Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Single-cell Signatures Reveal Immune Dysregulation in Nontuberculous Mycobacterial Pulmonary Disease

Jee Youn Oh
Korea University Guro Hospital
Sanghyun Roh
Soonchunhyang University
Jun Hee Lee
Korea University Guro Hospital
Juwhan Choi
Korea University Guro Hospital
Young Kyung Ko
Korea University Guro Hospital
Min Ha Jeon
Korea University Guro Hospital
Sun-Young Kang
Soonchunhyang University
Hyun Myung Doo
Korea University Guro Hospital
Jeong-An Gim
vitastar@sch.ac.kr

Soonchunhyang University

Research Article

Keywords: nontuberculous mycobacterial pulmonary disease, bronchoalveolar lavage fluid, single-cell RNA sequencing, diagnostic biomarker, immune dysregulation

Posted Date: June 1st, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9670100/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Background

Nontuberculous mycobacterial (NTM) pulmonary disease is characterized by chronic infection and dysregulated immunity. However, cell type-specific transcriptional changes driving disease progression remain poorly defined.

Methods

We performed single-cell RNA sequencing (scRNA-seq) on bronchoalveolar lavage fluid (BALF) samples from individuals with NTM pulmonary disease and integrated these data with external control BALF single-cell datasets. Cell clustering, reference-based annotation, differential expression analysis, and KEGG pathway enrichment were performed, and selected transcriptional signatures were validated using in vitro NTM-infected cellular models.

Results

Distinct transcriptional profiles were identified in alveolar epithelial and innate immune cells. Alveolar type II cells showed increased expression of KRT7, MUC1, LAMP3, and ABCA3, with reduced ATP1A1 and ATP1B1. Monocytes exhibited higher CCR2, SELL, and TLR2, but lower LYZ and CX3CR1. Macrophages displayed decreased CD74 and SIGLEC1, among other immune-regulatory genes. These alterations were reproduced in in vitro NTM-infected alveolar type II cells, monocytes, and macrophages, confirming disease-relevant expression signatures.

Conclusions

Our integrated single-cell and in vitro transcriptomic analyses reveal consistent dysregulation of immune and epithelial cells in NTM pulmonary disease. These findings suggest impaired antigen presentation and reduced immune activation, particularly in macrophage and monocyte populations. The concordance between in vivo and in vitro data highlights the robustness of these transcriptional signatures and identifies potential targets for immunomodulatory therapy in NTM infection.

Background

Nontuberculous mycobacterial (NTM) disease has emerged as an important cause of chronic pulmonary infection, particularly among individuals with structural lung abnormalities or impaired host defense mechanisms (1, 2). Unlike Mycobacterium tuberculosis, NTM species are ubiquitous environmental organisms commonly present in soil and water, and exposure is frequent in the general population (1).

However, only a subset of exposed individuals develops clinically significant disease, underscoring the critical role of host susceptibility and immune regulation in disease pathogenesis (2).

The clinical course of NTM disease is highly heterogeneous. Disease onset, progression, treatment response, and prognosis vary considerably depending on complex interactions between host immune status, underlying lung structure, and the virulence and adaptability of specific NTM strains (3, 4). In many individuals, NTM colonization remains asymptomatic or self-limited, whereas in others, it results in progressive, chronic pulmonary infection that is difficult to eradicate despite prolonged antimicrobial therapy (2, 5). This variability poses significant challenges for timely diagnosis, accurate risk stratification, and personalized therapeutic decision-making.

Recent advances in transcriptomic technologies, particularly single-cell RNA sequencing (scRNA-seq), have revolutionized our ability to characterize immune responses at single-cell resolution (6). Unlike bulk RNA sequencing, scRNA-seq enables the dissection of cellular heterogeneity, identification of rare or functionally distinct immune cell subsets, and delineation of dynamic activation or exhaustion states that might be obscured in population-level analyses (6, 7). This level of resolution is particularly relevant in NTM disease, where subtle immune dysregulation—rather than overt immunodeficiency—may determine whether environmental exposure leads to benign colonization or progressive infection (8).

Despite its growing utility in infectious and immunological research, the application of scRNA-seq to NTM disease remains limited. Comprehensive, cell type-specific characterization of immune responses in individuals with NTM infection is still lacking, hindering the identification of reliable biomarkers for disease activity, progression, and prognosis (3). Elucidating transcriptional signatures associated with impaired pathogen clearance, chronic inflammation, or immune exhaustion may provide critical mechanistic insights and inform the development of improved diagnostic strategies and host-directed therapeutic approaches (9).

In this study, we conducted single-cell RNA sequencing of BALF-derived immune and epithelial cells from individuals with NTM pulmonary disease. By characterizing immune cell populations, their functional states, and associated molecular pathways, we aimed to delineate the transcriptional landscapes underlying immune dysregulation in NTM disease. Specifically, we applied scRNA-seq to BALF-derived cells from patients with NTM pulmonary disease and external control BALF datasets to define cell type-specific transcriptional programs and identify key pathways associated with impaired pathogen clearance and chronic inflammation.

Methods

Study design and participants

This study was conducted in accordance with the principles outlined in the Declaration of Helsinki. Ethical approval was obtained from the Institutional Review Board of Korea University Guro Hospital (approval no. 2017GR0012). Written informed consent was obtained from all participants before

enrollment, following a detailed explanation of the study's objectives, procedures, potential risks, and anticipated benefits. Participants were informed of their right to withdraw from the study at any time without consequences for their current or future care. Participant confidentiality and anonymity were maintained throughout the study. All patients were recruited by the Department of Pulmonary, Allergy and Critical Care Medicine at Korea University Guro Hospital to undergo fiberoptic bronchoscopy and BALF collection for diagnostic purposes. BALF was collected according to standard clinical protocols. The procedure involved inserting a bronchoscope into the target bronchial segment, instilling 30–50 mL of sterile saline, and aspirating the fluid using low negative pressure to minimize mucosal trauma. The procedure was performed under sedation, and the recovered fluid was collected in sterile containers with gentle suction while continuously monitoring for potential complications. Patients with a peripheral arterial oxygen saturation below 90% on room air were excluded from the study. Additional exclusion criteria included the presence of a tracheostomy, dependence on home oxygen therapy, use of mechanical or noninvasive ventilation, nasal or nasopharyngeal disease, inability to provide clear verbal communication, and pregnancy. scRNA-seq was performed on 13 BALF samples.

External control datasets and comparison design

BALF specimens from the GEO datasets GSE197440 (n = 4) and GSE145926 (n = 3) served as external controls. Unless otherwise specified, all analyses were based on the comparison between Ctrl and NTM groups. This design was consistently applied across single-cell landscape characterization, cell type annotation and visualization, and differential expression analyses. In addition, targeted pathway enrichment analyses were conducted within macrophage-lineage compartments—ALV and MMM—using curated NTM-associated gene sets.

Sample processing and quality control for library preparation

BALF samples were processed immediately after collection. Following filtration to remove mucus and cellular debris, cells were pelleted and washed with phosphate-buffered saline containing 0.04% bovine serum albumin. Red blood cells, when present, were removed by lysis. The resulting cell suspensions were adjusted to a target concentration of approximately 2×10^6 to 1×10^7 cells/mL, and cell viability was assessed using acridine orange/propidium iodide (AO/PI) staining with an automated cell counter. Only samples demonstrating adequate viability and sufficient cell yield were used for downstream analyses. Single-cell libraries were prepared according to the BD Rhapsody Whole Transcriptome Analysis (WTA) and sample tag protocols. Cell suspensions were labeled with antibody-derived sample tags, filtered, and multiplexed (up to five samples per run, approximately 30,000 cells per sample).

