

MUUMI: an R package for statistical and network-based meta-analysis for Multi-omics data Integration

Simo Inkala¹, Michele Fratello¹, Giusy del Giudice^{1,3}, Giorgia Migliaccio¹, Angela Serra^{1,2,3}, Dario Greco^{1,3,4}, Antonio Federico^{1,2,3}

¹ Finnish Hub for Development and Validation of Integrated Approaches (FHAIVE), Faculty of Medicine and Health Technology, Tampere University, 33100, Tampere, Finland

² Tampere Institute for Advanced Study, Tampere University, 33100, Tampere, Finland

³ Division of Pharmaceutical Biosciences, Faculty of Pharmacy, University of Helsinki, 00100, Helsinki, Finland

⁴ Institute of Biotechnology, University of Helsinki, 00100, Helsinki, Finland

Data and code to reproduce the following case studies are available on Zenodo (<https://zenodo.org/records/15019060>).

Case Study 1

Experimental setup and data used

To showcase the functionalities of MUUMI, we performed a case study on idiopathic pulmonary fibrosis (IPF) by analysing samples deriving from the integration of 17 pre-processed and harmonized public transcriptomics datasets. Among all the included datasets, 11 derive from RNA sequencing (RNA-Seq) while 6 from microarray experiments (Table S1). The collected data include both fibrotic lung biopsy samples of IPF patients and biopsy samples of healthy individuals. Datasets are made publicly available in Zenodo (<https://zenodo.org/doi/10.5281/zenodo.10692128>). We employed ESPERANTO, a R/Shiny application, to efficiently curate and harmonize metadata using its semi-supervised approach, streamlining the metadata curation process (1). This case study aims to perform the integration of single omics datasets and carry out both statistical and network-based meta-analysis. The resulting data can then be utilised to uncover molecular vulnerabilities and dysregulated biological processes underlying the IPF phenotype.

GEO dataset ID	Platform	Citation	Number of IPF samples	Number of healthy samples
GSE150910	Illumina NovaSeq 6000	(2)	103	103
GSE213001	Illumina HiSeq 3000	(3)	62	41
GSE124685	Ion Torrent Proton	(4)	49	35

GSE199949	Illumina HiSeq 4000	(5)	26	16
GSE92592	Illumina HiSeq 2000	(6)	20	19
GSE199152	Illumina HiSeq 2500	Unpublished	20	4
GSE53845	Agilent-014850 4x44K G4112F	(7)	40	8
GSE166036	Illumina HiSeq 4000	(8)	10	4
GSE184316	Ion Torrent Proton	Unpublished	40	24
GSE110147	Affymetrix 1.0 ST Array	(9)	22	11
GSE21369	Affymetrix U133 Plus 2.0	(10)	11	6
GSE24206	Affymetrix U133 Plus 2.0 Array	(11)	17	6
GSE72073	Affymetrix Transcriptome Array 2.0	(12)	5	3
GSE169500	Ion Torrent PGM	(13)	20	10
GSE99621	Illumina HiSeq 2500	(14)	18	8
GSE10667	Agilent-014850 4x44K G4112F	(15)	31	15
GSE138283	Illumina NextSeq 550	(16)	12	5

Table S1 - Publicly available transcriptomics datasets included in the case study.

Phenotypic and molecular basis of IPF

IPF is a chronic, progressive, devastating, and age-related interstitial lung disease (ILD) that is irreversible and typically fatal. IPF has an unknown origin and only a few treatment options that are limited in their efficacy (17). In IPF, progressive formation of scar tissue in the supportive interstitium of the lungs occurs, causing breathing to become increasingly difficult (18). Histopathologically IPF typically exhibits subpleural and paraseptal fibrosis, honeycombing, and regions of less affected and normal parenchyma (17). The proposed pathogenesis connects IPF to a lung

susceptible to aging-related changes, which is subject to repetitive alveolar injuries triggered by factors such as inhaled cigarette smoke, microaspiration, nanomaterials, gastroesophageal reflux, or viruses (17,19). These injuries can lead to type I and type II epithelial cell death. Following micro-injuries and epithelial cell apoptosis, there is an increased vascular permeability to proteins like fibrinogen and fibronectin, leading to the formation of a wound clot. Subsequently, there is migration and proliferation of bronchiolar and alveolar epithelial cells, representing a rapid response of the respiratory tissue attempting self-repair. Abnormally activated epithelial cells release various chemokines, cytokines, and epidermal growth factors, attracting fibroblasts, immune system cells such as alveolar macrophages, and monocytes that differentiate into macrophages. Moreover, these cells secrete TGF- β 1, promoting epithelial-mesenchymal transition, extracellular matrix remodeling, and fibroblast differentiation into myofibroblasts. The heterogeneous macrophage population also secretes chemokines and growth factors like TGF- β , inducing the growth of fibrotic tissue in the extracellular matrix. Positive feedback loops contribute to the progressive expansion of the fibrotic tissue (20–22).

Identification of robust IPF genes by statistical meta-analysis

For each dataset indicated in Table S1, we conducted differential gene expression analysis between IPF and healthy samples. For RNA-Seq datasets, gene expression data normalisation and differential gene expression analysis were performed using DESeq2 version 1.24.0 (23). DNA microarray data preprocessing was carried out by using the eUTOPIA software, which implements the limma package for expression value normalisation and performed differential gene expression analysis by applying quantile normalisation and linear model (24), (25). We adjusted nominal p-values in both RNA-Seq and microarray datasets by using the Benjamini & Hochberg method (26). In order to identify robust IPF-associated genes, we integrated the results of the differential expression analysis by performing a statistical-based meta-analysis utilising the function "*run_ensembl_metanalysis*", whose input consists of adjusted p-values from all the datasets, with gene symbols as row names. All three methods included in our pipeline, namely "*effect_size*," "*pvalue*," and "*rank_product*," were utilized in the ensemble analysis. 9371 genes were included in the analysis. The full meta-analysis gene rank is available at <https://zenodo.org/records/15019060>.

Among the highest-ranked genes, we identified several members of the COL family: COL3A1, which holds the highest overall ranking in the analysis, along with COL14A1, COL15A1, and COL1A2 which are also in the top 20. The transcriptional profile of lung fibroblasts can influence the extracellular matrix (ECM) proteins encoded by these genes, leading to alterations in the ECM of lungs affected by IPF (27). Accumulating evidence suggests that the mechanical interactions between fibroblasts and the stiffened ECM create a feedforward mechanism, contributing to the persistence and progression of pulmonary fibrosis (28,29). A study by Huang, S. et al. 2022 (30) revealed that Asporin (ASPN) promotes the differentiation of lung myofibroblasts induced by TGF- β by facilitating Rab11-dependent recycling of T β RI. ASPN, a member of the small leucine-rich proteoglycan family known for its crucial roles in tissue injury and regeneration, is highly expressed in various tumor types and has the capacity to upregulate TGF- β 1. TGF- β induces fibrotic tissue growth in the ECM (31). Elevated protein levels of MMP7 have been observed in IPF compared to samples from healthy controls, and it serves as a predictive biomarker for disease progression in IPF (32). In individuals affected by IPF, the expression of

matrix metalloproteinases (MMPs) is dysregulated, resulting in substantial architectural remodeling in the lung microenvironment. MMPs have been considered potential therapeutic targets for IPF (33). Yu, G., et al. 2018 (34) found that the activity and expression of iodothyronine deiodinase 2 (DIO2), was higher in the lungs of patients with IPF compared to control individuals and correlated with disease severity. Based on these findings, we can confirm that the genes ranked by the *run_ensembl_metanalysis* function hold biological relevance in IPF, indicating the efficacy of our approach in prioritizing genes resulting from differential analysis of multiple omics datasets.

