Lee *et al.*
European Journal of Medical Research (2026) 31:299
https://doi.org/10.1186/s40001-026-03987-9

European Journal
of Medical Research

RESEARCH

Open Access

Radiologic phenotype-specific transcriptomic signatures in lung tissues from patients with *Mycobacterium avium* complex pulmonary disease

Ju Mi Lee[1†], Su-Young Kim[2†], Seong Mi Moon[3], Sung Jae Shin[1*] and Byung Woo Jhun[2*]

Abstract

Background The incidence of *Mycobacterium avium* complex (MAC)-pulmonary disease (PD) is increasing in South Korea, posing significant diagnostic and therapeutic challenges. Treatment guidelines recommend initiating therapy after serial computed tomography monitoring. Patients with the nodular bronchiectatic (NB) form often respond positively to drug therapy, whereas those with the fibrocavitary (FC) form frequently experience persistent disease despite treatment. Identifying phenotype-specific transcriptomic biomarkers could improve early diagnosis and inform personalized therapeutic strategies.

Methods We utilized surgically resected lung specimens from 21 MAC-PD patients, a valuable clinical resource, as lung surgery is uncommon in MAC-PD management. Each patient provided paired samples of affected and unaffected lung tissues, enabling direct transcriptomic comparisons. Quantitative RNA sequencing was performed on samples from 11 NB and 10 FC cases. Comprehensive bioinformatics and in silico analyses, including gene ontology (GO) and protein–protein interaction (PPI) network analyses, were conducted to identify key diagnostic signatures and biological pathways.

Results RNA sequencing revealed distinct and shared transcriptomic signatures correlated with radiological phenotypes. GO and PPI analyses identified significant gene clusters involved in B cell proliferation and immune regulation across both NB and FC forms. Additionally, NB-specific signatures highlighted genes predominantly regulating antimicrobial immune responses, while FC-specific signatures enriched genes related to extracellular matrix remodeling.

Conclusions This study is the first to characterize transcriptomic differences between MAC-PD phenotypes using paired lung tissue samples. Although the identified transcriptomic markers require functional validation, their strong correlation with radiologic subtypes provides preliminary evidence supporting their potential diagnostic value. These findings lay the groundwork for precision diagnostics in MAC-PD and require further validation in larger patient cohorts and through functional assays.

[†]Ju Mi Lee and Su-Young Kim have contributed equally to this work.

*Correspondence:
Sung Jae Shin
sjshin@yuhs.ac
Byung Woo Jhun
byungwoo.jhun@gmail.com
Full list of author information is available at the end of the article

Keywords *Mycobacterium avium* complex-pulmonary disease, Radiologic phenotypes, nodular bronciectatic, fibrocavitary, Transcriptomic profiling

Introduction

Nontuberculous mycobacteria (NTM) comprise a varied array of mycobacterial species, distinct from the *Mycobacterium tuberculosis* complex and *M. leprae* [1]. Within NTM, *M. avium* complex (MAC), including *M. avium* and *M. intracellulare*, stands as a predominant causative agent of pulmonary disease (PD) in South Korea, mirroring its prevalence in other nations [2, 3].

MAC is predominantly observed in pulmonary samples, further emphasizing its significance in the context of respiratory health [4, 5]. Radiologically, MAC-PD can be dichotomously classified into two distinct types: nodular bronchiectatic (NB) and fibrocavitary (FC) [6, 7]. A more granular categorization reveals four specific types: noncavitary NB (NC-NB), cavitary NB (C-NB), FC, and the exceedingly rare unclassifiable type [6, 7]. The NB form has been frequently observed as a characteristic clinical phenotype in middle-aged or older women with no prior chronic pulmonary disease or smoking history [8]. Conversely, the FC variant is renowned among middle-aged or older men with a notable background of smoking and alcohol consumption, often concomitant with underlying conditions like chronic pulmonary disease (e.g., chronic obstructive pulmonary disease, surgical lung resection, and cystic fibrosis) or a past bout of pulmonary tuberculosis [9]. Notably, the NB form tends to exhibit a comparatively slower progression, with radiological changes evolving gradually over several years in certain patients, necessitating prolonged and vigilant follow-up [9, 10].

Radiologically different forms exhibited discernible differences in treatment outcomes. Notably, the achievement of favorable outcomes was recognized to be notably higher in those with the NC-NB form (88%), eclipsing rates observed in C-NB and FC forms (78% and 76%, respectively) [6]. Characteristically, individuals with the NB types were more prone to experiencing recurrence or reinfection [11, 12]. In contrast, mortality was significantly higher in the C-NB and FC forms than in the NC-NB forms across multiple observational cohorts worldwide [2, 13–17]. Despite these clinical disparities, the underlying molecular mechanisms associated with specific radiological phenotypes remain largely uncharacterized, with transcriptomic studies of human lung tissues being particularly limited.

Transcriptomic profiling allows for a systematic assessment of the functional and immunological state of lung tissue, capturing coordinated changes of host immune activation, tissue remodeling, and inflammatory pathways that may not be readily inferred by single-marker analyses. This is particularly relevant in MAC-PD, where radiological heterogeneity is thought to reflect complex, locally regulated host responses rather than differences in pathogen burden alone. Accordingly, this approach facilitates the identification of biological processes functionally linked to distinct MAC-PD phenotypes. Therefore, a bioinformatic approach is required to delineate the molecular signatures associated with these distinct prognoses.

Hence, in this investigation, mRNA sequencing was conducted on lung tissues extracted from 21 MAC-PD patients undergoing lobectomy. We aimed to elucidate factors associated with the disease progression across various radiological forms. A comparison of transcriptomes was performed, encompassing 11 sections from NB forms-affected regions, 10 sections from FC forms-affected regions, and 21 sections from areas unaffected by MAC infection. Collectively, these findings could potentially lead to the prediction of prognostic markers and therapeutic targets in the field of MAC-PD.

Methods

Study participants and collected tissue specimen

This study is a gene-based data analysis using lung tissue samples and data from 21 NTM-PD patients who consented to tissue collection at Samsung Medical Center, Seoul, South Korea. The study was approved by the Institutional Review Board of the Samsung Medical Center

Table 1 Baseline characteristics of 21 patients who underwent surgical resection

	N = 21
Age years, median (IQR)	56 (52–62)
Female sex, N (%)	11 (52)
Radiological images, N (%)	
NB type	11 (52)
FC type	10 (48)
Etiologic organism, N (%) *M. avium*, N (%) *M. intracellulare*, N (%)	8 (38) 13 (62)
Treatment duration, months, median (IQR)	25 (15–34)
MIC of clarithromycin, MIC_{50} (range, mg/ml)	1 (0.5–4)
MAC culture positivity without mutation, N (%)	21 (100)
Culture conversion 12 months after surgery, N (%)	18 (86)

IQR, Interquartile range; NB, nodular bronchiectatic; FC, fibrocavitary; MIC, minimum inhibitory concentration; MIC_{50}, MIC at which 50% of isolates are inhibited; MAC, *Mycobacterium avium* complex