Sequencing and read processing

Single cells were captured on magnetic beads, lysed, and their mRNA was reverse-transcribed into complementary DNA. Amplification was performed using random priming, and unique molecular

identifiers (UMIs) were incorporated during polymerase chain reaction. The WTA and sample tag libraries were pooled and sequenced on an Illumina HiSeq X Ten platform with 151 paired-end cycles.

Illumina base call files (*.bcl) were converted into FASTQ format using bcl2fastq2 (v2.20.0). Reads were filtered for both quality and length, and R2 sequences were aligned to the human reference genome (GRCh38) using STAR (v2.5.2). Cell barcodes and UMIs were annotated, and artifacts were corrected using recursive substitution error correction. Putative cells were identified by second-derivative analysis, and demultiplexing was performed using BD's sample determination algorithm. High-confidence singlets were retained for gene expression matrix generation. The final outputs included BAM files, UMI-collapsed count matrices, and QC metrics.

Software and packages

All computational analyses were conducted using the R 4.5.0 statistical environment. Single-cell preprocessing, data integration, clustering, visualization, and differential expression analyses were primarily performed using the Seurat package. Reference-based cell type annotation was executed with SingleR, using reference datasets provided by the celldex resource—namely, the HPCA and Monaco Immune references. KEGG pathway enrichment analyses were conducted with clusterProfiler, with human gene annotation and identifier conversion supported by org.Hs.eg.db. Data handling, summarization, and figure generation were performed using common R data manipulation packages such as dplyr and tidyr as needed.

Input data and quality control

Gene-by-cell count matrices were imported into the single-cell analysis framework. QC filtering was applied to remove low-quality cells and technical artifacts based on total UMI counts, number of detected genes, and mitochondrial transcript fraction. Cells were retained only if they satisfied all of the following criteria: total UMI counts (nCount_RNA) > 200, detected genes (nFeature_RNA) < 2,500, and mitochondrial transcript fraction (percent.mt) < 15%. Genes detected in fewer than three cells were excluded prior to downstream analysis to reduce sparsity-associated noise.

Normalization and variance stabilization

Following QC filtering, gene expression values were normalized and variance-stabilized to minimize library size-related effects and enhance comparability across single cells. The mitochondrial transcript fraction was regressed out during normalization to mitigate confounding effects arising from differences in mitochondrial content. All downstream analyses were performed using variable genes identified during the variance stabilization step.

Dataset integration

To enable joint analysis across samples and reduce batch-dependent technical variation, datasets were integrated using the established data integration framework implemented in the Seurat workflow. This approach produced a shared low-dimensional representation suitable for cross-sample comparison and downstream analyses. The integrated dataset was subsequently used for dimensionality reduction, clustering, visualization, and additional computational analyses.

Dimensionality reduction, clustering, and visualization

Dimensionality reduction was performed using principal component analysis (PCA) on the integrated dataset. A cell–cell nearest-neighbor graph was constructed in PCA space and used for unsupervised, graph-based clustering. Analyses were conducted using the first ten principal components (PCs 1–10), with cluster granularity controlled by setting the resolution parameter to 0.5. Two-dimensional visualization was generated using UMAP based on the same PCA dimensions (PCs 1–10).

Cell type annotation

Cell type annotation was performed using a combined strategy. First, clusters were interpreted manually based on canonical marker gene expression patterns. Second, reference-based annotation was carried out using SingleR with curated reference datasets from the celldex package, including the HPCA and Monaco Immune references. Final cell type assignments were refined by integrating reference-based predictions with marker gene expression patterns and cluster relationships within the integrated UMAP embedding.

Marker gene identification and visualization

Cluster-specific marker genes were identified by comparing each cluster against all other cells using a standard differential expression framework. Candidate markers were ranked based on average log2 fold change and statistical significance, and representative genes were selected for downstream visualization and interpretation. Marker gene expression patterns were visualized using UMAP feature plots and dot plots to assess both cluster specificity and relative expression levels. Dot plots summarize the proportion of cells expressing each gene alongside the average expression level within each cluster or cell type.

To provide an integrated overview of cluster-level transcriptional programs, heatmaps were generated using the top-ranked marker genes for each cluster. Specifically, the highest-ranking marker genes per cluster, ordered by average log2 fold change, were visualized as scaled heatmaps to enable direct comparison of relative expression patterns across clusters. This approach facilitated the identification of distinct transcriptional signatures and supported the interpretation of cellular heterogeneity within the integrated single-cell dataset.

Differential expression analysis (Ctrl vs. NTM)

Differential expression analyses were performed between NTM and Ctrl groups to identify NTM-associated transcriptional alterations. Genes were tested within a standard single-cell differential

expression framework using inclusion thresholds of min.pct = 0.25 and log2FC threshold = 0.25. To define robust NTM-associated gene sets and select representative genes for visualization and interpretation, additional filtering criteria were applied: adjusted p-value < 0.05 and |log2FC| > 0.5. Multiple hypothesis testing was controlled using adjusted p-values derived from false discovery rate (FDR) correction.

KEGG pathway enrichment analysis (ALV and MMM)

To interpret pathway-level transcriptional programs within macrophage-lineage compartments, KEGG ORA was performed separately for ALV and MMM using curated NTM-associated gene sets for each compartment. Gene identifiers were converted into KEGG-compatible identifiers using org.Hs.eg.db, and enrichment analysis was conducted with clusterProfiler. Multiple hypothesis testing was controlled using adjusted p-values derived from FDR correction. Results were visualized as KEGG dot plots, where dot size represents the number of genes mapped to each pathway, and dot color indicates adjusted statistical significance.

In vitro validation models and statistical analysis

In vitro validation was performed using A549 human alveolar type II-like epithelial cells, THP-1 monocytes, and THP-1-derived M1 macrophages exposed to NTM. mRNA expression levels of selected genes were measured in NTM-infected cells to validate transcriptional signatures observed in patient-derived BALF scRNA-seq data. Relative expression levels were normalized to GAPDH and are presented as mean ± standard error of the mean from three to five independent experiments. Comparisons against the control condition were performed using one-way ANOVA with Dunnett's post hoc test, with $p < 0.05$ considered statistically significant.

Statistical considerations and reproducibility

Multiple testing correction was applied where appropriate, including genome-wide differential expression and pathway enrichment analyses. Filtering thresholds and key analysis parameters—such as QC thresholds, PCA dimensions, and clustering resolution—were maintained consistently across all samples to ensure methodological comparability and analytical reproducibility.

Results

Single-cell RNA sequencing landscape

A total of 13 BALF specimens from individuals with NTM pulmonary disease were analyzed, while BALF specimens from the GEO datasets GSE197440 (n = 4) and GSE145926 (n = 3) served as controls. Following quality control and normalization, 47,167 cells were subjected to dimensionality reduction using the Seurat package. After filtering, 10,804 NTM-derived cells and 36,363 control cells were retained for downstream analyses. Violin plots for all 20 individual specimens (13 NTM and 7 controls) are presented in Fig. S1.

Uniform Manifold Approximation and Projection (UMAP) was applied to summarize the global transcriptomic structure of BALF-derived single cells obtained from 13 NTM and 7 control specimens (Fig. 1a–b). When cells were colored according to sample origin (Fig. 1a), the embedding revealed that cells from multiple donors occupied overlapping regions of the UMAP space, with donor-specific contributions distributed across the principal manifolds rather than forming distinct, sample-exclusive clusters. This pattern is consistent with effective integration of multi-sample data while preserving interindividual biological variability. When the same embedding was colored by experimental group (Fig. 1b), NTM and control cells exhibited considerable overlap across the primary UMAP structure, indicating that both groups share several broad transcriptional states. However, differences in point density between groups across specific regions of the embedding suggest that NTM disease is associated with altered abundance of particular cellular states rather than a uniform shift in the entire cell population. Subsequently, unsupervised clustering of the integrated dataset identified 16 transcriptionally distinct clusters (labeled 0–15; Fig. 1c). To determine whether cluster representation varied by condition, we visualized the UMAP separately for control and NTM samples while maintaining the same cluster color scheme (Fig. 1d–e). This stratified view demonstrates that certain clusters are more prominently represented in one condition than the other, whereas many clusters are present in both groups.

