Article

Discovery Medicine 2024; 36(185): 1268–1279
https://doi.org/10.24976/Discov.Med.202436185.117

Identification of a Diagnostic Multiomics-Based Biomarker Cluster in a Mouse Model of Pulmonary Tuberculosis

Hongtai Zhang[1,2,†], Mingmin Shi[3,†], Lan Yu[2], Fanlei Ran[3,4], Nan Zheng[4], Xingyun Wang[4], Yi Liu[5], Chuanyou Li[2], Donghua Li[6], Jinlong Li[7], Lijun Bi[3,4,8,*], Zhilong Wu[1,*]

[1]Foshan Fourth People's Hospital, 528000 Foshan, Guangdong, China
[2]Beijing Center for Disease Prevention and Control, 100035 Beijing, China
[3]University of Chinese Academy of Sciences, 100049 Beijing, China
[4]Key Laboratory of RNA Biology and State Key Laboratory of Biomacromolecules, CAS Center of Excellence in Biomacromolecules, Institute of Biophysics, Chinese Academy of Sciences, 100101 Beijing, China
[5]Biobank of Beijing Chest Hospital, Beijing Tuberculosis and Thoracic Tumor Research Institute/Beijing Chest Hospital, Capital Medical University, 101125 Beijing, China
[6]College of Pharmacy and Bioengineering, Shenyang University of Chemical Technology, 110142 Shenyang, Liaoning, China
[7]TB Healthcare Co., Ltd., 528300 Foshan, Guangdong, China
[8]Guangzhou National Laboratory, 510005 Guangzhou, Guangdong, China
*Correspondence: blj@ibp.ac.cn (Lijun Bi); wuzlfs@163.com (Zhilong Wu)
[†]These authors contributed equally.
Published: 20 June 2024

Background: Tuberculosis (TB) stands as the second most prevalent infectious agent-related cause of death worldwide in 2022, trailing only COVID-19. With 1.13 million reported deaths, this figure is more than half of the mortality associated with human immunodeficiency virus/acquired immunodeficiency syndrome (HIV/AIDS), which accounted for 0.63 million deaths. Diagnosing *Mycobacterium tuberculosis* (MTB) infection remains a formidable challenge due to the inability to isolate and detect MTB in sputum and within the human body. The absence of universally reliable diagnostic criteria for MTB infection globally poses a significant obstacle to preventing the progression of tuberculosis from the MTB infection stage.

Methods: In this study, our objective was to formulate a diagnostic biomarker cluster capable of discerning the progression of MTB infection and disease. This was achieved through a comprehensive joint multiomics analysis, encompassing transcriptome, proteome, and metabolome, conducted on lung tissue samples obtained from both normal control mice and those infected with MTB.

Results: A total of 1690 differentially expressed genes and 94 differentially expressed proteins were systematically screened. From this pool, 10 core genes were singled out. Additionally, eight long non-coding ribonucleic acids and eight metabolites linked to these core genes were identified to establish a cohesive cluster of biomarkers. This multiomics-based biomarker cluster demonstrated its capability to differentiate uninfected samples from MTB-infected samples effectively in both principle component analysis and the construction of a random forest model.

Conclusion: The outcomes of our study strongly suggest that the multiomics-based biomarker cluster holds significant potential for enhancing the diagnosis of MTB infection.

Keywords: biomarker; tuberculosis; *Mycobacterium tuberculosis*; multiomics; metabolite; lncRNA

Introduction

Tuberculosis (TB), ranking among the top 10 global causes of death, is a chronic infectious disease caused by *Mycobacterium tuberculosis* (MTB), primarily affecting the lungs [1]. Approximately one-fourth of the world's population carries an MTB infection, with 5%–10% of those individuals at risk of developing TB [2]. The primary mode of MTB transmission is through sputum droplets from infected individuals [3]. In the early stages, 10% of MTB-infected patients exhibit clinical symptoms such as cough and sputum production, with the majority having latent infections [4]. Understanding the pathogenesis leading from asymptomatic infection to active TB has been a longstanding focus of research.

Early diagnosis of MTB infection is crucial, particularly given the limitations of methods focused on mycobacterium pathogens, such as acid-fast bacteria staining [5] and sputum culture, in detecting latent infections. The World Health Organization (WHO) recommends non-pathogen-

based detection methods to enhance clinical identification of MTB infection [6]. Host immune response biomarkers offer valuable insights to address this challenge.

In clinical settings, immunology-based approaches, including the tuberculin derivative skin test or purified protein derivative, have long been employed for TB diagnosis [7]. Similarly, the T cell-based TB-specific antigen response interferon-γ release test [8], which detects MTB-specific antibodies but exhibits low sensitivity and specificity [9]. Despite the emergence of novel biomarkers like non-coding ribonucleic acids (ncRNA) such as *MIR3945HG*, *NEAT1*, *PCED1B-AS1*, *LOC152742*, and *ENST00000354432* [10], as well as metabolites like fatty acids, amino acids, palmitic acid, phosphatidylcholine, and lysophosphatidylcholine [11], precise determination of MTB infection based on a single immunological biomarker remains challenging.

In contrast to the aforementioned methods, multiomics technology holds promise for a more comprehensive understanding of MTB infection characteristics by integrating information from the transcriptome, genome, proteome, and metabolome. This hypothesis draws inspiration from the significant strides made in tumor research utilizing multiomics [12]. For instance, a comprehensive analysis of genomic, transcriptomic, and proteomic data unveiled abnormal gene expression linked to tumor progression in prostate cancer patients [13]. Similarly, a study combining proteomic and metabolomic data identified specific metabolite changes in plasma samples of systemic lupus erythematosus patients, aiding in the identification of disease-associated molecular markers and potential therapeutic targets; another study integrating proteomic and metabolomic data, detailed cell behavior post-viral infection [14,15]. Surprisingly, in the context of TB infection and morbidity, multiomics methods have not yet been extensively employed to detect immune marker groups. Existing biomarkers associated with TB infection have primarily been screened using single technologies, including transcriptomics [16], proteomics [17], and metabolomics [18]. Leveraging multiomics methods to explore immune marker groups has the potential to significantly contribute to both early detection and disease control of TB.

Hence, the focus of this study was to explore an immune biomarkers cluster using transcriptome, proteome, and metabolome data to assess MTB infection. Lung tissue samples were procured from six normal and six MTB-infected mice, and differentially expressed genes, proteins, and metabolites were identified between MTB-infected and non-infected samples. Additionally, a set of immune markers, including non-coding RNAs, coding RNAs, proteins, and metabolites, associated with MTB infection in mice were identified. These findings present a promising avenue for enhancing MTB-infected diagnostics in humans.

Materials and Methods

Bacterial Culture

MTB H37Rv strains (American Type Culture Collection, 25618) were cultivated in Middlebrook 7H9 medium (Becton Dickinson, 271310), supplemented with 0.5% glycerol, 0.05% Tween-80, and 10% oleic acid-albumin-dextrose-catalase. MTB working stocks were meticulously prepared during the mid-log phase, with the optical density at 600 nm ranging between 0.6 and 0.8.

Animal Care and Sample Collection

Female C57BL/6 mice, aged 6–8 weeks, were procured from Cyagen Biosciences, Inc. (Suzhou, China) and were maintained under controlled Animal Biosafety Level 3 barrier conditions in specialized animal care facilities. This study received approval from the Animal Experiment Ethics Committee of Beijing Chest Hospital (No. XK2023-040). Mice were randomly assigned to either the infection or control groups. In the infection group, three animals were intravenously injected via the tail vein with approximately 1×10^6 CFU of MTB H37Rv. The experiments adhered to the guidelines outlined by the Chinese Council on Animal Care. After a 10-week duration, all mice were anesthetized by intraperitoneal injection of a 1% pentobarbital sodium solution (50 mg/kg) and sacrificed by cervical dislocation. Their lung tissue samples were collected.

