

ORIGINAL ARTICLE

Multi-omics integration reveals a nonlinear signature that precedes progression of lung fibrosis

Céline Pattaroni^{1,a}  , Christina Begka^{1,a}, Bailey Cardwell¹, Jade Jaffar¹, Matthew Macowan¹, Nicola L Harris¹, Glen P Westall^{1,2,b} & Benjamin J Marsland^{1,b}

¹Department of Immunology, School of Translational Medicine, Monash University, Melbourne, VIC, Australia

²Department of Respiratory Medicine, Alfred Hospital, Melbourne, VIC, Australia

Correspondence

C Pattaroni, Department of Immunology,
School of Translational Medicine, Monash
University, Melbourne, VIC, Australia
E-mail: celine.pattaroni@monash.edu

^aThese authors contributed equally to
this work as joint first authors.

^bJoint last authors.

Received 28 August 2023;

Revised 11 December 2023

and 9 January 2024;

Accepted 9 January 2024

doi: 10.1002/cti2.1485

Clinical & Translational Immunology

2024; 13: e1485

Abstract

Objectives. Idiopathic pulmonary fibrosis (IPF) is a devastating progressive interstitial lung disease with poor outcomes. While decades of research have shed light on pathophysiological mechanisms associated with the disease, our understanding of the early molecular events driving IPF and its progression is limited. With this study, we aimed to model the leading edge of fibrosis using a data-driven approach. **Methods.** Multiple omics modalities (transcriptomics, metabolomics and lipidomics) of healthy and IPF lung explants representing different stages of fibrosis were combined using an unbiased approach. Multi-Omics Factor Analysis of datasets revealed latent factors specifically linked with established fibrotic disease (Factor1) and disease progression (Factor2). **Results.** Features characterising Factor1 comprised well-established hallmarks of fibrotic disease such as defects in surfactant, epithelial–mesenchymal transition, extracellular matrix deposition, mitochondrial dysfunction and purine metabolism. Comparatively, Factor2 identified a signature revealing a nonlinear trajectory towards disease progression. Molecular features characterising Factor2 included genes related to transcriptional regulation of cell differentiation, ciliogenesis and a subset of lipids from the endocannabinoid class. Machine learning models, trained upon the top transcriptomics features of each factor, accurately predicted disease status and progression when tested on two independent datasets. **Conclusion.** This multi-omics integrative approach has revealed a unique signature which may represent the inflection point in disease progression, representing a promising avenue for the identification of therapeutic targets aimed at addressing the progressive nature of the disease.

Keywords: disease progression, lipidomics, metabolomics, multi-omics, pulmonary fibrosis, transcriptomics

INTRODUCTION

Idiopathic pulmonary fibrosis (IPF) is a severe lung disease of unknown aetiology characterised by

progressive lung scarring. Although the pathophysiological mechanisms associated with the end stages of the disease are well established, the early events preceding the first symptoms and

diagnosis of the disease remain unclear. The current postulate is that repeated micro-injuries to the lung epithelium cause a dysfunctional phenotype leading to an aberrant wound healing process.¹ While an aberrant epithelial phenotype is a central feature of the disease, a paradigm shift is needed in the way we consider IPF pathophysiology. Indeed, the mechanisms responsible for disease initiation (e.g. epithelial cell injury) may differ from those of disease progression. The complexity of the relationship between risk factors,^{2,3} timing and activation of interconnected cellular pathways prompts us to move away from the current reductionist view of perceiving IPF pathogenesis as gradual and linear. In fact, it has recently been proposed that the aberrant activation of pathways in IPF might follow a chaotic dynamic pattern, with small changes in baseline conditions being sufficient to trigger a long-term fibrotic response.⁴

High-throughput molecular profiling technologies have allowed for unprecedented characterisation of human lung diseases using omics datasets, including transcriptomics, metabolomics and lipidomics, among others.⁵ Transcriptomics has particularly expanded our understanding of IPF pathophysiology by revealing gene expression patterns associated with fibrosis such as matrix metalloproteinases (MMPs), activation of (myo)fibroblasts, WNT signalling, host defence and metabolism.⁶ Meanwhile, metabolic characterisation of the lung microenvironment has shown alterations in amino acid (e.g. related to arginine)^{7–9} and glucose metabolism (e.g. shift towards lactic acid production), as well as changes in the beta-oxidation pathway involving fatty acids (FA).¹⁰ Furthermore, lipids such as S1P and LPA may also contribute to IPF development.¹¹ Multi-omics studies aiming to simultaneously interrogate different layers of biomolecules have started to emerge, revealing the power of combining data modalities. However, compared with other respiratory diseases such as COPD/asthma or in the context of lung transplantation,¹² integrative multi-omics studies aiming to investigate IPF are scarce¹³ and often limited to animal models.¹⁴

In IPF, tissue remodelling typically begins at the basal posterior edges of the lungs and progresses apically, creating a unique opportunity to study disease progression by sampling different areas of the lung. Two studies^{15,16} have revealed that macroscopically normal IPF tissue exhibits notable gene expression differences when compared to tissue from non-IPF individuals. Both studies identified sets of genes with altered expression levels preceding the

macroscopic evidence of fibrosis. Notably, these two previous investigations exclusively focused on gene expression. With this study, we aimed to model the leading edge of fibrosis using an unbiased multi-omics data-driven approach on structurally normal, pre-fibrotic and fibrotic lung explant samples from IPF patients and non-diseased controls (NDC).

RESULTS

Idiopathic pulmonary fibrosis lung explants exhibit distinct molecular profiles when compared to controls

Three omics layers from lung explants taken from patients with IPF ($n = 6$) and NDC ($n = 5$) were simultaneously investigated. Patients' characteristics and key sample metadata are presented in Supplementary table 1. Transcriptomics (RNA sequencing), as well as untargeted metabolomics and lipidomics (LCMS) datasets, were generated for multi-omics data integration and machine learning (Figure 1a). For each lung explant, a sample from the base (fibrotic, late-stage of IPF) and a sample from the apex (pre-fibrotic, early stage of IPF) were collected. In order to characterise the extent of fibrosis prior to molecular profiling, histological grading was performed using the Ashcroft scale (Figure 1b). Ashcroft scores from NDC samples were significantly lower than their topological IPF counterparts (Wilcoxon test FDR-corrected P -value $< 0.01^{**}$ for apex and $< 0.01^{**}$ for base), suggesting that fibrosis has already been initiated in IPF apices (representative histology slides presented in Supplementary figure 1). Differences in IPF apex and IPF base scores were variable (Wilcoxon test FDR-corrected P -value = 0.1). Principal component analysis (PCA) of normalised transcriptomics (Figure 1c), metabolomics (Figure 1d) and lipidomics (Figure 1e) completely separated IPF from NDC samples. Standard differential expression (transcriptomics) and intensity (metabolomics, lipidomics) testing on individual omics revealed marked differences with disease with 36%, 39% and 39% of the total features showing differential abundance in the transcriptomics, metabolomics and lipidomics datasets, respectively (Supplementary figure 2a, b). Differential gene expression patterns between the apex and base of IPF samples were observed (Supplementary figure 2c), a distinction not observed in NDC samples or metabolomics and lipidomics analysis. In

