



OPEN A bioinformatics-driven approach to identify biomarkers and elucidate the pathogenesis of type 2 diabetes concurrent with pulmonary tuberculosis

Yan Liu¹, Yonglan Pu², Jie Wang²✉, Zhiyong Li², Songliang Liu² & Shenjie Tang³✉

Type 2 diabetes (T2DM) co-existing with pulmonary tuberculosis (PTB) is associated with increased rates of treatment failure and mortality. Therefore, greater understanding of the occurrence and prevalence of this comorbidity and research to address the prevention and treatment of PTB in patients with T2DM (PTB + T2DM) have become paramount. Weighted gene co-expression network analysis (WGCNA) and Gene Ontology and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses were employed to identify key gene modules and functions related to PTB + T2DM. Immune cell infiltration and drug sensitivity were compared between PTB + T2DM patients and healthy controls (HCs), with a bioinformatic approach. Several key genes were chosen for in vitro expression assays using quantitative real-time PCR (qRT-PCR), western blotting (WB), and enzyme-linked immunosorbent assay (ELISA). Compared to HCs and T2DM-only patients, PTB + T2DM patients showed upregulated expression of complement component C1q. WGCNA identified five crucial genes associated with PTB + T2DM: *C1QA*, *CD248*, *LINC00278*, *MMP8*, and *MMP9*. Multiscale embedded gene co-expression network analysis further identified *FN1*. The main KEGG pathways in PTB + T2DM patients were related to extracellular matrix-receptor interaction, the interleukin-17 signaling pathway, the AGE-RAGE signaling pathway in diabetic complications, the PI3K-Akt signaling pathway, and neutrophil extracellular trap formation. Receiver operating characteristic (ROC) analysis indicated that *CD248*, *MMP8*, *MMP9*, *LINC00278*, and *C1QA* have potential as diagnostic markers for PTB + T2DM. The expression levels of *C1QA*, *LINC00278*, *MMP8*, and *MMP9* were significantly higher, and that of *CD248* was significantly lower, in PTB + T2DM patients than in HCs. A network comprising highly correlated hub genes and microRNAs revealed the following interactions: *C1QA* with hsa-miR-363-5p, hsa-miR-671-5p, and hsa-miR-25-5p; *CD248* with *COL1A2*, *COL1A1*, and *COL4A1*; *MMP8* with hsa-miR-539-5p, *MMP9*, and *CEACAM8*; and *MMP9* with *FN1*, *MMP8*, hsa-miR-29b-3p, hsa-miR-942-3p, hsa-miR-302-5p, and hsa-miR-133a-5p. Seven drugs (ERK_440_1713, JAK_8517_1739, Palbociclib_1054, PLX.4720_1036, Savolitinib_1936, Selumetinib_1736, and VX.11e_2096) exhibited significant sensitivity in patients with high-expression or low-expression of *C1QA*. ELISA, qRT-PCR, and WB analyses confirmed the upregulated expression of *C1QA*, *MMP8*, and *MMP9* in the peripheral blood of PTB + T2DM patients. This study elucidated the intricate molecular connections between PTB and T2DM and identified potential shared targets. Five genes (*C1QA*, *MMP8*, *MMP9*, *CD248*, and *LINC00278*) have potential as diagnostic markers for PTB + T2DM, and three genes (*C1QA*, *MMP8*, and *MMP9*) were upregulated in the peripheral blood of PTB + T2DM patients. Our findings may serve as a valuable reference for future research and clinical applications.

Keywords Pulmonary tuberculosis, Type 2 diabetes, Weighted gene co-expression network analysis, C1Q, C1QA, MMP8, MMP9

¹Clinical Medical Center for Tuberculosis, Beijing Chest Hospital, Capital Medical University, Beijing, China.

²Department of Infectious Diseases, Taicang Affiliated Hospital of Soochow University, The First People's Hospital of Taicang, Taicang, China. ³Clinical Medical Center for Tuberculosis & Beijing Tuberculosis Thoracic Tumor Research

Institute, Beijing Chest Hospital, Capital Medical University, Beijing 100091, China. ✉email: wangjie@yeah.net; tangsj1106@hotmail.com

Tuberculosis (TB) is a serious, long-term infectious disease caused by *Mycobacterium tuberculosis* (MTB) and is a significant public health issue worldwide. Pulmonary tuberculosis (PTB), the primary manifestation of TB, mainly impacts the respiratory system, specifically the lungs. According to the Global Tuberculosis report of 2022, there were an estimated 10.60 million newly diagnosed cases of TB worldwide in 2021, resulting in an incidence rate of 134 per 100,000 individuals¹. Despite significant efforts that have contributed to the decline of the global TB epidemic, there is a pressing need to further alleviate the TB burden and attain the objectives outlined in the World Health Organization's "End TB strategy".

Diabetes mellitus (DM), a group of metabolic diseases characterized by elevated blood sugar levels, is recognized as a significant risk factor for PTB¹. The prevalence of DM is steadily rising. In 2019, there were over 460 million individuals with diabetes, and this number is projected to reach 700 million by 2045². Type 2 DM (T2DM) constitutes approximately 90% of all diabetes cases and is attributed to insulin resistance and inadequate insulin secretion³. T2DM frequently gives rise to detrimental microvascular and macrovascular complications, including chronic kidney disease, hypertension, and myocardial infarction⁴. Previous meta-analyses demonstrated that DM patients are two to four times more likely to develop active TB compared to individuals without DM⁵. This association may involve compromised innate and adaptive immune system functioning caused by inadequate glycemic control over extended periods. Additionally, coexisting DM in individuals with PTB is linked to higher rates of treatment failure and mortality⁶. Hence, greater comprehension of the occurrence and prevalence of the comorbidity of T2DM with PTB, particularly in regions with a heavy burden of TB, is of paramount importance.

This study utilized weighted gene co-expression network analysis (WGCNA) to identify the key gene modules associated with T2DM and PTB. Subsequently, targeted genes were extracted. To verify the function of key genes for PTB with T2DM (PTB + T2DM), Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses were conducted. Furthermore, PTB + T2DM patients and healthy controls (HCs) were compared for associated immune cell subsets, in silico, via the CIBERSORT tool, and for treatment efficacy prediction covering specific drugs. Receiver operating characteristic (ROC) analysis was performed to determine the suitability of the detected genes to serve as diagnostic markers. The expression of selected genes was further examined in vitro by quantitative real-time PCR (qRT-PCR), western blotting (WB), and enzyme-linked immunosorbent assay (ELISA). The findings of this study offer preliminary insights into the potential impact of peripheral blood C1q on immune evasion by MTB, potentially suggesting its utility as a biomarker.

Materials and methods

Clinical sample collection

The study included patients who visited the Infectious Disease Department of Taicang First People's Hospital between January and June 2023. Approval for the collection of peripheral blood samples and relevant clinical information was obtained from the Biomedical Research Ethics Committee of Taicang First People's Hospital (Ethics Approval No. 2022-ky-050) and the Helsinki Declaration of 1975, as revised in 2008. Written informed consent was obtained from all the patients or their family members. Eligible subjects were divided into four groups: healthy volunteer control group (HC, $n = 9$), PTB patient group (PTB, $n = 12$), T2DM patient group (T2DM, $n = 13$), and PTB with T2DM patient group (PTB + T2DM, $n = 6$). The inclusion criteria for the HC group were as follows: individuals with no potential TB infection and no symptoms related to cough, sputum, or fever; normal chest imaging; negativity for the enzyme-linked immunospot assay; positivity for the IFN release assay and the tuberculin skin test; and no evidence of active TB. For the PTB group, diagnosis and classification were based on the 2001 revised "Guidelines for the Diagnosis and Treatment of Pulmonary Tuberculosis" and "Diagnostic Criteria and Classification of Tuberculosis", taking into account epidemiological history, clinical signs and symptoms, and diagnostic confirmation based on microbiology and pathology results. Inclusion criteria for the T2DM group followed the "Chinese Type 2 Diabetes Prevention and Treatment Guidelines (2020 edition)" and included typical symptoms of DM (polyuria, polydipsia, unexplained weight loss), random blood glucose ≥ 11.1 mmol/L or fasting blood glucose ≥ 11.1 mmol/L, with an age between 18 and 65 years. The PTB + T2DM group included individuals meeting both the diagnostic criteria for PTB and T2DM. Exclusion criteria included the following: 24-hour urine protein quantification > 3.5 g, primary glomerulonephritis or secondary nephritis other than diabetic nephropathy, positive urine protein due to urinary tract infection, severe primary diseases of the heart, brain, lungs, liver (alanine transaminase [ALT] > 1.5 times the upper normal limit), and blood, and other specific health conditions. Peripheral blood samples from the three groups were collected to validate the effectiveness of potential biomarkers identified through bioinformatics analysis and observe changes in the levels of specific inflammatory factors.

Clarification of the relationship between C1q expression levels and PTB + T2DM

In order to elucidate the critical functional genes implicated in the progression of PTB + T2DM, we identified and selected a publicly available dataset, GSE181143 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE181143>), from the GEO database using the search terms "Tuberculosis; Diabetes" for subsequent bioinformatics analyses. This dataset includes pulmonary tuberculosis patients recruited from two sites in India and one site in Brazil, segregated based on the presence or absence of diabetes mellitus, with blood RNA sequencing conducted at baseline, 2 months, and 6 months into treatment. The dataset encompasses a total of 290 patient samples, comprising diabetic ($n = 55$), pulmonary tuberculosis ($n = 89$), pulmonary tuberculosis with diabetes ($n = 71$), and healthy volunteer control ($n = 75$) groups^{11,12}. Differential expression analysis of the four groups was

performed using the R limma package, filtered by cutoffs of $|\log_2 \text{ fold change}| > 2$ and false discovery rate < 0.05 . The results for C1QA are displayed as a volcano diagram.

Weighted gene coexpression network construction and identification of significant modules

WGCNA is a powerful approach for identifying functionally co-expressed gene modules associated with complex human diseases. WGCNA segments the gene co-expression networks of intricate biological processes into highly correlated feature modules. These modules represent sets of genes that undergo coordinated changes, offering insights into the underlying mechanisms of specific biological processes and facilitating the discovery of potential biomarkers. The process involves transforming the gene expression profile matrix into a gene similarity matrix based on Pearson correlation, followed by conversion into an adjacency matrix. The selection of the minimum possible β value is crucial, ensuring that the adjacency matrix adheres to scale-free topology criteria, as per the represented scale-free gene co-expression topological algorithm. Topological Overlap Matrix (TOM) similarity and dissimilarity modules are then generated using TOM and dissTOM, respectively. In this study, module identification was carried out using dynamic tree cut, with a specified minimum module size of 30. Modules with high similarity scores were merged, employing a threshold of 0.25. Gene significance values were then employed to assess the associations between individual genes and comorbid PTB disease in T2DM. Strong associations between gene significance and module membership (defined as the correlation between the module eigengene and the gene expression profile for each module) were considered to suggest that key genes within the modules were closely linked to the trait. These values can be leveraged to construct a network and identify central genes.