Transcriptome Generation

For the identification of RNA biomarker candidates associated with TB, lung tissue samples from the mice were promptly dissected, snap-frozen in liquid nitrogen, ground into a powder in a mortar, and then transferred to a ribonuclease (RNase)-free tube. Total RNA extraction was carried out using TRIzol reagent (Invitrogen, Carlsbad, CA, USA) in accordance with the manufacturer's guidelines. Subsequently, the obtained RNA was resuspended in RNase-free water and stored at –80 °C. The purity of RNA was assessed using the kaiao K5500® Spectrophotometer (Kaiao, Beijing, China), while the RNA concentration and integrity were evaluated using the RNA Nano 6000 Assay kit on the Bioanalyzer 2100 system (Agilent Technologies, Palo Alto, CA, USA). Enriched messenger RNA (mRNA) underwent fragmentation to construct a high-quality RNA sequencing library, employing KAPA Stranded RNA-Seq kits per the manufacturer's instructions. Sequencing of the prepared libraries was executed on the Illumina HiSeq X Ten platform (Illumina, San Diego, CA, USA).

Metabolome Generation

To identify metabolite biomarker candidates associated with TB, mice lung tissue samples were thawed at room temperature and transferred into 1.5 mL Eppendorf (EP) tubes. Subsequently, the tissues were homogenized at 45 Hz for 1 minute, repeated twice, in an extraction solvent

(80% methanol [v/v]) containing steel beads. After centrifugation, 350 µL of the supernatant was transferred into 1.5 mL EP tubes and dried completely in a vacuum concentrator without heating.

Next, 30 µL of methoxyamination hydrochloride (20 mg/mL in pyridine) and 40 µL of the N,O-Bis(trimethylsilyl) trifluoroacetamide reagent (1% trimethylsilyl chloride, v/v) were sequentially added to each sample. The samples were incubated for 30 minutes at 80 °C and then for 1.5 hours at 70 °C. All samples underwent further analysis using an Agilent 7890 gas chromatograph system coupled with a Pegasus high throughput (HT) time-of-flight (TOF) mass spectrometer. Data preprocessing and annotation were performed using the Chroma TOF 4.3X software (LECO Corporation, Saint Joseph, MI, USA) and the LECO-Fiehn Rtx5 database.

Proteome Generation

To identify protein biomarker candidates associated with TB, approximately 100 mg of frozen mice lung tissue sample was homogenized in 500 µL of extraction buffer (8 M urea, 1 mM phenylmethylsulfonyl fluoride, 1 × Roche complete mini protease inhibitor cocktail, pH 7.4) for 5 minutes. The protein content in each sample lysate was determined using a bicinchoninic acid assay kit (Thermo Fisher, Waltham, MA, USA, 23235). The lysate underwent reduction with 10 mM dithiothreitol and alkylation with 100 mM iodoacetamide. Enzyme lysate (50 mM NH_4HCO_3, 5% ceric ammonium nitrate (CAN), 10 ng/µL trypsin, double distilled (dd)H_2O) was added to digest the samples into peptides. Subsequently, the peptides were desalted using an extraction solution (5% formic acid, 50% CAN, ddH_2O) through vacuum freeze-drying.

For Tandem Mass Tag (TMT) labeling, a TMT Kit (Thermo Fisher, Waltham, MA, USA, 90064B) was utilized to recombine the peptides in 0.5 tetradecyltrimethylammonium bromide following the manufacturer's protocol. The mixture was then incubated for 2 hours at room temperature (37 °C) for TMT labeling. The labeled peptides were separated through high-pH reversed-phase high-performance liquid chromatography and analyzed using mass spectrometry.

Study Participants

To validate the applicability of the identified mice RNA biomarker candidates in patients with TB, whole blood samples were collected from additional healthy donors (n = 3) and pulmonary TB patients (n = 6) undergoing examination at Beijing Chest Hospital, Capital Medical University (Beijing, China). Healthy donors, confirmed to be MTB-negative serologically, and TB patients, who had not received any anti-TB treatment at the time of examination, were adults over 18 years of age. All study participants provided written informed consent. The use of deidentified patient information and the collection of clinical samples were approved by the Ethics Committee of the Beijing Chest Hospital, Capital Medical University (Beijing, China) (No. BJCH-2019-0301-018).

RNA Extraction and Quantitative Polymerase Chain Reaction (PCR)

Peripheral blood mononuclear cells (PBMCs) were obtained from the remaining blood samples of healthy donors and patients with TB, as examined at Beijing Chest Hospital. The isolation of peripheral PBMCs involved density gradient centrifugation, utilizing a density reagent (Tianjin Yeyang Huake Biological Technology Co., Ltd., Tianjin, China, HY2015), following the manufacturer's instructions. Subsequently, RNA extraction from PBMCs was carried out using the Total RNA Extraction Kit (deoxyribonuclease I) (GenePool, Beijing, China, GPQ1801) in accordance with the provided instructions. Complementary deoxyribonucleic acid (cDNA) synthesis was achieved using an mRNA cDNA Synthesis Kit (GenePool, Beijing, China, GPQ1803) following the manufacturer's protocol.

For real-time PCR analysis, an mRNA/long ncRNA (lncRNA) quantitative PCR Kit (GenePool, Beijing, China, GPQ1808) was employed on a QuantStudio7 instrument (Thermo Fisher, Waltham, MA, USA, A43162). Each reaction, conducted in a total volume of 20 µL, included an initial denaturation at 95 °C for 30 seconds, followed by 45 cycles of amplification and quantification at 95 °C for 5 seconds and 60 °C for 30 seconds. Gene expression levels were normalized to the internal reference expression of glyceraldehyde-3-phosphate dehydrogenase. The relative mRNA levels were calculated using the $2^{-\Delta\Delta Ct}$ method. The primer sequences utilized are provided in **Supplementary Table 1**.

RNA Sequence Analysis

RNA sequencing was conducted on six normal and six pulmonary TB samples. The raw sequencing data, after filtering, were aligned to the mouse reference genome (GCF_000001635.27_GRCm39_genomic.fna) using Hisat2 version 2.2.1 (https://daehwankimlab.github.io/hisat2/). Subsequently, transcriptome quantification was performed using HTseq-count version 2.0.7 (https://pypi.python.org/pypi/HTSeq), yielding the count value for each gene.

The analysis of differentially expressed genes (DEGs) was carried out using the linear models for microarray data (limma) package [19]. The criteria for selecting DEGs between TB mice and controls included $|\log2(\text{fold change})| > 1$ and a p-value < 0.05. The outcomes of the DEG analysis were visually represented in a volcano plot using the "ggplot2" package. DEGs were categorized into long noncoding RNAs (lncRNAs) and messenger RNA (mRNA), denoted as differentially expressed lncRNAs (lnc_DEGs) and differentially expressed mRNAs (m_DEGs), respectively.

Proteome Analysis

The proteomic data consisted of 10 samples, including two biological replicates. The mean value of these replicates was considered the quantitative value for each sample. The dataset comprised six samples from normal mice and three from individuals with pulmonary TB. Notably, three of these samples were matched with the RNA sequencing samples.

Orthogonal Partial Least Squares Discriminant Analysis (OPLS-DA) was employed to identify significant independent variables distinguishing between TB and normal mice. The roPLS package facilitated this analysis, and variable important in projection (VIP) values were computed using the OPLS-DA difference test method [20]. VIP represents the variable weight in the OPLS-DA model, measuring the strength of influence and explanatory power of each protein/metabolite accumulation difference on the classification discrimination of sample groups. A commonly accepted screening criterion is VIP $\geq$1. Differentially expressed proteins (DEPs) between TB and control mice were identified based on the following selection criteria: |log2(fold change)| $>$1, p-value $<$ 0.05, and VIP $\geq$1.