Identification of functionally relevant genes and pathways derived from statistical-based meta-analysis

We performed Gene Set Enrichment Analysis (GSEA) utilizing the function *compute_gsea* for the ranked gene list that was obtained from the meta-analysis. Gene sets were retrieved from the Molecular Signatures Database (MSigDB) (35). To ensure meaningful results, the minimum pathway size was set to 10 genes and the number of permutations to 100,000. The 15 most enriched pathways are represented in Table S2.

Pathway	p-val	p-adj	ES	NES	nMoreExtreme	Size
Extracellular Matrix Organization	1.00E-05	0.003423	0.458414	1.683331	0	185
Regulation of Insulin Like Growth Factor IGF Transport and Uptake by Insulin Like Growth Factor Binding Proteins IGFBPs	1.00E-05	0.003423	0.502973	1.716711	0	64
Collagen Degradation	1.00E-05	0.003423	0.587627	1.894219	0	37
Assembly of Collagen Fibrils and Other Multimeric Structures	2.00E-05	0.005132	0.537653	1.763733	1	43
ECM Proteoglycans	4.00E-05	0.007379	0.512437	1.704437	3	49
Collagen Biosynthesis and Modifying Enzymes	5.00E-05	0.007379	0.527783	1.731356	4	43
Molecules Associated with Elastic Fibres	5.03E-05	0.007379	0.633992	1.930158	4	24
Degradation of the Extracellular Matrix	7.00E-05	0.00802	0.454145	1.58119	6	81

Collagen Chain Trimerization	7.03E-05	0.00802	0.603954	1.860309	6	26
Reactome Collagen Formation	8.00E-05	0.008208	0.482114	1.63859	7	61
Metabolism of Steroids	9.00E-05	0.008394	0.449728	1.573069	8	86
Integrin Cell Surface Interactions	1.10E-04	0.009405	0.489075	1.642485	10	54
Elastic Fibre Formation	1.60E-04	0.012312	0.565026	1.774937	15	30
Class A1 Rhodopsin Like Receptors	1.70E-04	0.012312	0.438244	1.536992	16	89
GPRC Ligand Binding	1.80E-04	0.012312	0.405372	1.454706	17	124

Table S2 – Top 15 enriched Reactome pathways in GSEA analysis.

The *compute_gsea_thresh* function computes a threshold for the statistical meta-analysis gene rank based on the GSEA enrichment score. In this case study, the threshold was set to the top 2860 genes out of 9371 genes.

The GSEA analysis results of MUUMI (Table S2) successfully identified significant functional pathways associated with IPF. Notably, pathways related to ECM remodeling were found to be significantly enriched among the ranked genes in the meta-analysis. This emphasizes the prevalent occurrence of ECM remodeling in lung diseases such as IPF (36,37). "Extracellular Matrix Organization" and "Collagen Formation" pathways highlight the structural aspects of the ECM, emphasizing the importance of collagen assembly and fibrillogenesis in the pathogenesis of IPF. Additionally, pathways such as "Collagen Degradation" and "Degradation of The Extracellular Matrix" imply dysregulation in the turnover of ECM components, possibly contributing to fibrotic changes observed in IPF. Pathways like "Regulation of Insulin-Like Growth Factor IGF Transport" and "Integrin Cell Surface Interactions" suggest potential connections between growth factor signaling and cellular adhesion, emphasizing the complex interplay between signaling pathways and fibrotic processes. Moreover, pathways related to elastic fibers and G protein-coupled receptor (GPCR) ligand binding shed light on the involvement of these structural and signaling elements in the context of IPF pathology. The enrichment of pathways associated with metabolism, such as "Metabolism of Steroids," suggests a broader impact on cellular homeostasis. The results of the GSEA analysis demonstrate that MUUMI is capable of recapitulating biologically relevant pathways based on the statistical meta-analysis.

Mitigation of technical effects prior multi-study data integration

Technical nuisance can heavily affect the results of omics data, leading to artifacts in the results and their interpretation. Such a phenomenon is known as batch effect. Batch effect can derive from protocols, reagents, equipment, laboratory conditions, sample collection, storage, preparation methods, time, and so forth. Therefore, since gene expression datasets utilised in our case study derive from different studies, it is essential to mitigate the batch effect before carrying out subsequent integrative

analyses, such as inferring network relationships (38). To carry out batch adjustment, considering the diverse origins of the gene expression datasets, it was essential to carry out the dataset integration separately for DNA microarray and RNA-Seq datasets. The batch adjustment was carried out with the “*multi_studies_adjust*” function which takes as an input the integrated expression matrix, the sample labels, and the batch labels. The embedded functionality to carry out batch adjustment in the “*multi_studies_adjust*” utilizes the “*pamr.batchadjust*” from the “*pamr*” package (39). Figure S1 shows the outcome of the batch effect adjustment on both RNA-Sequencing (A) and DNA microarray (B) datasets. Panels A1 and B1 represent the disease samples while panels A2 and B2 represent the healthy samples.

