BMC Infectious Diseases

RESEARCH

Open Access

Identification of discriminative plasma protein biomarkers for recent *Mycobacterium tuberculosis* infection: a pilot data-independent acquisition mass spectrometry-based proteomics study

Yutong Han[2,3†], Jinyan Shi[4†], Qiao Liu[1], Yan Shao[1], Limei Zhu[1], Leonardo Martinez[5], Biao Xu[2,3*] and Chen Cheng[1*]

Abstract

Background Recently acquired *Mycobacterium tuberculosis* (*M.tb*) infection is associated with a higher risk of progression to active tuberculosis (TB), yet current diagnostic tools cannot distinguish recent from remote latent TB infection (LTBI).

Methods Plasma samples from 26 individuals with recently acquired LTBI, 26 with remotely acquired LTBI and 26 bacteriologically confirmed TB patients, were analyzed using a data-independent acquisition mass spectrometry (DIA-MS)-based proteomics approach. Differentially expressed proteins (DEPs) and enriched biological pathways associated with infection statuses were identified.

Results Principal component analysis revealed that individuals with recently acquired LTBI and those with active TB exhibited comparable proteomic profiles, while the group with remotely acquired LTBI demonstrated significantly distinct expression patterns. Differential abundance analysis revealed 470 DEPs between the remotely and recently acquired LTBI groups, and 171 between the recent LTBI and active TB groups. These dysregulated proteins were primarily enriched in pathways related to complement and coagulation cascades, cytoskeletal remodeling, and cell adhesion. By integrating hub proteins from both DEPs and WGCNA modules, and utilizing LASSO-based feature selection, a five-protein panel was identified, comprising ACTR3, ITGB2, RELN, TUBB1, and ITGB5. This panel demonstrated a robust capacity to differentiate between infection statuses, with AUC values exceeding 0.90 across all pairwise group comparisons. Furthermore, the identified signature also displayed significant differential expression across four independent transcriptomic datasets.

†Yutong Han and Jinyan Shi contributed equally to this work.

*Correspondence:
Biao Xu
bxu@shmu.edu.cn
Chen Cheng
chencheng128@gmail.com

Full list of author information is available at the end of the article

Conclusion These findings highlight the potential utility of plasma protein signatures in identifying individuals with recently acquired LTBI, providing a rationale for risk stratification to improve targeted LTBI interventions.

Clinical trial number Not applicable.

Keywords Tuberculosis, Latent tuberculosis infection, Proteomics, Biomarker, Mass spectrometry

Introduction

Tuberculosis (TB) remains a significant global health concern, causing over 1 million deaths annually [1]. Approximately one-quarter of the world's population is estimated to be infected with *Mycobacterium* tuberculosis (*M.tb*), referred to as latent tuberculosis infection (LTBI) [2]. Although the lifetime risk of reactivation in individuals with LTBI is estimated at 5–10%, recent evidence indicated that the risk is highest during the first few years after infection, compared to those with established remote LTBI [3]. Notably, recent acquisition of *M.tb* infection is a pivotal risk factor for progression to active disease among otherwise healthy individuals with LTBI [4, 5]. However, current LTBI diagnostics, including the tuberculin skin test (TST) and interferon-γ release assays (IGRAs), assess host immune memory, resulting in persistently positive results irrespective of the infection timeline [6, 7]. Furthermore, both TST and IGRA are prone to unexplained conversion and reversion over time, making serial test results difficult to interpret [8]. Such diagnostic limitations not only hinder the implementation of target interventions to disrupt community transmission chains, but also lead to unnecessary TB preventive therapy (TPT) [9]. There is an urgent need for a robust diagnostic capable of distinguishing recently infected LTBI individuals, thereby enabling risk-stratified TPT and advancing the WHO End TB Strategy [10].

In light of these diagnostic limitations, multiple studies have sought to characterize host–based signatures that could serve as indicators of recent *M.tb* acquisition [9]. Several immunological markers, such as HLA-DR expression on $CD4^+$ T cells and TNF-α–only effector T cell subsets, have been proposed as potential indicators of recent infection [11–13]. Latency antigens, including Rv2626c and Rv2628, have also demonstrated promise in distinguishing recent from remote LTBI based on antigen-specific IFN-γ responses [14, 15]. In parallel, molecular approaches utilizing blood transcriptomic profiling have gained momentum. A study by Kwan et al. identified 186 gene signatures that were differentially expressed between household contacts of TB patients and healthy individuals without recent exposure, underscoring the feasibility of transcriptomic tools to detect recent *M.tb* infection events [16]. Another study revealed that transcriptomic responses in murine and macaque models may reflect time since infection, underscoring the clinical potential of molecular methods in providing valuable insights for recent infection identification [17].

While these immune- and transcriptome-based approaches have advanced insights into LTBI heterogeneity and show promise for identifying recent *M.tb* infection, their translational applicability remains limited. These assays often rely on the isolation of peripheral blood mononuclear cells or techniques such as flow cytometry and cytokine assays, which require stringent pre-analytical conditions, specialized laboratory infrastructure, and complex analytical workflows. Such requirements limit their applicability for large-scale population screening and point-of-care test (POCT) application, particularly in resource-limited, high-burden settings [9].

In contrast, plasma protein quantification offers a more practical and scalable alternative. Protein-based biomarkers are well-suited to POCT development, as exemplified by lateral flow tests utilizing capillary blood obtained via finger prick [18], offering a promising avenue for the development of practical diagnostics to identify individuals with recently acquired *M.tb* infection. Indeed, profound changes in abundance of many plasma proteins have been reported in TB patients [19, 20]. However, to date, few studies have leveraged high-throughput, untargeted plasma proteomics to characterize protein expression differences associated with the timing of *M.tb* infection.

In this study, we applied data-independent acquisition mass spectrometry (DIA-MS) -based proteomics to profile plasma samples from individuals with recently acquired LTBI, remotely acquired LTBI, and active TB. We aimed to identify candidate biomarkers indicative of recent infection and to explore underlying biological mechanisms, thereby providing insights into risk-stratified TB preventive strategies.

Methods

Study participants and sample collection

This study enrolled participants from Jiangsu Province, China, including individuals with active TB, recently acquired LTBI, and remotely acquired LTBI. Active TB cases were bacteriologically confirmed through positive culture of *M.tb* from sputum samples and were recruited from the Fourth People's Hospital of Lianyungang in 2022. LTBI individuals were identified based on positive QuantiFERON-TB Gold in Tube (QFT) test results

(TB Ag–Nil > 0.7 IU/ml, which was considered a strong positive result), and in the absence of symptoms or other clinical signs of TB and with normal chest radiography. Participants were excluded if they were pregnant, under 18 years of age, had liver or renal dysfunction, or were experiencing an acute exacerbation of a chronic disease.