To further understand the biological functions of the identified DEPs between TB and normal mice, Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses were conducted using the "clusterProfiler" package [21,22]. GO analysis focuses on the enrichment of biological processes (BPs), cell components (CCs), and molecular functions (MFs). The threshold for significance was set at a p-value of $<$0.05.

Metabolome Analysis

The metabolome data encompassed 12 samples, comprising 6 normal samples and 6 pulmonary TB samples, perfectly aligned with the RNA sequencing samples. The dataset comprised a total of 844 metabolites. Differentially expressed metabolites (DEMs) between TB and control mice were identified using the following criteria: |log2(fold change)| $>$1, p-value $<$ 0.05, and VIP $\geq$1.

Multiomics Analysis

The identification of candidate mRNA and protein candidates involved determining the intersection of upregulated differentially expressed mRNAs (m_DEGs) and DEPs, as well as the intersection of downregulated m_DEGs and DEPs.

For further analysis, Pearson correlation analysis was conducted between candidate proteins and differentially expressed metabolites (DEMs). Key metabolites were singled out based on the criteria of |correlation coefficient| $>$0.8 and a p-value $<$ 0.05. Subsequently, a network depicting the relationships between candidate proteins and key metabolites was constructed using Cytoscape [23].

Additionally, Pearson correlation analysis of the candidate mRNAs and lnc_DEGs was performed. Key lncRNAs were identified using the criteria of |correlation coefficient| $>$0.95 and a p-value $<$ 0.05.

Omics Analysis of Key Metabolites, lncRNAs, and mRNAs

The biological functions of the identified key metabolites were analyzed using the Metabolomics Analyst (Metabo-Analyst) online tool available at http://www.metaboanalyst.ca/. Subsequently, a random forest model was constructed based on the key metabolites, long non-coding RNAs (lncRNAs), and messenger RNAs (mRNAs) using the randomForest software package (https://www.stat.berkeley.edu/~breiman/RandomForests/). Random forests consist of classification and regression trees, simple models that employ binary splits on predictor variables to determine outcome predictions [24]. In this context, they predict the importance of key metabolites/lncRNAs or assess the contribution of key metabolites, lncRNAs, and mRNAs to prediction accuracy through the Gini index. The Gini score is calculated for every tree, averaged across the entire ensemble, and divided by the standard deviation over the entire ensemble. Hence, a higher Gini score indicates greater importance of the key metabolite, lncRNA, or mRNA.

Statistical Analysis

The data analysis was conducted using R software (version 4.1.0) (https://www.r-project.org/). A t-test was employed for comparisons between the two groups. Statistical significance was determined based on differences with a p-value $<$ 0.05.

Results

DEGs, DEPs, and DEMs between TB and Control Samples

The design flow chart of this study is shown in Fig. 1.

A comprehensive total of 1690 DEGs were identified as statistically significant between the TB and control samples. This set comprised 552 lnc_DEGs and 1011 m_DEGs. Among them, 476 genes were upregulated, while 1214 were downregulated in the TB group compared to the control group (Fig. 2A,a, **Supplementary Table 2**). The expression patterns of the top 100 DEGs in each sample are visually represented in a heatmap (Fig. 2A,b).

To identify DEPs, an OPLS-DA model was established to depict the relationship between protein expression and sample categories, successfully predicting the sample categories with a clear separation observed between TB and control samples (Fig. 2B,a). A total of 94 DEPs were identified, all of which exhibited upregulation in the TB group (Fig. 2B,b, **Supplementary Table 3**). The heatmap visualizes the expression of DEPs in TB and control samples (Fig. 2B,c), illustrating their elevated expression in the TB samples.

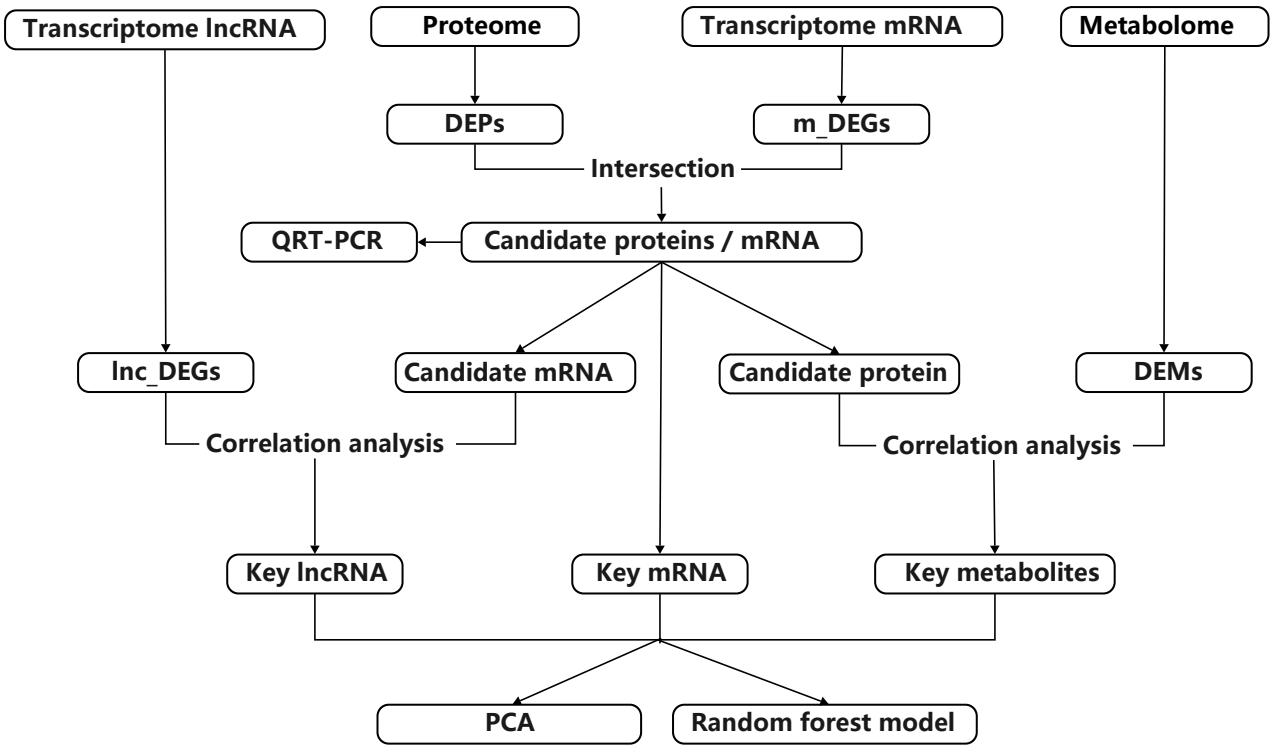

Fig. 1. Study design flowchart. First, we identified lnc_DEGs, m_DEGs, DEMs, and DEPS between TB and normal mice. The mRNA_DEGs and DEMs were intersected to obtain candidate proteins and mRNAs, and key metabolites were obtained based on the correlation between candidate proteins and DEPs. Then, key lncRNAs were determined based on the correlation between candidate mRNAs and lnc_DEGs. Finally, the resulting key metabolites, lncRNAs, and key mRNAs were combined to conduct PCA and build a random forest model. lncRNA, long non-coding RNA; mRNA, messenger RNA; DEPs, differentially expressed proteins; DEGs, differentially expressed genes; m_DEGs, differentially expressed mRNAs; qRT-PCR, quantitative reverse transcription polymerase chain reaction; lnc_DEGs, differentially expressed lncRNAs; DEMs, differentially expressed metabolites; PCA, Principal Component Analysis.

