European Review for Medical and Pharmacological Sciences 2022; 26: 8777-8787

IFNG and IFNGR1 polymorphisms are associated with tuberculosis: a case-control study

S.-Q. WU, X.-J. DING, Q.-L. YANG, M.-G. WANG, J.-Q. HE

Department of Respiratory and Critical Care Medicine, West China Hospital, Sichuan University, Chengdu, Sichuan, China

Abstract. – OBJECTIVE: Previous studies suggested that single-nucleotide polymorphisms (SNPs) of interferon gamma (IFNG) and its receptor IFNGR1 may be involved in the pathogenesis of tuberculosis (TB). We aimed to examine the association of IFNG gene polymorphisms with TB in the Tibetan population and use the machine learning method to establish a clinical prediction model of TB.

PATIENTS AND METHODS: A total of 613 TB patients and 603 healthy controls were selected for the study. Associations between SNPs and TB were analyzed using logistic regression, adjusted for sex and age. Clinical data and SNPs were integrated to construct a TB prediction model using random forest (RF) machine learning.

RESULTS: For IFNG, rs1861494 CT was a protective factor against TB compared with TT genotype ($p = 0.010$). The rs1861494 C allele was a protective factor for TB ($p = 0.010$). For IFNGR1, the rs3799488 C allele reduced the risk of TB by 30% ($p < 0.001$). rs9376267 CT ($p = 0.005$) and TT ($p = 0.001$) genotypes were protective factors for TB. Compared with the rs1327475 GG genotype, the frequency of the GA genotype in the case group significantly differed from the controls ($p = 0.013$). rs2234711 GA ($p < 0.001$), AA ($p < 0.001$) genotype and A ($p < 0.001$) alleles were also associated with TB. Finally, five markers are identified using the RF model. The area under the curve (AUC) reaches 0.6 in the training set and 0.59 in the test set.

CONCLUSIONS: Our study found that IFNG and IFNGR1 gene polymorphisms were associated with TB in a Tibetan population. The results also demonstrate the potential of clinical-SNPs as diagnostic tools for TB.

Key Words:
Tuberculosis, IFNG, IFNGR1, Random forest.

Introduction

Tuberculosis (TB) is a chronic infectious disease and a major global health problem. TB is one of the leading causes of death worldwide, especially in Asia and Africa, and its fatality rate is second to the human immunodeficiency virus (HIV). The World Health Organization (WHO) 2021 reported that there were 9.87 million new TB cases globally in 2020 and 1.28 million TB-related deaths among HIV-negative people[1]. China ranks second in TB cases among countries with a high TB burden after India. However, only 5-10% of people infected with Mycobacterium TB (MTB) develop TB. Many reasons affect the outcome of TB infection, such as previous vaccination, exposure to microbes, malnutrition, and co-infection with other pathogens[2]. Recently, a series of case-control studies have found that genetic polymorphisms were associated with TB[3-5].

Anthropological studies on genes showed susceptibility to infectious disease is associated with genetic polymorphisms[6]. Furthermore, genetic variants play a crucial role in the progression of TB in humans[7]. Studies have shown that genetic heterogeneity contributes 39-71% to the development of TB[8]. Other factors that influence TB progression include adaptive immunity and innate immunity. It is suggested that immune-related genes, including IL1B, IL6 and TNF, contribute to the development of TB[9]. Therefore, the genetic factors associated with TB may be due to population-based differences in innate and adaptive immunity[10]. TB can be regulated by various immune cells and depends on the interaction of cytokines secreted by these cells[11].

The previous study suggests that the interferon gamma (IFNG) plays an essential role in TB progression[12]. IFNG is secreted by natural and T cells and is a critical T helper type 1 cytokine. IFNG knockout mice infected with MTB exhibited relatively higher MTB bacilli loads, while the expression level of reactive nitrogen intermediates decreased[13]. IFNG levels are elevated and can activate macrophages in the presence of MTB infection[14]. Additionally, IFNG polymorphism is associated with TB in different populations[15].

Corresponding Author: Jian-Qing He, Ph.D, MD; e-mail: jianqhe@gmail.com; jianqing_he@scu.edu.cn

IFNG exerts its effects by binding to two IFNG receptors (IFNGR1 and IFNGR2), and further by triggering a signaling cascade[16]. IFNG homodimers lead to receptor dimerization by interacting with both receptors. IFNGR1 is located on chromosome 6q23.4 and encodes the IFNGR ligand-binding chain (alpha). IFNGR1 was associated with multiple diseases such as chronic prostatitis[17], Behçet's disease[18], and gastric cancer[19]. As a critical gene in the IFNG signaling pathway, IFNGR1 was found to be associated with the pathogenesis of TB in previous study[20]. Defects in the IFNGR1 gene significantly increase the risk of MTB infection. Furthermore, it was suggested that IFNGR1 polymorphisms are associated with TB susceptibility in different populations[15,21].

Although the statistical approach can correct for the effects of unrandom allocation or confounding factors, it does not consider the potential interactions between variables. Machine learning is a new technique that can be applied in medical practice to help diagnose and determine the prognosis of diseases. It may be an innovative new way to predict TB[22]. Machine learning methods have been applied to diagnose and prognosis various diseases, including anti-TB drug-induced hepatotoxicity (ATDH), ovarian, breast and liver cancers[23,24].

We considered that our previous study found IFNG and IFNGR1 gene polymorphisms to be associated with TB but lack validation in independent populations. Therefore, in this study, we validated the association results in the Tibetan population. At the same time, we use the random forest (RF) machine learning method to predict TB.