Individuals with LTBI were grouped into recently acquired or remotely acquired LTBI based on epidemiological exposure history to *M.tb.*

- **Recently Acquired LTBI**: Individuals were considered to have recently acquired LTBI if they had household close contact with a bacteriologically confirmed pulmonary TB case within 3 months prior to recruitment. These participants were identified through contact investigations conducted between 2022 and 2023 in Jiangsu Province, in accordance with the Chinese National Tuberculosis Control Guidelines. All participants had close and prolonged contact with the index case during the defined exposure window, such as sharing the same living space, and none had a prior history of TB diagnosis or treatment.
- **Remotely Acquired LTBI**: Individuals were considered to have remotely acquired LTBI if they had no known history of contact with active TB cases and tested QFT-positive during a routine health examination in 2023 in Danyang, Jiangsu Province. All participants had continuously resided in Danyang, a region where TB incidence markedly declined over the past decade (from approximately 60 to 20 cases per 100,000) [21].

In total, 26 bacteriologically positive TB cases, 26 individuals with recently acquired LTBI and 26 with remotely acquired LTBI were included. Among the remotely acquired LTBI individuals, 21 individuals had also tested QFT-positive in a large-scale LTBI prevalence survey conducted in the same region in 2013. The demographic characteristics of the participants are presented in Table S1 (Appendix, Table S1). And the individual QFT assay results for all LTBI participants are provided in Table S2 (Appendix, Table S2). For each participant, 3–5 ml of peripheral blood was drawn before any treatment. The whole blood collected in EDTA tubes was centrifuged to extract plasma, which was then divided into storage tubes and stored at -80 °C.

Liquid phase separation and mass spectrometry

Plasma samples were thawed on ice and depleted of high-abundance proteins using High Select™ Top14 resin. Proteins were denatured in 8 M urea, reduced with 10 mM DTT (37 °C, 30 min), and alkylated with 20 mM IAM (room temperature, 30 min, dark), followed by buffer exchange to 50 mM ammonium bicarbonate. Proteins were digested sequentially with LysC (1 µg, 37 °C, 2 h) and trypsin (1 µg, 37 °C, overnight). Digestion was quenched with 10% TFA. Peptides were desalted using a SOLAµ HRP plate, eluted with 80% ACN/0.1% TFA, dried, and reconstituted in 0.1% formic acid to 0.5 µg/µL. Two micrograms of peptides were injected for LC–MS/MS analysis. Equal amounts of peptides from all samples were pooled as QC to monitor instrument performance. LC-MS/MS data were acquired using an Ultimate™ 3000RSLC system and Q Exactive HF-X mass spectrometer. Peptide separation employed an Acclaim PepMap RSLC C18 analytical column (2 µm, 100 Å, 75 µm i.d. × 25 cm) with a 65-minute gradient at a 400 nL/min flow rate. Mass spectrometry was performed in data-independent acquisition (DIA) mode, with parameters detailed in Table S3 (Appendix, Table S3). The resulting DIA files were processed using DIA-NN v1.8.0 software, utilizing default parameter settings. A comprehensive plasma proteomic spectral library was used for peptide and protein identification, and the false discovery rate (FDR) for peptide and protein identification was set to 1%. Proteins with at least one matched peptide were considered identified, consistent with previous DIA-based studies [22–25]. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium (https://proteomecentral.proteomexchange.org) via the iProX partner repository [26, 27] with the dataset identifier PXD074267.

Bioinformatics and statistics analysis

Proteomic data pre-processing and differential expression analysis

For processing and analyzing protein abundance data, we employed a standardized workflow based on the 'DEP2' R package, specifically designed for mass spectrometry-based quantitative proteomics data [28]. Proteins detected in at least 70% of samples within each group were retained for further analysis. The data were log2-transformed and normalized using the variance stabilizing normalization method. Missing values were imputed using a random forest-based approach. Principal component analysis (PCA) was conducted to assess sample clustering and variability. Differential expression analysis was performed for pairwise group comparisons [29]. Proteins with a fold change (FC) > 1.5 or < 1/1.5 and a false discovery rate (FDR) < 0.05 were considered differentially expressed proteins (DEPs).

Functional enrichment analysis

Functional annotation and pathway analysis for differentially expressed proteins (DEPs) were performed using

Gene Ontology (GO) enrichment analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway annotation via the cluster Profiler R package [30]. Terms with an FDR-adjusted *P* value below 0.05 were considered significant.

Protein–protein interaction (PPI) network construction

The protein–protein interaction (PPI) network for DEPs was constructed using the STRING database (https://string-db.org), with an interaction confidence score ≥ 0.4 as the cutoff. The network was visualized using Cytoscape software (3.10.2) and top hub proteins were identified using the MCC algorithm implemented in the CytoHubba plugin.

Weighted correlation network analysis (WGCNA)

WGCNA was performed using the WGCNA package in R to identify clusters of highly correlated proteins and explore their associations with phenotypic traits [31]. Modules were defined using dynamic tree cutting and merged based on eigengene similarity. Hub proteins within each module were identified based on high module membership (|MM| > 0.8) and high gene significance (|GS| > 0.5).

Machine learning analysis for feature selection

Feature selection was performed using the least absolute shrinkage and selection operator (LASSO) regression model. Nested cross-validation was employed for feature selection and model performance evaluation. A 10-fold outer cross-validation split the dataset into training and testing folds, while an inner cross-validation within the training set optimized the LASSO penalty parameter (λ). Proteins with nonzero coefficients in each outer fold were recorded, and features selected in at least 7 out of 10 folds were retained to construct the final biomarker panel. To ensure comparability across features, all protein features were standardized (z-scored) prior to modelling.

To account for potential multicollinearity among the proteins, an Elastic Net regularized multinomial classification model was subsequently constructed. Elastic Net provides more stable coefficient estimation for correlated predictors while retaining a degree of sparsity. Hyperparameters were optimized using leave-one-out cross-validation (LOOCV) with a predefined tuning grid, in which the mixing parameter α ranged from 0 to 1 in increments of 0.1, and the regularization parameter λ was explored over 20 logarithmically spaced values from 10^{-3} to 1. The optimal model was obtained at $\alpha = 0.1$ and $\lambda = 0.026$. The multinomial model was implemented using type. multinomial = "grouped" in glmnet, applying a grouped penalty across outcome classes to ensure consistent feature selection for all three groups. It should be noted that LASSO-based feature selection was restricted to the predefined candidate protein set and performed strictly within the training folds of the nested cross-validation framework, without any proteome-wide filtering during cross-validation.

Model discriminative performance was evaluated by computing pairwise area under the receiver operating characteristic curve (AUC) values between each pair of groups, along with 95% confidence intervals (CIs) estimated using the DeLong method. All analyses were performed using the glmnet, pROC, and caret packages in R.

Transcriptomic validation using public TB datasets

To validate the protein signature, we analyzed four independent transcriptomic datasets from the curatedTBData R package (GSE101705, GSE152218, GSE28623, and GSE19444). These datasets included individuals with LTBI who were either recent immigrants from TB-endemic regions or close contacts of active TB patients, and are therefore considered at higher risk of recent *M.tb* infection. In each dataset, we first standardized gene expression values using Z-score transformation and assessed the differential expression of individual candidate genes between LTBI and active TB groups using the Wilcoxon rank-sum test. Additionally, single-sample gene set enrichment analysis (ssGSEA) was performed to quantify the overall enrichment of the protein signature in each sample, and group-level differences in enrichment scores were similarly evaluated [32].