Functional analysis was conducted to explore the biological roles of DEPs. Notably, 40 KEGG pathways and 601 GO terms exhibited significant enrichment (**Supplementary Table 4**). DEPs were predominantly associated with immune-related KEGG pathways, including antigen processing and presentation, phagosomes, Epstein–Barr virus infection, human immunodeficiency virus 1 infection, cell adhesion molecules, and herpes simplex virus 1 infection (Fig. 2B,d). Consistent findings were observed in GO enrichment, with the top BPs linked to immunity, such as the response to interferon-gamma, antigen processing and presentation of exogenous antigens, antigen processing and presentation of peptide antigens, regulation of the response to biotic stimuli, adaptive immune response based on somatic recombination of immune receptors built from immunoglobulin superfamily domains, and the defense response to viruses (Fig. 2B,e).

Likewise, the relationship between metabolite expression and sample categories was modeled using OPLS-DA to predict the sample categories, resulting in a clear separation between TB and control samples (Fig. 2C,a). A total of 60 DEMs were identified between TB and control samples, consisting of 31 upregulated and 29 downregulated metabolites in TB samples relative to control samples (Fig. 2C,b, **Supplementary Table 5**). The differential abundances of DEMs in each TB or control sample are depicted in a heatmap (Fig. 2C,c).

Candidate Key Genes

Initially, the upregulated and downregulated m_DEGs were intersected with DEPs to identify candidate key genes and their corresponding proteins in TB. Ten candidate key genes and proteins were identified in TB, including beta-2 microglobulin (B2M), cluster of differentiation (CD)74, guanylate binding protein (GBP)2, GBP4, histocompatibility 2, class II antigen A, beta 1 (H2-Ab1), hematopoietic cell-specific Lyn substrate 1 (HCLS1), lymphocyte cytosolic protein 1 (LCP1), prosaposin (PSAP), sterile alpha motif domain and helical domain-containing protein 1 (SAMHD1), and signal transducer and activator of transcription 1 (STAT1) (Fig. 3A).

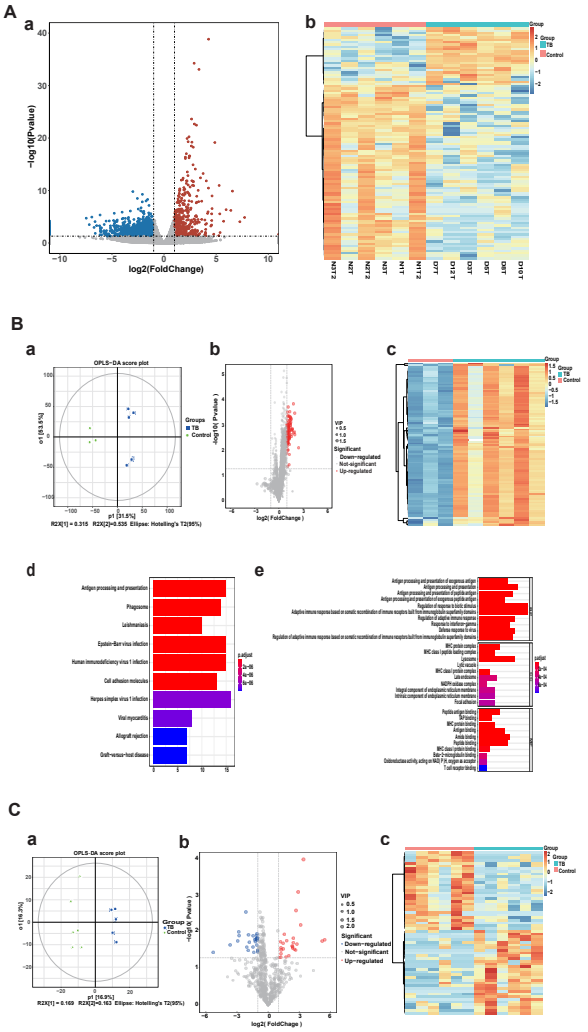

Fig. 2. Transcriptomic, proteomic, and metabolomic differential analyses. (A,a) Differentially expressed genes in TB and normal samples are shown in the volcano plot, with red dots representing significantly upregulated genes in TB samples, blue dots representing significantly downregulated genes, and gray dots representing non-significantly regulated genes. (A,b) Heatmap of the top 100 differentially expressed genes in TB and normal samples, with larger expression corresponding to the darker color, with red indicating high expression and blue indicating low expression. (B,a) OPLS-DA score chart where the abscissa shows the predicted principal components and the gap between groups; the ordinate shows the orthogonal principal components and the gap within the groups; the blue point indicates the TB sample, and the green point indicates the normal samples. (B,b) Differentially expressed proteins in TB and normal samples are shown in a volcano plot, with red dots representing significantly upregulated proteins in TB samples, blue dots representing significantly downregulated proteins, and gray dots representing non-significantly upregulated proteins. (B,c) Heatmap of differential protein expression in TB and normal samples, with higher expression corresponding to the darker color, with red indicating high expression and blue indicating low expression. (B,d) Top 10 KEGG pathways of the differentially expressed proteins. (B,e) Top 10 GO-BP, GO-CC, and GO-MF pathways of the differentially expressed proteins. (C,a) OPLS-DA score chart, where the abscissa shows the predicted principal components and the gap between groups; the ordinate shows the orthogonal principal components and the gap within the group; the blue point indicates the tuberculosis sample, and the green point indicates the normal samples. (C,b) Differentially expressed metabolites in TB and normal samples are shown in the volcano plot, with red dots representing significantly upregulated metabolites in TB samples, blue dots representing significantly downregulated metabolites, and gray dots representing metabolites that were not significantly regulated. (C,c) Heatmap of differential metabolite expression in TB and normal samples, with higher expression corresponding to the darker color, with red indicating high expression and blue indicating low expression. TB, tuberculosis; OPLS-DA, Orthogonal Partial Least Squares Discriminant Analysis; KEGG, Kyoto Encyclopedia of Genes and Genomes; GO, Gene Ontology; BP, biological process; CC, cell component; MF, molecular function; VIP, variable important in projection.

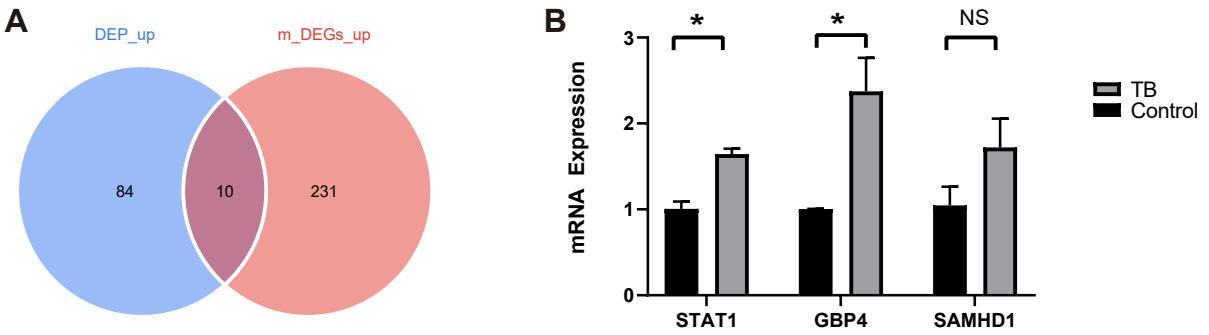

Fig. 3. Identification of candidate genes and their validation using qRT-PCR analysis. (A) Venn diagram of upregulated differential proteins and mRNAs. (B) qRT-PCR results showing the expression levels of *STAT1*, *GBP4*, and *SAMHD1* in TB patients were higher than those in healthy controls. TB, tuberculosis; DEG, differentially expressed gene; DEP, differentially expressed protein; up, upregulated; *STAT1*, signal transducer and activator of transcription 1; *GBP4*, guanylate binding protein 4; *SAMHD1*, sterile alpha motif domain and helical domain-containing protein 1; *, $p < 0.05$; NS, No significance.

