Human Molecular Genetics, 2017, Vol. 26, No. 23 4752–4763

doi: 10.1093/hmg/ddx365
Advance Access Publication Date: 26 September 2017
Association Studies Article

ASSOCIATION STUDIES ARTICLE

Discovery of susceptibility loci associated with tuberculosis in Han Chinese

Hui Qi[1,†], Yong-Biao Zhang[2,†], Lin Sun[1,†], Cheng Chen[3,†], Biao Xu[4,5], Fang Xu[1], Jia-Wen Liu[6], Jin-Cheng Liu[7], Chen Chen[7], Wei-Wei Jiao[1], Chen Shen[1], Jing Xiao[1], Jie-Qiong Li[1], Ya-Jie Guo[1], Yong-Hong Wang[1], Qin-Jing Li[1], Qing-Qin Yin[1], Ying-Jia Li[1], Ting Wang[1], Xing-Yun Wang[1], Ming-Liang Gu[2], Jun Yu[2] and A-Dong Shen[1,*]

[1]Beijing Key Laboratory of Pediatric Respiratory Infection Diseases, Key Laboratory of Major Diseases in Children, Ministry of Education, National Clinical Research Center for Respiratory Diseases, National Key Discipline of Pediatrics (Capital Medical University), Beijing Pediatric Research Institute, Beijing Children's Hospital, Capital Medical University, National Center for Children's Health, Beijing 100045, China, [2]Chinese Academy of Sciences and Key Laboratory of Genome Science and Information, Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing 100101, China, [3]Department of Chronic Communicable Disease, Center for Disease Control and Prevention, Jiangsu 210009, China, [4]School of Public Health, Fudan University, Shanghai 200433, China, [5]Department of Public Health Sciences (Global Health/IHCAR), Karolinska Institute, S-17177 Stockholm, Sweden, [6]Beijing Geriatric Hospital, Beijing 100095, China and [7]Tuberculosis Hospital of Shaanxi Province 710100, Shaanxi Province, China

*To whom correspondence should be addressed. Tel: +86010 59616980; Fax: +86010 59718662; Email: shenad16@hotmail.com

Abstract

Genome-wide association studies (GWASs) have revealed the worldwide heterogeneity of genetic factors in tuberculosis (TB) susceptibility. Despite having the third highest global TB burden, no TB-related GWAS has been performed in China. Here, we performed the first three-stage GWAS on TB in the Han Chinese population. In the stage 1 (discovery stage), after quality control, 691 388 SNPs present in 972 TB patients and 1537 controls were retained. After replication on an additional 3460 TB patients and 4862 controls (stages 2 and 3), we identified three significant loci associated with TB, the most significant of which was rs4240897 (logistic regression $P = 1.41 \times 10^{-11}$, odds ratio = 0.79). The aforementioned three SNPs were harbored by *MFN2*, *RGS12* and human leukocyte antigen class II beta chain paralogue encoding genes, all of which are candidate immune genes associated with TB. Our findings provide new insight into the genetic background of TB in the Han Chinese population.

Introduction

Tuberculosis (TB), a disease caused by infection with *Mycobacterium tuberculosis* (MTB), is a major global health problem and a leading cause of death worldwide. The World Health Organization estimated that there were 10.4 million new cases of TB and 1.4 million deaths from TB in 2015 (1). However, TB

[†]These authors contributed equally to this work.

Received: April 2, 2017. **Revised:** September 12, 2017. **Accepted:** September 19, 2017

patients are far fewer in number than people infected with MTB worldwide, as evidenced by the fact that only 10% of the MTB-infected population develop clinical TB (2). Many factors, including the differential virulence of MTB strains, and host factors such as malnutrition, HIV infection and immunosuppression can all affect the development of TB (3). Case observations, twin studies and mouse infection models have indicated that host genetic factors are important in determining susceptibility to TB (4,5).

The molecular nature of genetic susceptibility to TB has been explored by genome-wide association studies (GWASs) and candidate gene association studies. GWASs on TB have been conducted among African populations including Ghanaians, Gambians, Ugandans and Tanzanians (6–8), among European populations including Russians and Icelanders (9,10), and among Asian populations including Thais, Japanese and Indonesians (11,12). Among African populations, three independent GWASs identified TB-associated single-nucleotide polymorphisms (SNPs) located on chromosomes 18q11, 11p13 and 5q33, respectively (6–8). Among European populations, TB-associated SNPs were detected in the *ASAP1* gene on chromosome 8q24 and in the region harboring genes encoding the class II human leukocyte antigens (HLAs) (9–10). Although statistically significant loci were detected in African and European populations, no locus of genome-wide significance was detected among Asian populations (11,12). This finding may be explained, at least in part, by population specificity and inadequate sample sizes, as only 1102 TB patients in total were involved in the GWAS of the Thai and Japanese populations (11). A high-quality GWAS on TB in the Asian population, which has the highest TB burden in the world, is therefore urgently needed (1).

China has the third highest global TB burden in the world after India and Indonesia (1). A national epidemiological survey on the Chinese population showed that the infection rate of MTB was around 13–20% (13). Here, we report the first GWAS on TB in the Han Chinese population, with the aim of exploring the genetic characteristics of TB. Three significant loci were found to be associated with TB in the studied population.

Results

Loci associated with TB in the Han Chinese population

The GWAS was conducted in three stages: a discovery stage and two replication stages. In stage 1 (discovery stage), we genotyped 900 015 genetic variants in the genomes of 1008 Chinese TB patients and 1538 Chinese controls. After quality control (QC), 691 388 SNPs present in 972 TB patients and 1537 controls were retained. We then performed logistic regression (LR) to estimate the association between the retained SNPs and TB. In stage 2, we genotyped 45 loci from the SNPs whose *P*-values suggested an association with TB in the discovery stage (LR $P < 7 \times 10^{-5}$) from an additional 2304 cases and 2108 controls for replication. After QC, 41 QC-passed SNPs of 2278 cases and 2097 controls remained for analysis. In stage 3, we genotyped the nine most significant loci, which were replicated by stage 2 in 1156 cases and 2754 controls, the SNPs of 1060 cases and 2752 controls remained for analysis.

To avoid spurious associations from population stratification, the LR analyses were performed using the first five eigenvectors from the principal component analysis, as well as age and sex as covariates in stage 1 and the geographic regions (northern and southern China) of the subjects as covariates in the replication stages. The genomic inflation factor (λ_{GC}) was 1.017, indicating that minimal inflation of the genome-wide data from population stratification had occurred.

LR analysis of the combined patient and control data identified three genome-wide significant TB-associated SNPs (the Bonferroni-corrected significance threshold was LR $P < 5 \times 10^{-8}$) (14–16), rs4240897 (LR $P = 1.41 \times 10^{-11}$, odds ratio (OR) = 0.79), rs41553512 (LR $P = 7.93 \times 10^{-11}$, OR = 2.14) and rs2269497 (LR $P = 3.37 \times 10^{-8}$, OR = 1.51). Meta-analysis was performed to determine the heterogeneity of the TB-associated SNPs among the three stages. Significant heterogeneity was found for rs4240897 (1.31×10^{-2}), but no significant heterogeneity was found for rs41553512 (1.03×10^{-1}) and rs2269497 (1.17×10^{-1}) (the heterogeneity significance threshold was set at 0.05) (Fig. 1, Table 1). Regional plots of the above-mentioned three loci showed that these loci (rs4240897, rs2269497 and rs41553512) were harbored by the candidate genes *MFN2*, *RGS12* and HLA class II beta chain paralogue coding genes (Fig. 2). Besides GWAS routine analysis, we used SNP2HLA to perform the imputation on the HLA region and predict classical HLA alleles for all samples of stage 1. SNP2HLA predicted 33 classical HLA alleles and 320 amino acids from the genotype data derived from our chips. We then performed association tests on the classical HLA alleles and amino acids respectively and found that HLA-DQB1 0201 (LR $P = 4.31 \times 10^{-4}$, Bonferroni $P = 2.80 \times 10^{-3}$) was significantly associated with TB (Supplementary Material, Table S1).

Imputation and phenotypic variance estimation

To identify any additional genetic factors associated with TB, we imputed the untyped variants using the genotyped data and the haplotype information from the 1000 Genomes Project (17). After imputation, we obtained 7 333 833 QC-passed variants with minor allele frequencies of >0.01. We then performed single-point LR and joint analysis on the QC-passed SNPs with the first five eigenvectors from the principal component analysis, as well as age and sex as covariates. We found that 1054 loci had GWAS *P*-values $< 1 \times 10^{-4}$ in 120 continuous regions (the distance between two adjacent SNPs being less than 10 kb), thereby providing additional information about the genetic characteristics of TB in the Han Chinese population (Supplementary Material, Data S1).

SNPs significantly associated with TB explained 2.33% of the phenotypic variance based on the condition that the prevalence rate of the bacteriologically-confirmed TB cases was 11.9 per 10 000 in China (Table 1) (1). The joint effect of all the 691 388 genotyped SNPs and the imputed 7 333 833 variants explained 23% and 38% of the phenotypic variance observed in this study, respectively, and the heritability of TB can reach 0.38 in Chinese.

