



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

Characterization and evaluation
of the gene expression profile
for nontuberculous mycobacterial
pulmonary disease

Youngmok Park

Department of Medicine

The Graduate School, Yonsei University

Characterization and evaluation
of the gene expression profile
for nontuberculous mycobacterial
pulmonary disease

Directed by Professor Young Ae Kang

The Doctoral Dissertation
submitted to the Department of Medicine,
the Graduate School of Yonsei University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Medical Science

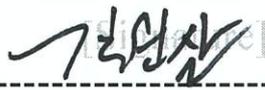
Youngmok Park

December 2023

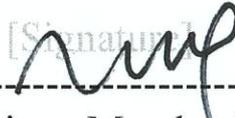
This certifies that the Doctoral Dissertation of
Youngmok Park is approved.



Thesis Supervisor: Young Ae Kang



Thesis Committee Member#1: Young Sam Kim



Thesis Committee Member#2: Hye-Jeong Lee



Thesis Committee Member#3: Heon Yung Gee



Thesis Committee Member#4: Chang Ki Kim

The Graduate School
Yonsei University

December 2023

ACKNOWLEDGEMENTS

I would like to express my utmost gratitude to my esteemed supervisor, Prof. Young Ae Kang, whose warm and supportive demeanor has been invaluable throughout my pursuit of a Doctorate degree.

I am deeply grateful to the esteemed members of my committee, Prof. Young Sam Kim, Prof. Hye-Jeong Lee, Prof. Heon Yung Gee, and Dr. Chang Ki Kim, for their unwavering guidance, encouragement, invaluable advice, and meticulous revisions throughout the entire process of my dissertation.

I am immensely grateful to Ji Won Hong and Eunsol Ahn for their invaluable assistance in the analysis. Without their help, this task would have been insurmountable.

I would like to express my heartfelt gratitude to my beloved father and mother, who have always stood by my side, as well as to my father-in-law and mother-in-law, who have been my sincere supporters.

Lastly, I would like to express my love and appreciation to my wife, Kyung En Chung, who has never hesitated to offer her unwavering support and make countless sacrifices, and my daughter, Haram Park, who has been an endless source of inspiration and motivation.

I shall give glory to God and strive to become a competent doctor, bringing pride to my loved ones.

Youngmok Park

<TABLE OF CONTENTS>

ABSTRACT.....	iv
I. INTRODUCTION.....	1
II. MATERIALS AND METHODS.....	2
1. Study population and blood sample collection	2
2. Clinical data of study participants.....	2
3. RNA sequencing	3
4. Mapping reads and expression calculation	4
5. Identification of DEGs and functional enrichment analysis	4
6. Immune cell deconvolution.....	5
7. Statistical analysis	5
III. RESULTS	6
1. Baseline characteristics	6
2. Principal component analysis.....	7
3. DEG analysis.....	8
4. Protein-protein connection	16
5. Ontology and pathway enrichment analysis	19
6. Immune cell deconvolution.....	24
7. Analysis of previously reported genes related to NTM-PD.....	25
8. <i>PARK2</i> gene as a diagnostic biomarker	27
IV. DISCUSSION.....	29
V. CONCLUSION.....	33
REFERENCES	34
ABSTRACT(IN KOREAN).....	40

LIST OF FIGURES

Figure 1. Principal component analysis depicting the unsupervised clustering of the case and control groups	9
Figure 2. Principal component analysis using 65 DEGs between the case and control groups	14
Figure 3. Heatmap generated using 65 DEGs between the case and control groups	15
Figure 4. Principal component analysis using 65 DEGs between the case and control groups	16
Figure 5. Principal component analysis using 65 DEGs between the case and control groups	17
Figure 6. Co-expression protein network construction	18
Figure 7. Modules of protein-protein interaction networks	19
Figure 8. DAVID functional GO analysis	21
Figure 9. Enrichment analysis	22
Figure10. Proportions of 22 types of immune cells between the case and control groups	25
Figure 11. Principal component analysis using previously reported gene sets between the case and control groups	26
Figure 12. The <i>PARK2</i> gene for diagnosing NTM-PD	28
Figure 13. Expression levels of the <i>PARK2</i> gene between	

pre-treatment and post-treatment samples of NTM-PD·····29

LIST OF TABLES

Table 1. Baseline characteristics of the study population ·····	7
Table 2. Upregulated genes between the case and control groups ···	11
Table 3. Downregulated genes between the case and control groups·	12
Table 4. Modules of protein-protein interaction networks ·····	20
Table 5. GO enrichment analysis conducted on the DEGs between the case and control groups ·····	23
Table 6. Pathway enrichment analysis conducted on the DEGs between the case and control groups ·····	24
Table 7. Analysis of genes previously reported to be associated with NTM-PD in the current study samples ·····	27

ABSTRACT

**Characterization and evaluation of the gene expression profile for
nontuberculous mycobacterial pulmonary disease**

Youngmok Park

*Department of Medicine
The Graduate School, Yonsei University*

(Directed by Professor Young Ae Kang)

Background: Nontuberculous mycobacteria (NTM) are environmental organisms that primarily cause pulmonary disease (PD). The incidence of NTM-PD is increasing globally; however, the current diagnosis and treatment methods are far from optimal. We performed ribonucleic acid (RNA) sequencing to explore gene expression profiles and identify potential biomarkers in individuals with NTM-PD.

Methods: We collected peripheral blood samples from individuals with NTM-PD and healthy controls at a tertiary referral center in South Korea. Additional samples were obtained from the case group after the completion of NTM treatment. RNA sequencing was performed on the samples, and differentially expressed genes were identified and subjected to functional enrichment analysis. We employed immune cell deconvolution techniques to quantitatively analyze the cellular composition of 22 types of immune cells.

Results: We enrolled 26 participants with NTM-PD (median age, 58.0 years; 84.6% female; *Mycobacterium avium* complex, 76.9%) and 22 healthy controls (median age, 58.5 years; 90.9% female). We identified 21 upregulated and 44 downregulated differentially expressed genes in the case group compared to the control group. Gene

ontology and pathway enrichment analyses showed that the upregulated genes were related to autophagy in individuals with NTM-PD. According to the immune cell deconvolution analysis, neutrophils were the most common immune cells in both groups, accounting for 7.0% (interquartile range [IQR], 5.4%–9.2%) in the NTM-PD group and 6.6% (IQR, 5.5%–8.7%; $P = 0.644$) in the control group. The proportions of other immune cell types were also similar between the two groups ($P > 0.05$ for all).

The *PARK2* gene, which is linked to the ubiquitination pathway, was downregulated in the study group (fold change, -1.314 , $P = 0.047$). The expression levels of the *PARK2* gene remained unaltered in the five samples collected after microbiologic cure, suggesting that the *PARK2* gene is associated with host susceptibility rather than the outcomes of infection or inflammation. The expression levels of the *PARK2* gene showed promise as a potential diagnostic biomarker with an area under the receiver operating characteristic curve of 0.813 (95% confidence interval, 0.694–0.932).

Conclusion: This study identified genetic signatures associated with NTM-PD infection in a cohort of Korean patients. The downregulation of the *PARK2* gene in individuals with NTM-PD could provide an opportunity for the development of biomarkers, making it a potential candidate gene for diagnosis.

Key words: nontuberculous mycobacteria, gene expression profiling, ribonucleic acid

Characterization and evaluation of the gene expression profile for nontuberculous mycobacterial pulmonary disease

Youngmok Park

Department of Medicine
The Graduate School, Yonsei University

(Directed by Professor Young Ae Kang)

I. INTRODUCTION

Nontuberculous mycobacteria (NTM) are ubiquitous environmental organisms in natural and drinking water systems, pools and hot tubs, biofilms, and soil.¹ They can affect various tissues and body fluids, mainly causing pulmonary diseases (PD).² The worldwide incidence and prevalence rates of NTM-PD are increasing, affecting both immuno-compromised and immuno-competent patients.^{3,4} Furthermore, the distribution of NTM varies depending on geographic location, primarily due to environmental factors.⁵ Accordingly, the characteristics of patients with NTM-PD differ across regions.

NTM-PD poses numerous challenges for physicians. The diagnosis of NTM-PD is a complex process that requires repeated culture results of certain NTM species, radiographic correlation of chest image, and related symptoms.^{2,6} The clinical courses of NTM-PD are diverse and unpredictable; some cases progress rapidly while others remain stable without treatment or even experience spontaneous remission.⁷ The factors that determine treatment response are not fully understood, and the timing of treatment initiation and choice of regimen are not definitive based on current knowledge.⁸ Consequently, the treatment outcome is often disappointing. The success rate of treating *Mycobacterium avium* complex (MAC) PD ranges from 55% to 65%,⁹ while that of

treating *Mycobacterium abscessus* complex (MABC) PD ranges from 24% to 46%, which is considered incurable.¹⁰

Researchers have investigated biomarkers of NTM-PD to improve diagnostic barriers and monitor treatment response. Potential biomarkers such as Interleukin 17, Carbohydrate antigen 19-9, anti-glycopeptidolipid IgA, and anti-interferon (IFN)- γ autoantibody titer have been identified.^{11,12} However, the available data is limited, and no standardized biomarker has been established.

We hypothesized that there would be differences in blood gene expression profiles between patients with NTM-PD and healthy controls. Therefore, the objective of this study was to explore the genetic signature of NTM-PD and identify potential blood biomarkers through ribonucleic acid (RNA) sequencing.

II. MATERIALS AND METHODS

1. Study population and blood sample collection

We collected peripheral blood samples from participants with NTM-PD and healthy controls between 2015 and 2019. We gathered 2.5 mL of peripheral blood into the PAXgene RNA tube (Becton Dickinson and Co., Franklin Lakes, NJ, USA) and froze at -80°C until analysis. In participants with NTM-PD, blood samples were stored before and at the end of treatment. We excluded individuals with malignancies, end-stage renal disease, and human immunodeficiency virus infection from the study. Healthy controls were defined as individuals without respiratory symptoms, abnormalities in chest X-rays, and a medical history of chronic lung diseases.

The institutional review board at Severance Hospital approved the protocol of this study (NTM-PD cohort: 4-2017-0958, healthy control cohort: 4-2010-0213, 4-2014-1108). All participants provided written informed consent.

2. Clinical data of study participants

The clinical data of the participants was collected from their medical records. The

medical charts were reviewed, taking into account factors such as age, sex, body mass index, and underlying diseases.

Participants with NTM-PD were diagnosed based on the guideline. Upon reviewing the chest computed tomography, radiologic types were classified as nodular bronchiectatic (NB) or fibrocavitary (FC). The severity of the disease was determined by the extent of lung involvement and sputum acid-fast bacilli smear results. The treatment outcomes were defined by the NTM-NET consensus statement.¹³

3. RNA sequencing

RNA was extracted using the provided instructions with the PAXgene Blood RNA Kit (PreAnalytiX, Hombrechtikon, Switzerland). The purity of the RNA was assessed by analyzing 1 μ L of the total RNA extract on a NanoDrop 8000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA). The integrity of the total RNA was evaluated using an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA) with an RNA integrity number value. RNA with an optical density of 260/280 \geq 1.8 and an RNA integrity number \geq 7 were selected for the subsequent experiments.

The total RNA sequencing libraries were prepared according to the manufacturer's instructions (Illumina TruSeq Stranded Total RNA Sample Prep Kit with Ribo-Zero Globin, Illumina, San Diego, CA, USA). The process involved removing ribosomal RNA from 500 ng of total RNA using Ribo-Zero Globin reagent, utilizing biotinylated probes to selectively bind rRNA species. After purification, the rRNA-depleted total RNA was fragmented into small pieces using divalent cations at an elevated temperature. The resulting cleaved RNA fragments were converted into first-strand cDNA using reverse transcriptase and random primers. Subsequently, second-strand cDNA was synthesized using DNA Polymerase I and RNase H. These cDNA fragments were modified with a single 'A' base and ligated with an adapter. The products were purified and enriched through PCR to create the final cDNA library.