To validate these findings, three randomly selected biomarkers, namely STAT1, GBP4, and SAMHD1, were assessed in TB samples using quantitative reverse transcription polymerase chain reaction (qRT-PCR). Consistent with the predictions, the expression levels of STAT1 ($p = 0.024$) and GBP4 ($p = 0.024$) in patients with TB were significantly higher than those in healthy controls (Fig. 3B). However, no significant difference in the expression levels of SAMHD1 ($p = 0.26$) was observed between the two groups.

Key Metabolites

The construction of a protein-metabolite network involved calculating the correlations between the 10 candidate proteins and 60 DEMs. Statistically significant proteins associated with TB were screened, resulting in the identification of eight known metabolites and one unknown metabolite, namely Analyte 127, Analyte 314, 3-methylglutaric acid, oxalacetic acid, N-methyl-L-glutamic acid 2, ciliatine, 2-deoxy-D-galactose 2, and indole-3-acetamide 1. These metabolites exhibited correlations with seven key proteins (Fig. 4A, **Supplementary Table 6**).

To unravel the biological functions of the key metabolites, they were subjected to enrichment analysis using the online tool MetaboAnalyst, revealing enrichment in 48 items (**Supplementary Table 7**). Key metabolites were primarily involved in the biosynthesis of primary bile acids, glycosaminoglycans, steroids, and unsaturated fatty acids, as well as in the metabolism of phenylalanine (Fig. 4B). The samples were further clustered based on the abundance of key metabolites, revealing consistent higher levels in control samples compared to TB samples (Fig. 4C). Notably, Analyte 127 was not identified and, consequently, not included in the heatmap. Additionally, the correlations among the seven key metabolites were calculated, revealing that Analyte 314 exhibited the most significant positive correlation with 3-methylglutaric acid (Fig. 4D).

Key lncRNAs

Correlations between the 10 candidate mRNAs and lnc_DEGs were computed, and statistically significant mRNAs associated with TB were screened (**Supplementary Table 8**). Eight known lncRNAs were identified: Gm36329, Gm31443, Gm14023, 9330175E14Rik, hyaluronan synthase 2, opposite strand (Has2os), serine carboxypeptidase 1, opposite strand (Scpep1os), Gm52174, and Gm46595. Clustering of samples based on the expression of key lncRNAs revealed a consistent and markedly higher expression in TB samples compared to control samples (Fig. 5A). Additionally, correlations among these eight key lncRNAs were calculated, highlighting that Gm14023 exhibited the most significant positive correlation with Has2os (Fig. 5B).

Combined Analysis of Key lncRNAs, mRNAs, and Metabolites

A co-expression network was established by analyzing the correlation between key metabolites, key lncRNAs, and key mRNAs (Fig. 6A). Principal Component Analysis (PCA) results of key metabolites, key lncRNAs, and key mRNAs (Fig. 6B) indicated a clear separation between TB and control samples. Key metabolites, along with their contribution degrees, were utilized to construct a random forest model. With approximately 10 decision tree data, as indicated in **Supplementary Table 9**, the overall forecast error rate was 0.0833, demonstrating the model's high accuracy in distinguishing normal samples from TB samples. The

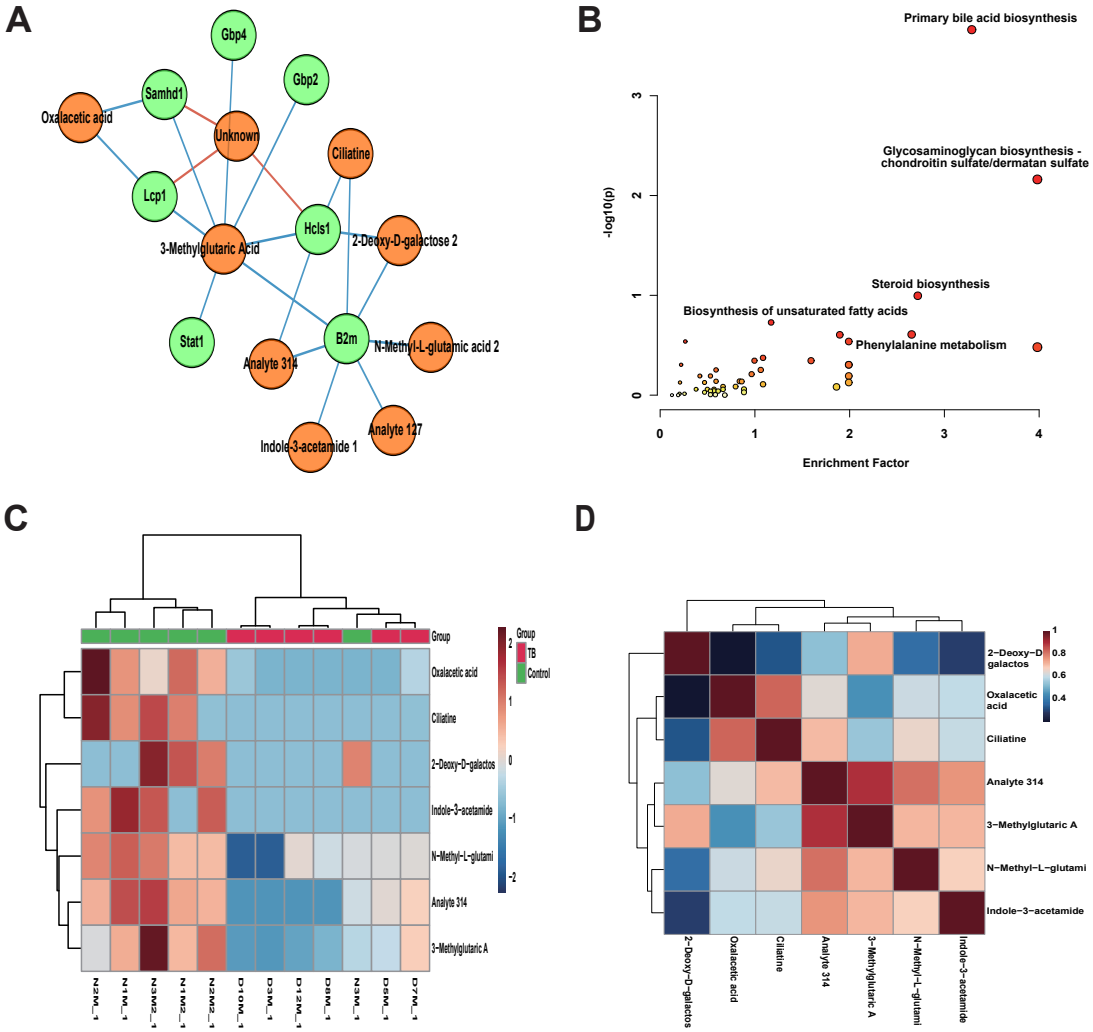

Fig. 4. Omics analysis of the key metabolites. (A) Network of candidate proteins and differential expression metabolites. The green ellipse indicates proteins, and the orange rectangle indicates metabolites. Lines indicate a correlation (red = positive, blue = negative). (B) Bubble diagram of the pathways enriched by key metabolites, where each point represents each enriched pathway and the abscissa represents the enrichment score; the greater the enrichment score, the higher the enrichment degree. (C) Heatmap of key metabolites, with larger expression corresponding to the darker color, with red indicating high expression and blue indicating low expression. (D) Heatmap of the key metabolite correlation where color indicates the magnitude of the correlation (red = high; blue = low).

contribution of these key metabolites, lncRNAs, and mRNAs to the model's classification accuracy was identified and visually displayed based on feature peaks and univariate statistical analysis (Fig. 6C).

Discussion

In this study, we employed transcriptome, proteome, and metabolome methods to unearth a panel of immune markers, comprising 10 mRNAs, 8 lncRNAs, and 8 metabolites. This comprehensive approach aimed to assess MTB infection and unveil the course of the disease.