Quality control of single-cell RNA sequencing data

Quality control (QC) metrics were evaluated to assess both cell-level and feature-level characteristics across samples. The total number of high-quality cells retained after filtering (nCount_RNA > 200, nFeature_RNA < 2,500, and percent.mt < 15) differed between experimental groups, with a greater number of cells retained from the Ctrl group compared with the NTM group (Fig. S2a). Density distributions of total RNA counts (nCount_RNA) and detected genes per cell (nFeature_RNA) exhibited comparable overall profiles between groups, with threshold cutoffs for cell retention indicated by vertical reference lines (Figs. S2b–c). Scatterplot analysis of nCount_RNA versus nFeature_RNA demonstrated a strong positive correlation, with mitochondrial gene percentages (percent.mt) color-coded to identify low-quality cells (Fig. S2d). Furthermore, identification of the most variable genes across all cells revealed key transcriptional features contributing to subsequent clustering and differential expression analyses (Fig. S2e). Collectively, these results confirm that the dataset maintained high-quality, biologically coherent single-cell transcriptomes suitable for downstream integration and clustering analyses.

Distinct spatial expression patterns of epithelial and myeloid marker genes in BALF single cells

To further characterize cell type-specific transcriptional patterns in NTM pulmonary disease, we visualized the spatial distribution of representative epithelial and innate immune markers using UMAP-based FeaturePlot analysis (Fig. S3). Alveolar epithelial cell markers—including AQP5, KRT7, MUC1, and LAMP3—showed localized expression within epithelial-enriched regions of the embedding. Ion transport-related genes (ATP1A1 and ATP1B1) exhibited comparatively lower expression intensity in subsets of

these epithelial-associated areas, consistent with the altered epithelial-state signatures identified in subsequent differential expression analyses.

Innate immune markers displayed organized expression patterns across myeloid-dominant regions. Monocyte-associated markers (CCR2 and SELL) were enriched within monocyte-like regions, whereas LYZ, CX3CR1, and FCGR3A demonstrated broader but heterogeneous expression across the myeloid compartment.

Macrophage-associated antigen presentation and phagocytic regulators—including CD74, MERTK, ITGAX, FCGR2B, CD36, CD14, and CD163—exhibited region-specific variation within macrophage-enriched areas, with several markers displaying reduced expression intensity in distinct macrophage-associated regions. Notably, apparent group-related patterns observed in FeaturePlots should be interpreted alongside group-stratified visualizations and differential expression analyses, as they may reflect differences in cell density rather than true transcriptional divergence.

Annotation of single-cell transcriptomes by HPCA and Monaco references

To refine cell type identification across BALF samples, we performed reference-based annotation using two widely utilized databases: the Human Primary Cell Atlas (HPCA; Fig. 2a) and the Monaco Immune Reference (Fig. 2b). UMAP visualization of HPCA-derived labels revealed extensive transcriptional heterogeneity, encompassing epithelial, fibroblast, endothelial, and multiple immune cell populations. Innate immune subsets, such as macrophages and neutrophils, occupied discrete regions on the UMAP, whereas lymphoid populations—including T, NK, and B cells—were distributed across distinct yet partially overlapping clusters, reflecting the complex cellular composition of the BALF microenvironment.

Annotation using the Monaco reference provided an immune-focused classification framework and allowed finer resolution among lymphocyte subsets, clearly distinguishing CD4 + T cells, CD8 + T cells, NK cells, and B cells (Fig. 2b). Myeloid populations, including monocytes, dendritic cells, and neutrophils, also exhibited more clearly defined boundaries under Monaco-based labeling. Overall, the two annotation strategies were complementary: HPCA captured broader stromal and epithelial cell types, whereas Monaco provided improved granularity in characterizing immune cell heterogeneity. Collectively, these analyses highlight the diverse mixture of epithelial, stromal, and innate/adaptive immune subsets within BALF, establishing a reliable reference framework for downstream cluster interpretation.

Integrated scRNA-seq analysis and cluster-specific transcriptional profiles

Following QC filtering, scRNA-seq data from three independent BALF datasets were integrated using SCTransform-based normalization and anchor-based integration to minimize batch effects. Dimensionality reduction and unsupervised clustering identified transcriptionally distinct cell clusters within the integrated dataset.

To define the molecular characteristics of each cluster, differentially expressed genes were identified through cluster-specific marker analysis. The top 10 marker genes per cluster, ranked by average log2 fold change, were visualized in a heatmap (Fig. 3). This analysis revealed distinct transcriptional signatures across clusters, supporting the identification of transcriptionally defined cell populations in the integrated dataset. Notably, the expression patterns of several immune- and epithelial-associated marker genes exhibited heterogeneous distribution across clusters, indicating functional diversity within the BALF cellular compartment. When considered alongside group-stratified visualizations in subsequent figures, these cluster-specific transcriptional programs correspond to condition-associated variations in the abundance and function of specific BALF cell populations.

Condition-associated differential expression of key marker genes between Control and NTM

Following the identification of cluster-specific transcriptional signatures in the integrated single-cell RNA sequencing analysis, we examined whether representative marker genes differed between disease conditions (Fig. 4). Gene-level comparisons between Control and NTM samples focused on markers associated with epithelial function and myeloid or macrophage identity. Markers linked to epithelial differentiation and activation—KRT7, MUC1, LAMP3, AQP5, and ABCA3—as well as inflammatory or immune-associated genes—SELL, CCR2, TLR2, and FCGR3A—showed significantly higher expression in NTM samples compared with controls. These genes are involved in epithelial remodeling and stress responses, antigen presentation-related pathways, and inflammatory or chemokine signaling, consistent with enhanced epithelial-associated transcriptional activity in NTM disease. Conversely, genes characteristic of differentiated or homeostatic myeloid and macrophage populations—including LYZ, CD74, ATP1B1, FCGR2B, MERTK, CD36, ITGAX, CD14, SIGLEC1, CD163, ATP1A1, and CX3CR1—exhibited significantly lower expression in NTM samples. This reduction suggests attenuation and phenotypic alteration of resident macrophage-associated transcriptional programs under NTM conditions.

To contextualize these transcriptional shifts within specific cell clusters, we summarized canonical lineage and functional marker expression using dot plots across the 16 transcriptional clusters (0–15) (Fig. S4). In this cluster-resolved representation, epithelial markers—KRT7 and MUC1—were predominantly confined to epithelial-like clusters, whereas macrophage and myeloid markers—LYZ, CD14, CD163, and related genes including ITGAX, CX3CR1, SIGLEC1, FCGR2B, MERTK, and CD36—were enriched in distinct myeloid-dominant clusters. Antigen presentation-associated markers, including CD74 and LAMP3, exhibited cluster-restricted enrichment consistent with antigen-presenting cell (APC)-like states. Collectively, these findings indicate that the disease-associated differential expression patterns observed in Fig. 4 correspond to lineage-specific alterations in epithelial and myeloid transcriptional programs within the bronchoalveolar compartment.

KEGG pathway enrichment highlights distinct NTM-associated programs in ALV and MMM

To functionally interpret NTM-associated transcriptional signatures within key innate immune compartments, we performed Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway over-representation analysis (ORA) separately for alveolar macrophages (ALV) and monocyte-derived macrophages (MMM), and summarized the results as dot plots (Fig. 5). The enriched pathways revealed that NTM disease is associated with pathway-level remodeling in macrophage-lineage populations, encompassing inflammatory and immune signaling, host-response processes, and metabolic as well as cell death-related programs.