Patients and Methods

Cases and Controls

Patients with TB and healthy controls were from the People's Hospital of the Aba Tibetan Autonomous Prefecture. The diagnosis of TB mainly depends on the symptoms of patients, sputum culture/smear/TB-DNA positive results, imaging, and the response to anti-TB drug therapy. The diagnosis of TB is based on WHO guidelines[1]. All participants with cancer, HIV, immune-related diseases, and other lung diseases were excluded. Included investigators must sign an informed consent form indicating their willingness to participate in the study. Subsequently, professional nurses draw 5ml of blood from their peripheral veins and store it at -80°C after centrifugation. We use a DNA extraction kit (Axygen Scientific Inc, Union City, CA, USA) to extract DNA from peripheral blood based on the manufacturer's instructions. This study has been approved by the Ethics Committees of the West China Hospital of Sichuan University. Our research follows the Declaration of Helsinki.

In this study, we selected Tag-SNPs for genotyping. Tag-SNPs selection criteria and SNPs genotyping refer to our previous published studies[21]. To control the genotyping quality, we randomly selected 5% of the samples to repeat the genotyping.

Statistical Analysis

Differences in the distribution of continuous variables between the two groups were tested using the student's *t*-test. Using the X^2-test to Test Hardy Weinberg equilibrium (HWE) and dichotomous variables. Associations between SNPs and TB were analyzed using logistic regression, adjusted for sex and age. Haplotype and pairwise linkage disequilibrium (LD) calculations were performed using the SHEsis online software platform (http://analysis.bio-x.cn). The statistics were done in SPSS version 19 (IBM; Armonk, NY, USA). $p < 0.05$ was considered the cut-off value of statistical differences.

Clinical data and SNPs data were sorted into CSV files for RF analysis. Stratified sampling divided the data into training cohorts (70%) and test cohorts (30%). Score the importance of each feature. The prediction model was established by selecting the appropriate variables by performing cross-validation of ten folds repeated five times. ROC curve was used to evaluate the accuracy of the model. Random Forest package of R 4.1.2 software (R Foundation for Statistical Computing, Vienna, Austria) was used for RF analysis.

Results

Demographics of the Participants and Results of Quality Control

A total of 613 TB patients (mean age, 4.53 ± 14.54 years; 392 males and 221 females) and 603 healthy controls (mean age, 34.63 ± 13.85 years; 404 males and 199 females) were selected for the study (Table I). All participants are Tibetans. There was no significant difference in gender and age between the two groups ($p > 0.05$).

Table I. Demographic distribution of healthy controls and tuberculosis patients.

Parameter	Cases, n = 613	Controls, n = 603	*p*-value
Age, (years)*	34.53±14.54	34.63±13.85	0.909
Male, n (%)	392 (63.9%)	404 (67.0%)	0.145

*Data are presented as mean ± SD.

Polymorphisms of the Three Genes in the Two Groups

As shown in Table II, two tag-SNPs of IFNG and four tag-SNPs of IFNGR1 were identified in the study. All tag-SNPs in the control group conformed to HWE. Table III shows the results of the association analysis of SNPs and TB. For IFNG, we only found that the rs1861494 polymorphism was associated with TB. rs1861494 CT was a protective factor against TB compared with TT genotype (OR = 0.73, 95%CI: 0.57-0.93; $p = 0.010$). The rs1861494 C allele was a protective factor for TB (OR = 0.80, 95%CI: 0.68-0.95; $p = 0.010$). rs1861494 was also related to TB in the dominant model (OR = 0.72, 95%CI: 0.58-0.91; $p = 0.005$).

For IFNGR1, four tag-SNPs were associated with TB in different genetic models. The rs3799488 C allele reduced the risk of TB by 30% (OR = 0.70, 95%CI: 0.57-0.85; $p < 0.001$). Compared with the rs3799488 TT genotype, the frequencies of CC (OR = 0.87, 95%CI: 0.80-0.95; $p = 0.001$) and CT (OR = 0.66, 95%CI: 0.52-0.84; $p = 0.001$) decreased significantly in the case group. rs9376267 CT (OR = 0.70, 95%CI: 0.54-0.90; $p = 0.005$) and TT (OR = 0.58, 95%CI: 0.41-0.81; $p = 0.001$) genotype were protective factors for TB. The C (OR = 0.75, 95%CI: 0.64-0.88; $p = 0.001$) allele is also associated with TB. Compared with the rs1327475 GG genotype, the frequency of the GA genotype in the case group significantly differed from the controls (OR = 0.67, 95%CI: 0.48-0.92; $p = 0.013$). In addition, the frequency of the A allele decreased in the case group (OR = 0.74, 95%CI: 0.55-0.99; $p = 0.045$). Finally, rs2234711 GA (OR = 1.65, 95%CI: 1.26-2.16; $p < 0.001$), AA (OR = 2.13, 95%CI: 1.54-2.93; $p < 0.001$) genotype and A (OR = 1.48, 95%CI: 1.26-1.73; $p < 0.001$) alleles were also associated with TB.

LD Patterns and Haplotype Analysis

Haplotypes analysis showed that IFNG AC haplotype was associated with TB. For IFNGR1, we found that both CTGG and TCAG haplotypes were associated with TB (Table IV). LD analyses showed that all SNPs did not have high LD ($r^2 < 0.8$) (Figure 1).

Subgroup Analysis

Stratified analysis of the included SNPs was performed based on a cut-off of 25 years (Table V)[25, 26]. For IFNG, in the female subgroup, rs2069718 and rs1861494 were associated with TB. rs1861494 was associated with TB in individuals aged < 25 years. For IFNGR1, in the male subgroup, rs3799488, rs9376267 and rs2234711 were associated with TB. In the female subgroup, rs3799488, rs9376267, rs1327475 and rs2234711 were related to TB. rs3799488, rs9376267, rs1327475 and rs2234711 were also associated with TB in the age ≥ 25 subgroups of the included population.

Diagnostic Potential of Tuberculosis Based on Polymorphisms

Clinical information of all participants and SNPs of IFNG and IFNGR1 were analyzed by RF.

