Exploring of the shared gene signatures and molecular mechanism in COVID-19 and tuberculosis

Jiwei Guo
Weifang People's Hospital

Jing Li
Weifang People's Hospital

Mingyue Liu
Weifang People's Hospital

Chengpeng Gao
Weifang People's Hospital

Jianwei Fang

Zhiliang Wang
Weifang People's Hospital

Hao Xu
Weifang People's Hospital

Hengtai Bi
Weifang People's Hospital

Yousen Wang
Weifang People's Hospital

Junjie Jiang (njdjjj@126.com)
Weifang People's Hospital

Article

Keywords: Coronavirus Disease 2019, Tuberculosis, GBP5, IFITM3, Bioinformatics

Posted Date: August 28th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-3287354/v1

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

Background

After the Coronavirus Disease 2019 (COVID-19) pandemic, tuberculosis (TB) incidence has demonstrated a noticeable upswing, with the causative linkage and mechanistic crosstalk between these conditions remaining uncharted. This study endeavours to decipher the communal genetic elements and underlying molecular interplay underlying COVID-19 and TB.

Methods

The Gene Expression Omnibus (GEO) served as the repository for sourcing RNA sequence datasets pertinent to COVID-19 and TB. Leveraging the R software, the Weighted Gene Co-expression Network Analysis (WGCNA) and limma package facilitated the uncovering of a co-expression network intertwined with both COVID-19 and TB. Shared genes underwent enrichment analysis via ClueGO, while hub genes within the COVID-19 and TB context were identified through MCODE based on Cytoscape software. An array of machine learning algorithms – Random Forests (RF), Least Absolute Shrinkage and Selection Operator (LASSO) Logistic Regression, and Support Vector Machine-Recursive Feature Elimination (SVM-RFE) – guided the further isolation of key genes. We also constructed the nomograms, and assessed the predictive prowess by evaluating the Area under the curve (AUC), calibration curves, decision curve analysis (DCA) and clinical impact curves. The immune microenvironment (TIME) in TB was analyzed using CIBERSORT, allowing for the assessment of correlation between key genes and immune cells.

Results

WGCNA analyses and gene expression differences analysis based on the Limma divulged a set of 281 shared differential genes between TB and COVID-19. Enrichment analysis elucidated their association with a variety of biological functions and signaling pathways, such as response to interferon-γ, NOD-like receptor signaling pathway, and influenza A. Machine learning facilitated the identification of GBP5 and IFITM3 genes, which were subsequently fashioned into nomograms, exhibiting solid clinical relevance (AUC = 0.9854, Mean Absolute Error = 0.009). CIBERSORT analysis uncovered substantial shifts in multiple immune cells in TB, notably Macrophages.M1, Dendritic.cells.activated, and Neutrophils cells, which revealed strong correlation with the expression of GBP5 and IFITM3 in TB.

Conclusion

In this study, we obtained the shared gene between COVID-19 and TB and preliminarily revealed its function. In addition, GBP5 and IFITM3 could serve as key genes among the shared genes and were associated with a variety of immune cells.
Background

Coronavirus Disease 2019 (COVID-19) is a respiratory illness caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). It was first identified at the end of 2019 and has rapidly spread worldwide, significantly impacting human health and social development over the past three years[1]. According to real-time data from the World Health Organization (WHO) as of March 10, 2023, the cumulative number of confirmed COVID-19 cases globally has surpassed 760 million, despite more than 1.3 billion COVID-19 vaccine doses administered. Tragically, COVID-19 infection manifests in over 200 symptoms, affecting multiple organs, which can also lead to severe complications such as pneumonia, liver injury, heart damage, and thrombosis, often resulting in serious illnesses and fatalities. And has resulted in nearly 6 million deaths[2–4]. During the initial stage of infection, the virus replication and its consequential infliction of tissue damage play a pivotal role in determining the progression and severity of subsequent stages. In the secondary stage, immune cells are recruited, leading to localized and systemic inflammatory reactions, which may persist even after the virus is cleared, Autopsies of deceased COVID-19 patients have revealed that excessive damage to the immune system or blood vessels is the primary cause of organ failure [2, 5]. Furthermore, a subset of COVID-19 patients continues to experience various new, recurring, or persistent symptoms and clinical manifestations even beyond a 4-week period or longer following infection. These individuals often exert an interferon response, elevated levels of inflammatory cytokines, and cellular activation phenotypes [3]. Additionally, even after recovery, individuals may encounter diverse sequelae, including inflammation, microvascular thrombosis, vascular edema, and bleeding [1].