In ALV, KEGG enrichment prominently included pathways related to inflammatory signaling and host defense—such as interleukin-17 (IL-17) signaling—as well as pathways associated with broader cellular remodeling and stress responses (Fig. 5a).

In MMM, KEGG enrichment similarly captured immune, inflammatory, and host-response pathways—including pattern-recognition receptor signaling and phagosome-related processes—alongside pathways involved in lipid metabolism and immune-mediated tissue responses (Fig. 5b). Collectively, the ALV and MMM enrichment profiles support macrophage-lineage reprogramming in NTM pulmonary disease across multiple biological pathways identified at the systems level.

In vitro validation of cell type-specific transcriptional signatures

To validate the transcriptomic alterations identified in patient-derived bronchoalveolar lavage samples, we established in vitro infection models using A549 (human alveolar type II-like) epithelial cells, THP-1 monocytes, and THP-1-derived M1 macrophages exposed to NTM (Fig. 6a–c). These models were used to determine whether the observed gene expression patterns were directly induced by NTM infection.

For the alveolar type II-like epithelial response, NTM-infected A549 cells exhibited significant, time-dependent upregulation of ABCA3 and MUC1 mRNA expression (Fig. 6a). Notably, ABCA3 expression increased progressively, reaching peak significance at 96 hours post-infection ($p < 0.001$), supporting its role as a marker of infection-induced epithelial stress and surfactant metabolic dysregulation.

For monocyte activation and pathogen sensing, NTM infection of THP-1 monocytes resulted in a pronounced elevation of SELL and TLR2 expression (Fig. 6b). These findings closely mirrored the heightened pathogen-sensing and recruitment signatures observed in patient BALF samples. However, the downregulation of LYZ and CX3CR1 detected in in vivo scRNA-seq data was not reproduced in this in vitro setting, suggesting that additional microenvironmental factors and/or prolonged exposure may be required to recapitulate these transcriptomic features.

In THP-1-derived M1 macrophages, NTM infection induced a rapid and sustained decrease in CD74 mRNA levels, with significant downregulation observed as early as 24 hours post-infection ($p < 0.01$) (Fig. 6c). This marked reduction aligns with the impaired antigen presentation-associated signatures identified in the clinical dataset and highlights a key mechanism of NTM-mediated immune evasion.

Overall, the strong concordance between the clinical scRNA-seq data and the in vitro models supports a direct contribution of NTM exposure to these cell type-specific transcriptional alterations.

Discussion

Our study provides a comprehensive single-cell transcriptomic analysis of BALF from patients with NTM pulmonary disease, offering an in-depth view of the cellular and molecular mechanisms underlying this increasingly prevalent and heterogeneous condition (1, 2). By integrating in vivo scRNA-seq of patient-derived BALF with in vitro NTM infection models, we identified cell type-specific transcriptional programs and distinguished pathogen-driven immune alterations from generalized inflammatory responses (6, 7).

Alveolar type II (AT2) epithelial cells from NTM pulmonary disease samples exhibited pronounced upregulation of epithelial stress and surfactant-associated genes, including KRT7, MUC1, LAMP3, and ABCA3. These transcriptional changes are consistent with previous studies showing that AT2 cells function not only as progenitors for alveolar repair but also as active modulators of immune signaling during chronic lung injury and infection (10–12). The observed downregulation of ATP1A1 and ATP1B1—key regulators of epithelial ion transport and alveolar fluid clearance (13)—suggests disruption of epithelial homeostasis, which may further impair mucosal defense mechanisms in NTM disease.

Within the monocyte compartment, we observed increased expression of CCR2, SELL, and TLR2, indicating enhanced recruitment to the lung and heightened pathogen-sensing capacity (14, 15). However, the concurrent downregulation of LYZ and CX3CR1—genes critical for bactericidal activity and tissue surveillance—suggests functional polarization toward a less effective or potentially exhausted phenotype (16). This dissociation between immune cell recruitment and effector function may represent a mechanism by which NTM persists despite sustained innate immune activation.

Macrophages, which play a central role in mycobacterial containment and clearance, exhibited a transcriptional profile indicative of impaired antimicrobial and immunoregulatory capacity. Downregulation of CD74, SIGLEC1, MERTK, CD163, and FCGR2B suggests compromised antigen presentation, phagocytic function, and Fc receptor-mediated immune signaling (17–19). These observations are consistent with previous studies demonstrating macrophage dysfunction, impaired phagolysosomal maturation, and altered lipid metabolism in chronic mycobacterial infections (20, 21).

Pathway-level interpretation through KEGG enrichment analysis provided additional biological context for these transcriptional alterations (22–24). Enrichment of pathways related to chemokine signaling and phagosome formation supports persistent innate immune activation, while alterations in lipid metabolism-associated pathways align with impaired macrophage handling of intracellular mycobacteria, which utilize host lipid resources for survival (20). Importantly, the KEGG results reinforced cell type-specific transcriptional findings rather than introducing unrelated biological interpretations.

Inflammatory signaling pathways, including IL-17-associated responses, were modestly enriched, suggesting a maladaptive inflammatory milieu in chronic NTM disease. While IL-17-mediated signaling may contribute to early host defense, sustained activation has been linked to neutrophil-driven inflammation, airway damage, and fibrosis—features commonly observed in advanced stages of NTM pulmonary disease (3).

A major strength of this study lies in the validation of in vivo transcriptional signatures using in vitro NTM infection models. Core epithelial stress markers (MUC1, ABCA3), monocyte sensing and recruitment genes (SELL, TLR2), and macrophage antigen presentation molecules (CD74) were consistently upregulated following direct NTM exposure (25, 26). However, the absence of monocyte exhaustion features in vitro suggests that chronic environmental factors—such as prolonged cytokine stimulation, repeated antigen exposure, and complex multicellular interactions—may be required to fully reproduce the immune dysregulation observed in patients (8, 27).

Our findings align with previous studies reporting impaired phagocytosis, altered lipid metabolism, and dysregulated innate immune signaling in NTM pulmonary disease (3, 20, 21). By integrating single-cell resolution with pathway-level interpretation, this study highlights convergent immune dysfunction across epithelial and myeloid compartments. The shared suppression of antigen presentation and phagocytic pathways suggests potential targets for host-directed therapeutic strategies aimed at restoring immune competence (9, 28).

Several limitations should be acknowledged. The relatively small number of BALF samples may limit the generalizability of the findings, and the in vitro infection models cannot fully recapitulate the complexity of the lung microenvironment, including extracellular matrix interactions and long-term immune conditioning (4, 8). Furthermore, although the transcriptional alterations identified were robust, functional validation at the protein and cellular activity levels remains necessary.

Conclusions

This integrated single-cell and pathway-level analysis reveals coordinated immune dysregulation across epithelial and myeloid compartments in NTM pulmonary disease. The convergence of impaired phagocytosis, altered lipid metabolism, and chronic inflammatory signaling provides a mechanistic framework for understanding disease persistence and heterogeneity and offers a rational basis for the development of novel diagnostic biomarkers and host-directed therapeutic strategies.