Table II. Basic information of all SNPs in our study.

Gene/SNPs	chromosome	Location	Functional Consequence	MA	MAF	HWE
IFNG						
rs2069718	12	68550162	intron3	G	0.139	0.995
rs1861494	12	68551409	intron3	C	0.316	0.960
IFNGR1						
rs3799488	6	137519780	intron6	C	0.184	0.921
rs9376267	6	137531031	intron1	T	0.382	0.832
rs1327475	6	137536455	5'FLANKING	A	0.071	0.843
rs2234711	6	137540520	5'UTR_exon1	A	0.429	0.413

Abbreviation: SNP, single nucleotide polymorphism; MA, minor allele; MAF, minor allele frequency; HWE, Hardy Weinberg equilibrium.

Table III. Genotype distribution of *IFNG* and *IFNGR1* polymorphisms.

Gene/SNPs	Case (%), n = 613	Control (%), n = 603	p#	OR# (95% CI)
IFNG				
rs2069718A>G				
Genotype				
AA	429 (70.0)	447 (74.1)		
GA	164 (26.8)	144 (23.9)	0.193	1.19 (0.92-1.54)
GG	20 (3.3)	12 (2.0)	0.138	1.74 (0.84-3.60)
Allele				
A	1,022 (83.4)	1,038 (86.1)		
G	204 (16.6)	168 (13.9)	0.063	1.23 (0.99-1.54)
Genetic model				
Dominant			0.105	1.23 (0.96-1.58)
Recessive			0.171	1.66 (0.80-3.42)
rs1861494T>C				
Genotype				
TT	294 (48.0)	241 (40.0)		
CT	251 (40.9)	283 (46.9)	0.010	0.73 (0.57-0.93)
CC	68 (11.1)	79 (13.1)	0.063	0.71 (0.49-1.02)
Allele				
T	839 (68.4)	765 (63.4)		
C	387 (31.6)	441 (36.6)	0.010	0.80 (0.68-0.95)
Genetic model				
Dominant			0.005	0.72 (0.58-0.91)
Recessive			0.281	0.83 (0.59-1.17)
IFNGR1				
rs3799488T>C				
Genotype				
TT	411 (67.0)	343 (56.9)		
CT	179 (29.2)	226 (37.5)	0.001	0.66 (0.52-0.84)
CC	23 (3.8)	34 (5.6)	0.001	0.87 (0.80-0.95)
Allele				
T	1,001(81.6)	912 (75.6)		
C	225(18.4)	294 (24.4)	<0.001	0.70 (0.57-0.85)
Genetic model				
Dominant			<0.001	0.65 (0.51-0.82)
Recessive			0.121	0.65 (0.38-1.12)
rs9376267C>T				
Genotype				
CC	237 (38.7)	178 (29.5)		
CT	284 (46.3)	306 (50.7)	0.005	0.70 (0.54-0.90)
TT	92 (15.0)	119 (19.7)	0.001	0.58 (0.41-0.81)
Allele				
C	758 (61.8)	662 (54.9)		
T	468 (38.2)	544 (45.1)	0.001	0.75 (0.64-0.88)
Genetic model				
Dominant			0.001	0.67 (0.52-0.84)
Recessive			0.029	0.72 (0.53-0.97)
rs1327475G>A				
Genotype				
GG	532 (86.8)	494 (82.1)		
GA	75 (12.2)	104 (17.3)	0.013	0.67 (0.48-0.92)
AA	6 (1.0)	4 (0.7)	0.599	1.41 (0.39-5.03)
Allele				
G	1,139 (92.9)	1,092 (90.7)		
A	87 (7.1)	112 (9.3)	0.045	0.74 (0.55-0.99)
Genetic model				
Dominant			0.022	0.69 (0.51-0.95)
Recessive			0.547	1.48 (0.41-5.27)
rs2234711G>A				
Genotype				

Continued

Table III. Genotype distribution of *IFNG* and *IFNGR1* polymorphisms.

Gene/SNPs	Case (%), n = 613	Control (%), n = 603	p#	OR# (95% CI)
GG	137 (22.3)	204 (33.9)		
GA	307 (50.1)	279 (46.3)	<0.001	1.65 (1.26-2.16)
AA	169 (27.6)	119 (19.8)	<0.001	2.13 (1.54-2.93)
Allele				
G	581 (47.4)	687 (57.1)		
A	645 (52.6)	517 (42.9)	<0.001	1.48 (1.26-1.73)
Genetic model				
Dominant			<0.001	1.79 (1.38-2.3)
Recessive			0.001	1.55 (1.18-2.02)

SNPs, single nucleotide polymorphisms; CI, confidence interval; OR, odds ratio; #adjusted by age and sex status.

Briefly, a total of eight variables were included in the calculation. The model reached optimality when the decision tree = 500 (mtry = 2), and the error rate was 43.24% based on this parameter for classifying the training set data (Figure 2). Figure 2A ranks the critical variables, the larger the value, the more vital the importance of the variable. Among all SNPs in the model, rs1327475 was the most significant predictor of TB. Figure 2B shows that the error is the lowest when the first five variables are selected for model establishment. Figure 2C shows the random forest tree obtained by RF calculation. The X-axis represents the number of trees. Y-axis represents the cross-validation error. After model evaluation, the area under the ROC curve of the training set was 0.60 (0.58 - 0.63), and the area under the ROC curve of the test set was 0.59 (0.55 - 0.63). The results indicated that the model had moderate accuracy in predicting TB susceptibility (Figure 2D and Figure 2E). ROC curves' best accuracy, sensitivity and specificity were 0.58, 0.52 and 0.64, respectively.