The quality of the amplified libraries was confirmed through capillary electrophoresis

using Bioanalyzer (Agilent Technologies). After performing real-time polymerase chain reaction with SYBR Green PCR Master Mix (Applied Biosystems, Foster City, CA, USA), we combined the libraries that were index tagged in equimolar amounts into a pool. Finally, RNA sequencing was performed using a NovaSeq 6000 system (Illumina) following the provided protocols for 2x100 sequencing.

4. Mapping reads and expression calculation

The reads for each sample were mapped to the reference genome of the Human Genome Reference Consortium Human Build 37 (GRCh37, hg19)¹⁴ using Tophat (v2.0.13).¹⁵ The aligned results were added to Cuffdiff (v2.2.1)¹⁶ to calculate the transcript per million value and report differentially expressed genes (DEGs) while applying a false discovery rate (FDR) of 5% significance. For library normalization and dispersion estimation, geometric and pooled methods were used. The geometric method was used ('blind' when each condition had single replicates, or 'pooled' when multiple replicates were available).

5. Identification of DEGs and functional enrichment analysis

Two filtering processes were applied to detect DEGs between the case and control groups. Initially, using the Cuffdiff status code, only genes with an 'OK' status were extracted. The status code indicates whether there are sufficient reads in a specific locus for a reliable calculation of gene expression level. An 'OK' status indicates that the gene expression level was successfully calculated. For the second filtering process, a 2-fold change was calculated, and genes falling within the following ranges were selected.

$$\text{Upregulated: } \log_2[\text{case}] - \log_2[\text{control}] > \log_2(2) = 1$$

$$\text{Downregulated: } \log_2[\text{case}] - \log_2[\text{control}] < \log_2(1/2) = -1$$

Using the STRING (v1.7.0, <https://string-db.org>)¹⁷ database, the protein-protein connections were assessed among the DEGs, and protein interactions were plotted using Cytoscape (v3.8, <https://cytoscape.org>).

We identified human genes with relevant functional gene ontology (GO). For

integrative analysis, we employed both the Database for Annotation, Visualization, and Integrated Discovery (DAVID) functional annotation tool and the ClueGO (v2.5.8)/CluePedia (v1.5.8) plugin of Cytoscape to complementarily identify the DEGs involved in the GO terms and pathways. ClueGO combines GO terms and pathways from Kyoto Encyclopedia of Genes (KEGG), Reactome, and Wiki, providing a structured GO term or pathway network from the DEG dataset.¹⁸ In addition, the CluePedia integrates into the ClueGO network of terms/pathways, linking genes based on in silico and experimental information.¹⁹ To determine significance, we applied a threshold of P values < 0.05 for the study of molecular/biological/cellular function GO and enrichment of pathway analysis for DEGs.

6. Immune cell deconvolution

To analyze the composition of immune cells in our samples, we employed the CIBERSORTx platform, a computational method designed for characterizing the cell composition of complex tissues based on their gene expression profiles.²⁰ The analysis was performed using the LM22 signature matrix, which consists of 22 distinct immune cell subtypes, including B cells, T cells, natural killer cells, macrophages, dendritic cells, and myeloid subsets.²¹ Notably, the analysis was conducted in the absolute mode, which enables a more quantitative interpretation of the results by providing the exact proportions of each cell type.

7. Statistical analysis

The median values of the variables among participants with NTM-PD and the control group were compared using the Mann-Whitney U test. Categorical variables were compared through Fisher's exact test. Wilcoxon signed-rank tests were conducted to compare the median gene expression values between pre- and post-treatment samples.

Receiver operating characteristic (ROC) curves were generated to assess the clinical relevance of the identified markers, and the area under the curve (AUC) was calculated to

determine the optimal cutoff value and discriminatory capacity. The sensitivity, specificity, positive predictive value, and negative predictive value were evaluated based on the optimal cutoff value. Differences with a two-sided *P* value below 0.05 were considered statistically significant.

III. RESULTS

1. Baseline characteristics

A total of 26 participants with NTM-PD and 22 healthy controls were included in the study. Among the participants with NTM-PD, five additional samples were collected after completing the treatment, all of whom achieved microbiological cure. The baseline characteristics of the study participants are presented in Table 1. The study group exhibited lower weight and body mass index compared to the control group. Otherwise, the clinical characteristics were similar between the two groups.

Table 1. Baseline characteristics of the study population

	NTM-PD (N = 26)	Healthy controls (N = 22)	<i>P</i> -value
Age, years	58.0 [48.0–64.0]	58.5 [56.0–60.0]	0.967
Sex, female	22 (84.6)	20 (90.9)	0.827
Height, cm	159.8 [157.0–162.0]	157.0 [155.0–166.0]	0.101
Weight, kg	52.0 [49.6–57.0]	58.0 [54.0–63.0]	0.021
BMI, kg/m ²	20.8 ± 2.5	23.4 ± 3.1	0.002
BMI <18.5 kg/m ²	4 (15.4)	0 (0.0)	0.162
Ever smoker	3 (11.5)	0 (0.0)	0.295
Comorbidity			
Hypertension	4 (15.4)	5 (22.7)	0.781
Diabetes mellitus	1 (3.8)	1 (4.5)	>0.999
Bronchiectasis	24 (92.3)		

COPD	5 (19.2)
History of tuberculosis	7 (26.9)
Previous NTM treatment	3 (11.5)
Causative organism	
<i>M. avium</i> complex	20 (76.9)
<i>M. abscessus</i> complex	4 (15.4)
Others	2 (7.7)
Radiologic findings	
Non-cavitary NB	13 (50.0)
Cavitary NB	10 (38.5)
Fibrocavitary	3 (11.5)
Extent, ≥ 3 lobes	21 (80.8)
Sputum smear positivity	7 (26.9)
Presence of cavity	13 (50.0)

Note: Data are presented with median [interquartile range] or number (percent).

Abbreviations: BMI, body mass index; COPD, chronic obstructive pulmonary disease; NB, nodular bronchiectatic; NTM, nontuberculous mycobacteria; PD, pulmonary disease.

2. Principal component analysis

Figure 1 illustrates unsupervised principal component analysis (PCA) conducted on the case and control groups. PC1 captured 14.51% of the total variance, while PC2 accounted for 10.17%. However, these components were not able to effectively differentiate the two groups.

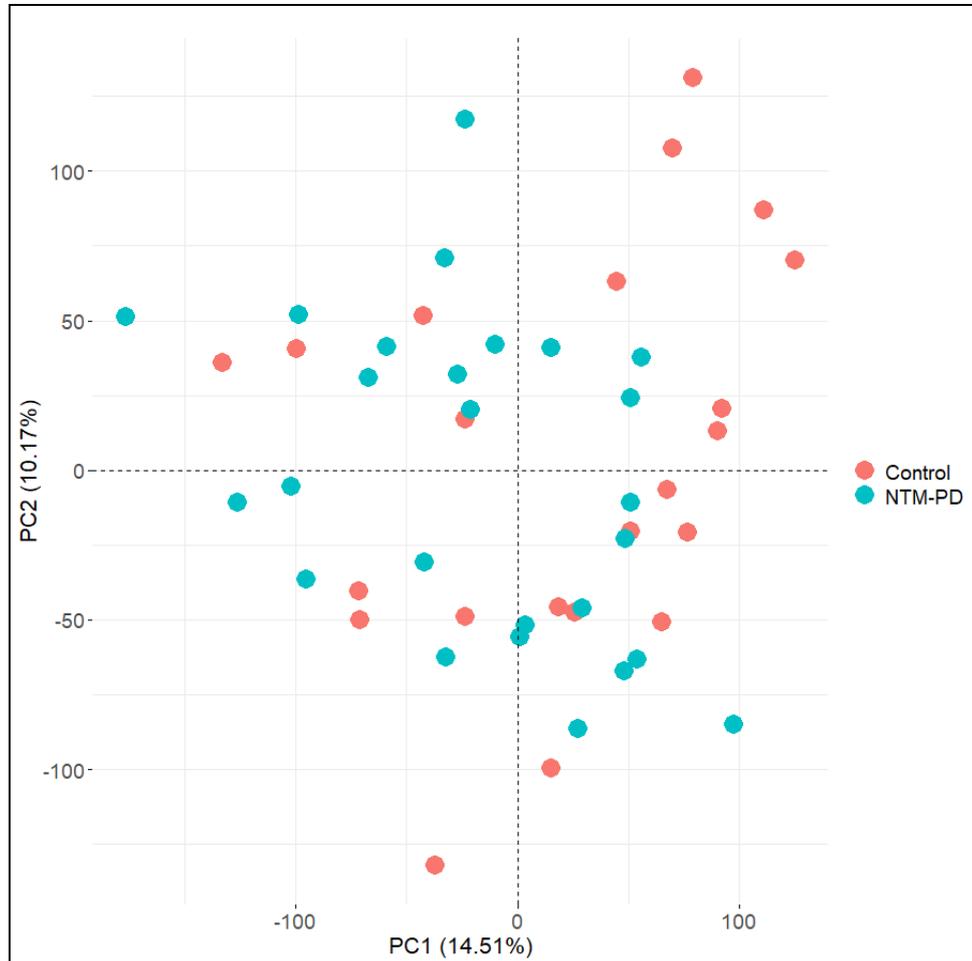


Figure 1. Principal component analysis depicting the unsupervised clustering of the case and control groups. Abbreviations: NTM-PD, nontuberculous mycobacterial pulmonary disease; PC, principal component.

3. DEG analysis

In participants with NTM-PD, we found 21 upregulated and 44 downregulated genes with an FDR below 0.05 compared with the control group. Tables 2 and 3 present the lists of upregulated and downregulated genes between the case and control groups.

Table 2. Upregulated genes between the case and control groups

Rank	Gene name	Description	Entrez ID	Fold change	P value
1	<i>GCSAML</i> (<i>C1orf150</i>)	germinal center associated signaling and motility like	148823	1.760	0.020
2	<i>MYCT1</i>	MYC target 1	80177	1.584	0.015
3	<i>SUCNR1</i>	succinate receptor 1	56670	1.564	0.047
4	<i>LEPR</i>	leptin receptor	3953	1.376	0.003
5	<i>CLIC2</i>	chloride intracellular channel 2	1193	1.305	0.047
6	<i>P2RY12</i>	purinergic receptor P2Y12	64805	1.296	0.020
7	<i>CISD2</i>	CDGSH iron sulfur domain 2	493856	1.262	0.019
8	<i>CCRL2</i>	C-C motif chemokine receptor like 2	9034	1.255	0.045
9	<i>C9orf40</i>	chromosome 9 open reading frame 40	55071	1.255	0.028
10	<i>ARL4A</i>	ADP ribosylation factor like GTPase 4A	10124	1.241	0.010
11	<i>MOSPD1</i>	motile sperm domain containing 1	56180	1.207	0.006
12	<i>STMP1</i> (<i>C7orf73</i>)	short transmembrane mitochondrial protein 1	647087	1.154	0.045
13	<i>YOD1</i>	YOD1 deubiquitinase	55432	1.148	0.020
14	<i>NSUN3</i>	NOP2/Sun RNA methyltransferase 3	63899	1.131	0.020
15	<i>CREG1</i>	cellular repressor of E1A stimulated genes 1	8804	1.113	0.018
16	<i>NT5C3</i> (<i>NT5C3A</i>)	5'-nucleotidase, cytosolic IIIA	51251	1.113	0.032
17	<i>YIPF6</i>	Yip1 domain family member 6	286451	1.089	0.020
18	<i>BNIP3L</i>	BCL2 interacting protein 3 like	665	1.083	0.046
19	<i>RWDD4</i>	RWD domain containing 4	201965	1.078	0.045
20	<i>STOM</i>	stomatin	2040	1.070	0.047
21	<i>RAB6A</i>	RAB6A, member RAS oncogene family	5870	1.058	0.030