Multiscale embedded gene co-expression network analysis (MEGENA) of key genes in PTB + T2DM

MEGENA was used to identify key genes in PTB + T2DM using the MEGENA routine in R. This analysis generates multiscale networks representing various potential interactions between genes. The protein-protein interaction (PPI) network was constructed using the online database search tool STRING (<https://www.string-db.org/>) with a threshold confidence score of ≥ 0.9 . Non-interacting protein nodes were excluded from the network. The PPI network was subsequently analyzed using Cytoscape (version: 3.8.0) to identify significant hub genes, which were further screened using the CytoHubba (version: 0.1) plug-in.

Functional enrichment analysis of the MEgreenyellow module gene set in PTB + T2DM

The MEgreenyellow module gene set from WGCNA of PTB + T2DM was analyzed using GO (<http://geneontology.org/>) and KEGG (<https://kegg.jp>)⁷. The bubble diagram was generated using the R package clusterProfiler.

ROC analysis and expression profiles of key genes in PTB + T2DM

We compared the predictive power of key genes in PTB + T2DM and determined the optimal threshold for each parameter. We utilized the ROC curve to analyze each parameter and identify the point where the sum of sensitivity and specificity was maximized to determine the optimal threshold for each parameter. The interpretation of the area under the curve (AUC) was based on the following criteria: AUC < 0.5 indicated lack of usefulness; AUC between 0.5 and 0.7 was considered poor; AUC between 0.7 and 0.9 was deemed good; and AUC between 0.9 and 1.0 was considered excellent. The expression profiles of key genes in our PTB + T2DM cohort were compared with expression data on HCs and PTB + T2DM patients from the public database GSE181143.

Construction of a highly correlated hub genes–microRNA (miRNA) network in PTB + T2DM

To explore the potential regulatory pathways of hub genes and miRNAs, we analyzed the PPI network using Cytoscape (version: 3.8.0). This analysis aimed to identify significant hub genes and miRNAs. Additionally, the PPI network was screened for hub genes using the CytoHubba (version: 0.1) plug-in.

Sensitivity analysis of treatment drugs in PTB + T2DM

The pRRophetic algorithm was utilized to evaluate differences in drug sensitivity between PTB + T2DM patients and HCs. Pharmacogenomics databases were employed to assess the response to pharmacotherapy in each patient from the public database GSE181143 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE181143>). Predictions were generated utilizing the R programming language package pRRophetic. Estimation of the half-maximal inhibitory concentration (IC₅₀) was performed utilizing the ridge regression method. A P value ≤ 0.05 was considered to indicate statistical significance.

In-silico immune cell association study

We evaluated the levels of infiltration by different types of immune cells in dataset GSE181143 using CIBERSORT (<https://cibersort.stanford.edu/>). The analysis excluded immune cells with expression values of 0. Heatmaps and boxplots were generated using the R packages ggplot2 and pheatmap. A comparison of the amounts of infiltrating immune cells and activating immune pathways between the two groups was conducted using the Wilcoxon test. A P value < 0.05 was considered to indicate statistical significance.

Human cytokine array analysis

Cytokines play important immune-related roles in both T2DM and PTB, and there is a possible interaction between T2DM and PTB at the cytokine level. For example, patients with T2DM may be more susceptible to MTB due to immune dysfunction. Meanwhile, PTB infection may also exacerbate insulin resistance and vascular complications in patients with T2DM. Therefore, the cytokine array analysis was further performed to explore the immunological link between T2DM and PTB. Specifically, within 2 h of peripheral blood collection, serum samples were collected without dilution, centrifuged at 2500 g for 10 min at 4 °C, and the supernatant was stored frozen for later use. After separation, all samples were stored at −20 °C for cytokine detection. The samples to be tested were washed twice with PBS buffer and diluted 3-fold, with 1 ml of sample volume loaded. Four serum samples from patients were randomly selected for testing in each group (each group containing 4 biological replicates). The obtained serum samples were analyzed using the RayBio® C-Series Human Cytokine Antibody Array C3 (RayBiotech, Norcross, GA, USA) kit, adhering to the manufacturer's instructions. Signals were scanned using the Axon GenePix laser scanner, and comprehensive data analysis was conducted using the RayBiotech data analysis software provided.

In vitro analyses

Expanding on the bioinformatics analyses, we explored the correlation between candidate key genes and the progression of T2DM combined with PTB. Peripheral blood samples collected from enrolled patients were analyzed by ELISA, qRT-PCR, and WB to assess changes in content and expression levels of *C1QA*, *MMP8*, and *MMP9* during the progression of T2DM combined with PTB. Within each group, we randomly selected 3 serum samples to detect inter-group differences in the transcription and expression levels of candidate research genes. ELISA detection kits from Cloud-Clone Corp (Wuhan, China) were used to evaluate the variations in *C1QA*, *MMP8*, and *MMP9* levels. For qRT-PCR assay, total RNA was extracted from peripheral blood samples using an RNA extraction kit (74534, QIAGEN), followed by reverse-transcription to cDNA using HiScript II Q Select RT SuperMix (Vazyme, Nanjing). Subsequently, qRT-PCR assay was carried out with SYBR Green PCR Master Mix (Invitrogen), and then the mRNA levels of genes were calculated by the method of $2^{-\Delta\Delta C_t}$. β -actin served as an endogenous control. Primer information is available in Supplementary Table 1. For WB assay, total protein from peripheral blood samples was extracted using a protein extraction kit (ab270054, Abcam, USA). After determination of concentration, samples were subjected to SDS-PAGE for protein separation and then transferred onto PVDF membranes. The membranes were received incubation with primary antibodies against *C1QA* (ab189922), *MMP8* (ab53017), and *MMP9* (ab76003, All from Abcam) at 4 °C overnight and then with HRP-coupled secondary antibody (ab6721) at room temperature for 2 h. The bands were visualized by an ECL detection kit (Tanon, Shanghai).

Statistical analysis

Mean values were compared for differences using one-way analysis of variance in SPSS 28.0 software. Statistical analysis and graphical representation of all quantitative data were performed in GraphPad Prism 9.0 software, and the mean values are presented \pm standard deviation (SD). The normality of the data was assessed using the Shapiro-Wilk method. For normally distributed data, inter-group comparisons were performed using the independent samples t-test, and results are presented as mean \pm SD. For non-normally distributed variables, the Wilcoxon rank-sum test was applied and results are presented as median values (interquartile range). The Mann-Whitney U test was employed to compare two groups of metric data, while the Kruskal-Wallis test was used to compare more than two groups of metric data. The chi-square test was used for intergroup comparisons of categorical variables. A significance level of $P \leq 0.05$ was considered to indicate statistical significance. Advanced scientific refinement has been applied to enhance the clarity and precision of the statement.

Results

Clinical characteristics of enrolled patients

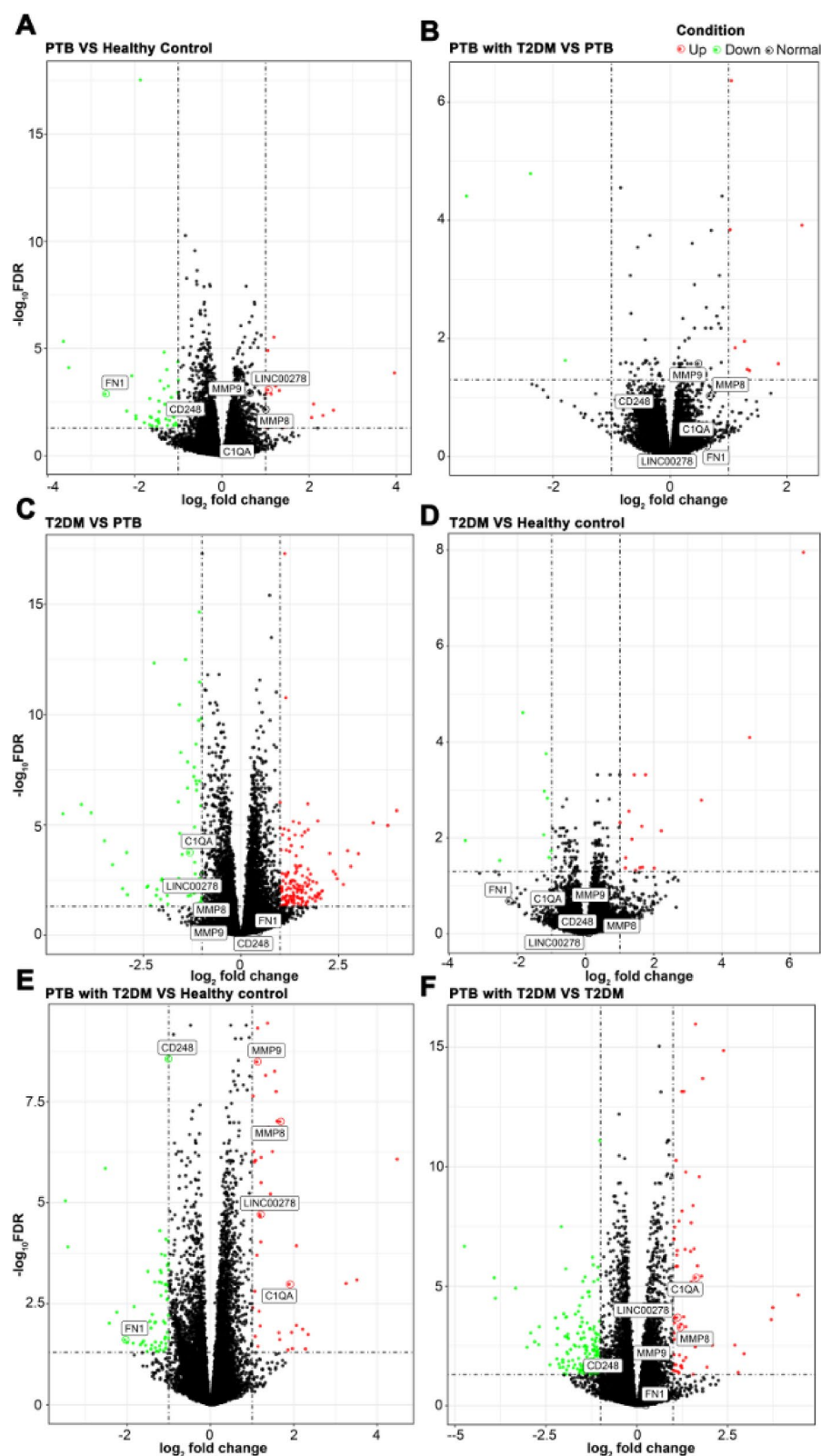
Between January and June 2023, we enrolled participants who visited the Infectious Disease Department of Taicang First People's Hospital. The study included 9 HCs, 12 patients with PTB, 13 patients with T2DM, and 6 patients with PTB + T2DM. Participants were selected on the basis of the inclusion and exclusion criteria described above. Analysis of the demographic and clinical characteristics of the enrollees are presented in Table 1. For variables with a normal distribution, inter-group comparisons were performed using the independent samples t-test, and results are presented as mean \pm SD. For non-normally distributed variables, the Wilcoxon rank-sum test was applied, and results are presented as median values with the respective interquartile ranges. The chi-square test was used for inter-group comparisons of categorical variables. There were no significant differences in demographic characteristics among the pairs of groups compared in Table 1: i.e., healthy controls vs. PTB and T2DM vs. PTB + T2DM. However, compared to the HCs, PTB patients exhibited a significantly higher monocyte count ($P = 0.036$). A similar trend was observed in PTB + T2DM patients, who had a significantly higher monocyte count than those with T2DM alone ($P = 0.017$). Additionally, compared to HCs, PTB patients had significantly lower levels of total protein ($P = 0.017$) and high-density lipoprotein ($P = 0.004$). Compared to those with T2DM alone, PTB + T2DM patients had statistically lower levels of Glutamic-pyruvic transaminase ($P = 0.002$) and Glutamic oxalacetic transaminase ($P = 0.014$).