Discussion

In this study, we assessed the association of IFNG and IFNGR1 gene polymorphisms with the risk of TB in a Tibetan population. The results of multiple logistic regression suggested that IFNG

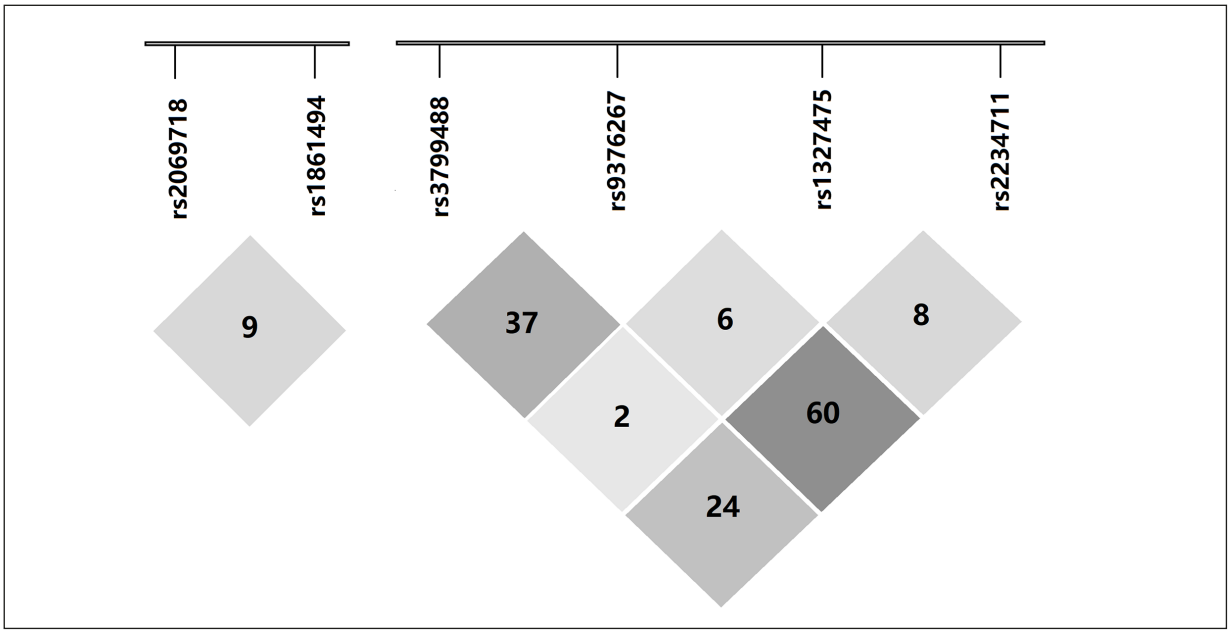

Figure 1. Pairwise linkage disequilibrium (LD) of IFNG and IFNGR1 gene polymorphisms. LD r2 values (range from 0 to 1) for all pairs of SNPs are presented as percentages. Shading from white to black indicates LD measured as r2 (range from 0 to 1).