Table 3. Downregulated genes between the case and control groups

Rank	Gene name	Description	Entrez ID	Fold change	P value
1	<i>MYBPH</i>	myosin binding protein H	4608	-3.501	0.032
2	<i>BFSP2</i>	beaded filament structural protein 2	8419	-2.912	0.003
3	<i>COL4A3</i>	collagen type IV alpha 3 chain	1285	-1.905	0.006
4	<i>PTPRB</i>	protein tyrosine phosphatase receptor type B	5787	-1.655	0.047
5	<i>COL4A4</i>	collagen type IV alpha 4 chain	1286	-1.529	0.003
6	<i>HNRNPA1P70</i> (<i>LOC341333</i>)	heterogeneous nuclear ribonucleoprotein A1 pseudogene 70	341333	-1.396	0.027
7	<i>PARK2</i>	parkin RBR E3 ubiquitin protein ligase	5071	-1.314	0.047
8	<i>PLXNA1</i>	Plexin A1	5361	-1.306	0.047
9	<i>ASSIP1</i>	Argininosuccinate synthetase 1 pseudogene 1	442167	-1.274	0.047
10	<i>PVT1</i>	Pvt1 oncogene	5820	-1.269	0.010
11	<i>LINC00544</i> (<i>LOC440131</i>)	long intergenic non-protein coding RNA 544	440131	-1.261	0.047
12	<i>ZC3H12B</i>	zinc finger CCCH-type containing 12B	340554	-1.244	0.047
13	<i>TMEM63A</i>	transmembrane protein 63A	9725	-1.170	0.005
14	<i>TSC1</i>	TSC complex subunit 1	7248	-1.123	0.047
15	<i>DOCK9</i>	dedicator of cytokinesis 9	23348	-1.116	0.020
16	<i>CTC1</i>	CST telomere replication complex component 1	80169	-1.116	0.047
17	<i>DGCR8</i>	DGCR8 microprocessor complex subunit	54487	-1.115	0.047
18	<i>PAN2</i>	poly(A) specific ribonuclease subunit PAN2	9924	-1.115	0.032
19	<i>SFI1</i>	SFI1 centrin binding protein	9814	-1.107	0.018
20	<i>LUC7L</i>	LUC7 like	55692	-1.101	0.047
21	<i>SGSM2</i>	small G protein signaling modulator 2	9905	-1.100	0.047

22	<i>ASXL1</i>	ASXL transcriptional regulator 1	171023	-1.100	0.018
23	<i>TRAF3</i>	TNF receptor associated factor 3	7187	-1.099	0.036
24	<i>RBM14</i>	RNA binding motif protein 14	10432	-1.096	0.047
25	<i>PLEC</i>	plectin	5339	-1.086	0.019
26	<i>CHD3</i>	chromodomain helicase DNA binding protein 3	1107	-1.085	0.038
27	<i>ZCCHC11</i>	terminal uridylyl transferase 4	23318	-1.081	0.047
28	<i>HIVEP2</i>	HIVEP zinc finger 2	3097	-1.069	0.020
29	<i>DIDO1</i>	death inducer-obliterator 1	11083	-1.067	0.005
30	<i>ANKZF1</i>	ankyrin repeat and zinc finger peptidyl tRNA hydrolase 1	55139	-1.066	0.028
31	<i>PBXIP1</i>	PBX homeobox interacting protein 1	57326	-1.064	0.032
32	<i>RPL36AL</i>	ribosomal protein L36a like	6166	-1.063	0.032
33	<i>STX16</i>	syntaxin 16	8675	-1.063	0.020
34	<i>SFSWAP</i>	splicing factor SWAP	6433	-1.061	0.031
35	<i>WHSC1L1</i>	nuclear receptor binding SET domain protein 3	54903	-1.054	0.027
36	<i>RASGRP2</i>	RAS guanyl releasing protein 2	10235	-1.052	0.032
37	<i>CELF1</i>	CUGBP Elav-like family member 1	10658	-1.052	0.027
38	<i>ELMO1</i>	engulfment and cell motility 1	9844	-1.051	0.047
39	<i>FAM120A</i>	family with sequence similarity 120 member A	23196	-1.051	0.044
40	<i>KLF6</i>	KLF transcription factor 6	1316	-1.047	0.047
41	<i>HMHA1</i>	Rho GTPase activating protein 45	23526	-1.044	0.047
42	<i>ENTPD4</i>	ectonucleoside triphosphate diphosphohydrolase 4	9583	-1.042	0.018
43	<i>ITSN2</i>	intersectin 2	50618	-1.035	0.047
44	<i>RBM5</i>	RNA binding motif protein 5	10181	1.033	0.049

PCA and a heatmap were generated using 65 DEGs, as shown in Figures 2 and 3. In Figure 2, PC1, accounting for 49.26% of the total variance, effectively distinguished the two groups. PC2 explained 10.43% of the variance, providing further separation, albeit to a lesser extent.

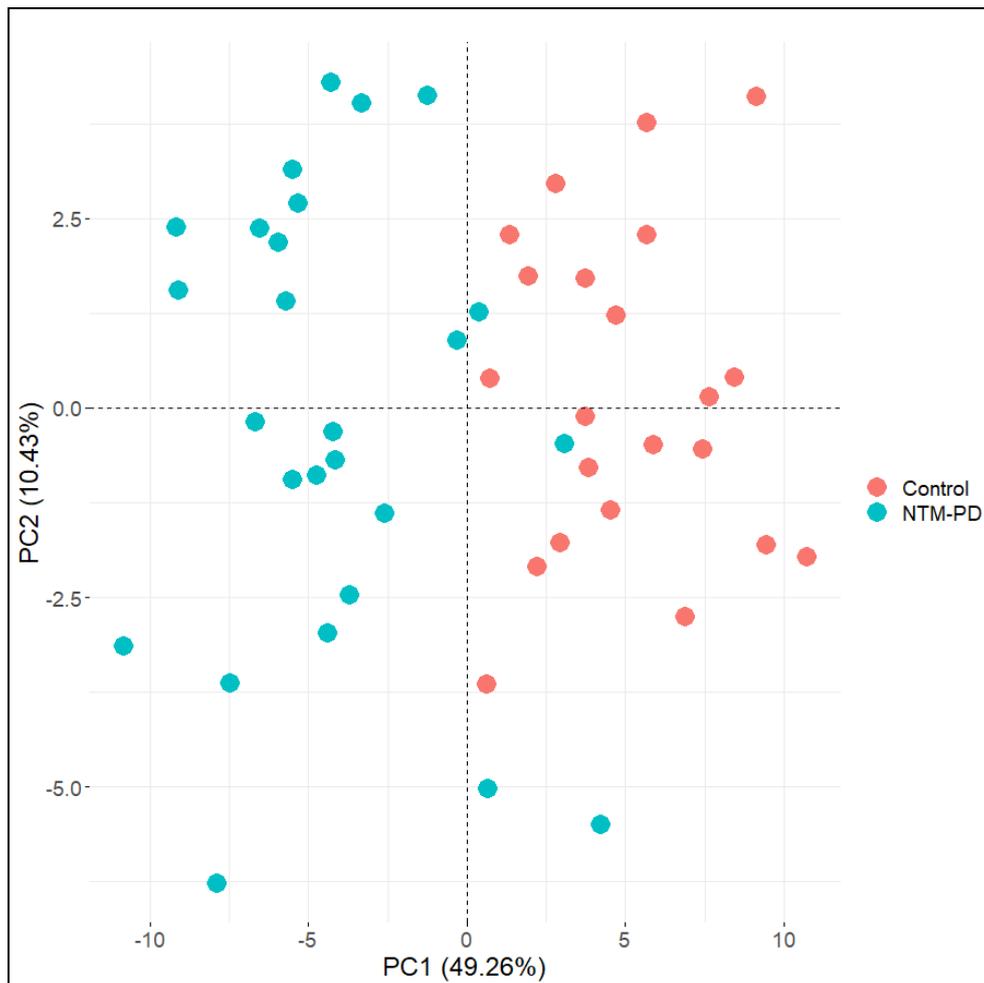


Figure 2. Principal component analysis using 65 DEGs between the case and control groups. Abbreviations: NTM-PD, nontuberculous mycobacterial pulmonary disease; PC, principal component.

However, there were no differences between healthy controls and subgroups of NTM-PD: subgroups by causative species (MAC, MABC, and others) and radiologic types (NB or FC) (Figures 4 and 5).

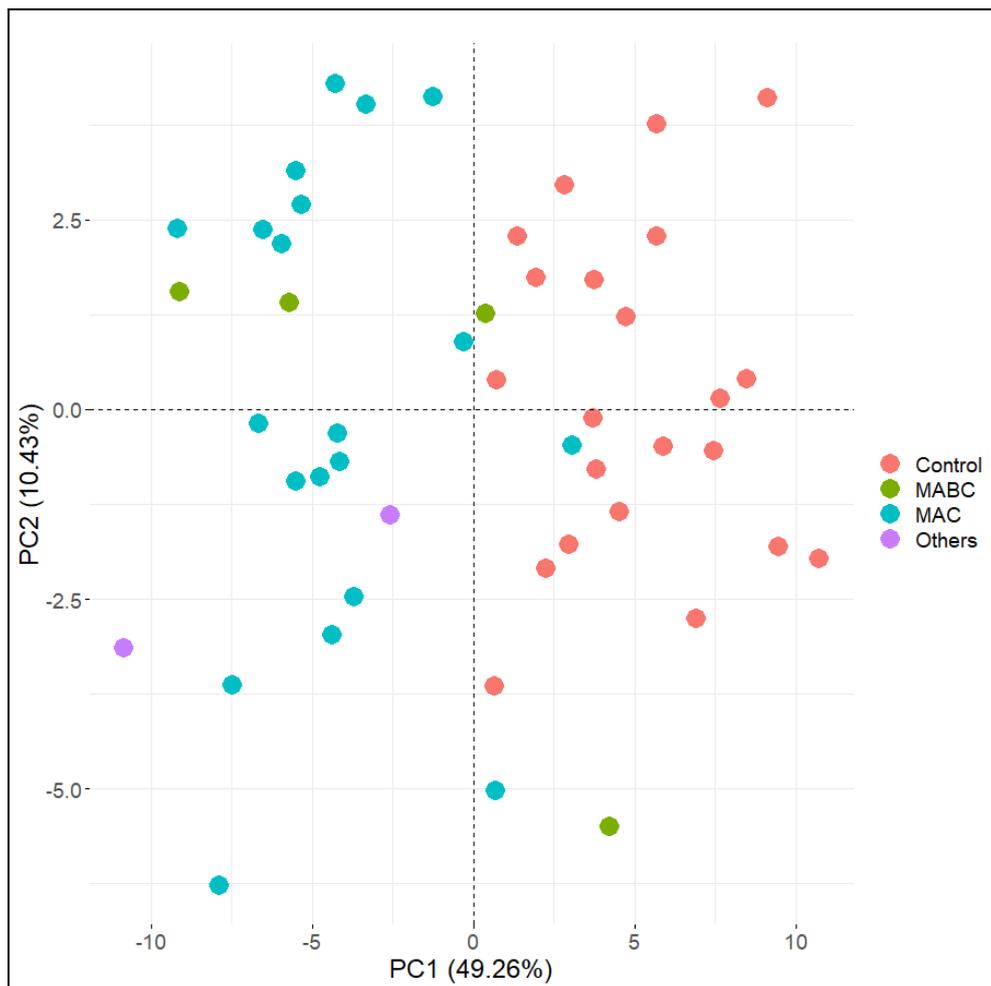


Figure 4. Principal component analysis using the 65 DEGs between the case and control groups. The case group was divided into subgroups based on the causative species. Abbreviations: MABC, *M. abscessus* complex; MAC, *M. avium* complex; PC, principal component.