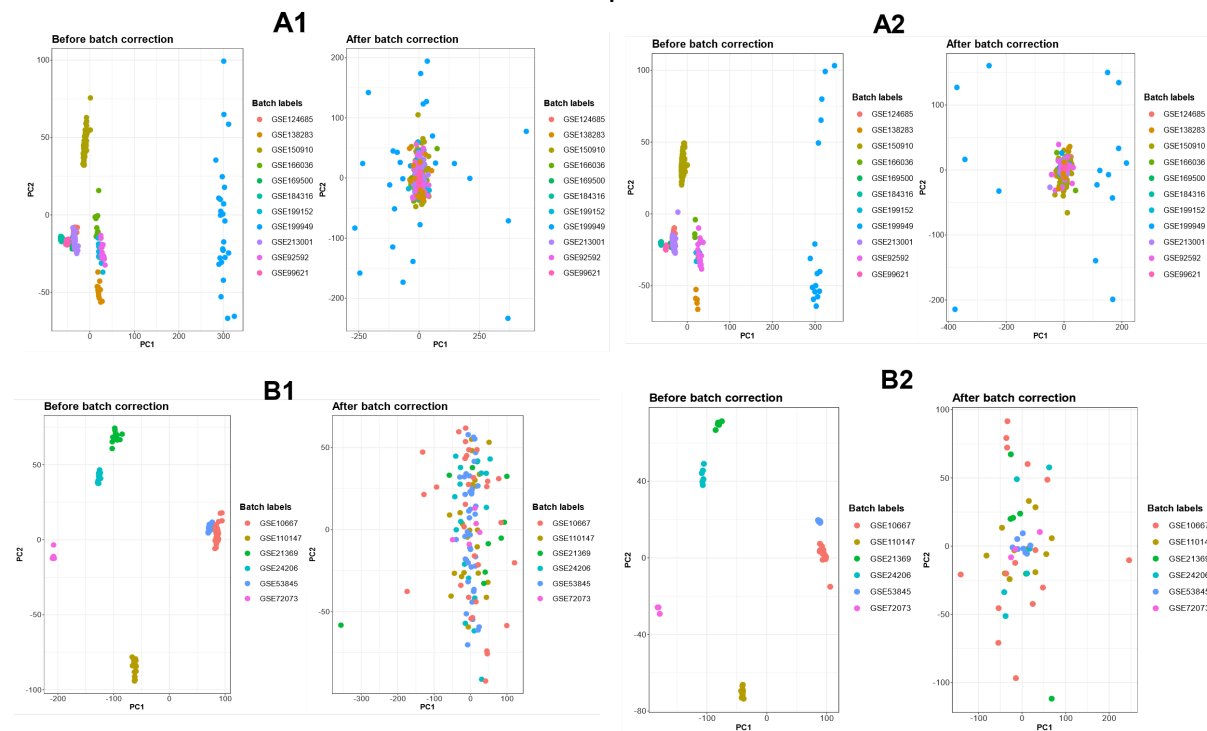


Figure S1 - A1 and A2 PCA plots represent RNA-Seq dataset samples both before and after batch correction, with the A1 charts corresponding to disease samples and A2 charts representing healthy samples. In contrast, B1 and B2 PCA plots represent microarray dataset samples before and after batch correction, with B1 diagrams representing disease samples and B2 diagrams illustrating healthy samples.

Integrated Datasets for Network-Based Meta-Analysis

Subsequently, we aimed at investigating co-expression patterns of the genes obtained from the meta-analysis. To do so, we inferred gene co-expression network models from IPF samples and healthy counterparts profiled through RNA-Seq datasets. To infer the networks from the normalised and batch-adjusted read count matrices, we firstly utilised the “*get_ranked_consensus_matrix*” function. Through this function we determined correlation measures between genes, based on their expression levels for both disease and healthy samples independently. In this case study, we computed Pearson correlation to build a squared adjacency matrix

underlying the network and then we set CLR as a network inference algorithm. We then completed the network inference through the "parse_edge_rank_matrix" function to systematically add edges to the graph so that all the nodes are connected. The output of this function is a binary squared matrix, where 1 indicates the presence of an edge between two nodes while 0 indicates that no edge is present between two nodes. We therefore converted the binary matrix to an iGraph object using the "get_iGraph" function. The number of vertices and edges in each network are represented in Table S3.

Functional characterisation of the networks

Community detection for each network was performed using the "get_modules" function, which implements the walktrap (40), spinglass (41), (42), louvain (43) and the greedy algorithm (44). For this case study, we employed the walktrap algorithm. The number of communities in each network is shown in Table S3. We conducted a functional annotation for disease and healthy modules using the "get_reactome_from_modules" function to identify significantly enriched biological pathways. To visualize the pathway enrichment results, we used the function "get_bubble_plot_from_pathways" to create bubble plots for each module, with the RNA-Seq network results displayed in Figure S2. Additionally, the "comparative_heatmap_for_modules" function was employed to plot both healthy and disease network modules on the same plot, with the comparative bubble plots for RNA-Seq networks shown in Figure S3.

RNA-seq disease network			Microarray disease network			RNA-seq healthy network			Microarray healthy network		
Modules	Vertices	Edges	Modules	Vertices	Edges	Modules	Vertices	Edges	Modules	Vertices	Edges
6	10460	899649	17	14437	612359	8	10460	1373581	6	14437	572961

Table S3 – Number of modules, vertices, and edges in RNA-Seq and microarray networks.