Overlap of TB- and leprosy-associated loci

Mycobacterium leprae, the agent responsible for leprosy, also causes serious mycobacterial infections on a worldwide scale, and induces a similar immune response to MTB (18,19). We looked for overlapping loci between our GWAS TB data and the published genetic associations for TB (6–12) and leprosy (19–25). Ninety-five previously reported TB- and leprosy-associated loci were found in our imputation data and five of these loci were nominally associated with Chinese TB (LR $P < 0.05$) (Supplementary Material, Data S2). Of these loci, rs955263 and rs4236914 were associated with TB in HIV-positive Ugandans and Tanzanians (8), and rs9270650, rs4574921, and rs10817758

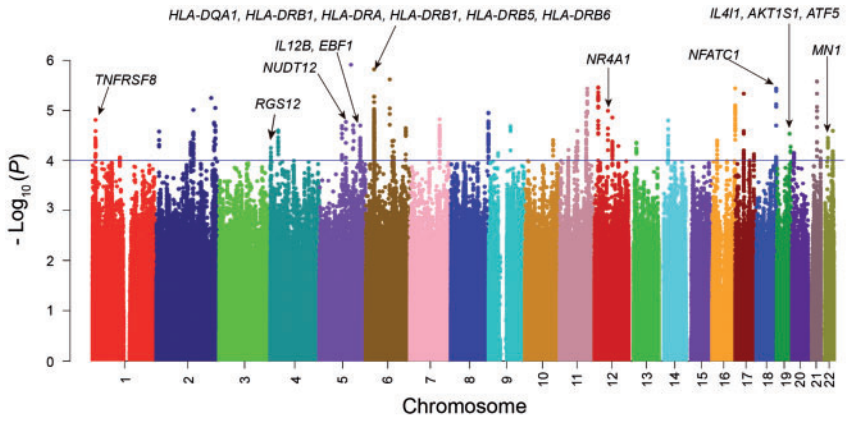

Figure 1. Distribution of candidate genes harboring SNPs that were selected after replication on the human genome. Manhattan plot of the LR *P*-values calculated from the genome-wide association study for the discovery stage. Data on 691, 388 SNPs that passed quality control steps were collected from 972 cases with TB and 1537 controls. The $-\log_{10}$ (LR *P*-value) of each SNP is shown as a function of its genomic position on autosomes (hg19). Each dot represents a SNP; and different colors are used for distinguishing chromosomes from 1 to 22. The solid blue line denotes a significance level of LR $P = 1 \times 10^{-4}$. Candidate genes harboring SNPs that were selected after replication are indicated.

were associated with leprosy in Indians and Chinese (19–21). Despite the LR *P*-values for rs955263, rs4236914 and rs4574921 being less than 0.05 in our research, the published risk alleles for these loci (rs955263-C, rs4236914-C and rs4574921-A) were not consistent with the risk alleles identified in our study (rs955263-T, rs4236914-T and rs4574921-G).

Meta-analysis

Using the imputed data from the Gambian (whole genome) and Indonesian (200-kb genomic regions of leading SNPs) populations, we performed genome-wide meta-analysis on the Chinese and Gambian populations and meta-analysis on all three populations (Chinese, Gambians and Indonesians) for the leading SNPs.

Initially, we carried out an imputation-based meta-analysis of TB for the Chinese and Gambian populations (6). Two hundred and twenty-two loci located in 56 continuous regions that possessed meta *P*-values of $<1 \times 10^{-4}$ and *P*-het of >0.05 were found in both populations (Chinese and Gambians) (Supplementary Material, Data S3, Fig. S1). As shown in Supplementary Material, Figure S1, in the 222 TB-associated loci, 92 loci were harbored by immune genes such as *SKAP2*, *LCP1*, *TAF4B* and the genes encoding the HLA class II region. In addition, we conducted another imputation-based meta-analysis for our lead SNPs of TB in the Chinese, Gambian (6) and Indonesian (12) populations. These three SNPs gave meta *P*-values in at least two of the populations. Among the three SNPs, rs4240897 and rs41553512 became more significant in meta-analysis than in the Chinese population alone. Significant heterogeneity was found for rs4240897 ($P\text{-het} = 2.36 \times 10^{-2}$) and rs2269497 ($P\text{-het} = 1.43 \times 10^{-3}$), but no significant heterogeneity was found for rs41553512 and rs2269497 ($P\text{-het} = 2.08 \times 10^{-1}$) (Supplementary Material, Data S4).

Functional annotation

Functional noncoding variants within gene regulatory elements may play roles in disease phenotypes by modulating gene expression levels. To predict the effects of variants on gene expression, we analysed all variants (284) with *P*-values of $<1 \times 10^{-4}$ within the flanking regions of the three significant loci using the software SeattleSeq (v138) (26) and HaploReg (v2) (27).

Of the SNP variants annotated using SeattleSeq, four were located in known transcription factor binding sites: one, a missense mutation (rs41542812) in *HLA-DQB1*, was classified as benign (PolyPhen score <0.15; Supplementary Material, Table S2), rs41553512 and rs1136744 were missense mutations in *HLA-DRB5*, and rs41552812 was a missense mutation in *HLA-DQB1*. These latter three SNPs were classified as probably damaging (PolyPhen score >0.85) (Supplementary Material, Table S2).

More than half of the SNPs were located in the gene expression regulatory motifs of enhancers and promoters, as determined using HaploReg, while 98 of the SNPs had a recorded effect on gene expression in the Genotype-Tissue Expression (GTEx) pilot (28) and the Genome-Wide Repository of Associations Between SNPs and Phenotypes databases (29) (Supplementary Material, Data S5). For the cell type-specific enhancer enrichment analyses, we conducted queries in HaploReg with the 284 SNPs and their linked SNPs ($r^2 = 1$), based on the epigenomic data from the Roadmap Epigenomics (30) and ENCyclopedia of DNA Elements (ENCODE) projects (31), and found that immune-related cell lines were enriched significantly, including T helper cells, monocytes and B cells (Chi-square test $P < 2 \times 10^{-6}$) (Supplementary Material, Table S3).

Chip-based gene expression profiling, RT-PCR verification and eQTL analysis

To investigate differences in the gene expression profiles between the TB and control groups, we used a human gene expression array on 15 additional peripheral blood mononuclear cell (PBMC) cases and 14 controls. We focused on 26 genes within the 400-kb flanking regions of the three significant SNPs (Table 1). The expression levels of the 26 genes are shown in Figure 3 and Supplementary Material, Data S6. Genes with significantly different expression levels were filtered according to a fold change (FC) (linear) of ≤ 0.67 (downregulated) or ≥ 1.5 (upregulated), and a Student's *t*-test *P*-value of <0.05. The expression level of *CLCN6* ($P = 5.28 \times 10^{-8}$, FC = 1.66) was significantly higher in the TB group than in the control group. In addition, the expression level of *MFN2* ($P = 2.81 \times 10^{-2}$, FC = 1.43) was nominally higher in the TB group than in the control group. The expression levels of *DOK7* ($P = 7.1 \times 10^{-4}$, FC = 0.54), *KIAA2013* ($P = 9.05 \times 10^{-7}$, FC = 0.47) and