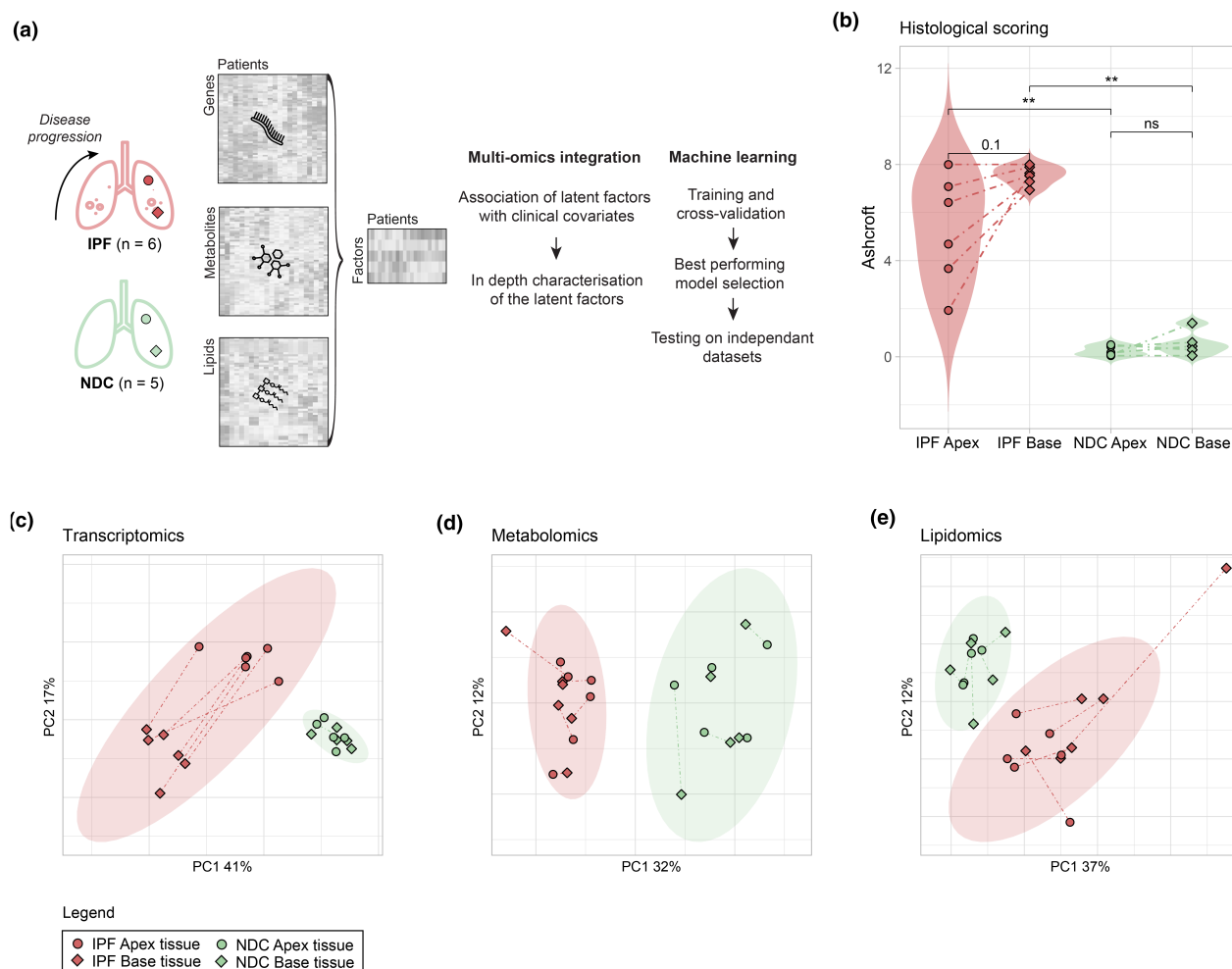


Figure 1. Study design and overall multi-omics data variance. **(a)** Multi-omics data (transcriptomics, metabolomics and lipidomics) from idiopathic pulmonary fibrosis (IPF) and non-diseased controls (NDC) apex and base lung samples were integrated in an unsupervised fashion. Factors correlating with clinical metadata were investigated and used as input for machine learning and validation on an independent dataset. **(b)** Violin plots of lung tissues histological scoring (Ashcroft score) comparing disease groups and lung tissue location. Principal component analysis (PCA) of normalised transcriptomics **(c)**, metabolomics **(d)** and lipidomics **(e)** data with ellipses representing the 95% confidence intervals for each group. Colours are representative of disease status with dashed lines linking the apex and base samples obtained from the same individual. Sample sizes are 6 IPF apex, 6 IPF base, 5 NDC apex and 5 NDC base. Statistics represent the result of FDR-corrected non-parametric Wilcoxon Rank Sum testing with FDR < 0.05, < 0.01 and < 0.001 represented as *, ** and ***, respectively.

addition, differential expression testing between apex samples from IPF against those from NDC revealed an upregulation of prototypical fibrotic remodelling genes, including MMPs and collagens, within the IPF apex (Supplementary figure 2d) indicating strong changes to cellular programmes before overt evidence of fibrosis. Given the marked differences observed by PCA and standard differential testing, we hypothesised that the integration of these multi-omics datasets, combined with the topological sampling reflective of different disease stages, could reveal molecular drivers underpinning progression of pulmonary fibrosis.

Multi-omics analysis identifies factors correlating with established disease and progression

We next aimed to integrate the different molecular layers characterising lung explant samples to uncover sources of variation that explain the progressive nature of the disease. We used Multi-Omics Factor Analysis (MOFA),¹⁷ a statistical framework enabling the discovery of latent factors capturing the major sources of variation across omics datasets in an unsupervised fashion. MOFA identified five latent factors that cumulatively

explained a significant proportion of variation in the transcriptomics (66%), metabolomics (36%) and lipidomics (37%) layers (Figure 2a). Factor1 explained most of the variance (R^2) and captured comparable proportions of variability across the different datasets (38% transcriptomics, 26% metabolomics and 25% lipidomics), which was also the case for Factor3 and 4 (Figure 2b). In contrast, Factor2 was mostly represented by the transcriptomics layer (16%) followed by lipidomics (4%). We next investigated the correlation between these factors and clinical variables and found that Factor1 was strongly associated with disease (Pearson correlation 0.93 P -value $< 0.001^{***}$), Ashcroft histological score (Pearson correlation 0.93 P -value $< 0.001^{***}$) and age (Pearson correlation 0.67 P -value $< 0.001^{***}$) (Figure 2c). Notably, Factor2 correlated with the lung tissue topographical location (apex versus base) (Pearson correlation 0.66 P -value $< 0.001^{***}$). Other significant correlations included smoking and BMI parameters with less pronounced correlation coefficients (Pearson correlation < 0.5). Differences in Factor1 values were further investigated by comparing disease and topological sampling subgroups. Factor1 completely separated samples from IPF and NDC (Wilcoxon test FDR-corrected P -value $< 0.1^{**}$) and significantly differed between IPF apex and base samples (Wilcoxon test FDR-corrected P -value $< 0.05^*$) (Figure 2d). Notably, when compared to histological scoring (Figure 1b), Factor1 values were able to better discriminate between IPF apex versus base. Marked differences between IPF apex and base samples were also observed for Factor2 (Wilcoxon test FDR-corrected P -value $< 0.1^{**}$) with no distribution overlap (Figure 2e). When considering topographical sampling as a proxy for the different stages of IPF, Factor1 values increased linearly with disease progression (Figure 2f), while a nonlinear U-shaped relationship was observed for Factor2 (Figure 2g). Altogether, the unbiased integration of multiple omics layers revealed key sources of co-variation captured by factors associated with fibrosis establishment (Factor1) and progression (Factor2) following distinct linear and nonlinear patterns, respectively.

Multi-omics Factor1 top features highlight distinct molecular pathways linked with established fibrosis

To investigate the molecular aetiology of the first two factors, we used the top 50 MOFA factors

loadings of each omics (Supplementary table 2). Maximum distance clustering of MOFA's Factor1 top genes separated samples based on disease status (IPF versus NDC) (Figure 3a). Gene pathways upregulated with disease included extracellular matrix (ECM) synthesis and deposition (*COL7A1*, *COL14A1*, *COL15A1*, *FNDC1*, *PI15*, *CTHRC1* and *THY1*), epithelial–mesenchymal (EMT) transition (*SFRP4*, *CPXM2* and *VEPH1*), and a profibrotic TGF β isoform (*TGFB3*). Comparatively, gene expression pathways increased in NDC samples were related to surfactant homeostasis (*SFTPA1*, *SFTPA2*, *SFTPD*, *SFTA2*, *ABCA3*, *MFSD2A*, *LAMP3*, *NAPSA*, *PLA2G4F* and *CACNA2D2*), lipid homeostasis and biosynthesis (*FASN*, *PCSK9* and *SCD*), EMT (*GKN2*) and TGF β pathway inhibition (*SMAD6*). Interestingly, a subset of genes (noted with 'm') characterising the first factor were linked to mitochondrial function (*SLC39A8*, *PLA2G4F*, *CACNA2D2*, *FASN*, *DIO2* and *PCSK9*) and predominantly downregulated in IPF samples. Similar to gene expression, maximum distance clustering of the top metabolites explaining Factor1's variation demonstrated a clear separation of samples based on disease status (Figure 3b). Differences in amino acid or derivatives were observed, either increased (e.g. citrulline, pyroglutamylglycine, N(6)-Methyllysine, gamma-Glutamylalanine and N-lactoyl-Tryptophan) or decreased (e.g. gabapentin and glutarylglutamine). Specifically, increased levels of glutathione metabolism intermediates (gamma-glutamylalanine, pyroglutamylglycine and L-cystine) and decreased levels of precursors (L-cystathionine and glutarylglutamine) and glutathione itself were observed in IPF tissues. The hypotaurine/taurine biosynthesis pathway, which partially overlaps with GSH metabolism, showed increased levels of several members in NDC samples (L-cystathionine, 3-sulfinolalanine and hypotaurine), except for L-Cystine, which was increased in IPF lung explants. Purine pathway members were among the top features for Factor1 (Figure 3b blue box). Adenine and adenosine were increased in NDC samples, while downstream molecules (xanthine, xanthosine and uric acid) were increased in IPF tissues. Finally, investigation of the top lipidomics features characterising Factor1 revealed a similar trend as the metabolomics dataset; a clear separation between IPF and NDC samples (Figure 3c). The majority of Factor1's top lipids were found to be increased in NDC samples, with 14 of them belonging to the phosphatidylglycerol (PG) family, the second most abundant surfactant