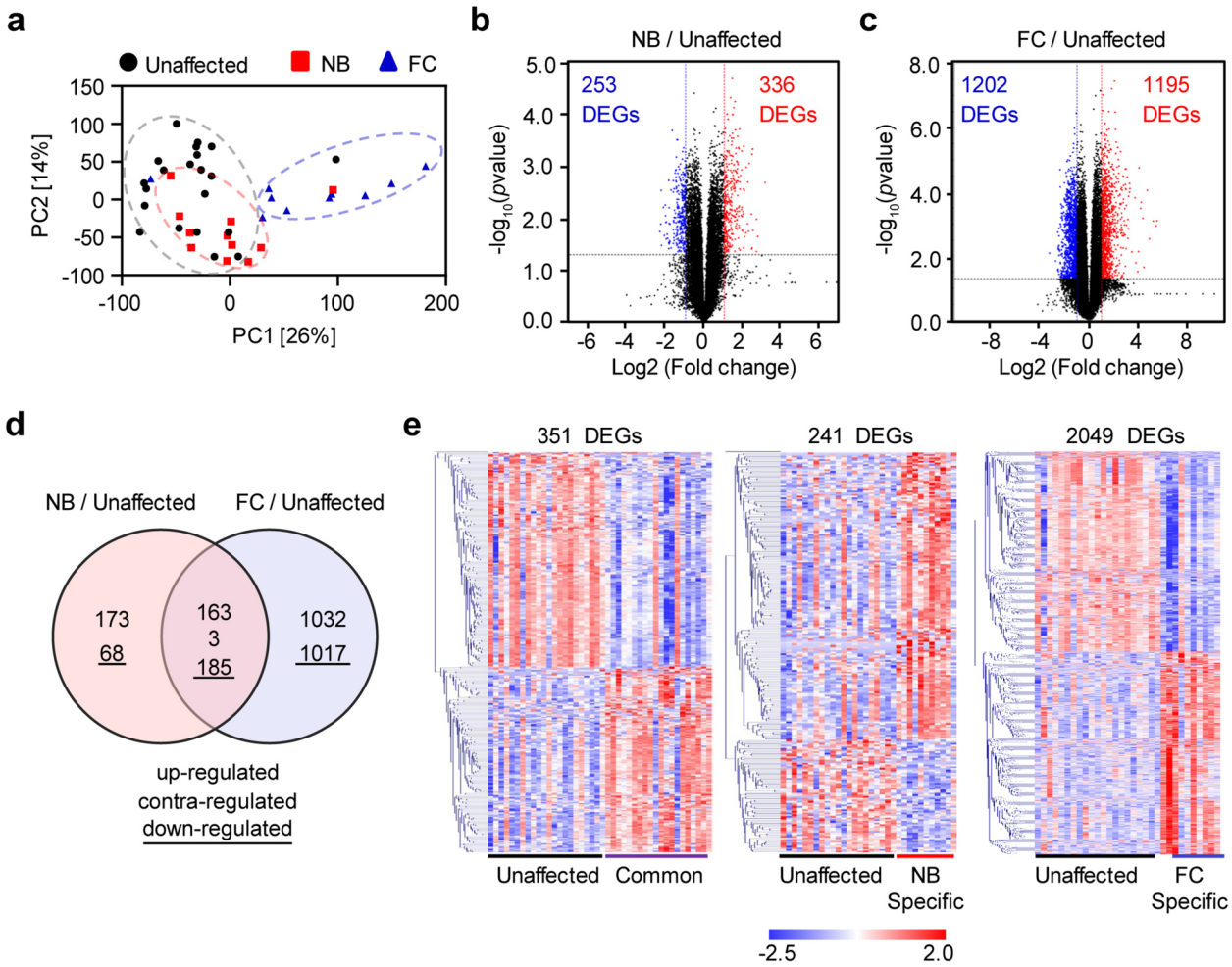


Fig. 1 Differences in lung DEGs between two different radiologic subtypes among 21 MAC-PD patients. **a** Principal component analysis (PCA) following determination of DEGs. Black circles represent unaffected samples from the entire cohort of MAC-PD patients (N = 21), red squares denote the NB form (N = 11), and blue triangles indicate the FC form (N = 10). Volcano plots illustrate the relationship between gene fold changes and statistical significance in MAC-PD, depicted as **b** NB form versus unaffected and **c** FC form versus unaffected. Down-regulated values are denoted in blue, whereas up-regulated values are indicated in red. **d** A Venn diagram illustrates the count of DEGs identified, highlighting those unique to or shared between the radiological phenotypes. **e** The heatmap displays the hierarchical clustering of expression patterns. All the DEGs adhere to the following criteria: |fold change| ≥ 2, $\log_2$ (fold change) ≥ 1, $p < 0.05$. DEGs, differentially expressed genes; MAC-PD, *M. avium* complex-pulmonary disease; NB, nodular bronchiectatic; FC, fibrocavitary; Common, common DEGs from NB and FC form samples

(IRB No. 2012-05-001). Those patients failed with the antibiotic therapy, but they were all susceptible to macrolide and underwent adjuvant surgical lung resection between July 2012 and April 2018 (Table 1). To examine biological variances between two distinct clinical phenotypes (NB or FC forms) based on unaffected sections, lung tissue was procured in pairs from each patient, as previously described [18]. Then, one sample originated from an affected lesion, and the other was extracted from an unaffected region. Resection criteria included consideration of the cavity lesion in FC form or the bronchiectasis lesion in NB form, along with the unaffected

(See figure on next page.)

Fig. 2 Top GO terms and KEGG pathways for 351 common DEGs in NB and FC forms. **a** GO analyses (including biological process [BP], cellular component [CC], and molecular function [MF]) and **b** KEGG pathway analysis revealed the top five and ten terms for up-regulated genes (163 DEGs) and the down-regulated genes (185 DEGs), respectively. The x-axis represents the gene ratio (the ratio of total enriched DEGs in the given terms), while the y-axis represents the categories. *p*-value (Fisher's exact test) is depicted as '-$\log_{10}$ (*p*value)', and the number of DEGs is indicated as 'count'. GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; Common, common DEGs from NB and FC form patient samples

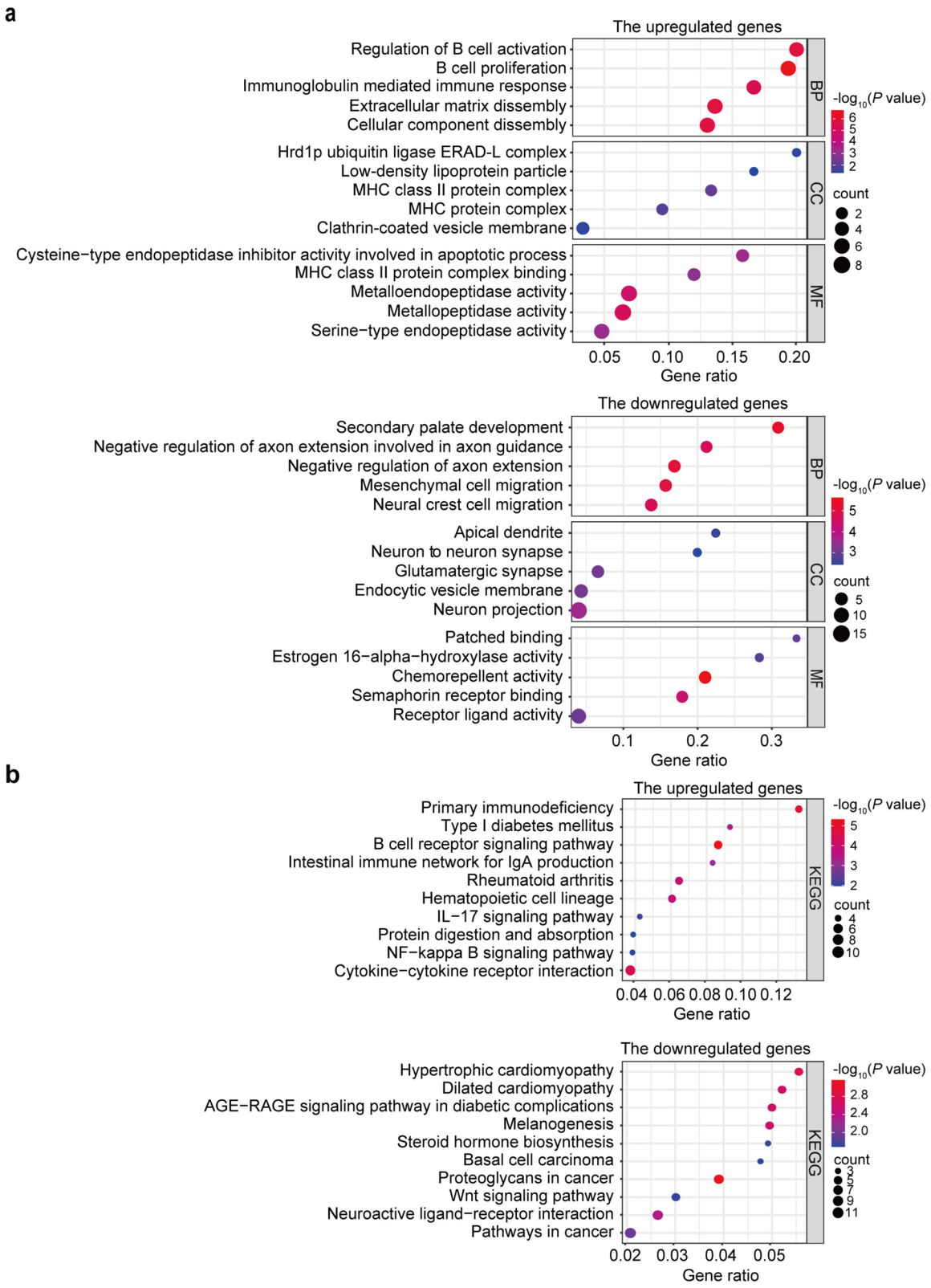


Fig. 2 (See legend on previous page.)

area, guided by preoperative computed tomography and macroscopic examination in the operating room, resulting in specimens sized 20 mm × 20 mm × 10 mm. These specimens were promptly submerged in cold phosphate-buffered saline, sectioned into smaller fragments, swiftly frozen using liquid nitrogen, and preserved at -80 °C.

RNA extraction from lung tissue and RNA sequencing experiments

For the sample collection, lung tissues were dissected and harvested in TRIzol reagent (Invitrogen, Waltham, MA, USA). RNA quality was assessed using an Agilent 2100 bioanalyzer with an RNA 6000 Nano Chip (Agilent Technologies, Amstelveen, The Netherlands), and RNA concentration was determined using an ND-2000 spectrophotometer (Thermo Fisher™, Waltham, MA, USA). The RNA Integrity Number (RIN) was measured with an Agilent 2100 Bioanalyzer, ensuring all samples had RIN ≥ 6.

