Scientific Research Publishing

Open Journal of Epidemiology, 2025, 15(3), 508-527
https://www.scirp.org/journal/ojepi
ISSN Online: 2165-7467
ISSN Print: 2165-7459

Bioinformatics Analyses Identify Shared Differentially Expressed Genes in Non-Tuberculosis and Tuberculosis Pulmonary Diseases

Chaoyue Liang[1*#], Hongyun Luo[2*], Ni Zhou[1], Wenmei Mu[1], Shouqiang Ma[1], Zhicheng Yang[3]

[1]Department of Critical Care Medicine, The Brain Hospital of Guangxi Zhuang Autonomous Region, Liuzhou, China
[2]Department of Respiratory Medicine, The Chest Hospital of Guangxi Zhuang Autonomous Region, Liuzhou, China
[3]Liuzhou KeyLab of Psychosis Treatment, The Brain Hospital of Guangxi Zhuang Autonomous Region, Liuzhou, China
Email: #yuellenmm@163.com, 1048564232@qq.com

How to cite this paper: Liang, C.Y., Luo, H.Y., Zhou, N., Mu, W.M., Ma, S.Q. and Yang, Z.C. (2025) Bioinformatics Analyses Identify Shared Differentially Expressed Genes in Non-Tuberculosis and Tuberculosis Pulmonary Diseases. *Open Journal of Epidemiology*, **15**, 508-527.
https://doi.org/10.4236/ojepi.2025.153032

Received: May 18, 2025
Accepted: June 21, 2025
Published: June 24, 2025

Abstract

Background: Nontuberculous mycobacteria (NTM) and Mycobacterium tuberculosis (TB) pulmonary infections share clinical features but have divergent outcomes, suggesting distinct host immune adaptations. **Methods:** We integrated transcriptomic datasets (GSE97298: 32 NTM vs. 9 controls; GSE83456: 45 TB vs. 61 controls) to identify shared and distinct molecular pathways. Differentially expressed genes (DEGs) were analyzed via limma ($|\log_2 FC| \geq 1.5$, FDR < 0.05), with functional enrichment (WebGestalt/Metascape) and PPI networks (STRING/Cytoscape). **Results:** We identified 48 shared DEGs with bidirectional regulation (e.g., LDHB: NTM ↓ vs. TB ↑; NOD2: NTM ↑ vs. TB ↓). Pathway analysis revealed neutrophil degranulation as a core-shared mechanism (FDR = 2.37×10^{-6}). ELANE and DEFA4 showed strong co-expression (Spearman *r* = 0.86, *p* < 0.001) linking to NETosis, while IRAK3 (innate immunity hub) and CD28 (adaptive node) emerged as context-dependent regulators. **Conclusion:** This study defines conserved neutrophil-driven immunopathology in mycobacterial infections and nominates IRAK3/CD28 for host-directed therapies.

Keywords

Non-Tuberculous Mycobacteria, Tuberculosis, Transcriptomics, Neutrophil Extracellular Traps, ELANE, DEFA4

*These authors contributed equally to this work.
#Corresponding author.

1. Introduction

Pulmonary infections caused by nontuberculous mycobacteria (NTM) and Mycobacterium tuberculosis (TB) represent a growing global health challenge, with overlapping clinical presentations but divergent therapeutic requirements [1]. While TB remains a leading infectious killer worldwide (10.6 million cases, 1.3 million deaths in 2022), NTM infections are increasingly recognized as emerging threats, with annual incidence rising by 8.2% among immunocompromised populations and chronic lung disease patients [2]. Despite shared anatomical tropism, these pathogens elicit distinct immune responses: TB drives robust Th1 polarization via IL-12/IFN-γ signaling, whereas NTM evades clearance through TLR2-dependent immunosuppressive cytokine production (e.g., IL-10) [3].

Current diagnostic paradigms relying on microbiological identification face critical limitations, with 34% of NTM cases misclassified as TB during initial assessment in multicenter cohorts [4]. This diagnostic ambiguity reflects incomplete understanding of their molecular signatures. While transcriptomic studies have characterized individual diseases—NTM blood signatures revealing IFNG suppression [5] and TB studies identifying NOD2 downregulation [6]—systematic comparisons of their host response landscapes remain absent. A 2023 review highlights this gap, noting that "no study has yet resolved shared molecular pathways with opposing regulatory trends across mycobacterial infections" [7].

Recent advances in bioinformatics now enable integrative analysis of public transcriptomic datasets to decode disease-specific adaptations. For instance, the RECON method developed for cross-cohort infectious disease analysis [8] has revealed conserved immune modules in viral infections. Leveraging such approaches, we hypothesized that comparative analysis of NTM and TB blood transcriptomes would reveal:

1) Conserved immune pathways underlying mycobacterial persistence

2) Bidirectional regulation of metabolic genes driving disease-specific outcomes

3) Novel neutrophil-mediated mechanisms contributing to immunopathology

4) Our study addresses three critical gaps:

The molecular basis for NTM's attenuated inflammation (e.g., TGF-β dominance) versus TB's NLRP3 inflammasome hyperactivation [9]. The role of understudied innate immune checkpoints (e.g., IRAK3, a negative TLR regulator) in chronic infection persistence [10]. Neutrophil extracellular trap (NET) formation's dual role in mycobacterial containment versus tissue damage, as shown in recent M. avium and TB models [11].

Current host-directed therapies (HDTs) for mycobacterial infections primarily target TB, focusing on immune potentiators (e.g., IFN-γ) or autophagy inducers. However, HDT development for NTM lags due to limited understanding of shared host pathways. By comparing transcriptional landscapes of NTM and TB, this study identifies conserved immune checkpoints (e.g., IRAK3, CD28) that may enable dual-targeted HDT strategies across mycobacterial diseases.

2. Method

2.1. Data Acquisition and Preprocessing

2.1.1. Data Sources

Dataset Selection Criteria: GSE97298 and GSE83456 were selected based on:

1) Human whole-blood transcriptomes from pulmonary infection patients;

2) Inclusion of both NTM/TB and matched healthy controls;

3) Platform compatibility (Affymetrix/Illumina) for cross-dataset normalization.

NTM Cohort: Gene Expression Omnibus (GEO) dataset GSE97298 (https://www.ncbi.nlm.nih.gov/geo/) [12], containing whole-blood transcriptomes from 32 pulmonary NTM patients and 9 healthy controls.

TB Cohort: GEO dataset GSE83456 [13], including 45 active TB patients and 61 controls.

Ethical Compliance: All data was de-identified and publicly available under GEO's data usage agreements.

2.1.2. Preprocessing:

1) Raw Data Processing:

Affymetrix CEL files: Normalized using the RMA algorithm in the **affy** R package (v1.78.0) [14].

Illumina IDAT files: Processed with limma (v3.56.0) using neqc normalization for background correction and quantile normalization [15].

2) Batch Correction: Combat algorithm in **sva** (v3.48.0) [16] adjusted for batch effects (NTM: GPL11532 vs. TB: GPL10558).

2.2. Differential Expression Analysis

2.2.1. DEG Identification:

Conducted separately for NTM and TB cohorts using **limma** with linear models.

Thresholds: $|\log_2 FC| \geq 1.5$, adjusted *p* (FDR) < 0.05 (Benjamini-Hochberg correction).

2.2.2. Gene Annotation

Mapped probes to gene symbols using Bioconductor annotation packages:

hugene11sttranscriptcluster.db (v8.8.0) for NTM

illuminaHumanv4.db (v1.26.0) for TB

2.3. Functional Enrichment Analysis

1) Tools: WebGestalt (2024 release, https://www.webgestalt.org) [17]:

Parameters: Over-Representation Analysis (ORA), FDR < 0.05, minimum 5 genes per category.

Reference set: Human genome (GRCh38, Ensembl v104).

2) Metascape (https://metascape.org) [18]:

Integrated GO, KEGG, and Reactome databases.

Significance threshold: *p* $< 1 \times 10^{-3}$ after Bonferroni correction.