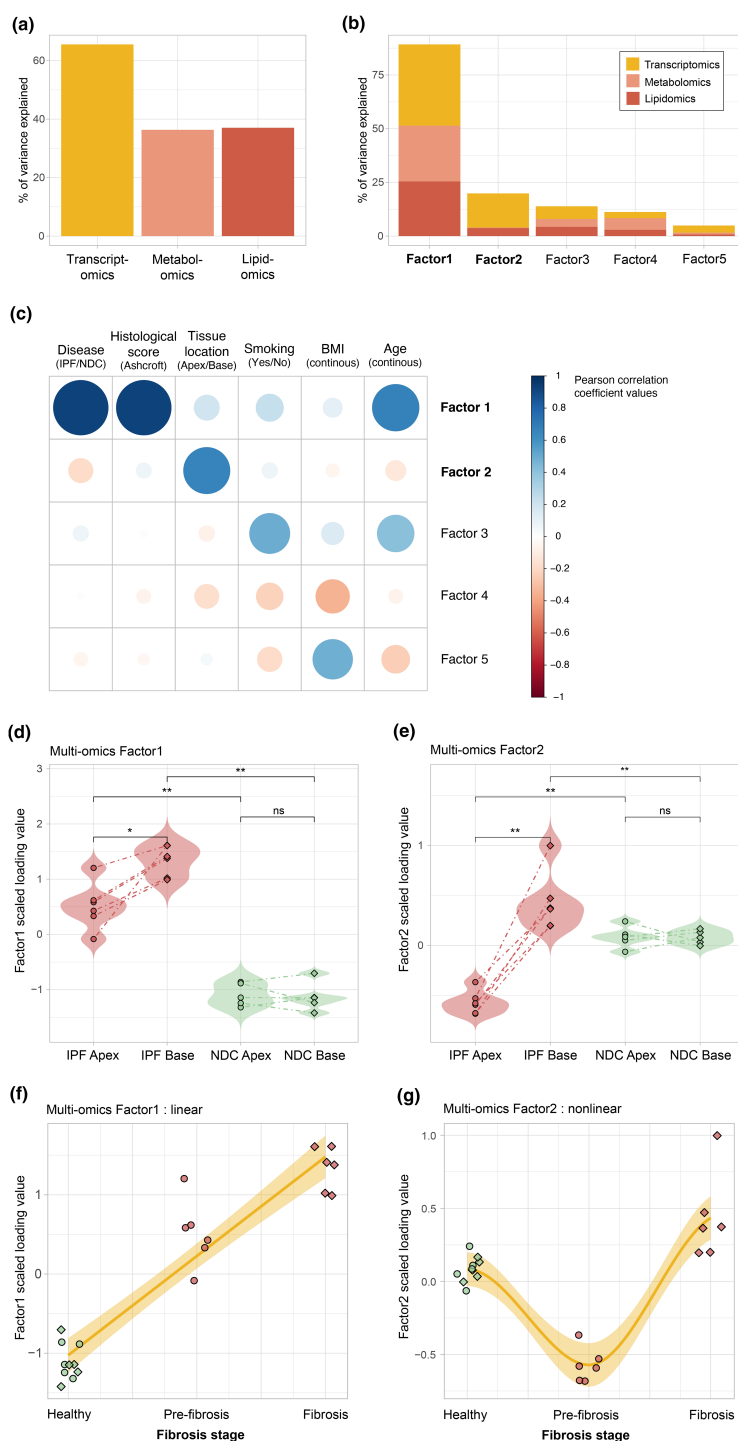


Figure 2. Unsupervised multi-omics data integration using Multi-Omics Factor Analysis (MOFA) and association with clinical covariates. **(a)** Cumulative proportions of total variance explained by each omics dataset. **(b)** Total variance explained by each omic view by individual factors. **(c)** Correlations between MOFA factors and clinical covariates with significant Pearson correlation coefficients (P -value < 0.05) represented in a colour scale from red (negative correlations) to blue (positive correlations). **(d)** Violin plots of MOFA Factor1 values comparing disease groups and lung tissue location. **(e)** Corresponding violin plots for MOFA Factor2 values. **(f)** Factor1 values in relation to fibrosis stage proxy (non-diseased controls (NDC), idiopathic pulmonary fibrosis (IPF) apex and IPF base) with lm fitting. **(g)** Factor2 values in relation to fibrosis stage proxy (NDC, IPF apex and IPF base) with loess curve fitting. Sample sizes are 6 IPF apex, 6 IPF base, 5 NDC apex and 5 NDC base. Statistics represent the result of FDR-corrected non-parametric Wilcoxon rank-sum testing with FDR < 0.05 , < 0.01 and < 0.001 represented as *, ** and ***, respectively.

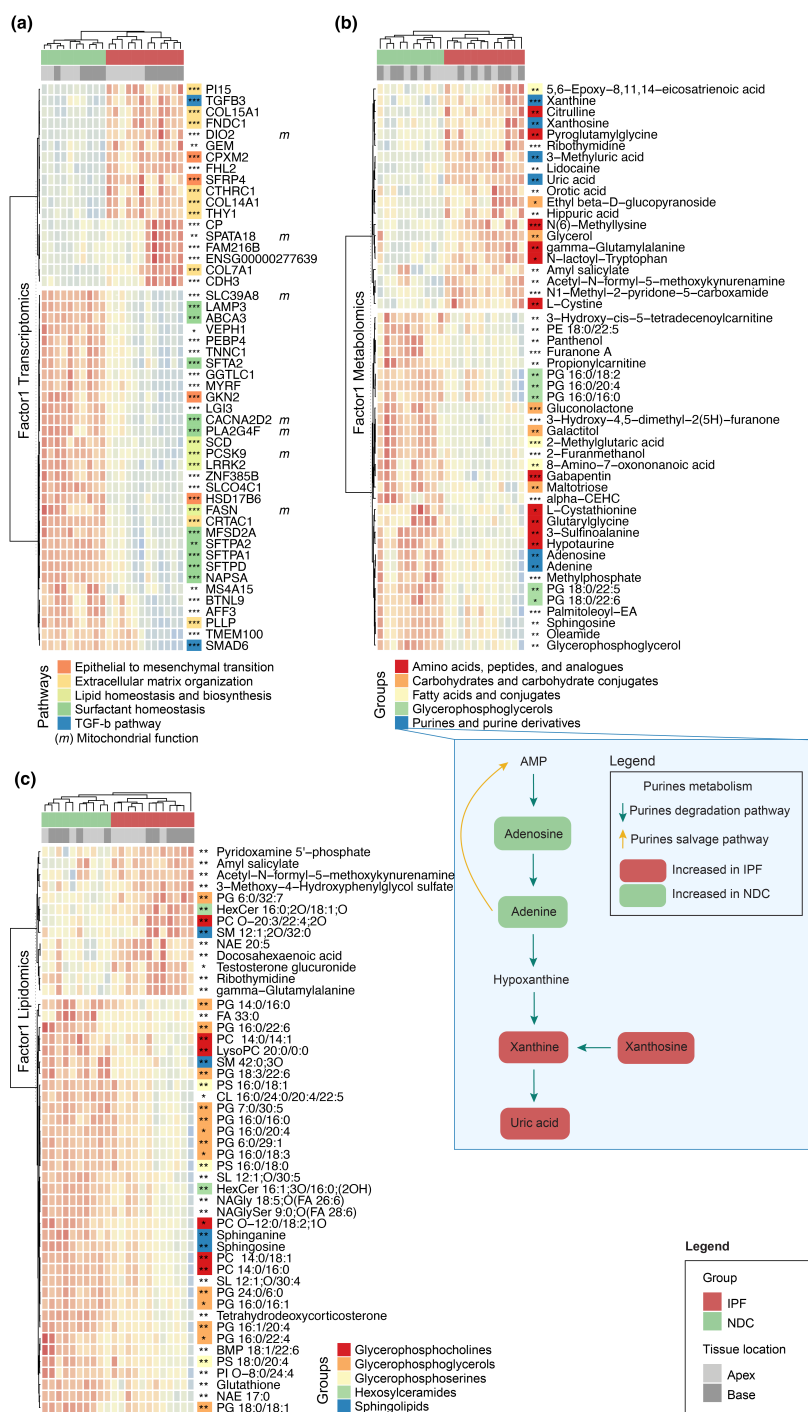


Figure 3. In-depth characterisation of Factor1. **(a)** Heatmap representing the top 50 genes with the highest loading values for Factor1, associated with disease and age. Each row represents a gene (z-score of normalised gene expression) and each column a sample. **(b)** Corresponding heatmap for metabolites. Each row represents a metabolite (z-score of normalised metabolites intensities). Purines metabolism pathway (degradation and salvage) with metabolites increased in non-diseased controls (NDC) (green) or idiopathic pulmonary fibrosis (IPF) (red) samples is highlighted in the orange box. **(c)** Corresponding heatmap for lipids. Each row represents a lipid (z-score of normalised lipids intensities). Column annotations depict disease groups and lung tissue locations. Row annotations refer to literature-based pathways for transcriptomics and metabolomics, and broad lipid classes for lipidomics. Sample sizes are 6 IPF apex, 6 IPF base, 5 NDC apex and 5 NDC base. Statistics represent the result of FDR-corrected limma differential gene expression testing for transcriptomics and lme4 for metabolomics and lipidomics with FDR < 0.05, < 0.01 and < 0.001 represented as *, ** and ***, respectively.

phospholipid.^{18,19} Other increased lipids in NDC samples included glycerophosphocholines, glycerophosphoserines and sphingolipids. In contrast, two hexosylceramides lipids family members were increased in IPF tissues. The antioxidant glutathione was increased in NDC explants. Altogether, investigation of the top features of Factor1 revealed well-established fibrosis pathways (ECM, EMT and TGF β), in some cases shared across multiple omics layers (e.g. surfactant in both the transcriptomics and lipidomics layers).