Results

Proteomic profiling and comparative differential protein analysis

A total of 2,122 proteins were identified across all samples, among which 1,570 proteins detected in more than 70% of the samples. PCA revealed markedly distinct protein abundance patterns between remotely acquired LTBI and the other two groups, while a substantial overlap was observed between recently acquired LTBI and TB groups (Fig. 1A).

Comparative proteomics analysis identified 470 DEPs (233 up-, 237 down-regulated) between remotely versus recently acquired LTBI, and 171 DEPs (93 up-, 78 down-regulated) between recently acquired LTBI and active TB, indicating a progressive narrowing of proteomic differences from remote infection through recent infection to active disease (Fig. 1B). Furthermore, 42 DEPs were found to be common across all three pairwise comparisons (Fig. 1C).

Functional enrichment and protein-protein interaction analysis

To characterize the biological functions and interactions of DEPs, functional enrichment analysis (GO and KEGG) and PPI network construction were performed.

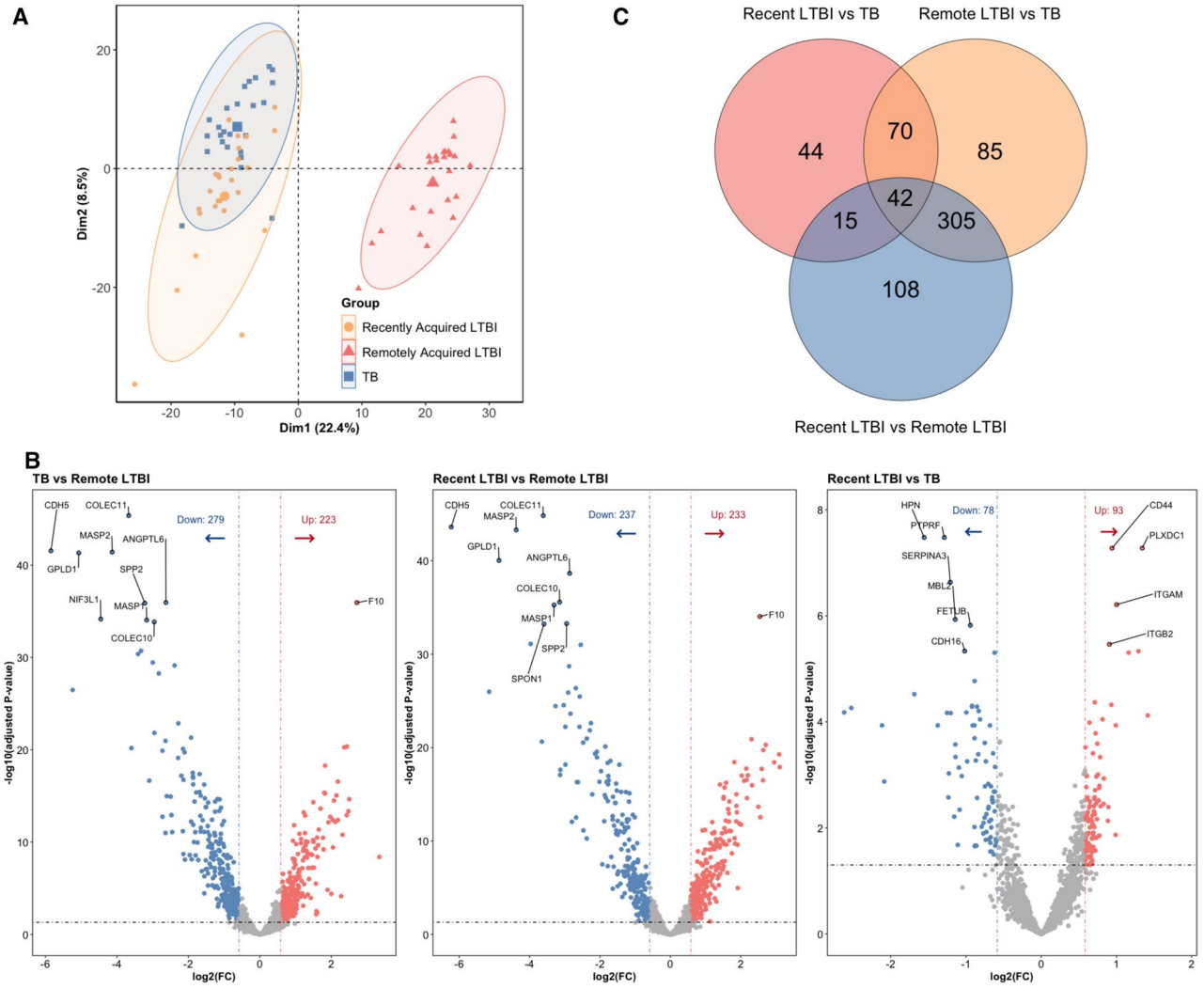

Fig. 1 Comparative analysis of plasma protein expression profiles across different participant groups. (**A**) Principal component analysis (PCA) reveals distinct clustering of plasma proteomic profiles among individuals with recently acquired LTBI, remotely acquired LTBI, and active TB. (**B**) Volcano plots showing differentially expressed proteins in pairwise comparisons: TB vs. remote LTBI (left), recent vs. remote LTBI (middle), and recent LTBI vs. TB (right). Significantly upregulated (red) and downregulated (blue) proteins are annotated. (**C**) Venn diagram displaying the overlap of differentially expressed proteins among the three pairwise comparisons

KEGG analysis identified complement and coagulation cascades, focal adhesion, regulation of actin cytoskeleton, phagosome, and ECM-receptor interaction pathways as consistently enriched across all pairwise comparisons, indicating that these pathways play essential roles in host responses to *M.tb* infection (Fig. 2A). GO biological process enrichment revealed distinct patterns among the top enriched terms across group comparisons, with regulation of body fluid levels enriched in recent vs. remote LTBI, complement activation in TB vs. remote LTBI, and cell-substrate adhesion in TB vs. recent LTBI (Fig. 2B). A PPI network constructed from 42 common DEPs identified hub proteins such as TLN1, ITGB2, ITGB5, FERMT3, and TUBA4A (Fig. 2C and D), which are mainly involved in integrin signaling, cytoskeletal remodeling, and immune cell migration, consistent with the biological processes highlighted by GO and KEGG enrichment analyses (Appendix, Table S4).