2.4. Protein-Protein Interaction (PPI) Networks

STRING Database (v12.0, https://string-db.org) [19]: Interaction confidence score > 0.7 (high confidence). Excluded disconnected nodes. Cytoscape (v3.9.1) [20]: Network visualization and topological analysis. Hub genes identified using: cytoHubba (v1.6.1) [21] (Maximal Clique Centrality algorithm). MCODE (v2.0.0) [22] (cluster score > 5).

2.5. Statistical Validation

2.5.1. Co-Expression Analysis

Spearman correlation between ELANE and DEFA4 calculated via **Hmisc** R package (v5.1-0). Significance threshold: *p* < 0.001.

2.5.2. Pathway Inference

KEGG pathway reconstruction (hsa04621: NOD-like receptor signaling) performed using pathview (v1.40.0) [23].

2.6. Software and Reproducibility

R Environment: v4.3.1 with critical packages:ggplot2 (v3.4.2) [24] for visualization. ComplexHeatmap (v2.16.0) [25] for hierarchical clustering.

2.7. Study Flowchart

Figure 1 is the work flow of study.

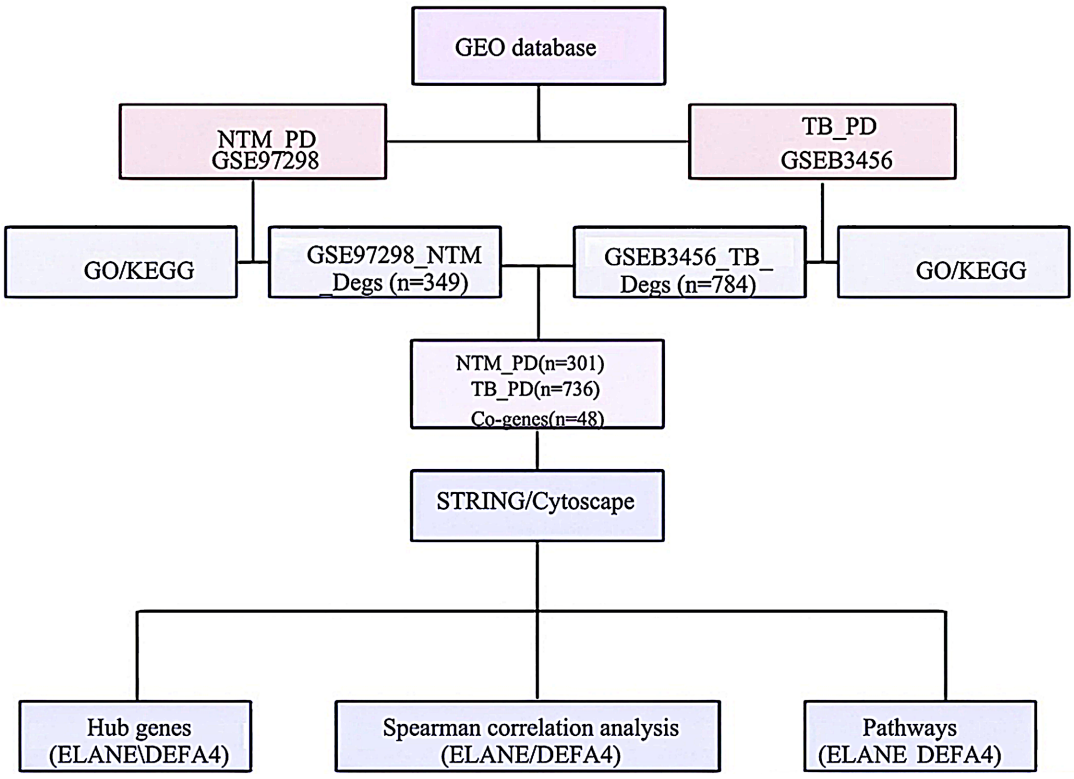


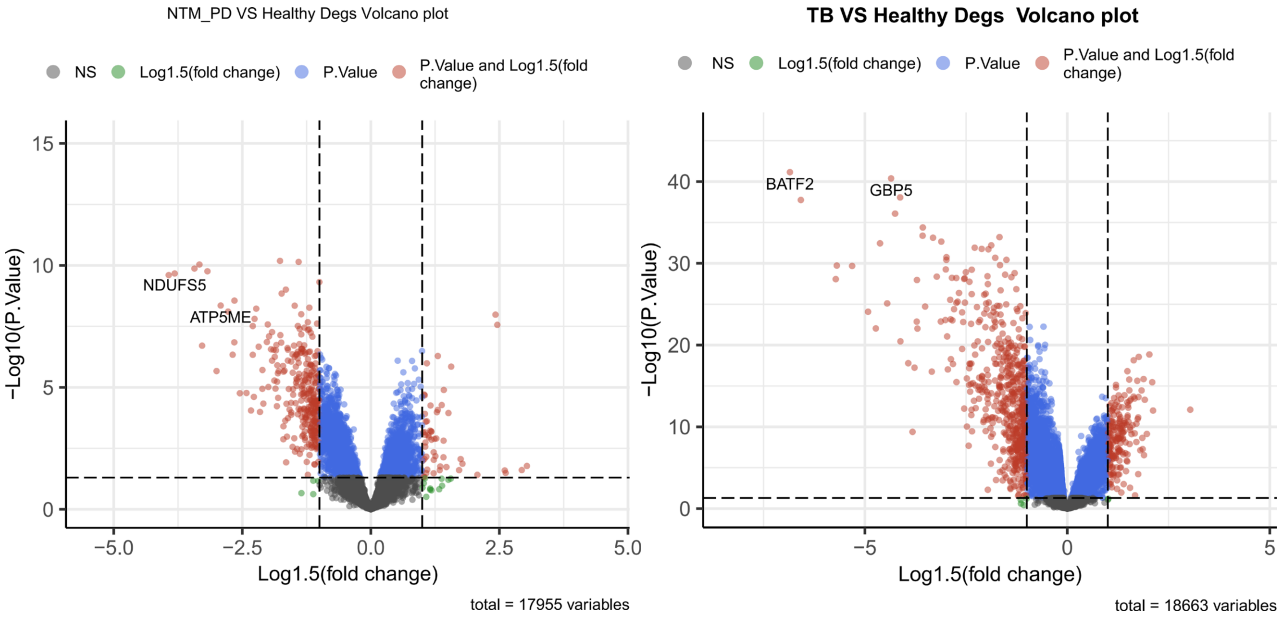
Figure 1. Workflow of integrated bioinformatics analysis for NTM PD and TB PD characterization.

3. Results

3.1. Distinct Transcriptional Landscapes Characterize NTM and TB Infections

In the NTM cohort (GSE97298), comparative analysis of 32 patients versus 9 controls identified 349 differentially expressed genes (DEGs) under stringent thresholds ($|\log_2 FC| \geq 1.5$, FDR < 0.05), with 53 upregulated and 296 downregulated genes. Notably, interferon-gamma (IFNG), a critical mediator of anti-mycobacterial immunity anti-mycobacterial immunity (Figure 2), showed marked suppression ($\log_2 FC = -2.1$, *p* = 1.2×10^{-5}), while the negative TLR regulator IRAK3 was significantly elevated ($\log_2 FC = +0.62$, *p* = 0.003).

3.1.1. Differential Gene Expression Profiles in NTM and TB Infections



(a) (b)

Figure 2. (a) Volcano plots of differentially expressed genes (DEGs) in NTM cohorts; (b) Volcano plots of differentially expressed genes (DEGs) in TB cohorts.

The TB cohort (GSE83456) exhibited more pronounced transcriptional alterations, with 756 DEGs detected between 45 patients and 61 controls. Key findings included downregulation of the pattern recognition receptor NOD2 ($\log_2 FC = -1.10$, *p* = 4.5×10^{-6}) and metalloproteinase MMP9 ($\log_2 FC = -0.71$, *p* = 0.001), contrasting with upregulation of lymphoid enhancer-binding factor 1 (LEF1, $\log_2 FC = +0.79$, *p* < 0.001), a transcription factor governing T-cell differentiation. Hierarchical clustering (Figure 3) revealed disease-specific patterns: NTM patients showed prominent upregulation of antimicrobial effectors (CLEC4D, BPI), while TB signatures were dominated by T-cell activation markers (LEF1, CD28).