Multi-omics Factor2 top features reveal molecular pathways specific to disease progression

Investigation of the top features explaining Factor2 revealed pathways differentiating IPF apex (pre-fibrosis) from base (fibrosis) samples. Two gene expression patterns were identified: (1) increased expression only in IPF base samples (Figure 4a bottom) and (2) increased expression in IPF base but decreased in IPF apex samples with intermediate levels in NDC samples (Figure 4a

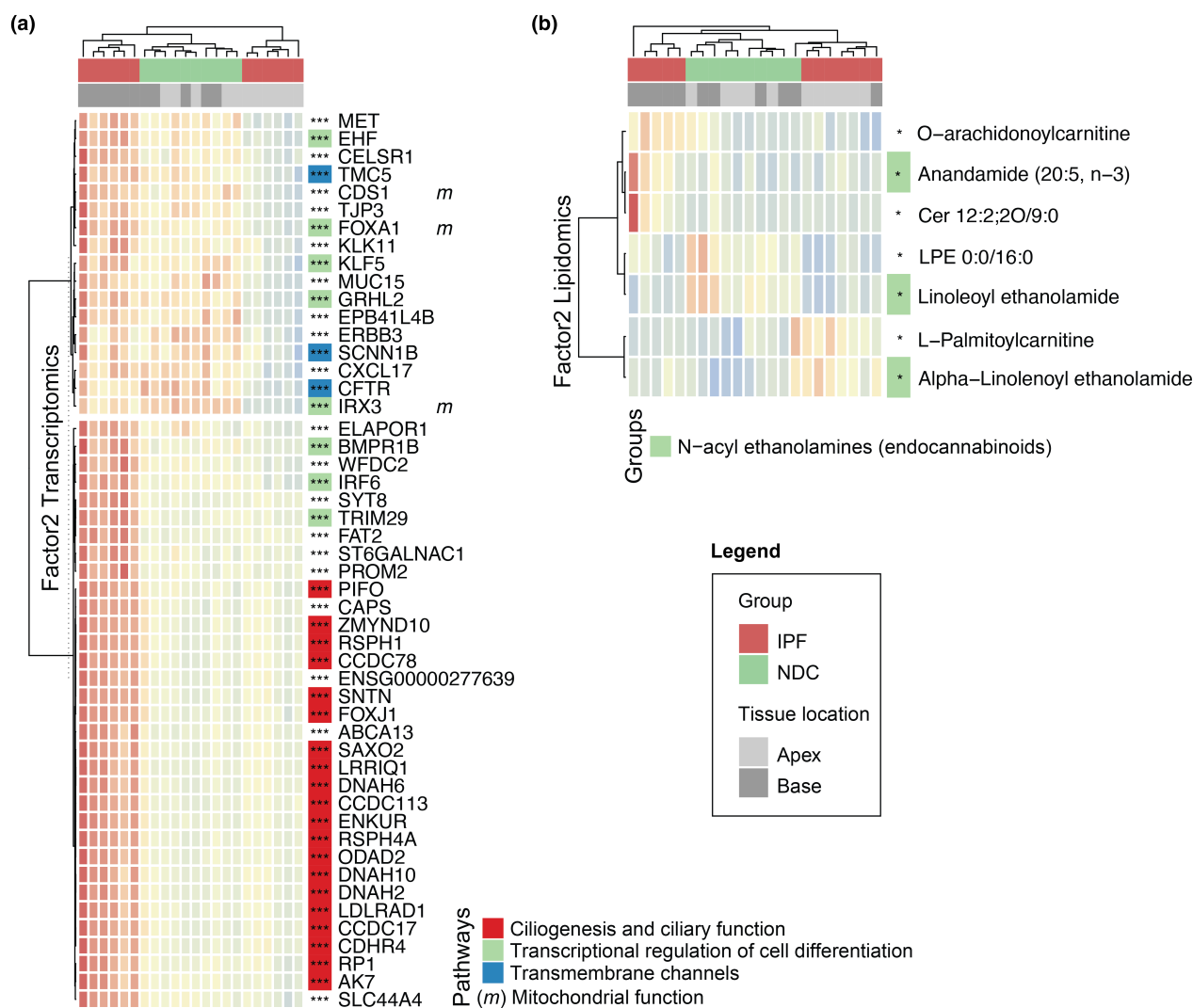


Figure 4. In-depth characterisation of Factor2. **(a)** Heatmap representing the top 50 genes with the highest loading values for Factor2, associated with disease progression. Each row represents a gene (z-score of normalised gene expression) and each column a sample. **(b)** Corresponding heatmap for lipids. Each row represents a lipid (z-score of normalised lipids intensities). Column annotations depict disease groups and lung tissue locations. Row annotations refer to literature-based pathways for transcriptomics and broad lipid classes for lipidomics. Sample sizes are 6 IPF apex, 6 IPF base, 5 NDC apex and 5 NDC base. Statistics represent the result of FDR-corrected limma differential gene expression testing for transcriptomics and lme4 for metabolomics and lipidomics with FDR < 0.05, < 0.01 and < 0.001 represented as *, ** and ***, respectively.

top). The second pattern explains the nonlinear pattern observed in Factor2. The first gene expression pattern found in IPF base samples was associated with ciliogenesis, including *FOXJ1* and genes involved in cilia structure (DNAH genes), motility (*SAXO2* and *SNTN*) and mucociliary clearance (*ODA2*, *LDLRAD1*, *RSPH4A*, among others). In contrast, the second pattern showed reduced expression of genes involved in cell differentiation regulation (*TRIM29*, *IRF6*, *EHF*, *IRX3*, among others) and transmembrane channel genes (*TMC5*, *CFTR* and *SCNN1B*) in pre-fibrotic IPF apex samples. Many of these genes play a role in embryogenesis (*FOXA1*, *MET*, *CELSR1*, *IRX3*, *KLF5*, *GRHL2* and *ERBB3*) and mitochondrial function (*FOXA1*, *CDS1* and *IRX3*). Top lipidomics features for Factor2 also revealed IPF progression-specific features, separating IPF bases and apices (Figure 4b). IPF bases showed elevated levels of long-chain acylcarnitines and O-arachidonoylcarnitine, and IPF apices had higher L-Palmitoylcarnitine, with NDC samples exhibiting low levels of these compounds. A notable pattern linked to endocannabinoid signalling was observed, varying with the disease progression status. Anandamide, which transforms into pro-inflammatory arachidonic acid via fatty acid amide hydrolase (FAAH), was increased in IPF bases. Conversely, linoleoylethanolamide (LEA), a natural FAAH inhibitor²⁰ reducing arachidonic acid production, was elevated in NDC tissues but low in IPF samples. Of note, α -Linoleoylethanolamide (α -LEA) levels were increased in IPF apices. In summary, fibrosis progression was associated with ciliogenesis, transcriptional regulation of cell differentiation and discrete changes in lipid profiles.

Machine learning algorithms predict fibrotic and pre-fibrotic signatures in independent datasets

Given the importance of the transcriptomics layer in MOFA-discovered Factor1 and Factor2, we evaluated the predictive performance of their top genes using machine learning (ML). To ensure accurate ML model training and performance estimation, the presented transcriptomics dataset was integrated with a larger published dataset from the same patient cohort.²¹ Four ML models were trained and compared using a leave-one-group-out cross-validation (LOGCV) approach. To validate the best-performing ML models, two

publicly available transcriptomics datasets from GEO archives that utilised differential site sampling to investigate IPF were reprocessed.^{15,16} These included macroscopic sampling of visually appearing scarred and normal regions of tissue¹⁵ and histological classification of samples from the apex, middle and base of explanted IPF lungs into stages of IPF progression¹⁶ (IPF1 = early-stage fibrosis, IPF2 = mid-stage fibrosis, IPF3 = end-stage fibrosis) alongside NDC samples (Figure 5a). The disease classification performance (Factor1) was first evaluated. All the ML models showed strong disease-discriminating performance properties, achieving an Area Under the Curve AUC of 1 for all tested models (Figure 5b). The k-nearest neighbours model had the highest mean accuracy (ACC: 0.99), the lowest false-positive rate (FPR: 0) and a true-positive rate of 1 (TPR: 1). Feature importance evaluation of the latter revealed genes with high discriminating power, including fibronectin gene *FNDC1*, followed by *VEPH1*, collagen genes *COL14A1* and *COL15A1*, surfactant protein genes *SFTPA1* and *SFTPA2*, among others (Figure 5c). The same model was then used to compare disease prediction in two independent datasets, generating a perfect confusion matrix observed for the first dataset (Figure 5d) and highly discerning classification for the other (Figure 5e). The prediction of disease progression using Factor2 top genes was evaluated next, categorising IPF samples into two classes based on the degree of fibrosis: early fibrosis (IPF apex) and late-stage fibrosis (IPF base). The stochastic gradient boosting ML model performed marginally better than other ML models with an AUC of 0.85, ACC of 0.76, FPR of 0.16 and TPR of 0.78 (Figure 5f). Feature importance evaluation revealed the transcription factor *IRX3* to be the most important for distinguishing between base and apex samples. Other strong predictors among Factor2 included several genes involved in epithelial functions (*PROM2*, *WFDC2*, *CFTR*, among others) (Figure 5g). External dataset validation showed concordant classification of Luzina et al. samples, indicating similarity between their 'normal' samples and the current study's apex samples (Figure 5h). McDonough et al. samples exhibited more variability (Figure 5i), with 'mid' and 'late-stage' classifications aligning with IPF base prediction, while some early-stage (IPF1) samples were split between base and apex predictions, potentially indicating early cellular dysfunction. Overall, validation of the MOFA-identified genes on

