

## Research Article

# RTP4, a Biomarker Associated with Diagnosing Pulmonary Tuberculosis and Pan-Cancer Analysis

Hao Li <sup>1</sup>, Qin Zhou <sup>2</sup>, ZhiXiang Ding,<sup>1</sup> and QingHai Wang<sup>1</sup>

<sup>1</sup>Department of Infectious Diseases, The First People's Hospital of Changde City, Changde, China

<sup>2</sup>Intensive Care Unit, The First People's Hospital of Changde City, Changde, China

Correspondence should be addressed to Hao Li; 13721056589@163.com

Received 9 November 2022; Revised 17 December 2022; Accepted 27 March 2023; Published 26 April 2023

Academic Editor: Jinghua Pan

Copyright © 2023 Hao Li et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**Background.** Pulmonary tuberculosis (PTB) is a global epidemic of infectious disease; the purpose of our study was to explore new potential biomarkers for the diagnosis of pulmonary tuberculosis and to use the biomarkers for further pan-cancer analysis. **Methods.** Four microarray gene expression sets were downloaded from the GEO public databases and conducted for further analysis. Healthy control (HC) samples and samples of pulmonary tuberculosis (PTB) were calculated with enrichment scores in folate biosynthesis pathways. The scores acted as a new phenotype combined with clinical information (control or PTB) for subsequent analysis. Weight gene coexpression network analysis (WGCNA) was used to seek the modules mostly related to PTB and folate biosynthesis in training sets. Twenty-nine coexistence genes were screened by intersecting the genes in the green-yellow module of GSE28623 and the brown module of GSE83456. We used the protein-protein interaction network analysis to narrow the gene range to search for hub genes. Then, we downloaded the unified and standardized pan-cancer data set from the UCSC database for correlations between biomarkers and prognosis and tumor stage differences. **Results.** Eventually, RTP4 was selected as a biomarker. To verify the reliability of this biomarker, an area under the ROC (AUC) was calculated in gene sets (GSE28623, GSE83456, and GSE34608). Lastly, to explore the difference in RTP4 expression before and after antituberculosis treatment, the GSE31348 gene set was enrolled to compare the expressions in weeks 0 and 26. The results showed significant differences between these two time points ( $p < 0.001$ ). RTP4 was significantly upregulated in the pulmonary tuberculosis group compared to the healthy control group in three gene sets and downregulated after antituberculosis therapy in one gene set. These results suggest that RTP4 can be used as a potential biomarker in diagnosing tuberculosis. The results of pan-cancer analysis showed that high expression of RTP4 in 4 tumor types was positively correlated with poor prognosis and high expression of RTP4 in 6 tumor types was negatively correlated with poor prognosis. We found significant differences in the expression of the RTP4 gene at different stages in 5 types of tumors. **Conclusion.** RTP4 might be a new potential biomarker for diagnosing pulmonary tuberculosis.

## 1. Introduction

Mycobacterium tuberculosis (MTB) is one of the significant causes of tuberculosis. According to the World Health Organization, about 5.8 million people worldwide were infected with TB in 2020 [1]. Mycobacterium tuberculosis infection has become a global public problem, especially in developing countries [2].

MTB can escape immune surveillance and kill by inhibiting oxidative stress, autophagy, and apoptosis of cells. It also can affect antigen presentation of antigen-

presenting cells (APC) by inhibiting the synthesis of histocompatibility complex molecules [3]. There are many clinical trials used to test for diagnosing TB, such as tuberculin skin test (TST), T-SPOT, smear microscope, the culture of Mycobacterium tuberculosis, and chest X-ray [4, 5]. But as a slow-growing intracellular parasitic bacteria [6], the above clinical test methods have certain false negative and hysteresis.

To search for sensitive indicators in Mycobacterium tuberculosis infection, we use bioinformatics methods to compare and analyze the spectrum of gene expression

spectrum of pulmonary tuberculosis and healthy people, hoping to find genes that play a significant role in the pathogenesis of PTB as potential biomarkers for diagnosing pulmonary tuberculosis and further explore the relationship between biomarkers and tumor prognosis and stage through pan-cancer analysis.

## 2. Materials and Methods

**2.1. Data Acquisition and Processing.** We downloaded the GSE28623 (GPL14133, Agilent-014850 Whole Human Genome Microarray), GSE83456 (GPL10558, Illumina HumanHT-12 V4.0 expression beadchip), GSE34608 (GPL6480, Agilent-014850 Whole Human Genome U133 Microarray), and GSE31348 (GPL570, Affymetrix Human Genome U133 Plus) from Gene Expression Omnibus (GEO, <https://www.ncbi.nlm.nih.gov/geo/>) by “GEOquery” R-package and extracted each expression profile information and Clinical phenotype of these gene sets. There were 46 pulmonary tuberculosis samples and 37 healthy control samples in GSE28623 and 45 pulmonary tuberculosis samples and 61 healthy control samples in GSE83456. Finally, 189 blood samples were enrolled as training sets. There were 8 pulmonary tuberculosis samples and 16 healthy control samples in GSE34608 and 135 pulmonary tuberculosis samples in GSE31348. In the end, 159 blood samples were enrolled as validation sets. After that, we downloaded the unified and standardized pan-cancer data set from the UCSC (<https://xenabrowser.net/>) database, TCGA Pan-cancer (PANCAN,  $N = 10535$ ,  $G = 60499$ ), extracted expression data of the RTP4 gene (ENSG00000136514) in each sample, and carried out a  $\log_2(x + 0.001)$  transformation, and finally, we also eliminated cancer types with fewer than 3 samples in a single cancer type and finally obtained the expression data of 26 cancer types.

**2.2. KEGG and Gene Set Variation Analysis.** The KEGG was a set of databases that included information about biological mechanisms, cellular processes, chemical substances, and diseases [7]. Gene set variation analysis (GSVA) was applied to explore different activity variations of the KEGG pathway (c2.cp.kegg v7.5.1, <http://www.gsea-msigdb.org/>) in GSE28623 and GSE83456 by using “GSVA” R package [8]. “limma” R package was used to determine the significant differences in GSVA enrichment between the healthy control (HC) group and pulmonary tuberculosis (PTB) group. After setting the threshold value (PTB versus HC,  $\log_2FC > 0.25$ , adjust  $p$  value  $< 0.05$ ) and taking the intersection of the training set (GSE28623 and GSE83456), we selected the folate biosynthesis pathway enrichment scores for further analysis with clinical phenotype.

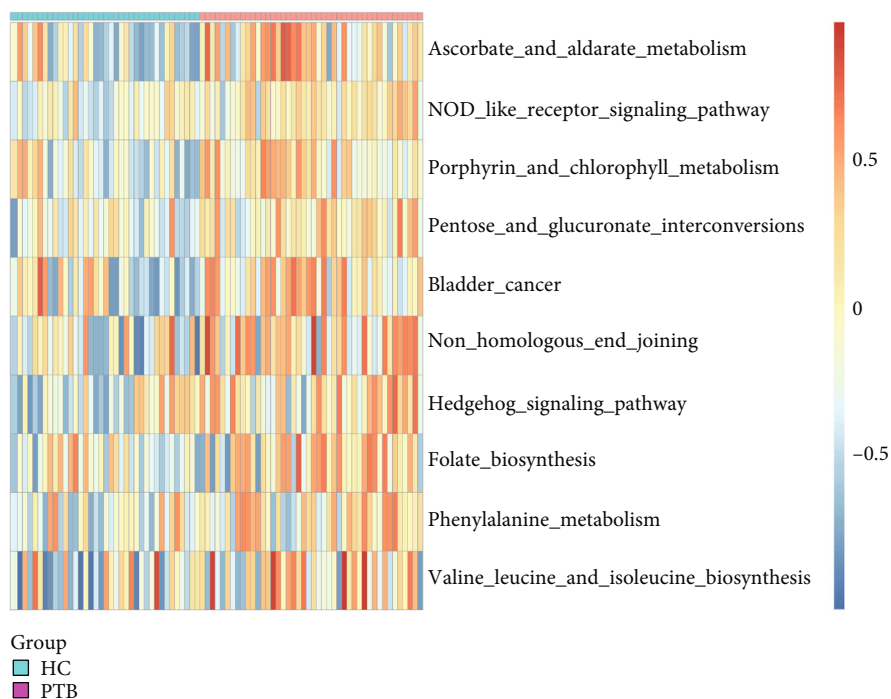
**2.3. Identification of Significant Modules and Weight Gene Coexpression Network Analysis.** Data selected in training sets were processed using the “WGCNA” R package [9]. After obtaining the gene expression matrix of GSE28623 and GSE83456 gene sets, we chose the genes that variance in the top quartile and used the hierarchical agglomerative

clustering (average link) to distinguish the outliers of each gene set. Then, we set each threshold to construct a scale-free network. After setting an approximate scale-free topology fit index above 0.85, we built an adjacency matrix and constructed the topological overlap matrix for searching the coexpression modules, which is a collection of genes with high topological overlap similarity and containing at least 30 genes. Genes in the same module usually have a higher degree of coexpression. The module eigengene (ME) represents the first principal component of modules and reflects the expression pattern of modules in each sample. After merging modules whose similarities were higher than 75%, 8 modules in GSE28623 and 13 modules in GSE83456 were identified according to the average hierarchical clustering and dynamic tree clipping algorithm. The green-yellow module of GSE28623 and brown module of GSE83456 were most relevant to the clinical phenotype and folate biosynthesis pathway enrichment.