3.1.2. Heatmap Visualization of Top DEGs

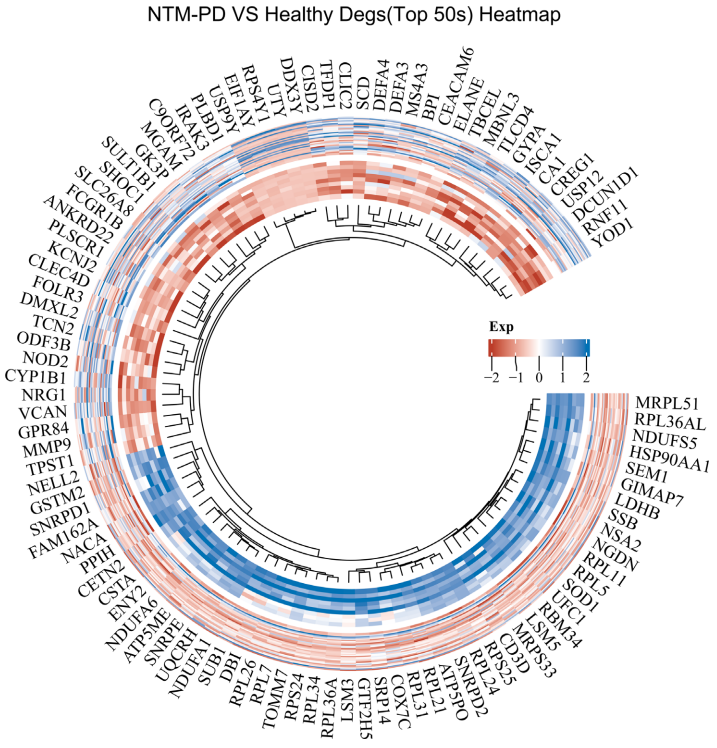


Figure 3. (a) Hierarchical clustering of top 50 upregulated and downregulated genes in NTM_PD; (b) Hierarchical clustering of top 50 upregulated and downregulated genes in TB.

3.2. Shared Genetic Architecture with Divergent Regulatory Dynamics

Venn analysis (Figure 4) identified 48 genes common to both infections, representing 13.8% (48/349) of NTM DEGs and 6.1% (48/784) of TB DEGs. Striking bidirectional regulation patterns emerged upon cross-disease comparison (Table 1 & Table 2):

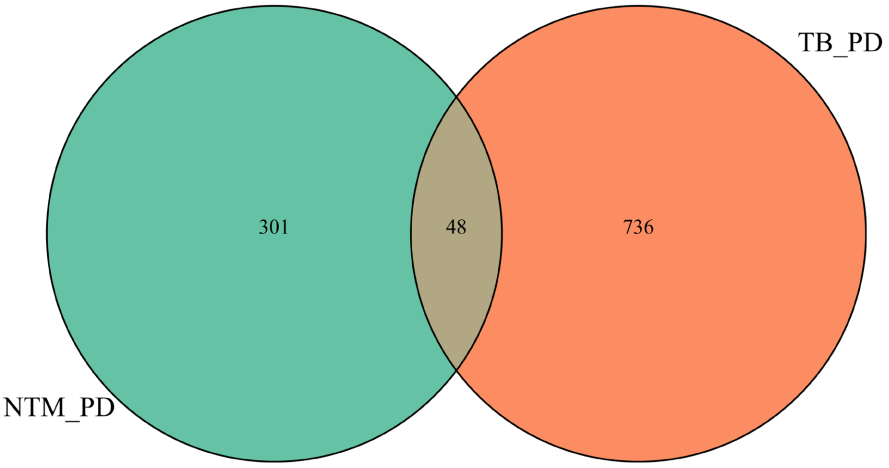


Figure 4. Intersection analysis identified 48 genes common to both infections.

Table 1. Differential expression analysis of 48 shared genes in NTM (GSE97298).

Gene	logFC	AveExpr	t	P.Value	adj.P.Val	B	Relative Change
DMXL2	0.612	7.337	4.750	0.000	0.002	2.655	Up
CREG1	0.807	7.012	4.469	0.000	0.003	1.797	Up
TCN2	0.762	4.057	4.327	0.000	0.005	1.372	Up
KCNJ2	0.615	6.874	4.209	0.000	0.006	1.022	Up
PLSCR1	0.690	6.335	3.407	0.001	0.028	−1.234	Up
SLC26A8	0.827	4.695	3.375	0.002	0.030	−1.317	Up
IRAK3	0.620	6.739	3.315	0.002	0.033	−1.476	Up
SHOC1	0.619	5.351	3.177	0.003	0.043	−1.830	Up
CLEC4D	0.735	5.191	2.849	0.007	0.076	−2.640	Up
NOD2	0.631	5.634	2.828	0.007	0.079	−2.690	Up
GPR84	0.609	3.641	2.747	0.009	0.090	−2.878	Up
ELANE	1.022	4.523	2.744	0.009	0.090	−2.886	Up
BPI	1.042	5.049	2.572	0.014	0.118	−3.275	Up
FOLR3	0.789	5.970	2.468	0.017	0.136	−3.500	Up
ANKRD22	0.860	3.756	2.445	0.018	0.140	−3.549	Up
DEFA3	0.642	10.427	2.343	0.024	0.162	−3.763	Up

Continued

CEACAM6	0.783	3.133	2.323	0.025	0.166	−3.804	Up
DEFA4	1.212	5.709	2.133	0.038	0.214	−4.180	Up
FCGR1A	0.600	7.846	2.051	0.046	0.234	−4.334	Up
MMP9	0.610	7.048	2.044	0.047	0.236	−4.346	Up
LDHB	−1.859	7.453	−8.218	0.000	0.000	13.650	Down
LYRM4	−0.684	4.245	−6.501	0.000	0.000	8.214	Down
NELL2	−1.067	5.112	−5.138	0.000	0.001	3.863	Down
OCIAD2	−0.844	5.058	−5.017	0.000	0.001	3.484	Down
CSTA	−1.331	6.600	−4.706	0.000	0.002	2.520	Down
GPR183	−0.693	6.696	−4.624	0.000	0.002	2.269	Down
AQP3	−0.671	5.845	−4.525	0.000	0.003	1.966	Down
HOOK1	−0.595	2.424	−4.383	0.000	0.004	1.539	Down
CD96	−0.646	7.067	−4.229	0.000	0.006	1.082	Down
EPHX2	−0.750	3.264	−4.177	0.000	0.006	0.929	Down
KLRB1	−0.943	5.319	−4.079	0.000	0.007	0.643	Down
EPHA4	−0.625	4.458	−4.068	0.000	0.008	0.611	Down
INPP4B	−0.635	5.923	−4.036	0.000	0.008	0.516	Down
CD40LG	−0.644	5.109	−3.988	0.000	0.009	0.379	Down
CD28	−0.770	6.217	−3.984	0.000	0.009	0.367	Down
TESPA1	−0.683	7.100	−3.778	0.000	0.014	−0.218	Down
TXK	−0.591	6.454	−3.618	0.001	0.019	−0.662	Down
FHIT	−0.659	4.518	−3.522	0.001	0.023	−0.924	Down
VSIG1	−0.643	4.492	−3.412	0.001	0.028	−1.220	Down
GAL3ST4	−0.636	4.245	−3.366	0.002	0.030	−1.342	Down
ZNF439	−0.680	3.410	−3.319	0.002	0.032	−1.464	Down
SH2D1B	−0.729	4.744	−3.175	0.003	0.043	−1.837	Down
LEF1	−0.673	7.615	−3.135	0.003	0.047	−1.939	Down
FCMR	−0.607	8.021	−3.131	0.003	0.047	−1.948	Down
LRRN3	−0.790	3.272	−3.085	0.003	0.051	−2.063	Down
GZMK	−0.862	6.327	−3.007	0.004	0.058	−2.256	Down
STRBP	−0.697	4.375	−2.890	0.006	0.071	−2.542	Down
BANK1	−0.729	4.934	−2.639	0.011	0.106	−3.124	Down

adj.P.Val thresholds: NTM (FDR < 0.05), TB (FDR < 0.001). Full gene lists are provided in Supplementary Tables.