Poly (A) + mRNA was isolated from 2 µg of total RNA using the poly (A) RNA Selection Kit (LEXOGEN, Inc., Austria). RNA-seq libraries were prepared using the SMARTer Stranded RNA-Seq Kit (Clontech Laboratories, Inc., Mountain View, CA, USA), involving cDNA synthesis and RNA fragmentation. Indexing was performed using Illumina indexes 1–12, and library enrichment was carried out via PCR amplification. High-throughput sequencing was performed as paired-end 100 bp reads on an Illumina HiSeq 2500 platform (Illumina, Inc., San Diego, CA, USA). Raw sequencing data underwent quality control using FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Subsequently, adapters and low-quality reads (< Q20) were eliminated, and trimmed reads were aligned to the reference genome with TopHat [19]. Gene expression levels were estimated based on read count (RC) and fragments per kb per million reads (FPKM) values using BEDTools [20] and Cufflinks [21]. Normalization was performed with EdgeR within R using Quantile normalization (http://www.R-project.org, R Development Core Team, 2016). Data mining and visualization were carried out using Excel-based Differentially Expressed Gene Analysis (ExDEGA, Ebiogen, Inc., Seoul, South Korea). The RNA-sequencing data have been deposited in the Gene Expression Omnibus (GEO) database with accession number GSE270278.

Gene and pathway enrichment

The Differentially Expressed Genes (DEGs) for the group exhibiting both forms versus unaffected sections were identified utilizing ExDEGA. DEGs were defined based on a multi-parameter filtering strategy that incorporated both statistical significance and expression magnitude. Specifically, genes were selected using the following criteria: a ≥ twofold change in expression, a nominal p-value < 0.05 (Student's t-test), and a normalized expression level ($\log_2$) ≥ 1 to exclude low-abundance transcripts. Hierarchical clustering analyses were executed with the MeV program, employing Euclidean distance correlation measurement with average linkage [22], and the results were visualized as heat maps. Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analyses were performed using the EnrichR tool (https://maayanlab.cloud/Enrichr/). Enriched terms were ranked according to nominal p-values as generated by EnrichR. Notably, ranking based on nominal p-values was identical to that obtained using false discovery rate (FDR)-adjusted p-values. Enrichment statistics, including gene overlap, nominal p-values, FDR-adjusted p-values, odds ratios, and combined scores integrating the p-values and the z-scores, were reported in Table S1-S6. For visualization, enrichment results were displayed as bubble plots through SRplot (https://www.bioinformatics.com.cn/en). In these plots, gene ratio, -$\log_{10}$(p-value), and gene count respectively indicate the enrichment magnitude, statistical significance, and the count of gene in each term [23].

Protein–protein interaction analysis

For the protein–protein interaction (PPI) analysis, our data were meticulously visualized and scrutinized using the Search Tool for the Retrieval of Interacting Genes/Proteins (STRING), the module of molecular complex detection (MCODE), and CytoHubba database networks. These tools function as plug-in applications within the Cytoscape program (version 3.10.1; https://cytoscape.org/), a publicly accessible software celebrated for its prowess in analyzing complex biological networks, offering graphical editing capabilities, and rendering intricate

(See figure on next page.)

Fig. 3 Top GO terms and KEGG pathways for 241 NB form-specific DEGs. **a** GO analyses (including BP, CC, and MF) and **b** KEGG pathway analysis revealed the top five and ten terms for up-regulated genes (173 DEGs) and the down-regulated genes (68 DEGs), respectively. The x-axis represents the gene ratio (the ratio of total enriched DEGs in the given terms), while the y-axis represents the categories. p-value (Fisher's exact test) is depicted as '– $\log_{10}$ ($p$value)', and the number of DEGs is indicated as 'count'. GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; NB, nodular bronchiectatic

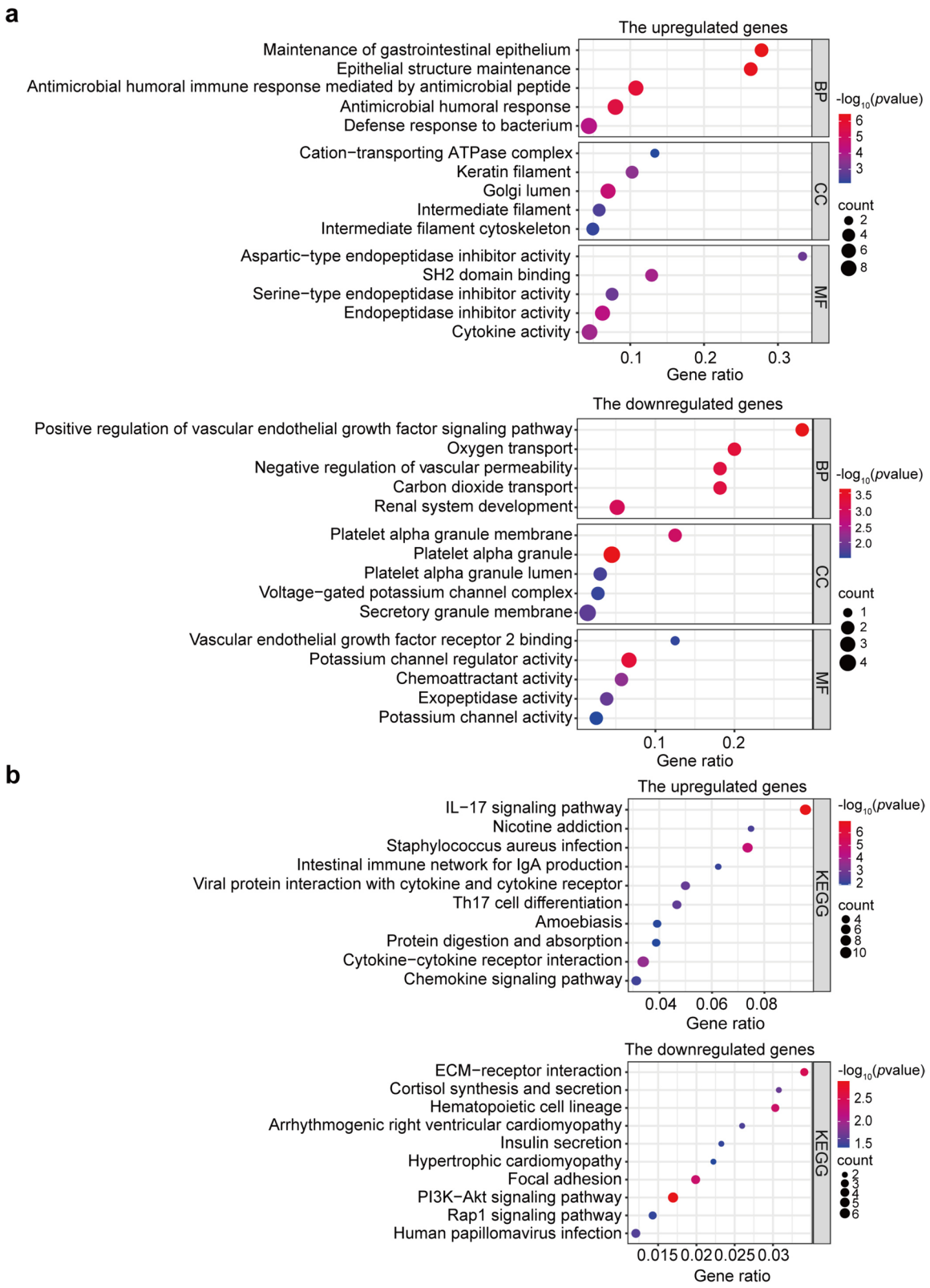


Fig. 3 (See legend on previous page.)

network representations. The Venn diagrams presented in Fig. S3 elucidate the intersection of the top 10 genes identified by three algorithms (Degree, MNC, and MCC), whereas the Venn diagrams in Figs. 5, 6 and 7 illustrate the intersection of pathways identified with genes from the top four modules.

Statistical analysis

The categorical outcomes of clinical information were depicted as frequencies and percentages, and continuous variables were presented as median and quartiles. Fisher's exact test compared categorical variables, and the Mann–Whitney U test compared continuous variables. Statistical significance was determined at $p < 0.05$ in two-sided tests.

Results

Study participants

Table 1 summarizes the characteristics of the 21 patients at the time of lung resection. The median age was 56 years, with 11 (52%) women and 10 men. Among them, 11 patients presented the NB clinical phenotype, while the remaining 10 exhibited the FC form. Most patients were infected with *M. intracellulare* (N = 13), compared to *M. avium* (N = 8). Up to the surgery, the median duration of drug therapy was 25 months, with no instances of macrolide resistance observed. Culture conversion status was achieved by 18 (86%) patients at the time of surgery. The median period between initial diagnosis and surgery was 2.2 years (ranging from 20 days to 6.9 years).