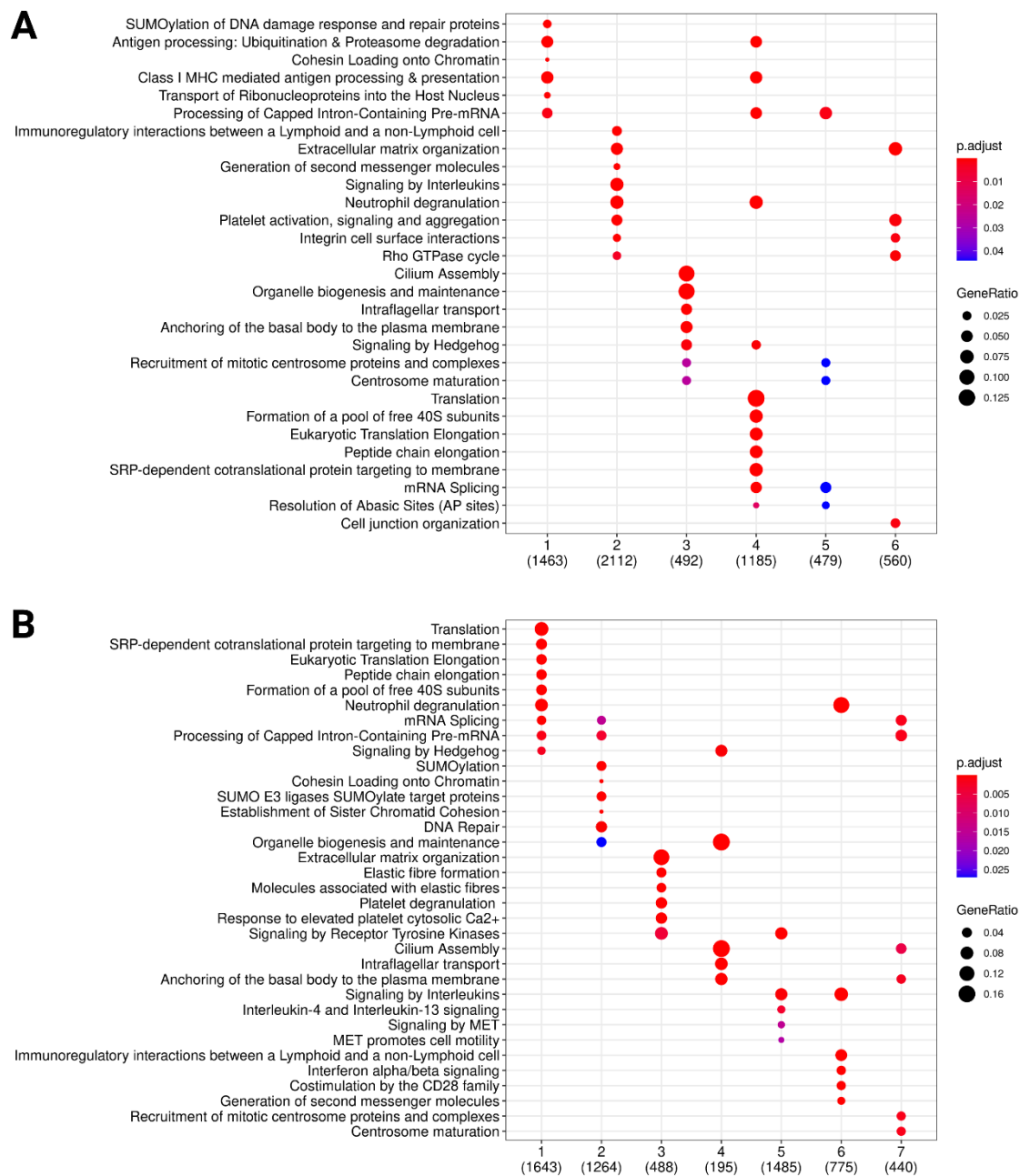


Figure S2 – Bubble plots of significant Reactome pathways in the RNA-Seq disease network (A) and RNA-Seq healthy network (B). X-axis represents the communities in the networks and Y-axis represents the Reactome pathways. The color of the bubble represents the significance of the pathway in the module and the size of the bubble represents the relative amount of genes in the module that are present in the Reactome pathway.

The network derived from the meta-analysis of RNA-sequencing data identified several modules associated with disease pathology. One of the primary disease-associated modules involves antigen presentation and processing, emphasizing the role of immune system activation in fibrosis (45). Also, immunoregulatory interactions between a Lymphoid and a non-Lymphoid cell Reactome pathway were significantly enriched in disease module two. Different chemokines are connected to the occurrence of IPF, and an uneven distribution of angiogenic chemokines has been correlated with vascular remodeling in the condition (45). Another significant module includes platelet activation, neutrophil degranulation, and extracellular compartment alterations, suggesting a strong inflammatory response and extracellular matrix remodeling, both critical in disease development. The ECM in the lungs consists of collagens, elastin, glycoproteins, and proteoglycans. These components play a central role in providing structural support for cells and ensuring the mechanical stability and elastic recoil essential for normal lung function. In IPF, a characteristic feature is the accumulation of myofibroblasts in clusters known as fibroblastic foci. This leads to a significant deposition of ECM within the interstitium, destroying lung architecture (46). Additionally, cilia impairment has been identified as a key factor, reinforcing its relevance in various respiratory conditions. The analysis also highlights significant alterations in translation, transcription, and mRNA metabolism, indicating broad regulatory changes at the molecular level that could contribute to disease-associated gene expression alterations. Furthermore, module five underscores disruptions in cell junction integrity and its links to the extracellular functions reported in module 2.

In contrast, the modules identified in healthy conditions highlight key biological processes crucial for homeostasis, including translation, transcription, and mRNA metabolism. Elastic fibers also emerge as an essential module, emphasizing their role in tissue elasticity and structural integrity. Notably, the control transcriptional profiles derive from healthy lungs, which therefore retain the majority of tissue functionalities. Cilia function remains a key feature in healthy conditions, mirroring its significance in disease when this function is altered.

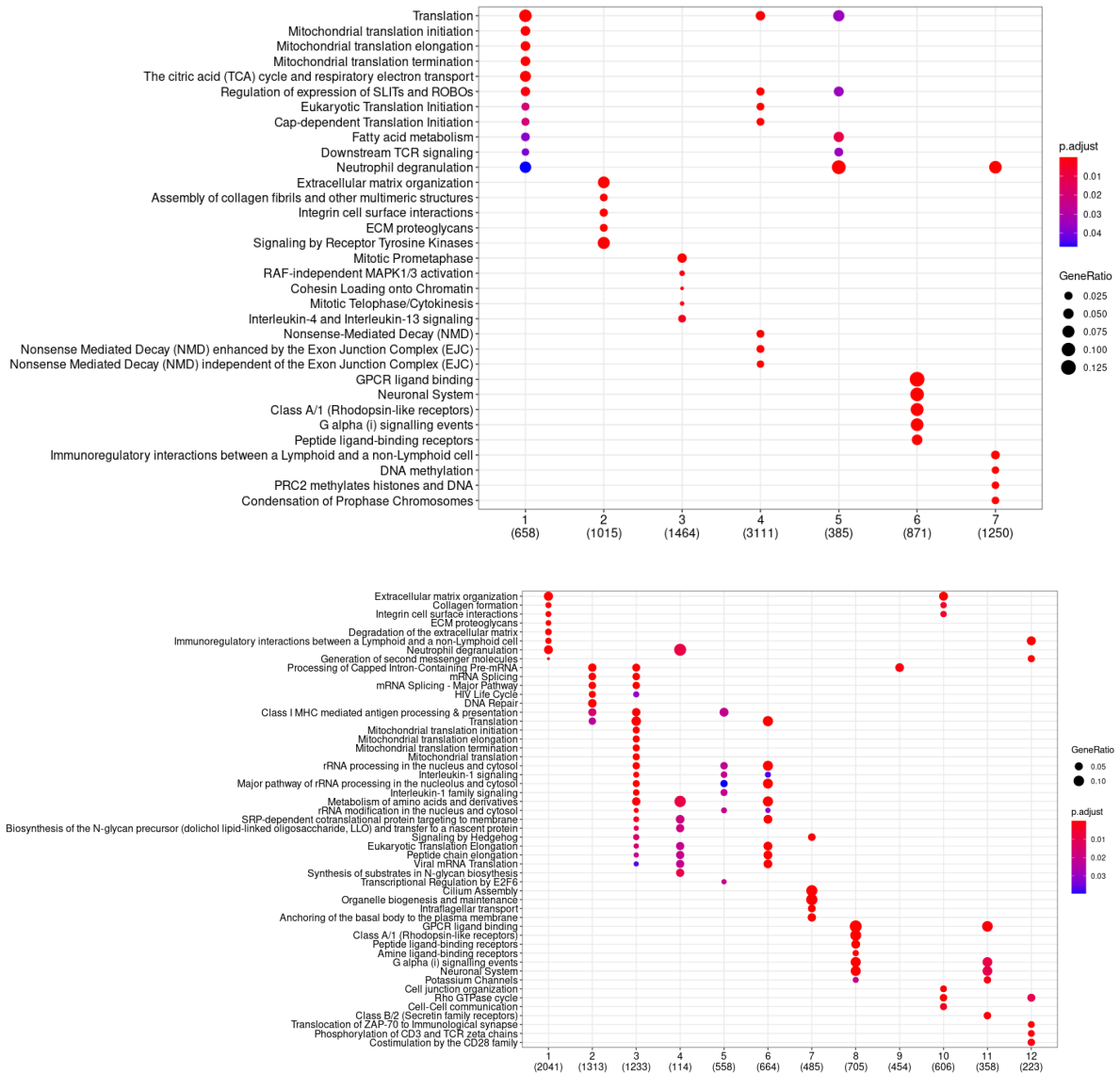


Figure S3 – Bubble plots of significant Reactome pathways in the integrated co-expression networks deriving from both RNA-Seq and microarrays. Panel A refers to the communities deriving from IPF samples and panel B refers to communities deriving from healthy samples. X-axis represents the community numbers and Y-axis represents the enriched pathways. The color of the bubble represents the significance of the pathway in the module and the size of the bubble represents the relative amount of genes in the module that are present in the pathway.