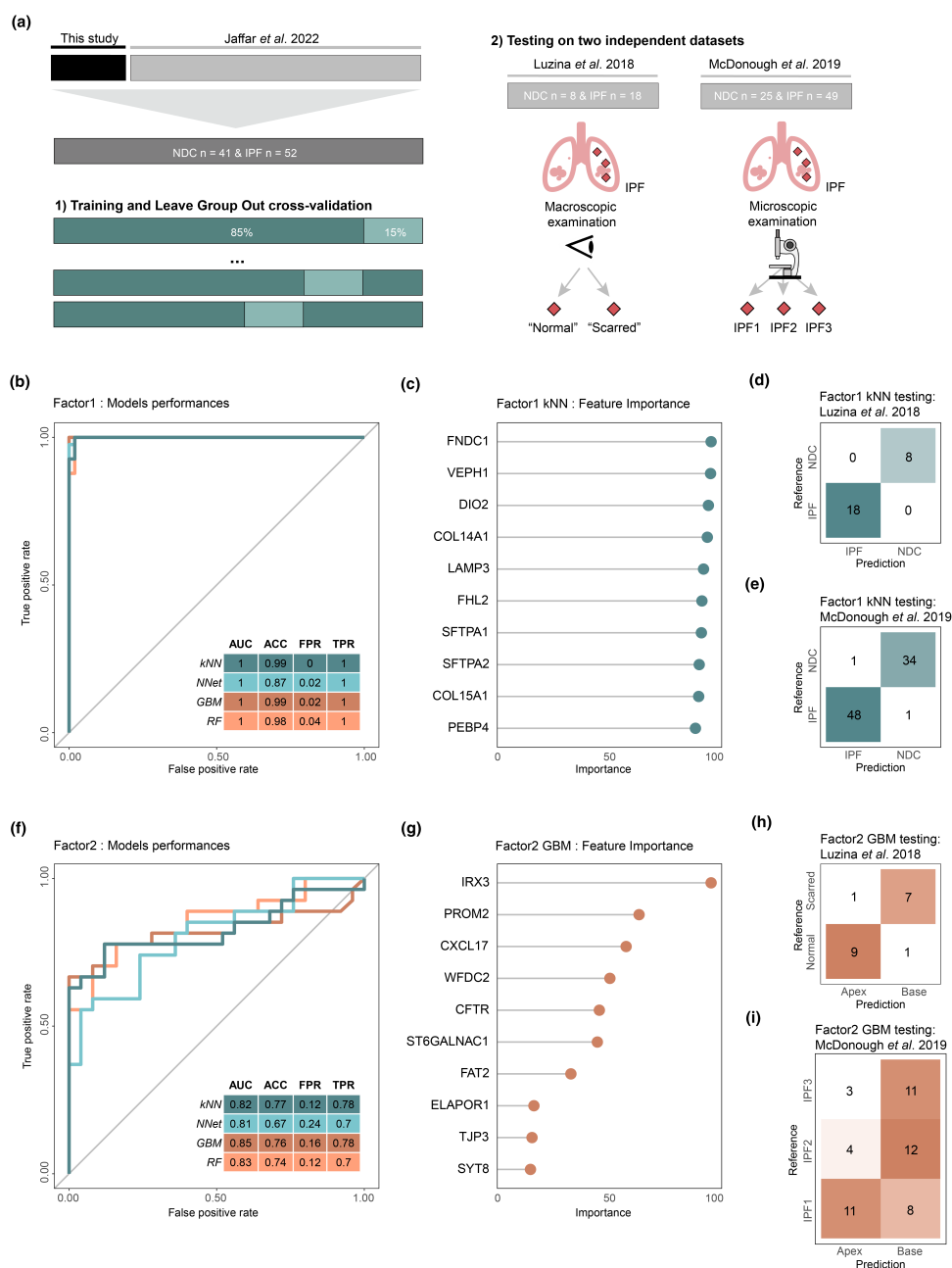


Figure 5. Machine learning and external validation using the transcriptomics data modality. **(a)** Workflow for machine learning models training and testing strategy using the current dataset with previously published data using a leave-one-group-out cross-validation (LOGCV) approach (left) followed by testing of the best-performing model on two independent testing datasets (right). The four machine learning algorithms used for both Factor1 and Factor2 were k-nearest neighbours (kNN), neural network (NNet), stochastic gradient boosting (GBM) and random forest (RF). The following metrics were used to assess each model's prediction capability: area under the receiver operating characteristic (ROC) curve (AUC), accuracy (ACC), false-positive rate (FPR) and true-positive rate (TPR). **(b)** ROC curves depicting the FPR against the TPR for the four different algorithms in discriminating disease status (idiopathic pulmonary fibrosis (IPF) versus non-diseased controls (NDC)) using the top 50 Factor1 genes as input. **(c)** Scaled features importance (genes) for the kNN classification model for Factor1. **(d, e)** Confusion matrices for the independent testing datasets using the kNN classification model for Factor1. **(f)** ROC curves for the four different machine learning algorithms for predicting IPF disease progression (IPF base versus IPF apex) using the top 50 Factor2 genes as input. **(g)** Scaled features importance (genes) for the stochastic gradient boosting classification model for Factor2. **(h, i)** Confusion matrices for the independent testing datasets using the stochastic gradient boosting classification model for Factor2. Sample sizes are **(b)** 41 NDC, 52 IPF; **(f)** 26 base, 26 apex; **(d)** 8 NDC, 18 IPF; **(e)** 35 NDC, 49 IPF; **(h)** 8 scarred IPF, 10 normal IPF; **(i)** 19 IPF1, 16 IPF2, 14 IPF3.

external datasets demonstrated their generalisability beyond the initial discovery dataset.

DISCUSSION

Using an unbiased multi-omics approach, we aimed to model the leading edge of fibrosis to identify novel therapeutic targets. MOFA identified five sources of co-variation, revealing molecular signals that might have been missed with single omic analysis. Factor1 represented variability that distinguished IPF from NDC tissues, while gradually and linearly increasing with disease stage in addition to correlating with histological scoring and age. The most important findings from the current study, however, revolve around Factor2, which discriminated between early (IPF apex) and later (IPF base) stages of IPF, following an unconventional nonlinear dynamic. Unlike Factor1, Factor2 was not associated with age and revealed pathophysiological events which may be driving a pre-fibrotic 'at risk' microenvironment towards pathological tissue remodelling.

In-depth characterisation of the molecular aetiology of Factor1 revealed a downregulation of surfactant signals observed in IPF in both the cellular transcripts and lipids compartments. This included key genes associated with surfactant biogenesis, transport and homeostasis—orchestrated by AEC2s—as well as lipids of the PG family. Indeed, PG lipids are the second most abundant lipid class in adult rodent surfactant²² and have been demonstrated to exhibit spatial localisation specifically within the airway epithelium.²³ As expected, TGF β transcript levels were increased in IPF samples while the expression of its negative regulator SMAD6²⁴ was decreased. This goes in line with the increased expression of genes related to EMT (*SFRP4*²⁵ and *CPXM2*²⁶), a process that contributes to fibrosis, in IPF explants compared with NDC. We also observed an upregulation of collagen and other matrix-associated genes (*CTHRC1*²⁷ and *FNDC1*²⁸ among others) previously identified as potential biomarkers of IPF.^{27,29–31} Notably, *FNDC1* was the most important feature predicting established disease in our ML model for Factor1. Further investigation of genes associated with surfactant and lipid homeostasis (*SPATA18*, *SLC39A8*, *PLA2G4F*, *CACNA2D2*, *FASN* and *PCSK9*) also revealed their implication in mitochondrial function. While the expression of these genes may be reflecting differences in cell composition (e.g. loss of AEC2s in IPF), their association with mitochondrial dysfunction

was further supported by observed alterations in the metabolomics modality. Indeed, oxidative stress and tissue hypoxia have been reported to induce the degradation of adenosine,^{32,33} which results in the release of certain by-products, including hypoxanthine, xanthine and uric acid.³⁴ Low adenosine and high hypoxanthine levels have been reported in IPF lungs,³⁵ and uric acid was identified as an important signal that triggers inflammation and fibrosis in mice.³⁶ Other affected metabolic pathways in IPF explants included a decreased hypotaurine/taurine metabolism signal, in line with previous studies demonstrating the protective role of hypotaurine/taurine metabolism against oxidant-induced lung injury and fibrosis.^{37,38} The decrease in GSH pathway precursors indicates a redox imbalance, suggesting vulnerability to metabolic inflammation and lung injury not only in established fibrosis (IPF base), as previously reported,³⁹ but also in the pre-fibrotic IPF apices. Both gene expression and metabolic signals should be interpreted in consideration of the confounding effect related to a shift in cell composition, such as well-described alterations in AEC2 and fibroblast proportions in IPF.⁴⁰ In summary, Factor1 revealed a multi-omics signature associated with AEC2 dysfunction, tissue remodelling, and oxidative stress in both fibrosed bases and pre-disease apices, indicating early dysfunction of the epithelial compartment in IPF. Given that the biological entities encapsulated within Factor1 demonstrate alterations during the early stages of the disease, they hold promise as potential biomarkers for early disease detection. Particularly, metabolites identified by Factor1 emerge as strong biomarker candidates as a result of their stability and ease of measurement.

Juxtaposed with the linear and gradual cellular events leading to fibrosis underlined by the first factor, we discovered a unique molecular signature linked to disease progression, characterised by a nonlinear dynamic (Factor2) predominantly explained by the transcriptomics layer. Ciliogenesis and cilia function signals were specifically increased in the latest stage of IPF. Primary cilia are not usually found in healthy adult alveoli, but they may briefly reappear during lung injury, possibly indicating their role in wound healing.^{41,42} High cilia-related gene expression in IPF lungs has been linked to increased honeycombing and activation of the SHH pathway,⁴³ which is known to be important in the lung during embryonic development.⁴⁴ Factor2 also highlighted an intriguing pattern of genes characterised by an intermediary expression in healthy tissue (NDC samples), a downregulation in pre-fibrosis (IPF apex) and an upregulation in established fibrosis (IPF base).