Identification of trait-associated modules and selection of hub proteins via WGCNA

A weighted protein co-expression network was constructed using WGCNA. A soft-thresholding power of $\beta = 3$ (scale-free $R^2 = 0.97$) was selected to ensure a scale-free network topology (Appendix, Figure S1A). Hierarchical clustering identified multiple modules of co-expressed proteins (Appendix, Figure S1B). A total of 11 modules were identified using hierarchical clustering and dynamic tree cutting (Appendix, Fig. S1C). Module–trait correlation analysis showed strong associations of the blue and turquoise modules with both recently and remotely acquired LTBI, while the brown and black

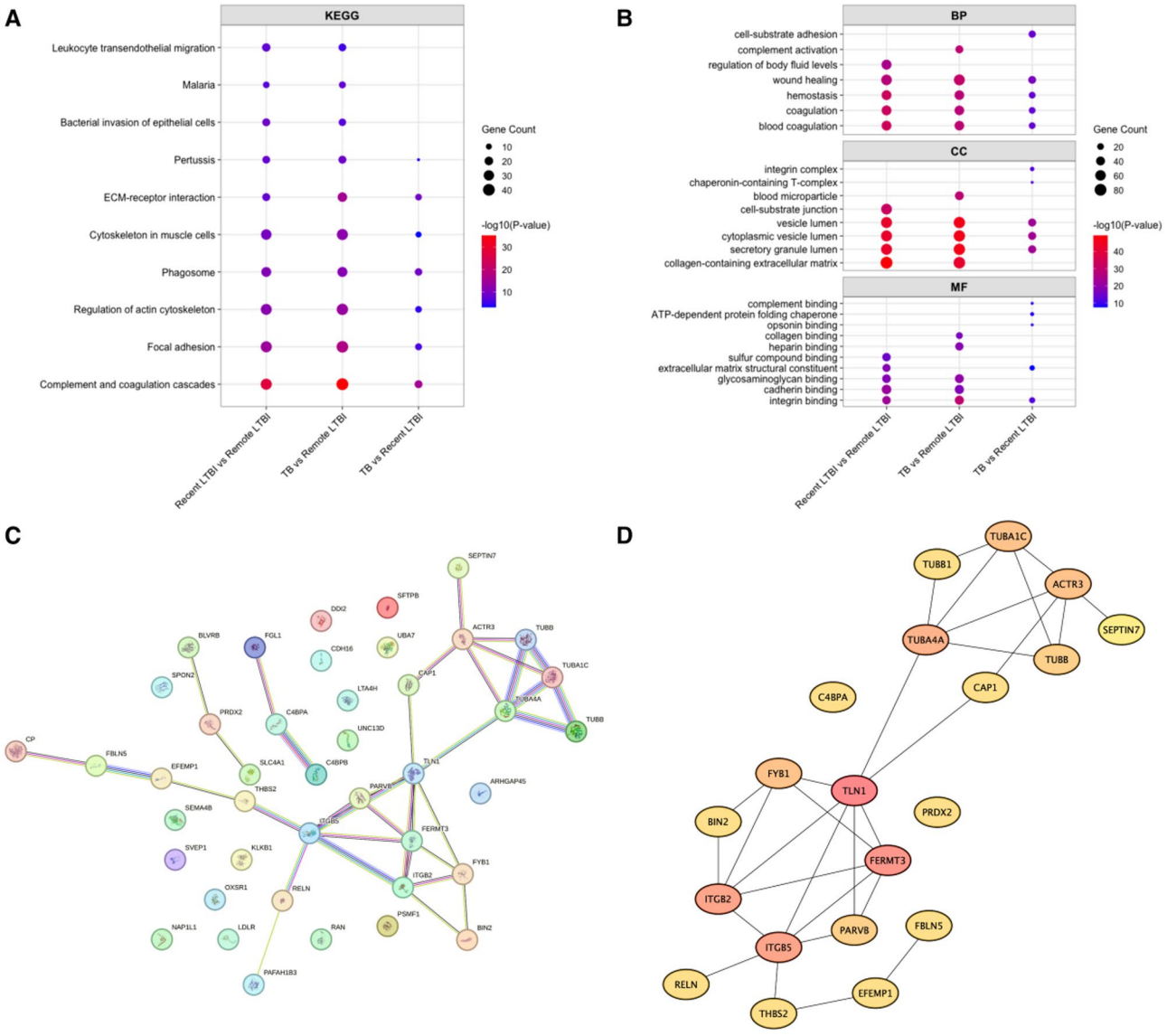

Fig. 2 Functional enrichment and network analysis of differentially expressed proteins (DEPs) among individuals with remotely acquired LTBI, recently acquired LTBI and TB. (**A**) KEGG pathway enrichment analysis of DEPs identified from three pairwise comparisons. (**B**) GO enrichment analysis of DEPs identified from three pairwise comparisons, including biological process (BP), cellular component (CC), and molecular function (MF). (**C**) Protein–protein interaction (PPI) network of the 42 overlapping differentially expressed proteins shared across all three group comparisons, constructed using the STRING database. (**D**) Hub protein module extracted from the PPI network in (**C**)

modules were significantly associated with active TB. Based on module–trait correlations, 117 hub proteins were identified from the blue, turquoise, brown, and black modules (Appendix, Figure S1D-F).

To further characterize the biological functions of these trait-associated modules, GO biological process and KEGG pathway enrichment analyses were performed. Proteins in the blue and turquoise modules were predominantly enriched in pathways related to complement and coagulation cascades, bacterial infection, and regulation of actin cytoskeleton, whereas proteins in the brown and black modules were enriched in cytoskeletal organization, cell adhesion, and Cajal body-related protein localization processes (Appendix, Figure S2).

Identification of key protein biomarkers through LASSO regression

To identify potential protein biomarkers associated with different statuses of *M.tb* infection, hub proteins derived from WGCNA modules and hub proteins identified from the PPI network based on DEPs were intersected, yielding a set of 12 candidate proteins (Appendix, Figure S3). Boxplot analysis was performed to assess the expression patterns of these 12 proteins across the three groups.

All 12 proteins exhibited significant differences among the groups (p value < 0.05), indicating their potential relevance for infection status discrimination (Fig. 3).

Subsequently, these 12 candidate proteins were subjected to feature selection using multinomial LASSO regression under a nested 10-fold cross-validation framework (Appendix, Figure S4A). Following cross-validation, five proteins, ACTR3, ITGB2, RELN, TUBB1, and ITGB5, were ultimately selected as the final biomarker panel (Appendix, Figure S4B). The LOOCV-based classification model using the five selected proteins achieved an AUCs for 0.981 (95%CI: 0.949-1.000) in TB vs. remote LTBI, 0.947 (95%CI: 0.882-1.000) in TB vs. recent LTBI, and 1.000 (95%CI: 1.000–1.000) for remote vs. recent LTBI (Fig. 4).

Validation of the 5-marker signature in transcriptomic datasets

To further validate the identified candidate biomarkers, we evaluated the expression patterns of the five-protein panel across four independent transcriptomic datasets. Notably, in all four datasets, LTBI individuals were recruited from populations at high risk of recent *M.tb* infection, including household contacts and recent immigrants from high TB-burden regions. ssGSEA revealed that the five-marker signature exhibited significant differences in enrichment scores between LTBI and active TB group (Appendix, Figure S5). Furthermore, differential expression analysis at the single-gene level revealed that although not all five genes exhibited statistically significant differences between LTBI and active TB across all datasets, each gene showed significant differential expression in at least one cohort (Appendix, Figure S6). These findings suggest that the five-protein panel demonstrates robust discriminatory capacity across transcriptomic and proteomic layers, supporting its potential as a biomarker set for identifying recent *M. tb* infection.