Relationship between C1q expression levels and PTB + T2DM

A volcano diagram was used to illustrate the differentially expressed genes (DEGs) that were upregulated and downregulated among the four groups from whole blood of *Homo sapiens* using the GSE181143 dataset: PTB ($n = 207$) vs. HC ($n = 90$) (Fig. 1A); PTB + T2DM ($n = 193$) vs. PTB ($n = 207$) (Fig. 1B); T2DM ($n = 70$) vs. PTB ($n = 207$) (Fig. 1C); T2DM ($n = 70$) vs. HC ($n = 90$) (Fig. 1D); PTB + T2DM ($n = 193$) vs. HC ($n = 90$) (Fig. 1E);

Item	Group					
	Healthy control	PTB	P value	T2DM	PTB + T2DM	P value
N	9	12		13	6	
Gender			0.161			0.625
Male	4(44.4)	10(83.3)		10(76.9)	6(100.0)	
Female	5(55.6)	2(16.7)		3(23.1)	0(0.0)	
Tobacco consumption			0.640			1.000
No	7(77.8)	7(58.3)		8(61.5)	4(66.7)	
Yes	2(22.2)	5(41.7)		5(38.5)	2(33.3)	
Alcohol consumption			1.000			1.000
No	8(88.9)	10(83.3)		10(76.9)	5(83.3)	
Yes	1(11.1)	2(16.7)		3(23.1)	1(16.7)	
Hypertensive disease			1.000			1.000
No	8(88.9)	10(83.3)		12(92.3)	6(100.0)	
Yes	1(11.1)	2(16.7)		1(7.7)	0(0.0)	
History of cardiovascular disease			0.513			0.326
No	9(100.0)	12(100.0)		13(100.0)	6(100.0)	
Baseline characteristics						
Age (years)	40.3 ± 7.7	39.8 ± 14.7	0.915	45.0 ± 12.6	41.0 ± 11.6	0.473
BMI	25.6(24.7,26.0)	25.1(21.0,26.1)	0.477	26.2 ± 2.0	25.8 ± 1.1	0.728
Weight (kg)	68.2 ± 10.6	68.0 ± 11.8	0.971	75.4 ± 8.6	74.0 ± 6.1	0.689
Systolic pressure	130.0(128.0,133.0)	131.0(126.8,136.5)	0.695	133.0 ± 5.4	130.7 ± 5.6	0.577
Diastolic pressure	74.0 ± 8.0	81.2 ± 9.7	0.084	74.4 ± 11.2	74.5 ± 6.1	0.951
White blood cell count	6.3(5.0,6.5)	6.4(5.4,7.6)	0.413	6.5 ± 1.7	7.1 ± 2.4	0.649
Neutrophile granulocyte	3.5(3.2,3.8)	4.2(3.3,5.0)	0.165	3.9 ± 1.2	4.5 ± 1.7	0.259
Mononuclear cell	0.4(0.4,0.5)	0.6(0.5,0.7)	0.036*	0.4(0.4,0.6)	0.7(0.5,0.8)	0.017*
Leukomonocyte	1.9 ± 0.5	1.6 ± 0.6	0.211	1.7(1.6,2.4)	1.6(1.3,1.8)	0.354
Eosinophilic granulocyte	0.09 ± 0.06	0.09 ± 0.06	0.984	0.08(0.06,0.20)	0.11(0.07,0.13)	0.756
Basophils	0.03(0.02,0.03)	0.03(0.02,0.03)	0.703	0.04 ± 0.02	0.04 ± 0.02	0.825
Red blood cell count	4.5 ± 0.6	4.9 ± 0.4	0.179	5.2(4.7,5.4)	5.2(4.9,5.4)	0.815
Blood platelet count	254.4 ± 51.1	290.2 ± 72.1	0.222	229.6 ± 55.7	263.0 ± 103.2	0.269
Biochemistry data						
Serum creatinine	59.8 ± 13.4	60.3 ± 18.1	0.952	62.4 ± 10.2	56.4 ± 11.7	0.293
Total cholesterol	5.1 ± 0.6	4.5 ± 0.8	0.050	5.3 ± 1.4	4.8 ± 1.1	0.428
Low density lipoprotein	3.0 ± 0.8	3.1 ± 0.6	0.822	3.4 ± 1.1	3.3 ± 0.8	0.827
High-density lipoprotein	1.4 ± 0.2	1.1 ± 0.2	0.004*	1.3 ± 0.3	1.2 ± 0.3	0.683
Triglyceride	1.6 ± 0.7	1.3 ± 0.3	0.180	1.8(0.8,2.2)	1.7(1.3,2.0)	0.846
Fasting blood-glucose	5.4 ± 0.4	5.3 ± 0.5	0.763	11.6 ± 3.9	11.5 ± 3.4	0.789
Blood uric acid	316.2 ± 82.5	325.8 ± 59.8	0.760	353.9 ± 61.2	350.6 ± 100.6	0.923
Total bilirubin	11.0 ± 3.3	12.9 ± 6.2	0.418	12.2 (11.3,12.8)	8.4 (7.4,11.1)	0.042
Hemoglobin	134.1 ± 16.2	135.5 ± 20.3	0.868	154.8 ± 12.4	149.7 ± 17.5	0.516
Total protein	72.7(72.4,78.2)	69.4(45.7,72.4)	0.017*	72.0(67.5,77.8)	67.3(65.9,70.9)	0.128
Glutamic-pyruvic transaminase	36.5 ± 18.2	25.4 ± 16.9	0.167	33.4 ± 10.8	17.0 ± 4.6	0.002*
Glutamic oxalacetic transaminase	30.8 ± 10.5	23.1 ± 8.6	0.081	22.0(18.9,39.0)	13.3(12.9,18.4)	0.014*
Alkaline phosphatase	70.1 ± 28.7	92.7 ± 30.6	0.101	105.5 ± 23.6	97.4 ± 20.8	0.653
Glutamyltransferase	38.8(26.6,45.1)	33.5(19.3,57.2)	0.831	32.6(22.2,53.6)	47.5(29.8,70.8)	0.177
Lactic dehydrogenase	175.1 ± 28.1	196.2 ± 35.8	0.160	167.0 ± 21.1	159.4 ± 33.3	0.558
Statistical analysis was conducted using R software (version 4.2.2). Normality of the data was assessed using the Shapiro-Wilk test. For normally distributed data, inter-group comparisons were performed using the independent samples t-test, and results are presented as mean ± SD. For non-normally distributed variables, the Wilcoxon rank-sum test was applied and results are presented as median values (interquartile range). Categorical variables were compared between groups using the chi-square test. A significance level of *P ≤ 0.05 was considered to indicate statistical significance.						

Table 1. Clinical and demographic characteristics of the HC, PTB, T2DM, and PTB + T2DM study groups.



and PTB + T2DM ($n = 193$) vs. T2DM ($n = 70$) (Fig. 1F). Notably, the expression of *C1QA* was upregulated in the PTB + T2DM group compared with that in the HC and T2DM groups (Fig. 1E and F, respectively) and was downregulated in the T2DM group compared with that in the PTB group (Fig. 1C).

Key genes involved in PTB comorbid with T2DM

WGCNA, a systems biology methodology, was next utilized to delineate sample sets in the GSE181143 dataset. Employing a soft-threshold power of 9, we generated a hierarchical clustering dendrogram to facilitate module identification (Fig. 2A). In WGCNA, genes exhibiting analogous expression profiles across samples are grouped

Fig. 1. Clarification of the relationship between C1q expression levels and PTB + T2DM. (A–F) We retrieved the RNA-seq chip data from the GSE181143 dataset, which integrates diabetes with pulmonary tuberculosis, from the GEO database utilizing the R package GEOquery. Subsequently, we conducted a comprehensive analysis of differential expression employing the R package limma. Lastly, we produced a visually informative volcano plot to elucidate the distinctive patterns of gene expression alterations within the dataset. Volcano plots of DEGs of the following group comparisons: T2DM ($n = 70$) vs. PTB ($n = 207$) (A); T2DM ($n = 70$) vs. HC ($n = 90$) (B); PTB + T2DM ($n = 193$) vs. HC ($n = 90$) (C); PTB + T2DM ($n = 193$) vs. T2DM ($n = 70$) (D); PTB ($n = 207$) vs. HC ($n = 90$) (E); and PTB + T2DM ($n = 193$) vs. PTB ($n = 207$) (F). The red spots represent upregulated genes with a \log_2 (fold change) of > 1 (right vertical dashed line) and false discovery rate < 0.05 (horizontal dashed line), and the green spots represent the downregulated genes with a \log_2 (fold change) of < -1 (left vertical dashed line) and false discovery rate < 0.05 .