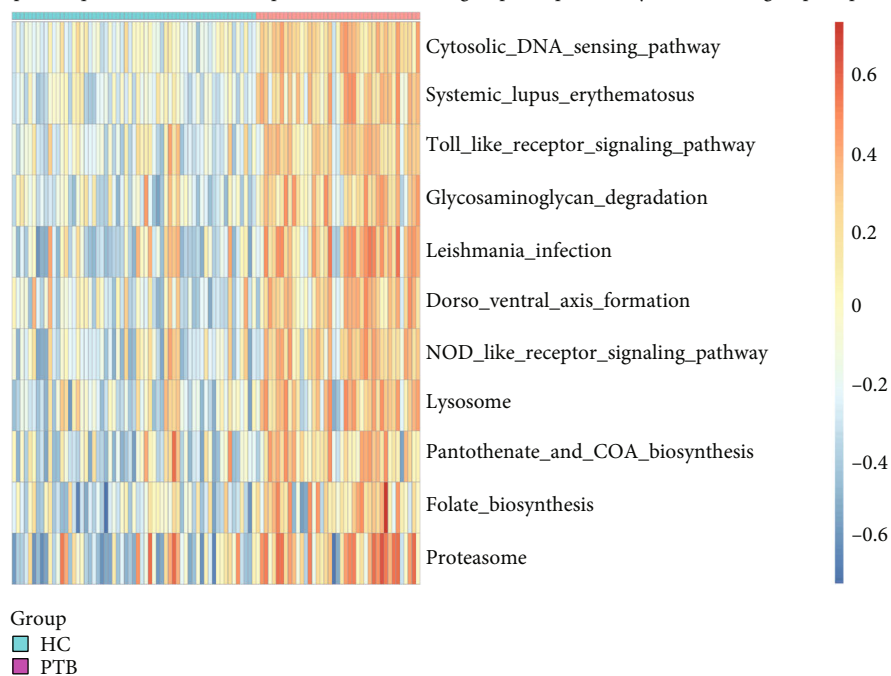
**2.4. Hub Gene Identification and PPI Network.** The coexisting genes in both modules were considered essential and uploaded to the Search Tool for the Retrieval of Interacting Genes (STRING) online database (<http://string-db.org>; version 11.5). A confidence score  $> 0.70$  was set as significant. After obtaining the protein-protein interaction (PPI) network data, we used the Cytoscape (version 3.9.1) to visualize the results. In order to verify the effectiveness of the hub genes, “pROC” and “ggplot2” were used to calculate and plot the area under curve (AUC) [10, 11].

**2.5. Pan-Cancer Analysis.** We used R software to calculate the expression difference between normal samples and tumor samples in each tumor to analyze the significance of the difference. The “Coxph” function of “survival” R package (version 3.4.0) was used to establish a Cox proportional hazard regression model to analyze the relationship between gene expression and prognosis in each tumor. After that, we calculated the expression difference of genes in different clinical stage samples in each tumor to explore the correlations between them.

**2.6. Statistical Analysis.** This study’s statistical analyses were carried out by R (version 4.1.0). Linear fitting and empirical Bayes, implemented in the “limma” R package, were used to test the difference between the HC and PTB groups [12]. Weight gene coexpression network analysis was used to allocate genes with similar expression patterns into different modules. The Pearson correlation analysis was applied to find correlations between the selected modules and clinical phenotype. The receiver operating characteristic (ROC) curve was used to evaluate the reliability of candidate biomarkers. It was considered statistically significant if the area under curve (AUC) was greater than 0.70. Student’s  $t$ -test was used to estimate the different expressions of RTP4 before and after antituberculosis therapy. All  $p < 0.05$  (bilateral) was considered statistically significant. The logarithmic ranking test was to obtain the tumor prognostic significance. The unpaired Student’s  $t$ -test was used for analyzing



(a) GSE28623: the different activated levels of Kyoto Encyclopedia of Genes and Genomes gene set between the healthy control and pulmonary tuberculosis groups. Turquoise and red lines represented the control group and pulmonary tuberculosis group, respectively



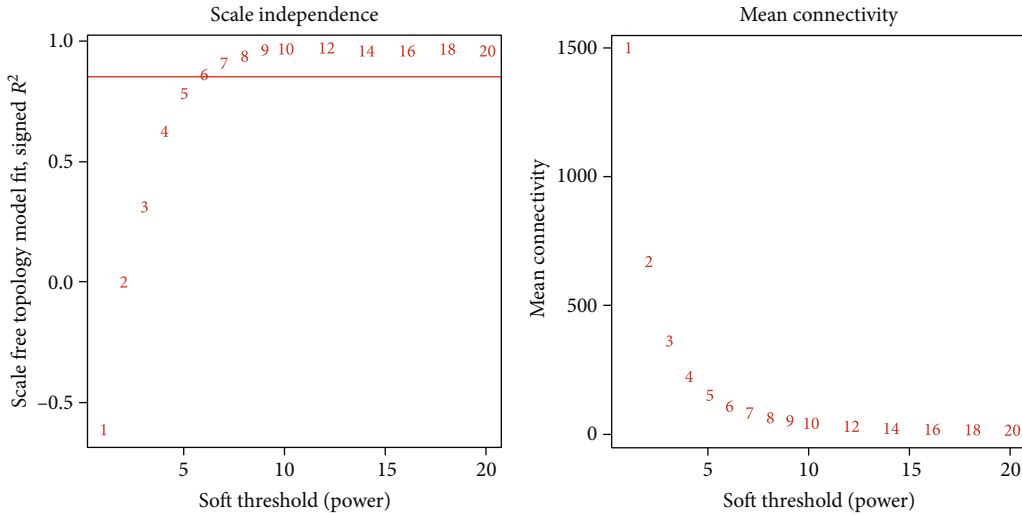
(b) GSE83456: the different activated levels of Kyoto Encyclopedia of Genes and Genomes gene set between the healthy control and pulmonary tuberculosis groups. Turquoise and red lines represented the control group and pulmonary tuberculosis group, respectively

FIGURE 1: Enrichment of differential KEGG pathway gene sets of training gene sets.

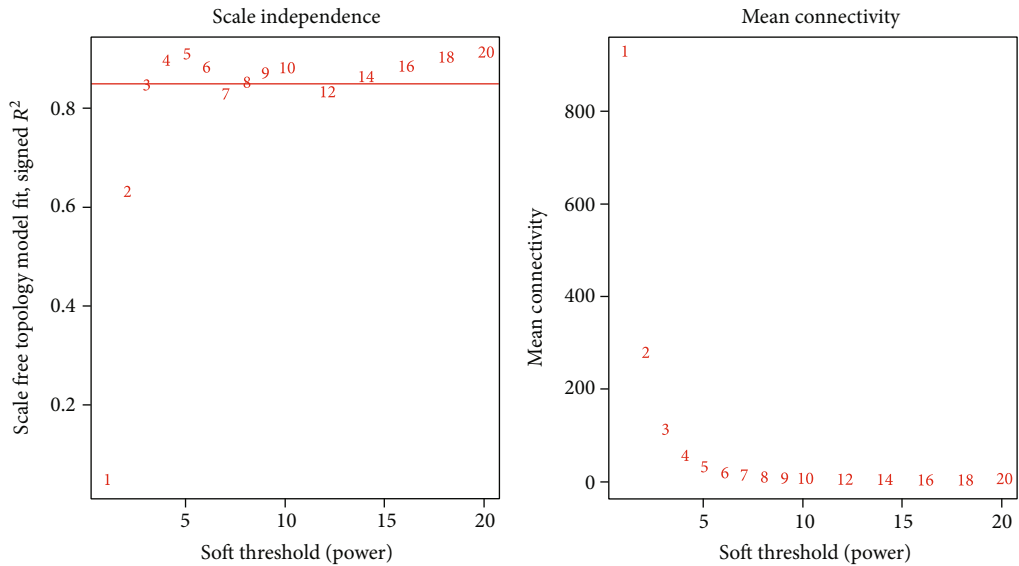
difference between each type of tumor groups. The unpaired Wilcoxon rank sum and signed rank tests were used to explore the significance of the difference between normal and tumor samples in each tumor type.

### 3. Results

3.1. Folate Biosynthesis Pathway Gene Set Scores and WGCNA. The gene expression of GSE28623 (19172 genes) and



(a) Analysis of the scale-free index and mean connectivity for various soft threshold powers in GSE28623



(b) Analysis of the scale-free index and mean connectivity for various soft threshold powers in GSE83456

FIGURE 2: Determination of soft threshold power in the WGCNA.