Discussion

In this study, we applied DIA-MS-based plasma proteomics to investigate host responses across different *M.tb* infection statuses. We observed distinct protein expression profiles between recent and remote LTBI, supporting the potential of proteomics to infer infection timing. Functional enrichment analysis revealed pathways related to complement activation, cytoskeletal

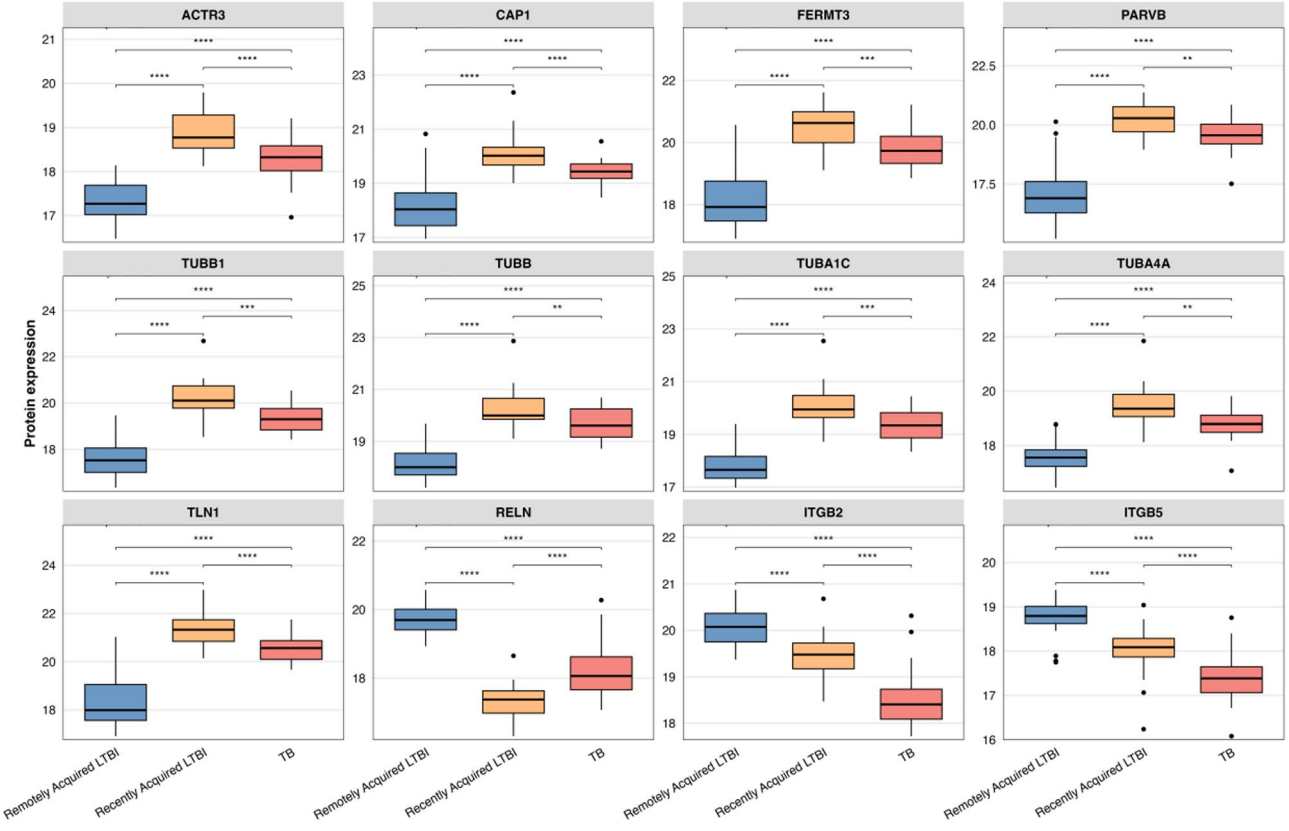

Fig. 3 Differential expression of candidate proteins across individuals with remotely acquired LTBI, recently acquired LTBI and TB. Boxplots display the protein expression levels of 12 selected candidate proteins among individuals with remotely acquired LTBI, recently acquired LTBI, and active TB. Statistical significance was assessed using Wilcoxon rank-sum tests with Benjamini-Hochberg correction for multiple comparisons P < 0.05 (*), P < 0.01 (**), P < 0.001 (***), and P < 0.0001 (****)

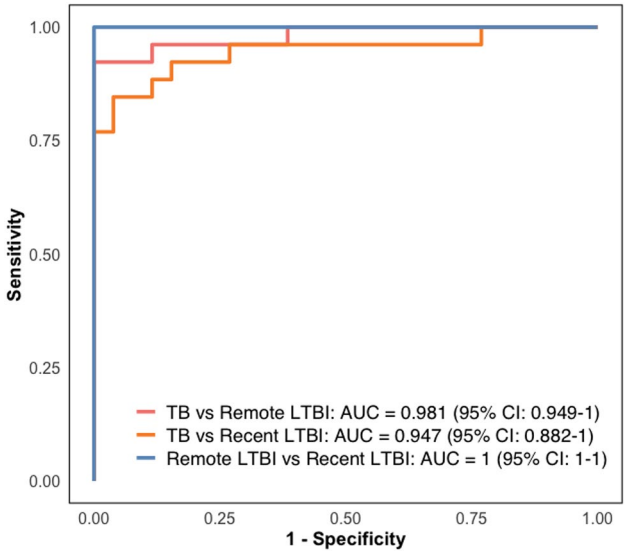

Fig. 4 Diagnostic performance of the five-protein signature panel based on LASSO classification. A multinomial LASSO regression model was trained using a five-protein panel (ACTR3, ITGB2, RELN, TUBB1, and ITGB5). Leave-one-out cross-validation (LOOCV) was applied to obtain predicted class probabilities for each sample. ROC curves and pairwise AUCs were generated to evaluate the ability of the protein panel to distinguish between TB, remotely acquired LTBI, and recently acquired LTBI.

remodeling, and cell adhesion, suggesting early immune and structural responses following infection. These findings offer preliminary insight into host-pathogen interactions during early infection and provide a foundation for developing protein-based diagnostic tools to support risk stratification and targeted LTBI treatment, particularly in high-burden settings where most individuals are already infected.

M.tb infection spans a spectrum from latent to active disease, with distinct host immune responses at each stage [6, 33]. In this study, pairwise comparisons among remote LTBI, recent LTBI, and active TB revealed substantial differences in both the number and biological functions of DEPs. To identify biomarkers of recent infection, we combined differential expression analysis, WGCNA, and machine learning [34, 35]. To enhance the robustness and biological interpretability of biomarker selection, we integrated hub proteins identified from both PPI networks and WGCNA modules. This approach allowed us to prioritize proteins that were not only centrally positioned in molecular interaction networks but also co-regulated within disease-relevant expression modules. These biologically informed candidate signatures were subsequently refined using LASSO regression with nested cross-validation, resulting in a stable and minimal protein signature with strong discriminatory.