Many of these genes are involved in transcriptional regulation of different processes, including EMT (*FOXA1*, *GRHL2* and *ERBB3*)^{45–47} or carcinogenesis (*EHF*, *TRIM29*, *KLF5* and *MET*),^{48–51} and also play a crucial role in embryonic lung development (*FOXA1*, *CELSR1*, *IRX3*, *KLF5* and *GRHL2*).^{52–56} Of particular interest is the transcription factor *IRX3*, the top most important feature of the best-performing ML model for Factor2, known not only for its roles in energy homeostasis⁵⁷ and mitochondrial function,⁵⁸ but also for orchestrating an array of cellular processes crucial for developmental biology, such as proximo-distal morphogenesis of the developing lung⁵⁴ or lung angiogenesis.⁵⁹ Factor2 top lipidomics features identified elevated long-chain acylcarnitines in IPF samples and in particular L-Palmitoylcarnitine in IPF apices, reflective of mitochondrial dysfunction. Indeed, acylcarnitines are produced during carnitine metabolism for FA oxidation in the mitochondria. Notably, their accumulation in serum and tissues has been associated with apoptosis via glutathione depletion and oxidative stress induced by mitochondrial dysfunction.^{60–63} This underlines the significance of the observed increase in L-Palmitoylcarnitine in IPF apices as a potential biomarker for mitochondrial dysfunction. Changes in lipids part of the endocannabinoid system, such as high anandamide levels in IPF base samples, were observed along Factor2. Studies have associated elevated anandamide levels with chronic tissue injury,^{64,65} the progression of schistosoma-induced liver fibrosis,⁶⁶ and pulmonary fibrosis progression in patients and a bleomycin mouse model of IPF⁶⁷ while modulating the endocannabinoid signalling showed potential in reducing collagen production and myofibroblast activation.⁶⁸ Interestingly, increased anandamide levels were only evident in IPF bases, while higher levels of LEA, which can attenuate anandamine hydrolysis to arachidonic acid,²⁰ were observed in NDC samples. Altogether, Factor2 demonstrated an altered lipidomic profile that aligned with our finding of decreased expression of mitochondria-related genes, such as *IRX3*, specifically in pre-fibrotic apices. For example, increased mitochondrial fragmentation and decreased mitochondrial complex IV activity have been observed with *in vitro* inhibition of *IRX3*.⁶⁹ We propose here a causal link between reduced *IRX3*, mitochondrial dysfunction and lipid metabolism inflammation that can potentially account for the early tissue injury and fibrosis spreading in IPF apices. This pattern could be indicative of dynamic shifts in cellular processes during different stages of IPF progression. One

hypothesis is that the intermediary expression in healthy tissue represents the equilibrium in lung epithelial function, while downregulation in pre-fibrosis might signify the early stages of injury and dysregulation, leading to the subsequent aberrant tissue repair response and upregulation in established fibrosis. Eventually, the system reaches a 'tipping point' where these become aberrantly reactivated (IPF base). This aligns with the idea proposed by Froidure et al.⁴ that such pathways may follow a nonlinear pattern governed by chaos. This type of dynamic is sensitive to initial conditions, and small changes can have large irreversible effects, such as the tissue remodelling seen in late-stage IPF.

While this study offers valuable insights into IPF progression, it is crucial to recognise its limitations, including the small sample size and potential variability in cellular composition across lung explant sampling sites. Nevertheless, our findings revealed strikingly clear and distinct molecular signatures between groups, underscoring robust and conserved mechanisms associated with disease progression. Additionally, we employed a machine learning approach to validate the reliability and generalisability of our gene expression findings. This included testing our predictive models on the only two other studies to our knowledge with comparable sample types.^{15,16} Indeed, our validation of Factor1 genes revealed a consistent signature that discriminated between NDC and IPF samples regardless of sample site or histopathological severity. In contrast, Factor2 genes displayed strong training performance which performed well in one of two datasets in validation. We hypothesise that the Factor2 signature may be particularly sensitive to sample handling and processing given the contrasting study-specific performance. Future work may allow careful dissemination of how and why these genes are initially downregulated in pre-fibrosed areas and then reactivated in extensively fibrosed regions.

Future investigations may focus on examining the cellular heterogeneity present in IPF lung explants from different disease stages, for example through the utilisation of single-cell sequencing. Notably, recent pioneering studies have laid the foundation for the IPF single-cell atlas,⁷⁰ paving the way for future investigations to delve into the cellular heterogeneity across various disease stages. Moreover, the integration of other emerging technologies, such as single-cell metabolomics,⁷¹ could offer valuable insights into the diverse molecular profiles and metabolic dynamics within specific cell populations. In summary, we identified

sources of multi-omics co-variations linked with IPF establishment and progression. We postulate that the multi-omics signature in IPF apices that was revealed by Factor2 represents the 'leading edge of fibrosis', an inflection point that marks a turning point in IPF progression.

METHODS

Study population/ethics and sampling

The study received ethical approval from the Alfred Human Ethics Committee (#468/14) and the Australian Red Cross Blood Service Ethics Committee, meeting the Australian National Statement on Ethical Conduct in Human Research (2007). Tissue samples were obtained from the Alfred Lung Fibrosis Biobank, and written informed consent was obtained from all participants or their families. Lung tissue was obtained from patients with end-stage interstitial lung disease or from diseased organ donors whose lungs were declined for transplantation.

Histological scoring

Lung tissues were fixed in 4% paraformaldehyde and embedded in paraffin blocks. Sections of 4 μ m were cut for Masson's trichrome staining. Blinded scoring of the lung histopathology was performed according to the Ashcroft scoring methodology.⁷²

Transcriptomics data acquisition and processing

Total RNA was extracted from lung tissues as described in Jaffar *et al.*²¹ Raw fastq files were processed with RNAsik v1.5.0,⁷³ aligned using STAR v2.5.2b against the GRCh38.104 reference genome and read counts were quantified to annotated genes using featureCounts.⁷⁴ Annotated protein-coding genes were retained and low expression genes filtered using the filterByExpr function from edgeR v3.36.0 before normalisation with vsn R package v3.62.0. For integrated analysis with previous datasets, raw fastq files were downloaded from the SRA from Jaffar *et al.* (GSE213001),²¹ Luzina *et al.* (GSE99621)¹⁵ and McDonough *et al.* (GSE124685).¹⁶ Files were processed using the above pipeline, including batch effect removal using ComBat-seq⁷⁵ with study as the batch effect and disease (NDC or IPF) as a biological condition.

LCMS sample processing and data acquisition

Frozen lung tissues were crushed with a stainless steel multisample pulveriser dipped in liquid nitrogen. The resulting powder was extracted using 80% MeOH with 1 μ M CHAPS, CAPS and PIPES internal standards for metabolomics and CHCl₃:MeOH:H₂O (1:3:1, v/v) spiked with 45 nM Cer(35:1) ceramide and 25 nM SM(35:1) and 10 μ M BHT for lipidomics. After vortexing, sonication and centrifugation,

the supernatant was transferred to Eppendorf tubes, and a pooled QC was prepared with blank samples for each analysis. Lipidomics samples were evaporated and solubilised in MeOH:BuOH:H₂O (4.5:4.5:1, v/v) mixture, vortexed, sonicated and centrifuged. LCMS data were acquired on a Q-Exactive Orbitrap mass spectrometer (Thermo Fisher) coupled with the Dionex Ultimate 3000 RS (Thermo Fisher) high-performance liquid chromatography system in a single batch and randomised to account for system drift.

LCMS data analysis

LCMS data were processed using the metabolome-lipidome-MSDIAL pipeline (see data and materials availability section). MS-DIAL v4.7 was used for peak detection and alignment against the MassBank v2021.02 and MS-DIAL internal lipid databases with the following parameters: minimum peak amplitude of 100 000, retention time tolerance parameter of 1 min and mass tolerance of 0.002 Da with gap-filling. Peak intensities were normalised using probabilistic quotient normalisation (PQN), followed by random forest missing data imputation and subsequent generalised logarithmic (glog) transformation using the pmp R package v1.6.0. Filtered peaks were mapped to the Human Metabolome database (HMDB) v4.202107 with a mass tolerance of 0.002 Da and annotated features were curated using a combination of the MS-DIAL fill percentage and signal-to-noise ratio values, along with visual confirmation.

Multi-omics data integration

Normalised transcriptomics, metabolomics and lipidomics matrices were used as an input for Multi-Omics Factor Analysis 2 (MOFA+) v1.4.0. This unsupervised Bayesian approach infers a low-dimensional representation of the data (latent factors) capturing overall sources of variability. Default parameters were used. Characterisation of the MOFA2 model included the quantification of the variance explained (R^2) by each factor in each view as well as the assessment of the top 50 weights in each omics modality, which can be interpreted as a measure of feature importance. Top feature characterisation into pathways was performed using a literature-based search (genes) or grouping into broad molecular categories (lipids and metabolites).

Machine learning

The top 50 genes with the highest weights in Factor1 and Factor2 were used as predictor variables for IPF and NDC, and early (IPF apex) and late-stage disease (IPF base) respectively. To increase the sample size for model training and performance evaluation, Jaffar *et al.*'s previously published dataset²¹ was included. Validation and performance estimations of four models were performed using a leave-group-out cross-validation approach using the caret R package, in which the data were folded 50 $\times$ into 85% training data and 15% validation data. The best-performing models, as determined using evalm from MLeval v0.3 R package, were selected based on the maximum mean area under the ROC curve and accuracy values. Feature importance was estimated using the varImp

function. The models were tested on two additional datasets,^{15,16} and a confusion matrix was generated for each dataset and factor.