DEG profiling in MAC infection-associated lesions of patients

We collected surgical specimens from a cohort of 21 macrolide-susceptible cases, categorized into 11 NB and 10 FC forms based on radiological phenotypes (Fig. S1). Due to the unavailability of healthy lung controls, paired samples from affected and unaffected lung sections were obtained after lobectomy. Following RNA extraction and library preparation, mRNA sequencing was performed (Fig. S2). DEGs between affected and unaffected lung sections were identified separately for NB and FC forms. Subsequent analyses included pathway enrichment, GO enrichment, PPI network construction, and module analysis to identify hub genes (Fig. S2).

The principal component analysis based on DEGs revealed clear separation among unaffected sections, NB forms and FC forms, with contribution rates of 26% and 14% from the first and second principal components, respectively (Fig. 1a). Using threshold criteria of |fold change| ≥ 2, normalized expression level ($\log_2$) ≥ 1, and p-value < 0.05, volcano plot analysis identified 589 DEGs between NB and unaffected samples, and 2397 DEGs between FC and unaffected samples (Fig. 1b,c).

To clarify the relationship between radiological phenotype-specific genes and disease progression, we compared DEGs from each phenotype pair, illustrated in Venn diagrams and hierarchical clustering heatmaps (Fig. 1d,e). Specifically, 163 upregulated, 3 contra-regulated, and 185 downregulated genes were common to both NB and FC forms. Additionally, among the NB-specific DEGs, 173 genes were upregulated and 68 were downregulated. In contrast, among the FC-specific DEGs, 1032 genes were upregulated, and 1017 were downregulated. (Fig. 1d).

GO enrichment and KEGG pathway for shared DEGs between NB and FC forms of MAC-PD patients

We conducted gene enrichment analyses based on GO terms and the KEGG pathway database using EnrichR to identify common characteristics between NB and FC phenotypes of MAC-PD. Figure 2 illustrates the top 5 enriched GO terms across Biological Processes (BP), Cellular Components (CC), and Molecular Functions (MF), alongside the top 10 KEGG pathways for both upregulated and downregulated DEGs. GOBP illustrated that upregulated DEGs were primarily associated with B cell activation and proliferation, and immunoglobulin-mediated immune response, while downregulated DEGs were mainly involved in secondary palate development and axon guidance regulation. GOCC analysis showed upregulated genes enriched in Hrd1p ubiquitin ligase ERAD-L complex, low-density lipoprotein particles, and MHC class II complex. Downregulated genes were linked to neuronal synapses. GOMF analysis indicated upregulated genes related to cysteine-type endopeptidase inhibition and MHC class II protein complex binding, while

(See figure on next page.)

Fig. 4 Top GO terms and KEGG pathways for 2049 FC form-specific DEGs. **a** GO analyses (including BP, CC, and MF) and **b** KEGG pathway analysis revealed the top five and ten terms for up-regulated genes (1032 DEGs) and the down-regulated genes (1017 DEGs), respectively. The x-axis represents the gene ratio (the ratio of total enriched DEGs in the given terms), while the y-axis represents the categories. *p*-value (Fisher's exact test) is depicted as '$-\log_{10}$ (*p*value)', and the number of DEGs is indicated as 'count'. GO, Gene Ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes; FC, fibrocavitary

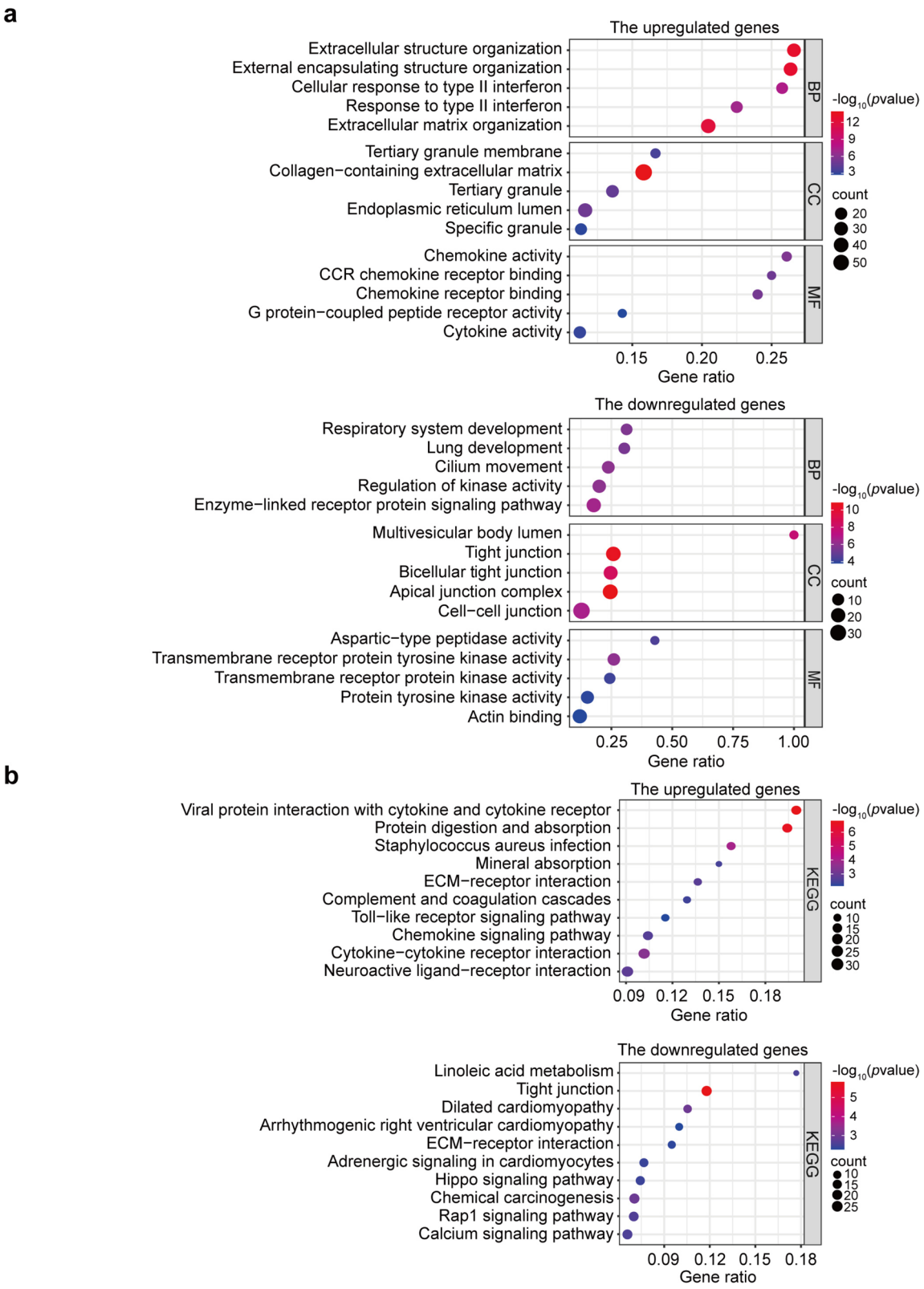


Fig. 4 (See legend on previous page.)

downregulated genes were associated with patched binding and estrogen metabolism (Fig. 2a and Table S1).

Consistent with the GO analyses, KEGG pathway analysis revealed enrichment of upregulated DEGs in primary immunodeficiency, type I diabetes mellitus, B cell receptor signaling, and IgA production. Downregulated DEGs were enriched in cardiomyopathies and AGE-RAGE signaling pathway in diabetic complications (Fig. 2b and Table S2).

GO enrichment and KEGG pathway for NB form-specific DEGs of MAC-PD patients

Next, NB-specific DEG analysis revealed that upregulated DEGs were enriched in GOBP terms related to gastrointestinal epithelium maintenance, epithelial structure integrity, and antimicrobial humoral responses. In contrast, downregulated DEGs were associated with vascular endothelial growth factor (VEGF) signaling and oxygen transport. GOCC highlighted upregulated genes in cation-transporting ATPase complexes and Golgi lumen, whereas downregulated genes were enriched in platelet alpha granules and potassium channel complexes. GOMF analysis showed enrichment of upregulated genes in protease inhibitor activities and SH2 domain binding, while downregulated genes were associated with VEGF receptor and potassium channel regulation (Fig. 3a and Table S3).

KEGG pathways for NB-specific DEGs prominently featured IL-17 signaling, nicotine addiction, cytokine interactions, and pathogen-related pathways. Downregulated DEGs were associated with extracellular matrix (ECM)-receptor interactions, cortisol synthesis, and hematopoietic lineage pathways (Fig. 3b and Table S4).

GO enrichment and KEGG pathway for FC form-specific DEGs of MAC-PD patients

Then, analysis of FC-specific DEGs demonstrated that upregulated genes were enriched in GOBP terms related to extracellular structure organization, encapsulating structure organization, and response to type II interferon. Downregulated DEGs were mainly associated with respiratory and lung development, as well as cilium movement. In GOCC analysis, upregulated genes were enriched in tertiary granule membranes and collagen-containing extracellular matrices, whereas downregulated genes were associated with cell–cell junctions, including bicellular tight junction and apical junction. GOMF analysis revealed upregulated DEGs related to chemokine activities and receptor interactions, with downregulated DEGs involved in transmembrane receptor kinase activities (Fig. 4a and Table S5).