In addition, as a preliminary analysis, PCA was performed as an unsupervised method to capture global trends in the dataset. PCA primarily reflects overall variance and broad patterns across all features, without considering group labels, and therefore may show substantial overlap between groups even when key discriminative features exist. In contrast, the supervised LASSO-based classifier selects specific features most informative for distinguishing the groups. Classification performance was evaluated using LOOCV, in which each sample is sequentially left out as a test case while the model is trained on the remaining samples. Thus, the PCA indicated substantial overlap between the recently acquired LTBI and active TB groups, while the remotely acquired LTBI group showed greater separation, consistent with a continuum of host immune responses across infection stages. In contrast, our supervised modeling framework focused on a small subset of biologically informed features that were most relevant for distinguishing the groups. Notably, as this study was based on a relatively small sample size and lacked independent external validation, the observed AUC values may be overestimated. Therefore, these findings require validation in larger, independent cohorts. At the same time, the consistent differential expression of the corresponding genes in public transcriptomic datasets provides convergent evidence supporting the relevance of these markers across different stages of M.tb infection [32].

In this study, we observed that the DEPs identified across all pairwise comparisons were consistently enriched in pathways related to actin cytoskeleton regulation, focal adhesion, and ECM-receptor interaction. Similar findings were reported by Li et al. who reported significant alterations in focal adhesion and actin cytoskeleton regulation pathways in response to *M.tb* infection [36]. The host cytoskeleton plays a crucial role in the pathogenicity of intracellular bacteria and viruses, particularly in their mechanisms for evading the host immune system [37, 38]. Dutta et al. demonstrated a link between actin organization and the efficiency of mycobacterial entry, showing that reduced levels of F-actin are significantly associated with diminished pathogen internalization [39]. In line with these findings, our study identified differential expression of ACTR3, a core component of the Arp2/3 complex involved in actin nucleation [40], among individuals with different infection status. In addition, we also observed altered expression of ITGB5, a member of the integrin family of adhesion molecules that plays a critical role in ECM–receptor interactions and focal adhesion signaling [41, 42]. These adhesion-related pathways are critical for immune cell positioning, tissue remodeling, and pathogen recognition during infection [43]. Furthermore, a recent systems genetics study identified ITGB5 as a major candidate gene associated with host necrotizing and inflammatory responses during *M.tb* infection and pulmonary TB disease progression. The findings further underscore a potential role for

ITGB5 in modulating host immune responses to *M.tb* infection [44].

Among the pathways consistently enriched across all pairwise comparisons in our study, the complement and coagulation cascades emerged as one of the prominent pathways. Notably, this pathway has also been frequently reported in previous TB proteomic studies, especially those using untargeted discovery approaches [20]. As a critical component of both the innate and adaptive immune systems [45], the complement system plays a key role in coordinating macrophage responses to *M. tb* infection. While activation of this system facilitates bacterial clearance, *M.tb* has developed various mechanisms to evade complement-mediated clearance and persist within macrophages [46]. Qi et al. reported that complement and coagulation cascades inhibited in patients with persist latent infection and enhanced in patients eliminating *M. tb*, further suggesting a dynamic role in host defense and immune control, particularly in the context of granuloma formation [47]. In line with these findings, our study observed significant downregulation of ITGB2, a key complement-related integrin, in individuals with recently acquired LTBI and active TB, potentially indicating compromised innate immune containment or pathogen immune evasion strategies [48]. Our findings are further supported by a salivary proteomics study by Mutavhatsindi et al. which found ACTR3 and ITGB2 as individual proteins that significantly differentiated active TB patients from individuals with other respiratory diseases. The repeated identification of these proteins across diverse sample types and TB phenotypes underscores their potential relevance in TB pathophysiology and biomarker discovery [49]. Furthermore, a population-based genetic study in Korea identified 10 SNPs in the ITGB2 gene significantly associated with TB susceptibility, highlight the potential influence of ITGB2 genetic polymorphisms on TB incident [50].

In addition to the aforementioned protein signatures, our final biomarker panel also included RELN and TUBB1, together pointing to a potential involvement of platelet-linked inflammatory and immune response in the spectrum of *M.tb* infection. RELN, traditionally known for its role in neuronal migration, has recently been implicated in platelet adhesion, coagulation, as well as permeability to leukocytes, with important implications for auto-inflammatory and auto-immune diseases such as arthritis, atherosclerosis, or cancer [51]. Elevated circulating RELN has been reported to promote vascular inflammation by increasing endothelial adhesiveness and permeability to leukocytes, potentially via NF-kB mediated expression of vascular adhesion molecules, endothelial activation, and may contribute to a prothrombotic milieu during systemic inflammation [52]. Notably, platelet–monocyte interactions and platelet activation have been proposed to play a role in TB immunopathogenesis and in shaping host immune responses to *M.tb* [53, 54]. Furthermore, Calvier et al. also revealed that circulating reelin promotes inflammation and modulates disease activity in acute and long COVID-19 cases [55]. TUBB1 is specifically expressed in platelets and megakaryocytes, and may be involved in proplatelet production and platelet release [56]. TUBB1 encodes the platelet and megakaryocyte enriched β1-tubulin, a key microtubule component required for proplatelet extension, platelet release, and maintenance of platelet shape. Pathogenic TUBB1 variants impair proplatelet formation and are associated with macrothrombocytopenia [57, 58]. Beyond platelet production, β1-tubulin–dependent cytoskeletal dynamics also influence platelet functional responses, and a functional TUBB1 polymorphism has been linked to reduced platelet aggregation [59]. Diseases associated with TUBB1 include bleeding disorders, macrothrombocytopenia, and cancer progression [60]. These findings highlight the potential relevance of RELN and TUBB1 in mediating immune and inflammatory responses, which may also be pertinent in the context of *M.tb* infection.

Our study has several limitations. First, the sample size was relatively small (n = 26 per group). As this was an exploratory study, a priori power analysis was not performed. Although we used nested cross-validation for feature selection and LOOCV for model evaluation, potential overfitting and limited generalizability of the five-protein panel cannot be excluded. Furthermore, the observed high AUC values may be influenced by the relatively small sample size and the lack of external validation. Validation of this five-protein panel in a larger, independent plasma cohort using clinically applicable technologies (e.g., ELISA or multiplex immunoassays) is necessary to confirm its diagnostic performance and translational potential, and may facilitate future translation into point-of-care testing. Second, the timing of *M.tb* infection could not be precisely determined, as latent infection is asymptomatic. As a cross-sectional pilot study, we did not conduct prospective follow-up and serial IGRA testing. In line with previous studies [13], recently acquired LTBI was inferred from recent household contact with a TB case, while remote LTBI was defined based on the absence of known exposure. However, exposure history may be subject to recall bias and misclassification bias, and the lack of a validated diagnostic tool for infection timing further limits classification accuracy. Despite these uncertainties, prior studies consistently show that household contacts face a significantly higher infection risk than the general population, supporting the likelihood of recent infection in this group [61]. We also screened the recently acquired LTBI group for any prior TB contact history and excluded those with

known previous exposure, while in the remotely acquired LTBI group, most of the participants had documented QFT positivity 10 years earlier, supporting remote acquisition. Furthermore, in this study, the TB Ag–Nil values were significantly higher in the recently acquired LTBI group than in the remotely acquired LTBI group (4.53 [3.07–8.21] vs. 1.40 [1.01–3.46] IU/mL; $P < 0.001$), providing biological support for the exposure-based classification and suggesting a continuum of immune responses across groups.