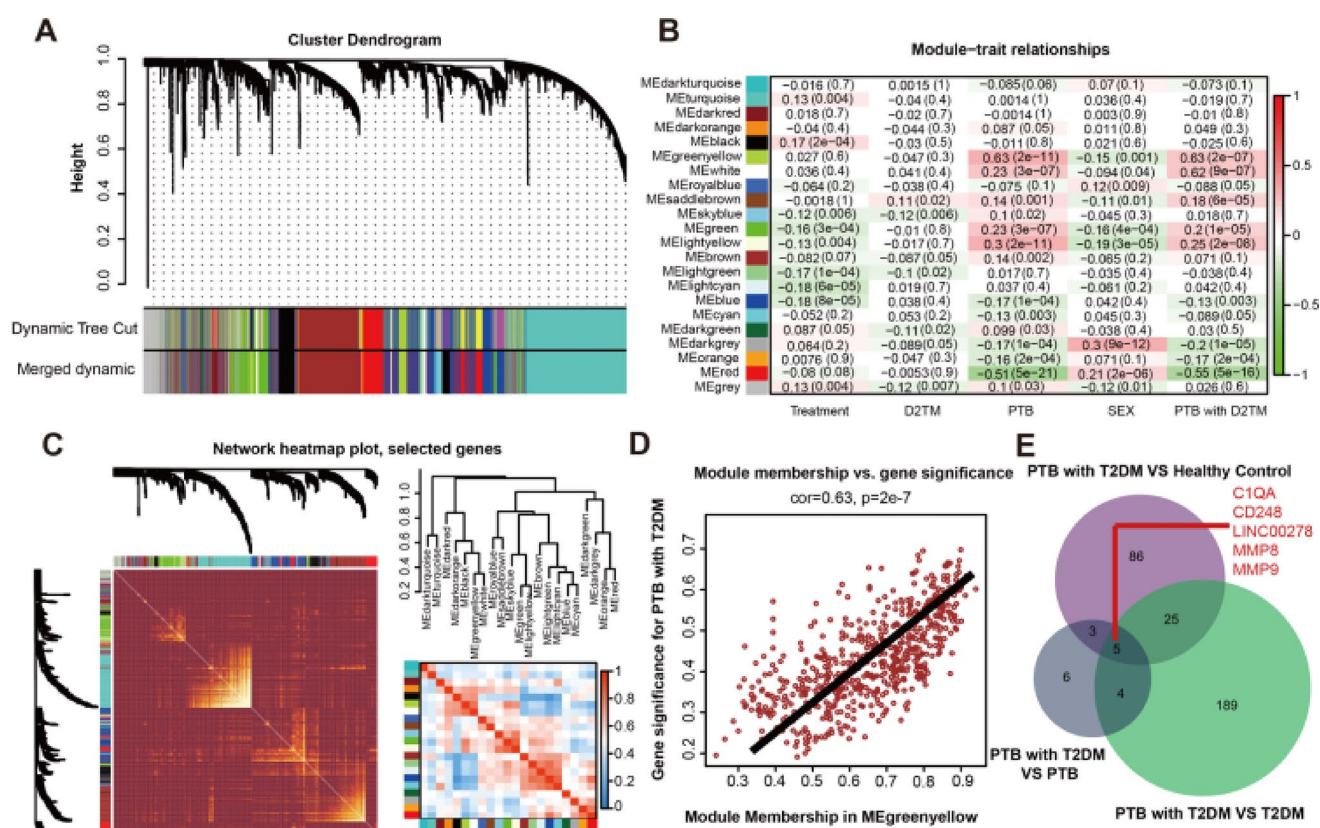


Fig. 2. Key genes of PTB + T2DM identified by WGCNA. Utilizing the publicly available dataset GSE181143, we employed WGCNA to identify and prioritize key gene modules closely associated with the progression of PTB in individuals with T2DM. This integrative approach allows for the systematic exploration of gene co-expression patterns, enabling the identification of biologically relevant gene modules that may serve as potential biomarkers or therapeutic targets for this complex disease phenotype. (A) Hierarchical clustering dendrogram for module identification. (B) Consensus eigengene networks. (C) Pearson's correlation analysis. (D) Correlation analysis of PTB + T2DM and the MEgreenyellow gene module. (E) Venn intersection of the three sets of genes.

into modules. The Module Eigengene (ME) represents the primary principal component within a WGCNA module. Our analytical findings demonstrate that the MEgreenyellow module serves as the predominant principal component within both the PTB and PTB + T2DM modules. The MEgreenyellow module displayed a significant association between PTB (coefficient = 0.63; $P < 0.001$) and PTB + T2DM (coefficient = 0.63; $P < 0.001$; Fig. 2B). The consensus eigengene networks exhibited a high level of agreement with the results of hierarchical clustering (Fig. 2C). Pearson's correlation analysis was employed to investigate the relationship between eigengene modules and clinical traits, demonstrating a highly significant correlation between gene significance for PTB + T2DM and the MEgreenyellow module (coefficient = 0.60; $P < 0.001$; Fig. 2D). Finally, using WGCNA, three sets of genes were intersected to construct a Venn diagram, resulting in the identification of five key genes in PTB + T2DM: *C1QA*, *CD248*, *LINC00278*, *MMP8*, and *MMP9* (Fig. 2D, E).

MEGENA of key genes in PTB + T2DM

Together with the MEgreenyellow module genes obtained from the WGCNA of the GSE181143 dataset, MEGENA was used to further identify key genes involved in PTB + T2DM. This analysis identified *LINC00278*, *MMP9*, *MMP8*, *C1QA*, *CD248*, and *FN1* as genes highly correlated with PTB + T2DM (Fig. 3A).

Using the MEgreenyellow module gene set within the GSE181143 dataset, we conducted a PPI network, revealing the involvement of the *LINC00278*, *MMP9*, *MMP8*, *C1QA*, and *FN1* proteins in PTB + T2DM (Fig. 3B). These results confirmed *C1QA*, *CD248*, *LINC00278*, *MMP8*, and *MMP9* as five genes with key involvement in PTB + T2DM.

Functional enrichment analysis of the MEgreenyellow module gene set in PTB + T2DM

To better understand the biological behaviors of the MEgreenyellow module gene set, we conducted an additional round of GO functional enrichment analysis, unveiling significant associations between these genes and multiple biological processes, including endodermal cell differentiation, antimicrobial humoral response, defense response to bacterium, extracellular matrix (ECM) organization, and immune system process (Fig. 4A). The main GO cellular components were specific granule lumen, azurophil granule lumen, ECM, azurophil granule, and extracellular space (Fig. 4B). The main GO molecular functions were platelet-derived growth factor binding, serine-type endopeptidase activity, ECM, ECM structural constituent conferring tensile strength, and ECM structural constituent (Fig. 4C). The main related KEGG pathways were related to ECM-receptor interaction, the IL – 17 signaling pathway, the AGE – RAGE signaling pathway in diabetic complications, the PI3 K – Akt signaling pathway, and neutrophil extracellular trap formation (Fig. 4D).

Expression profiles and ROC analysis of key genes in PTB + T2DM

In the above analysis of the GSE181143 dataset, the PTB + T2DM group exhibited higher expression levels of *C1QA*, *LINC00278*, *MMP8*, and *MMP9*, and lower levels of *CD248*, compared to the HC group (Fig. 5A–E) and the T2DM group (Fig. 5G–K). Using the five key genes identified above for ROC analysis, we found that the combination of these genes could serve as a diagnostic marker for PTB + T2DM, particularly compared to HCs. The AUC for this combined gene set reached 0.95, with the following single-gene AUCs: 0.93 for *MMP8*, 0.87 for *C1QA*, 0.79 for *MMP9*, 0.75 for *CD248*, and 0.73 for *LINC00278* (Fig. 5F). Although to a lower extent, *MMP9* and *C1QA* retained biomarker potential even when comparing PTB + T2DM to T2DM (Fig. 5L).

A highly correlated hub gene-miRNA network in PTB + T2DM

To investigate the potential regulatory pathways of hub genes, both upstream and downstream, a regulatory network linking hub genes and miRNAs was constructed. Comprehensive, proposed PPI networks illustrated the following relationships of the five key genes in PTB + T2DM with miRNAs: *C1QA* interacted with hsa-miR-

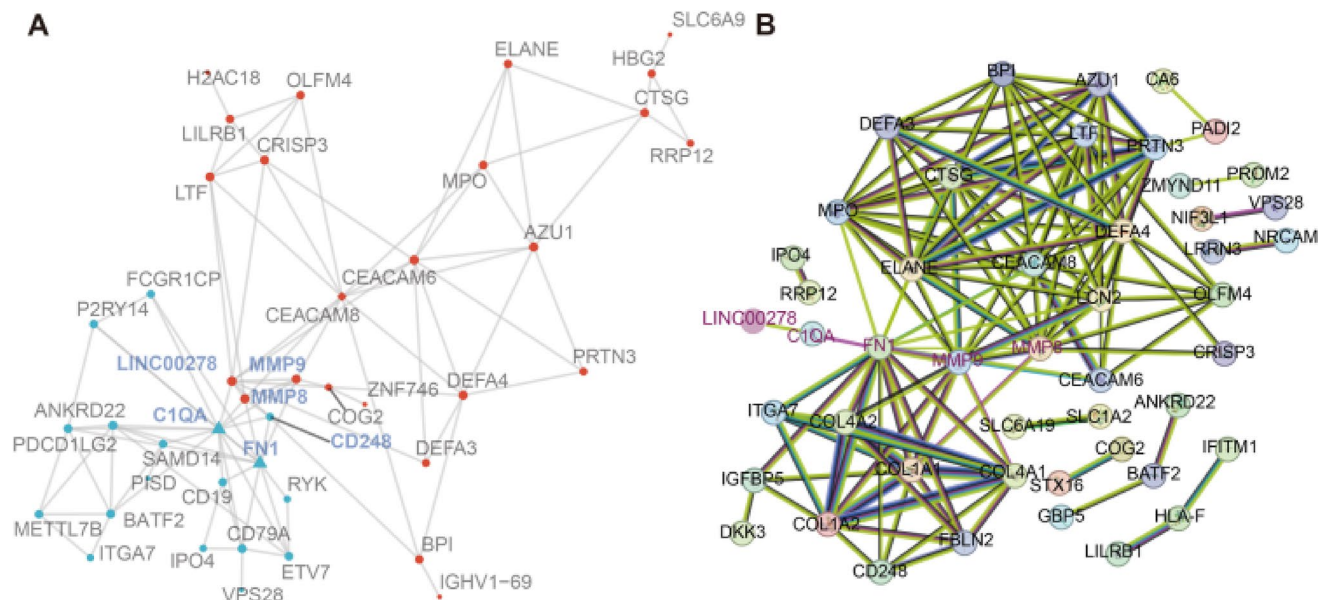


Fig. 3. Gene co-expression network analyses of key genes involved in comorbid PTB with T2DM. Based on the gene set identified from the MEgreenyellow module through WGCNA analysis of Fig. 2, we conducted multi-scale embedded gene co-expression network analysis and protein-protein interaction (PPI) network analysis to identify key genes closely associated with PTB + T2DM, particularly focusing on *LINC00278*, *MMP9*, *MMP8*, *C1QA*, *CD248*, and *FN1*. This comprehensive approach aims to uncover potential upstream and downstream interactors of these genes within biological networks, shedding light on their roles in the pathogenesis of PTB + T2DM and providing valuable insights for further investigation and therapeutic targeting. (A) MEGENA of key genes in PTB + T2DM. (B) The PPI protein network of the MEgreenyellow module gene set.