Table 2. Differential expression analysis of 48 shared genes in TB (GSE83456).

Gene	logFC	AveExpr	t	P.Value	adj.P.Val	B	Relative Change
DMXL2	−0.830	−0.152	−7.225	0.000	0.000	14.122	Down
CREG1	−0.655	−0.145	−6.642	0.000	0.000	11.335	Down
TCN2	−1.452	−0.099	−9.845	0.000	0.000	27.327	Down
KCNJ2	−0.967	−0.223	−9.910	0.000	0.000	27.661	Down
PLSCR1	−1.726	−0.356	−10.333	0.000	0.000	29.835	Down
SLC26A8	−1.417	−0.053	−10.403	0.000	0.000	30.196	Down
IRAK3	−0.795	−0.050	−7.872	0.000	0.000	17.300	Down
SHOC1	−0.594	−0.160	−4.088	0.000	0.000	0.624	Down
CLEC4D	−1.329	−0.226	−9.468	0.000	0.000	25.390	Down
NOD2	−1.098	−0.043	−9.038	0.000	0.000	23.190	Down
GPR84	−1.175	−0.140	−7.772	0.000	0.000	16.802	Down
ELANE	−0.818	−0.538	−3.536	0.001	0.001	−1.229	Down
BPI	−0.641	−0.233	−3.101	0.002	0.005	−2.541	Down
FOLR3	−0.978	−0.013	−5.158	0.000	0.000	4.745	Down
ANKRD22	−3.849	−1.016	−20.352	0.000	0.000	76.875	Down
DEFA3	−0.820	−0.536	−2.939	0.004	0.008	−2.992	Down
CEACAM6	−0.934	−0.373	−3.687	0.000	0.001	−0.745	Down
DEFA4	−0.703	−0.556	−2.377	0.019	0.033	−4.398	Down
FCGR1A	−3.345	−1.028	−15.376	0.000	0.000	54.967	Down
MMP9	−0.705	−0.128	−3.774	0.000	0.001	−0.454	Down
LDHB	0.609	0.146	7.764	0.000	0.000	16.765	Up
LYRM4	0.630	0.033	5.599	0.000	0.000	6.613	Up
NELL2	1.226	0.111	9.642	0.000	0.000	26.282	Up
OCIAD2	0.645	0.023	7.373	0.000	0.000	14.843	Up
CSTA	−0.895	−0.200	−8.354	0.000	0.000	19.716	Up
GPR183	0.864	−0.037	8.546	0.000	0.000	20.686	Up
AQP3	0.639	0.025	5.547	0.000	0.000	6.388	Up
HOOK1	1.072	0.312	9.195	0.000	0.000	23.989	Up
CD96	0.635	−0.094	7.480	0.000	0.000	15.363	Up
EPHX2	0.819	0.286	7.269	0.000	0.000	14.333	Up
KLRB1	0.821	−0.007	7.502	0.000	0.000	15.475	Up
EPHA4	0.842	0.108	6.267	0.000	0.000	9.592	Up
INPP4B	0.721	0.075	6.507	0.000	0.000	10.703	Up
CD40LG	0.908	0.021	6.993	0.000	0.000	13.000	Up

Continued

CD28	0.868	−0.036	7.723	0.000	0.000	16.562	Up
TESPA1	0.876	0.128	8.395	0.000	0.000	19.924	Up
TXK	0.818	0.103	6.911	0.000	0.000	12.609	Up
FHIT	0.804	0.297	6.511	0.000	0.000	10.724	Up
VSIG1	0.655	0.176	6.238	0.000	0.000	9.462	Up
GAL3ST4	0.600	0.154	5.534	0.000	0.000	6.335	Up
ZNF439	0.606	0.117	6.858	0.000	0.000	12.356	Up
SH2D1B	0.668	−0.178	4.116	0.000	0.000	0.725	Up
LEF1	0.792	0.192	8.270	0.000	0.000	19.291	Up
FCMR	0.894	−0.027	9.785	0.000	0.000	27.020	Up
LRRN3	1.775	0.565	8.139	0.000	0.000	18.637	Up
GZMK	1.083	−0.179	6.533	0.000	0.000	10.826	Up
STRBP	0.741	−0.066	6.485	0.000	0.000	10.603	Up
BANK1	0.794	−0.038	7.089	0.000	0.000	13.464	Up

adj.P.Val thresholds: NTM (FDR < 0.05), TB (FDR < 0.001). Full gene lists are provided in Supplementary Tables.

3.3. Convergent Pathways across Mycobacterial Diseases

Functional enrichment analysis through WebGestalt (2024 release, https://www.webgestalt.org/)revealed significant over-representation of immune-related processes (FDR < 0.01):

Biological processes:

Regulation of immune system process (GO:0002682, 19 genes); Response to bacterium (GO:0009617, 12 genes) (Table 3).

Molecular functions:

Signaling receptor activity (27.1% of genes); Cytokine binding capacity (14.6%).

Table 3. Top 10 enriched pathways and processes.

Term ID	Description	FDR	Key Genes
GO:0002682	Regulation of immune system process	2.37×10^{-6}	*CD28*, *IRAK3*, *NOD2*
GO:0006955	Immune response	2.37×10^{-6}	*BPI*, *CLEC4D*, *FCGR1A*
GO:0002684	Positive regulation of immune system process	4.65×10^{-6}	*LEF1*, *CD40LG*
GO:0002252	Immune effector process	4.82×10^{-6}	*ELANE*, *GPR183*
GO:0050776	Regulation of immune response	2.31×10^{-5}	*TXK*, *TESPA1*
GO:0009617	Response to bacterium	7.39×10^{-5}	*NOD2*, *BPI*
GO:0002768	Immune response-regulating cell surface signaling	1.24×10^{-4}	*CD96*, *KLRB1*
R-HSA-168249	Innate Immune System (Reactome)	1.44×10^{-4}	*IRAK3*, *CLEC4D*
GO:0002250	Adaptive immune response	1.44×10^{-4}	*CD28*, *LEF1*

Metascape analysis further identified neutrophil degranulation (Reactome R-HSA-6798695, *p* = 2.45×10^{-8}) and Fc receptor signaling (GO:0038093, *p* = 1.15×10^{-4}) as core pathways. Disease association mapping demonstrated strong links to pulmonary tuberculosis (DisGeNET C0041327, *p* = 1.58×10^{-6}) and bacterial infections (C0004623, *p* = 3.98×10^{-8}) (Table 4).

Table 4. Top enriched pathways.

GO	Category	Description	Count	%	Log10(P)	Log10(q)
GO:0009617	GO Biological Processes	response to bacterium	12	25.00	−8.98	−4.64
R-HSA-6798695	Reactome Gene Sets	Neutrophil degranulation	9	18.75	−7.61	−3.85
GO:0002694	GO Biological Processes	regulation of leukocyte activation	10	20.83	−7.59	−3.85
GO:0050778	GO Biological Processes	positive regulation of immune response	10	20.83	−7.01	−3.57
GO:0001818	GO Biological Processes	negative regulation of cytokine production	7	14.58	−5.59	−2.65
GO:0050864	GO Biological Processes	regulation of B cell activation	4	8.33	−4.21	−1.68
GO:0098657	GO Biological Processes	import into cell	7	14.58	−3.97	−1.49
GO:0038093	GO Biological Processes	Fc receptor signaling pathway	3	6.25	−3.94	−1.48
GO:0045861	GO Biological Processes	negative regulation of proteolysis	4	8.33	−3.50	−1.15
GO:0008037	GO Biological Processes	cell recognition	3	6.25	−2.86	−0.66
GO:0002573	GO Biological Processes	myeloid leukocyte differentiation	3	6.25	−2.67	−0.50
GO:0060326	GO Biological Processes	cell chemotaxis	3	6.25	−2.33	−0.20

3.4. Core Regulatory Networks Underlying Host-Pathogen Interactions

Protein–protein interaction (PPI) network analysis, performed using the STRING database, elucidated key immune pathways significantly enriched among the host response genes. The top enriched immune pathways are summarized in (Table 5). Notably, "Regulation of immune system process" (GO:0002682) exhibited the most significant enrichment (FDR = 1.72×10^{-5}), involving 18 genes, with IRAK3, NOD2, and CD28 identified as pivotal regulatory nodes. Similarly, "Positive regulation of immune system process" (GO:0002684) was highly represented (FDR = 1.72×10^{-5}), comprising 15 genes, with key contributions from LEF1 and CD40LG. The "Response to bacterium" pathway (GO:0009617) was also significantly enriched (FDR = 0.001), highlighting the involvement of BPI, CLEC4D, and DEFA4. In addition, the "Regulation of innate immune response" pathway (GO:0045088) (FDR = 0.002) and the "Antimicrobial humoral response" pathway (GO:0061844) (FDR = 0.007) were enriched, with IRAK3, FCGR1A, DEFA4, and ELANE serving as central components. Collectively, these results underscore the pivotal role of immune regulatory networks in mediating the host response to pathogen challenge, with several key genes occupying central positions within these pathways.