Following MTB infection, a total of 1690 DEGs, 94 DEPs, and 60 DEMs were identified. Through the intersection of upregulated and downregulated mDEGs and DEPs, we identified ten candidate mRNAs/proteins associated with MTB infection, namely B2M, CD74, GBP2, GBP4, H2-Ab1, HCLS1, LCP1, PSAP, SAMHD1, and STAT1. These markers play diverse roles in various aspects of infection immunity. For instance, B2M, CD74, HCLS1, LCP1, PSAP, and STAT1 have been reported to regulate the immune response and inflammation in TB infection [25]. Additionally, HCLS1, LCP1, and STAT1 are associated with the cell cycle and proliferation, influencing

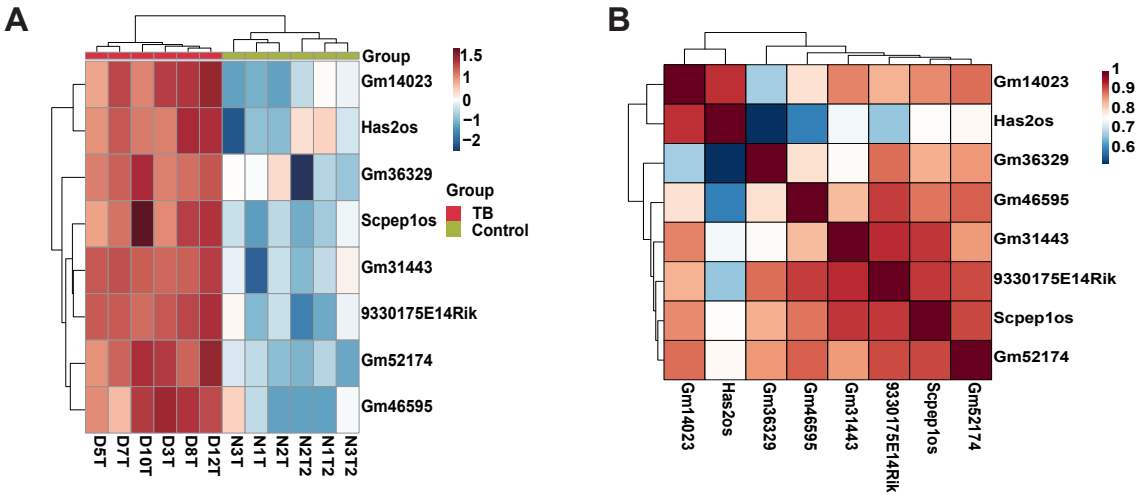

Fig. 5. Omics analysis of the key lncRNAs. (A) Heatmap of the key lncRNAs, where larger expression corresponds to a darker color, with red indicating high expression and blue indicating low expression. (B) Heatmap of the key lncRNA correlations, with color indicating the magnitude of the correlation; the correlation gradually decreases from red to blue.

T cell differentiation and the formation of immune memory. Furthermore, HCLS1, LCP1, and PSAP are involved in the cell death pathway, affecting the release of pathogens from cells infected with MTB [26]. Additionally, HCLS1, LCP1, GBP2, and GBP4 influence host defense against mycobacteria by affecting the cytoskeleton, cell motility, and production of antimicrobial peptides [26,27]. Moreover, HCLS1 and STAT1 interact with the Toll-like receptor pathway to recognize mycobacteria and elicit an immune response. B2M, CD74, and H2-Ab1 are involved in antigen presentation and CD4+ T cell activation, thereby influencing the host immune response to mycobacteria, while SAMHD1 is involved in autophagy by regulating nucleotide metabolism, maintaining intracellular homeostasis, and clearing pathogens [28].

From the 10 candidate mRNAs identified as key genes, correlation analysis with lncRNAs led to the identification of eight key lncRNAs, namely GM36329, GM31443, GM14023, 9330175E14Rik, Has2os, Scpep1os, GM52174, and GM46595. Additionally, correlation analysis between the 10 candidate proteins and DEPs revealed eight known metabolites associated with TB: Analyte 127, Analyte 314, 3-methylglutaric acid, oxalacetic acid, N-methyl-L-glutamic acid 2, ciliatine, 2-deoxy-D-galactose 2, and indole-3-acetamide 1. Notably, 2-deoxy-D-galactose 2 exhibited a significant correlation with the key proteins associated with pulmonary TB.

B2M, a component of the human leukocyte antigen β light chain, exhibits structural similarity to the immunoglobulin stable region and is widely distributed in human blood, urine, and cerebrospinal fluid. It actively participates in cellular immune processes *in vivo* [29]. B2M serves as a marker for evaluating immune activation in individuals with human immunodeficiency virus (HIV) infection. In HIV/TB co-infection cases, elevated levels of B2M have been observed compared to individuals with HIV infection alone and healthy controls. Additionally, it has been demonstrated that the interaction between early secreted antigenic target 6 kDa (ESAT-6) and B2M leads to a reduction in major histocompatibility complex class I antigens, contributing to the host's ability to limit the survival of MTB [30].

Among the 10 candidate mRNAs identified in this study, B2M held the top ranking. B2M transcripts showed associations with lncRNAs Gm14023, Has2os, and metabolites Analyte 127, Analyte 314, 3-methylglutamic acid, N-methyl-L-glutamic acid 2, ciliatine, 2-deoxy-D-galactose 2, and indole-3-acetamide 1, forming a distinct cluster. Furthermore, mRNA B2M was associated with lncRNA Gm46595 and metabolite N-methyl-L-glutamic acid 2, creating another cluster. The association of mRNA B2M with HCLS1 was also noted, linked with metabolite Analyte 314 and lncRNAs Scpep1os, Gm52174, and Gm46595, forming additional clusters. These findings suggest a crucial role for B2M in the context of MTB infection.

Among the screened mRNAs, STAT1 can effectively regulate cell growth, differentiation, apoptosis, and immune function [31]. It also plays an important role in immune defense against TB infection and is related to the interferon signaling pathway. Recent study has shown that STAT1 and its related molecules are potential biomarkers of MTB infection [32]. In our results, STAT1 clustered with the lncRNAs Gm31443 and 9330175E14Rik and the metabolite 3-methylglutamic acid, which helped to identify MTB infection.

In our study, 3-methylglutaric acid demonstrated the strongest association with mRNA in MTB-infected lung tissue. Notably, 3-methylglutaric acid metabolites exhibited