Conclusion

In conclusion, this study highlights the potential of plasma proteomic signatures for identifying individuals with recent *M.tb* infection and provides novel insights into the biological mechanisms underlying early stages of infection. Our results provided a foundation for developing POCT suitable for large-scale population screening, particularly in high-burden settings. Such applications could enhance TB risk stratification, facilitate targeted preventive therapy, and help to interrupt TB transmission chains, thereby reducing the burden of tuberculosis and contributing to global TB control efforts.

Abbreviations

M.tb	Mycobacterium tuberculosis
TB	Tuberculosis
LTBI	Latent tuberculosis infection
DIA-MS	Data-independent acquisition mass spectrometry
DEPs	Differentially expressed proteins
TST	Tuberculin skin test
IGRAs	Interferon-γ release assays
TPT	Tuberculosis preventive therapy
POCT	Point-of-care test
QFT	QuantiFERON-TB Gold in Tube
PCA	Principal component analysis
FC	Fold change
FDR	False discovery rate
GO	Gene Ontology
KEGG	Kyoto Encyclopedia of Genes and Genomes
PPI	Protein–protein interaction
WGCNA	Weighted correlation network analysis
MM	Module membership
GS	Gene significance
LASSO	Least absolute shrinkage and selection operator
LOOCV	Leave-one-out cross-validation
AUC	Area under the receiver operating characteristic curve
CIs	Confidence intervals
ssGSEA	Single-sample gene set enrichment analysis

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12879-026-13179-9.

Supplementary Material 1

Acknowledgements

We thank for all participants and the support from faculties and staffs in study sites. We gratefully acknowledge the technical support from the Nanjing YiKe Population Health Research Institute for proteomic experiments and analyses in this study.

Author contributions

Y.H., C.C. and B.X. participated in the conception and design of the study. J.S., Q.L., Y.S., L.Z., and C.C. were involved in participant recruitment and sample collection. Y.H. and J.S. analyzed the data and drafted the manuscript. C.C., B.X., L.Z., Q.L., and L.M. participated in the analysis and interpretation of the data. All authors contributed to the study and have read and approved the final manuscript.

Funding

This work was supported by the Jiangsu Commission of Health (Grant M2020040 to C.C.), Jiangsu Provincial Medical Key Discipline (Grant ZDXK202250 to C.C.), and the National Natural Science Foundation of China (Grant 82173581 to B.X.). The funder had no role in study design, data collection and analysis, decision to publish, or manuscript preparation.

Data availability

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium (https://proteomecentral.proteomexchange.org) via the iProX partner repository with the dataset identifier PXD074267.

Declarations

Ethics approval and consent to participate

This study was performed in line with the principles of the Declaration of Helsinki. Ethical approval was obtained from the institutional review board of Jiangsu provincial center for disease control and prevention, and written informed consent was obtained from each participant (No. JSJK2023-B029-02).

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

[1]Department of Chronic Communicable Disease, Center for Disease Control and Prevention of Jiangsu Province, Nanjing, Jiangsu Province, People's Republic of China
[2]Department of Epidemiology, School of Public Health, Fudan University, Shanghai, People's Republic of China
[3]Key Laboratory of Health Technology Assessment, National Health Commission of the People's Republic of China, Fudan University, Shanghai, People's Republic of China
[4]The Fourth People's Hospital of Lianyungang City, Lianyungang, Jiangsu Province, People's Republic of China
[5]Department of Epidemiology, School of Public Health, Boston University, Boston, MA, USA

Received: 26 August 2025 / Accepted: 20 March 2026
Published online: 01 April 2026

References

1. World Health Organization. Global Tuberculosis Report 2024. 2024.
2. World Health Organization. Global Tuberculosis Report 2022. 2022.
3. Menzies NA, et al. Time Since Infection and Risks of Future Disease for Individuals with Mycobacterium tuberculosis Infection in the United States. Epidemiology. 2021;32(1):70–8.
4. Houben RM, Dodd PJ. The Global Burden of Latent Tuberculosis Infection: A Re-estimation Using Mathematical Modelling. PLoS Med. 2016;13(10):e1002152.
5. Andrews JR, et al. Risk of progression to active tuberculosis following reinfection with Mycobacterium tuberculosis. Clin Infect Dis. 2012;54(6):784–91.
6. Pai M, et al. Tuberculosis Nat Rev Dis Primers. 2016;2:16076.
7. Hamada Y, et al. Tests for tuberculosis infection: landscape analysis. Eur Respir J. 2021;58(5).
8. Zhang H, et al. Reversion of QuantiFERON-TB Gold In-Tube test in individuals with and without prophylactic treatment for latent tuberculosis infection: A systematic review and meta-analysis. J Infect. 2018;77(4):276–82.