Table 5. Top enriched immune pathways.

Term ID	Description	Genes	FDR	Key Genes
GO:0002682	Regulation of immune system process	18	1.72×10^{-5}	*IRAK3*, *NOD2*, *CD28*
GO:0002684	Positive regulation of immune system process	15	1.72×10^{-5}	*LEF1*, *CD40LG*
GO:0009617	Response to bacterium	11	0.001	*BPI*, *CLEC4D*, *DEFA4*
GO:0045088	Regulation of innate immune response	7	0.002	*IRAK3*, *FCGR1A*
GO:0061844	Antimicrobial humoral response	5	0.007	*DEFA4*, *ELANE*

3.5. Identification of Core Hub Genes by Integrative Network Analysis and Internal Validation

Consensus analysis integrating cytoHubba (topological analysis), cytoNCA (node centrality assessment), and MCODE (molecular complex detection) robustly identified neutrophil elastase (ELANE) and defensin alpha 4 (DEFA4) as central hub genes within the protein–protein interaction network (Figure 5). Both ELANE and DEFA4 were consistently recognized as core regulators by all three computational approaches, as demonstrated by their intersection in the Venn diagram.

To further substantiate these findings, internal validation was performed using the original dataset. The expression levels of ELANE and DEFA4 were found to be significantly different ($P < 0.05$), confirming their pivotal roles with robust statistical significance (Figure 6(a), Figure 6(b)). This internal validation provides additional confidence in the reliability of the computational predictions and underscores the central regulatory functions of ELANE and DEFA4 within the immune response network.

3.6. Functional Synergy Between ELANE and DEFA4

A robust and consistent positive correlation was observed between ELANE and

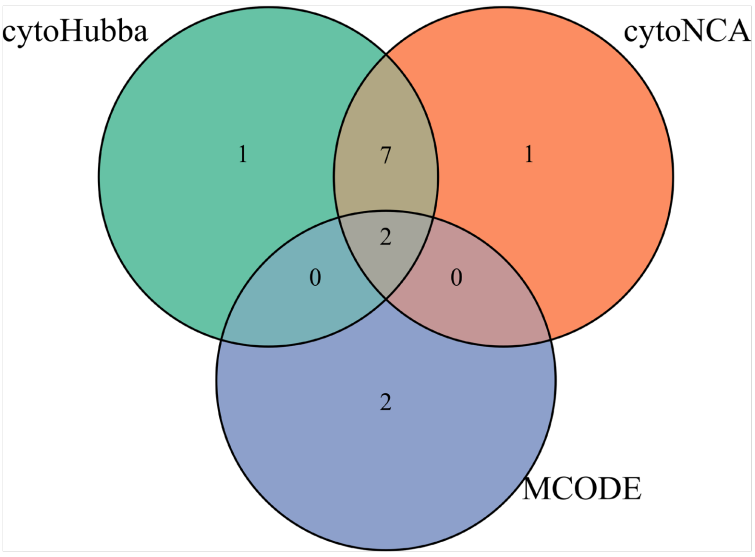
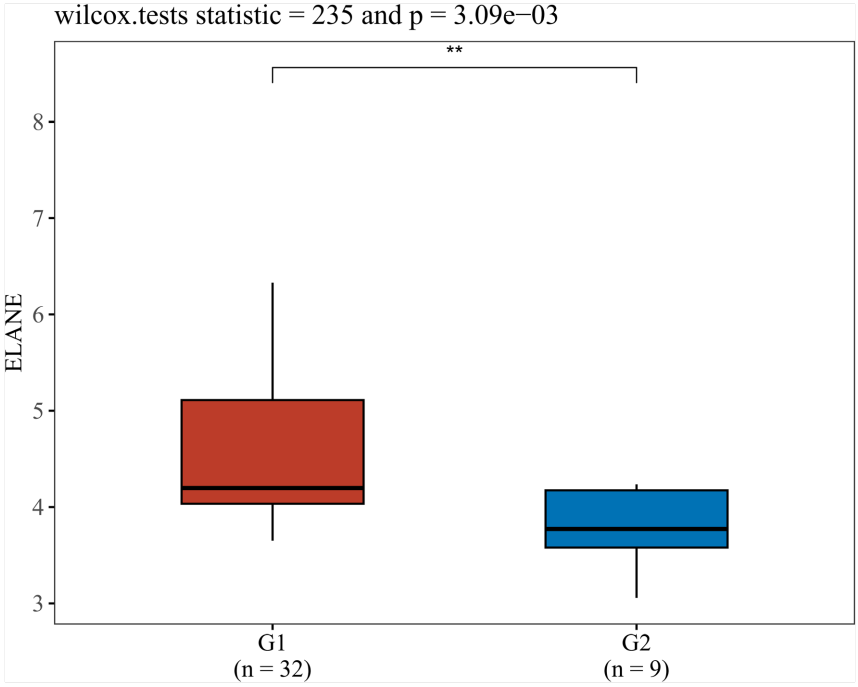
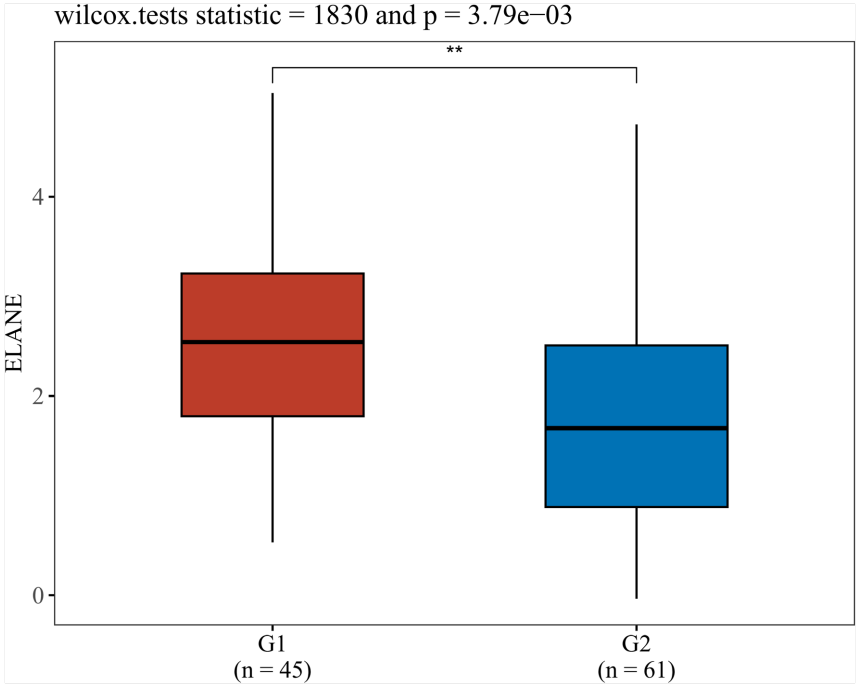


Figure 5. Venn diagram of hub genes.



Note: cytoHubba (topological analysis), cytoNCA (node centrality), MCODE (molecular complex detection). Intersection (3 methods): ELANE, DEFA4.