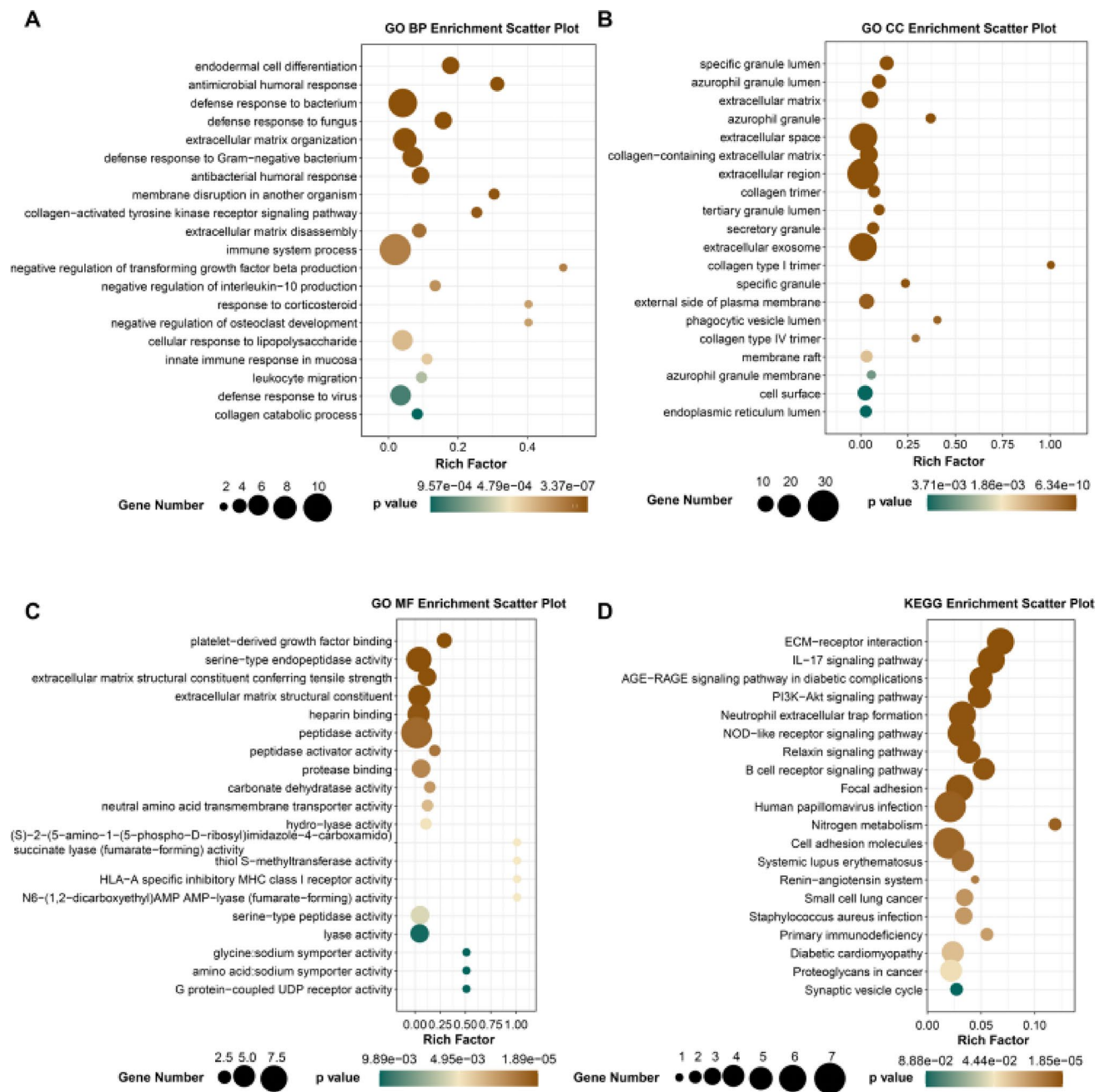


Fig. 4. Gene ontology analysis of the MEgreenyellow module gene-set in PTB + T2DM. The MEgreenyellow module was defined as by Fig. 2. (A) GO functions described as biological processes (BP; A), cellular components (CC; B), and molecular functions (MF; C) of the MEgreenyellow module gene set in PTB + T2DM. (D) KEGG pathways of the MEgreenyellow module gene set in PTB + T2DM⁷.

363-5p, hsa-miR-671-5p, and hsa-miR-25-5p; *CD248* interacted with *COL1A2*, *COL1A1*, and *COL4A1*; *MMP8* interacted with hsa-miR-539-5p, *MMP9*, and *CEACAM8*; and *MMP9* interacted with *FN1*, *MMP8*, hsa-miR-29b-3p, hsa-miR-942-3p, hsa-miR-302-5p, and hsa-miR-133a-5p (Fig. 6).

Relationship between drug sensitivity and C1QA expression in PTB + T2DM patients

Seven drugs, namely ERK_440_1713, JAK_8517_1739, Palbociclib_1054, PLX.4720_1036, Savolitinib_1936, Selumetinib_1736, and VX.11e_2096, were chosen to predict drug sensitivity on the basis of high or low C1QA expression levels in patients with PTB + T2DM ($n = 193$) from the public database GSE181143 (Fig. 7). Interestingly, this analysis revealed significant sensitivity to all seven drugs in both the high-expression and low-expression modes of C1QA (Fig. 7). Furthermore, it was observed that high C1QA expression conferred greater sensitivity than low C1QA expression (Fig. 7).

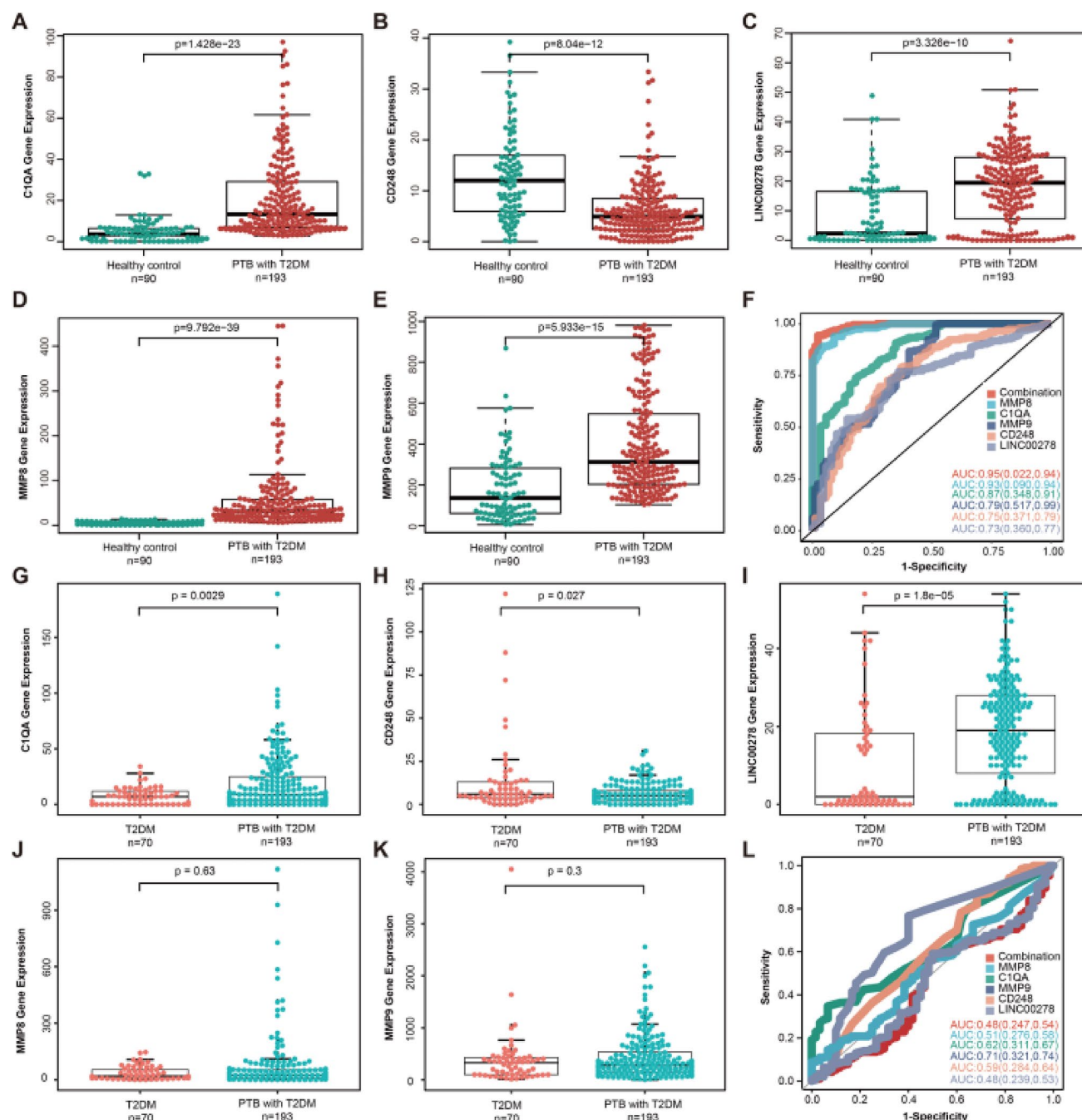


Fig. 5. Analysis of the biomarker potential of select candidate genes to detect PTB + T2DM. Data was bioinformatically analysed as shown in Fig. 2. (A–E) Comparative expression of *C1QA*, *MMP8*, *MMP9*, *LINC00278*, and *CD248* in the PTB + T2DM ($n=193$) and HC ($n=90$) groups. (F) ROC analysis of expression of *C1QA*, *MMP8*, *MMP9*, *LINC00278*, and *CD248*, and the five genes combined between PTB + T2DM ($n=193$) and HC ($n=90$) groups. (G–K) Comparative expression of *C1QA*, *MMP8*, *MMP9*, *LINC00278*, and *CD248* in the PTB + T2DM ($n=193$) and T2DM ($n=70$) groups. (L) ROC analysis of expression of *C1QA*, *MMP8*, *MMP9*, *LINC00278*, and *CD248*, and the five genes combined between PTB + T2DM ($n=193$) and T2DM ($n=70$) groups.