Subsequently, we showcase the late integration functionalities of MUUMI by integrating transcriptomics data deriving from both RNA-Seq and DNA microarrays. First, we identified the communities in each of the inferred networks for both of the technologies. We then used the “*aggregate_communities*” function to independently aggregate communities from RNA-Seq and DNA microarray experiments for both disease and healthy networks. The number of modules for the aggregated communities was determined by averaging and rounding to the nearest integer based on the module counts in the RNA-Seq and DNA microarray networks Table S3.

By incorporating DNA microarray data alongside RNA-Seq, the merged meta-analysis yielded cleaner and more consolidated results. Notably, functional redundancies observed in RNA-Seq disease modules, such as modules 2 and 5, appear condensed into a single module, streamlining the interpretation of immune-related and extracellular processes. Previously unobserved functions, such as the impact on mitochondria, have now emerged as key elements. This finding aligns with recent research demonstrating that mitochondrial quality significantly influences pulmonary fibrosis (PF) progression (47). Further considerations suggest that mitochondrial function plays a crucial role in disease progression. The emergence of mitochondrial dysfunction as a significant factor provides a novel perspective on pulmonary fibrosis pathogenesis. This insight is supported by recent findings highlighting mitochondrial quality as a key determinant of disease severity. Additionally, interleukin signaling, particularly IL-4 and IL-13, plays a crucial role in immune regulation and tissue maintenance. Furthermore, interferons, CD28, and other interleukins emerge as critical immune modulators maintaining a healthy physiological state.

Additionally, the role of nonsense-mediated mRNA decay (NMD) has been identified in module 4, implicating various mechanisms of gene expression regulation. NMD serves as a critical surveillance pathway for controlling mRNA quality and abundance. Gene expression deregulation through various epigenetic mechanisms is a known contributor to idiopathic pulmonary fibrosis (IPF). Interestingly, NMD based approaches have recently emerged as a therapeutic target in cystic fibrosis.

This functional interpretation offers valuable insights into the biological processes underlying disease states. The integration of multiple datasets not only refines the understanding of previously identified functionalities but also reveals novel functional associations, particularly with mitochondrial dysfunction and RNA surveillance mechanisms. Overall, the diverse array of pathways highlights the intricate molecular landscape involved in IPF, providing potential targets for therapeutic intervention and a deeper understanding of the disease mechanisms.

Finally, we computed the contributions of each of RNA-Seq and microarray technologies to the inference of the integrated communities. The contribution of each data source is calculated as the sum of its weights across all genes, normalized by the total weight for each community. These weights are derived using non-negative matrix factorization (NMF), which integrates gene communities identified in each data source and condition. Notably, some integrated communities exhibit relatively balanced contributions from all data sources (e.g., communities 1–7 in the diseased samples and community 1 in the healthy samples), indicating that these communities are shared across multiple data views. In contrast, the remaining communities show more uneven contributions, suggesting that these integrated communities are uniquely driven by a single data source. These findings underscore the strength of this integrative approach in both consolidating redundant communities across different data modalities and capturing unique, dataset-specific information that would otherwise remain undetected.

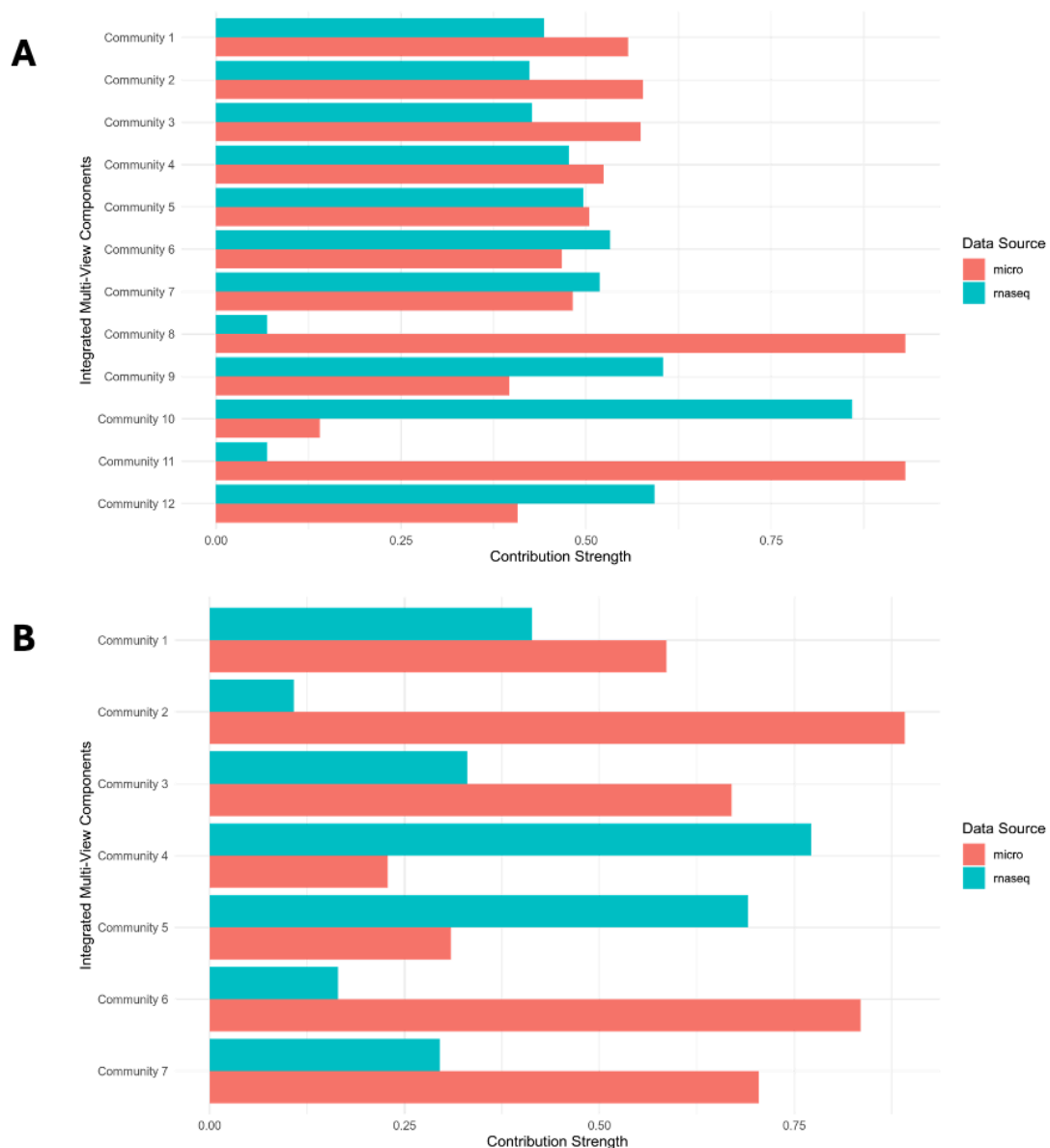


Figure S4 - Contribution of each individual data source to the integrated communities in both disease (panel A) and healthy (panel B) samples.