Statistical analysis and data visualisation

Downstream analyses and figures were generated using R statistical software v4.1.1. Differences between groups were addressed using non-parametric Wilcoxon rank-sum testing with FDR correction for multiple testing, and PCA was performed on normalised transcriptomics, metabolomics and lipidomics matrices using the prcomp function. Transcriptomics differential expression analysis was performed using the R package limma v3.56.2 to identify differentially expressed genes (FDR adjusted *P*-values < 0.05) to compare disease groups, apex/base sampling and disease progression. For the analysis of differentially abundant metabolites and lipids, linear modelling (lm function in stats v3.4.1) was used for each feature followed by FDR adjustment of *P*-values. Pearson correlations between factors and covariates were calculated using the correlate_factors_with_covariates function in the MOFA v1.4.0 R package (significant correlations with *P*-value < 0.05 are presented). Data visualisation was performed using ggplot2 R package v3.3.5 and heatmaps drawn with ComplexHeatmap v2.10.0 R package with clustering method complete for rows and Maximum for columns. A fixed random-number seed value of 2 was set for reproducibility.

ACKNOWLEDGMENTS

Histological processing and staining were performed by the Monash Histology Platform. We thank Professor Catriona McLean of Alfred's Anatomical Pathology Department for assistance in obtaining IPF tissue. We also thank Professors Darren Creek and Dovile Anderson of the Monash Proteomics and Metabolomics Facility for their help with LC-MS data acquisition. The ALF Biobank is supported by the National Health and Medical Research Council for Research Excellence in Pulmonary Fibrosis. Bailey Cardwell is supported by an Australian Government Research Training Program (RTP) Scholarship. Benjamin J Marsland and Nicola L Harris are supported by the National Health and Medical Research Council (NHMRC) Australia, Senior Research Fellowships. This work was supported by funding from The Baker Foundation, Melbourne, Australia awarded to Benjamin J Marsland and Glen P Westall. Open access publishing facilitated by Monash University, as part of the Wiley - Monash University agreement via the Council of Australian University Librarians.

AUTHOR CONTRIBUTIONS

Céline Pattaroni: Formal analysis; visualization; writing – original draft. **Christina Begka:** Data curation; formal analysis; writing – original draft. **Bailey Cardwell:** Formal analysis; visualization; writing – original draft. **Jade Jaffar:** Investigation; methodology. **Matthew Macowan:** Methodology. **Nicola L Harris:** Writing – review and editing. **Glen P Westall:**

Conceptualization; funding acquisition; writing – review and editing. **Benjamin J Marsland:** Conceptualization; funding acquisition; writing – review and editing.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

Raw transcriptomics data for this study are freely available as NCBI GEO Dataset GSE213001. Pipeline for LCMS data processing are available at <https://github.com/respiratory-immunology-lab/metabolome-lipidome-MSDIAL>. Annotated normalised metabolomics and lipidomics intensity tables are available as Mendeley datasets (doi: [10.17632/4zr8jjmzhy.1](https://doi.org/10.17632/4zr8jjmzhy.1)).

REFERENCES

1. Sgalla G, Iovene B, Calvello M, Ori M, Varone F, Richeldi L. Idiopathic pulmonary fibrosis: Pathogenesis and management. *Respir Res* 2018; **19**: 32.
2. Kropski JA, Blackwell TS, Loyd JE. The genetic basis of idiopathic pulmonary fibrosis. *Eur Respir J* 2015; **45**: 1717–1727.
3. Park Y, Ahn C, Kim T-H. Occupational and environmental risk factors of idiopathic pulmonary fibrosis: A systematic review and meta-analyses. *Sci Rep* 2021; **11**: 4318.
4. Froidure A, Marchal-Duval E, Homps-Legrand M et al. Chaotic activation of developmental signalling pathways drives idiopathic pulmonary fibrosis. *Eur Respir Rev* 2020; **29**: 190140.
5. Singh S, Natalini JG, Segal LN. Lung microbial-host interface through the lens of multi-omics. *Mucosal Immunol* 2022; **15**: 837–845.
6. Vukmirovic M, Kaminski N. Impact of transcriptomics on our understanding of pulmonary fibrosis. *Front Med* 2018; **5**: 87.
7. Durante W, Liao L, Reyna SV, Peyton KJ, Schafer AI. Transforming growth factor β (1) stimulates L-arginine transport and metabolism in vascular smooth muscle cells: Role in polyamine and collagen synthesis. *Circulation* 2001; **103**: 1121–1127.
8. Hesse M, Modolell M, La Flamme AC et al. Differential regulation of nitric oxide synthase-2 and arginase-1 by type 1/type 2 cytokines in vivo: Granulomatous pathology is shaped by the pattern of L-arginine metabolism. *J Immunol Baltim Md* 1950; **2001**: 6533–6544.
9. Endo M, Oyadomari S, Terasaki Y et al. Induction of arginase I and II in bleomycin-induced fibrosis of mouse lung. *Am J Physiol Lung Cell Mol Physiol* 2003; **285**: L313–L321.
10. Geng J, Liu Y, Dai H, Wang C. Fatty acid metabolism and idiopathic pulmonary fibrosis. *Front Physiol* 2021; **12**: 794629.
11. Shea BS, Tager AM. Role of the lysophospholipid mediators lysophosphatidic acid and sphingosine 1-phosphate in lung fibrosis. *Proc Am Thorac Soc* 2012; **9**: 102–110.

12. Watzenboeck ML, Gorki A-D, Quattrone F et al. Multi-omics profiling predicts allograft function after lung transplantation. *Eur Respir J* 2022; **59**: 2003292.
13. Zhao YD, Yin L, Archer S et al. Metabolic heterogeneity of idiopathic pulmonary fibrosis: A metabolomic study. *BMJ Open Respir Res* 2017; **4**: e000183.
14. Nojima Y, Takeda Y, Maeda Y et al. Metabolomic analysis of fibrotic mice combined with public RNA-Seq human lung data reveal potential diagnostic biomarker candidates for lung fibrosis. *FEBS Open Bio* 2020; **10**: 2427–2436.
15. Luzina IG, Salcedo MV, Rojas-Peña ML et al. Transcriptomic evidence of immune activation in macroscopically normal-appearing and scarred lung tissues in idiopathic pulmonary fibrosis. *Cell Immunol* 2018; **325**: 1–13.
16. McDonough JE, Ahangari F, Li Q et al. Transcriptional regulatory model of fibrosis progression in the human lung. *JCI Insight* 2019; **4**: 131597.
17. Argelaguet R, Arnol D, Bredikhin D et al. MOFA+: A statistical framework for comprehensive integration of multi-modal single-cell data. *Genome Biol* 2020; **21**: 111.
18. Wright SM, Hockey PM, Enhorning G et al. Altered airway surfactant phospholipid composition and reduced lung function in asthma. *J Appl Physiol Bethesda Md* 1985; **2000**: 1283–1292.
19. Kandasamy P, Zarini S, Chan ED, Leslie CC, Murphy RC, Voelker DR. Pulmonary surfactant phosphatidylglycerol inhibits mycoplasma pneumoniae-stimulated eicosanoid production from human and mouse macrophages. *J Biol Chem* 2011; **286**: 7841–7853.
20. Maccarrone M, van der Stelt M, Rossi A, Veldink GA, Vliegthart JF, Agrò AF. Anandamide hydrolysis by human cells in culture and brain. *J Biol Chem* 1998; **273**: 32332–32339.
21. Jaffar J, Wong M, Fishbein GA et al. Matrix metalloproteinase-7 is increased in lung bases but not apices in idiopathic pulmonary fibrosis. *ERJ Open Res* 2022; **8**: 191-02022.
22. Dexter A, Steven RT, Patel A et al. Imaging drugs, metabolites and biomarkers in rodent lung: A DESI MS strategy for the evaluation of drug-induced lipidosis. *Anal Bioanal Chem* 2019; **411**: 8023–8032.
23. Lukowski JK, Olson H, Velickovic M et al. An optimized approach and inflation media for obtaining complimentary mass spectrometry-based omics data from human lung tissue. *Front Mol Biosci* 2022; **9**: 1022775.
24. Imamura T, Takase M, Nishihara A et al. Smad6 inhibits signalling by the TGF- β superfamily. *Nature* 1997; **389**: 622–626.
25. Tinazzi I, Mulipa P, Colato C et al. SFRP4 expression is linked to immune-driven fibrotic conditions, correlates with skin and lung fibrosis in SSc and a potential EMT biomarker. *J Clin Med* 2021; **10**: 5820.
26. Zhao X, Li R, Wang Q, Wu M, Wang Y. Overexpression of carboxypeptidase X M14 family member 2 predicts an unfavorable prognosis and promotes proliferation and migration of osteosarcoma. *Diagn Pathol* 2019; **14**: 118.
27. Tsukui T, Sun K-H, Wetter JB et al. Collagen-producing lung cell atlas identifies multiple subsets with distinct localization and relevance to fibrosis. *Nat Commun* 2020; **11**: 1920.
28. Yang F, Hou Z-F, Zhu H-Y et al. Catalpol protects against pulmonary fibrosis through inhibiting TGF- β 1/Smad3 and Wnt/ β -catenin signaling pathways. *Front Pharmacol* 2020; **11**: 594139.
29. Tsitoura E, Trachalaki A, Vasarmidi E et al. Collagen 1a1 expression by airway macrophages increases in fibrotic ILDs and is associated with FVC decline and increased mortality. *Front Immunol* 2021; **12**: e645548.
30. Adams TS, Schupp JC, Poli S et al. Single-cell RNA-seq reveals ectopic and aberrant lung-resident cell populations in idiopathic pulmonary fibrosis. *Sci Adv* 2020; **6**: eaba1983.
31. Wan H, Huang X, Cong P et al. Identification of hub genes and pathways associated with idiopathic pulmonary fibrosis via bioinformatics analysis. *Front Mol Biosci* 2021; **8**: 711239.
32. Mentzer RM, Rubio R, Berne RM. Release of adenosine by hypoxic canine lung tissue and its possible role in pulmonary circulation. *Am J Physiol* 1975; **229**: 1625–1631.
33. Woolliscroft JO, Colfer H, Fox IH. Hyperuricemia in acute illness: A poor prognostic sign. *Am J Med* 1982; **72**: 58–62.
34. Elsayed NM, Nakashima JM, Postlethwait EM. Measurement of uric acid as a marker of oxygen tension in the lung. *Arch Biochem Biophys* 1993; **302**: 228–232.
35. Kang YP, Lee SB, Lee J et al. Metabolic profiling regarding pathogenesis of idiopathic pulmonary fibrosis. *J Proteome Res* 2016; **15**: 1717–1724.
36. Gasse P, Riteau N, Charron S et al. Uric acid is a danger signal activating NALP3 Inflammasome in lung injury inflammation and fibrosis. *Am J Respir Crit Care Med* 2009; **179**: 903–913.
37. Gordon RE, Heller RF, Heller RF. Taurine protection of lungs in hamster models of oxidant injury: A morphologic time study of paraquat and bleomycin treatment. *Adv Exp Med Biol* 1992; **315**: 319–328.
38. Chang H, Meng H, Bai W, Meng Q. A metabolomic approach to elucidate the inhibitory effects of baicalin in pulmonary fibrosis. *Pharm Biol* 2021; **59**: 1014–1023.
39. Beeh KM, Beier J, Haas IC, Kornmann O, Micke P, Buhl R. Glutathione deficiency of the lower respiratory tract in patients with idiopathic pulmonary fibrosis. *Eur Respir J* 2002; **19**: 1119–1123.
40. Selman M, Pardo A. Role of epithelial cells in idiopathic pulmonary fibrosis. *Proc Am Thorac Soc* 2006; **3**: 364–372.
41. Look DC, Walter MJ, Williamson MR et al. Effects of paramyxoviral infection on airway epithelial cell Foxj1 expression, ciliogenesis, and mucociliary function. *Am J Pathol* 2001; **159**: 2055–2069.
42. Jain R, Pan J, Driscoll JA et al. Temporal relationship between primary and motile ciliogenesis in airway epithelial cells. *Am J Respir Cell Mol Biol* 2010; **43**: 731–739.
43. Lee J, Oh DH, Park KC et al. Increased primary cilia in idiopathic pulmonary fibrosis. *Mol Cells* 2018; **41**: 224–233.
44. Pepicelli CV, Lewis PM, McMahon AP. Sonic hedgehog regulates branching morphogenesis in the mammalian lung. *Curr Biol* 1998; **8**: 1083–1086.
45. BenAayed-Guerfali D, Dabbèche-Bouricha E, Ayadi W et al. Association of FOXA1 and EMT markers (Twist1 and E-cadherin) in breast cancer. *Mol Biol Rep* 2019; **46**: 3247–3255.