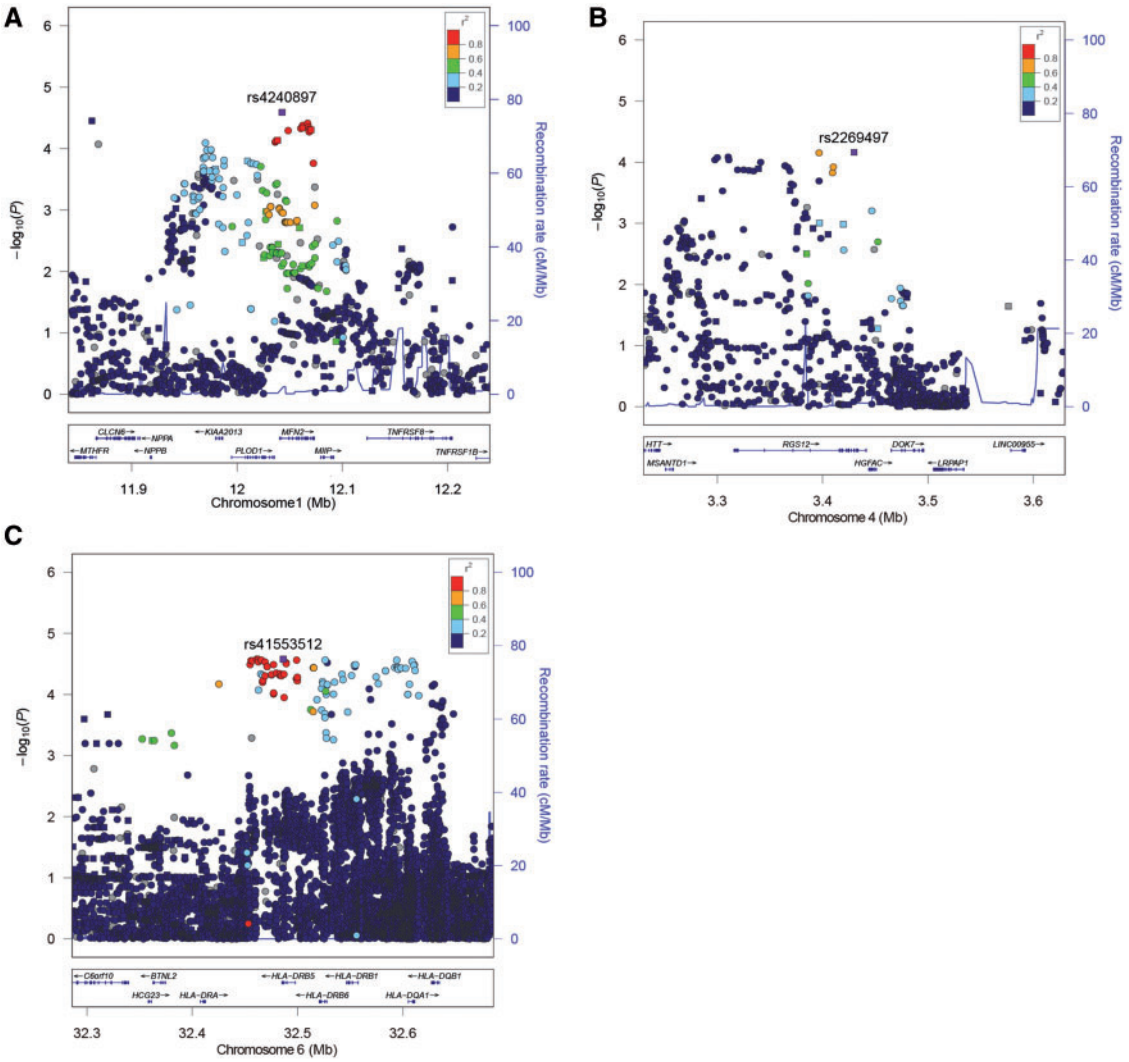

Figure 2. Regional distribution of three loci significantly associated with TB. Regional plots showing the association of TB risk with four loci, rs4240897 (**A**), rs41553512 (**B**) and rs2269497 (**C**), harbored by candidate TB-related immune genes. Each point represents a SNP plotted with its $-\log_{10}$P-value as a function of genomic position (hg19). Imputation analysis is shown by circles and direct genotyping by squares. In each regional plot, the purple symbol denotes the lead SNP, and its name is shown at the top of each plot. The color coding of the rest of the SNPs shows their LD with the lead SNP: red, $r^2 \geq 0.8$; gold, $0.6 \leq r^2 < 0.8$; green, $0.4 \leq r^2 < 0.6$; cyan, $0.2 \leq r^2 < 0.4$; blue, $r^2 < 0.2$; gray, r^2 unknown. Recombination rates are estimated from the Asian (ASN) populations of the 1000 Genomes Project (Nov 2014). Gene annotations are taken from the UCSC genome browser and some genes are not shown.

TNFRSF8 ($P = 8.39 \times 10^{-3}$, FC =0.63) were significantly lower in the TB group than in the control group. To verify the expression difference in the aforementioned five candidate genes, we conducted RT-PCR on the newly enrolled 22 PBMC case samples and 14 control samples. In RT-PCR verification, the expression levels of *CLCN6* ($P = 4.67 \times 10^{-3}$) and *MFN2* ($P = 1.12 \times 10^{-3}$) were significantly higher in the TB group than in the control group, and the expression levels of *KIAA2013* ($P = 2.39 \times 10^{-2}$) and *TNFRSF8* ($P = 3.84 \times 10^{-2}$) were significantly lower in the TB group than in the control group (Supplementary Material, Fig. S2). Overall, four candidate genes within the 400-kb flanking regions of the three significant SNPs had significantly different expression levels between the two groups.

To explore the eQTL of the SNPs, we queried all genes within the flanking regions of the three significant loci to the GTEx portal. The expression of the *MTHFR*, *KIAA2013*, *MFN2*, *RGS12*, *HLA-DRA*, *HLA-DQA1* and *HLA-DRB6* genes were significantly influenced by the 29 SNPs (GTEx *P*-value $< 1 \times 10^{-4}$). The most significant SNP detected in our study, rs4240897, was found to significantly influence the expression of *MFN2* (4.67×10^{-27}) (Supplementary Material, Table S4).

Pathway enrichment analysis

For the gene annotation enrichment analysis and functional annotation clustering, we analysed the aforementioned 26 genes within the 400-kb flanking regions of the three significant SNPs using the Database for Annotation, Visualization and Integrated Discovery (DAVID) v6.7 (32) (Supplementary Material, Table S5). Six DAVID-defined clusters showed significant enrichment scores (Fisher's exact test $P < 0.05$), including antigen processing and presentation, co-stimulation and activation of T cells, cell-cell adhesion, cellular response to stimulus, signal transduction, regulation of homeostatic process, and immunoglobulin production. Candidate genes in the HLA regions were enriched

Table 1. TB-associated SNPs identified from 4310 cases and 6386 controls from China

Lead SNPs	Position	Alleles	Stage 1 (972 cases vs 1537 controls)			Stage 2 (2278 cases vs 2097 controls)			Stage 3 (1060 cases vs 2752 controls)			Meta analysis on three stages (4310 cases vs 6386 controls)			P-het	Varexpl (%)	Candidate genes
			MAF (case/con)	OR (95% CI)	P	MAF (case/con)	OR (95% CI)	P	MAF (case/con)	OR (95% CI)	P	MAF (case/con)	OR (95% CI)	P			
rs4240897	Chr1: 12042755	A/G	0.438/0.502	0.77 (0.69–0.87)	2.57E-05	0.424/0.500	0.76 (0.70–0.83)	9.38E-09	0.490/0.5152	0.91 (0.82–1.01)	5.40E-02	0.447/0.507	0.79 (0.75–0.83)	1.41E-11	0.0131	0.9	*MTHFR*, *CLCN6*, *NPPA*, *NPPB*, *KIAA2013*, *PLOD1*, *MFN2*, *MIIP*, *TNFRSF8*, and *TNFRSF1B*
rs41553512	Chr6: 32486402	A/G	0.040/0.020	2.06 (1.46–2.89)	2.64E-05	0.036/0.017	2.23 (1.68–2.96)	1.40E-06	0.024/0.015	1.55 (1.08–2.21)	1.56E-02	0.033/0.016	2.14 (1.78–2.57)	7.93E-11	0.1034	0.85	*C6OF10*, *BTNL2*, *HCG23*, *HLA-DRA*, *HLA-DRB5*, *HLA-DRB1*, *HLA-DRB6*, *HLA-DQB1*, and *HLA-DQA1*
rs2269497	Chr4: 3429856	G/A	0.100/0.064	1.63 (1.32–2.00)	6.88E-05	0.084/0.058	1.47 (1.24–1.73)	1.02E-04	0.061/0.050	1.25 (1.00–1.55)	5.03E-02	0.082/0.056	1.51 (1.35–1.68)	3.37E-08	0.1174	0.58	*HTT*, *MSANTD1*, *RGS12*, *HGF*, *ACDOK7*, *LRPAP1*, and *LINC00955*

SNP, single-nucleotide polymorphism; Position, Physical position on chromsome (hg19), MAF, OR, Odds ratio for the minor allele; CI, confidence interval; *P*-values, *P*-values calculated by logistic regression; Combined, 4310 cases and 6386 controls; Stage 1, 972 cases and 1537 controls; Stage 2, 2278 cases and 2097 controls; Stage 3, 1060 cases vs 2752 controls; Het *P*, *P* value from the heterogenetity test based on GWAS (genome-wide association study) and replication study; Varexpl, variance in liability to TB explained by the locus at the prevalence rate of 11.9/10 000 in China (1).

in all of the six clusters. Additionally, *MFN2* and *TNFRSF8* were enriched in the cellular response to stimulus and signal transduction. Therefore, genes within the 400-kb flanking regions of the three significant SNPs may participate in the pathogenesis of TB mainly through activation of immune cells and signal transduction.

Discussion

Genetic factors play an important role in the outcome of infection with MTB. Previous GWAS have investigated genetic susceptibility to TB in African (6–8), European (9,10) and Asian populations (11,12). Although one previous candidate association study revealed one suggestive associated locus (rs2057178) with Chinese TB (33), GWAS is still needed to investigate the genetic characteristics of TB in the Chinese population. In our study, we found that the candidate genes harboring the three identified significant loci were associated with the etiology of TB. First, TB-associated SNPs were located within the flanking regions of the three significant loci enriched in the enhancers of immune-related cell lines, which may be involved in the immune response to TB. Second, gene expression profiling revealed that many of the candidate genes, such as *TNFRSF8* and *MFN2*, were differently expressed in the TB and control groups. Third, gene-annotation enrichment analyses and functional annotation clustering of the 26 genes within the 400-kb flanking regions of three significant SNPs indicated that the genes were enriched in the activation of immune cells and signal transduction. Thus, our data indicate that an association exists between the aforementioned candidate genes and TB at different levels.