KEGG pathway analysis identified upregulated DEGs associated with receptor interactions (e.g., viral protein-cytokine receptor, ECM-receptor, Toll-like receptor, and cytokine-cytokine receptor) and protein digestion and absorption. Downregulated DEGs were enriched in linoleic acid metabolism, tight junction, and cardiomyopathy pathways (Fig. 4b and Table S6).

Modules and key hub genes for shared DEGs between NB and FC forms of MAC-PD patients

To identify critical gene networks, DEGs were visualized using STRING to construct a PPI network. This analysis yield 299 nodes and 606 edges (interaction score > 0.4). Gene modules were identified using the MCODE plugin (degree cutoff = 2, node score cutoff = 0.2, k-core = 2, max depth = 100, score ≥ 4, nodes ≥ 6). Module 1, the most significant, consisted of 19 nodes and 147 edges (score = 16.333; Fig. 5a). A heatmap indicated that Module 1 predominantly included upregulated genes. KEGG pathway analysis of Module 1 revealed significant enrichment in seven pathways (Fig. 5b, Table S7). CytoHubba analysis, employing three commonly used algorithms (Degree, MNC, and MCC), identified eight consistent hub genes (CD19, CD79A, MS4A1, CD79B, CD27, BLK, FCRLA, and FCRL5), primarily involved in B cell receptor signaling and primary immunodeficiency pathways, and significantly upregulated in affected tissues of both groups (Fig. S3).

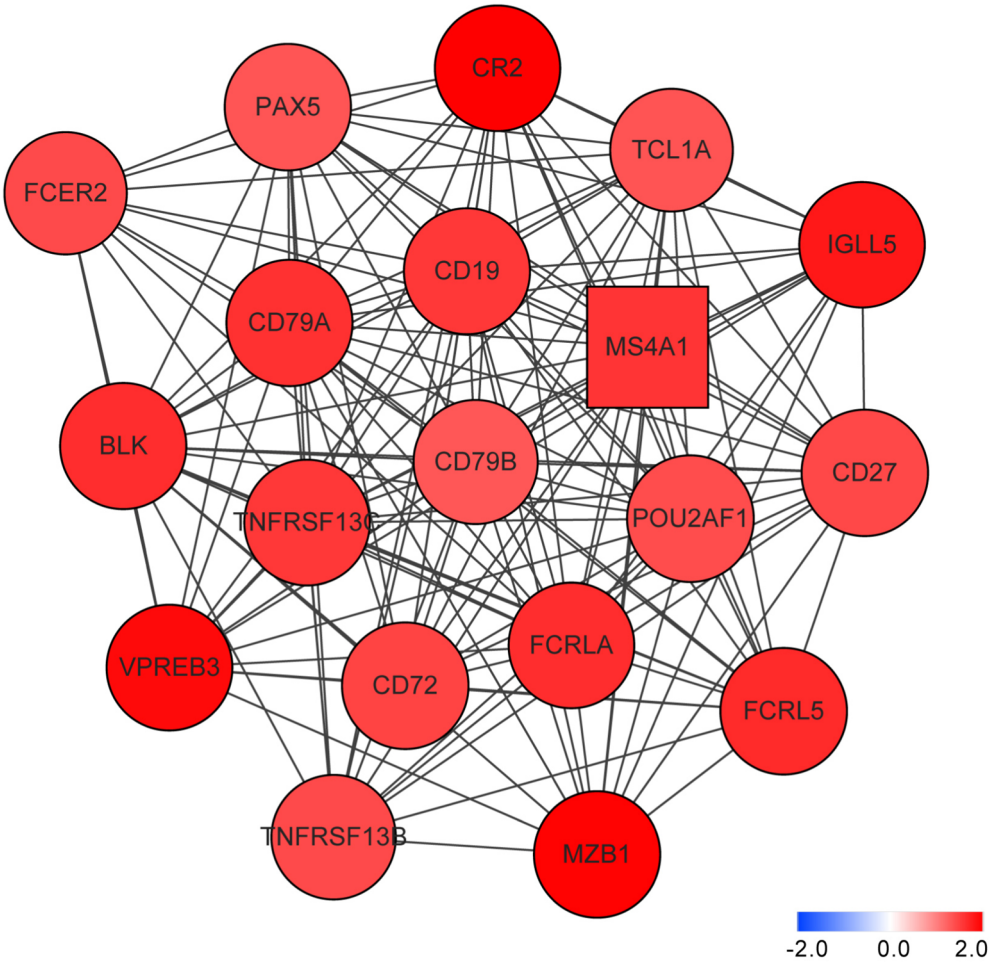
Modules and key hub genes for NB form-specific DEGs of MAC-PD patients

NB-specific DEGs generated a PPI network comprising 203 nodes and 372 edges. MCODE analysis identified four significant gene modules (Fig. 6a–d). Module 1 (nodes = 8, edges = 28, score = 8), Module 2 (nodes = 10, edges = 28, score = 6.222), Module 3 (nodes = 15, edges = 38, score = 5.429), and Module 4 (nodes = 8,

(See figure on next page.)

Fig. 5 MCODE analysis of 351 common DEGs in NB and FC forms. **a** Module 1 (nodes = 19, edges = 147, score 16.333). Module with node ≥ 6 and MCODE score ≥ 10 were selected (degree cutoff = 2, node score cutoff = 0.2, k-core = 2, max depth = 100). The seed protein is shown as a squared node and is responsible for forming the clusters. The node colors indicates log2(fold change) from red to blue; node size reflects *p*-value. **b** KEGG pathway analysis of module genes were performed using EnrichR and significant terms are listed with genes. MCODE, Module of molecular complex detection; PPI, protein–protein interaction; Common, common DEGs from NB and FC form patient samples; KEGG, Kyoto Encyclopedia of Genes and Genomes

a



b

KEGG Pathway	ID	Gene
B cell receptor signaling pathway	hsa04662	CD79B;CD79A;CR2;CD72;CD19
Primary immunodeficiency	hsa05340	CD79A;TNFRSF13B;CD19;TNFRSF13C
Hematopoietic cell lineage	hsa04640	FCER2;CR2;CD19;MS4A1
Epstein-Barr virus infection	hsa05169	FCER2;CR2;CD19
Intestinal immune network for IgA production	hsa04672	TNFRSF13B;TNFRSF13C
Cytokine-cytokine receptor interaction	hsa04061	TNFRSF13B;CD27;TNFRSF13C
PI3K-Akt signaling pathway	hsa04151	TCL1A;CD19

Fig. 5 (See legend on previous page.)

edges = 16, score = 4.571) predominantly included upregulated genes. KEGG analysis showed enrichment in multiple pathways, notably the IL-17 signaling pathway consistently enriched across all modules (Fig. 6E, Table S8). CytoHubba analysis highlighted six hub genes (IFNG, CXCR3, CXCL9, CXCR5, MUC5B, and MUC5AC), significantly upregulated in the NB form, and linked to antimicrobial immune responses, chemokine activities, and the IL-17 signaling pathway (Fig. S4).

Modules and key hub genes for FC form-specific DEGs of MAC-PD patients

FC-specific DEGs resulted in a PPI network of 1606 nodes and 10,317 edges. MCODE analysis revealed 11 significant gene modules, with the top four selected based on module scores (Fig. 7a–d). Module 1 (nodes = 47, edges = 1055, score of 45.87), Module 2 (nodes = 22, edges = 230, score = 21.9), Module 3 (nodes = 37, edges = 264, score = 14.667), and Module 4 (nodes = 74, edges = 416, score = 11.397) showed extensive KEGG pathway enrichment, particularly in ECM-receptor interaction and cancer pathway (Fig. 7E, Table S9). CytoHubba analysis identified ten hub genes (GAPDH, EGFR, FN1, CDH1, ERBB2, COL1A1, FGF2, MMP2, ITGAM, and CXCR4), consistently detected by at least two algorithms (Fig. S5). Among these, FN1, COL1A1, and MMP2 were specifically upregulated in FC form, while CDH1, ERBB2, and FGF2 were downregulated. EGFR and CXCR4 were upregulated in both forms (Fig. S5).