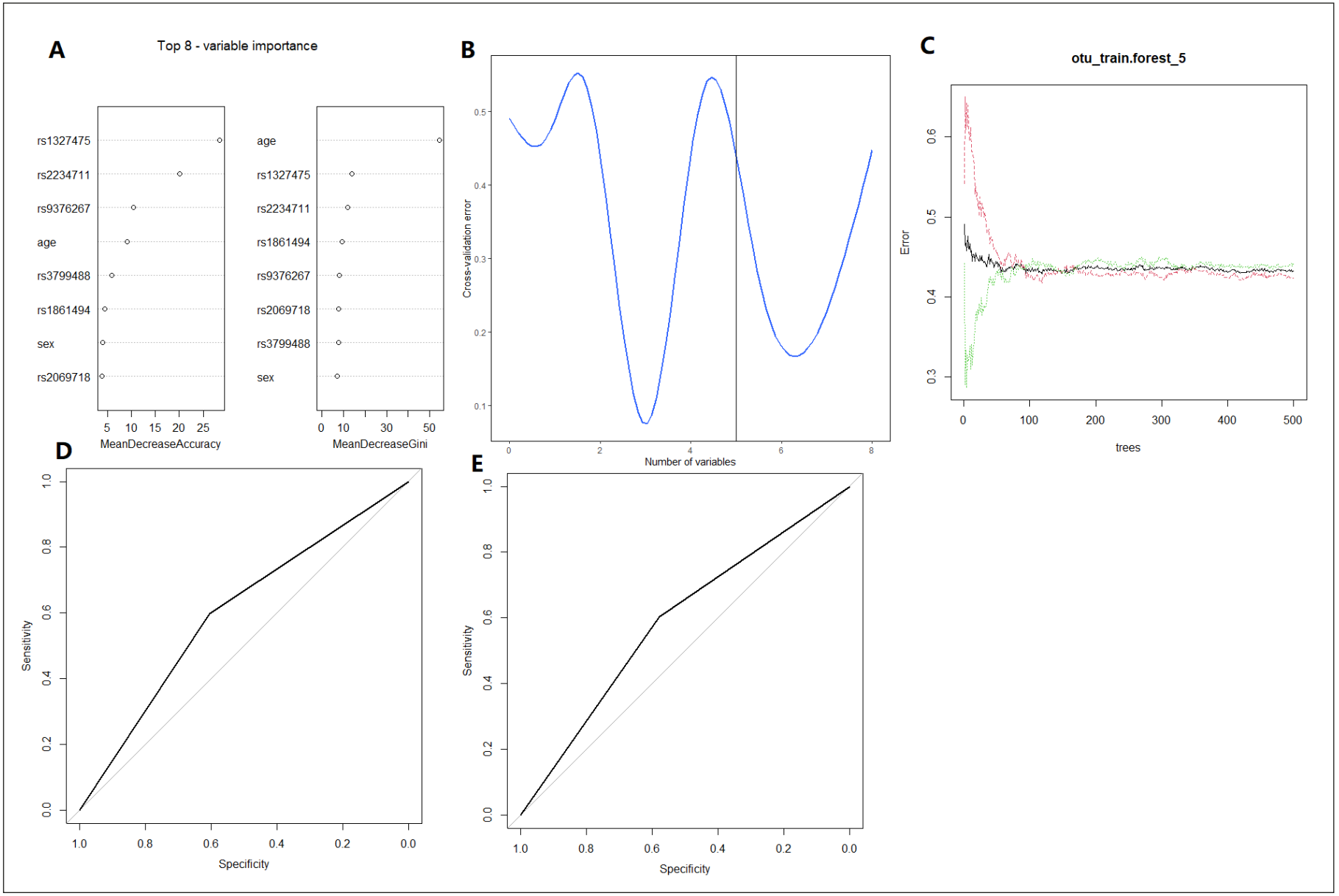

Figure 2. ***A***, Parameter importance score chart. The Mean Decrease Accuracy and Mean Decrease Gini value of the first eight critical variables, the larger the value is, the more important the index is; **B**, Cross verification curve. The lowest errors were found when the first five variables were selected for model building; **C**, Random forest trees. The X-axis represents the number of trees. Y-axis represents the cross-validation error. The lower dashed line represents the control group error, the upper dashed line represents the experimental group error, and the middle dashed line represents all sample errors; **D**, Training set ROC curve; **E**, test set ROC curve.

Table IV. Haplotype analyses in this study.

Gene/haplotype	Case (%), n=1,226	Control (%), n=1,204	p	OR (95% CI)
IFNG				
AC	386.9 (31.6)	441.0 (36.6)	0.009	0.80 (0.68-0.95)
AT	635.1 (51.8)	597.1 (49.5)	0.258	1.20 (0.94-1.29)
GT	203.9 (16.6)	168.0 (13.9)	0.064	1.23 (0.99-1.54)
Other*	0.1 (0.0)	0.1 (0.0)		
IFNGR1				
CTGG	223.9 (18.3)	294.0 (24.4)	<0.001	0.70 (0.57-0.85)
TCAG	87.0 (7.1)	111.9 (9.3)	0.052	0.75 (0.56-1.00)
TCGA	634.7 (51.8)	510.9 (42.4)	<0.001	1.47 (1.25-1.73)
TCGG	35.2 (2.9)	37.2 (3.1)	0.763	0.93 (0.58-1.49)
TTGG	233.9 (19.1)	243.8 (20.3)	0.497	0.93 (0.76-1.14)
Other*	11.4 (0.9)	6.17 (0.5)		

CI, confidence interval; OR, odds ratio. *Those lowest frequency threshold (LFT) < 0.03 were pooled in this part.

and IFNGR1 gene polymorphisms were associated with TB. At the same time, the RF method combined with clinical data and SNPs was used to construct a TB prediction model.

When the human body is infected with MTB, the innate immune response expressed by natural killer cells (NK) and NK T cells producing IFNG will be activated. Once specific antigen immunity is established, CD4 and CD8 T cells will secrete IFNG[27]. A study has shown that IFNG knockout mice are more susceptible to MTB infection than wild type mice[13]. In addition to mouse studies, a clinical study has shown that IFNG plays a vital role in human MTB infection[28]. The above evidence points to the critical role of IFNG in TB. Therefore, IFNG gene polymorphism has become a reliable candidate marker for TB.

Studies have shown that IFNG is a candidate gene for TB. However, the results have been inconsistent. rs2069718 located in intron 3 was reportedly related to Chronic prostatitis/chronic pelvic pain syndrome[17] and TB[29]. However, in our study, rs2069718 was not associated with TB. It has been proposed that rs1861494 can alter gene transcription and further functionally alter IFNG expression levels[30]. rs1861494 has been reported to be associated with various diseases, including IgA nephropathy[31], inflammatory bowel disease[32] and asthma[33]. Several studies have explored the relationship between rs1861494 and TB, but the results are inconsistent. A study in Argentina showed that rs1861494 was related to TB in a dominant model[34]. Another study has shown that the GG genotype is associated with TB[29]. Our previous study in the Han population revealed that the C allele is a risk factor for TB[21]. However, the C allele is a protective factor for TB in the Tibetan population in this study. The results of these differences may be attributed to ethnic differences.

The IFNG signaling pathway is regulated by the ligand binding to IFNGR1. Some scholars[35] have proposed that IFNGR1 gene mutations may be associated with MTB infection. rs3799488 has been shown to be associated with HBV, and CT/CC genotypes were a high-risk factor for HBV[36]. Another study[37] showed that rs3799488 was associated with rectal cancer. In this study, CT, CC and the C allele were protective factors for TB[38]. rs9376267 located in intron 1 is associated with TB risk under a recessive model[39]. rs9376267 is a protective factor of TB in this study under different gene models. rs1327475 has been shown to be a risk factor for TB in a previous study[40]. Contrary to their results, our study showed that rs1327475 was a protective factor for TB. rs2234711 is located in the 5'-UTR region of IFNGR1 and encodes human IFNGR1 ligand binding chain 1. It has been found that the conversion of T to C in the promoter region of rs2234711 may reduce the expression level of IFNGR1 on the cell surface[41]. A study in Africa found that the minor allele of rs2234711 was relatively low in the TB group, suggesting its protective role in TB[42]. Another study in China[43] showed that the rs2234711 C allele is a protective factor against TB. Our findings are basically consistent with them. However, the results of IFNGR1 gene polymorphisms in Tibetan are inconsistent with our previous studies. The difference results need to consider racial differences and may be attributed to the difference in minimum allele frequency.

Table V. Subgroup analysis of *IFNG* and *IFNGR1* polymorphisms and TB.