Immune cell differences between PTB + T2DM patients and HCs in silico

The GSE181143 dataset was utilized to generate immune cell enrichment scores for 22 distinct immune cell types, aiming to explore the associations of each with PTB + T2DM (Fig. 8A). T follicular helper cells and resting mast cells showed increased involvement in PTB + T2DM patients compared to HCs. Conversely, PTB + T2DM patients exhibited lower involvement of regulatory T cells and activated dendritic cells than HCs (Fig. 8A), using a bioinformatic approach via the CIBERSORT tool.

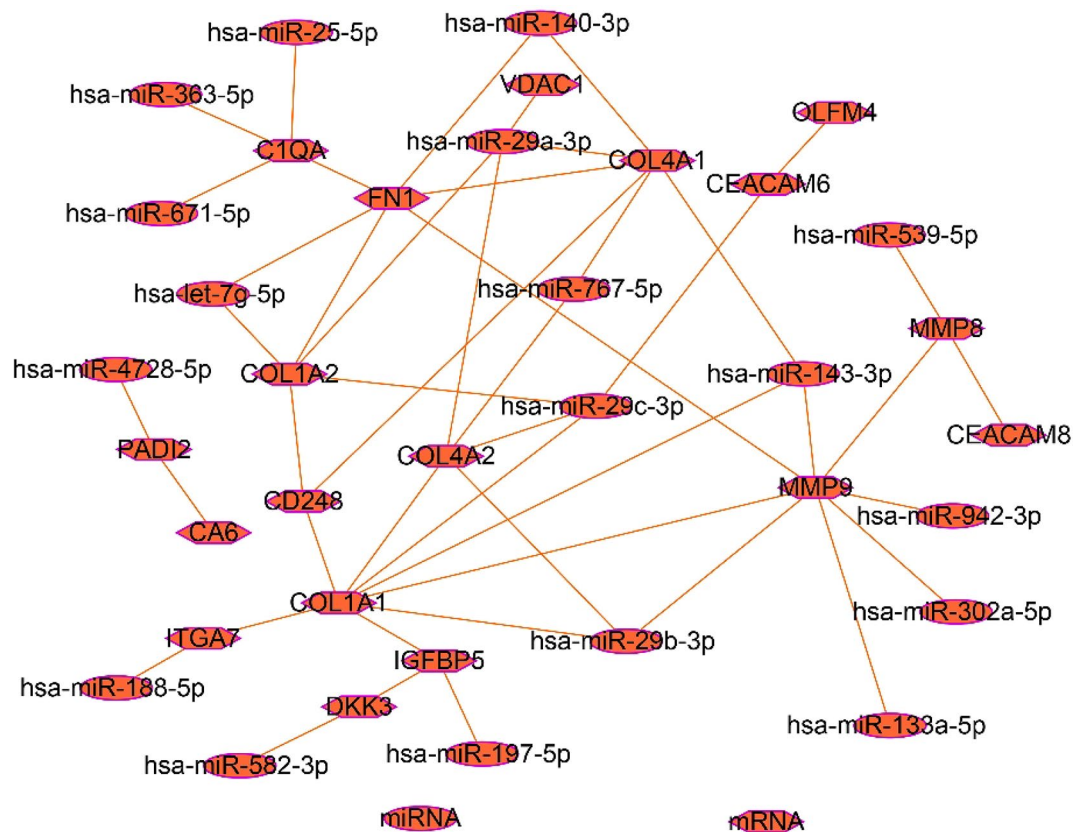


Fig. 6. Highly correlated hub gene-miRNA network involved in PTB + T2DM. Utilizing the MEGreenyellow module genes extracted via WGCNA (refer to Fig. 2), MEGENA was deployed to delve deeper into pivotal genes implicated in PTB + T2DM. This analysis unveiled LINC00278, MMP9, MMP8, C1QA, CD248, and FN1 as genes strongly correlated with PTB + T2DM. The five key genes correlated with miRNAs. *C1QA* correlated with hsa-miR-363-5p, hsa-miR-671-5p, and hsa-miR-25-5p; *CD248* correlated with *COL1A2*, *COL1A1*, and *COL4A1*; *MMP8* correlated with hsa-miR-539-5p, *MMP9*, and *CEACAM8*; and *MMP9* correlated with *FN1*, *MMP8*, hsa-miR-29b-3p, hsa-miR-942-3p, hsa-miR-302a-5p, and hsa-miR-133a-5p.

Simultaneously, we conducted an analysis of the immune cell enrichment scores for these 22 immune cell types between the PTB + T2DM and T2DM groups, aiming to explore the differences in immune involvement between the two groups. The results revealed that, compared to the T2DM group, PTB + T2DM exhibited significantly higher proportions of activated Dendritic cells, Macrophage M0, Monocytes, Neutrophils, Plasma cells, and other cell types, while the proportions of involvement in naïve B cells, NK cells, T cells CD4 memory resting, T cells CD4 naïve, T cells CD8, and other cells were significantly lower (Fig. 8B). These data suggest the presence of specific immune cell patterns in PTB + T2DM patients.

The correlation analysis presented of specific immune cell patterns in Fig. 8C examined the relationships between six genes (*CD248*, *MMP8*, *MMP9*, *LINC00278*, *C1QA*, and *FN1*) and various immune cell types. The separation of PTB + T2DM patients and HCs was demonstrated through principal component analysis (PCA; Fig. 8D). Based on Spearman's correlation, there were positive correlations between the expression of *C1QA* and M2 macrophages ($R = 0.59$; $P < 0.001$) and the expression of *LINC00278* and M2 macrophages ($R = 0.43$; $P < 0.001$) (Fig. 8E).

Expression profiles of candidate biomarkers in PTB + T2DM patients

To validate the bioinformatics-identified candidate biomarkers at the clinical level, we conducted experimental verification of the 40 samples collected from the four study groups, from Table 1. Additionally, we explored the correlation between C1q levels and the immune-metabolic characteristics induced by MTB in patients with PTB + T2DM. Immune features were assessed using an array analysis of the following inflammatory cytokines in patient serum samples: transforming growth factor (TGF)- β 1, interleukin (IL)-28 A, IL-23, IL-17, IL-17 F, IL-22, IFN γ , IL-21, IL-2, macrophage inflammatory protein (MIP)-3 α , IL-10, IL-5, IL-9, IL-12-p70, IL-3, IL-4, IL-13, IL-1 β , MIP-1 γ , granulocyte colony-stimulating factor (G-CSF), MIP-1 α , TNF α , IL-1 α , IL-6, monocyte chemoattractant protein (MCP)-1, and granulocyte-macrophage colony-stimulating factor (GM-CSF). The results are presented in Fig. 9A, with detailed statistical analysis provided in Supplementary Table 2. Comparison of the isolated PTB group with the HC group revealed significant changes in the levels of IL-21, IL-10, IL-5, IL-9, IL-12-p70, IL-3, IL-4, IL-1 β , MIP-1 γ , G-CSF, MIP-1 α , TNF α , IL-1 α , IL-6, MCP-1, and GM-CSF. The isolated

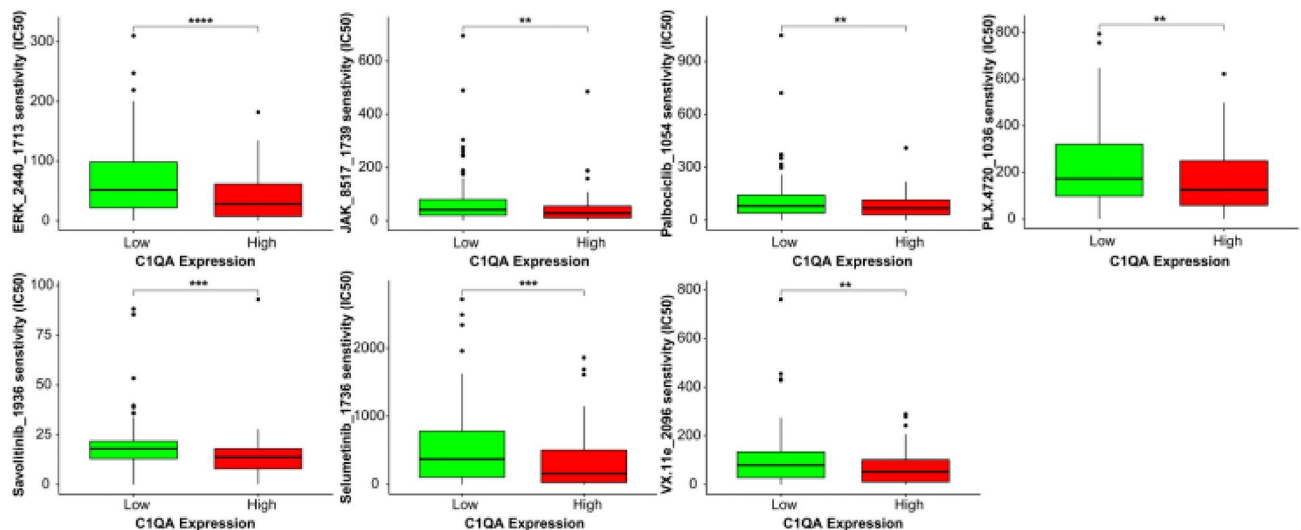


Fig. 7. Sensitivity analysis of seven drugs in relation to *C1QA* expression in PTB + T2DM. Based on the expression data of *C1QA* in the PTB + T2DM group from GSE181143, samples were divided into high and low expression groups by sorting the expression data from highest to lowest and selecting the top 25% and bottom 25% intervals, respectively. The pRRophetic algorithm was then employed to assess disparities in drug sensitivity between these two expression groups of *C1QA*. Among which seven drugs, namely ERK_440_1713, JAK_8517_1739, Palbociclib_1054, PLX.4720_1036, Savolitinib_1936, Selumetinib_1736, and VX.11e_2096, were chosen to predict drug sensitivity on the basis of high or low *C1QA* expression levels in patients with PTB + T2DM.

T2DM and PTB groups showed significant differences in the levels of IL-23, IL-21, MIP-3 α , IL-3, IL-4, IL-1 β , MIP-1 γ , G-CSF, MIP-1 α , TNF α , IL-1 α , and MCP-1, while the isolated PTB and comorbid PTB + T2DM groups exhibited significant differences in the levels of TGF- β 1, IL-23, IL-17 F, IL-22, IL-10, IL-5, IL-9, IL-12-p70, IL-1 β , MIP-1 γ , G-CSF, MIP-1 α , TNF α , IL-1 α , IL-6, MCP-1, and GM-CSF. These results indicated the differential expression of various cytokines in patients with T2DM combined with PTB, isolated PTB, and isolated T2DM, suggesting the existence of distinct immune microenvironments in these patient groups.