GSE83456 (20937 genes) was obtained after data preprocessing. After excluding latent tuberculosis, extrapulmonary tuberculosis, and sarcoid samples, we had 83 samples left in GSE28623 and 106 samples in GSE83456. We used an agglomerative hierarchical clustering algorithm to exclude 2 PTB samples of GSE28623 and 4 PTB samples of GSE83456 because there were outliers. Subsequently, the top quartile variance genes were selected (GSE28623, 4982 genes; GSE83456, 5234 genes). Through gene set variation analysis, we found that the KEGG pathways activated in GSE28623 (PTB group versus HC group) were mainly associated with ascorbate and aldarate metabolism, NOD-like receptor signaling, porphyrin and chlorophyll metabolism, pentose and glucuronate interconversions, bladder cancer, nonhomologous and joining, hedgehog signaling, folate biosynthesis, phenylalanine metabolism, valine leucine, and isoleucine biosynthesis. However, the main KEGG pathways

activated in GSE83456 (PTB group versus HC group) were cytosolic DNA sensing, systemic lupus erythematosus, toll-like receptor signaling, glycosaminoglycan degradation, leishmania infection, dorso-ventral axis formation, NOD-like receptor signaling, lysosome, pantothenate and CoA biosynthesis, folate biosynthesis, and proteasome. We observed the same activation trend both in GSE28623 ( $\log_2 \text{FC} = 0.286$ ,  $\text{adj.} p < 0.001$ ) and GSE83456 ( $\log_2 \text{FC} = 0.325$ ,  $\text{adj.} p < 0.001$ ) of folate biosynthesis pathway. The differential KEGG pathway gene set enrichment results were depicted by a heat map (Figure 1). We set the threshold power at 6 (GSE28623,  $R^2 = 0.850$ ) and 4 (GSE83456,  $R^2 = 0.895$ ) respectively based on an approximate scale-free topology fit index of above 0.85 for each gene set. This network conforms to the power-law distribution closer to the real biological network state (Figures 2(a) and 2(b)). Gene dendrograms and module colors of the training sets are shown in Figure 3. After

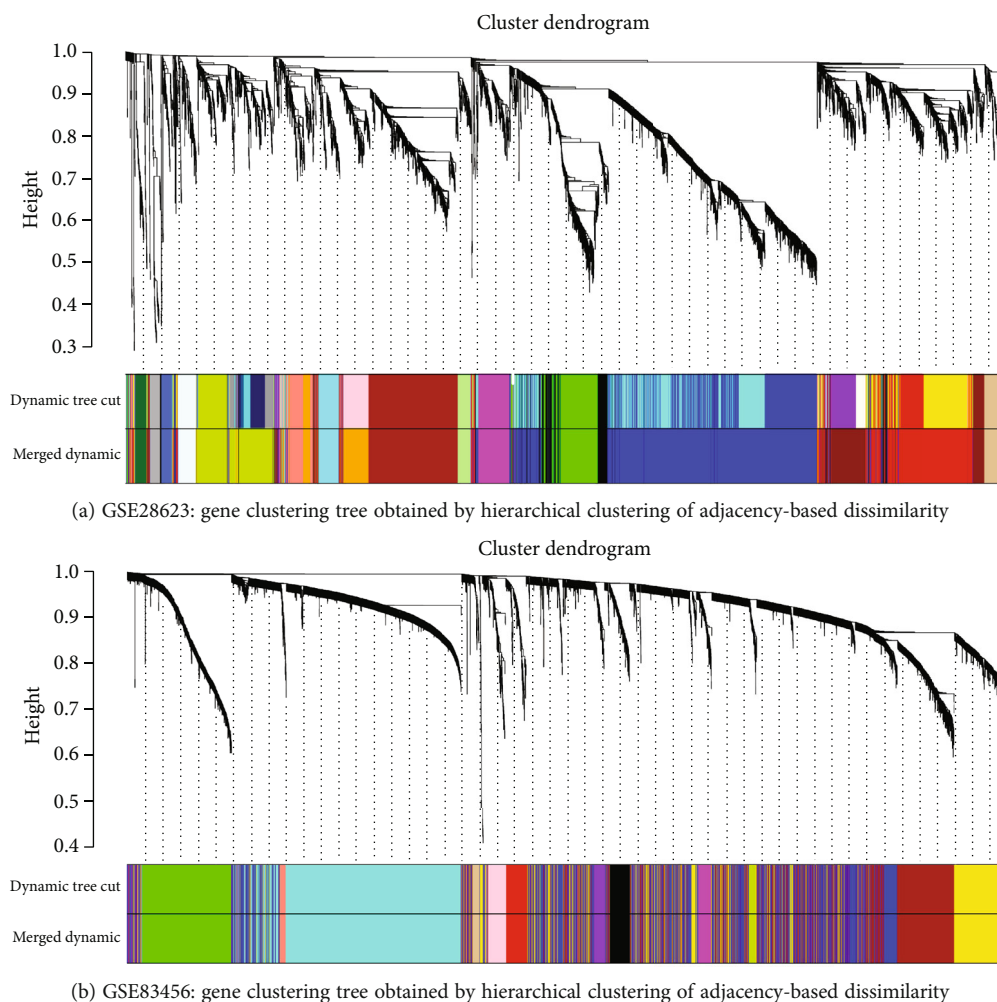


FIGURE 3: Dendrogram of all differentially expressed genes clustered based on the measurement of dissimilarity (1-TOM). The color band shows the results obtained from the automatic single-block analysis. Each color band of dynamic tree cut represents a cluster of a collection of genes with high topological overlap similarity and has minimal genes of 30. Merged dynamic represents the merged modules of dynamic tree cut, in which the dissimilarity degree is below 75%.

obtaining the different modules, we screened out that the most associated with pulmonary tuberculosis. After that, we selected the most relevant folate biosynthesis pathway of them. By carefully comparing, we found that the green-yellow module in GSE28623 and the brown module in GSE83456 were highly related to pulmonary tuberculosis and the folate biosynthesis pathway. Finally, we selected these two modules for further analysis (Figure 4).

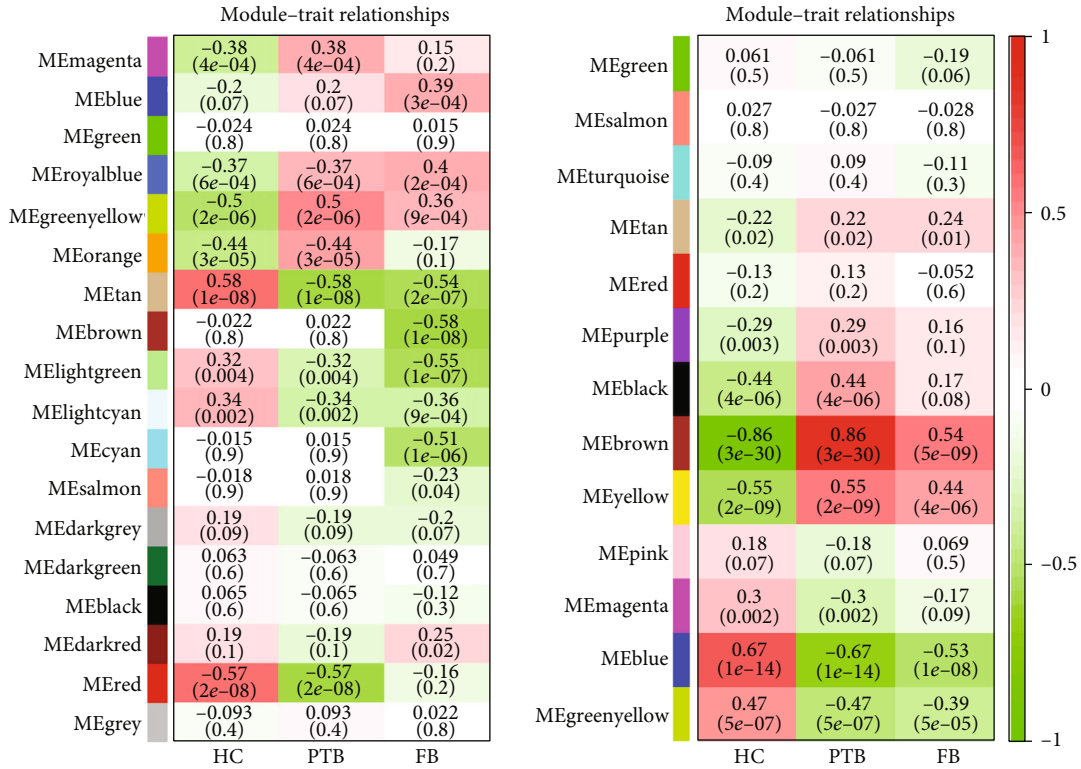
**3.2. PPI Network Analysis and Hub Gene Identification.** In the intersection of the two modules, we obtained 29 coexisting genes, which were illustrated by the Venn map (Figure 5). After using the STRING online database, eight nodes and sixteen edges were observed. Six genes (OAS2, SAMD9L, RSAD2, RTP4, CD151, and BATF) were filtered in the PPI network complex based on the setting confidence score (Figure 6).