Gene/SNPs	Genetic model	p#	OR#(95% CI)
IFNG			
rs2069718A>G	allele		
Male		0.529	1.10 (0.82-1.48)
Female		0.033	1.44 (1.03-2.01)
<25		0.197	1.29 (0.88-1.89)
≥25		0.202	1.20 (0.91-1.57)
rs1861494T>C	allele		
Male		0.164	0.85 (0.68-1.07)
Female		0.013	0.73 (0.57-0.94)
<25		0.002	0.62 (0.46-0.83)
≥25		0.264	0.89 (0.73-1.09)
IFNGR1			
rs3799488T>C	allele		
Male		0.028	0.74 (0.57-0.97)
Female		0.003	0.65 (0.49-0.86)
<25		0.077	0.73 (0.51-1.04)
≥25		0.001	0.68 (0.54-0.86)
rs9376267C>T	allele		
Male		0.023	0.77 (0.62-0.97)
Female		0.006	0.71 (0.56-0.91)
<25		0.642	0.93 (0.70-1.25)
≥25		<0.001	0.68 (0.56-0.82)
rs1327475G>A	allele		
Male		0.319	1.19 (0.85-1.67)
Female		0.005	1.69 (1.17-2.43)
<25		0.298	1.28 (0.81-2.03)
≥25		0.014	1.45 (1.08-1.94)
rs2234711G>A	allele		
Male		0.003	1.39 (1.13-1.72)
Female		0.001	1.52 (1.12-1.93)
<25		0.129	1.25 (0.94-1.65)
≥25		<0.001	1.54 (1.27-1.86)

SNP, single nucleotide polymorphism; CI, confidence interval; OR, odds ratio. #adjusted by age and sex status.

At the same time, we performed haplotype analysis. IFNG AC and IFNGR1 CTGG haplotypes are protective factors against TB. IFNGR1 TCGA haplotype is a risk factor for TB. In addition, we also conducted subgroup analysis according to gender and age. Interestingly, rs3799488, rs9376267 and rs2234711 were associated with TB in the female group. In the male group, rs2069718, rs1861494, rs3799488, rs9376267, rs1327475 and rs2234711 were associated with TB. In the age < 25 groups, only rs1861494 was related to TB. In the age ≥ 25 groups, rs3799488, rs9376267, rs1327475, and rs2234711 were associated with TB.

Based on the clinical and genomic data, previous researchers[23] compared the accuracy of various machine learning methods in predicting ATDH. The artificial neural network with clinical and genomic factors showed the best prediction performance, with an accuracy of 88.67% and a sensitivity of 80% in the test set. Another study examined the gut microbiotas in patients with chronic kidney disease. Five best microbial markers were identified using an RF model, with an area under the curve of 0.99 and 0.95 in the discovery and validation cohorts, respectively[44]. These studies illustrate the successful application of machine learning methods to predict the occurrence of adverse events in multifactorial diseases. Some studies[45-47] also use random forest machine learning methods to diagnose TB or latent TB infection. In contrast to previous studies, our study combined clinical and SNPs data to establish a TB prediction model using machine learning for the first time. We analyzes multiple variables based on the RF model, scores the importance of each variable, and obtains the optimal combination of variables to construct the TB prediction model. The results of the training set and test set were consistent, indicating reliable results and specific clinical application values.

Our study also has some limitations. First, the SNPs associated with TB lacked functional validation. Furthermore, the area under the ROC curve in the training and test sets was not very high, and the prediction model was only moderately accurate in predicting TB disease. The results need to be validated in a larger population. Finally, multiple machine learning methods were not used to analyze and compare clinical data and SNPs in TB.

Conclusions

Our study found that IFNG and IFNGR1 gene polymorphisms were associated with TB in a Tibetan population group. The RF combined model with clinical data and genetic risk factors generated the best prediction in TB.

Conflicts of Interest

The authors declare that they have no conflict of interests.

Acknowledgments

None.

Informed Consent

All participants signed a consent form.

Authors' Contribution

Conceived and designed the experiments: JQH. Analyzed the data: SQW JQH. Contributed reagents/materials/analysis tools: SQW QLY XJD. Wrote the paper: SQW JQH. Obtained ethical permission for the use of urine when taking samples: MGW XJD..

Funding

The National Natural Science Foundation of China [grant number 81870015] supported the research. The Medical Science and Technology Project of Sichuan Provincial Health Commission [grant number 21PJ025] supported the research. Post-Doctor Research Project, West China Hospital, Sichuan University [grant number 2021HXBH075] supported the research.

ORCID ID

Ming-Gui Wang: https://orcid.org/0000-0002-4132-2925.