The candidate genes harboring SNPs that were significantly associated with TB were *MFN2*, *RGS12* and HLA class II beta chain paralogue encoding genes (*HLA-DQA1*, *HLA-DQB1*, *HLA-DRA*, *HLA-DRB1*, *HLA-DRB5* and *HLA-DRB6*). These genes participate in many steps of the immune response to TB as discussed below.

The MFN2 protein is a key factor in mitochondrial fusion and mitochondrial metabolism. In a meta-analysis of GWASs, rs2336384 of *MFN2* was identified to be significantly associated with the platelet count and mean platelet volume in European populations (34). Additionally, researchers found that platelets drove macrophage differentiation into epithelioid-like multinucleated giant foam cells in tuberculosis granulomas (35). In our study, rs4240897, an intron mutation in *MFN2*, was found to significantly influence the expression of *MFN2* by GTEx. Moreover, the expression level of *MFN2* was significantly higher in the TB group than in the control group. Thus, it is possible that rs4240897 may affect platelet count and macrophage differentiation by moderating the expression and function of *MFN2*.

In the 80 kD downstream of rs4240897, *TNFRSF8* is another important immune gene that has been reported to be involved in the immune response to TB. TNFSF8/TNFRSF8 signaling plays an augmentative role in the production of IFN-γ in Th1 cells in response to *Mycobacterium bovis* bacillus Calmette–Guerin (BCG) infection (36). During genetic research into leprosy, 18 cis-expression gene quantitative trait loci of *TNFSF8* were found to be associated with Type 1 reactions (T1R) in Vietnamese and Brazilian populations, which are pathological inflammatory responses that are the main cause of nerve damage for leprosy patients (37). In addition, the association of rs6478108, one of 18 loci, with T1R was more pronounced in younger leprosy patients (<30 years old) (38). Consequently, variants of TNFSF8/TNFRSF8 may be important for susceptibility to mycobacterial diseases. In our study, as rs4240897 locates near *TNFRSF8* and

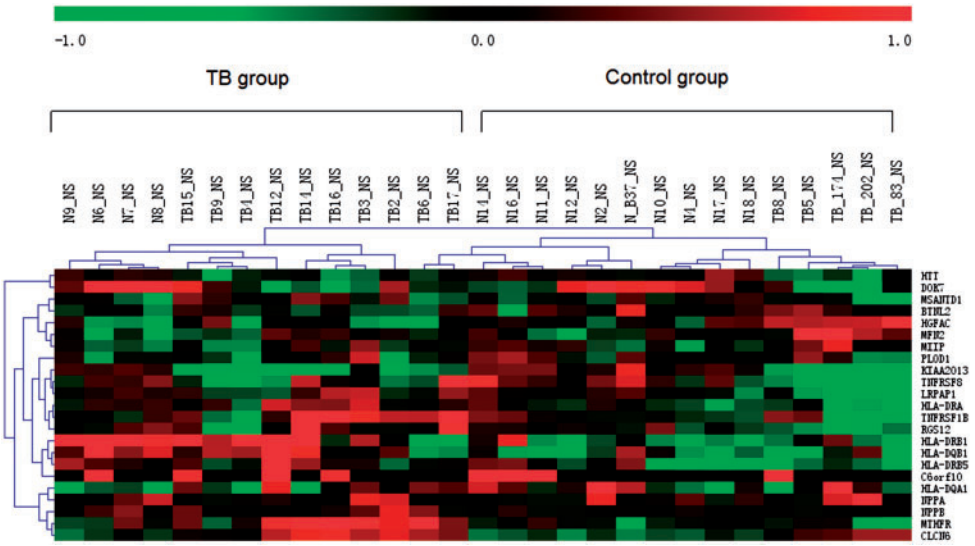

Figure 3. Differences in the expression of candidate genes within the 400-kb flanking regions of the three significant SNPs in TB patients and controls. The heatmap represents the results of a two-way hierarchical clustering of mRNA and samples. Each row represents an mRNA, and each column represents the sample tested. Red represents mRNAs with an expression level above the mean, and green represents mRNAs with an expression level below the mean.

the expression level of *TNFRSF8* is significantly lower in the TB group than in the control group, we cannot rule out the possibility that rs4240897 may affect the function or expression of TNFRSF8 in TB.

HLA-restricted $CD4^+$ and $CD8^+$ T cells recognize many MTB-derived epitopes in MTB-infected populations (39). HLA-DQA1, HLA-DQB1, HLA-DRA, HLA-DRB1, HLA-DRB5 and HLA-DRB6 are HLA class II beta chain paralogues (10). Polymorphisms in *HLA-DRB1* and *HLA-DQB1* are associated with TB, not only in Europeans, but also in Africans and South Americans (10,40,41). In our study, rs41553512, a missense mutation in *HLA-DRB5*, was significantly associated with TB and was classified as damaging. Thus, rs41553512 probably has an effect on antigen presentation by changing the amino acid in HLA-DRB5.

RGS12 can enhance the deactivation of G-protein α subunits to reduce the activation of downstream effectors of G-protein-coupled receptors (GPCRs) (42). Chemokine receptors are involved in the migration and function of immune cells and are the most important GPCRs in these cells (42). The role played by rs2269497 and *RGS12* in TB is not clear, but cell migration and chemokine receptor signaling is an area worthy of future research.

To explore the GWAS hotspots for mycobacterial diseases (as represented by TB and leprosy), we compared our GWAS TB data with the published data of TB (6–12) and leprosy (19–25), and found five previously reported loci (LR $P < 0.05$) were nominally associated with Chinese TB. These loci may reflect a predisposition to infectious diseases caused by mycobacterial infection. However, the published risk alleles for rs955263 (8), rs4236914 (8) and rs4574921 (21) are not consistent with the risk alleles discovered in our study. This inconsistency may result from population specificity, the strong selective pressure exerted by MTB (19–21), and/or different features of TB and leprosy (43). Although TB and leprosy are both caused by mycobacteria, many differences exist between the two diseases (44). Both TB and leprosy are two ancient diseases, which have been identified as infecting humans 9000 and 4000 years ago, respectively (45). As effective treatments to these two diseases were only established in the recent decades, TB and leprosy have been responsible for a large number of deaths throughout history. In the past, untreated TB was associated with a high mortality rate, whereas untreated leprosy had a low mortality rate but a high disability rate due to peripheral neuropathy (44). As a result, individuals who had genetic mutations associated with TB susceptibility were more likely to die from TB, whereas individuals with leprosy were more likely to survive and have the opportunity to propagate their mutations. Overall, a considerable difference exists in the selection pressures of these two diseases and this difference may contribute to the inconsistent genetic findings between TB and leprosy.

Our imputation-based meta-analysis of TB in Chinese and Gambian populations identified the candidate immune genes *HLA-DRA*, *HLA-DRB1*, *HLA-DRB5*, *HLA-DRB6*, *HLA-DQA1*, *HLA-DQB1*, *HLA-DQA2*, *HLA-DQB2*, *SKAP2*, *LCP1* and *TAF4B*. These genes are associated with TB progression, possibly by affecting antigen presentation (10), macrophage infiltration (46), cell migration (47), or cell maturation (48). However, we did not find any SNPs of genome-wide significance in the aforementioned meta-analysis and more work is needed to explore the functions of SNPs and candidate genes. In addition, we conducted another imputation-based meta-analysis for our three significant SNPs of TB in Chinese, Gambian (6) and Indonesian (12) populations. Our results indicated that rs4240897 and rs41553512 became more significant in meta-analysis with the enrolled Indonesian population. However, rs2269497 became less significant in meta-analysis with the enrolled Gambian population. We propose that the differences in population specificity and environmental risk factors among Gambian and Asian populations may contribute to the unsatisfactory meta results for rs2269497. Taken together, the meta-analysis data derived from different populations (particularly Asian populations) provides further support for the association between TB and the significant SNPs found in our GWAS.

Our research provides novel insights into the genetic characteristics of TB in the Han Chinese population. However, our study had some limitations and our results should be verified by further research. First, the geographical distribution of cases and controls was unmatched in stage 3, which may partly

contribute to the unsatisfactory results of stage 3. In addition, about 70% of samples for stage 3 originated from southern China, whereas most of the samples (>80%) for stages 1 and 2 were from northern China. The significant heterogeneity of rs4240897 was likely due to the fact that samples of stage 3 comprised different Chinese ethnic background in comparison with samples of stage 1 and 2. The nominal association of stage 3 replication suggested that more samples should be investigated to depict the possible differences in pathogenesis between TB patients from northern and southern China. Second, in addition to the Gambian and Indonesian populations, meta-analysis of worldwide TB data is needed to validate the associated loci. Third, a larger independent sample of the Han Chinese population should be employed to replicate the present GWAS. Fourthly, as a genetic study of infectious disease, research into human genomics and pathogen genomics should be conducted concurrently to analyse the interaction between pathogen and host. Despite these limitations, our findings not only present new risk loci for TB, but also reveal the complexity and specificity of the genetic characteristics of TB susceptibility.