9. Ding YE, et al. Advancement in diagnostic approaches for latent tuberculosis: distinguishing recent from remote infections. One Health Outlook. 2025;7(1):19.
10. Partnership ST. Global Plan to End TB 2023–2030. Available from: https://www.stoptb.org/global-plan-to-end-tb/global-plan-to-end-tb-2023-2030.
11. Mpande CAM, et al. Antigen-Specific T-Cell Activation Distinguishes between Recent and Remote Tuberculosis Infection. Am J Respir Crit Care Med. 2021;203(12):1556–65.
12. Mpande CAM, et al. Immune profiling of Mycobacterium tuberculosis-specific T cells in recent and remote infection. EBioMedicine. 2021;64:103233.
13. Halliday A, et al. Stratification of Latent Mycobacterium tuberculosis Infection by Cellular Immune Profiling. J Infect Dis. 2017;215(9):1480–7.
14. Amiano NO, et al. IFN-γ and IgG responses to Mycobacterium tuberculosis latency antigen Rv2626c differentiate remote from recent tuberculosis infection. Sci Rep. 2020;10(1):7472.
15. Goletti D, et al. Response to Rv2628 latency antigen associates with cured tuberculosis and remote infection. Eur Respir J. 2010;36(1):135–42.
16. Kwan PKW, et al. A blood RNA transcript signature for TB exposure in household contacts. BMC Infect Dis. 2020;20(1):403.
17. Ault RC, et al. Blood RNA signatures predict recent tuberculosis exposure in mice, macaques and humans. Sci Rep. 2020;10(1):16873.
18. Penn-Nicholson A, et al. Discovery and validation of a prognostic proteomic signature for tuberculosis progression: A prospective cohort study. PLoS Med. 2019;16(4):e1002781.
19. MacLean E, et al. A systematic review of biomarkers to detect active tuberculosis. Nat Microbiol. 2019;4(5):748–58.
20. Mousavian Z, Källenius G, Sundling C. From simple to complex: Protein-based biomarker discovery in tuberculosis. Eur J Immunol. 2023;53(12):e2350485.
21. Cao X, et al. The dynamic change of tuberculosis infection prevalence in rural residents: 10-year follow-up of a population-based, multicentre cohort study from China. Lancet Reg Health West Pac. 2025;56:101509.
22. Chapman JR, et al. Validation of a Mass Spectrometry-Based Proteomics Molecular Pathology Assay. Mol Cell Proteom. 2026;25(1):101487.
23. Allgoewer K, et al. New Proteomic Signatures to Distinguish Between Zika and Dengue Infections. Mol Cell Proteom. 2021;20:100052.
24. Tian Y, et al. Generation of mature epicardium derived from human-induced pluripotent stem cells via inhibition of mTOR signaling. Nat Commun. 2025;16(1):5902.
25. Krishnamurthy S, et al. Recombinant Protein Spectral Library (rPSL) DIA-MS method improves identification and quantification of low-abundance cancer-associated and kynurenine pathway proteins. Commun Chem. 2025;8(1):141.
26. Ma J, et al. iProX: an integrated proteome resource. Nucleic Acids Res. 2019;47(D1):D1211–7.
27. Chen T, et al. iProX in 2021: connecting proteomics data sharing with big data. Nucleic Acids Res. 2022;50(D1):D1522–7.
28. Feng Z, et al. DEP2: an upgraded comprehensive analysis toolkit for quantitative proteomics data. Bioinformatics. 2023;39(8).
29. D'Angelo G, et al. Statistical Models for the Analysis of Isobaric Tags Multiplexed Quantitative Proteomics. J Proteome Res. 2017;16(9):3124–36.
30. Yu G, et al. clusterProfiler: an R package for comparing biological themes among gene clusters. Omics. 2012;16(5):284–7.
31. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008;9:559.
32. Mousavian Z, et al. A protein signature associated with active tuberculosis identified by plasma profiling and network-based analysis. iScience. 2022;25(12):105652.
33. Zhuang L, et al. Mycobacterium tuberculosis: immune response, biomarkers, and therapeutic intervention. MedComm (2020) 2024;5(1):e419.
34. Schiff HF, et al. Integrated plasma proteomics identifies tuberculosis-specific diagnostic biomarkers. JCI Insight. 2024;9(8).
35. Sarmah DT, et al. Latent tuberculosis and computational biology: A less-talked affair. Prog Biophys Mol Biol. 2023;178:17–31.
36. Li P, et al. Comparative transcriptomics reveals common and strain-specific responses of human macrophages to infection with Mycobacterium tuberculosis and Mycobacterium bovis BCG. Microb Pathog. 2024;189:106593.
37. Song OR, et al. ArfGAP1 restricts Mycobacterium tuberculosis entry by controlling the actin cytoskeleton. EMBO Rep. 2018;19(1):29–42.
38. Stradal TEB, Schelhaas M. Actin dynamics in host-pathogen interaction. FEBS Lett. 2018;592(22):3658–69.
39. Dutta A, et al. Integrity of the Actin Cytoskeleton of Host Macrophages is Necessary for Mycobacterial Entry. J Membr Biol. 2022;255(4–5):623–32.
40. Hu H, et al. ACTR3 promotes cell migration and invasion by inducing epithelial mesenchymal transition in pancreatic ductal adenocarcinoma. J Gastrointest Oncol. 2021;12(5):2325–33.
41. Chen Z, et al. ITGB5 is a prognostic factor in colorectal cancer and promotes cancer progression and metastasis through the Wnt signaling pathway. Sci Rep. 2025;15(1):9225.
42. Kanchanawong P, Calderwood DA. Organization, dynamics and mechanoregulation of integrin-mediated cell-ECM adhesions. Nat Rev Mol Cell Biol. 2023;24(2):142–61.
43. Patarroyo M. Adhesion molecules mediating recruitment of monocytes to inflamed tissue. Immunobiology. 1994;191(4–5):474–7.
44. Gatti DM, et al. Systems genetics uncover new loci containing functional gene candidates in Mycobacterium tuberculosis-infected Diversity Outbred mice. PLoS Pathog. 2024;20(6):e1011915.
45. Li Z, et al. Assessing the risk of TB progression: Advances in blood-based biomarker research. Microbiol Res. 2025;292:128038.
46. Jagatia H, Tsolaki AG. The role of complement system and the immune response to tuberculosis infection. Med (Kaunas). 2021;57(2).
47. Qi X, et al. Transcriptional profiling of human peripheral blood mononuclear cells in household contacts of pulmonary tuberculosis patients provides insights into mechanisms of Mycobacterium tuberculosis control and elimination. Emerg Microbes Infect. 2024;13(1):2295387.
48. Couto Alves A, et al. Dysregulation of complement system and CD4 + T cell activation pathways implicated in allergic response. PLoS ONE. 2013;8(10):e74821.
49. Mutavhatsindi H, et al. Identification of novel salivary candidate protein biomarkers for tuberculosis diagnosis: A preliminary biomarker discovery study. Tuberculosis (Edinb). 2021;130:102118.
50. Hyun-Seok Jin S-IL, Park S. Association between ITGB2 Genetic Polymorphisms and Tuberculosis. Koreanal J Clin Lab Sci. 2018;50(2):118–25.
51. Alexander A, Herz J, Calvier L. Reelin through the years: From brain development to inflammation. Cell Rep. 2023;42(6):112669.
52. Calvier L, et al. Reelin depletion protects against autoimmune encephalomyelitis by decreasing vascular adhesion of leukocytes. Sci Transl Med. 2020;12(556).
53. Kullaya V, et al. Platelet-monocyte interaction in Mycobacterium tuberculosis infection. Tuberculosis (Edinb). 2018;111:86–93.
54. Kirwan DE, Chong DLW, Friedland JS. Platelet Activation and the Immune Response to Tuberculosis. Front Immunol. 2021;12:631696.
55. Calvier L, et al. Circulating Reelin promotes inflammation and modulates disease activity in acute and long COVID-19 cases. Front Immunol. 2023;14:1185748.
56. Palma-Barqueros V, et al. Expanding the genetic spectrum of TUBB1-related thrombocytopenia. Blood Adv. 2021;5(24):5453–67.
57. Patel SR, Hartwig JH, Italiano JE Jr. The biogenesis of platelets from megakaryocyte proplatelets. J Clin Invest. 2005;115(12):3348–54.
58. Freson K, et al. The TUBB1 Q43P functional polymorphism reduces the risk of cardiovascular disease in men by modulating platelet function and structure. Blood. 2005;106(7):2356–62.
59. Cuenca-Zamora EJ, et al. Tubulin in platelets: when the shape matters. Int J Mol Sci. 2019;20(14).
60. Zhu Z, et al. TUBB, a robust biomarker with satisfying abilities in diagnosis, prognosis, and immune regulation via a comprehensive pan-cancer analysis. Front Mol Biosci. 2024;11:1365655.
61. Pinto P, et al. Incidence and risk factors of tuberculosis among 420 854 household contacts of patients with tuberculosis in the 100 Million Brazilian Cohort (2004-18): a cohort study. Lancet Infect Dis. 2024;24(1):46–56.

Publisher's note