Tuberculosis (TB) is a chronic infectious disease caused by Mycobacterium tuberculosis (Mtb). It is the most prevalent form of TB and the second leading infectious disease fatality worldwide, surpassing HIV and AIDS. In 2021, approximately 10.6 million people were infected with TB, resulting in 1.6 million deaths [6]. The World Health Organization (WHO) aims to reduce TB-related deaths by 95% and incidence rates by 90% by 2035 through collaborative efforts [7]. Notably, the incidence of TB has exerted a significant decline during the COVID-19 pandemic, attributable to factors such as travel restrictions, reduced mobility, decreased population movements, and improved health awareness [8]. However, as COVID-19-related policies are gradually lifted, the incidence of TB is on the rise. Over the next 5 years, it is projected that the number of new TB cases may increase by 6.3 million, leading to an additional 1.4 million deaths [9]. Early predictions based on mathematical models during the initial stages of the COVID-19 outbreak also indicated that even short-term disruptions caused by COVID-19 could result in an increase in the number of TB cases and deaths over the subsequent five years [10, 11].

Several factors contribute to the increased incidence of TB during the COVID-19 pandemic. Firstly, the prioritization of medical resources for COVID-19 screening and treatment, coupled with reduced mobility and decreased healthcare-seeking behavior, has led to a significant decline in TB detection and screening. Moreover, infection with the SARS-CoV-2 can make individuals more susceptible to TB infection or reactivate latent TB. Similarly, existing TB can worsen the outcomes of COVID-19[5]. Secondly, Bacillus Calmette-Guérin (BCG) vaccination is an important measure in TB prevention. Studies have shown that
BCG vaccination enhances the immune response, but the BCG vaccination rate has decreased significantly during the COVID-19 pandemic. This decline in vaccination coverage may contribute to the rapid spread of TB bacteria. Interestingly, BCG vaccination also appears to have some positive effects in the prevention and treatment of COVID-19. Countries with long-standing BCG vaccination programs have reported fewer infections and deaths from COVID-19 [5]. Thirdly, a weakened immune system increases the risk of TB [12]. The immunosuppression caused by COVID-19, particularly the depletion of CD4 + T cells, can trigger the reactivation of latent TB. The significant reduction in CD4 + T cells results in decreased production of effector cell factors such as interleukin (IL)-2, IL-4, IL-5, and IL-13. This depletion of immune cells can lead to the progression of latent TB infection to active disease, thereby further burdening global TB prevention and control efforts [9].

In short, there is a causal relationship between COVID-19 and TB in various aspects, supported by evidence of epidemiological and immune regulatory mechanisms. However, there are no reports on common gene regulatory features between COVID-19 and TB. The present study aimed to discover the shared genes of COVID-19 and TB, and explore the potential biological functions of their effects, and further tap into key genes and molecular mechanisms through a variety of approaches (Fig. 1).

**Methods**

**Dataset download and process**

The gene expression profiles were obtained from Gene Expression Omnibus (GEO) (http://www.ncbi.nlm.nih.gov/geo/) database. The COVID-19 data is GSE152075 containing the nasopharyngeal swabs mRNA expression of 430 individuals with SARS-CoV-2 and 54 negative controls. The TB data is GSE83456 containing the blood mRNA expression of 61 healthy controls and 106 TB (45 pulmonary TB and 61 extra-pulmonary TB).

**Weighted correlation network analysis (WGCNA)**

Gene expression data of TB and COVID-19 were analyzed by the “WGCNA” package based on R (v4.2.1) [13]. Genes were removed which mean expression were lower than 0.5. The optimal soft threshold (6 for TB and 17 for COVID-19) and gene-gene correlation matrix were used to build the adjacency matrix. The adjacency matrix was then converted into the topological overlap matrix (TOM). Then, the dynamic tree-cutting were performed to the module identification. Finally, the module eigengene and the correlation were calculated to identify clinical-related modules.