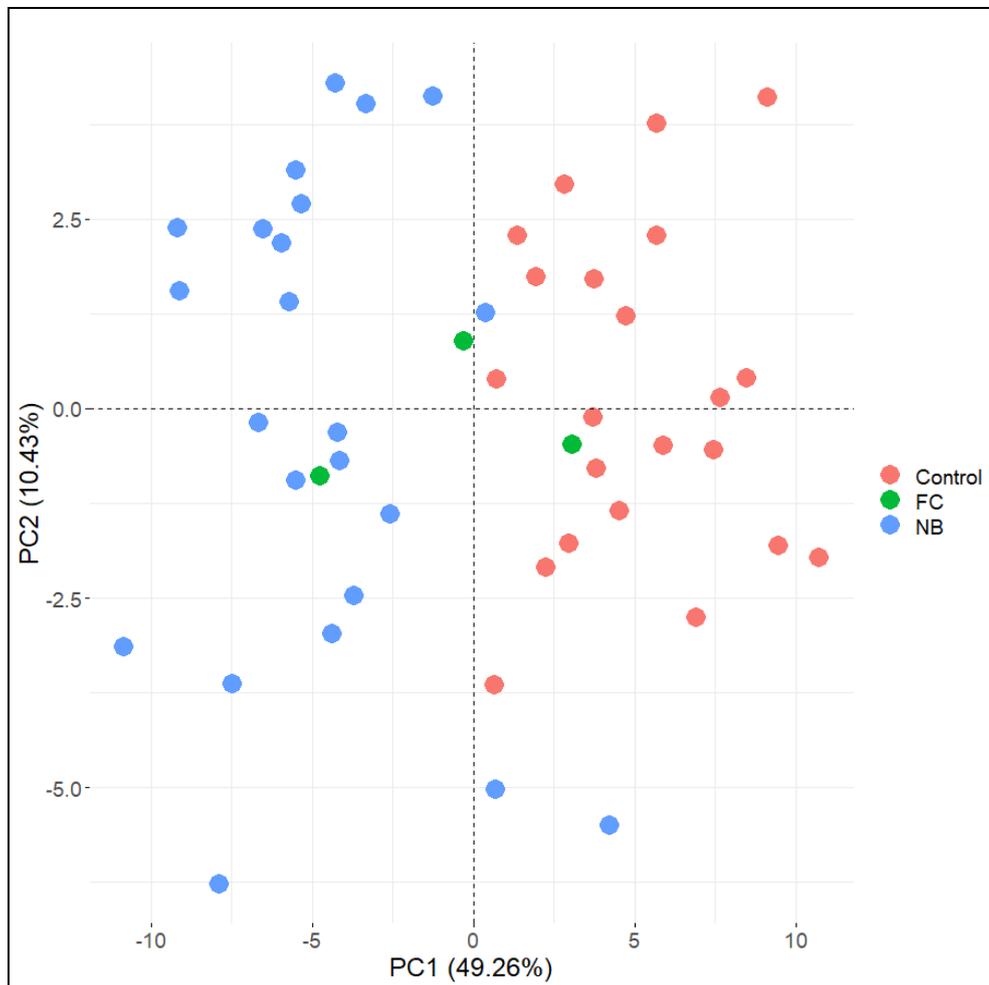


Figure 5. Principal component analysis using 65 DEGs between the case and control groups. The case group was divided into subgroups based on the radiologic types. Abbreviations: FC, fibrocavitary; NB, nodular bronchiectatic; PC, principal component.

4. Protein-protein connection

The protein-protein connections were assessed among the DEGs, and the protein interactions were plotted, as shown in Figures 6 and 7 and Table 4. A confidence level of 0.15 was set as the minimum requirement for interaction scores.

Figure 6 represents the protein-protein interaction network with 53 nodes and 92 edges. The nodes denote the number of proteins, while the edges represent their interactions.

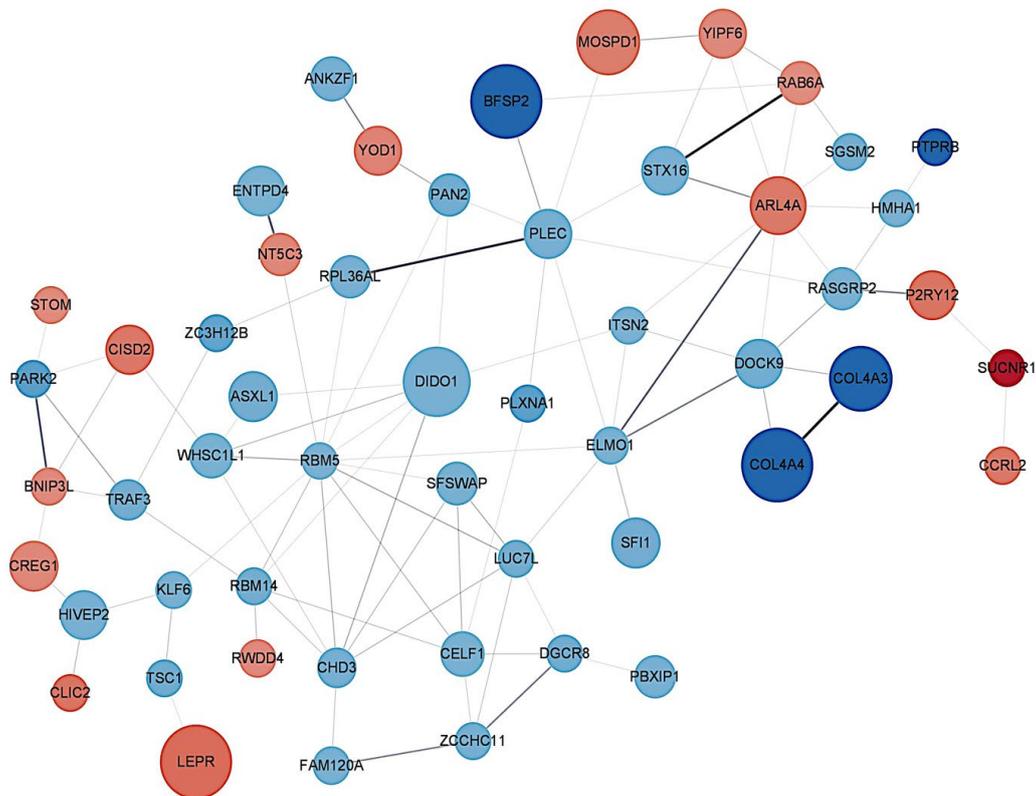


Figure 6. Co-expression protein network construction. Red nodes represent significantly upregulated genes; blue nodes represent significantly downregulated genes. Node size is inversely related to the P value; edge color and edge width are directly related to the confidence score.

The MCODE plugin of Cytoscape was used to interpret the closely interlinked regions in clusters from the network of proteins. The cluster finding parameters included a degree cutoff of 2 to exclude loops, a node score cutoff of 0.2, a kappa score of 2, and a max depth of 100, which limits the cluster size for co-expressing networks. Four clusters were identified, which were related to immune response, intracellular transportation, and GTPase regulation. Figure 7 and Table 4 display each cluster and its associated genes.

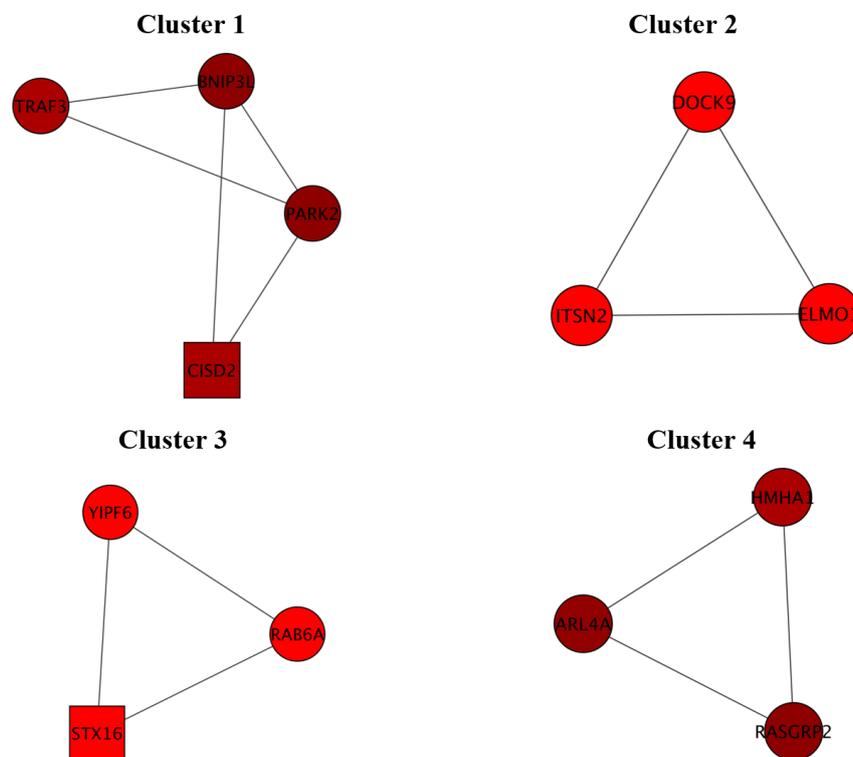


Figure 7. Modules of protein-protein interaction networks. The molecular complex detection (MCODE) plugin of Cytoscape was used for the analysis. Node shapes indicate the cluster status of the nodes. A square represents the seed (highest scoring node in the cluster), and a circle represents clustered proteins. Node color represents the node score; a range from black to red indicates the MCODE computed node scores (lowest to highest, respectively).

Table 4. Modules of protein-protein interaction networks

Cluster	Description	Score	Nodes	Edges	Node IDs
1	Mitophagy, Immune response	3.333	4	5	BNIP3L, CISD2, PARK2, TRAF3
2	Regulation of GTPase activity	3	3	3	DOCK9, ELMO1, ITSN2
3	Retrograde transport, endosome to Golgi, trans-Golgi network	3	3	3	RAB6A, STX16, YIPF6
4	Regulation of GTPase activity (closely linked to trans-Golgi network)	3	3	3	ARL4A, HMHA1, RASGRP2

5. Ontology and pathway enrichment analysis

To further understand which aspects of the gene signature were differentially expressed between the NTM-PD and control groups, the DAVID functional annotation tool was used to analyze the GO terms (Figure 8). Notably, regulation of autophagy (GO:0010506) was upregulated in the case group.