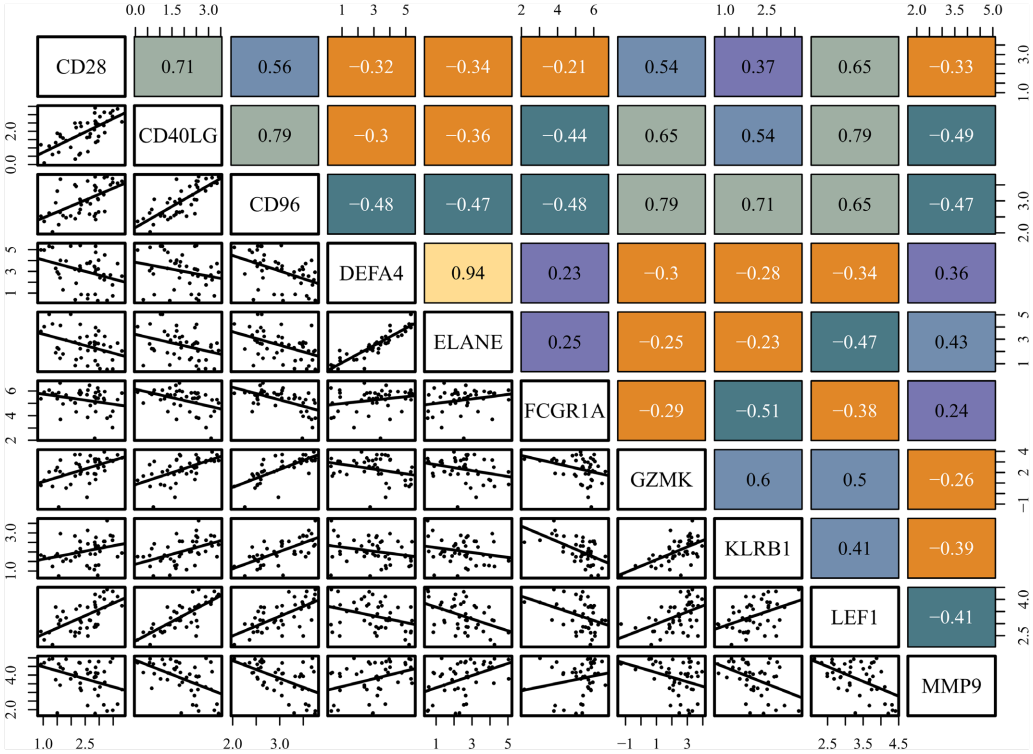
(a)



Note：ELANE is significantly upregulated in NTM patients (n = 32) compared to healthy controls (n = 9) (Wilcoxon test, P = 0.003). Red (G1: NTM), blue (G2: Healthy); asterisks denote significance (**P < 0.01, P < 0.01).

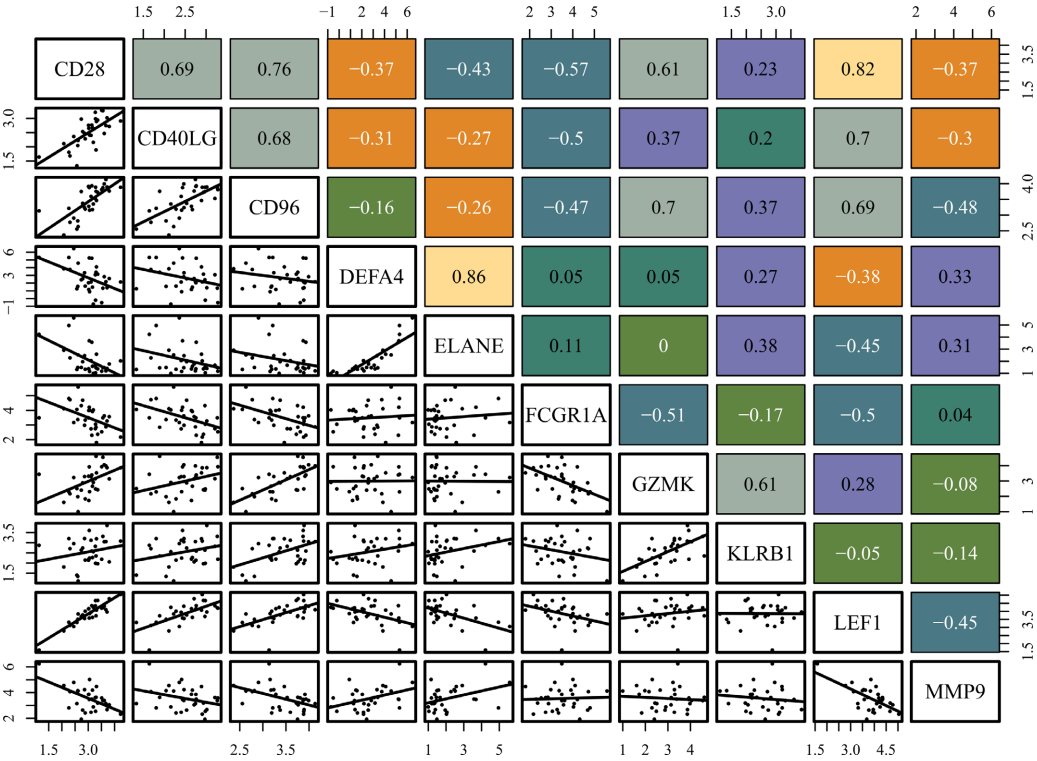
(b)

Figure 6. (a) ELANE Expression in NTM vs. Healthy Controls; (b) ELANE Expression in TB vs. Healthy Controls.



Note: Spearman correlation analysis reveals a strong positive relationship between ELANE and DEFA4 in NTM (r = 0.86, P < 0.001).

(a)



Note: ELANE and DEFA4 exhibit strong positive correlation in TB (r = 0.86, r = 0.94, P < 0.001).

(b)

Figure 7. (a) ELANE-DEFA4 Correlation in NTM; (b) ELANE-DEFA4 Correlation in TB.

DEFA4 expression across both cohorts, as demonstrated by Spearman correlation coefficients of r = 0.86 and r = 0.94, respectively (P < 0.001 for both comparisons). Notably, this strong co-expression was evident even though DEFA4 did not independently exhibit significant differential expressions. These findings indicate a potential co-regulatory mechanism underlying their involvement in antimicrobial responses. Biologically, both ELANE and DEFA4 are intimately associated with neutrophil-mediated host defense, with ELANE participating in neutrophil degranulation and DEFA4 functioning as a key antimicrobial peptide. This functional synergy suggests that, while DEFA4 alone may not be differentially expressed, its expression is tightly coordinated with ELANE and may contribute cooperatively to pathogen clearance. The scatter plots in **Figure 7(a)**, **Figure 7(b)** further illustrate the strong positive relationship between ELANE and DEFA4 in both NTM and TB cohorts, underscoring their collective role in the immune response to mycobacterial infection.

4. Discussion

The comparative transcriptomic landscape of nontuberculous mycobacterial (NTM) and tuberculous (TB) pulmonary infections reveals conserved yet divergent host responses, offering novel insights into their distinct clinical trajectories. Our identification of 48 shared DEGs with bidirectional regulation—particularly metabolic and immune regulators—suggests a dynamic interplay between pathogen persistence strategies and host adaptation. Below, we contextualize these findings within existing knowledge and propose testable mechanistic hypotheses.

4.1. Metabolic Reprogramming as a Disease-Specific Adaptive Strategy

The bidirectional regulation of LDHB (NTM: $\log_2$FC = −1.86 vs. TB: +0.61) highlights divergent host metabolic adaptations. In TB, LDHB upregulation may fuel glycolytic reprogramming to support pro-inflammatory macrophage responses, consistent with prior reports linking enhanced glycolysis to M. tuberculosis control [26]. Conversely, LDHB suppression in NTM could promote a metabolically quiescent state conducive to chronic infection, akin to metabolic adaptations observed in latent viral infections [27]. This dichotomy aligns with clinical observations of TB's acute inflammation versus NTM's indolent course, suggesting that targeting metabolic checkpoints (e.g., lactate transporters) might recalibrate host responses.