Next, we verified the correlations between the expression levels of *C1QA*, *MMP8*, and *MMP9* in peripheral blood and the occurrence of PTB + T2DM using ELISA and qRT-PCR. *C1QA*, *MMP8*, and *MMP9* were upregulated both as protein levels, determined by ELISA (Fig. 9B), and as mRNA expression (Fig. 9C) in individuals with PTB + T2DM compared to T2DM patients and HCs. Furthermore, WB analysis of the protein expression of *C1QA*, *MMP8*, and *MMP9* revealed a similar pattern to the results obtained from ELISA and qRT-PCR analyses (Fig. 9D, E).

Discussion

DM is recognized as a significant predisposing factor for the occurrence of PTB. T2DM accounts for approximately 90% of all diabetes cases³. Understanding the molecular pathogenesis of the comorbidity of T2DM with PTB is crucially important for regions heavily affected by TB, which require continued research to specifically address the prevention and treatment of PTB in individuals with T2DM.

In this study, the expression of *C1QA* was found to be upregulated in PTB + T2DM patients compared to that in HCs and in those with T2DM only. Using Venn intersection analysis of three gene sets identified using WGCNA, we discovered five crucial genes associated with PTB + T2DM: *C1QA*, *CD248*, *LINC00278*, *MMP8*, and *MMP9*. MEGENA identified those five genes and one additional gene: *FN1*. The main KEGG pathways of the MEGreenyellow module gene set in PTB + T2DM were related to ECM-receptor interaction, the IL-17 signaling pathway, the AGE-RAGE signaling pathway in diabetic complications, the PI3 K-Akt signaling pathway, and neutrophil extracellular trap formation. ROC analysis demonstrated the utility of these five genes as diagnostic markers for PTB or T2DM, especially in comparison to HCs. The expression levels of *C1QA*, *LINC00278*, *MMP8*, and *MMP9* were significantly higher, and that of *CD248* was significantly lower, in PTB + T2DM patients than in HCs. Network interaction analysis of hub genes and miRNAs linked *C1QA* with hsa-miR-363-5p, hsa-miR-671-5p, and hsa-miR-25-5p; *CD248* with *COL1 A2*, *COL1 A1*, and *COL4 A1*; *MMP8* with hsa-miR-539-5p, *MMP9*, and *CEACAM8*. Finally, *MMP9* was linked with *FN1*, *MMP8*, hsa-miR-29b-3p, hsa-miR-942-3p, hsa-miR-302-5p, and hsa-miR-133a-5p. However, future studies are needed to investigate whether the above miRNAs target directly genes identified in such network interactions. Furthermore, our bioinformatic analysis of drug sensitivity prediction using publicly available datasets revealed that the seven drugs examined may exhibit significant sensitivity in both high-expression and low-expression modes of *C1QA*. The findings from ELISA, qRT-PCR, and WB analyses confirmed that the expression levels of *C1QA*, *MMP8*, and *MMP9* were upregulated in the peripheral blood of patients with PTB + T2DM. However, the cohorts from the publicly

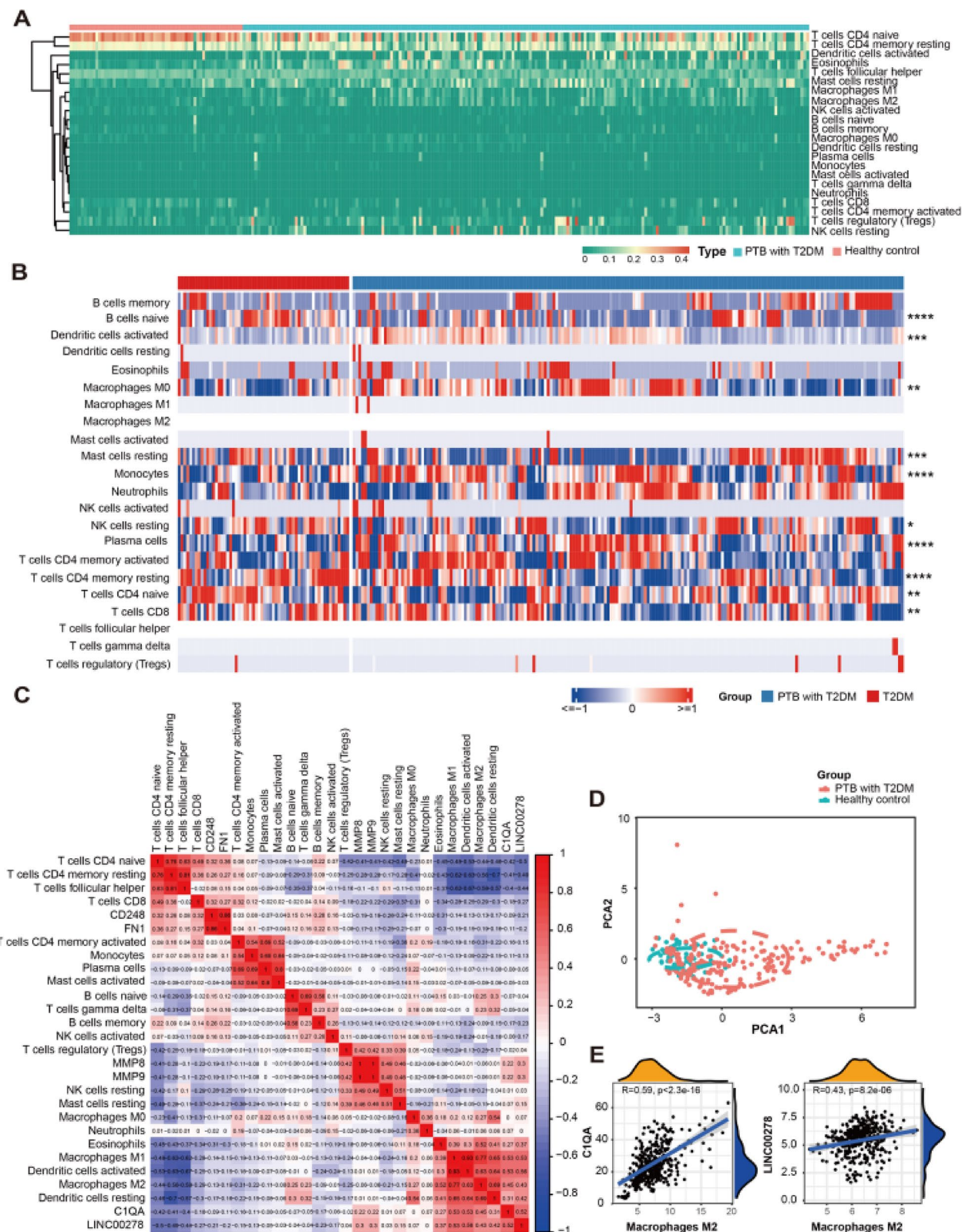


Fig. 8. Comparative bioinformatic analyses of immune cell enrichment in the PTB + T2DM and HC groups. (A–B) Enrichment scores of 22 distinct immune cell types in the two groups (PTB with T2DM vs. Healthy control and PTB with T2DM vs. T2DM group). (C) Correlation analysis between six key genes in PTB + T2DM and various immune cell types. (D) PCA was performed using the PCA expression plot feature in RStudio, data dimensionality of the data set was reduced, allowing the creation of a two-dimensional scatter plot for an initial sample distribution assessment. (E) Spearman's correlation analysis of *C1QA* and *LINC00278* expression with M2 macrophages. The values on the x-axis represent the relative enrichment level of M2 macrophages, and the values on the y-axis represent the expression levels of *C1QA* and *LINC00278*.