Materials and Methods

Samples

All the individuals in our study were Han Chinese. The diagnosis of TB cases was consistent with previously published criteria (49–51).The diagnosis of TB is based on a the following factors: (i) etiology or pathology results (Acid-Fast Bacilli Stain or culture); (ii) clinical presentation (symptoms or signs); (iii) imaging (chest radiography or computed tomography scan); (iv) contact history (family and close contact); (v) purified protein derivative (PPD) skin tests or interferon gamma release assay (IGRA) positive results; (vi) positive clinical response to anti-TB therapy; (vii) except other diseases, such as the pneumonia, tumor, inflammatory diseases and soon. Clinical TB could be diagnosed if positive features of (ii)–(iii) plus either two of (iv)–(vii) were present. Bacteriologically confirmed TB could be diagnosed if positive features of (i) plus (ii) and/or (iii).

Pulmonary TB: Patients with exclusively intrathoracic involvement (i.e. confined to the lung parenchyma, pleura and intrathoracic lymph nodes) were considered to have pulmonary TB (50,52). Extrapulmonary TB: Patients with pulmonary involvement who also had extension of the disease to organs or tissues outside the thorax were classified as having extrapulmonary TB (50,52).

Controls were defined as individuals with no history of TB, normal radiographic findings, and negative purified protein derivative (PPD) skin tests (<5 mm). In addition, as described previously, the controls had no infectious diseases or history of any inflammatory or autoimmune disease (53).

In the stage 1, we enrolled 1008 TB patients (645 males and 363 females with a mean age of 22.5 ± 17.2 years) from the Beijing Children's Hospital, the Beijing Geriatric Hospital and the Tuberculosis Hospital in Shaanxi Province as a case cohort for this study. Geographic location records were available for 992 cases, 86.9% of which were from northern China (according to a previously described population structure boundary (54)). The control group contained 1538 individuals 1024 of which were males and 514 females. The mean age of controls was 24.5 ± 18.6 years. The Controls were enrolled at physical examination centers located in northern and southern China, 84.2% being from northern China. The population structure of case and control groups was not significantly different (two-tailed Chi-square test $P = 0.058$). After QC, data of 972 cases and 1537 controls were remained; demographic information and clinical characteristics of these individuals were shown in Supplementary Material, Table S6.

In stage 2, we enrolled 2304 TB patients (1370 males and 934 females, mean age 31.6 ± 16.0 years) and 2108 controls (1388 males and 720 females, mean age 39.5 ± 18.2 years) from the same hospitals and physical examination centers used in the stage 1. After QC, data of 2278 cases and 2097 controls were remained; demographic information and clinical characteristics of these individuals were also shown in Supplementary Material, Table S6.

In stage 3, we enrolled 1156 TB patients (726 males and 430 females, mean age 46.1 ± 23.2 years) and 2754 controls (1601 males and 1153 females, mean age 42.3 ± 22.5 years). Among them, 764 cases and 764 controls were from Center for Disease Control and Prevention of Jiangsu Province (33), and the remaining samples were collected from the same hospitals and physical examination centers as stage 1 and stage 2. After genotyping 9 loci from stage 2 and QC, data of 1060 cases and 2752 controls were remained; demographic information and clinical characteristics of these individuals were summarized in Supplementary Material, Table S6.

All participants provided informed consent. This project was reviewed and approved by the Ethics Committees of the Beijing Children's Hospital, the Beijing Geriatric Hospital, the Tuberculosis Hospital in Shaanxi Province, the Beijing Institute of Genomics, Chinese Academy of Sciences and the Center for Disease Control and Prevention of Jiangsu Province.

Genetic power calculation

Sample size power was evaluated using CaTS software (55). In multiplicative model mode, we set the parameters as follows: case number: 972, control number: 1537, and prevalence of bacteriologically confirmed TB cases was 0.00119 (1).We next evaluated the power to obtain 0.05, 1×10^{-4}, and 5×10^{-8} significance levels at disease allele frequencies (DAFs) of 0.05, 0.10 and 0.15 (Supplementary Material, Fig. S3). Although the power is limited with the current sample size, there is still an 80% chance of obtaining genome-wide significant SNPs ($P = 5 \times 10^{-8}$) with a genetic relative risk (GRR) = 2.0 and DAF = 0.05, or GRR = 1.8 and DAF = 0.1, or GRR = 1.5 and DAF = 0.15.

Genotyping and quality control

We collected blood samples from the TB and control groups. Samples were stored at −80 °C. Genomic DNA was extracted from 400 μl EDTA anti-coagulated whole blood samples using QIAamp DNA Blood Midi Kits (Qiagen, Duesseldorf, Germany). The DNA sample concentration was 50 ng/μlin the discovery stage and 10 – 30 ng/μl in the replication stage. DNA samples were quantified using a NanoDrop ND-1000 spectrophotometer (Thermo, MA, USA) and agarose gel electrophoresis. In the discovery stage, 1008 cases and 1538 control samples were genotyped by Bio Miao Biological Technology (Beijing, China) using the HumanOmniZhongHua-8 v1.1 BeadChip (Illumina, CA, USA) in accordance with the manufacturer's specifications. The investigators were blinded to the group allocation of chips during the genotyping. Genotypes of 900, 015 SNPs were analysed using the genotyping module of GenomeStudio v3.0 (Illumina). In total, 2526 DNA samples were successfully genotyped at a call rate >99.8%; the genotype call threshold (boundary for calling a genotype relative to its associated cluster) was 0.15. The genotype reoccurrence rate for 12 duplicated individuals was 99.99% on average.

To obtain high-quality data for the GWAS, we pruned the discovery stage data set using the following criteria: sample call rate >99%; SNP call rate >95%; and a threshold for Hardy–Weinberg equilibrium (HWE) P value of 0.001 in the control cohort. We also calculated genome-wide identity by descent (IBD) for each pair of samples so that closely related individuals could be excluded. We found one pair in the case cohort and one pair in the control cohort with an IBD >0.05, and randomly removed one individual from each pair from the cohorts. We retained SNPs with minor allele frequencies >0.01 due to the limited power for rare variants in this association study. In addition, we extracted genotype data for the YRI, CEU, JPT, CHB and CHS populations from the 1000 Genomes Project (17) and performed principal components analysis on these samples along with our genotyped samples using the smart PCA package (56). We clustered together the Asian populations (i.e. CHB, CHS, JPT and our samples) and found that the Chinese samples were well separated from the Japanese samples (Supplementary Material, Fig. S4).Three outliers within the TB cohort (based on genome-wide identity-by-state) were removed from subsequent analyses. The final data set used for GWAS analyses included 972 cases and 1537 controls, with 691 388 SNPs. In addition, we have submitted our microarray data to the GEO (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi? acc=GSE83397).

Genotype imputing

We used PLINK software to flip SNPs from the reverse strand to the forward strand and pre-phased the haplotypes in each chromosome using the SHAPEIT algorithm (57). Untyped SNPs of Gambian, Indonesian (200-kb flanking regions of the lead SNPs) and our GWAS data, were imputed based on the 1000 Genomes Project phase I integrated variant set (b37; October 2014) using IMPUTE2 (58). A strict cutoff (info >0.85, Fisher's exact test $P > 0.001$ for HWE) was set for post-imputation SNP filtering to remove poorly imputed SNPs.

Selection of genomic loci for replication

To select genomic loci for replication, the genomic distribution of genotyped SNPs that were suggestively associated with TB (LR $P < 7 \times 10^{-5}$) was examined. As SNPs within the same contiguous genomic region tend to form 'association peaks' due to strong linkage disequilibrium (LD), the most significant SNP in each region was selected to represent the rest of the selected SNPs. In addition, we checked whether other 'isolated TB-associated SNPs' scattered on the chromosome were in strong LD ($r^2 > 0.8$) with surrounding SNPs (400-Kb window) in the Han Chinese from the 1000 Genomes Project. If there were no other genotyped SNPs in strong LD with the 'isolated TB-associated SNP', the SNP was retained as a potentially true TB-associated SNP. In the end, 45 SNPs were selected in this manner for the replication study.

In stage 2, these 45SNPs were then genotyped in the additional 2304 cases and 2108 controls in the replication set by Bio Miao Biological Technology (Beijing, China) using a MassARRAY system (Sequenom, CA, USA). The investigators were blinded to the group allocation during the genotyping in MassARRAY. Twenty-six case and 11 control samples that had more than 5% missing genotypes were removed from the data analysis. Of the 45 SNPs, 41 had less than 5% missing genotypes and showed no deviation from HWE (Fisher's exact test $P > 0.001$) in the control samples.

In stage 3, we genotyped nine most significant loci which were replicated by stage 2 in 1156 cases and 2754 controls with the same method and platform of stage 2. Ninety-six case and two control samples that had more than 5% missing genotypes were removed from the data analysis. All of these nine SNPs had less than 5% missing genotypes and showed no deviation from HWE (Fisher's exact test $P > 0.001$) in the control samples.