Case Study 2

Multi-omics data integration to uncover the plastic nature of macrophages upon polarization stimuli

This case study demonstrates the potential of MUUMI to integrate multi-omics datasets. Integrating multi-omics data is essential to gain a comprehensive understanding of complex biological systems, as it reveals the intricate regulatory mechanisms operating across different molecular layers. Combining transcriptomics, epigenomics, proteomics, and other omics layers provides insights into biological system regulation that single-omics approaches may overlook (48).

In this case study, we analyzed publicly available bulk RNA-Seq and DNA methylation array data previously collected from Gene Expression Omnibus (GEO), reported with the GEOID: GSE273628. In this study, Migliaccio et al. (49) explored the regulatory relationship between DNA methylation and gene expression in macrophages exposed to different polarization stimuli. Cells were differentiated with 30.9 ng/ml of phorbol 12-myristate 13-acetate (PMA) for 48h. After PMA differentiation, cells are exposed to two different cytokines cocktails and fresh media for 24h, 48h, and 72h. In particular, to simulate a pro-inflammatory environment the cells are exposed to LPS 10 pg/ml and INF γ 20 ng/ml; to induce an anti-inflammatory phenotype instead, the cells are exposed to IL-13 20 ng/ml and IL-4 20 ng/ml.

To gain a comprehensive understanding of the capabilities of the tool and its potential to uncover mechanisms underlying molecular responses to different polarization stimuli, we analyzed the epigenomic and transcriptomic profiles of macrophages stimulated with LPS-INF γ and IL4-IL13. To comply with the findings of Migliaccio et al. 2024, which emphasize the slower kinetics of DNA methylation compared to transcriptional changes, we selected the 72-hour time point for our analysis. This timeframe allows for the detection of more established epigenetic modifications, including those that may have originated at earlier time points.

To integrate these datasets, we utilized the *snf_based_integration* function implemented in MUUMI. Through our analysis, we generated two fusion networks: one for the LPS-INF γ and one for the IL4-IL13-treated macrophages. Such networks are completely connected, and each edge indicates the similarity values between genes. The similarity parameter assesses the degree of concordance between the expression patterns of gene pairs in both the transcriptomic and epigenomic layers. A higher similarity score suggests that the genes in a pair tend to exhibit similar regulatory patterns in transcriptomics and epigenomics layers. Both of the networks are composed of 12,940 nodes and 163,443,600 edges. To retain only the most biologically meaningful connections and make the computation faster, we pruned both of the networks through the function *prune_snf_network*, by setting the 90th percentile as a threshold of significance. The pruned LPS-INF γ network encompasses 8,502,486 edges while the IL4-IL13 one 8,500,640 edges.

Subsequently, we ranked the network edges based on their weights and performed a functional annotation using KEGG pathways through ESEA (Enrichment Set Enrichment Analysis). By performing an enrichment analysis on edges with high similarity scores, we can uncover biological processes that are strongly coordinated at gene expression and epigenetic level. To identify the most significant biological processes and pathways associated with each condition, we performed separate enrichment analysis on the LPS-INF γ and IL4-IL13 networks. The most significant enriched terms (adjusted pvalue <0.05) are presented in Figure S5.

Our integrated network analysis of the LPS-INF γ condition reveals a strong correlation between genes involved in TLR-receptor and TNF signaling pathways (50), consistent with the findings of Migliaccio et al. at the transcriptome level. LPS activates TLR4, initiating MyD88-dependent signaling that drives NF- κ B activation and TNF- α expression, amplifying inflammation through autocrine and paracrine effects (51).

In addition, this integrated network further demonstrates the crucial role of metabolic pathways in gene regulation (52), (53), as also observed by Migliaccio et al. Notably, genes implicated in chronic inflammatory diseases, such as rheumatoid arthritis, also exhibit high similarity within this integrated network. The presence of these epigenetic modifications suggests that alterations in the epigenetic regulation of

these genes may contribute to the development and maintenance of chronic inflammatory states. In contrast to the LPS-INF γ network, fewer pathways appear enriched within the IL4-IL13 network, demonstrating a milder response to IL4-IL13 stimulation compared to LPS-INF γ . Nevertheless, the presence of terms related to immune pathways suggests the existence of distinct gene groups exhibiting high correlation across both regulatory layers within this integrated network.

By facilitating network-based meta-analyses, MUUMI offers a versatile platform that enables researchers to identify crucial biological pathways, uncover synergistic regulatory effects, and effectively pinpoint key molecular targets. By employing the network analysis features of MUUMI, researchers can integrate omics networks, providing valuable insights into gene sets exhibiting high concordance in all regulatory layers.

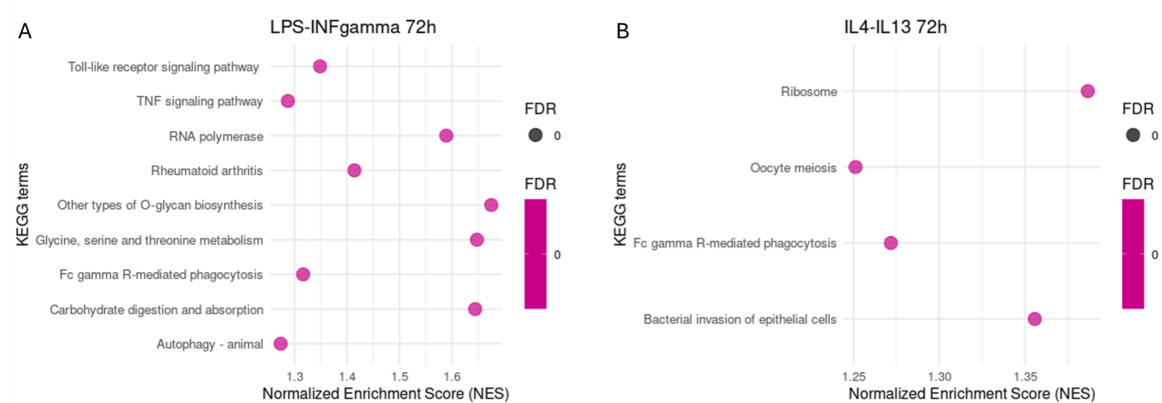


Figure S5 - Results of the Edge Set Enrichment Analysis (ESEA) on the multi-omics integrated networks of LPS-INF γ -treated macrophages (panel A) and the IL4-IL13-treated macrophages (panel B).