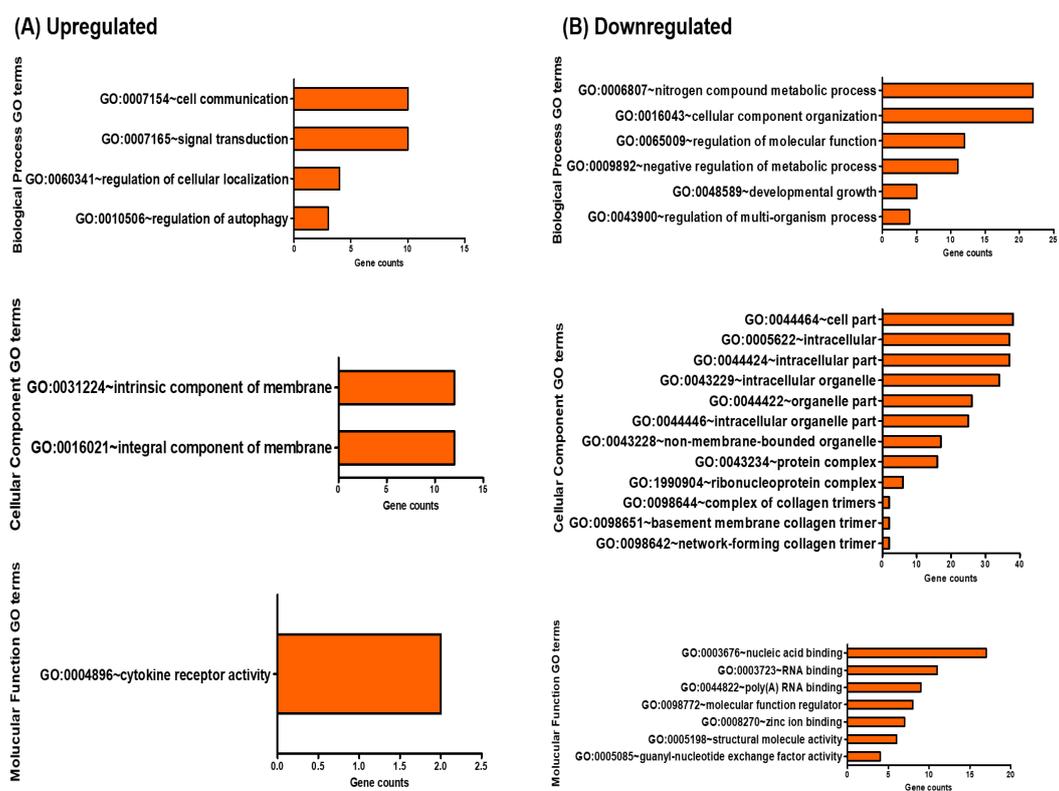


Figure 8. DAVID functional GO analysis. GO analysis with (A) 21 upregulated DEGs and (B) 44 downregulated DEGs.

Figure 9 presents the gene set enrichment analysis with GO terms and pathway terms. The connectivity of pathways in the network is determined by functional nodes and edges that are shared among the DEGs with a kappa score of 0.4. The enrichment shows only significant pathways ($P \leq 0.05$). The size of the nodes represents the values of $P \leq 0.05$. The color of the nodes indicates the specific functional class they are associated with. Detailed information on each node is provided in Tables 5 and 6.

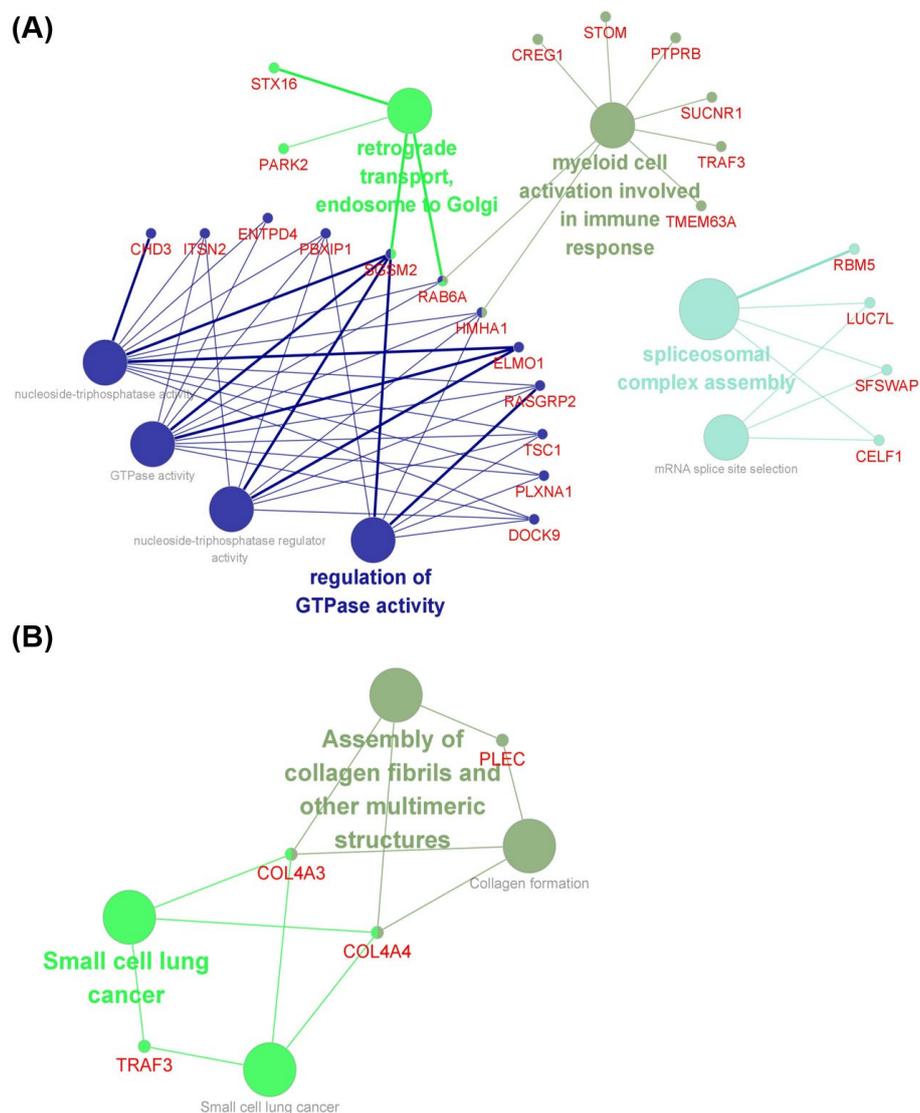


Figure 9. Enrichment analysis. (A) Enrichment by GO terms using the ClueGO/CluePedia plugin from Cytoscape. Vital molecular functions and biological processes associated with the DEGs are shown, along with the specific gene interactions. B) Enrichment by pathway terms visualized using the ClueGo/CluePedia plugin from Cytoscape. The plugin provides a comprehensive analysis of DEGs, including KEGG, REACTOME, and Wiki pathways.

Table 5. GO enrichment analysis conducted on the DEGs between the case and control groups

Source	Node color	Term	Genes	P value
GO_Biological Process		GO:0000245_spliceosomal complex assembly	<i>CELF1, LUC7L, RBM5, SFSWAP</i>	3.95E-05
GO_Biological Process		GO:0043087_regulation of GTPase activity	<i>ARHGAP45, DOCK9, ELMO1, ITSN2, PBXIP1, RASGRP2, SGSM2, TSC1</i>	1.10E-04
GO_Molecular Function		GO:0003924_GTPase activity	<i>ARHGAP45, DOCK9, ELMO1, ENTPD4, ITSN2, PBXIP1, PLXNA1, RAB6A, RASGRP2, SGSM2, TSC1</i>	1.35E-04
GO_Biological Process		GO:0006376_mRNA splice site selection	<i>CELF1, LUC7L, SFSWAP</i>	1.49E-04
GO_Molecular Function		GO:0017111_nucleoside-triphosphatase activity	<i>ARHGAP45, CHD3, DOCK9, ELMO1, ENTPD4, ITSN2, PBXIP1, PLXNA1, RAB6A, RASGRP2, SGSM2, TSC1</i>	1.67E-04
GO_Biological Process		GO:0042147_retrograde transport, endosome to Golgi	<i>PRKN, RAB6A, SGSM2, STX16</i>	2.52E-04
GO_Biological Process		GO:0060589_nucleoside-triphosphatase regulator activity	<i>ARHGAP45, DOCK9, ELMO1, ITSN2, PBXIP1, RASGRP2, SGSM2, TSC1</i>	2.77E-04
GO_Biological Process		GO:0002275_myeloid cell activation involved in immune response	<i>ARHGAP45, CREG1, PTPRB, RAB6A, STOM, SUCNR1, TMEM63A, TRAF3</i>	5.12E-04

Table 6. Pathway enrichment analysis conducted on the DEGs between the case and control groups

Source	Node color	Term	Genes	P value
REACTOME _Pathways		R-HSA:2022090_Assembly of collagen fibrils and other multimeric structures	<i>COL4A3</i> , <i>COL4A4</i> , <i>PLEC</i>	1.43E-03
REACTOME _Pathways		R-HSA:1474290_Collagen formation	<i>COL4A3</i> , <i>COL4A4</i> , <i>PLEC</i>	4.32E-03
KEGG _Pathways		KEGG:05222_Small cell lung cancer	<i>COL4A3</i> , <i>COL4A4</i> , <i>TRAF3</i>	4.59E-03
Wiki _Pathways		WP:4658_Small cell lung cancer	<i>COL4A3</i> , <i>COL4A4</i> , <i>TRAF3</i>	5.48E-03

6. Immune cell deconvolution

The differential distribution of immune cells between the NTM-PD and control groups was visualized using box plots in Figure 10. Each box plot represents the interquartile range (IQR) of the cell proportions, with the line inside the box indicating the median. The whiskers extend to 1.5 times the IQR. Any individual data points beyond the whiskers are considered outliers.

Neutrophil was the most common cell type in both groups (NTM-PD, 7.0% [IQR, 5.4%–9.2%]; Controls, 6.6% [IQR, 5.5%–8.7%]; $P = 0.644$), followed by monocytes (NTM-PD, 3.3% [IQR, 2.9%–3.8%]; Controls, 3.4% [IQR, 3.1%–3.9%]; $P = 0.601$). The proportions of other immune cell types were also similar between the two groups ($P > 0.05$ for all).

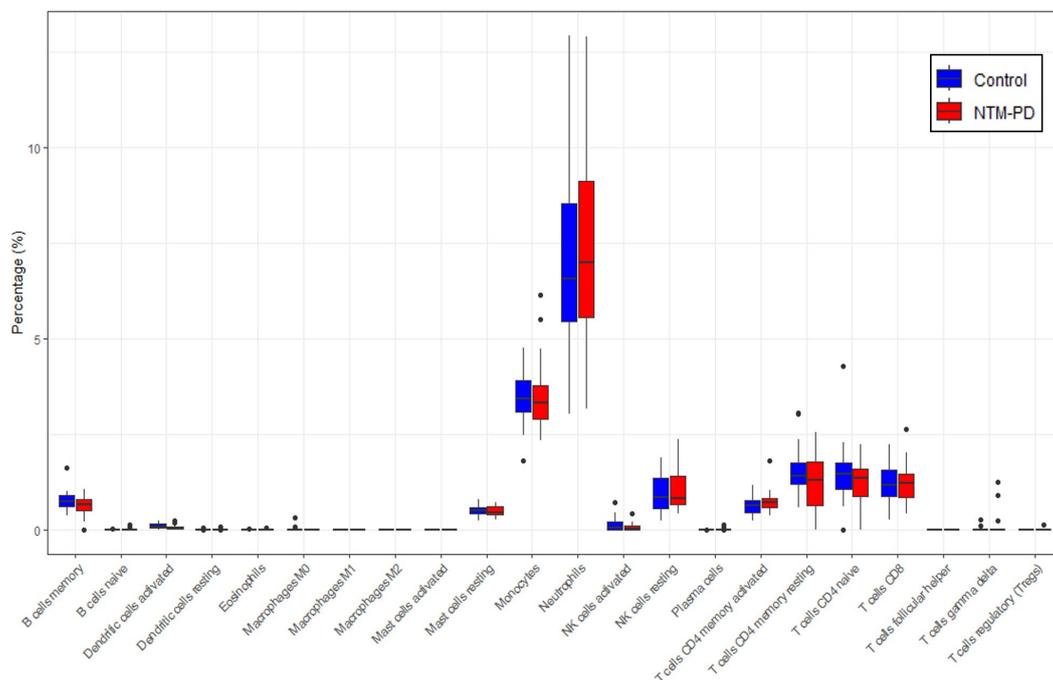


Figure 10. Proportions of 22 types of immune cells between the case and control groups. Abbreviations: NTM-PD, nontuberculous mycobacterial pulmonary disease.