We then screen one (RTP4) of them according to the area under curve both in training sets (GSE28623, AUC = 0.719; GSE83456, AUC = 0.964, Figures 7(a) and 7(b)) and

the validation cohort (GSE34608, AUC = 0.903 Figure 7(c)) as the hub gene. To make the outcome persuasive, we chose the gene set (GSE31384) to validate diversity expression before and after antituberculosis therapy and compare the differential expression of RTP4 between the HC and PTB groups in GSE28623 and GSE83456. As shown in Figure 8, the expression of RTP4 was significantly downregulated after 26 weeks of antituberculosis treatment. We also found that there were significantly upregulated in the PTB group (Figures 9 and 10).

**3.3. Pan-Cancer Analysis.** In the expression difference between normal samples and tumor samples in each tumor, we found that RTP4 was significantly upregulated in 13 types of tumor tissues such as GBM (tumor:  $2.22 \pm 1.14$ , normal:  $0.3 \pm 0.44$ ,  $p = 5.4e - 4$ ), GBMLGG (tumor:  $1.70 \pm 1.09$ , normal:  $0.35 \pm 0.44$ ,  $p = 3.2e - 3$ ), LGG (tumor:  $1.54 \pm 1.02$ , normal:  $0.35 \pm 0.44$ ,  $p = 5.5e - 3$ ), ESCA (tumor:  $2.78 \pm 1.49$ , normal:  $1.00 \pm 1.39$ ,  $p = 1.5e - 4$ ), STES (tumor:  $2.87 \pm 1.24$ , normal:  $1.63 \pm 1.20$ ,  $p = 1.7e - 10$ ), KIRP





(a) The module-trait relationships of the GSE28623 gene set

(b) The module-trait relationships of the GSE83456 gene set

FIGURE 4: Each row corresponds to the module eigengene column to a clinical trait. Each cell contains the corresponding correlation and  $p$  value, red for positive and green for negative correlations.

(tumor:  $2.78 \pm 1.34$ , normal:  $2.00 \pm 0.76$ ,  $p = 4.2e - 13$ ), KIPAN (tumor:  $3.15 \pm 1.27$ , normal:  $2.00 \pm 0.76$ ,  $p = 7.1e - 28$ ), STAD (tumor:  $2.91 \pm 1.12$ , normal:  $1.86 \pm 1.06$ ,  $p = 1.1e - 7$ ), HNSC (tumor:  $3.07 \pm 1.50$ , normal:  $1.34 \pm 1.51$ ,  $p = 6.5e - 11$ ), KIRC (tumor:  $3.44 \pm 1.10$ , normal:  $2.00 \pm 0.76$ ,  $p = 6.7e - 37$ ), LIHC (tumor:  $2.29 \pm 1.29$ , normal:  $1.50 \pm 1.19$ ,  $p = 1.3e - 6$ ), KICH (tumor:  $2.47 \pm 1.48$ , normal:  $2.00 \pm 0.76$ ,  $p = 2.8e - 3$ ), and CHOL (tumor:  $3.66 \pm 1.28$ , normal:  $1.25 \pm 0.56$ ,  $p = 6.0e - 7$ ) and significantly downregulated in 4 types of tumor tissues such as LUAD (tumor:  $2.19 \pm 1.22$ , normal:  $2.71 \pm 0.56$ ,  $p = 1.9e - 6$ ), COADREAD (tumor:  $2.08 \pm 1.13$ , normal:  $2.51 \pm 0.57$ ,  $p = 0.02$ ), COADREAD (tumor:  $2.02 \pm 1.10$ , normal:  $2.50 \pm 0.55$ ,  $p = 2.1e - 3$ ), and PRAD (tumor:  $1.50 \pm 1.27$ , normal:  $2.22 \pm 0.75$ ,  $p = 1.7e - 6$ ) (Figure 11) [13].

In the analysis process of the relationship between expression level and tumor prognosis, we found that the poor prognosis of 4 tumors was positively correlated with the high expression of RTP4 (GBMLGG ( $N = 619$ ,  $p = 1.5e - 13$ , HR = 1.58 (1.40, 1.78)), LGG ( $N = 474$ ,  $p = 2.5e - 5$ , HR = 1.49 (1.24, 1.79)), LAML ( $N = 144$ ,  $p = 1.2e - 3$ , HR = 1.30 (1.11, 1.53)), and PAAD ( $N = 172$ ,  $p = 8.8e - 4$ , HR = 1.44 (1.16, 1.78))), and the poor prognosis of 6 tumors was positively correlated with the low expression of RTP4 (SARC ( $N = 254$ ,  $p = 9.7e - 8$ , HR = 0.72 (0.64, 0.81)), KIRC ( $N = 515$ ,  $p = 3.2e - 3$ , HR = 0.82 (0.72, 0.94)), THCA ( $N = 501$ ,  $p = 0.01$ , HR = 0.43 (0.23, 0.83)), MESO ( $N = 84$ ,  $p = 6.1e - 6$ , HR = 0.68 (0.58, 0.81)), SKCM-M

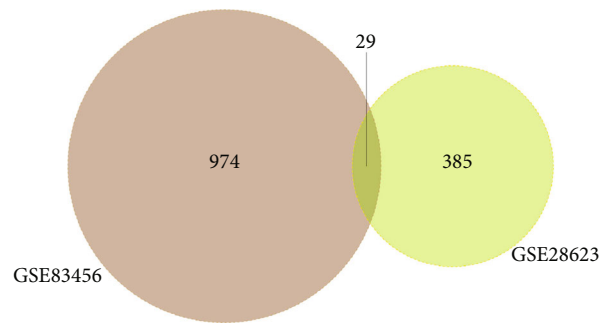


FIGURE 5: Coexisting genes exist in yellow-green modules (GSE28623) and brown modules (GSE83456). The intersection of the two modules.

( $N = 347$ ,  $p = 1.1e - 3$ , HR = 0.85 (0.78, 0.94)), and SKCM ( $N = 444$ ,  $p = 4.2e - 3$ , HR = 0.88 (0.80, 0.96))) (Figures 12 and 13).

After exploring the expression difference of the RTP4 gene in different clinical stages in each type of tumor, we found significant differences among the 5 types of tumors. (HNSC (stage I = 27, II = 82, III = 93, IV = 316) ( $p = 0.03$ ), LIHC (stage I = 169, II = 86, III = 85, IV = 5) ( $p = 0.04$ ), THCA (stage I = 283, II = 52, III = 112, IV = 55) ( $p = 2.7e - 3$ ), PAAD (stage I = 21, II = 147, III = 3, IV = 4) ( $p = 0.02$ ), and BLCA (stage II = 130, III = 140, IV = 133) ( $p = 0.02$ )) (Figure 14).

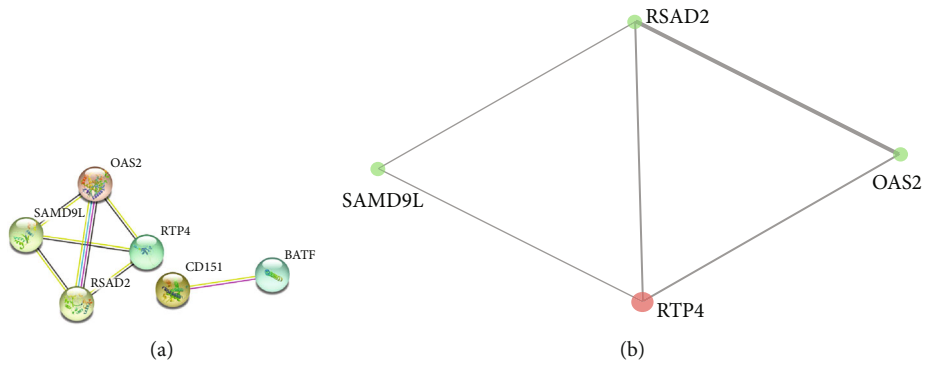


FIGURE 6: PPI network. (a) The interaction between 6 intersections coexisted; only 4 genes had the interaction. (b) The interaction between hub genes.