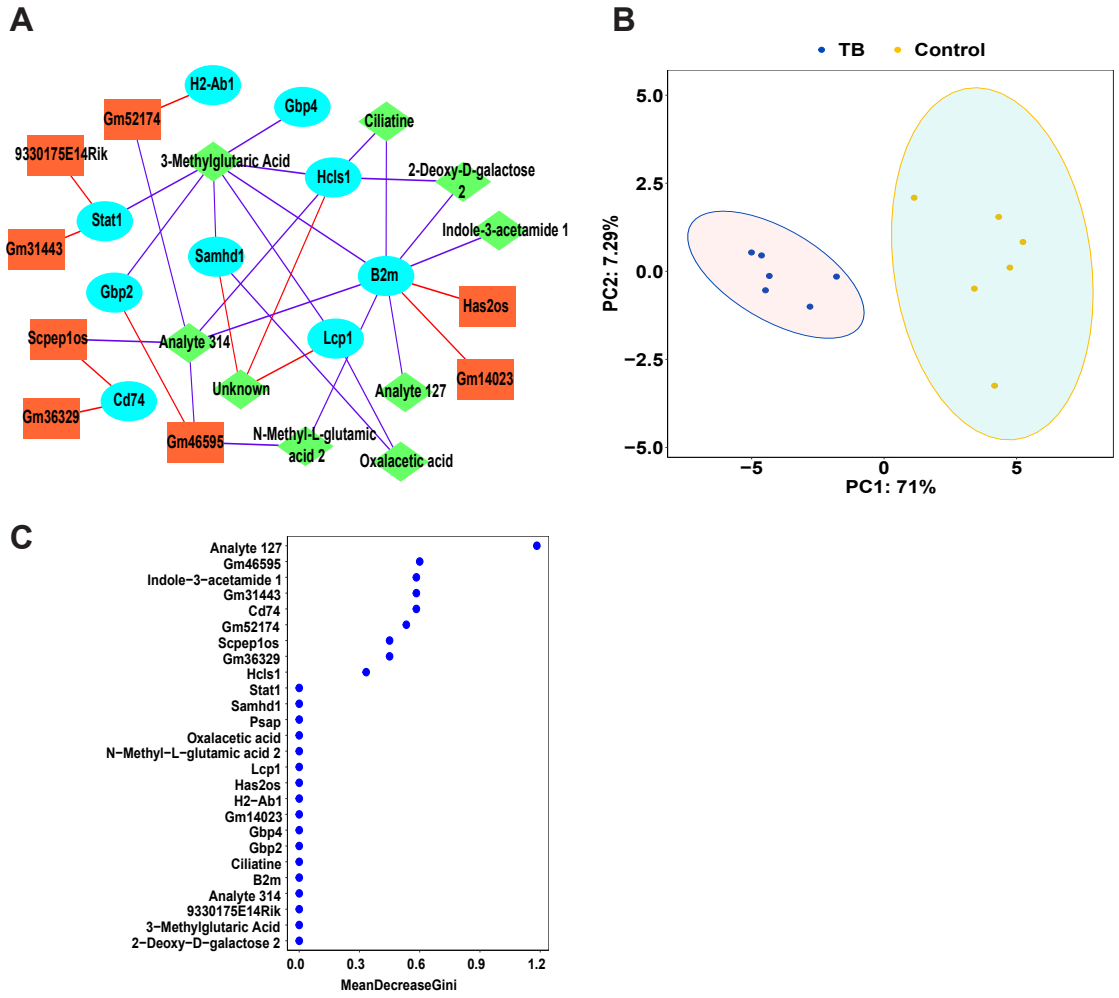

Fig. 6. Combined analysis of key lncRNAs, mRNAs, and metabolites. (A) The network of key mRNAs, lncRNAs, and metabolites. Blue ellipses indicate key mRNAs and orange rectangles indicate key lncRNAs, where lines indicate correlations, green diamonds indicate metabolites, red lines indicate positive correlations, and blue lines indicate negative correlations. (B) PCA score plots of key mRNAs, lncRNAs, and metabolites; blue dots indicate tuberculosis samples and orange dots indicate normal samples. (C) Key metabolites, lncRNAs, and mRNAs were used to construct a random forest model, with their respective contribution degrees.

significant correlations with seven mRNAs, namely B2M, GBP2, GBP4, HCLS1, LCP1, SAMHD1, and STAT1. As a dicarboxylic acid associated with leucine metabolism; 3-Methylglutaric acid accumulation in blood and urine is indicative of certain metabolic disorders [33], making it a potential biomarker for diagnosing and monitoring mitochondrial dysfunction. MTB infection may also affect 3-methylglutaric acid metabolism.

This study has certain limitations. The analysis focused on the infected areas of mouse lungs for multiomics analysis, which, while more sensitive to the host response, may not always yield a direct match in a readily available biological sample like blood. Additionally, the mouse model used in the experiments demonstrated the transcription of only a subset of core genes in a small number of human blood samples. The equivalence of response between the mouse model and human TB has not been thoroughly justified for biomarker research. The panel of immune biomarkers (10 mRNAs, 8 lncRNAs, and 8 metabolites) is restricted to MTB-infected and normal control mice, and further research is necessary to validate these findings in other sample types, such as blood, and in human patients with TB or those infected with MTB.

In MTB infection, alterations related to RNA transcription, protein translation, or metabolite production are not independent. The immune marker cluster, based on multiomics, assesses MTB infection and illuminates the disease's progression from diverse perspectives simultaneously. In comparison to pinpointing a singular biomarker, immune marker clusters offer greater precision, thereby enhancing the diagnosis of MTB infection and tracking disease progression.

Conclusion

In this study, a collection of immune marker panels (comprising 10 mRNAs, 8 metabolites, and 8 lncRNAs) linked to MTB infection was identified using marker panel analysis in a mouse model of MTB infection, drawing on mRNA, proteins, metabolites, and lncRNAs. Employing PCA and a random forest model, TB and normal samples demonstrated clear differentiation (with an error rate of 8.33%). This holds substantial promise for diagnosing MTB infection in humans based on a multiomics-based biomarker cluster.

Availability of Data and Materials

The raw sequence data reported in this paper have been deposited in the Genome Sequence Archive (Genomics, Proteomics & Bioinformatics 2021) at the National Genomics Data Center (Nucleic Acids Res 2022), China National Center for Bioinformation/Beijing Institute of Genomics, Chinese Academy of Sciences (GSA: CRA012503), and are publicly accessible at https://ngdc.cncb.ac.cn/gsa. The data reported in this paper have been deposited in the OMIX, China National Center for Bioinformation/Beijing Institute of Genomics, Chinese Academy of Sciences (https://ngdc.cncb.ac.cn/omix: accession numbers OMIX004924 and OMIX004927).

Author Contributions

Conceptualization, HZ and LB; methodology, HZ, MS, NZ, and XW; validation, MS, NZ, and XW; formal analysis, HZ, FR, DL, LY, and JL; resources, YL and CL; writing-original draft preparation, HZ and MS; writing-review and editing, HZ; supervision, ZW and LB; project administration, ZW and LB; funding acquisition, YL, CL, DL, ZW and LB; acquisition of data, FR and YL; analysis and interpretation of data, CL and ZW. All authors have read and agreed to the submitted version of the manuscript. All authors contributed to significant editorial changes in the manuscript. All authors have participated sufficiently in the work and agreed to be accountable for all aspects of the work.

Ethics Approval and Consent to Participate

This study received approval from the Animal Experiment Ethics Committee of Beijing Chest Hospital (No. XK2023-040). The use of deidentified patient information and the collection of clinical samples were approved by the Ethics Committee of the Beijing Chest Hospital, Capital Medical University (Beijing, China) (No. BJCH-2019-0301-018). The research follows international and national regulations in accordance with the Declaration of Helsinki or any other relevant set of ethical principles. All study participants provided written informed consent.

Acknowledgment

We gratefully acknowledge the help of Miao Tian (Biobank of Beijing Chest Hospital, Beijing Tuberculosis and Thoracic Tumor Research Institute/Beijing Chest Hospital, Capital Medical University, Tongzhou District, Beijing, China) for technical help in mice experiments with MTB infection.

Funding

This work was supported by the “Dengfeng Plan” High-level Hospital Construction Opening Project of Foshan Fourth People’s Hospital [grant number FSSYKF-2020007 to LB]; The Strategic Priority Research Program of the Chinese Academy of Sciences [grant number XDB29020000 to LB]; the Key Research and Development Program of Xinjiang Uygur Autonomous Region [grant number 2022B03032-3 to LB]; The Capital Medical Development Special Foundation [grant number 2020-2-1042 to YL]; The Capital Health Research and Department of Special [grant number 2022-1G-3012 to CL]; The Scientific Research Project of Liaoning Education Department [grant number 2021(LJKZ0444) to DL].

Conflict of Interest

Jinlong Li is the consultant of TB Healthcare Co., Ltd. All other authors have reported no conflicts relevant to the contents of this paper to disclose.

Supplementary Material

Supplementary material associated with this article can be found, in the online version, at https://doi.org/10.24976/Discov.Med.202436185.117.

References

[1] World Health Organization. Global tuberculosis report 2022. World Health Organization: Geneva. 2022.