Limitations of the study

This study utilised multi-omics profiles obtained from public repositories. The availability of clinical data was usually insufficient. The limited detailed clinical information in many datasets manifests challenges to the predictive capacity of this study. It inhibits our ability to infer certain clinical parameters for patients with IPF, such as disease severity, time elapsed since the initial diagnosis, and the anatomical location of the samples. Additionally, the lack of information concerning the current administration of active pharmacological therapy to the patients is a noteworthy limitation. The transcriptional signatures shaping the topology of network models may be impacted by ongoing or terminated pharmacological therapies, potentially interfering with the pharmacological footprint investigated in this study. Consequently, the lack of comprehensive sample characterization restricts the translational potential of this research. The lack of randomization and data FAIRness in many datasets was problematic. The minimum standards for omics data often result in poor usability due to incomplete characterization of the experimental design and execution, as well as the lack of description regarding potential systematic effects caused by reagents, microarrays, and other factors (54).

Bibliography

1. Di Lieto E, Serra A, Inkala SI, Saarimäki LA, Del Giudice G, Fratello M, et al. ESPERANTO: a GLP-field sEmi-SuPERvised toxicogenomics metadAta curation TOOL. *Bioinformatics*. 2023 Jun 1;39(6).
2. Furusawa H, Cardwell JH, Okamoto T, Walts AD, Konigsberg IR, Kurche JS, et al. Chronic Hypersensitivity Pneumonitis, an Interstitial Lung Disease with Distinct Molecular Signatures. *Am J Respir Crit Care Med*. 2020 Nov 15;202(10):1430–44.
3. Jaffar J, Wong M, Fishbein GA, Alhamdoosh M, McMillan L, Gamell-Fulla C, et al. Matrix metalloproteinase-7 is increased in lung bases but not apices in idiopathic pulmonary fibrosis. *ERJ Open Research*. 2022 Oct 24;8(4).
4. McDonough JE, Ahangari F, Li Q, Jain S, Verleden SE, Herazo-Maya J, et al. Transcriptional regulatory model of fibrosis progression in the human lung. *JCI Insight*. 2019 Nov 14;4(22):e131597.
5. Huang Y, Guzy R, Ma S-F, Bonham CA, Jou J, Schulte JJ, et al. Central lung gene expression associates with myofibroblast features in idiopathic pulmonary fibrosis. *BMJ Open Respir Res*. 2023 Feb;10(1).
6. Schafer M, White T, Iijima K, Mazula D, Passos J, Kirkland JL, et al. CELLULAR SENESENCE DRIVES FIBROTIC PULMONARY DISEASE. *Innov Aging*. 2017 Jul 1;1(suppl_1):135–135.
7. DePianto DJ, Chandriani S, Abbas AR, Jia G, N'Diaye EN, Caplazi P, et al. Heterogeneous gene expression signatures correspond to distinct lung pathologies and biomarkers of disease severity in idiopathic pulmonary fibrosis. *Thorax*. 2015 Jan;70(1):48–56.

8. DePianto DJ, Heiden JAV, Morshead KB, Sun K-H, Modrusan Z, Teng G, et al. Molecular mapping of interstitial lung disease reveals a phenotypically distinct senescent basal epithelial cell population. *JCI Insight*. 2021 Apr 22;6(8).
9. Cecchini MJ, Hosein K, Howlett CJ, Joseph M, Mura M. Comprehensive gene expression profiling identifies distinct and overlapping transcriptional profiles in non-specific interstitial pneumonia and idiopathic pulmonary fibrosis. *Respir Res*. 2018 Aug 15;19(1):153.
10. Cho J-H, Gelinas R, Wang K, Etheridge A, Piper MG, Batte K, et al. Systems biology of interstitial lung diseases: integration of mRNA and microRNA expression changes. *BMC Med Genomics*. 2011 Jan 17;4:8.
11. Meltzer EB, Barry WT, D'Amico TA, Davis RD, Lin SS, Onaitis MW, et al. Bayesian probit regression model for the diagnosis of pulmonary fibrosis: proof-of-principle. *BMC Med Genomics*. 2011 Oct 5;4:70.
12. Geng J, Huang X, Li Y, Xu X, Li S, Jiang D, et al. Down-regulation of USP13 mediates phenotype transformation of fibroblasts in idiopathic pulmonary fibrosis. *Respir Res*. 2015 Oct 9;16:124.
13. Brereton CJ, Yao L, Davies ER, Zhou Y, Vukmirovic M, Bell JA, et al. Pseudohypoxic HIF pathway activation dysregulates collagen structure-function in human lung fibrosis. *eLife*. 2022 Feb 21;11.
14. Luzina IG, Salcedo MV, Rojas-Peña ML, Wyman AE, Galvin JR, Sachdeva A, et al. Transcriptomic evidence of immune activation in macroscopically normal-appearing and scarred lung tissues in idiopathic pulmonary fibrosis. *Cell Immunol*. 2018 Mar;325:1–13.
15. Konishi K, Gibson KF, Lindell KO, Richards TJ, Zhang Y, Dhir R, et al. Gene expression profiles of acute exacerbations of idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med*. 2009 Jul 15;180(2):167–75.
16. Yin Q, Strong MJ, Zhuang Y, Flemington EK, Kaminski N, de Andrade JA, et al. Assessment of viral RNA in idiopathic pulmonary fibrosis using RNA-seq. *BMC Pulm Med*. 2020 Apr 3;20(1):81.
17. King TE, Pardo A, Selman M. Idiopathic pulmonary fibrosis. *Lancet*. 2011 Dec 3;378(9807):1949–61.
18. Martinez FJ, Collard HR, Pardo A, Raghu G, Richeldi L, Selman M, et al. Idiopathic pulmonary fibrosis. *Nat Rev Dis Primers*. 2017 Oct 20;3:17074.
19. Wuyts WA, Agostini C, Antoniou KM, Bouros D, Chambers RC, Cottin V, et al. The pathogenesis of pulmonary fibrosis: a moving target. *Eur Respir J*. 2013 May;41(5):1207–18.
20. Misharin AV, Morales-Nebreda L, Reyfman PA, Cuda CM, Walter JM, McQuattie-Pimentel AC, et al. Monocyte-derived alveolar macrophages drive lung fibrosis and persist in the lung over the life span. *J Exp Med*. 2017 Aug 7;214(8):2387–404.
21. Saarimäki LA, Kinaret PAS, Scala G, del Giudice G, Federico A, Serra A, et al. Toxicogenomics analysis of dynamic dose-response in macrophages highlights molecular alterations relevant for multi-walled carbon nanotube-induced lung fibrosis.