46. Chung VY, Tan TZ, Tan M et al. GRHL2-miR-200-ZEB1 maintains the epithelial status of ovarian cancer through transcriptional regulation and histone modification. *Sci Rep* 2016; **6**: 19943.
47. Stella GM, Gentile A, Balderacchi A, Meloni F, Milan M, Benvenuti S. Ockham's razor for the MET-driven invasive growth linking idiopathic pulmonary fibrosis and cancer. *J Transl Med* 2016; **14**: 256.
48. Albino D, Longoni N, Curti L et al. ESE3/EHF controls epithelial cell differentiation and its loss leads to prostate tumors with mesenchymal and stem-like features. *Cancer Res* 2012; **72**: 2889–2900.
49. Zhou Z-Y, Yang G-Y, Zhou J, Yu M-H. Significance of TRIM29 and β -catenin expression in non-small-cell lung cancer. *J Chin Med Assoc* 2012; **75**: 269–274.
50. Li Q, Dong Z, Zhou F, Cai X, Gao Y, Wang L-W. Krüppel-like factor 5 promotes lung tumorigenesis through upregulation of Sox4. *Cell Physiol Biochem* 2014; **33**: 1–10.
51. Drilon A, Cappuzzo F, Ou S-HI, Camidge DR. Targeting MET in lung cancer: Will expectations finally be MET? *J Thorac Oncol* 2017; **12**: 15–26.
52. Wan H, Dingle S, Xu Y et al. Compensatory roles of Foxa1 and Foxa2 during lung morphogenesis. *J Biol Chem* 2005; **280**: 13809–13816.
53. Yates LL, Schnatwinkel C, Murdoch JN et al. The PCP genes *Celsr1* and *Vangl2* are required for normal lung branching morphogenesis. *Hum Mol Genet* 2010; **19**: 2251–2267.
54. van Tuyl M, Liu J, Groenman F et al. Iroquois genes influence proximo-distal morphogenesis during rat lung development. *Am J Physiol Lung Cell Mol Physiol* 2006; **290**: L777–L789.
55. Wan H, Luo F, Wert SE et al. Kruppel-like factor 5 is required for perinatal lung morphogenesis and function. *Dev Camb Engl* 2008; **135**: 2563–2572.
56. Kersbergen A, Best SA, Dworkin S et al. Lung morphogenesis is orchestrated through Grainyhead-like 2 (Grhl2) transcriptional programs. *Dev Biol* 2018; **443**: 1–9.
57. Zhang Z, Wu Q, He Y et al. *IRX3* overexpression enhances *Ucp1* expression *in vivo*. *Front Endocrinol* 2021; **12**: 634191.
58. Bjune J-I, Dyer L, Røslund GV et al. The homeobox factor *Irx3* maintains adipogenic identity. *Metabolism* 2020; **103**: 154014.
59. Scarlett K, Pattabiraman V, Barnett P, Liu D, Anderson LM. The proangiogenic effect of Iroquois homeobox transcription factor *Irx3* in human microvascular endothelial cells. *J Biol Chem* 2015; **290**: 6303–6315.
60. Noland RC, Koves TR, Seiler SE et al. Carnitine insufficiency caused by aging and overnutrition compromises mitochondrial performance and metabolic control. *J Biol Chem* 2009; **284**: 22840–22852.
61. Weng L, Gong Y, Culver J et al. Presence of arachidonoyl-carnitine is associated with adverse cardiometabolic responses in hypertensive patients treated with atenolol. *Metabolomics* 2016; **12**: 160.
62. Enooku K, Nakagawa H, Fujiwara N et al. Altered serum acylcarnitine profile is associated with the status of nonalcoholic fatty liver disease (NAFLD) and NAFLD-related hepatocellular carcinoma. *Sci Rep* 2019; **9**: 10663.
63. Turnbull PC, Hughes MC, Perry CGR. The fatty acid derivative palmitoylcarnitine abrogates colorectal cancer cell survival by depleting glutathione. *Am J Physiol-Cell Physiol* 2019; **317**: C1278–C1288.
64. Osei-Hyiaman D, DePetrillo M, Pacher P et al. Endocannabinoid activation at hepatic CB1 receptors stimulates fatty acid synthesis and contributes to diet-induced obesity. *J Clin Invest* 2005; **115**: 1298–1305.
65. Jourdan T, Szanda G, Rosenberg AZ et al. Overactive cannabinoid 1 receptor in podocytes drives type 2 diabetic nephropathy. *Proc Natl Acad Sci USA* 2014; **111**: E5420–E5428.
66. Liu H, Gao X, Duan R et al. Endocannabinoids anandamide and its cannabinoid receptors in liver fibrosis after murine schistosomiasis. *J Huazhong Univ Sci Technol Med Sci* 2009; **29**: 182–186.
67. Cinar R, Gochuico BR, Iyer MR et al. Cannabinoid CB1 receptor overactivity contributes to the pathogenesis of idiopathic pulmonary fibrosis. *JCI Insight* 2017; **2**: e92281.
68. Marquart S, Zerr P, Akhmetshina A et al. Inactivation of the cannabinoid receptor CB1 prevents leukocyte infiltration and experimental fibrosis. *Arthritis Rheum* 2010; **62**: 3467–3476.
69. Ma Y, Chen G, Yi J et al. *IRX3* plays an important role in the pathogenesis of metabolic-associated fatty liver disease by regulating hepatic lipid metabolism. *Front Endocrinol* 2022; **13**: 895593.
70. Neumark N, Cosme C, Rose K-A, Kaminski N. The idiopathic pulmonary fibrosis cell atlas. *Am J Physiol Lung Cell Mol Physiol* 2020; **319**: L887–L892.
71. Guo S, Zhang C, Le A. The limitless applications of single-cell metabolomics. *Curr Opin Biotechnol* 2021; **71**: 115–122.
72. Ashcroft T, Simpson JM, Timbrell V. Simple method of estimating severity of pulmonary fibrosis on a numerical scale. *J Clin Pathol* 1988; **41**: 467–470.
73. Tsyganov K, Perry A, Archer S, Powell D. RNAsik: A pipeline for complete and reproducible RNA-seq analysis that runs anywhere with speed and ease. *J Open Source Softw* 2018; **3**: 583.
74. Liao Y, Smyth GK, Shi W. featureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinforma Oxf Engl* 2014; **30**: 923–930.
75. Zhang Y, Parmigiani G, Johnson WE. ComBat-seq: Batch effect adjustment for RNA-seq count data. *NAR Genom Bioinform* 2020; **2**: lqaa078.

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.



This is an open access article under the terms of the [Creative Commons Attribution](#) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.