7. Analysis of previously reported genes related to NTM-PD

Cowan *et al.*²² reported 213 DEGs in whole-blood samples from participants with NTM-PD compared to controls. We applied these gene sets to the current data, but no distinct clustering patterns were observed in the PCA (Figure 11).

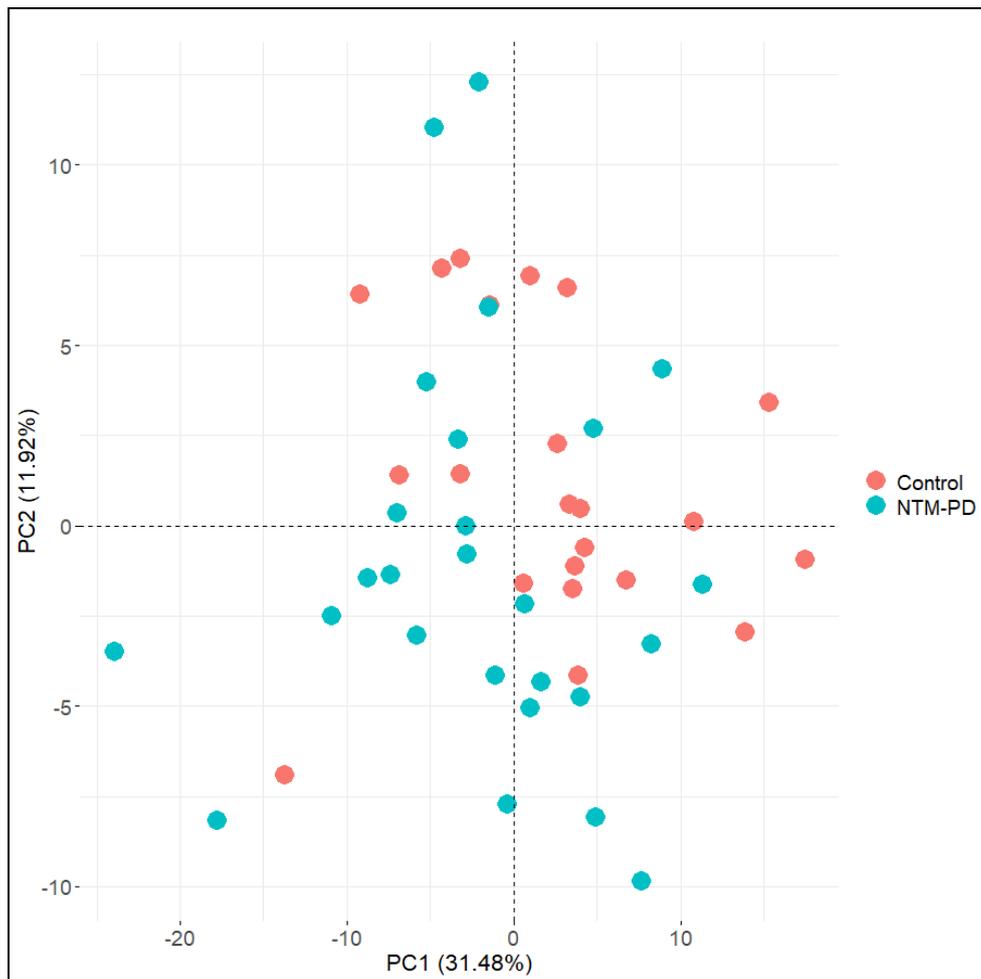


Figure 11. Principal component analysis using previously reported gene sets between the case and control groups. Abbreviations: NTM-PD, nontuberculous mycobacterial pulmonary disease; PC, principal component.

We also examined the expression levels of previously reported NTM-PD-related genes.²³ We explored the FDR and *P*-value of 44 genes from previous studies, but none showed statistical significance in distinguishing between NTM-PD patients and healthy controls in the current data (Table 7).

Table 7. Analysis of genes previously reported to be associated with NTM-PD in the current study samples

Gene	FDR	<i>P</i> -value	Gene	FDR	<i>P</i> -value
<i>GATA2</i>	0.188	0.003	<i>TIGIT</i>	0.860	0.514
<i>LDHB</i>	0.232	0.009	<i>FLJ45825</i>	0.871	0.530
<i>PSPH</i>	0.311	0.019	<i>MST1R</i>	0.878	0.540
<i>NELL2</i>	0.409	0.036	<i>ANKRD6</i>	0.885	0.552
<i>SLC29A1</i>	0.462	0.051	<i>XCL1</i>	0.902	0.578
<i>GZMK</i>	0.511	0.137	<i>MAP2K4</i>	0.919	0.610
<i>IFNGR1</i>	0.511	0.117	<i>ISG15</i>	0.944	0.653
<i>MPEG1</i>	0.511	0.074	<i>STK17A</i>	0.971	0.706
<i>MUC12</i>	0.511	0.102	<i>XCL2</i>	0.975	0.714
<i>PPIH</i>	0.511	0.090	<i>TPBG</i>	0.989	0.740
<i>NFATC2</i>	0.530	0.161	<i>CFTR</i>	1	0.818
<i>IFNG</i>	0.540	0.170	<i>FAHD2A</i>	1	0.862
<i>TTK</i>	0.540	0.171	<i>IFNGR2</i>	1	0.954
<i>IL12RB1</i>	0.598	0.212	<i>IL2RB</i>	1	0.917
<i>RCOR3</i>	0.652	0.258	<i>IRF8</i>	1	0.834
<i>AK5</i>	0.664	0.275	<i>PMS2P1</i>	1	0.891
<i>KRT83</i>	0.683	0.293	<i>SAMD3</i>	1	0.806
<i>ORC3</i>	0.704	0.314	<i>SLC11A1</i>	1	0.998
<i>PZP</i>	0.722	0.332	<i>STAT1</i>	1	0.958
<i>CRTAM</i>	0.734	0.344	<i>TARP</i>	1	0.989
<i>A2M</i>	0.736	0.347	<i>GUSBP14</i>	NA	NA
<i>FCRL3</i>	0.841	0.488	<i>IFNLRI</i>	NA	NA

NA, not available.

8. *PARK2* gene as a diagnostic biomarker

After reviewing the results and previous studies, we found that the expression level of the *PARK2* gene could differentiate samples with NTM-PD from healthy controls. Figure 12A presents the expression levels of the *PARK2* gene in the study samples. The ROC curve of the *PARK2* gene is presented in Figure 12B. The AUC was 0.813 (95% confidence interval 0.694–0.932), suggesting a satisfactory discriminatory ability (sensitivity, 61.5%; specificity, 95.5%; positive predictive value, 32.3%; negative predictive value, 5.9%).

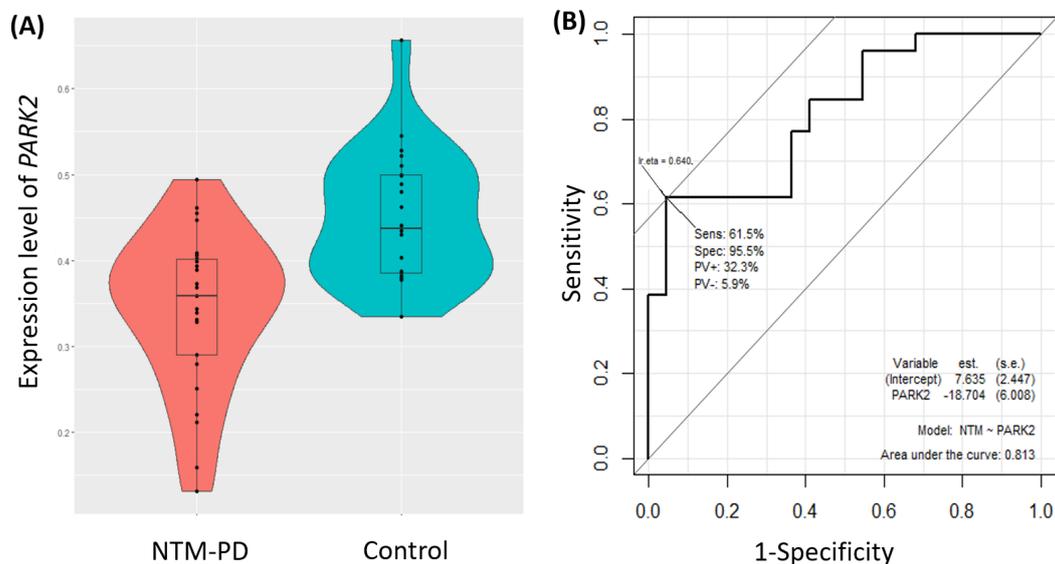


Figure 12. The *PARK2* gene for diagnosing NTM-PD. (A) Expression levels of *PARK2* in individuals with NTM-PD and healthy controls ($P = 0.047$). (B) The receiver operating characteristic curves depicting the predictive potential of the *PARK2* gene in classifying a sample as NTM-PD or control. Abbreviations: NTM-PD, nontuberculous mycobacterial pulmonary disease.

However, the expression levels of the *PARK2* gene between pre-treatment and post-treatment samples from participants with NTM-PD were similar (Figure 13).

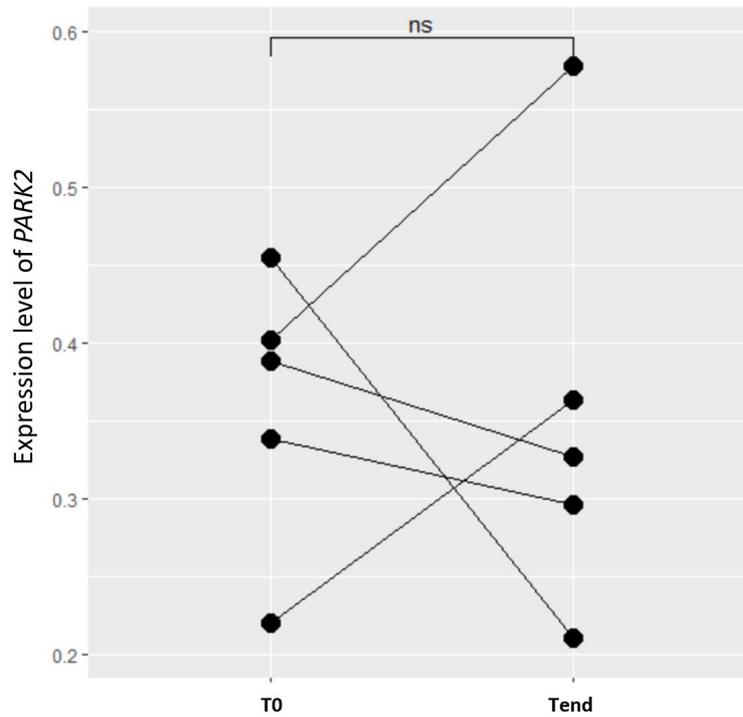


Figure 13. Expression levels of the *PARK2* gene between pre-treatment and post-treatment samples of NTM-PD. T0 indicates samples collected at the beginning of the treatment, while Tend represents samples collected at the end of the treatment. The two groups had no statistical differences ($P > 0.05$).

IV. DISCUSSION

To the best of our knowledge, this is the first study comparing the RNA expression profile of whole blood between the NTM-PD and control groups in a Korean patient cohort. We identified 65 DEGs in the participants with NTM-PD compared to the healthy controls. The *PARK2* gene was found to be downregulated in participants with NTM-PD, suggesting its potential as a diagnostic biomarker.