**Screening of shared genes and their functions in TB and COVID-19**

The shared genes in TB and COVID19 modules with positive correlation coefficients were overlapped using R packages “VennDiagram” (R 4.2.1). And we conducted the differentially expressed gene (DEG) analysis using the R packages “limma” (R v4.2.1), and adjusted p value < 0.05 was considered as the
cutoff value. Then the overlap genes were represented by venn diagram. And the volcano maps were made using R packages “ggplot2” (R v4.2.1). ClueGO is a bioinformatics tools to analyze and interpret biological data, and provide a comprehensive and intuitive visualization[14]. And we performed the Gene Ontology (GO) analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis by the ClueGO (Cytoscape v3.9.1) to further explore the biological function of these genes. In addition, we used R packages “DOSE” (R v4.2.1) for Disease Ontology (DO) analysis.

**Protein–protein interaction (PPI) network construction**

The PPI network of the overlapped DEGs was annotated with the help of the Search Tool for the Retrieval of Interacting Genes (STRING) online database (http://string-db.org). The PPI network was constructed using only those interactions that had been empirically validated and had a total score that was higher than 0.4. Then the hub PPI networks were created using the “MCODE” algorithm with default settings in Cytoscape (v3.9.1). Finally, we used the R package “clusterProfiler (v4.4.4)” (R v4.2.1) for GO and KEGG analyze based on the genes in hub PPI networks[15], the significance threshold was set as an adjusted p value < 0.05, and the above analysis results are presented as a bar chart and bubble chart by the “ggplot2” package (R v4.2.1).

**Screening for Potential Pharmacological Targets**

The connectivity map (CMAP) is an extensively utilized online pharmacogenomic database that archives gene expression data obtained from cultured cells subjected to individual treatments with diverse chemicals, encompassing an array of phytochemicals. This valuable resource serves as a comprehensive catalog, providing insights into the intricate relationship between chemical compounds and their effects on gene expression profiles in a controlled laboratory setting. (https://clue.io/) [16]. The up/down-regulated genes were introduced into CMAP to obtain the relevant small molecule compounds and their scores (from −100 to 100), then sorting them according to the scores to screen for suitable compounds.

**Machine learning identifies key genes**

3 machine-learning algorithms, including random forests (RF)[17], least absolute shrinkage and selection operator (LASSO) logistic regression[18], and support vector machine-recursive feature elimination (SVM-RFE)[19], were used to screened the key genes of the TB. The RF analysis was implemented through the “randomForest” R package in this study. And “mlbench” R package was used to the SVM-RFE analysis. This study carried out LASSO logistic regression investigation with the R package “glmnet”, and minimal lambda was considered optimal[20]. Finally, the overlap genes of the 3 machine-learning algorithms were represented by venn diagram. To further verify the reliability of the results, we built the Nomogram model using the “nomogramFormula” R package and measure it using the receiver operating characteristic (ROC) curve, and calculated the area under the curve (AUC) values. Statistical significance was determined with a P value of less than 0.05.

**Evaluation and correlation analysis of Immune Cell Subtype Distribution**
CIBERSORT is an analytical tool to provide an estimation of the abundances of member cell types in a mixed cell population [21]. And we used the “CIBERSORT” R package for this analysis after normalizing the data using the “limma” R package, and subsequently visualized the results using “ggplot2” R package. Wilcoxon test was used to screen different immune cell infiltration. Spearman correlation test was used to analyze the correlation between the immune score and the expression level of genes. And P < 0.05 was considered statistically significant. The results were visualized by the “ggplot2” R package.