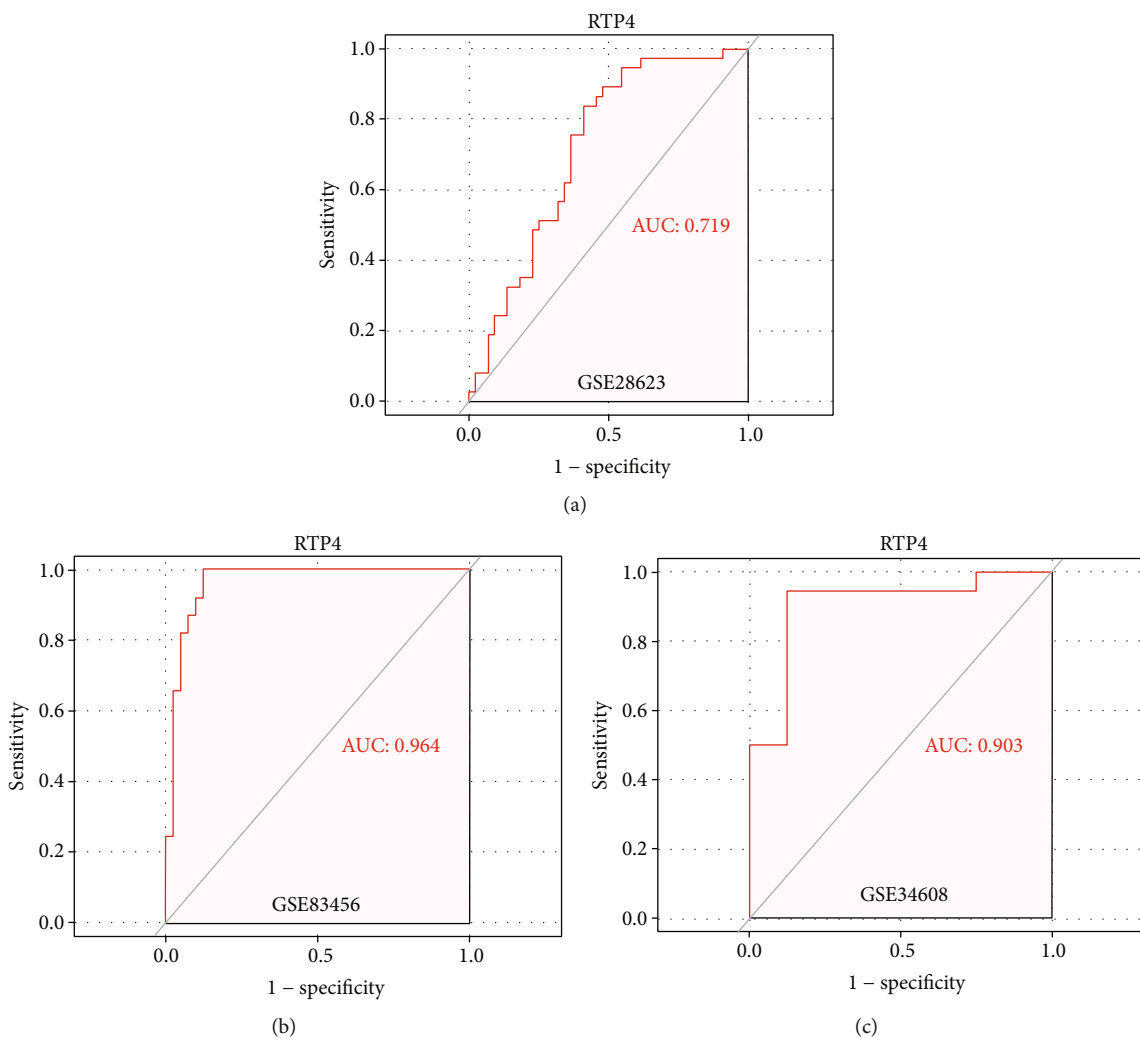


FIGURE 7: The different AUCs of RTP4 expression. AUC plot of RTP4 expression in GSE28623, GSE83456, and GSE34608, respectively.

#### 4. Discussion

MTB can infect many tissues and organs of the human body, causing various tuberculosis-related diseases, such as primary and secondary tuberculosis, tuberculous enteritis,

osteoarticular tuberculosis, tuberculous pleurisy, tubercular meningitis, and tuberculous lymphadenitis. Some diseases can be fatal because of delays in diagnosis and treatment.

Tuberculosis is a common infectious disease in developing countries. There are some clinical tests in diagnosing this

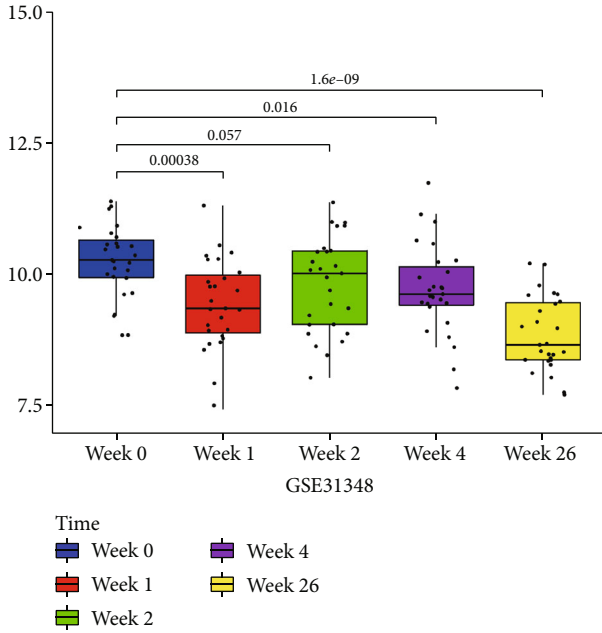


FIGURE 8: Diversity expression of RTP4 before and after antituberculosis therapy (five time points). RTP4 expression of different time points (week 0, week 1, week 2, week 4, and week 26) in GSE31348.

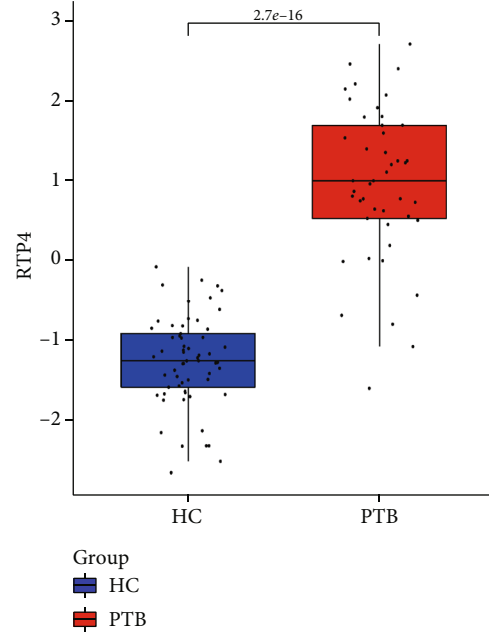


FIGURE 10: The different expressions of RTP4 between the HC group and PTB group in GSE83456.

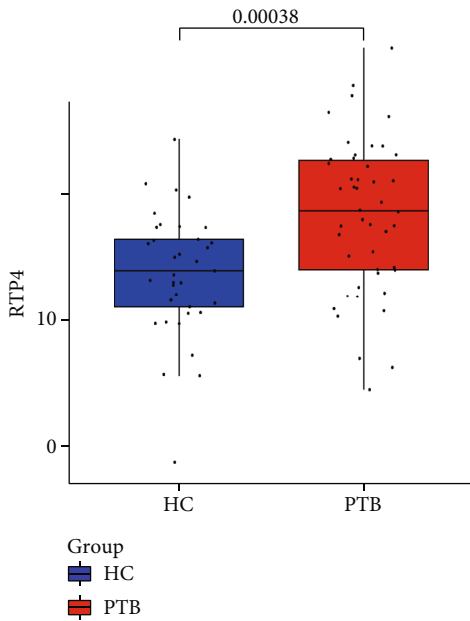


FIGURE 9: The different expressions of RTP4 between the HC group and PTB group in GSE28623.

disease, but those ways have somehow false negatives [14–16]. The histopathologic biopsy is an excellent method to diagnose tuberculosis infection. Due to the difficulty of sampling, some tissues in vivo and patients cannot tolerate the discomfort during the sampling process, so this method is not often used in clinical diagnosis. Sputum culture is considered the gold standard for diagnosing PTB infections.

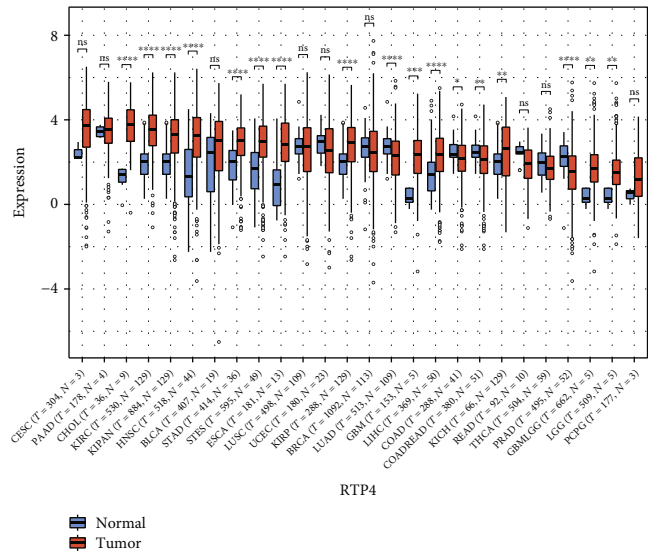


FIGURE 11: The different expressions of RTP4 in 26 types of tumors between normal samples and tumor samples. \*0.05, \*\*0.01, \*\*\*0.001, and \*\*\*\*0.0001. Red: tumor group. Blue: normal group.