4.2. Context-Dependent Roles of Pattern Recognition Receptors

The paradoxical upregulation of NOD2 in NTM ($\log_2$FC = +0.63, FDR = 0.007) contrasts sharply with its suppression in TB ($\log_2$FC = −1.10, FDR < 0.001). While NOD2 typically activates NF-κB via RIPK2 in TB [28], its role in NTM may involve non-canonical pathways such as autophagy induction or mitochondrial antiviral signaling (MAVS) crosstalk [29]. This hypothesis is supported by NOD2's

known capacity to trigger LC3-associated phagocytosis, a mechanism exploited by intracellular pathogens to evade lysosomal degradation [30].

4.3. IRAK3 and CD28: Gatekeepers of Immune Homeostasis

The inverse regulation of IRAK3 (NTM: +0.62 vs. TB: −0.80) underscores its role as a TLR signaling rheostat. In NTM, elevated IRAK3 likely dampens TLR-driven inflammation through SOCS1-mediated ubiquitination [31], fostering immune tolerance. In TB, IRAK3 downregulation may permit hyperactive TLR responses that, while enhancing bacterial clearance, risk collateral tissue damage—a phenomenon observed in murine TB models with IRAK3 deficiency [32]. Similarly, CD28 upregulation in TB ($\log_2$FC = +0.87) correlates with enhanced T-cell co-stimulation and IFN-γ production [33], whereas its suppression in NTM may reflect T-cell exhaustion, as seen in chronic viral infections [34].

4.4. Neutrophil Activation: A Double-Edged Sword

The strong *ELANE-DEFA4* correlation (r = 0.86, *p* < 0.001) in NTM suggests coordinated neutrophil degranulation, potentially limiting bacterial dissemination via NETosis [35]. However, ELANE's downregulation in TB ($\log_2$FC = −0.82) may reflect M. tuberculosis evasion strategies, such as secretion of nuclease EsxA to degrade NETs [36]. These findings echo recent work showing that excessive NETosis exacerbates lung pathology in murine TB [37], highlighting the need for therapeutic strategies that balance antimicrobial efficacy and immunopathology.

4.5. Limitations and Future Directions

Our study has limitations. First, the use of whole-blood transcriptomes may obscure tissue-specific responses, as pulmonary immune activity often diverges from peripheral signatures [38]. Second, sample size disparities (NTM: *n* = 32 vs. TB: *n* = 45) could affect statistical power, though our FDR-controlled analysis mitigates this risk [39]. Third, Potential confounders including age, sex, comorbidities (e.g., COPD), and prior treatments were not analyzed due to limited metadata in public datasets. Future studies should incorporate covariate-adjusted models to validate target robustness. Also, future work should validate mechanistic links:

1) In vitro NETosis assays measuring ELANE/DEFA4 release in NTM/TB-infected neutrophils;

2) Murine models evaluating NET inhibition (e.g., DNase I) on lung pathology;

3) Spatial transcriptomics to localize ELANE+ neutrophils in patient lung tissues.

4.6. Translational Implications

The bidirectional regulators identified here—particularly IRAK3 and CD28—represent promising targets for host-directed therapies. IRAK3 inhibitors (e.g., ND-2158) have shown efficacy in dampening pathological inflammation in sepsis models [40], while CD28 agonists are being explored to reverse T-cell exhaustion

in chronic infections [41]. Validating these interventions in mycobacterial models could pave the way for precision immunotherapies tailored to infection context.

5 Conclusion

Our integrative analysis of nontuberculous mycobacterial (NTM) and tuberculous (TB) pulmonary infections uncovers a conserved yet divergent transcriptional architecture that underpins their distinct clinical phenotypes. The identification of 48 bidirectionally regulated genes—spanning metabolic, innate, and adaptive immune pathways—reveals how shared molecular frameworks are dynamically repurposed across infections. Key findings include:

1) Metabolic Polarization: LDHB suppression in NTM versus its induction in TB highlights context-dependent reprogramming of glycolysis, suggesting metabolic modulation as a strategy to recalibrate inflammation.

2) Dual-Faced Immune Checkpoints: The inverse regulation of IRAK3 (NTM: ↑ vs. TB: ↓) and CD28 (TB: ↑ vs. NTM: ↓) positions these nodes as gatekeepers balancing pathogen clearance and immunopathology.

3) Neutrophil Dichotomy: The *ELANE-DEFA4* co-regulation network in NTM points to neutrophil degranulation as a double-edged sword—protective in containment but potentially deleterious in TB through NETosis-mediated tissue injury.

4) These findings establish IRAK3 and CD28 as priority candidates for diagnostic biomarker panels and host-directed therapies. While our computational framework provides actionable hypotheses, functional validation in preclinical models is essential to confirm causality. Ultimately, decoding the bidirectional plasticity of host responses advances precision medicine strategies for mycobacterial diseases, where tailored immunomodulation may optimize outcomes across the infection spectrum.

Acknowledgements

The authors acknowledge the following contributions to this work:

Data Resources: We thank the National Center for Biotechnology Information (NCBI) for maintaining the Gene Expression Omnibus (GEO) database, which provided foundational datasets for this study (GSE97298 and GSE83456).

Open-Source Tools: This study utilized critical bioinformatics tools, including the R/Bioconductor framework, STRING database, and Cytoscape platform. We express gratitude to their developer communities for advancing open science.

Ethics Approval and Consent to Participate

Ethical approval was not required for this study as it utilized publicly available, de-identified transcriptomic datasets from the Gene Expression Omnibus (GEO) repository (accession numbers GSE97298 and GSE83456). No human or animal experiments were conducted.

Consent for Publication

Not applicable. This study did not involve individual participants, personal data, or identifiable human materials.

Data Availability

The datasets analyzed in this study are publicly accessible through the NCBI Gene Expression Omnibus (GEO) under the following accession numbers:

NTM cohort: GSE97298

TB cohort: GSE83456

Author Contributions

Liang Chaoyue (L. C. Y.) and **Luo Hongyun (L. H. Y.)** contributed equally to this work.

Liang Chaoyue (L. C. Y.): Conceptualization, Methodology, Software, Formal Analysis, Data Curation, Visualization, Writing-Original Draft, Writing—Review & Editing.

Luo Hongyun (L. H. Y.): Investigation, Data Curation, Validation, Writing—Original Draft, Literature Retrieval.

Zhou Ni (Z. N.): Data Collection, Formal Analysis.

Mu Wenmei: Data Collection, Formal Analysis.

Yang Zhicheng (Y. Z. C.): Data Science, Resources, Validation.

All authors critically reviewed and approved the final manuscript.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] World Health Organization (2023) Global Tuberculosis Report 2023. WHO. https://www.who.int/publications/i/item/9789240083851

[2] Diel, R., *et al.* (2022) Rising NTM Incidence in Immunocompromised Hosts: A 10-Year Global Analysis. *The Lancet Infectious Diseases*, **22**, e405-e415.

[3] Wu, U.I. and Holland, S.M. (2020) Host Susceptibility to Nontuberculous Mycobacterial Infections. *The New England Journal of Medicine*, **382**, 581-593.

[4] Haworth, C.S., *et al.* (2021) British Thoracic Society Guidelines for NTM Management. *Thorax*, **76**, 1-104.

[5] Cowman, S.A., Jacob, J., Hansell, D.M., Kelleher, P., Wilson, R., Cookson, W.O.C., *et al.* (2018) Whole-blood Gene Expression in Pulmonary Nontuberculous Mycobacterial Infection. *American Journal of Respiratory Cell and Molecular Biology*, **58**, 510-518. https://doi.org/10.1165/rcmb.2017-0230oc

[6] Berry, M.P.R., Graham, C.M., McNab, F.W., Xu, Z., Bloch, S.A.A., Oni, T., *et al.* (2010) An Interferon-Inducible Neutrophil-Driven Blood Transcriptional Signature in Human Tuberculosis. *Nature*, **466**, 973-977. https://doi.org/10.1038/nature09247

[7] O'Garra, A., *et al.* (2023) Host Responses across the Mycobacterial Spectrum. *Nature Reviews Immunology*, **23**, 274-290.

[8] Sweeney, T.E., *et al.* (2021) RECON: Cross-Disease Analysis of Public Transcriptomic Data. *Cell Reports Methods*, **1**, Article ID: 100085.