Abbreviations

ALV
alveolar macrophages
APC
antigen-presenting cell
AT2

alveolar type II

BAL

bronchoalveolar lavage

BALF

bronchoalveolar lavage fluid

Ctrl

control

FDR

false discovery rate

GEO

Gene Expression Omnibus

HPCA

Human Primary Cell Atlas

IL-17

interleukin-17

KEGG

Kyoto Encyclopedia of Genes and Genomes

MMM

monocyte-derived macrophages

NTM

nontuberculous mycobacterial

ORA

over-representation analysis

PCA

principal component analysis

QC

quality control

scRNA-seq

single-cell RNA sequencing

SEM

standard error of the mean

UMAP

Uniform Manifold Approximation and Projection

UMI

unique molecular identifier

WTA

Whole Transcriptome Analysis

Declarations

Ethics approval and consent to participate

This study was approved by the Institutional Review Board of Korea University Guro Hospital (approval no. 2017GR0012) and was conducted in accordance with the principles of the Declaration of Helsinki. Written informed consent was obtained from all subjects involved in the study.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Funding

This research was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean Government (Ministry of Science and ICT, MSIT) (grant no. 2022R1C1C1009250); the Global Learning & Academic Research Institution for Master's, PhD Students, and Postdoctoral Researchers (G-LAMP) Program of the NRF grant funded by the Ministry of Education (grant no. RS-2025-25441283); and Korea University Guro Hospital (Korea Research-Driven Hospital) (grant no. O2513591). The funding agencies had no role in the design of the study; in the collection, analysis, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Author Contribution

Conceptualization, J.Y.O., Y.K.K., and J.-A.G.; methodology, J.Y.O. and J.-A.G.; formal analysis, J.Y.O., Y.K.K., M.H.J., and H.M.D.; investigation, Y.K.K., M.H.J., S.-Y.K., and H.M.D.; resources, J.Y.O., Y.K.K., and M.H.J.; data curation, J.Y.O. and Y.K.K.; writing—original draft preparation, J.Y.O., Y.K.K., S.-Y.K., and J.-A.G.; writing—review and editing, M.H.J., H.M.D., and S.R.; visualization, S.R., S.-Y.K., and J.-A.G.; supervision, J.Y.O.; project administration, J.-A.G.; funding acquisition, J.Y.O. All authors have read and approved the final manuscript.

Acknowledgements

Not applicable.

Data Availability

The datasets generated and analyzed during the current study are available through the Gene Expression Omnibus (GEO) under accession number GSE278953. Correspondence and requests for data access should be directed to the corresponding author.

References

1. Falkinham JO. Environmental sources of nontuberculous mycobacteria. Clin Chest Med. 2015;36:35–41.
2. Daley CL, Iaccarino JM, Lange C, Cambau E, Wallace RJ, Andrejak C, et al. Treatment of nontuberculous mycobacterial pulmonary disease: an official ATS/ERS/ESCMID/IDSA clinical practice guideline. Clin Infect Dis. 2020;71:e1–36.
3. Cowman S, van Ingen J, Griffith DE, Loebinger MR. Nontuberculous mycobacterial pulmonary disease. Eur Respir J. 2019;54:1900250.
4. Honda JR, Knight V, Chan ED. Pathogenesis and risk factors for nontuberculous mycobacterial lung disease. Clin Chest Med. 2015;36:1–11.
5. 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.
6. Papalexi E, Satija R. Single-cell RNA sequencing to explore immune cell heterogeneity. Nat Rev Immunol. 2018;18:35–45.
7. Keren L, Bosse M, Marquez D, Angoshtari R, Jain S, Varma S, et al. A structured tumor-immune microenvironment in triple negative breast cancer revealed by multiplexed ion beam imaging. Cell. 2018;174:1373–e138719.
8. Hussell T, Bell TJ. Alveolar macrophages: plasticity in a tissue-specific context. Nat Rev Immunol. 2014;14:81–93.
9. Zumla A, Rao M, Wallis RS, Kaufmann SHE, Rustomjee R, Mwaba P, et al. Host-directed therapies for infectious diseases: current status, recent progress, and future prospects. Lancet Infect Dis. 2016;16:e47–63.
10. Barkauskas CE, Cronce MJ, Rackley CR, Bowie EJ, Keene DR, Stripp BR, et al. Type 2 alveolar cells are stem cells in adult lung. J Clin Invest. 2013;123:3025–36.
11. Zacharias WJ, Frank DB, Zepp JA, Morley MP, Alkhaleel FA, Kong J, et al. Regeneration of the lung alveolus by an evolutionarily conserved epithelial progenitor. Nature. 2018;555:251–5.
12. Travaglini KJ, Nabhan AN, Penland L, Sinha R, Gillich A, Sit RV, et al. A molecular cell atlas of the human lung from single-cell RNA sequencing. Nature. 2020;587:619–25.
13. Ridge KM, Olivera WG, Saldías F, et al. Alveolar type I cells express the α2 Na,K-ATPase, which contributes to lung liquid clearance. Am J Physiol Lung Cell Mol Physiol. 2003;284:L685–95.
14. Shi C, Pamer EG. Monocyte recruitment during infection and inflammation. Nat Rev Immunol. 2011;11:762–74.

15. Underhill DM, Goodridge HS. Information processing during phagocytosis. Nat Rev Immunol. 2012;12:492–502.
16. Guilliams M, Mildner A, Yona S. Developmental and functional heterogeneity of monocytes. Immunity. 2018;49:595–613.
17. Tailleux L, Schwartz O, Herrmann JL, Pivert E, Jackson M, Amara A, et al. DC-SIGN is the major Mycobacterium tuberculosis receptor on human dendritic cells. J Exp Med. 2003;197:121–7.
18. Russell DG, Cardona PJ, Kim MJ, Allain S, Altare F. Foamy macrophages and the progression of the human tuberculosis granuloma. Nat Immunol. 2009;10:943–8.
19. Schlesinger LS. Macrophage phagocytosis of virulent but not attenuated strains of Mycobacterium tuberculosis is mediated by mannose receptors in addition to complement receptors. J Leukoc Biol. 1993;54:559–65.
20. Flynn JL, Chan J. Immunology of tuberculosis. Annu Rev Immunol. 2001;19:93–129.
21. Roux AL, Viljoen A, Bah A, Simeone R, Bernut A, Laencina L, et al. The distinct fate of smooth and rough Mycobacterium abscessus variants inside macrophages. Open Biol. 2016;6:160185.
22. Kanehisa M, Goto S. KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000;28:27–30.
23. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. Omics. 2012;16:284–7.
24. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. 2005;102:15545–50.
25. Cohen SB, Gern BH, Delahaye JL, Adams KN, Plumlee CR, Winkler JK, et al. Alveolar macrophages provide an early Mycobacterium tuberculosis niche and initiate dissemination. Cell Host Microbe. 2018;24:439–e4464.
26. Cadena AM, Fortune SM, Flynn JL. Heterogeneity in tuberculosis. Nat Rev Immunol. 2017;17:691–702.
27. Mayer-Barber KD, Barber DL. Innate and adaptive cellular immune responses to Mycobacterium tuberculosis infection. Cold Spring Harb Perspect Med. 2015;5:a018424.
28. Wynn TA, Vannella KM. Macrophages in tissue repair, regeneration, and fibrosis. Immunity. 2016;44:450–62.