However, as a slow-growing bacteria, clinical doctors may be able to get results in 2-3 months. We conducted this study to reduce the time of diagnosing and improve the diagnostic accuracy of PTB.

Several previous researchers have reported some potential biomarkers for tuberculosis [17, 18]. Our study is aimed at investigating new biomarkers for diagnosing pulmonary tuberculosis related to folate synthesis, and RTP4 was finally identified in training and validated in the validation set.



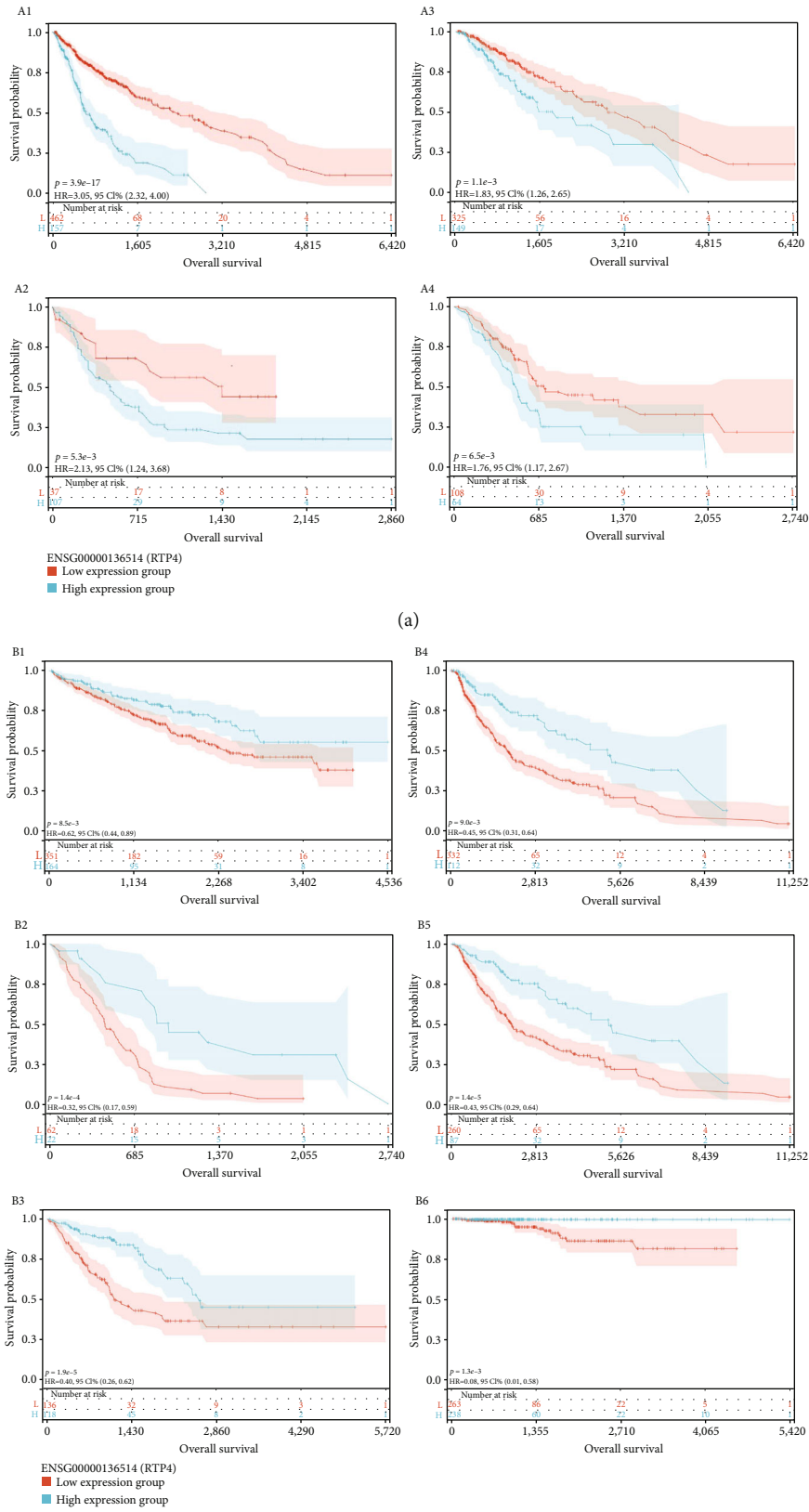


FIGURE 12: (A1–A4) The survival probability of the RTP4 low expression group and high expression group in GBMLGG, LAMA, LGG, and PAAD. (B1–B6) The survival probability of RTP4 low expression group and high expression group in KIRC, MESO, SARC, SKCM. Red: low expression group. Blue: high expression group.

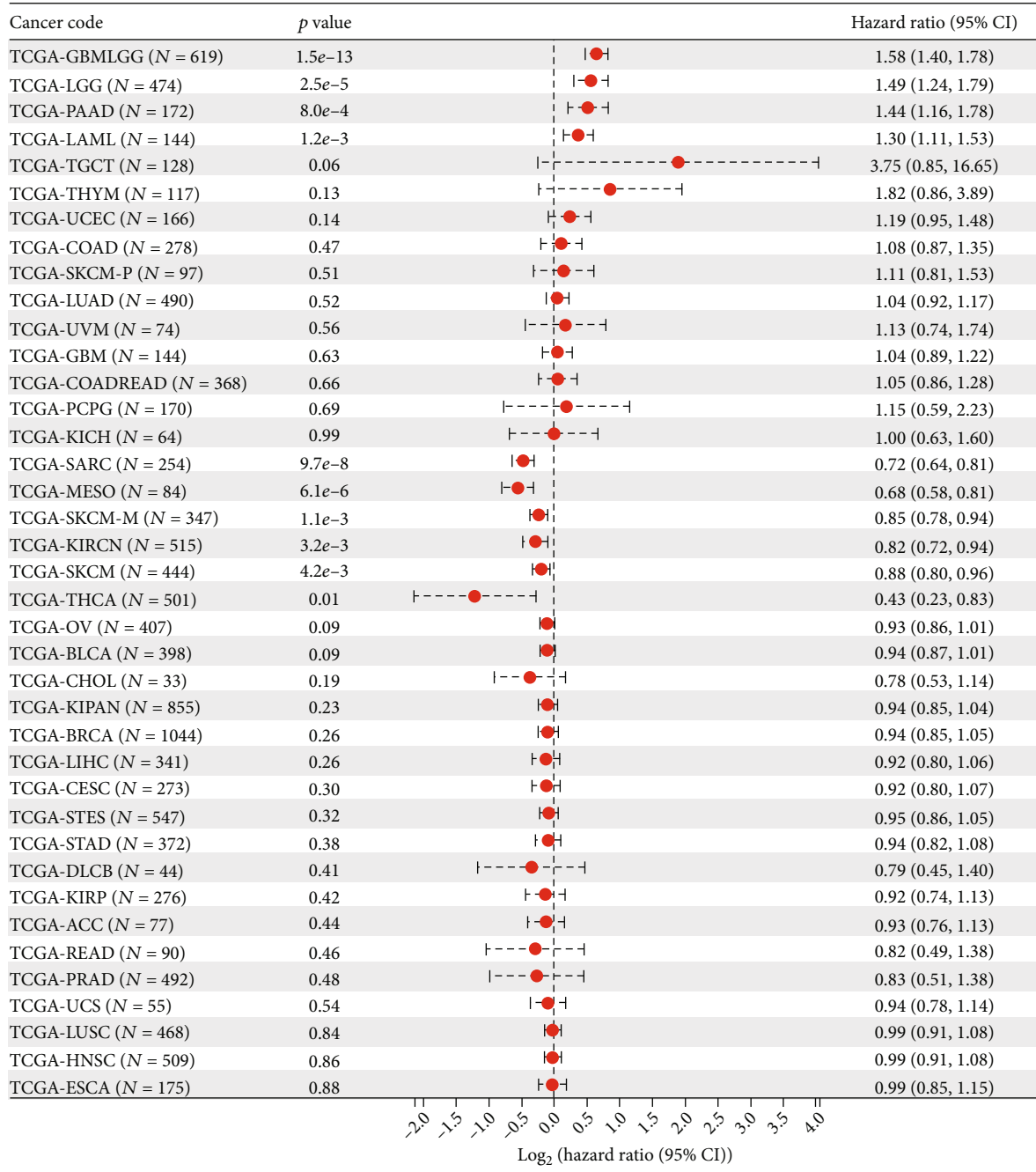


FIGURE 13: The relationship between RTP4 gene expression and prognosis in 39 types of tumors.

Folate acid is known as vitamin B9. It is necessary for the body to yield amino acids, RNA, and DNA [19]. It has previously been reported that lacking folic acid may lead to neuropathy and neoplastic diseases [20–22]. Folic acid is also an element for *M. tuberculosis*. It is believed that folate can affect the synthesis of purines and thymidine by regulating one-carbon transfer reactions as an essential factor, which is vital for bacteria [23, 24]. In recent years, there has been renewed interest in synthesizing antifolate acid drugs due to the increase in clinical drug-resistant tuberculosis cases [25, 26].