Genetic predisposition profoundly impacts the development, progression, and morbidities of many diseases. Bronchiectasis is a common comorbidity in patients with NTM-PD. The onset of bronchiectasis is linked to a reduction in gene expression associated with cell adhesion and an increase in gene expression related to inflammation.²⁴ Additionally, there is a decrease in the expression of genes in the Wnt signaling pathway, along with an elevation in the expression of genes related to ciliogenesis.²⁴ Notably, genotype-based categorization and treatment of bronchiectasis substantially improve its clinical course, especially in cases of cystic fibrosis and primary ciliary dyskinesia.²⁵ Chronic obstructive pulmonary disease (COPD), another common condition combined with NTM-PD, is a heterogeneous condition affected by early-life risk factors, individual and social factors, the external environment, and harmful exposures.²⁶ Lung development genes, which are considered among the critical early-life risk factors, regulate adult lung function and contribute to the development of both restrictive and obstructive lung function.²⁷ The heritability of lung function and COPD is estimated to be 38–50%.²⁶ Therefore, investigating the genetic background of NTM-PD holds promise in explaining its pathobiology, developing biomarkers, and improving clinical outcomes.

Blood-based transcriptomic analysis is widely utilized in various diseases, including asthma, acute leukemia, and inflammatory bowel diseases.²⁸⁻³⁰ In infectious diseases, diverse external stimuli also lead to changes in mRNA expression through transcriptional responses. As direct pathogen detection is not always possible in individuals with various conditions, several studies have investigated blood-based transcriptomic signatures in

infectious diseases, including tuberculosis.³¹ Whole-blood genetic signatures have shown promising diagnostic performance in assessing the risk of tuberculosis progression and monitoring treatment response.³²⁻³⁴ Moreover, a point-of-care triage test for tuberculosis using fingerstick blood has achieved the minimum target product profile set by the World Health Organization (at least 90% sensitivity and 70% specificity) in the interim analysis.³⁵

Only a few studies have investigated genetic signatures in NTM-PD. Matsuyama *et al.*³⁶ conducted RNA sequencing on NTM-infected human respiratory epithelium. They reported that genes related to cilia were downregulated, while those related to cytokines, chemokines, and cholesterol biosynthesis were upregulated in NTM-infected epithelium. Cowman *et al.*²² performed a microarray analysis of whole-blood gene expression on 25 participants with NTM-PD and 27 controls. They reported the downregulation of 213 transcripts associated with T cell signaling, including the *IFNG* (IFN- γ) gene, which plays an essential role in antimycobacterial immunity. Cho *et al.*²³ conducted a genome-wide association study on 403 participants with NTM-PD and 306 healthy controls in Korea. They indicated that expression levels of the proapoptotic *STK17A* (serine/threonine kinase 17a) gene may be associated with susceptibility to NTM-PD.

However, there is limited evidence for the genetic characteristics of NTM-PD, and the results are inconsistent. The *IFNG* gene, which was found to be downregulated in the study by Cowman *et al.*,²² was not identified as a differentially expressed gene by Cho *et al.*²³ (FDR = 0.709, $P = 0.185$). The alteration in the expression of the *STK17A* gene, reported by Cho *et al.*,²³ was not significant in the results of Cowman *et al.*²² In the current study, the expression levels of previously reported NTM-PD-related genes, including *IFNG* and *STK17A*, were similar between the case and control groups (Table 7). Moreover, the gene lists from Cowman *et al.*²² could not distinguish between the case and control groups in the current study (Figure 11). Only one gene, *DOCK9* (dedicator of cytokinesis 9), overlapped between the current study and the study by Cowman *et al.*²²

Limited commonality among genetic signatures was also observed in the studies of

tuberculosis; 563 out of 721 genes were detected only once among 30 previously published studies.³² There are several possible reasons for this inconsistency. First, research on human cells exhibits diversity in the types of samples used, such as peripheral blood mononuclear cells, broncho-alveolar lavage fluid, and whole blood.³⁷ Second, the study populations show variations in age, infecting species, comorbidities, stage of treatment, and geographic location. Third, the substances used to stimulate cytokine assay studies are not consistent and include phytohemagglutinin, lipopolysaccharide, and neutralized bacteria.³⁷ Therefore, caution is needed when comparing and interpreting the results, and further integrated analysis would be required.

One of the notable findings from the current study is that *PARK2* was downregulated in participants with NTM-PD. Mutations in the *PARK2* gene increase the risk of developing Parkinson's disease.³⁸ However, polymorphisms in the regulatory region of *PARK2* lead to reduced expression of the *PARK2* protein, known as Parkin, which has been linked to a higher susceptibility to intracellular pathogens such as *Mycobacterium leprae*, *Mycobacterium ulcerans*, and *Salmonella enterica* serovar Typhi.³⁹⁻⁴²

Autophagy serves as an innate immune response mechanism for eliminating intracellular pathogens.⁴³ Parkin, an E3 ubiquitin-ligase, plays a role in this process; Parkin-mediated ubiquitination recruits ubiquitin-adaptors, promoting autophagic targeting of mycobacteria.⁴⁴ Parkin also influences T-cell stimulation in the mitochondrial antigen presentation pathway.⁴⁵ Moreover, the downregulation of Parkin results in decreased interleukin-6 (IL-6) and monocyte chemoattractant protein 1 (MCP-1/*CCL2*) production, suggesting an influence of Parkin in multiple immune-related pathways.⁴⁶ Therefore, the decreased expression of *PARK2* may be associated with the development of NTM-PD.

The expression levels of the *PARK2* gene differed between the NTM-PD and control groups (Figure 12), but they did not change after treatment, with favorable outcomes (Figure 13). Therefore, the decreased levels of *PARK2* could be associated with the host's susceptibility rather than a response to infection or disease severity.

Various genes were also differentially expressed between the case and control groups (Tables 2 and 3). One of these genes, *RAB6*, a small GTPase regulating endosomal trafficking pathways, presents ligands from *Mycobacterium tuberculosis* to mucosal-associated invariant T cells as an early response to infection.⁴⁷ *RAB6A* expression was upregulated in the case group, possibly resulting from NTM infection. Another upregulated gene, *P2RY12*, is considered to be a critical player in the inflammatory response.⁴⁸

A few genes were linked to COPD, a common comorbidity of NTM-PD. Type IV collagen, the most abundant nonfibrillar collagen in the lung, is linked to basement membrane integrity. The degradation of *COL4A3* has been associated with the disease activity of asthma and COPD.^{49,50} As extracellular matrix proteins in the lung play a critical role in the adhesion and invasion of various pathogens, the decreased levels of *COL4A3* and *COL4A4* in this study may be related to NTM infection, although the evidence from previous studies is insufficient.⁵¹ *LEPR* gene polymorphisms are linked to lung function decline in COPD,⁵² and *NSUN3* gene is associated with lung cancer development in COPD.⁵³ *HIVEP2* gene was included in a blood-based transcriptomic risk score for COPD, which was associated with COPD, lung function decline, and COPD-related traits.⁵⁴ Further investigation would be required to identify the role of the above genes in the pathobiology of NTM-PD.

This study has several limitations. First, the case-control, observational, cross-sectional design does not allow for any conclusions about causality. Second, the number of participants was limited to those from a single center in Korea. Therefore, further validation studies are needed to investigate our findings in experimental settings involving diverse ethnic populations. Third, we could not identify disease-specific genes associated with NTM-PD, as a decreased level of the *PARK2* gene has been reported in infections caused by other intracellular pathogens.

V. CONCLUSION

We have identified genetic signatures associated with NTM-PD in a cohort of Korean patients. The downregulated *PARK2* gene could potentially serve as a biomarker for diagnosis, offering further development opportunities. Investigating the interaction between genes and metabolites may provide valuable insights for novel approaches to diagnosing NTM-PD.

REFERENCES

1. Jeon D. Infection Source and Epidemiology of Nontuberculous Mycobacterial Lung Disease. *Tuberc Respir Dis* 2018;81.
2. Griffith DE, Aksamit T, Brown-Elliott BA, Catanzaro A, Daley C, Gordin F, et al. An official ATS/IDSA statement: diagnosis, treatment, and prevention of nontuberculous mycobacterial diseases. *Am J Respir Crit Care Med* 2007;175:367-416.
3. Prevots DR, Marras TK. Epidemiology of human pulmonary infection with nontuberculous mycobacteria: a review. *Clin Chest Med* 2015;36:13-34.
4. Lee H, Myung W, Koh WJ, Moon SM, Jhun BW. Epidemiology of Nontuberculous Mycobacterial Infection, South Korea, 2007-2016. *Emerg Infect Dis* 2019;25:569-72.
5. Hoefsloot W, van Ingen J, Andrejak C, Angeby K, Bauriaud R, Bemer P, et al. The geographic diversity of nontuberculous mycobacteria isolated from pulmonary samples: an NTM-NET collaborative study. *Eur Respir J* 2013;42:1604-13.
6. Daley CL, Iaccarino JM, Lange C, Cambau E, Wallace RJ, Jr., Andrejak C, et al. Treatment of nontuberculous mycobacterial pulmonary disease: an official ATS/ERS/ESCMID/IDSA clinical practice guideline. *Eur Respir J* 2020;56.
7. Kwon YS, Koh WJ, Daley CL. Treatment of Mycobacterium avium Complex Pulmonary Disease. *Tuberc Respir Dis (Seoul)* 2019;82:15-26.
8. Kwon YS, Koh WJ. Diagnosis and Treatment of Nontuberculous Mycobacterial Lung Disease. *J Korean Med Sci* 2016;31:649-59.
9. Kwak N, Park J, Kim E, Lee CH, Han SK, Yim JJ. Treatment Outcomes of Mycobacterium avium Complex Lung Disease: A Systematic Review and Meta-analysis. *Clin Infect Dis* 2017;65:1077-84.
10. Pasipanodya JG, Ogbonna D, Ferro BE, Magomedze G, Srivastava S, Deshpande D, et al. Systematic Review and Meta-analyses of the Effect of

- Chemotherapy on Pulmonary Mycobacterium abscessus Outcomes and Disease Recurrence. *Antimicrob Agents Chemother* 2017;61.
11. Vinnard C, Mezocho A, Oakland H, Klingsberg R, Hansen-Flaschen J, Hamilton K. Assessing Response to Therapy for Nontuberculous Mycobacterial Lung Disease: Quo Vadis? *Front Microbiol* 2018;9:2813.
 12. Yoshizawa K, Aoki A, Shima K, Tanabe Y, Koya T, Hasegawa T, et al. Serum Anti-interferon-gamma Autoantibody Titer as a Potential Biomarker of Disseminated Non-tuberculous Mycobacterial Infection. *J Clin Immunol* 2020; doi:10.1007/s10875-020-00762-1.
 13. van Ingen J, Aksamit T, Andrejak C, Bottger EC, Cambau E, Daley CL, et al. Treatment outcome definitions in nontuberculous mycobacterial pulmonary disease: an NTM-NET consensus statement. *Eur Respir J* 2018;51.
 14. Dreszer TR, Karolchik D, Zweig AS, Hinrichs AS, Raney BJ, Kuhn RM, et al. The UCSC Genome Browser database: extensions and updates 2011. *Nucleic Acids Res* 2012;40:D918-23.
 15. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 2009;25:1105-11.
 16. Trapnell C, Hendrickson DG, Sauvageau M, Goff L, Rinn JL, Pachter L. Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nat Biotechnol* 2013;31:46-53.
 17. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res* 2019;47:D607-d13.
 18. Bindea G, Mlecnik B, Hackl H, Charoentong P, Tosolini M, Kirilovsky A, et al. ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics* 2009;25:1091-3.
 19. Bindea G, Galon J, Mlecnik B. CluePedia Cytoscape plugin: pathway insights