Association testing

The associations between SNP genotypes and the TB trait were estimated by applying a LR algorithm in PLINK (v1.9) (57). To handle the population stratification of the samples, LR was performed on all the SNPs with covariates of the first five principal components from the principal components analysis. A quantile-quantile (Q-Q) plot of this test is shown in Supplementary Material, Fig. S5, where λ_{GC} was 1.017 (based on median Chi-square). We constructed Manhattan plots using qqman (59). Bonferroni correction was used for multiple comparisons, and the threshold for genome-wide significance was set at LR P-value $< 5 \times 10^{-8}$ (14–16). We visualized regional association and linkage disequilibrium using LocusZoom (60). For the replication study, to avoid spurious associations caused by population stratification, we categorized the samples into northern and southern Chinese groups and included groups as a covariate in the LR association test. Combined analyses on the discovery and replication data were carried out using METAL (61) with the following parameters: EFFECT, Beta; Weights in P-value Based Analysis, sample size; and heterogeneity, Cochran's Q-test.

Association analysis between HLA alleles and TB

We used SNP2HLA to impute the genotype of untyped variant surrounding leading SNP within HLA (62). We then predict the classic HLA alleles, as well as amino acid sequences of classical allele and performed logistic regression association test on them.

Joint multiple-SNP analysis

To determine the interactions of SNPs within a gene or a defined haplotype block, we performed joint analyses on gene or block sets composed of QC-passed SNPs. We defined two sets, a 17 856 gene set harboring 311 130 SNPs and a 120 399 block set harboring 595 643 SNPs. To estimate joint effects, we used the SKAT package (63) and implemented multiple LR with the first five eigenvectors from principal components analysis as covariates and with PolyPhen scores (64) as the weight of each SNP. The thresholds for adjustment of multiple tests were set as 2.8×10^{-6} (0.05/17 856 sets) and 4.15×10^{-7} (0.05/120 399 sets) for the gene set and haplotype set regressions, respectively.

Meta-analysis

We just obtain the imputed data of Gambians (whole genome) and Indonesians (200-kb genomic regions of leading SNPs). With obtained data, we performed genome-wide meta-analysis on Chinese and Gambians and meta-analysis on the three populations (Chinese, Gambians and Indonesians) for the leading SNPs.

Initially, we carried out an imputation-based meta-analysis of TB in Chinese and Gambian populations (6) using METAL (61) with the following parameters: EFFECT, Beta; Weights in P-value

Based Analysis, sample size; and heterogeneity, Cochran's Q-test. Variants with meta *P*-values $<1\times10^{-4}$and *P*-het >0.05 in meta-analysis were given in Supplementary Material, Data S3. In addition, we conducted another imputation-based meta-analysis for our nine lead SNPs of TB in Chinese, Gambians (6) and Indonesians (12) in the same way. The heterogeneity significance threshold was set at 0.05. The meta-analysis results of three significant SNPs were given in Supplementary Material, Data S4.

Functional annotation

TB-associated variants (GWAS $P<1\times10^{-4}$ of typed and imputed variants) in the flanking regions of the 9 associated loci were annotated using SeattleSeq (v138, http://snp.gs.washington.edu/SeattleSeqAnnotation138/) (26) and HaploReg (v2, http://www.broadinstitute.org/mammals/haploreg/haploreg_v2.php) (27). For SeattleSeq, variants that might have functional effects (Supplementary Material, Table S2) were retained. For HaploReg, the LD calculation was based on the Asian (ASN) populations from the 1000 Genomes Project (phase I), and the LD threshold (r^2) was set at 1.0. We performed enrichment analyses on cell type-specific enhancers based on the ENCODE database with the 1000 Genomes Project ASN data as the background set (Supplementary Material, Table S3). The Chi-square test was applied to calculate the significance level of enhancer-enriched immune-related cells.

Estimation of variance in TB susceptibility explained by associated SNPs

The GCTA package (65) was used to estimate the variance in TB susceptibility that could be explained by either the associated SNPs or all genotyped SNPs, with the prevalence of bacteriologically confirmed TB cases in China was 11.9 per 10 000 (1). For each associated locus, an SNP set composed of SNPs with *P*-values <0.05 flanking the 400-kb region of the lead SNP was used to estimate the phenotypic variance that could be explained.

RNA extraction and purification

We enrolled 37 TB blood samples and 28 control blood samples, and used Ficoll-Hypaque Solution (Hao Yang, Shanghai, China) to distinguish PBMCs from granulocytes.15 TB PBMCs and 14 control PBMCs were used in gene expression chips, while the other 22 TB PBMCs and 14 control PBMCs were used in RT-PCR verification. The criteria used for inclusion of the PBMC samples from TB patients were as follows: (i) the patients were the initial untreated cases and, (ii) the samples were collected on the day of diagnosis. Total RNA was extracted and purified using a miRNeasy Mini Kit (Cat#217004, Qiagen) following the manufacturer's instructions. The RNA integrity number was checked using an Agilent Bioanalyzer 2100 (Agilent Technologies, CA, US).

RNA profiling and data analysis

Gene expression profiling of 26 genes within the 400-kb regions of three significant SNPs was conducted using gene expression chips (SBC human 4×180k lncRNA Microarray v6, ID: 074348, Agilent) in Shanghai Biotechnology Corporation according to the manufacturer's instructions. The RNA samples were extracted from PBMCs of TB patients and healthy controls and then amplified, labeled, hybridized and scanned according to the manufacturer's protocols.

Data were extracted using Feature Extraction software 10.7 (Agilent Technologies). Raw data were normalized using the quantile algorithm in the GeneSpring Software 11.0 (Agilent Technologies). Criteria used to identify genes significantly expressed genes were: fold change (FC) (linear) ≤0.67 (downregulation) or ≥1.5 (upregulation) and student's t-test *P*-value <0.05.

RT-PCR verification

After RNA extraction, 500 ng of the RNA was reverse transcribed into cDNA using ReverTra Ace qPCR RT Kit (TOYOBO, Osaka, Japan), and then quantitative real-time PCR was carried out in an7900 HT Sequence Detection System (ABI, Massachusetts, USA) using the ABI Power SYBR Green PCR Master Mix (ABI, Massachusetts, USA)in accordance with the manufacturer's instructions. The thermal cycling conditions were: 2 min at 95 °C for initial denaturation, followed by 40 cycles of 15 s at 95 °C, 60 s at 60 °C for amplification, and 15 s at 95 °C, 15 s at 60 °C and 15 s at 95 °C for melting curve analysis. The primers of *CLCN6*, *DOK7*, *KIAA2013*, *TNFRSF8*, *MFN2* and *GAPDH* were listed in Supplementary Material, Table S7.

eQTL analysis

We queried all genes within the flanking regions of the three significant loci to GTEx portal (https://gtexportal.org/home/) and obtained all cis-eQTLs of them (66). We just kept those cis-eQTLs as potential candidates if their GWAS *P*-value less than 1×10^{-4} in stage 1 of our study (Supplementary Material, Table S4).

Gene set enrichment analysis

The pathway enrichment analysis was analysed by DAVID v6.7 (https://david.ncifcrf.gov/) (32). In the functional annotation analysis, modified Fisher's exact test was used to determine the significance of gene-term enrichment with a cutoff value at $P=0.05$. In the clustering of functional annotations, the Enrichment Score (ES) was used to rank the overall enrichment of the annotation groups. ES is a modified Fisher Exact *P*-value. When members of two independent groups can fall into one of two mutually exclusive categories, Fisher Exact test is used to determine whether the proportions of those falling into each category differ by group. In DAVID annotation system, Fisher Exact is adopted to measure the gene-enrichment in annotation terms. The value is defined as minus log transformation on the average *P*-values of each annotation term and was set at 1.3 (non-log scale 0.05) for significance. Additionally, a classification stringency parameter was used in the functional annotation clustering to control the fuzzy clustering of DAVID, and we used the high stringency for tight, clean and smaller numbers of clusters.

Supplementary Material

Supplementary Material is available at *HMG* online.

Acknowledgements

The GWAS data on TB in the Gambian population were kindly provided by Professor A.V.S. Hill, Dr. V. Naranbhai (Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford,

UK), and the International Tuberculosis Host Genetics Consortium. The Indonesian TB GWAS data were kindly provided by Indonesian TB GWAS team (Png E, Alisjahbana B, Sahiratmadja E, Marzuki S, Nelwan R, Adnan I, van de Vosse E, Hibberd ML, van Crevel R, Ottenhoff TH, Seielstad M) .The writing of manuscript was kindly reviewed and edited by Joy Fleming associate professor (Institute of Biophysics, Chinese Academy of Sciences) and Liwen Bianji, Edanz Group China (www.liwenbianji.cn).

Conflict of Interest statement. None declared.