Results

Co-expression modules in TB and COVID-19

Firstly, the application of WGCNA was employed on both the TB and COVID-19 datasets. In the COVID-19 dataset, we identified 8 module-trait relationships based on Spearman correlation coefficients. Notably, the blue module showed a significant positive correlation with COVID-19 (correlation coefficient r = 0.11, P value = 0.01) (Fig. 2A and C), which consisted of 27,919 genes, and a total of 5,304 DEGs were identified within it (adjust P < 0.05) (Fig. 2A and B). In the TB datasets, we identified 5 module-trait relationships. Among these, the magenta module exhibited a significant positive correlation with TB (correlation coefficient r = 0.47, P value = 7e-21) (Fig. 1B and D), which comprised 20,704 genes including 881 DEGs (adjust P < 0.05) (Fig. 2B and C). These findings indicated potential molecular interrelationships between COVID-19 and TB, as revealed by the identified module-trait correlations. Notably, the genes of the magenta module within the TB dataset and the blue module in the COVID-19 dataset appear to play significant roles in these interrelationships.

The function of the shared genes in TB and COVID-19

In our analysis of the shared genes in TB and COVID-19, we found 17,802 genes that were common to both conditions (Fig. 3A). Among these, a total of 281 genes were identified as DEGs in both COVID-19 and TB (Fig. 3C). These genes are of particular interest as they may be involved in the pathogenesis of both diseases. To gain further insights into the potential role of these shared genes, we performed GO and KEGG enrichment analyses. The GO analysis revealed their involvement in various biological processes, including the type I interferon signaling pathway, response to interferon-γ, defense response to viruses, regulation of response to biotic stimuli, IL-1 production, and regulation of response to cytokine stimuli (Fig. 3D). The KEGG analysis indicated their potential impact on pathways such as the NOD-like receptor signaling pathway, NF-κB signaling pathway, and Kaposi sarcoma-associated herpesvirus infection (Fig. 3E). Additionally, DO analysis showed associations between these genes and various diseases, including coronavirus infectious disease and tuberculosis (Fig. 3F).

To explore the protein-level interactions among these 281 genes, we constructed PPI networks (Fig. 4A) and identified three core networks using MCODE (Fig. 4B-D). Further enrichment analysis of genes within these core networks revealed their involvement in biological processes such as the response to viruses, inflammasome complex formation, and double-stranded RNA binding (Fig. 4E). The KEGG enrichment
analysis highlighted pathways including NOD-like receptor signaling, Influenza A, Coronavirus disease - COVID-19, Hepatitis C (Fig. 4E). Moreover, we utilized the CMAP to screen for small molecule inhibitors targeting the upregulated genes in the TB group expression within the core network. We ranked and selected the top 10 drugs based on their connectivity scores, including 3-matida, catechin, BML-190, naringin, NSC-119889, alitretinoin, thiorphan, erythromycin, BU-224, and mosapride (Table 1). These small molecule compounds hold promise as potential therapeutic agents for TB caused by COVID-19.

<table>
<thead>
<tr>
<th>Small molecule inhibitor</th>
<th>Score</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>3-matida</td>
<td>99.93</td>
<td>Glutamate receptor antagonist</td>
</tr>
<tr>
<td>catechin</td>
<td>99.93</td>
<td>Beta secretase inhibitor</td>
</tr>
<tr>
<td>BML-190</td>
<td>99.93</td>
<td>Cannabinoid receptor inverse agonist</td>
</tr>
<tr>
<td>naringin</td>
<td>99.93</td>
<td>Cytochrome P450 inhibitor</td>
</tr>
<tr>
<td>NSC-119889</td>
<td>99.93</td>
<td>Protein synthesis inhibitor</td>
</tr>
<tr>
<td>alitretinoin</td>
<td>99.89</td>
<td>Retinoid receptor agonist</td>
</tr>
<tr>
<td>thiorphan</td>
<td>99.89</td>
<td>Membrane metalloendopeptidase inhibitor</td>
</tr>
<tr>
<td>erythromycin</td>
<td>99.89</td>
<td>NFkB pathway inhibitor</td>
</tr>
<tr>
<td>BU-224</td>
<td>99.89</td>
<td>Imidazoline receptor ligand</td>
</tr>
<tr>
<td>mosapride</td>
<td>99.86</td>
<td>Serotonin receptor agonist</td>
</tr>
</tbody>
</table>

**Table 1**

TOP 10 small molecule inhibitor of the hub gene according to the CMAP

Machine learning identifies the key genes

In our analysis, we employed three machine learning algorithms, including RF model (Fig. 5A), Lasso regression (Fig. 5B and C), and SVM-RFE (Fig. 5D and E) to identify signature genes from the pool of 281 differential genes. Finally, Lasso regression identified 22 genes, while RF was used as a classifier for interaction prediction, resulting in the selection of the top 20 genes for further analysis, and SVM-RFE identified 48 genes (Fig. 5F). Among these, two genes, IFITM3 and GBP5, were consistently identified by all 3 algorithms (Fig. 5F). Both the 2 genes showed increased expression in the TB group (Fig. 5G).