Discussion

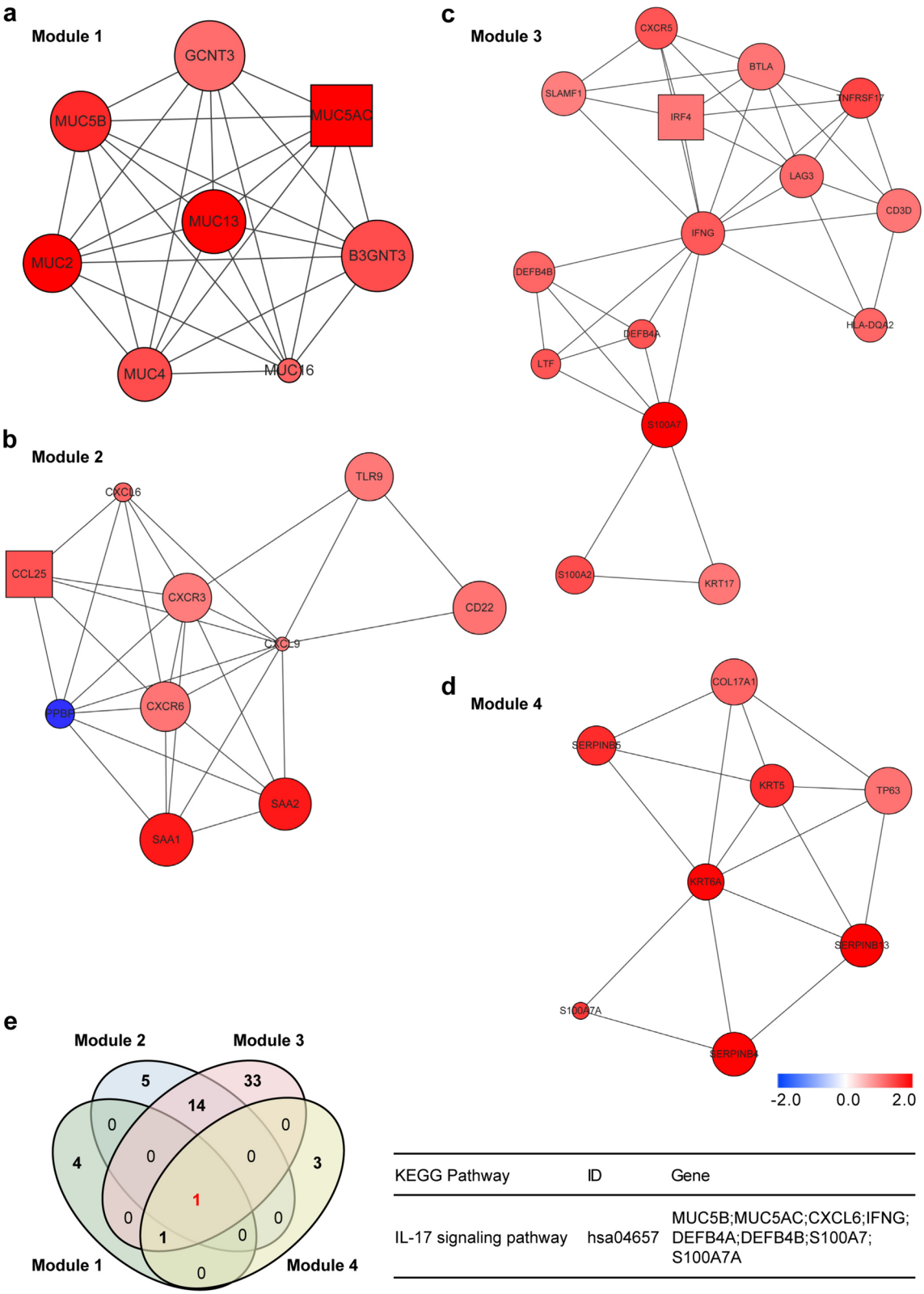
This study explored the distinct radiological phenotypes of MAC pulmonary infection, specifically the NB and FC forms, aiming to elucidate the critical genes underpinning their unique and shared molecular characteristics. We conducted a comprehensive transcriptome analysis to identify functional pathways and candidate genes within lung tissues of MAC-PD patients categorized by radiological phenotypes, marking the first such comparative investigation. Key findings from this study are as follows: (i) Genes related to B cell proliferation and activation were prominently expressed in both NB and FC forms. (ii) Genes implicated in epithelial cell functions and humoral antimicrobial immune responses were exclusively enriched in the NB form. (iii) Genes related to extracellular structure and cytokine-cytokine receptor interaction were distinctly expressed in the FC form.

Our comparative analysis of candidate genes identified in previous studies within our cohort reveals intriguing insights. Initially, the common characteristics between NB and FC forms of surgical sections, when normalized with unaffected sections, show that DEGs are predominantly expressed in the B cell receptor signaling pathway and primary immunodeficiency. This is particularly evident in genes associated with autosomal recessive agammaglobulinemia (CD79A and CD79B) and common variable immunodeficiency (CD19, TNFRSF13B, and TNFRSF13C) (Fig. 2, 5, and Fig. S4). These findings align with previous research suggesting that primary immunodeficiency can predispose individuals to NTM disease [24–28]. Furthermore, B cells play a crucial role in autoimmune diseases, notably rheumatoid arthritis (RA) [29, 30]. In line with this, our study also shows the enrichment of RA-related DEGs (Fig. 2b), indicating a potential connection between RA and the progression and mortality of NTM disease [31, 32]. Despite these correlations, the key factors underlying this relationship remain elusive, suggesting that our DEGs could serve as valuable biomarkers for further investigation.

In preceding investigations, the epithelium emerged as pivotal for defending the respiratory tract and influencing susceptibility to NTM-PD, notably through the intricate workings of cilia [33–35]. Moreover, Matsuyama et al. (2018) unveiled heightened expression of downstream molecules within the IL-17 signaling pathway in respiratory epithelium infected with MAC, even in the absence of direct IL-17 regulation [33]. Echoing these revelations, our findings delineate a distinctive surge in gene expression associated with IL-17 signaling immune responses and epithelial structure within the lung tissues of the NB form (MUC5B, MUC5AC, and IFNG) despite the absence of significant enrichment in IL-17 genes. Of particular note, genes implicated in antimicrobial responses, encompassing autoimmune pathology, neutrophil recruitment, and defense against extracellular pathogens, such as IFNG, CXCR3, CXCL9, and CXCR5, exhibit enrichment (Fig. 3, 6, and Fig. S4). Thus, our transcriptomic data substantiates the inflammatory role of the respiratory epithelium. Therefore, the functional significance of these findings needs further investigation.

(See figure on next page.)

Fig. 6 MCODE analysis of 241 NB form-specific DEGs. **a** Module 1 (nodes = 8, edges = 28, score = 8), **b** Module 2 (nodes = 10, edges = 28, score = 6.222), **c** Module 3 (nodes = 15, edges = 38, score = 5.429), **d** Module 4 (nodes = 8, edges = 16, score = 4.571). Modules were identified using the same criteria as in Fig. 5. **e** KEGG pathways analysis of module genes were performed and a Venn diagram shows a common pathway term shared across the modules, with associated gene listed. MCODE, Module of molecular complex detection; PPI, protein–protein interaction; NB, nodular bronchiectatic; KEGG, Kyoto Encyclopedia of Genes and Genomes



KEGG Pathway	ID	Gene
IL-17 signaling pathway	hsa04657	MUC5B;MUC5AC;CXCL6;IFNG; DEFB4A;DEFB4B;S100A7; S100A7A

Fig. 6 (See legend on previous page.)