Funding

National Natural Science Foundation of China (81571950 to A-D.S. and 31371347 to Y-B.Z.), Beijing Natural Science Foundation (7121007 to A-D.S.) and Beijing Nova Program Interdisciplinary Studies Cooperative Projects (xxjc201611 to A-D.S.). The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

1. World Health Organization. Global tuberculosis report 2016. Geneva 27, Switzerland (2016).
2. Arend, S.M., Engelhard, A.C., Groot, G., de Boer, K., Andersen, P., Ottenhoff, T.H. and van Dissel, J.T. (2001) Tuberculin skin testing compared with T-cell responses to Mycobacterium tuberculosis-specific and nonspecific antigens for detection of latent infection in persons with recent tuberculosis contact. *Clin. Diagn. Lab. Immun.*, **8**, 1089–1096.
3. Mathema, B., Kurepina, N.E., Bifani, P.J. and Kreiswirth, B.N. (2006) Molecular epidemiology of tuberculosis: current insights. *Clin. Microbiol. Rev.*, **19**, 658–685.
4. Moller, M. and Hoal, E.G. (2010) Current findings, challenges and novel approaches in human genetic susceptibility to tuberculosis. *Tuberculosis*, **90**, 71–83.
5. Apt, A. and Kramnik, I. (2009) Man and mouse TB: contradictions and solutions. *Tuberculosis*, **89**, 195–198.
6. Thye, T., Vannberg, F.O., Wong, S.H., Owusu-Dabo, E., Osei, I., Gyapong, J., Sirugo, G., Sisay-Joof, F., Enimil, A., Chinbuah, M.A. *et al.* (2010) Genome-wide association analyses identifies a susceptibility locus for tuberculosis on chromosome 18q11.2. *Nat. Genet.*, **42**, 739–741.
7. Thye, T., Owusu-Dabo, E., Vannberg, F.O., van Crevel, R., Curtis, J., Sahiratmadja, E., Balabanova, Y., Ehmen, C., Muntau, B., Ruge, G. *et al.* (2012) Common variants at 11p13 are associated with susceptibility to tuberculosis. *Nat. Genet.*, **44**, 257–259.
8. Sobota, R.S., Stein, C.M., Kodaman, N., Scheinfeldt, L.B., Maro, I., Wieland-Alter, W., Igo, R.J., Magohe, A., Malone, L.L., Chervenak, K. *et al.* (2016) A locus at 5q33.3 confers resistance to tuberculosis in highly susceptible individuals. *Am. J. Hum. Genet.*, **98**, 514–524.
9. Curtis, J., Luo, Y., Zenner, H.L., Cuchet-Lourenco, D., Wu, C., Lo, K., Maes, M., Alisaac, A., Stebbings, E., Liu, J.Z. *et al.* (2015) Susceptibility to tuberculosis is associated with variants in the ASAP1 gene encoding a regulator of dendritic cell migration. *Nat. Genet.*, **47**, 523–527.
10. Sveinbjornsson, G., Gudbjartsson, D.F., Halldorsson, B.V., Kristinsson, K.G., Gottfredsson, M., Barrett, J.C., Gudmundsson, L.J., Blondal, K., Gylfason, A., Gudjonsson, S.A. *et al.* (2016) HLA class II sequence variants influence tuberculosis risk in populations of European ancestry. *Nat. Genet.*, **48**, 318–322.
11. Mahasirimongkol, S., Yanai, H., Mushiroda, T., Promphittayarat, W., Wattanapokayakit, S., Phromjai, J., Yuliwulandari, R., Wichukchinda, N., Yowang, A., Yamada, N. *et al.* (2012) Genome-wide association studies of tuberculosis in Asians identify distinct at-risk locus for young tuberculosis. *J. Hum. Genet.*, **57**, 363–367.
12. Png, E., Alisjahbana, B., Sahiratmadja, E., Marzuki, S., Nelwan, R., Balabanova, Y., Nikolayevskyy, V., Drobniewski, F., Nejentsev, S., Adnan, I. *et al.* (2012) A genome wide association study of pulmonary tuberculosis susceptibility in Indonesians. *BMC. Med. Genet.*, **13**, 5.
13. Gao, L., Lu, W., Bai, L., Wang, X., Xu, J., Catanzaro, A., Cardenas, V., Li, X., Yang, Y., Du, J. *et al.* (2015) Latent tuberculosis infection in rural China: baseline results of a population-based, multicentre, prospective cohort study. *Lancet. Infect. Dis.*, **15**, 310–319.
14. Hinds, D.A., McMahon, G., Kiefer, A.K., Do, C.B., Eriksson, N., Evans, D.M., St, P.B., Ring, S.M., Mountain, J.L., Francke, U. *et al.* (2013) A genome-wide association meta-analysis of self-reported allergy identifies shared and allergy-specific susceptibility loci. *Nat. Genet.*, **45**, 907–911.
15. Anttila, V., Winsvold, B.S., Gormley, P., Kurth, T., Bettella, F., McMahon, G., Kallela, M., Malik, R., de Vries, B., Terwindt, G. *et al.* (2013) Genome-wide meta-analysis identifies new susceptibility loci for migraine. *Nat. Genet.*, **45**, 912–917.
16. Voight, B.F., Scott, L.J., Steinthorsdottir, V., Morris, A.P., Dina, C., Welch, R.P., Zeggini, E., Huth, C., Aulchenko, Y.S., Thorleifsson, G. *et al.* (2010) Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. *Nat. Genet.*, **42**, 579–589.
17. Auton, A., Abecasis, G.R., Altshuler, D.M., Durbin, R.M., Abecasis, G.R., Bentley, D.R., Chakravarti, A., Clark, A.G., Donnelly, P., Eichler, E.E. *et al.* (2015) A global reference for human genetic variation. *Nature*, **526**, 68–74.
18. Britton, W.J. and Lockwood, D.N. (2004) Leprosy. *Lancet*, **363**, 1209–1219.
19. Liu, H., Irwanto, A., Fu, X., Yu, G., Yu, Y., Sun, Y., Wang, C., Wang, Z., Okada, Y., Low, H. *et al.* (2015) Discovery of six new susceptibility loci and analysis of pleiotropic effects in leprosy. *Nat. Genet.*, **47**, 267–271.
20. Wong, S.H., Gochhait, S., Malhotra, D., Pettersson, F.H., Teo, Y.Y., Khor, C.C., Rautanen, A., Chapman, S.J., Mills, T.C., Srivastava, A. *et al.* (2010) Leprosy and the adaptation of human toll-like receptor 1. *PLOS. Pathog.*, **6**, e1000979.
21. Zhang, F.R., Huang, W., Chen, S.M., Sun, L.D., Liu, H., Li, Y., Cui, Y., Yan, X.X., Yang, H.T., Yang, R.D. *et al.* (2009) Genomewide association study of leprosy. *N. Engl. J. Med.*, **361**, 2609–2618.
22. Zhang, F., Liu, H., Chen, S., Low, H., Sun, L., Cui, Y., Chu, T., Li, Y., Fu, X., Yu, Y. *et al.* (2011) Identification of two new loci at IL23R and RAB32 that influence susceptibility to leprosy. *Nat. Genet.*, **43**, 1247–1251.
23. Liu, H., Bao, F., Irwanto, A., Fu, X., Lu, N., Yu, G., Yu, Y., Sun, Y., Low, H., Li, Y. *et al.* (2013) An association study of TOLL and CARD with leprosy susceptibility in Chinese population. *Hum. Mol. Genet.*, **22**, 4430–4437.
24. Wang, Z., Sun, Y., Fu, X., Yu, G., Wang, C., Bao, F., Yue, Z., Li, J., Sun, L., Irwanto, A. *et al.* (2016) A large-scale genome-wide association and meta-analysis identified four novel susceptibility loci for leprosy. *Nat. Commun.*, **7**, 13760.
25. Liu, H., Irwanto, A., Tian, H., Fu, X., Yu, Y., Yu, G., Low, H., Chu, T., Li, Y., Shi, B. *et al.* (2012) Identification of IL18RAP/IL18R1 and IL12B as leprosy risk genes demonstrates shared pathogenesis between inflammation and infectious diseases. *Am. J. Hum. Genet.*, **91**, 935–941.