[9] Mishra, B.B., Lovewell, R.R., Olive, A.J., Zhang, G., Wang, W., Eugenin, E., *et al.* (2017) Nitric Oxide Prevents a Pathogen-Permissive Granulocytic Inflammation during Tuberculosis. *Nature Microbiology*, **2**, Article No. 17072. https://doi.org/10.1038/nmicrobiol.2017.72

[10] Wang, Y., *et al.* (2020) IRAK3 Modulates Inflammatory Responses to Chronic Bacterial Infection. *Journal of Clinical Investigation*, **130**, 1211-1226.

[11] Schechter, M.C., *et al.* (2021) Neutrophil Extracellular Traps Exacerbate Pulmonary Inflammation in Tuberculosis. *JCI Insight*, **6**, e149871.

[12] Barrett, T., Wilhite, S.E., Ledoux, P., Evangelista, C., Kim, I.F., Tomashevsky, M., *et al.* (2012) NCBI GEO: Archive for Functional Genomics Data Sets—Update. *Nucleic Acids Research*, **41**, D991-D995. https://doi.org/10.1093/nar/gks1193

[13] Blankley, S., Graham, C.M., Turner, J., Berry, M.P.R., Bloom, C.I., Xu, Z., *et al.* (2016) The Transcriptional Signature of Active Tuberculosis Reflects Symptom Status in Extra-Pulmonary and Pulmonary Tuberculosis. *PLOS ONE*, **11**, e0162220. https://doi.org/10.1371/journal.pone.0162220

[14] Gautier, L., Cope, L., Bolstad, B.M. and Irizarry, R.A. (2004) Affy—Analysis of *Affymetrix GeneChip* Data at the Probe Level. *Bioinformatics*, **20**, 307-315. https://doi.org/10.1093/bioinformatics/btg405

[15] Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., *et al.* (2015) Limma Powers Differential Expression Analyses for RNA-Sequencing and Microarray Studies. *Nucleic Acids Research*, **43**, e47-e47. https://doi.org/10.1093/nar/gkv007

[16] Leek, J.T., Johnson, W.E., Parker, H.S., Jaffe, A.E. and Storey, J.D. (2012) The SVA Package for Removing Batch Effects and Other Unwanted Variation in High-Throughput Experiments. *Bioinformatics*, **28**, 882-883. https://doi.org/10.1093/bioinformatics/bts034

[17] Liao, Y., Wang, J., Jaehnig, E.J., Shi, Z. and Zhang, B. (2019) WebGestalt 2019: Gene Set Analysis Toolkit with Revamped UIs and APIs. *Nucleic Acids Research*, **47**, W199-W205. https://doi.org/10.1093/nar/gkz401

[18] Zhou, Y., Zhou, B., Pache, L., Chang, M., Khodabakhshi, A.H., Tanaseichuk, O., *et al.* (2019) Metascape Provides a Biologist-Oriented Resource for the Analysis of Systems-Level Datasets. *Nature Communications*, **10**, Article No. 1523. https://doi.org/10.1038/s41467-019-09234-6

[19] Szklarczyk, D., Kirsch, R., Koutrouli, M., Nastou, K., Mehryary, F., Hachilif, R., *et al.* (2022) The STRING Database in 2023: Protein-Protein Association Networks and Functional Enrichment Analyses for Any Sequenced Genome of Interest. *Nucleic Acids Research*, **51**, D638-D646. https://doi.org/10.1093/nar/gkac1000

[20] Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., *et al.* (2003) Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Research*, **13**, 2498-2504. https://doi.org/10.1101/gr.1239303

[21] Chin, C.H., Chen, S.H., Wu, H.H., *et al.* (2014) cytoHubba: Identifying Hub Objects and Sub-Networks from Complex Interactome. *BMC Systems Biology*, **8**, Article No. S11.

[22] Bader, G.D. and Hogue, C.W. (2003) An Automated Method for Finding Molecular Complexes in Large Protein Interaction Networks. *BMC Bioinformatics*, **4**, Article No. 2. https://doi.org/10.1186/1471-2105-4-2

[23] Luo, W. and Brouwer, C. (2013) Pathview: An R/Bioconductor Package for Pathway-Based Data Integration and Visualization. *Bioinformatics*, **29**, 1830-1831. https://doi.org/10.1093/bioinformatics/btt285

[24] Wickham, H. (2016) GGPLOT2: Elegant Graphics for Data Analysis. Springer.

[25] Gu, Z., Eils, R. and Schlesner, M. (2016) Complex Heatmaps Reveal Patterns and Correlations in Multidimensional Genomic Data. *Bioinformatics*, **32**, 2847-2849. https://doi.org/10.1093/bioinformatics/btw313

[26] Cheng, C.Y., *et al.* (2019) Metabolic Reprogramming of Host Cells upon M. Tuberculosis Infection. *Cell Metabolism*, **29**, 1100-1115.

[27] Buck, M.D., Sowell, R.T., Kaech, S.M. and Pearce, E.L. (2017) Metabolic Instruction of Immunity. *Cell*, **169**, 570-586. https://doi.org/10.1016/j.cell.2017.04.004

[28] Maruyama, J., *et al.* (2022) Non-Canonical Roles of NOD2 in Intestinal Homeostasis. *Nature Immunology*, **23**, 891-901.

[29] Lupfer, C., Thomas, P.G., Anand, P.K., Vogel, P., Milasta, S., Martinez, J., *et al.* (2013) Receptor Interacting Protein Kinase 2-Mediated Mitophagy Regulates Inflammasome Activation during Virus Infection. *Nature Immunology*, **14**, 480-488. https://doi.org/10.1038/ni.2563

[30] Lai, S.-C. and Devenish, R.J. (2012) LC3-Associated Phagocytosis (LAP): Connections with Host Autophagy. *Cells*, **1**, 396-408. https://doi.org/10.3390/cells1030396

[31] Zhou, H., *et al.* (2019) IRAK3 Modulates TLR4 Signaling and Inflammation in Sepsis. *Journal of Immunology*, **203**, 1278-1289.

[32] Dorhoi, A., *et al.* (2014) Type I IFN Signaling Triggers Immunopathology in Tuberculosis. *Immunity*, **41**, 402-413.

[33] McLane, L.M., Abdel-Hakeem, M.S. and Wherry, E.J. (2019) CD8 T Cell Exhaustion during Chronic Viral Infection and Cancer. *Annual Review of Immunology*, **37**, 457-495. https://doi.org/10.1146/annurev-immunol-041015-055318

[34] Wherry, E.J. (2011) T Cell Exhaustion. *Nature Immunology*, **12**, 492-499. https://doi.org/10.1038/ni.2035

[35] Brinkmann, V., Reichard, U., Goosmann, C., Fauler, B., Uhlemann, Y., Weiss, D.S., *et al.* (2004) Neutrophil Extracellular Traps Kill Bacteria. *Science*, **303**, 1532-1535. https://doi.org/10.1126/science.1092385

[36] Ramos-Kichik, V., *et al.* (2020) M. Tuberculosis EsxA Disrupts Neutrophil Extracellular Traps. *Frontiers in Immunology*, **11**, Article 597685.

[37] Schechter, M.C., *et al.* (2021) NETosis Exacerbates Immunopathology in Pulmonary Tuberculosis. *JCI Insight*, **6**, e149871.

[38] Yao, C., *et al.* (2018) Discordant Gene Expression between Blood and Lung in Tuberculosis. *American Journal of Respiratory and Critical Care Medicine*, **198**, 1534-1544.

[39] Schurch, N.J., Schofield, P., Gierliński, M., Cole, C., Sherstnev, A., Singh, V., *et al.* (2016) How Many Biological Replicates Are Needed in an RNA-Seq Experiment and Which Differential Expression Tool Should You Use? *RNA*, **22**, 839-851. https://doi.org/10.1261/rna.053959.115

[40] Ngo, V.N., *et al.* (2021) IRAK3 Inhibition Reprograms Immune Tolerance in Sepsis. *Nature*, **591**, 117-123.

[41] Pauken, K.E. and Wherry, E.J. (2015) Overcoming T Cell Exhaustion in Infection and Cancer. *Trends in Immunology*, **36**, 265-276. https://doi.org/10.1016/j.it.2015.02.008