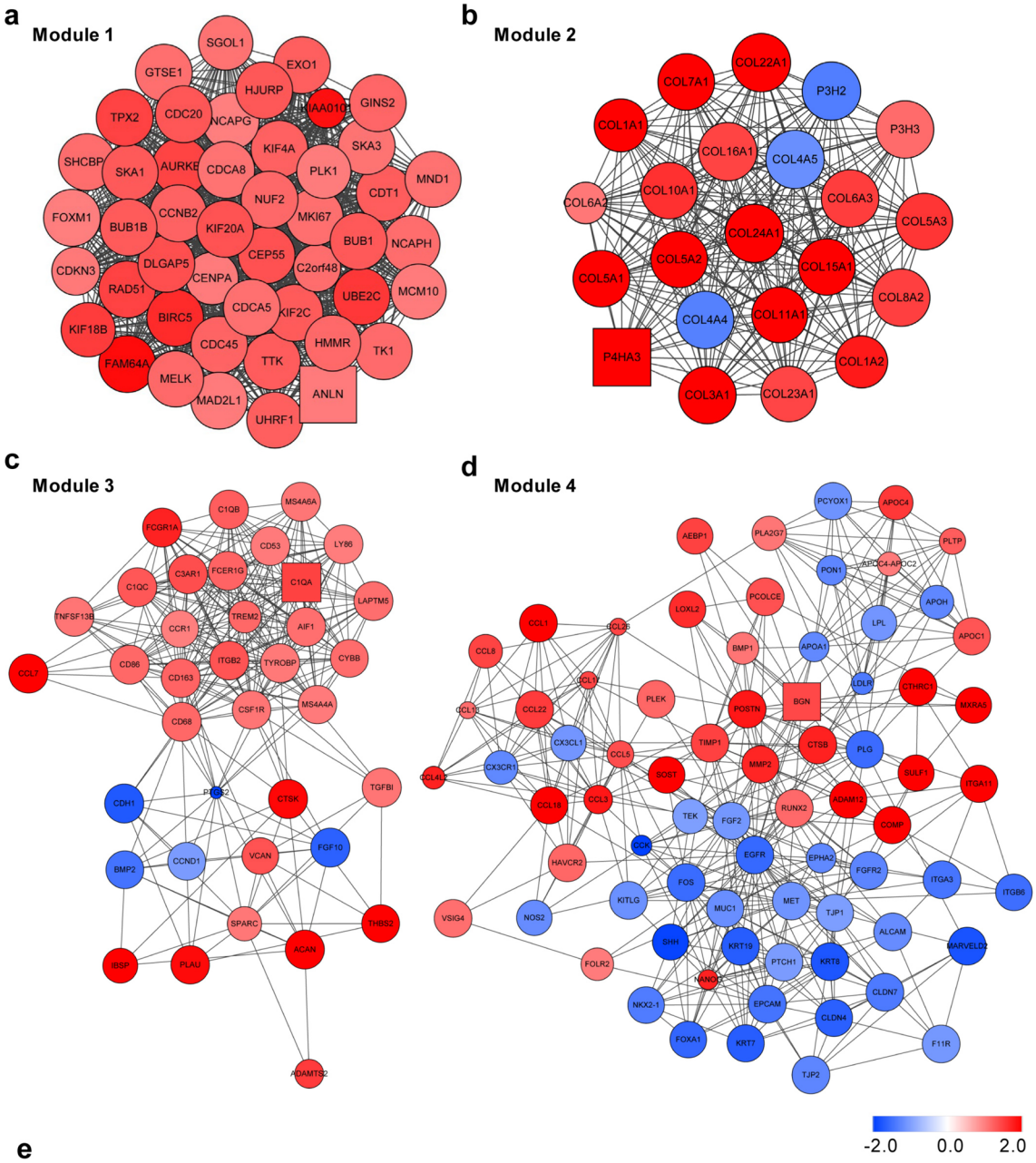
Due to the destruction of lung tissue in MAC-PD patients with cavitation, it is unsurprising that genes differentially expressed in FC-form patients are highly enriched in those related to the ECM and membrane-tight junctions. Collagen, a well-known and abundant component of the ECM in the large airways, blood vessels, and alveolar interstitium of the lungs [36], plays a significant role here. Notably, the genes associated with collagen and tight junctions are predominantly enriched, as illustrated in Fig. 3, 6, and Supplemental Data Fig. S6. Lung cavities frequently occur in tuberculosis and lung cancer, typically involving granuloma formation, ECM dissolution, and subsequent fibrosis [37, 38]. This sequence can be modified by excessive collagen deposition from fibroblasts and other cells [39]. Recent studies suggest that chronic inflammation may contribute to cancer development, linking mycobacteria and cancer [39–44]. Given the similarities between the FC form of MAC-PD, tuberculosis, and lung cancer [45], lung remodeling in MAC-PD is likely to enrich genes related to ECM-receptor interactions and cancer-related pathways.

Our study has several acknowledged limitations. First, the analysis, utilizing a non-randomized retrospective design, was conducted at a sole referral center. Consequently, our investigation is confined to the lung tissues of a specific group of surgical patients. Second, we did not genotype isolates from patients who did not achieve negative culture conversion after surgery, making it challenging to differentiate between a relapse of residual bacteria and reinfection. Third, since our study focused on patients with MAC infection, claiming that unaffected areas are entirely free of bacteria poses a challenge. Given the extended treatment duration of over a year for these patients, it is difficult to exclude the possibility that regions considered unaffected may be areas where bacteria have been eradicated or inflammation has subsided due to treatment. Forth, as these samples are lung lobe tissues from surgically treated patients, there may be differences in target genes or pathways compared to patients who achieved mycobacterial eradication through natural healing or drug therapy. Fifth, this study relied on bulk transcriptomic profiling of lung tissues rather than single-cell or spatial transcriptomic approaches. Therefore, the gene expression changes cannot be attributed to specific cell types and should be interpreted at the level of the lung microenvironment rather than cell type-specific transcriptional regulation. Finally, because the analyses were performed on resected lung tissues representing advanced and treatment-refractory disease, the identified transcriptional signatures may reflect disease-associated outcomes rather than causal relationships. Accordingly, functional validation of the identified pathways and candidate genes was not performed in this study, and further investigations integrating higher-resolution transcriptomic approaches and targeted functional assays will be required to establish causal relationships and to validate the biological relevance of the identified molecular signatures.

Collectively, these results offer initial insights with significant features for evaluating MAC-PD radiographically different types (NB and FC forms) based on the dissection of affected and unaffected sections within lung tissue. The findings from this exploratory cohort hold significance for future diagnostic biomarker

(See figure on next page.)

Fig. 7 MCODE analysis of 2049 FC form-specific DEGs. **a** Module 1 (nodes = 47, edges = 1055, score = 45.87), **b** module 2 (nodes = 22, edges = 230, score = 21.9), **c** module 3 (nodes = 37, edges = 264, score = 14.667), **d** module 4 (nodes = 74, edges = 416, score = 11.397). Modules were identified using the same criteria as in Fig. 5. **e** KEGG pathways analysis of module genes were performed and a Venn diagram highlights two common KEGG terms, and the related genes are listed. MCODE, Module of molecular complex detection; PPI, protein–protein interaction; FC, fibrocavitary; KEGG, Kyoto Encyclopedia of Genes and Genomes



KEGG Pathway	ID	Gene
ECM-receptor interaction	hsa04512	HMMR;COL1A1;COL1A2;COL4A4;COL6A2;COL4A5;COL6A3;IBSP;THBS2;COMP;ITGA3;ITGA11;ITGB6
Pathways in cancer	hsa05200	RAD51;BIRC5;COL4A4;COL4A5;CSF1R;BMP2;CCND1;CDH1;PTGS2;FGF10;SHH;KITLG;NOS2;ITGA3;MMP2;PTCH1;FOS;FGF2;MET;EGFR;FGFR2

Fig. 7 (See legend on previous page.)

discovery related to radiological types and can serve as a validation dataset for other research groups.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s40001-026-03987-9.

Additional file 1.

Author contributions

Byung Woo Jhun and Sung Jae Shin conceived, designed, and supervised the work; Ju Mi Lee and Su-Young Kim analyzed the data and interpreted the results. Su-Young Kim, Seong Mi Moon, and Byung Woo Jhun provided lung tissue samples from MAC-PD patients and analyzed data. Ju Mi Lee and Su-Young Kim drafted and prepared the manuscript while Byung Woo Jhun and Sung Jae Shin were involved in the editing of the manuscript. All authors read and approved the final version.

Funding

This work was supported by the National Research Foundation of Korea (NRF) grant, funded by the Korea government (MSIT) (RS-2024-00405542), by a grant of the Korea Health Technology R&D Project through the Korea HealthIndustry Development Institute (KHIDI), funded by the Ministry of Health & Welfare (RS-2025-02193034), and by the National Institute of Health (NIH) research project (2024-ER2107-02), South Korea.

Data availability

All datasets presented in the study are included in the article/Supplementary Material. Raw data and processed data of RNA sequencing have been deposited in the GEO with accession number GSE270278.

Declarations

Ethics approval and consent to participate

The study was approved by the Institutional Review Board of the Samsung Medical Center (IRB No. 2012-05-001). Written informed consent was obtained from all patients participating in the study.

Competing interest

The authors declare no competing interests.

Author details

[1]Department of Microbiology, Institute for Immunology and Immunological Disease, Graduate School of Medical Science, Yonsei University College of Medicine, Brain Korea 21 Project, Seoul, South Korea. [2]Division of Pulmonary and Critical Care Medicine, Department of Medicine, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, South Korea. [3]Division of Pulmonary, Allergy and Critical Care Medicine, Department of Internal Medicine, Chung-Ang University Gwangmyeong Hospital, Gwangmyeong, South Korea.