The RTP4 is known as a member of the receptor transport protein (RTP), which participates in exporting odorant and taste receptors to the cell surface [27, 28]. Previous studies have suggested that RTP4 helps the GPCR oligomer properly assemble in the endoplasmic reticulum, promotes the dimerization of receptors in the Golgi apparatus, and decreases the ubiquitination of the heterodimers [29]. Other studies have shown that the RTP4 was overexpressed in some connective tissue disorders and parasitic infections [30, 31] and correlated with some cancers [32, 33]. That RTP4 induced by IFN-I could explain this phenomenon and is associated with immune responses.

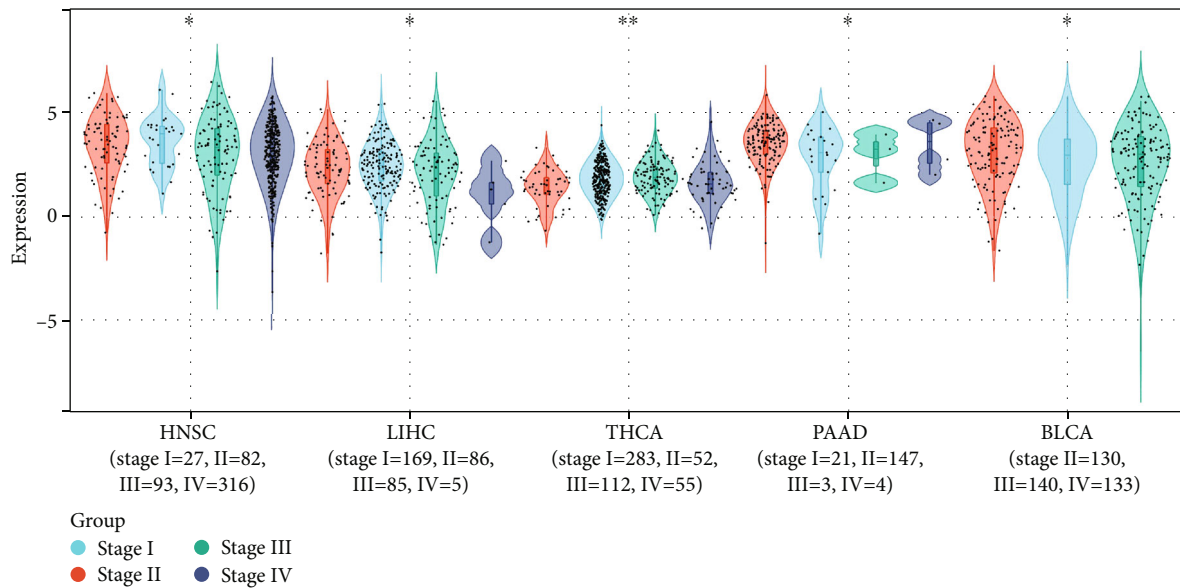


FIGURE 14: Differences in the expression of RTP4 in different stages in 5 types of tumors. \*0.05, \*\*0.01, \*\*\*0.001, and \*\*\*\*0.0001.

In our research, we also found several other enriched signaling pathways related to complications of PTB. Previous studies have implicated the NOD-like signaling pathway and Toll-like receptor signaling pathways in idiopathic pulmonary fibrosis [34]. The Hedgehog signaling pathway was reported to be associated with lung cancer. Abnormal activation of this pathway is thought to be related to the development of lung cancer and promotes malignant lung cancer progression with poor prognostic outcomes [35, 36]. These may explain some PTB patients' clinical progression of pulmonary fibrosis and lung cancer. However, in this study, we only found that the folate biosynthesis signaling pathways were most associated with the PTB group in training gene sets.

Some previous articles have reported a correlation between the RTP4 gene and certain characteristics of tumors. For example, methylated RTP4 can be considered as a new biomarker for precise diagnosis and treatment of prostate cancer [37] and RTP4 can be considered as an independent indicator to judge the prognosis of oral cancer [38]. RTP4 has also been reported that it is significantly associated with immune cell infiltration and immune checkpoint encoding genes (PDCD1, TIM-3, and LAG3) in melanoma [33].

We worked on seeking the different genes as biomarkers of diagnosing PTB infection through four GEO databases and used a method combining gene set variation analysis with weight gene coexpression network analysis. There were some limitations in this study. Firstly, the validated samples were less than the training samples, and more gene chip samples of patients with TB were validated from the experimental verification. To further explore the relationship between the expression of RTP4 and tumors, we performed a pan-cancer analysis based on TCGA database. It was also found that RTP4 expression was different between some tumor and normal tissues, and it was correlated with the overall survival time and stage of some tumors.

To conclude, our results suggest that RTP4 can be used as a biomarker to identify tuberculosis infection, providing a new perspective on TB diagnosis in clinical.

## Abbreviations

PTB:	Pulmonary tuberculosis
HC:	Healthy control
WGCNA:	Weight gene coexpression network analysis
ROC:	Receiver operating characteristic
GSEA:	Gene set variation analysis
PPI:	Protein-protein interaction network
MTB:	Mycobacterium tuberculosis
APC:	Antigen-presenting cells
TST:	Tuberculin skin test
STRING:	Search Tool for the Retrieval of Interacting Genes
AUC:	Area under curve
KEGG:	Kyoto Encyclopedia of Genes and Genomes
ACC:	Adrenocortical carcinoma
BLCA:	Bladder urothelial carcinoma
BRCA:	Breast invasive carcinoma
CEC:	Cervical squamous cell carcinoma and endocervical adenocarcinoma
CHOL:	Cholangiocarcinoma
COAD:	Colon adenocarcinoma
COADREAD:	Colon adenocarcinoma/rectum adenocarcinoma esophageal carcinoma
DLBC:	Lymphoid neoplasm diffuse large B-cell lymphoma
ESCA:	Esophageal carcinoma
FPPP:	FFPE pilot phase II
GBM:	Glioblastoma multiforme
GBMLGG:	Glioma
HNSC:	Head and neck squamous cell carcinoma
KICH:	Kidney chromophobe
KIPAN:	Pan-kidney cohort (KICH+KIRC+KIRP)

KIRC:	Kidney renal clear cell carcinoma
KIRP:	Kidney renal papillary cell carcinoma
LAML:	Acute myeloid leukemia
LGG:	Brain lower grade glioma
LIHC:	Liver hepatocellular carcinoma
LUAD:	Lung adenocarcinoma
LUSC:	Lung squamous cell carcinoma
MESO:	Mesothelioma
OV:	Ovarian serous cystadenocarcinoma
PAAD:	Pancreatic adenocarcinoma
PCPG:	Pheochromocytoma and paraganglioma
PRAD:	Prostate adenocarcinoma
READ:	Rectum adenocarcinoma
SARC:	Sarcoma
STAD:	Stomach adenocarcinoma
SKCM:	Skin cutaneous melanoma
STES:	Stomach and esophageal carcinoma
TGCT:	Testicular germ cell tumors
THCA:	Thyroid carcinoma
THYM:	Thymoma
UCEC:	Uterine corpus endometrial carcinoma
UCS:	Uterine carcinosarcoma
UVM:	Uveal melanoma.

## Data Availability

Data can be obtained from GEO (<https://www.ncbi.nlm.nih.gov/geo/>) and TCGA (<https://www.cancer.gov/ccg/research/genome-sequencing/tcga>).

## Ethical Approval

This study was approved by the Ethics Committee of The First People's Hospital of Changde City.

## Consent

Written informed consent for publication was obtained from all participants.

## Disclosure

The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

## Conflicts of Interest

The authors have no conflicts of interest to declare.

## Authors' Contributions

LH and ZQ performed analyses and drafted the manuscript. LH searched and downloaded the original gene sets from GEO. LH and ZQ contributed to statistical analyses. LH, DZX, and WQH edited the pictures. LH conceived and supervised the study. All authors contributed to writing the manuscript. All authors reviewed and approved the final manuscript. Hao Li and Qin Zhou contributed equally to this work.

## Acknowledgments

We sincerely thank Huihuang Li (Central South University, Changsha City, Hunan Province, China) for generously sharing his experience and codes. We sincerely thank all participants in the study.

