicians and MDTs to establish the diagnosis of IPF in less-invasive TBLB, particularly for patients without a clear radiological diagnosis.

Key points

- An RNA analysis of TBLBs coupled with a machine learning algorithm can identify a UIP histological pattern with a high specificity and sensitivity.
- This technique can be particularly useful for improving multidisciplinary diagnosis of IPF on patients without a definitive UIP pattern on HRCT, thus preventing the need for SLB.

Affiliations

Andrea Crespo¹, Tiago Alfaro², Vivien Somogyi^{3,4}, Michael Kreuter³

¹Pneumology Service, Rio Hortega University Hospital, Valladolid, Spain. ²Centre of Pneumology, Faculty of Medicine, University of Coimbra, Coimbra, Portugal. ³Center for Interstitial and Rare Lung Diseases, Thoraxklinik, University of Heidelberg, Germany and German Center for Lung Research (DZL), Heidelberg, Germany. ⁴Dept of Pulmonology, Semmelweis University, Budapest, Hungary.

Conflict of interest

None declared

References

1. Raghu G, Remy-Jardin M, Myers JL, *et al.* Diagnosis of idiopathic pulmonary fibrosis. An Official ATS/ERS/JRS/ALAT clinical practice guideline. *Am J Respir Crit Care Med* 2018; 198: e44–e68.
2. Hutchinson J, Fogarty A, Hubbard R, *et al.* Global incidence and mortality of idiopathic pulmonary fibrosis: a systematic review. *Eur Respir J* 2015; 46: 795–806.
3. Kreuter M, Swigris J, Pittrow D, *et al.* The clinical course of idiopathic pulmonary fibrosis and its association to quality of life over time: longitudinal data from the INSIGHTS-IPF registry. *Respir Res* 2019; 20: 59.
4. Fernandez Perez ER, Daniels CE, Schroeder DR, *et al.* Incidence, prevalence, and clinical course of idiopathic pulmonary fibrosis: a population-based study. *Chest* 2010; 137: 129–137.
5. 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.
6. 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.
7. Lynch DA, Sverzellati N, Travis WD, *et al.* Diagnostic criteria for idiopathic pulmonary fibrosis: a Fleischner Society White Paper. *Lancet Respir Med* 2018; 6: 138–153.
8. Hunninghake GW, Zimmerman MB, Schwartz DA, *et al.* Utility of a lung biopsy for the diagnosis of idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2001; 164: 193–196.
9. Tomassetti S, Wells AU, Costabel U, *et al.* Bronchoscopic lung cryobiopsy increases diagnostic confidence in the multidisciplinary diagnosis of idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2016; 193: 745–752.
10. Flaherty KR, King TE, Raghu G, *et al.* Idiopathic interstitial pneumonia: what is the effect of a multidisciplinary approach to diagnosis? *Am J Respir Crit Care Med* 2004; 170: 904–910.
11. Hetzel J, Maldonado F, Ravaglia C, *et al.* Transbronchial cryobiopsies for the diagnosis of diffuse parenchymal lung diseases: expert statement from the cryobiopsy working group on safety and utility and a call for standardization of the procedure. *Respiration* 2018; 95: 188–200.
12. Pesci A, Ricchiuti E, Ruggiero R, *et al.* Bronchoalveolar lavage in idiopathic pulmonary fibrosis: what does it tell us? *Respir Med* 2010; 104: Suppl. 1, S70–S73.
13. Raghu G, Flaherty KR, Lederer DJ, *et al.* Use of a molecular classifier to identify usual interstitial pneumonia in conventional transbronchial lung biopsy samples: a prospective validation study. *Lancet Respir Med* 2019; 7: 487–496.
14. 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.
15. Newton CA, Molyneaux PL, Oldham JM. Clinical genetics in interstitial lung disease. *Front Med (Lausanne)* 2018; 5: 116.
16. Wang J, Lesko M, Badri MH, *et al.* Lung microbiome and host immune tone in subjects with idiopathic pulmonary fibrosis treated with inhaled interferon-gamma. *ERJ Open Res* 2017; 3: 00008-2017.
17. 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.
18. Vukmirovic M, Kaminski N. Impact of transcriptomics on our understanding of pulmonary fibrosis. *Front Med (Lausanne)* 2018; 5: 87.
19. 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.
20. Choi Y, Lu J, Hu Z, *et al.* Analytical performance of Envisia: a genomic classifier for usual interstitial pneumonia. *BMC Pulm Med* 2017; 17: 141.
21. Pankratz DG, Choi Y, Imtiaz U, *et al.* Usual interstitial pneumonia can be detected in transbronchial biopsies using machine learning. *Ann Am Thorac Soc* 2017; 14: 1646–1654.
22. Troy L, Grainge C, Corte TJ, *et al.* Transbronchial lung cryobiopsy for interstitial lung disease diagnosis: results of the COLDICE Study. *Eur Respir J* 2019; 54: Suppl. 63, RCT1886.
23. Wang P, Jones KD, Urisman A, *et al.* Pathologic findings and prognosis in a large prospective cohort of chronic hypersensitivity pneumonitis. *Chest* 2017; 152: 502–509.
24. Wong AW, Ryerson CJ, Guler SA. Progression of fibrosing interstitial lung disease. *Respir Res* 2020; 21: 32.