NanoImpact. 2020 Nov;100274.

22. Sugeir S, de Moraes AG. Bronchoscopy in the intensive care unit. In: Demetriades D, Inaba K, Lumb PD, editors. Atlas of critical care procedures. Cham: Springer International Publishing; 2018. p. 49–55.
23. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 2014;15(12):550.
24. Marwah VS, Scala G, Kinaret PAS, Serra A, Alenius H, Fortino V, et al. eUTOPIA: solUTion for Omics data Preprocessing and Analysis. *Source Code Biol Med.* 2019 Jan 29;14:1.
25. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 2015 Apr 20;43(7):e47.
26. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological).* 1995 Jan;57(1):289–300.
27. Zolak JS, de Andrade JA. Idiopathic pulmonary fibrosis. *Immunol Allergy Clin North Am.* 2012 Nov;32(4):473–85.
28. Zhou Z, Qu J, He L, Zhu Y, Yang S-Z, Zhang F, et al. Stiff matrix instigates type I collagen biogenesis by mammalian cleavage factor I complex-mediated alternative polyadenylation. *JCI Insight.* 2020 Feb 13;5(3).
29. Mei Q, Liu Z, Zuo H, Yang Z, Qu J. Idiopathic pulmonary fibrosis: an update on pathogenesis. *Front Pharmacol.* 2021;12:797292.
30. Huang S, Lai X, Yang L, Ye F, Huang C, Qiu Y, et al. Asporin Promotes TGF- β -induced Lung Myofibroblast Differentiation by Facilitating Rab11-Dependent Recycling of T β RI. *Am J Respir Cell Mol Biol.* 2022 Feb;66(2):158–70.
31. Frangogiannis N. Transforming growth factor- β in tissue fibrosis. *J Exp Med.* 2020 Mar 2;217(3):e20190103.
32. Bauer Y, White ES, de Bernard S, Cornelisse P, Leconte I, Morganti A, et al. MMP-7 is a predictive biomarker of disease progression in patients with idiopathic pulmonary fibrosis. *ERJ Open Research.* 2017 Jan;3(1).
33. Mahalanobish S, Saha S, Dutta S, Sil PC. Matrix metalloproteinase: An upcoming therapeutic approach for idiopathic pulmonary fibrosis. *Pharmacol Res.* 2020 Feb;152:104591.
34. Yu G, Tzouveleakis A, Wang R, Herazo-Maya JD, Ibarra GH, Srivastava A, et al. Thyroid hormone inhibits lung fibrosis in mice by improving epithelial mitochondrial function. *Nat Med.* 2018 Jan;24(1):39–49.
35. Liberzon A, Subramanian A, Pinchback R, Thorvaldsdóttir H, Tamayo P, Mesirov JP. Molecular signatures database (MSigDB) 3.0. *Bioinformatics.* 2011 Jun 15;27(12):1739–40.
36. Åhrman E, Hallgren O, Malmström L, Hedström U, Malmström A, Bjermer L, et al. Quantitative proteomic characterization of the lung extracellular matrix in chronic obstructive pulmonary disease and idiopathic pulmonary fibrosis. *J Proteomics.* 2018

Oct 30;189:23–33.

37. Qian W, Xia S, Yang X, Yu J, Guo B, Lin Z, et al. Complex involvement of the extracellular matrix, immune effect, and lipid metabolism in the development of idiopathic pulmonary fibrosis. *Front Mol Biosci*. 2021;8:800747.
38. Federico A, Hautanen V, Christian N, Kremer A, Serra A, Greco D. Preprocessed and Harmonised Transcriptomics Datasets for Psoriasis and Atopic Dermatitis. Zenodo. 2020;
39. T. Hastie, R. Tibshirani, Balasubramanian Narasimhan and Gil Chu. Pam: Prediction Analysis for Microarrays. R package version 1.56.1. <https://CRAN.R-project.org/package=pamr>. CRAN R-project; 2019.
40. Pons P, Latapy M. Computing communities in large networks using random walks. *J Graph Algorithms Appl*. 2006;10(2):191–218.
41. Reichardt J, Bornholdt S. Statistical mechanics of community detection. *Phys Rev E*. 2006 Jul 18;74(1).
42. Newman MEJ, Girvan M. Finding and evaluating community structure in networks. *Phys Rev E Stat Nonlin Soft Matter Phys*. 2004 Feb 26;69(2 Pt 2):026113.
43. Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E. Fast unfolding of communities in large networks. *J Stat Mech*. 2008 Oct 9;2008(10):P10008.
44. Clauset A, Newman MEJ, Moore C. Finding community structure in very large networks. *Phys Rev E*. 2004 Dec 6;70(6).
45. Butler MW, Keane MP. The role of immunity and inflammation in IPF pathogenesis. In: Meyer KC, Nathan SD, editors. *Idiopathic pulmonary fibrosis: A comprehensive clinical guide*. Cham: Springer International Publishing; 2019. p. 97–131.
46. Upagupta C, Shimbori C, Alsilmi R, Kolb M. Matrix abnormalities in pulmonary fibrosis. *Eur Respir Rev*. 2018 Jun 30;27(148).
47. Hara H, Kuwano K, Araya J. Mitochondrial quality control in COPD and IPF. *Cells*. 2018 Jul 24;7(8).
48. Hasin Y, Seldin M, Lusis A. Multi-omics approaches to disease. *Genome Biol*. 2017 May 5;18(1):83.
49. Migliaccio G, Morikka J, Del Giudice G, Vaani M, Möbus L, Serra A, et al. Methylation and transcriptomic profiling reveals short term and long term regulatory responses in polarized macrophages. *Comput Struct Biotechnol J*. 2024 Dec;25:143–52.
50. Ciesielska A, Matyjek M, Kwiatkowska K. TLR4 and CD14 trafficking and its influence on LPS-induced pro-inflammatory signaling. *Cell Mol Life Sci*. 2021 Feb;78(4):1233–61.
51. Kawai T, Akira S. The role of pattern-recognition receptors in innate immunity: update on Toll-like receptors. *Nat Immunol*. 2010 May;11(5):373–84.
52. Fanucchi S, Domínguez-Andrés J, Joosten LAB, Netea MG, Mhlanga MM. The intersection of epigenetics and metabolism in trained immunity. *Immunity*. 2021 Jan 12;54(1):32–43.
53. Lauterbach MA, Hanke JE, Serefidou M, Mangan MSJ, Kolbe C-C, Hess T, et al.

Toll-like Receptor Signaling Rewires Macrophage Metabolism and Promotes Histone Acetylation via ATP-Citrate Lyase. *Immunity*. 2019 Dec 17;51(6):997-1011.e7.

54. Saarimäki LA, Melagraki G, Afantitis A, Lynch I, Greco D. Prospects and challenges for FAIR toxicogenomics data. *Nat Nanotechnol*. 2022 Jan;17(1):17–8.