Figures

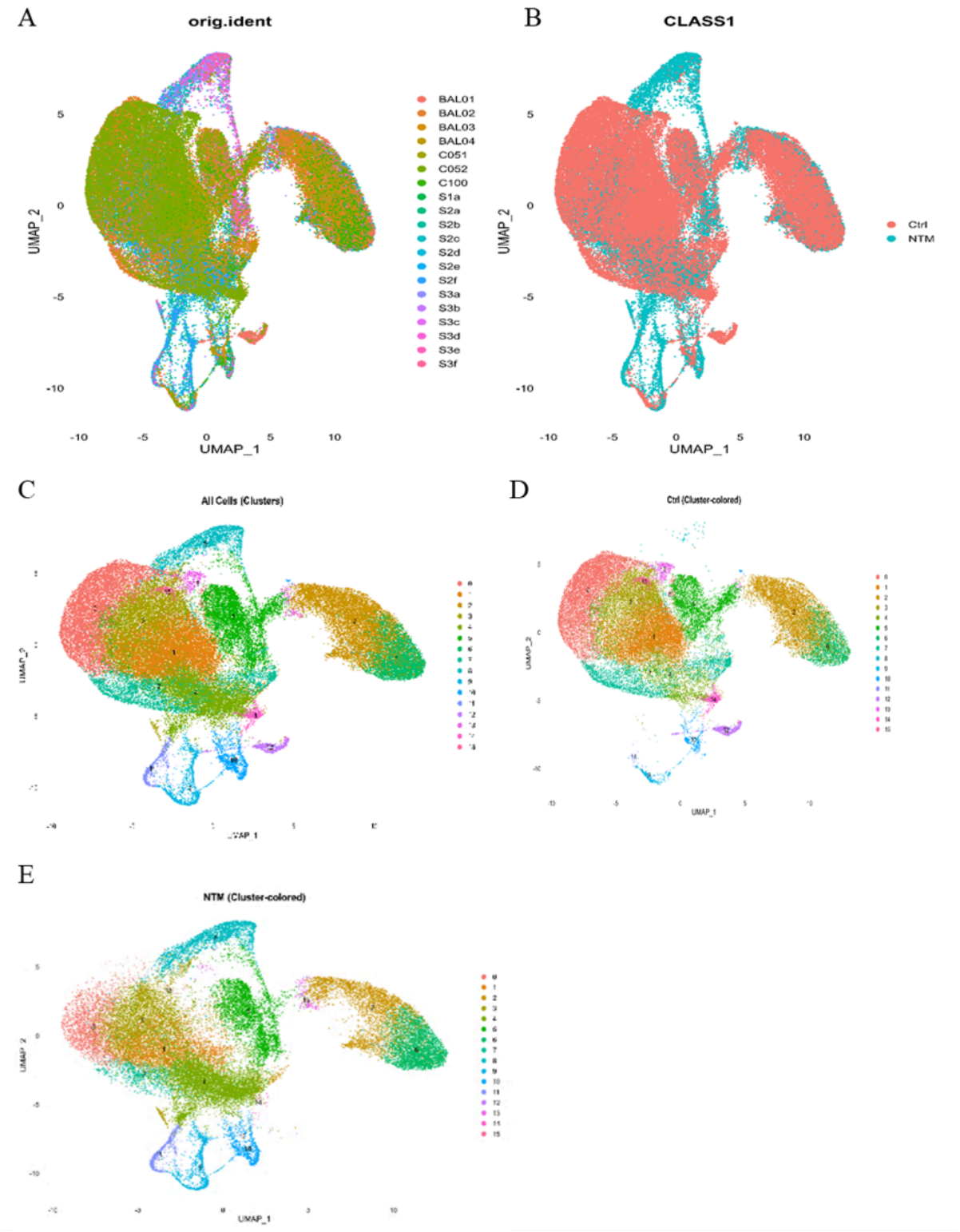


Figure 1

UMAP visualization and clustering patterns of BALF-derived single cells.

(A) Cells are colored according to their sample origin (orig.ident). (B) Cells are colored by experimental group (NTM vs. Ctrl; CLASS1). (C) UMAP plot showing the distribution of integrated BALF single cells, colored by cluster identity. (D–E) UMAP plots separated by clinical group (Ctrl and NTM) and colored by

the same cluster identities. Each point represents a single cell (n = 47,167). Visualization was generated using the Seurat DimPlot function with UMAP reduction (pt.size = 0.1, alpha = 0.3). Distinct clustering patterns indicate inter-sample consistency and group-dependent transcriptional variation. In the group-separated UMAPs (D–E), NTM and Ctrl cells exhibit partially overlapping distributions, with differences in relative cell density across regions of the embedding—consistent with condition-associated shifts in cellular composition and/or transcriptional states.

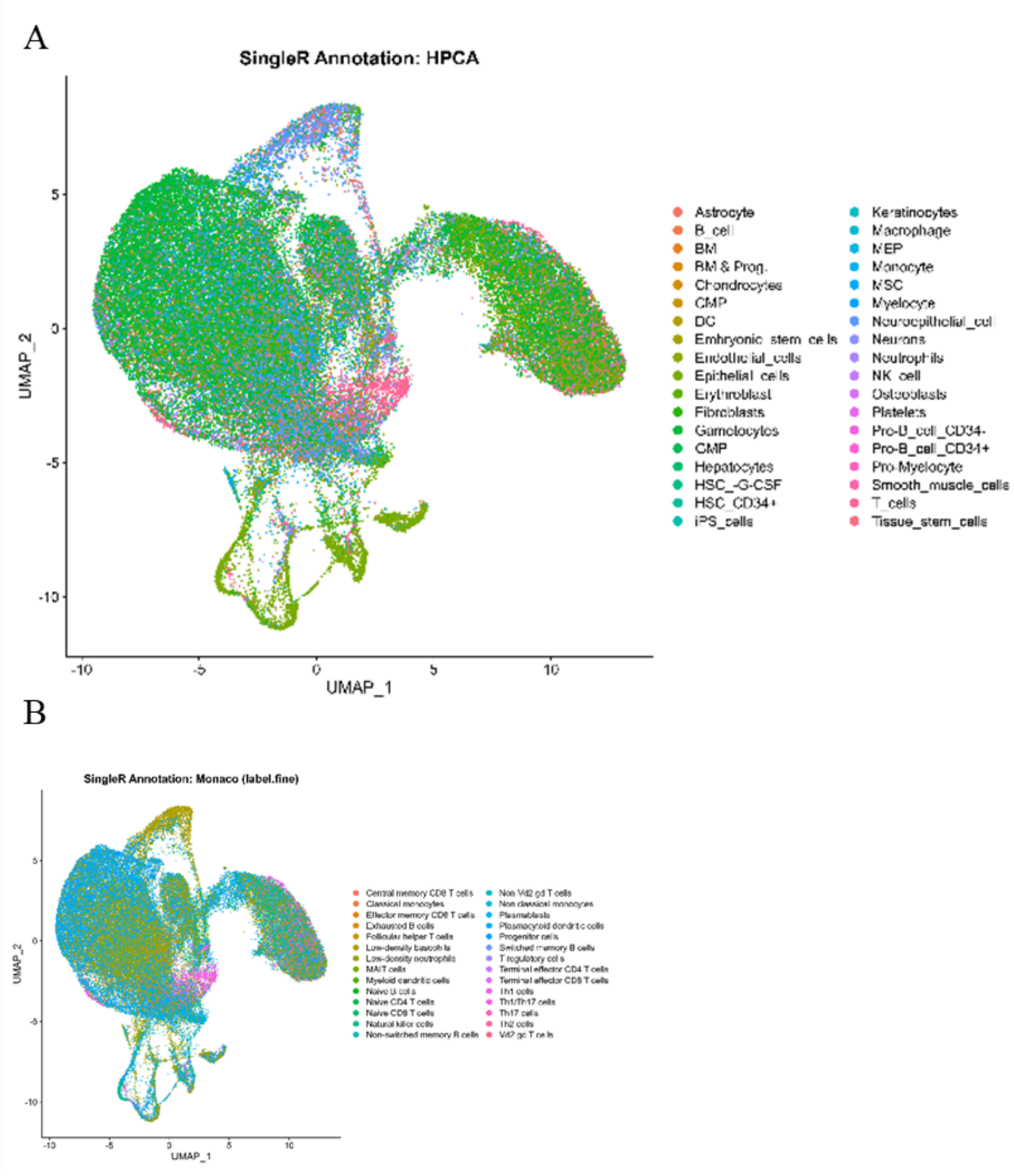


Figure 2

UMAP visualization of BALF single-cell transcriptomes annotated using HPCA and Monaco references.