[2] World Health Organization. WHO consolidated guidelines on tuberculosis: tuberculosis preventive treatment: Module 1: prevention. World Health Organization: Geneva. 2020.

[3] Calderwood CJ, Wilson JP, Fielding KL, Harris RC, Karat AS, Mansukhani R, *et al*. Dynamics of sputum conversion during effective tuberculosis treatment: A systematic review and meta-analysis. PLoS Medicine. 2021; 18: e1003566.

[4] Khabibullina NF, Kutuzova DM, Burmistrova IA, Lyadova IV. The Biological and Clinical Aspects of a Latent Tuberculosis Infection. Tropical Medicine and Infectious Disease. 2022; 7: 48.

[5] Qi J, Fan W. Study on the value of molecular biology combined with liquid MGIT culture method in clinical examination of mycobacterium tuberculosis. American Journal of Translational Research. 2021; 13: 9757–9763.

[6] Martinez L, Andrews JR. Improving Tuberculosis Case Finding in Persons Living with Advanced HIV through New Diagnostic Algorithms. American Journal of Respiratory and Critical Care Medicine. 2019; 199: 559–560.

[7] Pahal P, Pollard EJ, Sharma S. PPD Skin Test. StatPearls Publishing: Treasure Island (FL). 2023.

[8] Benachinmardi K, Sampath S, Rao M. Evaluation of a new interferon gamma release assay, in comparison to tuberculin skin tests and quantiferon tuberculosis goldplus for the detection of latent tuberculosis infection in children from a high tuberculosis burden setting. International Journal of Mycobacteriology. 2021; 10: 142–148.

[9] Steingart KR, Flores LL, Dendukuri N, Schiller I, Laal S, Ramsay A, *et al*. Commercial serological tests for the diagnosis of active pulmonary and extrapulmonary tuberculosis: an updated systematic review and meta-analysis. PLoS Medicine. 2011; 8: e1001062.

[10] Zhang X, Chen C, Xu Y. Long Non-coding RNAs in Tuberculosis: From Immunity to Biomarkers. Frontiers in Microbiology. 2022; 13: 883513.

[11] Salgado-Bustamante M, Rocha-Viggiano AK, Rivas-Santiago C, Magaña-Aquino M, López JA, López-Hernández Y. Metabolomics applied to the discovery of tuberculosis and diabetes mellitus biomarkers. Biomarkers in Medicine. 2018; 12: 1001–1013.

[12] Rusk N. Multi-omics single-cell analysis. Nature Methods. 2019; 16: 679.

[13] Yan Y, Zhou B, Lee YJ, You S, Freeman MR, Yang W. Box-Car and shotgun proteomic analyses reveal molecular networks regulated by UBR5 in prostate cancer. Proteomics. 2022; 22: e2100172.

[14] Bojkova D, Klann K, Koch B, Widera M, Krause D, Ciesek S, *et al*. Proteomics of SARS-CoV-2-infected host cells reveals therapy targets. Nature. 2020; 583: 469–472.

[15] El-Hawary SS, Rabeh MA, Raey MAE, El-Kadder EMA, Sobeh M, Abdelmohsen UR, *et al*. Metabolomic profiling of three Araucaria species, and their possible potential role against COVID-19. Journal of Biomolecular Structure & Dynamics. 2022; 40: 6426–6438.

[16] Alsulaimany FA, Zabermawi NMO, Almukadi H, Parambath SV, Shetty PJ, Vaidyanathan V, *et al*. Transcriptome-Based Molecular Networks Uncovered Interplay Between Druggable Genes of $CD8^+$ T Cells and Changes in Immune Cell Landscape in Patients with Pulmonary Tuberculosis. Frontiers in Medicine. 2022; 8: 812857.

[17] Penn-Nicholson A, Hraha T, Thompson EG, Sterling D, Mbandi SK, Wall KM, *et al*. Discovery and validation of a prognostic proteomic signature for tuberculosis progression: A prospective cohort study. PLoS Medicine. 2019; 16: e1002781.

[18] Cho Y, Park Y, Sim B, Kim J, Lee H, Cho SN, *et al*. Identification of serum biomarkers for active pulmonary tuberculosis using a targeted metabolomics approach. Scientific Reports. 2020; 10: 3825.

[19] Liu S, Wang Z, Zhu R, Wang F, Cheng Y, Liu Y. Three Differential Expression Analysis Methods for RNA Sequencing: limma, EdgeR, DESeq2. Journal of Visualized Experiments: JoVE. 2021.

[20] Thévenot EA, Roux A, Xu Y, Ezan E, Junot C. Analysis of the Human Adult Urinary Metabolome Variations with Age, Body Mass Index, and Gender by Implementing a Comprehensive Workflow for Univariate and OPLS Statistical Analyses. Journal of Proteome Research. 2015; 14: 3322–3335.

[21] Gene Ontology Consortium, Aleksander SA, Balhoff J, Carbon S, Cherry JM, Drabkin HJ, *et al*. The Gene Ontology knowledgebase in 2023. Genetics. 2023; 224: iyad031.

[22] Kanehisa M, Furumichi M, Sato Y, Kawashima M, Ishiguro-Watanabe M. KEGG for taxonomy-based analysis of pathways and genomes. Nucleic Acids Research. 2023; 51: D587–D592.

[23] Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, *et al*. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Research. 2003; 13: 2498–2504.

[24] Speiser JL, Miller ME, Tooze J, Ip E. A Comparison of Random Forest Variable Selection Methods for Classification Prediction Modeling. Expert Systems with Applications. 2019; 134: 93–101.

[25] O'Garra A, Redford PS, McNab FW, Bloom CI, Wilkinson RJ, Berry MPR. The immune response in tuberculosis. Annual Review of Immunology. 2013; 31: 475–527.

[26] Zhou JY, Szasz TP, Stewart-Hutchinson PJ, Sivapalan J, Todd EM, Deady LE, *et al*. L-Plastin promotes podosome longevity and supports macrophage motility. Molecular Immunology. 2016; 78: 79–88.

[27] Srivastava S, Ernst JD, Desvignes L. Beyond macrophages: the diversity of mononuclear cells in tuberculosis. Immunological Reviews. 2014; 262: 179–192.

[28] Xu B, Sui Q, Hu H, Hu X, Zhou X, Qian C, *et al*. SAMHD1 Attenuates Acute Inflammation by Maintaining Mitochondrial Function in Macrophages via Interaction with VDAC1. International Journal of Molecular Sciences. 2023; 24: 7888.

[29] Jha V, Rameshwaram NR, Janardhan S, Raman R, Sastry GN, Sharma V, *et al*. Uncovering Structural and Molecular Dynamics of ESAT-6:β2M Interaction: Asp53 of Human β2-Microglobulin Is Critical for the ESAT-6:β2M Complexation. Journal of Immunology (Baltimore, Md.: 1950). 2019; 203: 1918–1929.

[30] Dominici R, Finazzi D, Polito L, Oldoni E, Bugari G, Montanelli A, *et al*. Comparison of β2-microglobulin serum level between Alzheimer's patients, cognitive healthy and mild cognitive impaired individuals. Biomarkers: Biochemical Indicators of Exposure, Response, and Susceptibility to Chemicals. 2018; 23: 603–608.

[31] Butturini E, Carcereri de Prati A, Mariotto S. Redox Regulation of STAT1 and STAT3 Signaling. International Journal of Molecular Sciences. 2020; 21: 7034.

[32] Yi XH, Zhang B, Fu YR, Yi ZJ. STAT1 and its related molecules as potential biomarkers in Mycobacterium tuberculosis infection. Journal of Cellular and Molecular Medicine. 2020; 24: 2866–2878.

[33] Jones DE, Perez L, Ryan RO. 3-Methylglutaric acid in energy metabolism. Clinica Chimica Acta; International Journal of Clinical Chemistry. 2020; 502: 233–239.