- using integrated experimental and in silico data. *Bioinformatics* 2013;29:661-3.
20. Newman AM, Steen CB, Liu CL, Gentles AJ, Chaudhuri AA, Scherer F, et al. Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat Biotechnol* 2019;37:773-82.
 21. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, et al. Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods* 2015;12:453-7.
 22. Cowman SA, Jacob J, Hansell DM, Kelleher P, Wilson R, Cookson WOC, et al. Whole-Blood Gene Expression in Pulmonary Nontuberculous Mycobacterial Infection. *Am J Respir Cell Mol Biol* 2018;58:510-8.
 23. Cho J, Park K, Choi SM, Lee J, Lee CH, Lee JK, et al. Genome-wide association study of non-tuberculous mycobacterial pulmonary disease. *Thorax* 2020; doi:10.1136/thoraxjnl-2019-214430.
 24. Xu K, Diaz AA, Duan F, Lee M, Xiao X, Liu H, et al. Bronchial gene expression alterations associated with radiological bronchiectasis. *Eur Respir J* 2023;61.
 25. Flume PA, Chalmers JD, Olivier KN. Advances in bronchiectasis: endotyping, genetics, microbiome, and disease heterogeneity. *Lancet* 2018;392:880-90.
 26. Christenson SA, Smith BM, Bafadhel M, Putcha N. Chronic obstructive pulmonary disease. *Lancet* 2022;399:2227-42.
 27. Portas L, Pereira M, Shaheen SO, Wyss AB, London SJ, Burney PGJ, et al. Lung Development Genes and Adult Lung Function. *Am J Respir Crit Care Med* 2020;202:853-65.
 28. Singh A, Shannon CP, Kim YW, Yang CX, Balshaw R, Cohen Freue GV, et al. Novel Blood-based Transcriptional Biomarker Panels Predict the Late-Phase Asthmatic Response. *Am J Respir Crit Care Med* 2018;197:450-62.
 29. Biasci D, Lee JC, Noor NM, Pombal DR, Hou M, Lewis N, et al. A blood-based prognostic biomarker in IBD. *Gut* 2019;68:1386-95.
 30. Arindrarto W, Borràs DM, de Groen RAL, van den Berg RR, Locher IJ, van

- Diessen S, et al. Comprehensive diagnostics of acute myeloid leukemia by whole transcriptome RNA sequencing. *Leukemia* 2021;35:47-61.
31. Kim CH, Choi G, Lee J. Host Blood Transcriptional Signatures as Candidate Biomarkers for Predicting Progression to Active Tuberculosis. *Tuberc Respir Dis (Seoul)* 2023;86:94-101.
 32. Vargas R, Abbott L, Bower D, Frahm N, Shaffer M, Yu WH. Gene signature discovery and systematic validation across diverse clinical cohorts for TB prognosis and response to treatment. *PLoS Comput Biol* 2023;19:e1010770.
 33. Gupta RK, Turner CT, Venturini C, Esmail H, Rangaka MX, Copas A, et al. Concise whole blood transcriptional signatures for incipient tuberculosis: a systematic review and patient-level pooled meta-analysis. *The Lancet Respiratory Medicine* 2020; doi:10.1016/s2213-2600(19)30282-6.
 34. Turner CT, Gupta RK, Tsaliki E, Roe JK, Mondal P, Nyawo GR, et al. Blood transcriptional biomarkers for active pulmonary tuberculosis in a high-burden setting: a prospective, observational, diagnostic accuracy study. *The Lancet Respiratory Medicine* 2020; doi:10.1016/s2213-2600(19)30469-2.
 35. Sutherland JS, van der Spuy G, Gindeh A, Thuong NTT, Namuganga A, Owolabi O, et al. Diagnostic Accuracy of the Cepheid 3-gene Host Response Fingerstick Blood Test in a Prospective, Multi-site Study: Interim Results. *Clin Infect Dis* 2022;74:2136-41.
 36. 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;58:241-52.
 37. Ratnatunga CN, Lutzky VP, Kupz A, Doolan DL, Reid DW, Field M, et al. The Rise of Non-Tuberculosis Mycobacterial Lung Disease. *Frontiers in Immunology* 2020;11.
 38. Thomas B, Beal MF. Parkinson's disease. *Hum Mol Genet* 2007;16 Spec No. 2:R183-94.

39. Mira MT, Alcaïs A, Nguyen VT, Moraes MO, Di Flumeri C, Vu HT, et al. Susceptibility to leprosy is associated with PARK2 and PACRG. *Nature* 2004;427:636-40.
40. Chopra R, Ali S, Srivastava AK, Aggarwal S, Kumar B, Manvati S, et al. Mapping of PARK2 and PACRG overlapping regulatory region reveals LD structure and functional variants in association with leprosy in unrelated indian population groups. *PLoS Genet* 2013;9:e1003578.
41. Capela C, Dossou AD, Silva-Gomes R, Sopoh GE, Makoutode M, Menino JF, et al. Genetic Variation in Autophagy-Related Genes Influences the Risk and Phenotype of Buruli Ulcer. *PLoS Negl Trop Dis* 2016;10:e0004671.
42. Ali S, Vollaard AM, Widjaja S, Surjadi C, van de Vosse E, van Dissel JT. PARK2/PACRG polymorphisms and susceptibility to typhoid and paratyphoid fever. *Clin Exp Immunol* 2006;144:425-31.
43. Huang J, Brumell JH. Bacteria-autophagy interplay: a battle for survival. *Nat Rev Microbiol* 2014;12:101-14.
44. Manzanillo PS, Ayres JS, Watson RO, Collins AC, Souza G, Rae CS, et al. The ubiquitin ligase parkin mediates resistance to intracellular pathogens. *Nature* 2013;501:512-6.
45. Matheoud D, Sugiura A, Bellemare-Pelletier A, Laplante A, Rondeau C, Chemali M, et al. Parkinson's Disease-Related Proteins PINK1 and Parkin Repress Mitochondrial Antigen Presentation. *Cell* 2016;166:314-27.
46. de Léséleuc L, Orlova M, Cobat A, Girard M, Huong NT, Ba NN, et al. PARK2 mediates interleukin 6 and monocyte chemoattractant protein 1 production by human macrophages. *PLoS Negl Trop Dis* 2013;7:e2015.
47. Huber ME, Kurapova R, Heisler CM, Karamooz E, Tafesse FG, Harriff MJ. Rab6 regulates recycling and retrograde trafficking of MR1 molecules. *Sci Rep* 2020;10:20778.
48. Mansour A, Bachelot-Loza C, Nesseler N, Gaussem P, Gouin-Thibault I. P2Y12

- Inhibition beyond Thrombosis: Effects on Inflammation. *Int J Mol Sci* 2020;21.
49. Ronnow SR, Sand JMB, Langholm LL, Manon-Jensen T, Karsdal MA, Tal-Singer R, et al. Type IV collagen turnover is predictive of mortality in COPD: a comparison to fibrinogen in a prospective analysis of the ECLIPSE cohort. *Respir Res* 2019;20:63.
 50. Weckmann M, Bahmer T, Sand JM, Rank Rønnow S, Pech M, Vermeulen C, et al. COL4A3 is degraded in allergic asthma and degradation predicts response to anti-IgE therapy. *Eur Respir J* 2021;58.
 51. Singh B, Fleury C, Jalalvand F, Riesbeck K. Human pathogens utilize host extracellular matrix proteins laminin and collagen for adhesion and invasion of the host. *FEMS Microbiol Rev* 2012;36:1122-80.
 52. Hansel NN, Gao L, Rafaels NM, Mathias RA, Neptune ER, Tankersley C, et al. Leptin receptor polymorphisms and lung function decline in COPD. *Eur Respir J* 2009;34:103-10.
 53. Li Z, Xu D, Jing J, Wang J, Jiang M, Li F. Identification and Validation of Prognostic Markers for Lung Squamous Cell Carcinoma Associated with Chronic Obstructive Pulmonary Disease. *J Oncol* 2022;2022:4254195.
 54. Moll M, Boueiz A, Ghosh AJ, Saferali A, Lee S, Xu Z, et al. Development of a Blood-based Transcriptional Risk Score for Chronic Obstructive Pulmonary Disease. *Am J Respir Crit Care Med* 2022;205:161-70.

ABSTRACT(IN KOREAN)

비결핵항산균 폐질환 특이 유전자 발현 패턴 발굴

<지도교수 강 영 애>

연세대학교 대학원 의학과

박 영 목

배경: 비결핵 항산균은 주로 폐 질환을 일으키는 환경 균주이다. 비결핵 항산균 폐질환의 발생률은 전 세계적으로 증가하고 있으나 현재의 진단 및 치료는 한계점이 많다. 본 연구에서는 비결핵 항산균 폐질환 환자에서 유전자 발현 특성을 탐색하고 잠재적 바이오 마커를 식별하기 위해 RNA 서열 분석을 수행하였다.

방법: 우리는 대한민국의 3차 의료기관에서 비결핵 항산균 폐질환 환자와 건강한 대조군의 말초 혈액 표본을 수집하였다. 비결핵 항산균 환자군에서는 치료가 완료된 후에 추가 혈액 표본을 수집하였다. 수집한 표본에서 RNA 서열 분석을 수행하였으며, 발현량이 차이나는 유전자(differentially expressed gene)를 식별하였다. 또한 면역 세포 디콘볼루션(immune cell deconvolution)을 활용하여 22종류의 면역 세포의 구성을 정량적으로 분석하였다.

결과: 비결핵 항산균 폐질환 환자 26명(나이 중앙값, 58.0세; 여성, 84.6%; *M. avium* complex, 76.9%)과 건강한 대조군 22명(나이 중앙값, 58.5세; 여성, 90.9%)을 모집하였다. 대조군에 비해 비결핵 항산균 폐질환군에서 발현량이

증가한 21개의 유전자와 발현량이 감소한 44개의 유전자를 식별하였다. 온톨로지(ontology) 및 기능 강화 분석(functional enrichment analysis)결과 대조군에 비해 비결핵 항산균 폐질환 환자군에서 자가 포식(autophagy)과 관련된 유전자의 상승을 확인하였다. 면역 세포 디콘볼루션 결과에 따르면, 양 그룹 모두에서 중성구(neutrophil)의 비율이 가장 높았다. (비결핵 항산균 그룹 7.0%, 대조군 6.6%, $P = 0.644$) 다른 면역 세포 유형의 비율도 두 그룹 간에 유사하였다. ($P > 0.05$ for all)

*PARK2*는 유비퀴티네이션(ubiquitination)과 관련되어 있는 유전자로 비결핵 항산균 폐질환 환자군에서 발현량이 감소하였고 (Fold change -1.314, $P = 0.047$), 임상적으로 성공적 치료(microbiologic cure)를 마친 다섯 개의 혈액 표본에서도 그 발현량이 비슷하였다. *PARK2* 유전자가 감염이나 염증에 의한 결과라기 보다는 숙주의 감수성 (host susceptibility)와 관련이 있다는 것을 시사한다. *PARK2* 유전자의 발현 수준은 수신자 운영 특성 곡선의 면적(area under the receiver operating characteristic curve)이 0.813으로 (95% 신뢰구간 0.694—0.932) 진단의 바이오 마커로서 잠재력을 보였다.

결론: 이 연구는 한국의 비결핵 항산균 폐질환 환자의 유전적 특성을 탐색하였다. 비결핵 항산균 폐질환 환자에서 *PARK2* 유전자의 발현량 저하는 비결핵 항산균 폐질환의 진단을 위한 후보 유전자로서 바이오 마커 개발 기회를 제공할 수 있을 것이다.

핵심 되는 말: 비결핵항산균, 유전자 발현, 리보핵산