References

1) World Health O. Global tuberculosis report 2021 Geneva: World Health Organization, 2021.
2) van de Vosse E, Hoeve MA, Ottenhoff TH. Human genetics of intracellular infectious diseases: molecular and cellular immunity against mycobacteria and salmonellae. Lancet Infect Dis 2004; 4: 739-749.
3) El-Masry EA, Taher I, Hetta HF, Eldahdouh SS. Pulmonary tuberculosis susceptibility and association with Toll-Like receptor 2 Arg753Gln polymorphism. J Infect Dev Ctries 2022;16: 125-133.
4) Li B, Wen F, Wang Z. Correlation between polymorphism of vitamin D receptor TaqI and susceptibility to tuberculosis: An update meta-analysis. Medicine 2022; 101: e29127-e29127.
5) Li T, Han S. Association of Single Nucleotide Polymorphism rs17580 with Smoking and Pulmonary Tuberculosis. J Healthc Eng 2022; 2022: 6984403.
6) Shahsavar F, Varzi AM, Ahmadi SAY. A genomic study on distribution of human leukocyte antigen (HLA)-A and HLA-B alleles in Lak population of Iran. Genom Data 2017; 11: 3-6.
7) Wu S, Wang M, Wang Y, Zhang M, He JQ. Polymorphisms in the STAT4 gene and tuberculosis susceptibility in a Chinese Han population. Microb Pathog 2019; 128: 288-293.
8) Newport MJ, Goetghebuer T, Weiss HA, Whittle H, Siegrist CA, Marchant A; MRC Gambia Twin Study Group. Genetic regulation of immune responses to vaccines in early life. Genes Immun 2004; 5: 122-129.
9) Wu S, Wang MG, Wang Y, He JQ. Polymorphisms of cytokine genes and tuberculosis in two independent studies. Sci Rep 2019; 9: 2507.
10) Bulat-Kardum LJ, Etokebe GE, Lederer P, Balen S, Dembic Z. Genetic Polymorphisms in the Toll-like Receptor 10, Interleukin (IL)17A and IL17F Genes Differently Affect the Risk for Tuberculosis in Croatian Population. Scand J Immunol 2015; 82: 63-69.
11) Munk M, Emoto M. Functions of T-cell subsets and cytokines in mycobacterial infections. Eur Respir J Suppl 1995; 20: 668s-675s.
12) Collins HL, Kaufmann SH. The many faces of host responses to tuberculosis. Immunology 2001; 103: 1-9.
13) Cooper AM, Dalton DK, Stewart TA, Griffin JP, Russell DG, Orme IM. Disseminated tuberculosis in interferon gamma gene-disrupted mice. J Exp Med 1993; 178: 2243-2247.
14) Lee J, Kornfeld H. Interferon-γ Regulates the Death of M. tuberculosis-Infected Macrophages. J Cell Death 2010; 3: 1-11.
15) Wu S, Liu X, Wang Y, Zhang M, Wang M, He JQ. Genetic Polymorphisms of IFNG and IFNGR1 with Latent Tuberculosis Infection. Dis Markers 2019; 2019: 8410290.
16) Greenlund AC, Schreiber RD, Goeddel DV, Pennica D. Interferon-gamma induces receptor di-

merization in solution and on cells. J Biol Chem 1993; 268: 18103-18110.

17) Chen L, Chen J, Mo F, Bian Z, Jin C, Chen X, Liang C. Genetic Polymorphisms of IFNG, IFNGR1, and Androgen Receptor and Chronic Prostatitis/Chronic Pelvic Pain Syndrome in a Chinese Han Population. Dis Markers 2021; 2021: 2898336.

18) Ortiz Fernández L, Coit P, Yilmaz V, Yentür SP, Alibaz-Oner F, Aksu K, Erken E, Düzgün N, Keser G, Cefle A, Yazici A, Ergen A, Alpsoy E, Salvarani C, Casali B, Kısacık B, Kötter I, Henes J, Çınar M, Schaefer A, Nohutcu RM, Zhernakova A, Wijmenga C, Takeuchi F, Harihara S, Kaburaki T, Messedi M, Song YW, Kaşifoğlu T, Carmona FD, Guthridge JM, James JA, Martin J, González Escribano MF, Saruhan-Direskeneli G, Direskeneli H, Sawalha AH. Genetic Association of a Gain-of-Function IFNGR1 Polymorphism and the Intergenic Region LNCAROD/DKK1 With Behçet's Disease. Arthritis Rheumatol 2021; 73: 1244-1252.

19) Wu D, Zhang P, Ma J, Xu J, Yang L, Xu W, Que H, Chen M, Xu H. Serum biomarker panels for the diagnosis of gastric cancer. Cancer Med 2019; 8: 1576-1583.

20) Stein CM, Zalwango S, Chiunda AB, Millard C, Leontiev DV, Horvath AL, Cartier KC, Chervenak K, Boom WH, Elston RC, Mugerwa RD, Whalen CC, Iyengar SK. Linkage and association analysis of candidate genes for TB and TNFalpha cytokine expression: evidence for association with IFNGR1, IL-10, and TNF receptor 1 genes. Human Genetics 2007; 121: 663-673.

21) Wu S, Wang Y, Zhang M, Wang M, He JQ. Genetic variants in IFNG and IFNGR1 and tuberculosis susceptibility. Cytokine 2019; 123: 154775.

22) Kang S. Personalized prediction of drug efficacy for diabetes treatment via patient-level sequential modeling with neural networks. Artif Intell Med 2018; 85: 1-6.

23) Lai NH, Shen WC, Lee CN, Chang JC, Hsu MC, Kuo LN, Yu MC, Chen HY. Comparison of the predictive outcomes for anti-tuberculosis drug-induced hepatotoxicity by different machine learning techniques. Comput Methods Programs Biomed 2020; 188: 105307.

24) Ubaidillah SHSA, Sallehuddin R, Ali NA. Cancer detection using aritifical neural network and support vector machine: A comparative study. Jurnal Teknologi 2013; 65: 73-81.

25) Grant AV, El Baghdadi J, Sabri A, El Azbaoui S, Alaoui-Tahiri K, Rhorfi IA, Gharbaoui Y, Abid A, Benkirane M, Raharimanga V. Age-dependent association between pulmonary tuberculosis and common TOX variants in the 8q12-13 linkage region. Am J Hum Genet 2013; 92: 407-414.

26) Sabri A, Grant AV, Cosker K, El Azbaoui S, Abid A, Abderrahmani Rhorfi I, Souhi H, Janah H, Alaoui-Tahiri K, Gharbaoui Y. Association study of genes controlling IL-12-dependent IFN-γ immunity: STAT4 alleles increase risk of pulmonary tuberculosis in Morocco. J Infect Dis 2014; 210: 611-618.

27) Schoenborn JR, Wilson CB. Regulation of interferon-γ during innate and adaptive immune responses. Adv Immunol 2007; 96: 41-101.

28) Bhanothu V, Lakshmi V, Theophilus JP, Rozati R, Badhini P, Vijayalaxmi B. Investigation of Toll-like receptor-2 (2258G/A) and interferon gamma (+ 874T/A) gene polymorphisms among infertile women with female genital tuberculosis. PLoS One 2015; 10: e0130273.