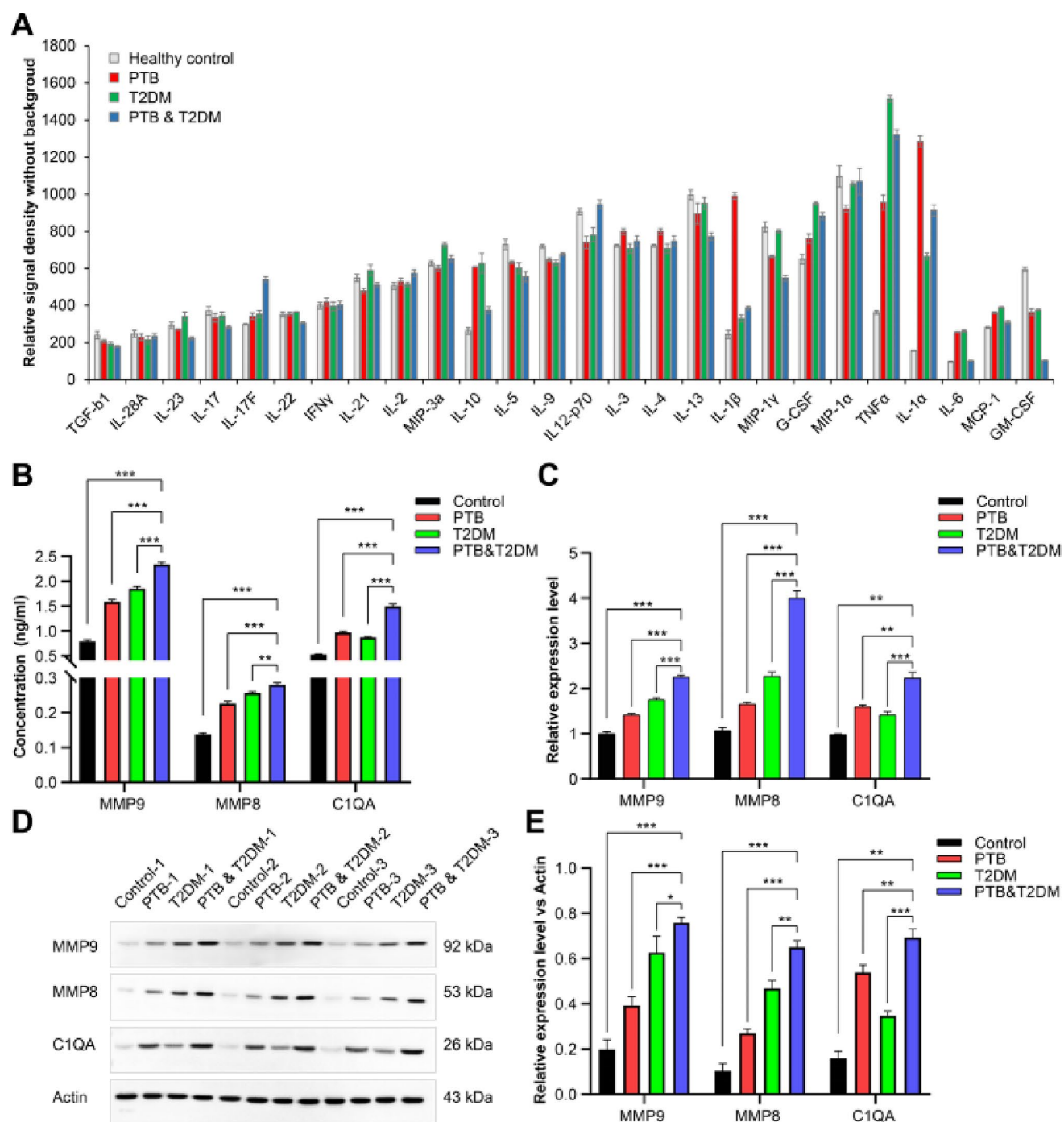


Fig. 9. Expression profile of candidate biomarkers in patients with PTB + T2DM. **(A)** Inflammation array results presented as means \pm SD of triplicate experiments; * P < 0.05, ** P < 0.01, *** P < 0.001. Each cytokine has four wells on the assay plate, allowing the collection of four valid signal density values. Statistical analysis of these density values enables a comparative assessment of the selected cytokine's concentration changes across these four groups. **(B–E)** Results of ELISA **(B)**, RT-qPCR **(C)**, and WB **(D, E)** analyses showing upregulation of C1QA, MMP8, and MMP9 expression at the mRNA and protein levels in PTB + T2DM patients. Data in graphs shown in **B, C,** and **E** expressed as mean values (bars) \pm standard error of the mean (error bars) of representative serum samples ($n = 3$) from each treatment group. Each experiment was repeated three times using three serum samples per group. Differences between means were compared using an two-tailed unpaired t-test for each marker; * P < 0.05; ** P < 0.01; *** P < 0.001.

available dataset (re-analysed in our Figs. 1, 2, 3, 4, 5, 6, 7 and 8) may differ from the patients we recruited and analysed in vitro (in Fig. 9), thus future studies with larger validation cohorts are needed to confirm these data.

Complement component C1q comprises three subunits (C1QA, C1QB, and C1QC) and possesses two functional domains: a globular head region and a collagen-like tail region. TB is linked to increased serum concentrations of C1q⁸. In one study of four geographically independent patient cohorts, those with active TB exhibited higher serum C1q levels than those with latent TB infection⁸. Following 6 months of TB treatment, however, the C1q levels become comparable to those of the control group, indicating an association with the disease rather than individual genetic susceptibility⁸. Significantly higher serum C1q levels are observed in patients with TB than in those with sarcoidosis or pneumonia, serving as an important discriminatory diagnostic criterion in clinical practice^{8,9}. Additionally, C1q levels are closely correlated with disease severity, decreasing after anti-TB treatment. C1q collaborates with MTB to promote the secretion of CCL2 and IFN- β by macrophages, both of which contribute to TB-induced tissue pathology¹⁰. We speculate that in line with our bioinformatic analysis, different cytokines may be secreted by the diverse immune cell types identified in specific groups from the GSE181143 mRNA dataset, in silico. In particular, there emerged a crucial regulatory role played by macrophages in the immune microenvironment of PBT + T2DM.

Neutrophil collagenase, also known as matrix metalloproteinase-8 (MMP-8), is the enzyme responsible for the breakdown of collagen. Elevated concentrations of MMP-8 in plasma correlate with positive sputum culture among individuals diagnosed with PTB¹¹. Gender-related differences in elevated plasma levels of MMP-8 have also been observed in individuals diagnosed with PTB¹². In addition, neutrophil-derived MMP-8 is responsible for the AMP-activated protein kinase-dependent destruction of the lung matrix in human PTB¹³.

MMP-9, a matrix metalloproteinase family member¹⁴, plays a crucial role in tissue remodeling and degradation of the ECM. Pleural TB is characterized by elevated MMP-9 activity, which is positively associated with the formation of granulomas¹⁵. TB elicits synergistic upregulation of MMP-9, resulting in enhanced MMP-9 secretion in epithelial cells¹⁶.

Fibronectin (FN1) binding protein and Ca²⁺ play a crucial role in mediating pathogenesis in MTB infection¹⁷. FN1 has also been shown to enhance MTB adhesion to alveolar macrophages in mice¹⁸. The development of pathological manifestations in PTB is linked to the presence of serum complement C4b, FN1, and prolidase¹⁹.

The glycoprotein endosialin (CD248) is predominantly expressed on the cell surfaces of pericytes and fibroblasts²⁰. It significantly contributes to the regulation of angiogenesis, tissue remodeling, and tumor progression²⁰. Notably, despite the potential importance of CD248 and LINC00278 in TB, their presence and impact have not been documented in recent TB studies.

This study successfully elucidated potential intricate molecular connections between PTB and T2DM and identified potential shared targets. Five genes (C1QA, MMP8, MMP9, CD248, and LINC00278) were shown to have potential as diagnostic markers for PTB + T2DM. Our findings offer preliminary insights into the impact of peripheral blood C1q on immune evasion by MTB, highlighting its potential utility as a biomarker for future research and clinical applications.

This study has several limitations. First, despite conducting a thorough evaluation and analysis across multiple platforms and databases, the conclusions are solely derived from in vitro analyses. Given the inherent limitations of computational analysis, additional network experiments are necessary to clarify the underlying mechanisms, and the results must be verified through clinical trials. Consequently, further exploration is required to elucidate the mechanisms of these pivotal genes in comorbid PTB and T2DM.

Conclusions

The current study successfully elucidated the intricate molecular connections between PTB and T2DM and identified potential shared targets. Five genes (C1QA, MMP8, MMP9, CD248, and LINC00278) were shown to have potential as diagnostic markers for comorbid PTB and T2DM. The expression levels of three of these genes, C1QA, MMP8, and MMP9, were upregulated in the peripheral blood of PTB + T2DM patients. Our findings offer preliminary insights into the impact of peripheral blood C1q on immune evasion by MTB, highlighting its potential utility as a biomarker for future research and clinical applications.

Data availability

The datasets generated and/or analysed during the current study are available in the GSE181143 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE181143>).

Received: 30 August 2024; Accepted: 2 May 2025

Published online: 15 May 2025

References

1. Bagcchi, S. WHO's global tuberculosis report 2022. *Lancet Microbe*. **4** (1), e20 (2023).
2. Saeedi, P. et al. Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: Results from the International Diabetes Federation Diabetes Atlas, 9th edition. *Diabetes Res Clin Pract.* ; 157: 107843. (2019).
3. Zheng, Y., Ley, S. H. & Hu, F. B. Global aetiology and epidemiology of type 2 diabetes mellitus and its complications. *Nat. Rev. Endocrinol.* **14** (2), 88–98 (2018).
4. Braunwald, E. Diabetes, heart failure, and renal dysfunction: the vicious circles. *Prog Cardiovasc. Dis.* **62** (4), 298–302 (2019).
5. Al-Rifai, R. H. et al. Association between diabetes mellitus and active tuberculosis: a systematic review and Meta-analysis. *PloS One*. **12** (11), e0187967 (2017).
6. Dooley, K. E. & Chaisson, R. E. Tuberculosis and diabetes mellitus: convergence of two epidemics. *Lancet Infect. Dis.* **9** (12), 737–746 (2009).
7. Ogata, H. et al. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **27** (1), 29–34 (1999).
8. Lubbers, R. et al. Complement component C1q as serum biomarker to detect active tuberculosis. *Front. Immunol.* **9**, 2427 (2018).

9. Revel, M., Sautès-Fridman, C., Fridman, W. H. & Roumenina, L. T. C1q + macrophages: passengers or drivers of cancer progression. *Trends Cancer*. **8** (7), 517–526 (2022).
10. Dijkman, K. et al. Systemic and pulmonary C1q as biomarker of progressive disease in experimental non-human primate tuberculosis. *Sci. Rep.* **10** (1), 6290 (2020).
11. Queiroz, A. T. L. et al. A multi-center, prospective cohort study of whole blood gene expression in the tuberculosis-diabetes interaction. *Sci. Rep.* **13** (1), 7769 (2023).
12. Vinhaes, C. L. et al. An integrative multi-omics approach to characterize interactions between tuberculosis and diabetes mellitus. *iScience* **27** (3), 109135 (2024).
13. Ong, C. W. et al. Neutrophil-Derived MMP-8 drives AMPK-Dependent matrix destruction in human pulmonary tuberculosis. *PLoS Pathog.* **11** (5), e1004917 (2015).
14. Chaturvedi, M. & Kaczmarek, L. Mmp-9 Inhibition: a therapeutic strategy in ischemic stroke. *Mol. Neurobiol.* **49** (1), 563–573 (2014).
15. Sheen, P. et al. High MMP-9 activity characterises pleural tuberculosis correlating with granuloma formation. *Eur. Respir J.* **33** (1), 134–141 (2009).
16. Elkington, P. T. et al. Synergistic up-regulation of epithelial cell matrix metalloproteinase-9 secretion in tuberculosis. *Am. J. Respir Cell. Mol. Biol.* **37** (4), 431–437 (2007).
17. Meena, P. R. & Monu, Meena, L. S. Fibronectin binding protein and Ca²⁺ play an access key role to mediate pathogenesis in Mycobacterium tuberculosis: an overview. *Biotechnol. Appl. Biochem.* **63** (6), 820–826 (2016).
18. Pasula, R., Wisniewski, P. & Martin, W. J. 2 nd. Fibronectin facilitates Mycobacterium tuberculosis attachment to murine alveolar macrophages. *Infect. Immun.* **70** (3), 1287–1292 (2002).
19. Wang, C. et al. Serum complement C4b, fibronectin, and prolidase are associated with the pathological changes of pulmonary tuberculosis. *BMC Infect. Dis.* **14**, 52 (2014).
20. Teicher, B. A. CD248: A therapeutic target in cancer and fibrotic diseases. *Oncotarget* **10** (9), 993–1009 (2019).

Acknowledgements

We extend our sincere appreciation to Beijing Genecome Medical Technology Co., Ltd. for their valuable advice and unwavering support throughout the data analysis phase of this research.

Author contributions

ST: Conceptualization, Funding acquisition, Writing — review & editing. YL: Formal analysis, Investigation, Writing – original draft, Validation. YL and YP: Methodology, Visualization. JW and ZL: Project administration. SL: Supervision.

Funding

This work was supported by the Beijing Key Clinical Specialty Project (20201214) and National Key R&D Program of China (2022YFC2304802).

Declarations

Competing interests

The authors declare no competing interests.

Ethics approval

The studies involving human participants were reviewed and approved by the Biomedical Research Ethics Committee of Taicang First People's Hospital (Ethics Approval No. 2022-ky-050). Written informed consent to participate in this study was provided by the patients.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-00928-0>.

Correspondence and requests for materials should be addressed to J.W. or S.T.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025