## Supplementary Materials

Supplementary materials The difference in gene expression between the PTB group and the HC group are shown in supplementary Figure 1, Figure 2, and Figure 3 (supplementary Figure 1: GSE28623, supplementary Figure 2: GSE83456, and supplementary Figure 3: GSE 34608). (*Supplementary 1*)

## References

- [1] Organization, W H, *Global Tuberculosis Report*, World health organization, 2021.
- [2] A. D. Harries and A. M. V. Kumar, "Challenges and progress with diagnosing pulmonary tuberculosis in low- and middle-income countries," *Diagnostics (Basel)*, vol. 8, no. 4, p. 78, 2018.
- [3] S. Chauhan, M. A. Mandell, and V. Deretic, "Mechanism of action of the tuberculosis and Crohn disease risk factor IRGM in autophagy," *Autophagy*, vol. 12, no. 2, pp. 429–431, 2016.
- [4] S. Toonkomdang, P. Phinyo, B. Phetsuksiri, J. Patumanond, J. Rudeeaneksin, and W. Klayut, "Pragmatic accuracy of an in-house loop-mediated isothermal amplification (LAMP) for diagnosis of pulmonary tuberculosis in a Thai community hospital," *PLoS One*, vol. 15, no. 7, article e0236496, 2020.
- [5] S. E. Murthy, F. Chatterjee, A. Crook et al., "Pretreatment chest X-ray severity and its relation to bacterial burden in smear positive pulmonary tuberculosis," *BMC Medicine*, vol. 16, no. 1, p. 1, 2018.
- [6] S. Cole, R. Brosch, J. Parkhill et al., "Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence," *Nature*, vol. 393, no. 6685, pp. 537–544, 1998.
- [7] M. Kanehisa and S. Goto, "KEGG: Kyoto Encyclopedia of Genes and Genomes," *Nucleic Acids Research*, vol. 28, no. 1, pp. 27–30, 2000.
- [8] S. Hanzelmann, R. Castelo, and J. Guinney, "GSVA: gene set variation analysis for microarray and RNA-seq data," *BMC Bioinformatics*, vol. 14, no. 1, p. 7, 2013.
- [9] P. Langfelder and S. Horvath, "WGCNA: an R package for weighted correlation network analysis," *BMC Bioinformatics*, vol. 9, no. 1, p. 559, 2008.
- [10] X. Robin, N. Turck, A. Hainard et al., "pROC: an open-source package for R and S+ to analyze and compare ROC curves," *BMC Bioinformatics*, vol. 12, no. 1, 2011.
- [11] H. Wickham, *ggplot2: Elegant Graphics for Data Analysis*, Springer-Verlag, New York, 2016.
- [12] M. E. Ritchie, B. Phipson, D. Wu et al., "Limma powers differential expression analyses for RNA-sequencing and microarray studies," *Nucleic Acids Research*, vol. 43, no. 7, article e47, 2015.
- [13] S. Yu, H. Wang, T. Liu, C. Liang, and J. Luo, "A knowledge-driven network for fine-grained relationship detection between miRNA and disease," *Briefings in Bioinformatics*, vol. 23, no. 3, article bbac058, 2022.

- [14] L. Di and Y. Li, "The risk factor of false-negative and false-positive for T-SPOT.TB in active tuberculosis," *Journal of Clinical Laboratory Analysis*, vol. 32, no. 2, article e22273, 2018.
- [15] P. D. Davies and M. Pai, "The diagnosis and misdiagnosis of tuberculosis," *The International Journal of Tuberculosis and Lung Disease*, vol. 12, no. 11, pp. 1226–1234, 2008.
- [16] C. M. Bauer, A. Schmahl, and M. Kreuter, "Imaging and laboratory diagnostics for tuberculosis," *Klinische Monatsblätter für Augenheilkunde*, vol. 233, no. 5, pp. 587–593, 2016.
- [17] S. Natarajan, M. Ranganathan, L. E. Hanna, and S. Tripathy, "Transcriptional profiling and deriving a seven-gene signature that discriminates active and latent tuberculosis: an integrative bioinformatics approach," *Genes (Basel)*, vol. 13, no. 4, p. 616, 2022.
- [18] Z. Wen, L. Wu, L. Wang et al., "Comprehensive genetic analysis of tuberculosis and identification of candidate biomarkers," *Frontiers in Genetics*, vol. 13, article 832739, 2022.
- [19] M. Fenech, "Folate (vitamin B9) and vitamin B12 and their function in the maintenance of nuclear and mitochondrial genome integrity," *Mutation Research*, vol. 733, no. 1–2, pp. 21–33, 2012.
- [20] R. J. Berry, "Lack of historical evidence to support folic acid exacerbation of the neuropathy caused by vitamin B12 deficiency," *The American Journal of Clinical Nutrition*, vol. 110, no. 3, pp. 554–561, 2019.
- [21] J. R. Rees, C. B. Morris, J. L. Peacock et al., "Unmetabolized folic acid, tetrahydrofolate, and colorectal adenoma risk," *Cancer Prevention Research (Philadelphia, Pa.)*, vol. 10, no. 8, pp. 451–458, 2017.
- [22] J. M. Gong, Y. Shen, W. W. Shan, and Y. X. He, "The association between MTHFR polymorphism and cervical cancer," *Scientific Reports*, vol. 8, no. 1, p. 7244, 2018.
- [23] M. R. Nixon, K. W. Saionz, M. S. Koo et al., "Folate pathway disruption leads to critical disruption of methionine derivatives in *Mycobacterium tuberculosis*," *Chemistry & Biology*, vol. 21, no. 7, pp. 819–830, 2014.
- [24] Z. B. Li, L. Y. Shi, Y. S. Han et al., "Pyridoxal phosphate, pyridoxamine phosphate, and folic acid based on ceRNA regulatory network as potential biomarkers for the diagnosis of pulmonary tuberculosis," *Infection, Genetics and Evolution*, vol. 99, article 105240, 2022.
- [25] K. A. Wolff and L. Nguyen, "Strategies for potentiation of ethionamide and folate antagonists against *Mycobacterium tuberculosis*," *Expert Review of Anti-Infective Therapy*, vol. 10, no. 9, pp. 971–981, 2012.
- [26] Y. Minato, J. M. Thiede, S. L. Kordus, E. J. McKlveen, B. J. Turman, and A. D. Baughn, "Mycobacterium tuberculosis folate metabolism and the mechanistic basis for para-aminosalicylic acid susceptibility and resistance," *Antimicrobial Agents and Chemotherapy*, vol. 59, no. 9, pp. 5097–5106, 2015.
- [27] M. Behrens, J. Bartelt, C. Reichling, M. Winnig, C. Kuhn, and W. Meyerhof, "Members of RTP and REEP gene families influence functional bitter taste receptor expression," *The Journal of Biological Chemistry*, vol. 281, no. 29, pp. 20650–20659, 2006.
- [28] H. Saito, M. Kubota, R. W. Roberts, Q. Chi, and H. Matsunami, "RTP family members induce functional expression of mammalian odorant receptors," *Cell*, vol. 119, no. 5, pp. 679–691, 2004.
- [29] F. M. Decaillot, R. Rozenfeld, A. Gupta, and L. A. Devi, "Cell surface targeting of mu-delta opioid receptor heterodimers by RTP4," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 105, no. 41, pp. 16045–16050, 2008.
- [30] G. Chen, L. Che, X. Cai, P. Zhu, and J. Ran, "Bioinformatic analysis identifies biomarkers and treatment targets in primary Sjogren's syndrome patients with fatigue," *BioMed Research International*, vol. 2022, Article ID 7697558, 10 pages, 2022.
- [31] X. He, A. W. Ashbrook, Y. Du et al., "RTP4 inhibits IFN-I response and enhances experimental cerebral malaria and neuropathology," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 117, no. 32, pp. 19465–19474, 2020.
- [32] Y. Peng, C. Liu, M. Li et al., "Identification of a prognostic and therapeutic immune signature associated with hepatocellular carcinoma," *Cancer Cell International*, vol. 21, no. 1, p. 98, 2021.
- [33] Y. Li, J. Qi, and J. Yang, "RTP4 is a novel prognosis-related hub gene in cutaneous melanoma," *Hereditas*, vol. 158, no. 1, p. 22, 2021.
- [34] H. Wang, Q. Xie, W. Ou-Yang, and M. Zhang, "Integrative analyses of genes associated with idiopathic pulmonary fibrosis," *Journal of Cellular Biochemistry*, vol. 120, no. 5, p. 8648, 2019.
- [35] C. Ma, K. Hu, I. Ullah, Q. K. Zheng, N. Zhang, and Z. G. Sun, "Molecular mechanisms involving the sonic hedgehog pathway in lung cancer therapy: recent advances," *Frontiers in Oncology*, vol. 12, article 729088, 2022.
- [36] S. Huang, L. Yang, Y. An et al., "Expression of hedgehog signaling molecules in lung cancer," *Acta Histochemica*, vol. 113, no. 5, pp. 564–569, 2011.
- [37] N. Xu, Y. P. Wu, Z. B. Ke et al., "Identification of key DNA methylation-driven genes in prostate adenocarcinoma: an integrative analysis of TCGA methylation data," *Journal of Translational Medicine*, vol. 17, no. 1, p. 311, 2019.
- [38] A. Reyimu, Y. Chen, X. Song, W. Zhou, J. Dai, and F. Jiang, "Identification of latent biomarkers in connection with progression and prognosis in oral cancer by comprehensive bioinformatics analysis," *World Journal of Surgical Oncology*, vol. 19, no. 1, p. 240, 2021.