Moreover, we constructed a nomogram for diagnosing TB based on the genes GBP5 and IFITM3 (Fig. 6A). Additionally, ROC curve analysis was performed, which showed an AUC value of 0.9854 (95% CI 0.9721–0.998) for the column line graph model. The AUC values of 0.9811 (95% CI 0.9636–0.9986) and 0.9653 (95% CI 0.9392–0.9913) were obtained for the other models, slightly lower than the AUC value of the column line graph model (Fig. 6B). Calibration plots exhibited excellent concordance between the predicted and observed values (Mean absolute error = 0.009) (Fig. 6C). Decision curve analysis (DCA) indicated that the column line graph model had better clinical benefits (Fig. 6D). Additionally, the clinical
impact curve also demonstrated that the column line graph model had superior clinical effectiveness (Fig. 6E). These findings highlight the clinical importance of GBP5 and IFITM3 in the diagnosis of TB.

**Relationship of the two genes with immune cells**

We used CIBERSORT to calculate the proportion of various immune cells in TB ((Fig. 7A)), and we observed significant differences ($P < 0.05$) in the cell populations of Macrophages.M0, Macrophages.M1, Macrophages.M2, Dendritic.cells.activated, Eosinophils, and Neutrophils between the TB and control groups(Fig. 7B). Subsequently, we conducted Spearman correlation analysis to explore the relationship between GBP5 and IFITM3 expression and these differential cell populations. The analysis revealed significant positive correlations between GBP5 and IFITM3 expression levels with Macrophages.M1, Dendritic.cells.activated, and Neutrophils cells, indicating that there is a strong association between the expression of GBP5 and IFITM3 and the presence or activity of these specific immune cell types in TB (Fig. 7C-D).

**Discussion**

An increasing body of evidence points towards an interaction between COVID-19 and TB, even the COVID-19 with potential negative implications for TB. In our study, we identified a substantial quantity of genes shared between TB and COVID-19, and these genes were found to be notably enriched in several biological processes, including the response to IFN-γ, viral defense, production of interleukins, and cytokine stimulation response. Moreover, these genes demonstrated a significant correlation with both NOD – like receptor signaling pathway, NF-κB signaling pathway. In our pursuit to enhance the precision of our results, we employed 3 distinct machine learning algorithms, Lasso regression, RF, and SVM-RFE to rigorously screen two immune response-related genes, GBP5 and IFITM3. Following the selection process, we constructed the nomogram, and further underscores the pivotal role of the expression levels of GBP5 and IFITM3 in TB. Considering the integral role of GBP5 and IFITM3 in immune response, we utilized the CIBERSORT algorithm to calculate the proportion of various immune cells present in TB, and to discern their correlation with the expression levels of GBP5 and IFITM3. Ultimately, we found that the expression levels of GBP5 and IFITM3 showed a significant and positive correlation with Macrophages.M1, Dendritic.cells.activated, and Neutrophils cells. The evidence supports the crucial involvement of these cell types in the immunological response to TB.

GBP5 and IFITM3 are two genes associated with immune response. GBP5 belongs to the GTPases superfamily of dynamins, and plays a vital roles in several crucial cellular processes, including signal transduction, translation, vesicular trafficking, and phagocytosis. Furthermore, GBP5 exhibits antiviral activity against viral infections in an IFN-dependent manner and servers as a critical modulator of antiviral immunity[22]. IFITM3 is an interferon-induced transmembrane protein and has been shown to function in defense against various pathogenic infections. The protein encoded by IFITM3 can prevent a variety of viral pathogens from entering cells. IFITM3 can effectively inhibit SARS-CoV-2 infection, and individuals with IFITM3 single nucleotide polymorphisms may be more prone to severe COVID-19 [23].
Furthermore, IFITM3 is linked with the infection and replication of Mtb, promoting the growth of Mtb in human monocytes and alveolar/epithelial cells [24].