29) Lee SW, Chuang TY, Huang HH, Lee KF, Chen TTW, Kao YH, Wu LSH. Interferon gamma polymorphisms associated with susceptibility to tuberculosis in a Han Taiwanese population. J Microbiol Immunol Infect 2015; 48: 376-380.

30) Chevillard C, Moukoko CE, Elwali N-EM, Bream JH, Kouriba B, Argiro L, Rahoud S, Mergani A, Henri S, Gaudart J. IFN-γ polymorphisms (IFN-γ+ 2109 and IFN-γ+ 3810) are associated with severe hepatic fibrosis in human hepatic schistosomiasis (Schistosoma mansoni). J Immunol 2003; 171: 5596-5601.

31) Gao J, Wei L, Liu X, Wang L, Niu D, Jin T, Yao G, Wang M, Yu Q, Fu R. Association between IFN-γ gene polymorphisms and IgA nephropathy in a Chinese Han population. Kidney Blood Press Res 2017; 42: 136-144.

32) Gonsky R, Deem RL, Landers CJ, Haritunians T, Yang S, Targan SR. IFNG rs1861494 polymorphism is associated with IBD disease severity and functional changes in both IFNG methylation and protein secretion. Inflamm Bowel Dis 2014; 20: 1794-1801.

33) Kumar A, Ghosh B. A single nucleotide polymorphism (A→G) in intron 3 of IFNγ gene is associated with asthma. Genes & Immunity 2008; 9: 294-301.

34) Rolandelli A, Pellegrini JM, Amiano NO, Santilli MC, Morelli MP, Castello FA, Tateosian NL, Levi A, Casco N, Palmero DJ. The IFNG rs1861494 single nucleotide polymorphism is associated with protection against tuberculosis disease in Argentina. Genes 2018; 9: 46.

35) Altare F, Jouanguy E, Newport M, Lamhamedi S, Fischer A, Levin M, Casanova JL. IFNgR1, a human mycobacterial susceptibility candidate gene. Bull Inst Pasteur 1997; 3: 143-146.

36) He D, Tao S, Guo S, Li M, Wu J, Huang H, Guo X, Yan G, Zhu P, Wang Y. Interaction of TLR-IFN and HLA polymorphisms on susceptibility of chronic HBV infection in Southwest Han Chinese. Liver Int 2015; 35: 1941-1949.

37) Slattery ML, Lundgreen A, Bondurant KL, Wolff RK. Interferon-signaling pathway: associations with colon and rectal cancer risk and subsequent survival. Carcinogenesis 2011; 32: 1660-1667. K, Boom WH, Elston RC, Mugerwa RD, Whalen CC, Iyengar SK. Linkage and association analysis of candidate genes for TB and TNFalpha cytokine expression: evidence for association with IFNGR1, IL-10, and TNF receptor 1 genes. Human Genetics 2007; 121: 663-673.

38) Slattery ML, Lundgreen A, Bondurant KL, Wolff RK. Interferon-signaling pathway: associations with colon and rectal cancer risk and subsequent survival. Carcinogenesis 2011; 32: 1660-1667.

39) Shin JG, Park BL, Kim LH, Namgoong S, Kim JO, Chang HS, Park JS, Jang AS, Park SW, Kim

DJ. Association study of polymorphisms in interferon-γ receptor genes with the risk of pulmonary tuberculosis. Mol Med Rep 2015; 12: 1568-1578.

40) Lü J, Pan H, Chen Y, Tang S, Feng Y, Qiu S, Zhang S, Wu L, Xu R, Peng X. Genetic polymorphisms of IFNG and IFNGR1 in association with the risk of pulmonary tuberculosis. Gene 2014; 543: 140-144.

41) Jüliger S, Bongartz M, Luty AJ, Kremsner PG, Kun JF. Functional analysis of a promoter variant of the gene encoding the interferon-gamma receptor chain I. Immunogenetics 2003; 54: 675-680.

42) Cooke GS, Campbell SJ, Sillah J, Gustafson P, Bah B, Sirugo G, Bennett S, McAdam KP, Sow O, Lienhardt C. Polymorphism within the interferon-γ/receptor complex is associated with pulmonary tuberculosis. Am J Respir Crit Care Med 2006; 174: 339-343.

43) Du F, Xie L, Zhang Y, Gao F, Zhang H, Chen W, Sun B, Sha W, Fang Y, Jia H. Prospective comparison of QFT-GIT and T-SPOT. TB assays for diagnosis of active tuberculosis. Scientific reports 2018; 8: 1-9.

44) Ren Z, Fan Y, Li A, Shen Q, Wu J, Ren L, Lu H, Ding S, Ren H, Liu C. Alterations of the human gut microbiome in chronic kidney disease. Adv Sci 2020; 7: 2001936.

45) Beccaria M, Mellors TR, Petion JS, Rees CA, Nasir M, Systrom HK, Sairistil JW, Jean-Juste MA, Rivera V, Lavoile K. Preliminary investigation of human exhaled breath for tuberculosis diagnosis by multidimensional gas chromatography-Time of flight mass spectrometry and machine learning. J Chromatogr B Biomed Appl 2018; 1074: 46-50.

46) Robison HM, Chapman CA, Zhou H, Erskine CL, Theel E, Peikert T, Lindestam Arlehamn CS, Sette A, Bushell C, Welge M. Risk assessment of latent tuberculosis infection through a multiplexed cytokine biosensor assay and machine learning feature selection. Sci Rep 2021; 11: 1-10.

47) Ren Z, Hu Y, Xu L. Identifying tuberculous pleural effusion using artificial intelligence machine learning algorithms. Respiratory Res 2019; 20: 1-9.