(A) UMAP plot showing SingleR-based cell type annotation using the HPCA reference, providing broad classification of epithelial, stromal, and immune cell populations. (B) UMAP plot showing SingleR-based annotation using the Monaco Immune Reference (label.fine), offering immune-focused labeling with finer resolution among lymphoid and myeloid subsets. Collectively, the HPCA and Monaco annotations reveal a mixture of epithelial and stromal cells, along with diverse innate and adaptive immune subsets in BALF, establishing a reference framework for downstream cluster interpretation.

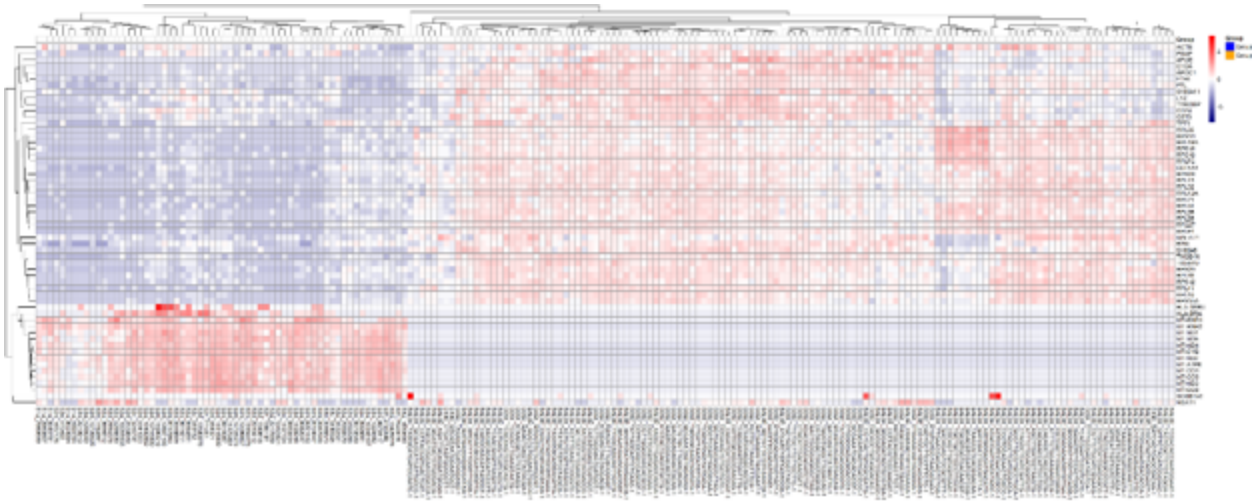


Figure 3

Heatmap of cluster-specific marker gene expression in integrated BALF scRNA-seq data.

The heatmap displays the expression profiles of the top 10 marker genes for each unsupervised cluster identified in the integrated BALF scRNA-seq dataset. Marker genes were selected based on average log2 fold change derived from cluster-specific differential expression analysis. Gene expression values were scaled across cells to emphasize relative differences in expression intensity. Columns represent individual cells ordered by cluster, and rows correspond to marker genes. The heatmap illustrates distinct transcriptional signatures across clusters and serves as a reference for interpreting condition-stratified differences in subsequent analyses.

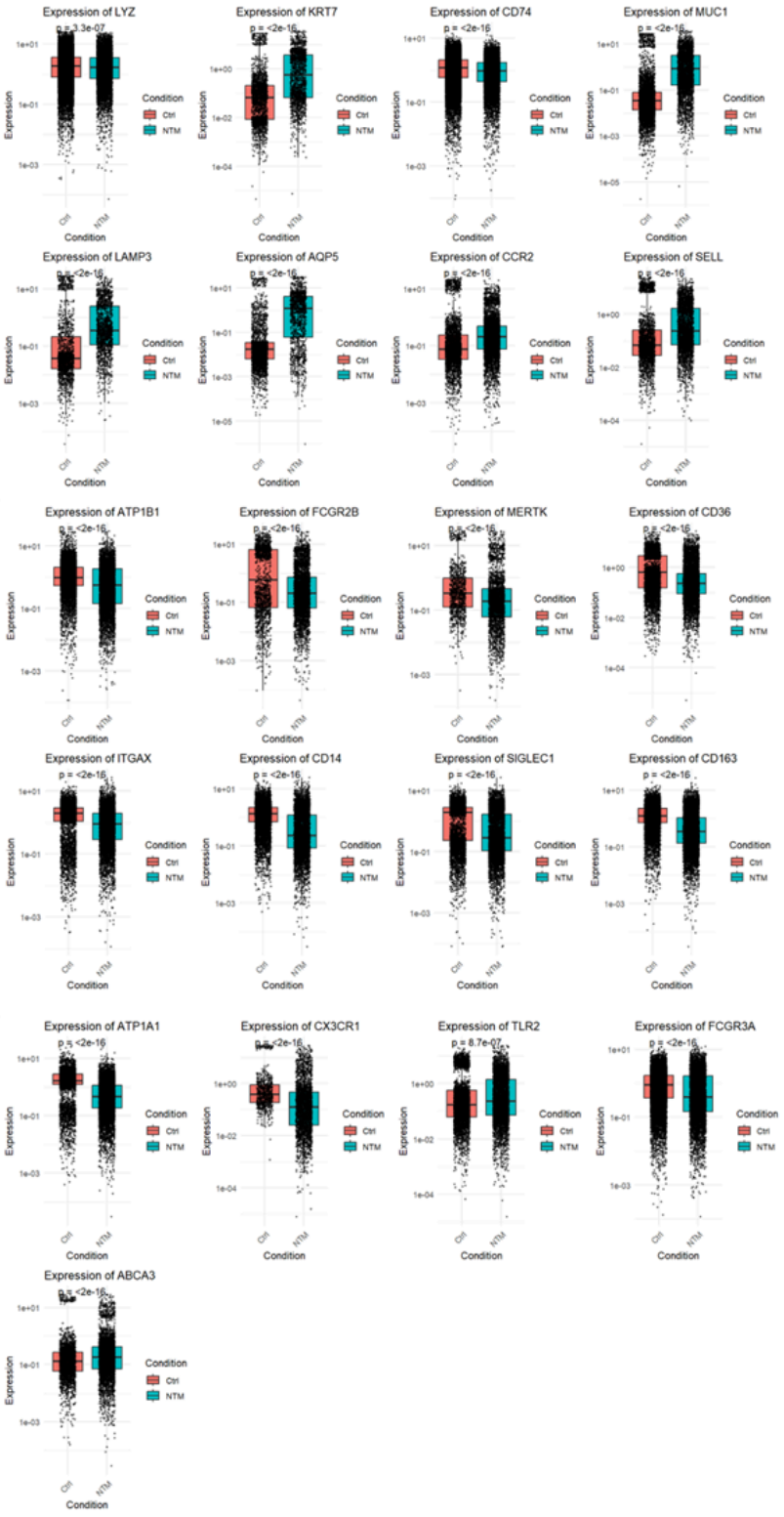


Figure 4

Differential expression of selected marker genes between Control and NTM samples.

Boxplots display single-cell expression levels of selected marker genes in Control and NTM samples. Each dot represents an individual cell, and boxes summarize the distribution within each group (Control, red; NTM, blue). Expression values are presented on a logarithmic scale. P-values for between-group

comparisons are shown for each gene. Overall, epithelial-associated genes (e.g., KRT7, MUC1, LAMP3, AQP5) show higher expression in NTM, whereas myeloid/macrophage-associated genes (e.g., LYZ, CD14, CD163) exhibit lower expression in NTM.