Macrophages, dendritic cells, and neutrophils are vital components of the immune system, each playing unique roles in defense against pathogens. Macrophages are large white blood cells present throughout the body’s tissues, characterized by phagocytic activity, cytokine secretion, and tissue repair functions. Dendritic cells are antigen-presenting cells primarily involved in antigen capture and presentation, and the activation of adaptive immune responses. Neutrophils are the first responders to infection or inflammatory reactions, with functions including phagocytosis, release of antimicrobial substances, and inflammation regulation. Pulmonary macrophages are a significant target of COVID-19[25–27]. In severe COVID-19 patients, macrophages accumulate in large numbers in bronchoalveolar lavage fluid [28], and the pro-inflammatory macrophage microenvironment largely determines the immune cell composition in the alveoli of these patients [29]. After SARS-CoV-2 infection, macrophages are typically activated, and their secretion of inflammatory cytokines (IL-6, IL-10, and IFN-γ) often becomes hyperactivated, leading to aberrant immune responses. This hyperactivation can cause tissue damage and exacerbate inflammation, leading to pathological inflammation [25, 30]. Furthermore, dendritic cells are involved in the immune response process to the virus in COVID-19 patients [31]. COVID-19 can cause exhaustion and functional impairment of DC cells, and this long-term damage may adversely affect secondary infections [32–34]. Neutrophils are key mediators of innate immune responses, and numerous studies indicate that neutrophils play a vital role in the pathophysiology of COVID-19. Especially in severe cases, an increased count of neutrophils in the infection site and blood is a characteristic feature of COVID-19 [35, 36]. In addition, there are significant changes in the composition and dysfunction of neutrophils in the bodies of COVID-19 patients, especially severe patients, with increased proportions of mature inactive neutrophils and immature neutrophils [37].

During the initial stage of Mycobacterium tuberculosis infection, various cells release pro-inflammatory cytokines and chemokines that induce immune cells to migrate to the site of infection, initiating granuloma formation and triggering host protective responses. Cytokines play a significant role in determining the outcome of Mtb infection [38]. Macrophages, dendritic cells, neutrophils, NK cells, and epithelial cells are key cells in the immune response to Mtb [39]. Mtb can be phagocytosed by macrophages in the early stage of infection. Still, it possesses the unique ability to survive and even proliferate within macrophages, protecting it from immune attack [40]. The changes in macrophages and dendritic cells in COVID-19 patients may provide favorable conditions for Mtb infection. GBP5 can affect the function and activity of macrophages by regulating cell signal transduction pathways and the expression of immune-related factors. For instance, in Rosacea, GBP5 biases macrophage polarization towards the M1 phenotype through the NF-κB signaling pathway [41], and GBP5 plays a crucial role in the lipopolysaccharide (LPS)-activated inflammatory immune response, serving as a marker gene for M1 macrophage polarization [22]. Neutrophils have a strong ability to control Mtb. Neutrophils stimulated by TNF can inhibit 50–95% of Mtb growth within 1 hour [42]. The imbalance and functional impairment of neutrophils caused by COVID-19 may facilitate Mtb infection.
In this study, we conducted an initial screening to identify shared genes between COVID-19 and tuberculosis (TB), as well as their concomitant influence on biological processes and pathways. Consequently, we have identified GBP5 and IFITM3 as potential biomarkers and therapeutic targets. We further found strong correlations between these two genes and Macrophages.M1, Dendritic.cells.activated, and Neutrophils cells. We hypothesize that changes in immune cells affected by COVID-19, especially Macrophages.M1, Dendritic.cells.activated, and Neutrophils cells, could be important factors promoting TB, where GBP5 and IFITM3 might play significant regulatory roles. However, due to the limitations of data provided by public databases and potential shortcomings of the analytical methods used in the present study, further molecular biology experiments and clinical trials are needed to validate these findings.