26. Ng, S.B., Turner, E.H., Robertson, P.D., Flygare, S.D., Bigham, A.W., Lee, C., Shaffer, T., Wong, M., Bhattacharjee, A., Eichler, E.E. *et al.* (2009) Targeted capture and massively parallel sequencing of 12 human exomes. *Nature*, **461**, 272–276.
27. Ward, L.D. and Kellis, M. (2012) HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res.*, **40**, D930–D934.
28. Ardlie, K.G., Deluca, D.S., Segre, A.V., Sullivan, T.J., Young, T.R., Gelfand, E.T., Trowbridge, C.A., Maller, J.B., Tukiainen, T., Lek, M. *et al.* (2015) Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science*, **348**, 648–660.
29. Eicher, J.D., Landowski, C., Stackhouse, B., Sloan, A., Chen, W., Jensen, N., Lien, J.P., Leslie, R. and Johnson, A.D. (2015) GRASP v2.0: an update on the genome-wide repository of associations between SNPs and phenotypes. *Nucleic Acids Res.*, **43**, D799–D804.
30. Bernstein, B.E., Stamatoyannopoulos, J.A., Costello, J.F., Ren, B., Milosavljevic, A., Meissner, A., Kellis, M., Marra, M.A., Beaudet, A.L., Ecker, J.R. *et al.* (2010) The NIH Roadmap Epigenomics Mapping Consortium. *Nat. Biotechnol.*, **28**, 1045–1048.
31. ENCODE Project Consortium. (2004) The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science*, **306**, 636–640.
32. Huang, D.W., Sherman, B.T. and Lempicki, R.A. (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.*, **4**, 44–57.
33. Chen, C., Zhao, Q., Hu, Y., Shao, Y., Li, G., Zhu, L., Lu, W. and Xu, B. (2016) A rare variant at 11p13 is associated with tuberculosis susceptibility in the Han Chinese population. *Sci. Rep.*, **6**, 24016.
34. Gieger, C., Radhakrishnan, A., Cvejic, A., Tang, W., Porcu, E., Pistis, G., Serbanovic-Canic, J., Elling, U., Goodall, A.H., Labrune, Y. *et al.* (2011) New gene functions in megakaryopoiesis and platelet formation. *Nature*, **480**, 201–208.
35. Feng, Y., Dorhoi, A., Mollenkopf, H.J., Yin, H., Dong, Z., Mao, L., Zhou, J., Bi, A., Weber, S., Maertzdorf, J. *et al.* (2014) Platelets direct monocyte differentiation into epithelioid-like multinucleated giant foam cells with suppressive capacity upon mycobacterial stimulation. *J. Infect. Dis.*, **210**, 1700–1710.
36. Tang, C., Yamada, H., Shibata, K., Muta, H., Wajjwalku, W., Podack, E.R. and Yoshikai, Y. (2008) A novel role of CD30L/CD30 signaling by T-T cell interaction in Th1 response against mycobacterial infection. *J. Immunol.*, **181**, 6316–6327.
37. Fava, V.M., Cobat, A., Van Thuc, N., Latini, A.C., Stefani, M.M., Belone, A.F., Ba, N.N., Orlova, M., Manry, J., Mira, M.T. *et al.* (2015) Association of TNFSF8 regulatory variants with excessive inflammatory responses but not leprosy per se. *J. Infect. Dis.*, **211**, 968–977.
38. Fava, V.M., Sales-Marques, C., Alcais, A., Moraes, M.O. and Schurr, E. (2017) Age-dependent association of TNFSF15/TNFSF8 variants and leprosy type 1 reaction. *front Immunol.*, **8**, 155.
39. Lindestam, A.C., Lewinsohn, D., Sette, A. and Lewinsohn, D. (2014) Antigens for CD4 and CD8 T cells in tuberculosis. *Cold. Spring. Harb. Perspect. Med.*, **4**, a018465.
40. Souza, D.L.D., Morishi, O.M., Porto, D.S.M., de Melo, S.C., Alves, D.A.V., Assumpcao, A.I., Boechat, A.L., Ramasawmy, R. and Sadahiro, A. (2016) Alleles of HLA-DRB1*04 associated with pulmonary tuberculosis in Amazon Brazilian Population. *Plos One*, **11**, e0147543.
41. Lombard, Z., Dalton, D.L., Venter, P.A., Williams, R.C. and Bornman, L. (2006) Association of HLA-DR, -DQ, and vitamin D receptor alleles and haplotypes with tuberculosis in the Venda of South Africa. *Hum. Immunol.*, **67**, 643–654.
42. Moratz, C., Harrison, K. and Kehrl, J.H. (2004) Regulation of chemokine-induced lymphocyte migration by RGS proteins. *Methods. Enzymol.*, **389**, 15–32.
43. Hussain, T. (2007) Leprosy and tuberculosis: an insight-review. *Crit. Rev. Microbiol.*, **33**, 15–66.
44. Scollard, D.M., Dacso, M.M. and Abad-Venida, M.L. (2015) Tuberculosis and leprosy: classical granulomatous diseases in the twenty-first century. *Dermatol. Clin.*, **33**, 541–562.
45. Rawson, T.M., Anjum, V., Hodgson, J., Rao, A.K., Murthy, K., Rao, P.S., Subbanna, J. and Rao, P.V. (2014) Leprosy and tuberculosis concomitant infection: a poorly understood, age-old relationship. *Lepr. Rev.*, **85**, 288–295.
46. Tanaka, M., Shimamura, S., Kuriyama, S., Maeda, D., Goto, A. and Aiba, N. (2016) SKAP2 promotes podosome formation to facilitate tumor-associated macrophage infiltration and metastatic progression. *Cancer. Res.*, **76**, 358–369.
47. Dubovsky, J.A., Chappell, D.L., Harrington, B.K., Agrawal, K., Andritsos, L.A., Flynn, J.M., Jones, J.A., Paulaitis, M.E., Bolon, B., Johnson, A.J. *et al.* (2013) Lymphocyte cytosolic protein 1 is a chronic lymphocytic leukemia membrane-associated antigen critical to niche homing. *Blood*, **122**, 3308–3316.
48. Matza, D., Wolstein, O., Dikstein, R. and Shachar, I. (2001) Invariant chain induces B cell maturation by activating a TAF(II)105-NF-kappaB-dependent transcription program. *J. Biol. Chem.*, **276**, 27203–27206.
49. World Health Organization. (2015) Treatment of Tuberculosis: Guidelines. Geneva 27, Switzerland.
50. Wu, X.R., Yin, Q.Q., Jiao, A.X., Xu, B.P., Sun, L., Jiao, W.W., Xiao, J., Miao, Q., Shen, C., Liu, F. *et al.* (2012) Pediatric tuberculosis at Beijing Children's Hospital: 2002-2010. *Pediatrics*, **130**, e1433–e1440.
51. China, M.O.H.O. (2008) *National TB Control Program (NTP) Guidelines in China.*, Beijing, China: Ministry of Health, China.
52. Sreeramareddy, C.T., Panduru, K.V., Verma, S.C., Joshi, H.S. and Bates, M.N. (2008) Comparison of pulmonary and extrapulmonary tuberculosis in Nepal- a hospital-based retrospective study. *Bmc. Infect. Dis.*, **8**, 8.
53. Qi, H., Sun, L., Jin, Y.Q., Shen, C., Chu, P., Wang, S.F., Yin, Q.Q., Qi, Z., Xu, F., Jiao, W.W. *et al.* (2014) rs2243268 and rs2243274 of Interleukin-4 (IL-4) gene are associated with reduced risk for extrapulmonary and severe tuberculosis in Chinese Han children. *Infect. Genet. Evol.*, **23**, 121–128.
54. Xu, S., Yin, X., Li, S., Jin, W., Lou, H., Yang, L., Gong, X., Wang, H., Shen, Y., Pan, X. *et al.* (2009) Genomic dissection of population substructure of Han Chinese and its implication in association studies. *Am. J. Hum. Genet.*, **85**, 762–774.
55. Skol, A.D., Scott, L.J., Abecasis, G.R. and Boehnke, M. (2006) Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. *Nat. Genet.*, **38**, 209–213.
56. Patterson, N., Price, A.L. and Reich, D. (2006) Population structure and eigenanalysis. *Plos. Genet.*, **2**, e190.
57. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J. *et al.* (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.*, **81**, 559–575.
58. Howie, B.N., Donnelly, P., Marchini, J. and Schork, N.J. (2009) A flexible and accurate genotype imputation method for the

next generation of genome-wide association studies. *PLOS. Genet.*, **5**, e1000529.
59. Turner, S.D. (2014) qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. *Biorxiv*.
60. Pruim, R.J., Welch, R.P., Sanna, S., Teslovich, T.M., Chines, P.S., Gliedt, T.P., Boehnke, M., Abecasis, G.R. and Willer, C.J. (2010) LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics*, **26**, 2336–2337.
61. Willer, C.J., Li, Y. and Abecasis, G.R. (2010) METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*, **26**, 2190–2191.
62. Pillai, N.E., Okada, Y., Saw, W.-Y., Ong, R.T.-H., Wang, X., Tantoso, E., Xu, W., Peterson, T.A., Bielawny, T., Ali, M. *et al.* (2014) Predicting HLA alleles from high-resolution SNP data in three Southeast Asian populations. *Hum. Mol. Genet.*, **23**, 4443–4451.
63. Wu, M.C., Lee, S., Cai, T., Li, Y., Boehnke, M. and Lin, X. (2011) Rare-variant association testing for sequencing data with the sequence kernel association test. *Am. J. Hum. Genet.*, **89**, 82–93.
64. Ramensky, V., Bork, P. and Sunyaev, S. (2002) Human non-synonymous SNPs: server and survey. *Nucleic. Acids. Res.*, **30**, 3894–3900.
65. Yang, J., Lee, S.H., Goddard, M.E. and Visscher, P.M. (2011) GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.*, **88**, 76–82.
66. GTEx Consortium. (2013) The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.*, **45**, 580–585.