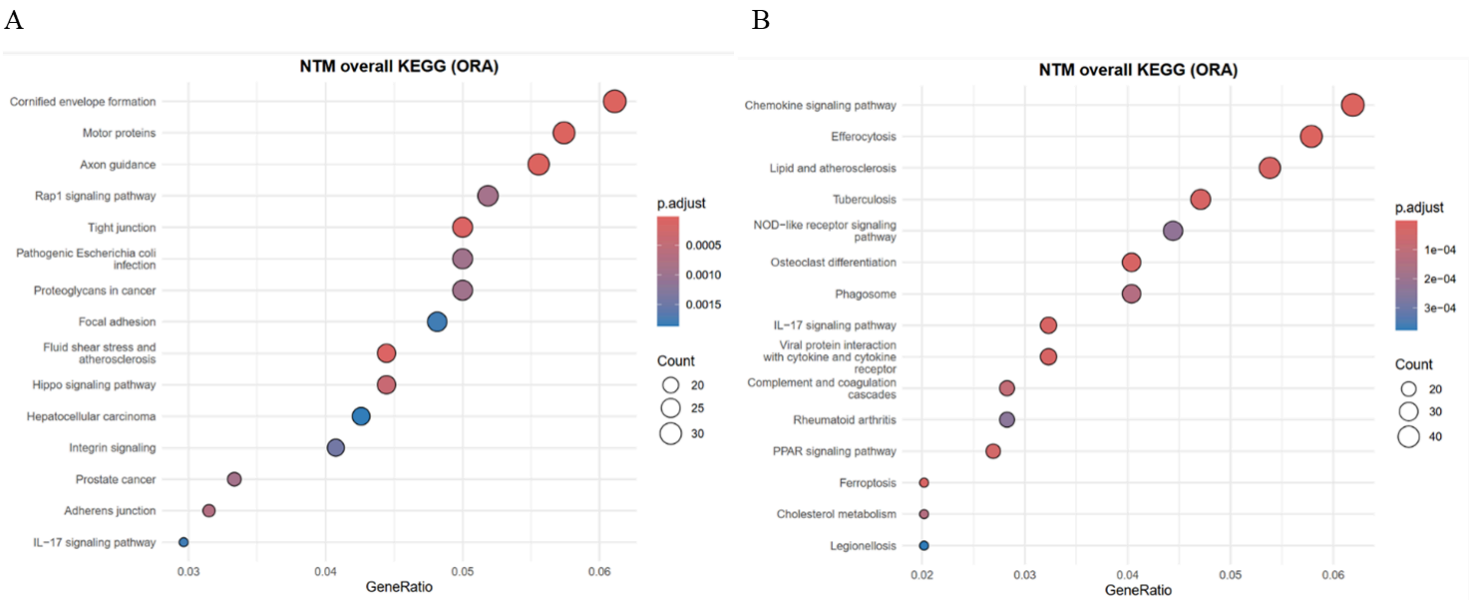


Figure 5

KEGG pathway enrichment (ORA) of NTM-associated gene signatures in macrophage-lineage populations.

KEGG ORA was conducted separately for ALV and MMM using NTM-associated gene sets defined from the Ctrl vs. NTM differential expression results within each compartment. Dot plots depict significantly enriched KEGG pathways; dot size denotes the number of genes mapped to each pathway (count), and dot color represents multiple testing-adjusted significance (adjusted p-value). (A) ALV KEGG enrichment dot plot. (B) MMM KEGG enrichment dot plot.

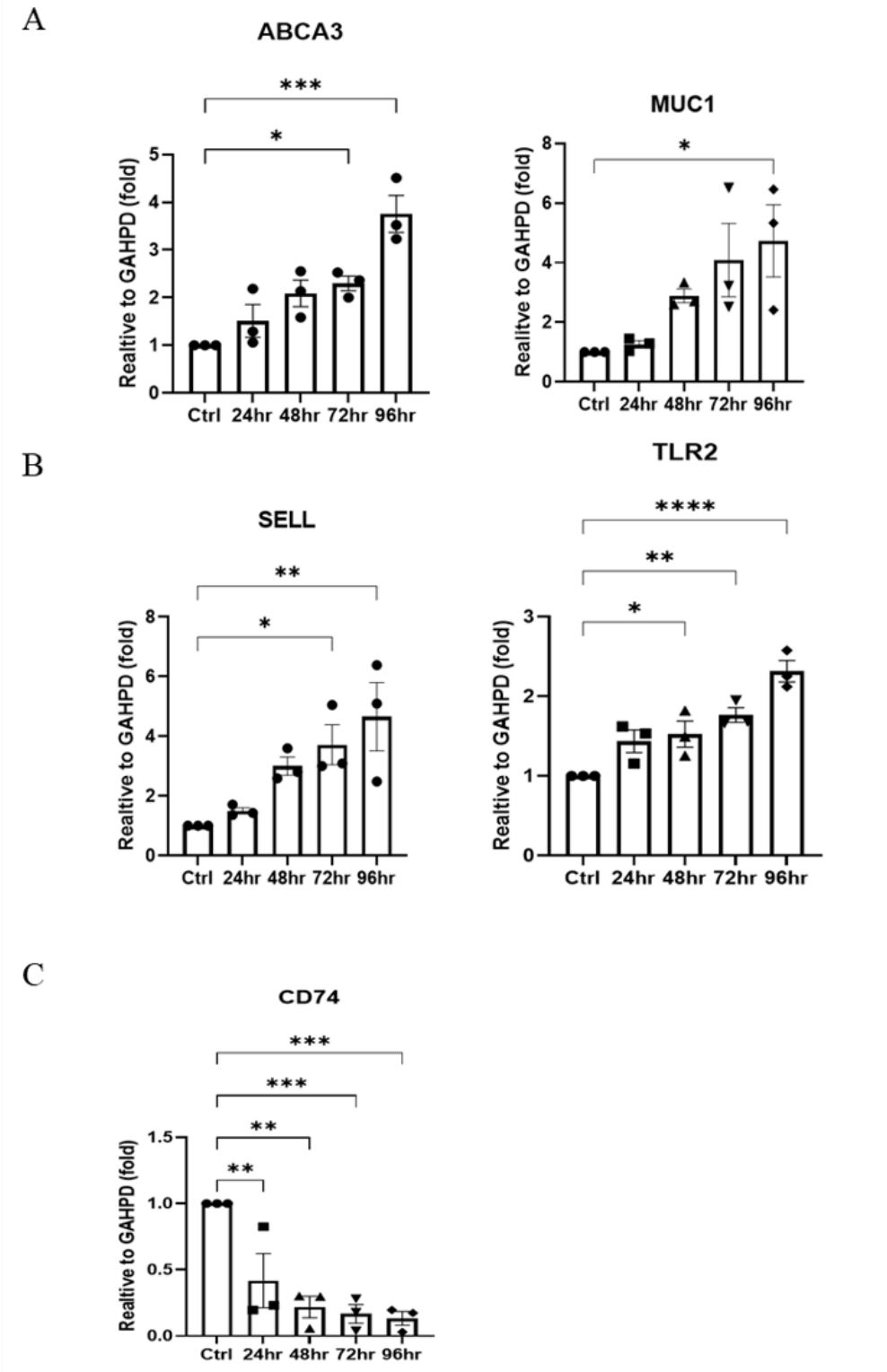


Figure 6

In vitro validation of NTM-induced transcriptional signatures in pulmonary epithelial and myeloid cell lines.

(A) mRNA expression of ABCA3 and MUC1 in human alveolar type II-like epithelial cells (A549) following NTM infection. (B) mRNA expression of SELL and TLR2 in NTM-infected human monocytes (THP-1). (C)

Time-dependent downregulation of CD74 in NTM-infected, THP-1-derived M1 macrophages. Relative expression levels were normalized to GAPDH and are presented as mean ± standard error of the mean (SEM) from three to five independent experiments. *p < 0.05, **p < 0.01, ***p < 0.001, ****p < 0.0001 versus Control (one-way ANOVA with Dunnett's post hoc test).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- FigureS1.png
- FigureS2.png
- FigureS3.png
- FigureS4.png