**Abbreviations**

COVID-19, Coronavirus Disease 2019;
SARA-CoV-2; Severe Acute Respiratory Syndrome Coronavirus 2;
WHO, World Health Organization;
TB, Tuberculosis;
Mtb, Mycobacterium Tuberculosis;
BCG, Bacillus Calmette-Guérin;
IL, Interleukin;
GEO, Gene Expression Omnibus;
WGCNA, Weighted Gene Co-expression Network Analysis;
TOM, Topological Overlap Matrix;
DEG, Differentially Expressed Gene;
GO, Gene Ontology;
KEGG, Kyoto Encyclopedia of Genes and Genomes;
DO, Disease Ontology;
PPI, Protein–protein interaction;
STRING, Search Tool for the Retrieval of Interacting Genes;
CMAP, Connectivity Map;
Declarations

During the preparation of this work the author(s) used ChatGPT 4.0 in order to improve readability and language. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Ethics approval and consent to participate

All the data used in the text are taken from public databases and therefore do not require ethics committee approval.

Availability of data and materials

All data generated or analysed during this study are included in this published article.

Competing interests

The authors declare that they have no competing interests.

Authors’ contributions

Junjie Jiang, Chengpeng Gao and Jiwei Guo conceived and designed the study. Jiwei Guo and Jing Li performed data analysis and prepared the manuscript. Junjie Jiang, Mingyue Liu, and Hao Xu also prepared the manuscript. Jianwei Fang and Zhiliang Wang conducted the bioinformatics and statistical analyses. Hengtai Bi and Youseng Wang provides guidance on statistical analysis and clinical knowledge. All authors contributed to the article and approved the submitted version.

Acknowledgments
The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Funding**

This work was supported by grants from Weifang City health and Family planning Commission Science research project (Grant No. WFWSJK-2020-201)

**References**


**Figures**
Figure 1

Workflow of the study. This figure illustrates the analytical methods and processes used in this study.
Figure 2

Identification of modules linked to clinical features of COVID-19 and TB by WGCNA. (A-B) Cluster dendrogram of co-expressed genes in COVID-19 (A) and TB (B); (C-D) Heap of module–trait relationships in COVID-19 (C) and TB (D).
Figure 3

Analysis of the shared genes and their functions in COVID-19 and TB. (A) Venn diagram depicting shared genes in CRC and COVID-19; (B) Volcano-plot representation of differential gene expression (DGE) in CRC and COVID-19; (C) Venn diagram depicting shared genes in CRC and COVID-19 with statistical significance (Adjust P-value < 0.05); (D) The network of GO terms in ClueGO. (E) The network of KEGG terms in ClueGO; (F) DO enrichment analysis.
Figure 4

PPI networks with statistically significant shared genes. (A) The PPI network among the shared genes; (B-D) The PPI network analyzed by MCODE; (E) GO and KEGG enrichment of the hub-genes in figure B-D.
Figure 5

Machine learning screens for key genes. (A) based on SVM-RFE to screen biomarkers; (B-C) LASSO logistic regression algorithm to screen diagnostic markers; (D-E) Based on RF algorithm to screen biomarkers; (F) Venn diagram showed the intersection of diagnostic markers obtained by the three algorithms; (G) Boxplot showed the expression of key genes between TB and control group based on Wilcoxon rank-sum test. (***, P < 0.001; **, P < 0.01; *, P < 0.05)
Figure 6

Key genes for TB diagnosis. (A) Nomogram is used to predict the occurrence of TB; (B) The ROC curve of the diagnostic efficacy verification; (C) Calibration curve; (D) DCA curves; (E) Clinical impact curve.
Figure 7

Analysis of immune cell infiltration of TB and correlation with key genes. (A-B) Immune cell infiltration, the proportion of each immune cell in each sample (A) and boxplots of the proportion of immune cells in TB and controls based on Wilcoxon rank-sum test. (C) Spearman correlation analysis among the 2 key genes and immune cells. (D) Correlation between GBP5, IFITM3 and Macrophages.M1, Dendritic.cells.activated, Neutrophils cells. (***, P < 0.001; **, P < 0.01; *, P < 0.05)