Received: 17 July 2025 Accepted: 25 January 2026
Published online: 31 January 2026

References

1. Jeon D. Infection source and epidemiology of nontuberculous mycobacterial lung disease. Tuberc Respir Dis (Seoul). 2019;82(2):94–101.
2. Jhun BW, Moon SM, Jeon K, Kwon OJ, Yoo H, Carriere KC, et al. Prognostic factors associated with long-term mortality in 1445 patients with nontuberculous mycobacterial pulmonary disease: a 15-year follow-up study. Eur Respir J. 2020;55(1).
3. Lee H, Myung W, Koh WJ, Moon SM, Jhun BW. Epidemiology of nontuberculous mycobacterial infection, South Korea, 2007-2016. Emerg Infect Dis. 2019;25(3):569–72.
4. Prevots DR, Marshall JE, Wagner D, Morimoto K. Global epidemiology of nontuberculous mycobacterial pulmonary disease: a review. Clin Chest Med. 2023;44(4):675–721.
5. Lee YM, Kim MJ, Kim YJ. Increasing trend of nontuberculous mycobacteria isolation in a referral clinical laboratory in South Korea. Medicina Kaunas. 2021. https://doi.org/10.3390/medicina57070720.
6. Koh WJ, Moon SM, Kim SY, Woo MA, Kim S, Jhun BW, et al. Outcomes of *Mycobacterium avium* complex lung disease based on clinical phenotype. Eur Respir J. 2017. https://doi.org/10.1183/13993003.02503-2016.
7. Lee JH, Park YE, Chong YP, Lee HJ, Shim TS, Jo KW. Radiologic subtypes and treatment outcome of unclassifiable type *Mycobacterium avium* complex pulmonary disease. J Korean Med Sci. 2023;38(3):e16.
8. Kartalija M, Ovrutsky AR, Bryan CL, Pott GB, Fantuzzi G, Thomas J, et al. Patients with nontuberculous mycobacterial lung disease exhibit unique body and immune phenotypes. Am J Respir Crit Care Med. 2013;187(2):197–205.
9. Kang EY. Radiologic diagnosis of nontuberculous mycobacterial pulmonary disease. J Korean Soc Radiol. 2021;82(4):838–50.
10. Kitada S, Uenami T, Yoshimura K, Tateishi Y, Miki K, Miki M, et al. Long-term radiographic outcome of nodular bronchiectatic *Mycobacterium avium* complex pulmonary disease. Int J Tuberc Lung Dis. 2012;16(5):660–4.
11. Zo S, Kim H, Kwon OJ, Jhun BW. Antibiotic maintenance and redevelopment of nontuberculous mycobacteria pulmonary disease after treatment of *Mycobacterium avium* complex pulmonary disease. Microbiol Spectr. 2022;10(4):e0108822.
12. Lee BY, Kim S, Hong Y, Lee SD, Kim WS, Kim DS, et al. Risk factors for recurrence after successful treatment of *Mycobacterium avium* complex lung disease. Antimicrob Agents Chemother. 2015;59(6):2972–7.
13. Okumura M, Iwai K, Ogata H, Ueyama M, Kubota M, Aoki M, et al. Clinical factors on cavitary and nodular bronchiectatic types in pulmonary *Mycobacterium avium* complex disease. Intern Med. 2008;47(16):1465–72.
14. Fleshner M, Olivier KN, Shaw PA, Adjemian J, Strollo S, Claypool RJ, et al. Mortality among patients with pulmonary non-tuberculous mycobacteria disease. Int J Tuberc Lung Dis. 2016;20(5):582–7.
15. Shu CC, Lee CH, Hsu CL, Wang JT, Wang JY, Yu CJ, et al. Clinical characteristics and prognosis of nontuberculous mycobacterial lung disease with different radiographic patterns. Lung. 2011;189(6):467–74.
16. Gochi M, Takayanagi N, Kanauchi T, Ishiguro T, Yanagisawa T, Sugita Y. Retrospective study of the predictors of mortality and radiographic deterioration in 782 patients with nodular/bronchiectatic *Mycobacterium avium* complex lung disease. BMJ Open. 2015;5(8):e008058.
17. Fukushima K, Kitada S, Abe Y, Yamamoto Y, Matsuki T, Kagawa H, et al. Long-term treatment outcome of progressive *Mycobacterium avium* complex pulmonary disease. J Clin Med. 2020;9(5):1315.
18. Kim BG, Kang N, Kim SY, Kim DH, Kim H, Kwon OJ, et al. The lung microbiota in nontuberculous mycobacterial pulmonary disease. PLoS ONE. 2023;18(5):e0285143.
19. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. Bioinformatics. 2009;25(9):1105–11.
20. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26(6):841–2.
21. Roberts A, Trapnell C, Donaghey J, Rinn JL, Pachter L. Improving RNA-Seq expression estimates by correcting for fragment bias. Genome Biol. 2011;12(3):R22.
22. Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci U S A. 1998;95(25):14863–8.
23. Tang D, Chen M, Huang X, Zhang G, Zeng L, Zhang G, et al. SRplot: a free online platform for data visualization and graphing. PLoS ONE. 2023;18(11):e0294236.
24. Reichenbach J, Rosenzweig S, Doffinger R, Dupuis S, Holland SM, Casanova JL. Mycobacterial diseases in primary immunodeficiencies. Curr Opin Allergy Clin Immunol. 2001;1(6):503–11.
25. Lee WI, Huang JL, Yeh KW, Jaing TH, Lin TY, Huang YC, et al. Immune defects in active mycobacterial diseases in patients with primary immunodeficiency diseases (PIDs). J Formos Med Assoc. 2011;110(12):750–8.
26. Sherkat R. Enigmas of primary immunodeficiency and mycobacterial infection in our territory. Allergy Asthma Clin Immunol. 2014;10(1):A30.
27. Galkina E, Kondratenko I, Bologov A, editors. Mycobacterial infections in primary immunodeficiency patients. Immune-Mediated Diseases; 2007. New York, NY: Springer New York.

28. van de Vosse E. Primary immunodeficiency leading to mycobacterial disease. Int J Mycobacteriol. 2015;4:63.
29. Mahendra A, Yang X, Abnouf S, Adolacion JRT, Park D, Soomro S, et al. Beyond autoantibodies: biologic roles of human autoreactive B cells in rheumatoid arthritis revealed by RNA-sequencing. Arthritis Rheumatol. 2019;71(4):529–41.
30. Silverman GJ, Carson DA. Roles of B cells in rheumatoid arthritis. Arthritis Res Ther. 2003;5 Suppl 4(Suppl 4):S1–6.
31. Kwak N, Moon J, Kim JY, Park JW, Yim JJ. Clinical course of nontuberculous mycobacterial pulmonary disease in patients with rheumatoid arthritis. Adv Rheumatol. 2024;64(1):20.
32. Lim DH, Kim YG, Shim TS, Jo KW, Ghang B, Ahn SM, et al. Nontuberculous mycobacterial infection in rheumatoid arthritis patients: a single-center experience in South Korea. Korean J Intern Med. 2017;32(6):1090–7.
33. Matsuyama M, Martins AJ, Shallom S, Kamenyeva O, Kashyap A, Sampaio EP, et al. Transcriptional response of respiratory epithelium to nontuberculous mycobacteria. Am J Respir Cell Mol Biol. 2018;58(2):241–52.
34. Szymanski EP, Leung JM, Fowler CJ, Haney C, Hsu AP, Chen F, et al. Pulmonary nontuberculous mycobacterial infection. A multisystem, multigenic disease. Am J Respir Crit Care Med. 2015;192(5):618–28.
35. Fowler CJ, Olivier KN, Leung JM, Smith CC, Huth AG, Root H, et al. Abnormal nasal nitric oxide production, ciliary beat frequency, and Toll-like receptor response in pulmonary nontuberculous mycobacterial disease epithelium. Am J Respir Crit Care Med. 2013;187(12):1374–81.
36. Davidson JM. Biochemistry and turnover of lung interstitium. Eur Respir J. 1990;3(9):1048–63.
37. Ong CW, Elkington PT, Friedland JS. Tuberculosis, pulmonary cavitation, and matrix metalloproteinases. Am J Respir Crit Care Med. 2014;190(1):9–18.
38. Chaudhuri MR. Primary pulmonary cavitating carcinomas. Thorax. 1973;28(3):354–66.
39. Dheda K, Booth H, Huggett JF, Johnson MA, Zumla A, Rook GA. Lung remodeling in pulmonary tuberculosis. J Infect Dis. 2005;192(7):1201–9.
40. Fol M, Kozinski P, Kulesza J, Bialecki P, Druszczynska M. Dual nature of relationship between mycobacteria and cancer. Int J Mol Sci. 2021;22(15).
41. Lande L, Peterson DD, Gogoi R, Daum G, Stampler K, Kwait R, et al. Association between pulmonary *Mycobacterium avium* complex infection and lung cancer. J Thorac Oncol. 2012;7(9):1345–51.
42. Brenner DR, McLaughlin JR, Hung RJ. Previous lung diseases and lung cancer risk: a systematic review and meta-analysis. PLoS ONE. 2011;6(3):e17479.
43. Yu YH, Liao CC, Hsu WH, Chen HJ, Liao WC, Muo CH, et al. Increased lung cancer risk among patients with pulmonary tuberculosis: a population cohort study. J Thorac Oncol. 2011;6(1):32–7.
44. Liang HY, Li XL, Yu XS, Guan P, Yin ZH, He QC, et al. Facts and fiction of the relationship between preexisting tuberculosis and lung cancer risk: a systematic review. Int J Cancer. 2009;125(12):2936–44.
45. Musaddaq B, Cleverley JR. Diagnosis of non-tuberculous mycobacterial pulmonary disease (NTM-PD): modern challenges. Br J Radiol. 2020;93(1106):20190768.

Publisher's Note